ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΟΔΥΣΣΕΑ Κ. ΚΟΥΒΕΛΗ

«Εκπαίδευση Βαθιών Νευρωνικών Δικτύων Στη Διεργασία Ανίχνευσης Εισβολών»

ΣΤΗΝ ΕΠΙΣΤΗΜΟΝΙΚΗ ΠΕΡΙΟΧΗ:
ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ- ΔΙΚΤΥΑ- ΠΛΗΡΟΦΟΡΙΚΗ

ΕΠΙΒΛΕΠΩΝ:

ΙΑΚΩΒΟΣ Σ. ΒΕΝΙΕΡΗΣ, ΚΑΘΗΓΗΤΗΣ, ΣΗΜΜΥ

ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ

Ιάκωβος Βενιέρης, καθηγητής, ΣΗΜΜΥ

Δημητρα-Θεοδώρα Κακλαμάνη, καθηγήτρια, ΣΗΜΜΥ

Αντώνιος Συμβώνης, καθηγητής ΣΕΜΦΕ

ΑΘΗΝΑ, ΙΟΥΛΙΟΣ-2024

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω τον καθηγητή Ι. Βενιέρη για την ευκαιρία που μου έδωσε με αυτήν τη διπλωματική εργασία.

Επίσης, θα ήθελα να εκφράσω την ευγνωμοσύνη μου προς τον υποψήφιο διδάκτορα Ι. Πανόπουλο. Η καθοδήγηση και η υποστήριξή του υπήρξαν αναντικατάστατες καθ' όλη τη διάρκεια της προσπάθειας μου.

Τέλος, ευχαριστώ την καθηγήτρια Δ.Θ. Κακλαμάνη και τον καθηγητή Α. Συμβώνη για τις πολύτιμες συμβουλές τους.

.....

Οδυσσέας Κουβέλης

1 Ιουλίου, 2024

,

Πρόλογος

Η ασφάλεια των πληροφοριακών συστημάτων αναδύεται ως ζήτημα καίριας σημασίας στην εποχή της ψηφιακής τεχνολογίας. Καθώς οι διαδικτυακές απειλές αυξάνονται και εξελίσσονται, η ανάγκη αναβάθμισης πρακτικών κυβερνοασφάλειας γίνεται επιτακτική.

Παράλληλα, οι ραγδαίες εξελίξεις στην πληροφορική και ιδιαίτερα στον τομέα της τεχνητής νοημοσύνης (artificial intelligence - ai), προσφέρουν ένα ευρύ φάσμα μεθοδολογιών για την αυτοματοποίηση και τη διαχείριση μαζικών δεδομένων. Συγκεκριμένα, η τομή της ai και της πληροφορικής δικτύων (networking) αποτελεί έναν ταχέως εξελισσόμενο τεχνολογικό τομέα, παρουσιάζοντας πλήθος εφαρμογών ενίσχυσης της απόδοσης και της ασφάλειας των δικτύων. Επί παραδείγματι αναφέρονται εφαρμογές αυτοματοποίησης και διοίκησης δικτύων (network management & automation), βελτιστοποίησης της δικτυακής ροής (traffic optimization) και διαχείρισης της ποιότητας υπηρεσίας (quality of service - QoS). Αυτές οι καινοτομίες καθιστούν τη συνέργεια της τεχνητής νοημοσύνης και των δικτύων θεμελιώδη συνιστώσα των τεχνολογιών πληροφορικής και επικοινωνιών του μέλλοντος. Μέσω της χρήσης αλγορίθμων μάθησης, βελτιστοποιείται η διαχείριση πόρων και η αποδοτικότητα, ενώ νέες μέθοδοι μηχανικής και βαθιάς μάθησης συγκροτούνται για την αντιμετώπιση δικτυακών προκλήσεων.

Η αυτοματοποιημένη διαδικασία ανίχνευσης και απόκρισης σε απειλές αποτελεί μία σημαντική εφαρμογή της AI στον τομέα της ασφάλειας δικτύων. Ως εκ τούτου, η παρούσα διπλωματική εργασία εξετάζει την ανάπτυξη ενός Συστήματος Ανίχνευσης Εισβολών (Intrusion Detection System) με τη χρήση τεχνικών βαθιάς μάθησης (deep learning), με στόχο την ανίχνευση και την πρόληψη κυβερνοεπιθέσεων μέσα σε ένα δίκτυο υπολογιστών. Συγκεκριμένα, το σύστημα αυτό θα βασίζεται σε μεθόδους επιβλεπόμενης μάθησης (supervised learning) και θα εκτελεί τη διεργασία της ταξινόμησης δικτυακής κίνησης. Με άλλα λόγια, το αντικείμενο της εργασίας είναι η ανάπτυξη ενός μηχανισμού ο οποίος δέχεται δεδομένα δικτυακής κίνησης ως είσοδο και αποφαίνεται εάν η κίνηση προέρχεται από καλοήθη (benign) δραστηριότητα ή αντιστοιχεί σε κακόβουλη (malicious) δραστηριότητα.

Περίληψη

Η παρούσα μελέτη εξετάζει αρχιτεκτονικές βαθιάς μάθησης για την ανίχνευση εισβολών στον τομέα της κυβερνοασφάλειας. Ο στόχος είναι η ανάπτυξη ενός μηχανισμού ταξινόμησης της δικτυακής κίνησης, ο οποίος, λαμβάνοντας δεδομένα δικτυακής ροής, θα αποφασίζει αν η ροή είναι καλοήθης (benign) ή κακόβουλη (malicious). Επίσης, αναπτύσσονται παραλλαγές για τον ακριβή προσδιορισμό του τύπου της κυβερνοαπειλής.

Εν γένει, η ανίχνευση εισβολών επιτυγχάνεται μέσω ταξινομητών. Οι υπό μελέτη ταξινομητές περιλαμβάνουν τρεις αρχιτεκτονικές βαθιάς μάθησης: Perceptrons Πολλαπλών Επιπέδων, Μετασχηματιστές, και Συνελικτικά Νευρωνικά Δίκτυα. Οι αρχιτεκτονικές αυτές παρουσιάζονται και αναλύονται ως προς τη δομή και τη λειτουργικότητά τους και υλοποιούνται είτε σε δυαδική είτε σε πολυκατηγορική μορφή, προσαρμοσμένες κατάλληλα για την ταξινόμηση. Παράλληλα, τα δεδομένα προεπεξεργάζονται με διάφορους τρόπους που δεν είναι κοινοί για όλες τις αρχιτεκτονικές. Περιλαμβάνονται τεχνικές όπως η SMOTE, επιλογή χαρακτηριστικών με τον αλγόριθμο XGBoost, και η κλιμάκωση δεδομένων με τον Quantile Transformer. Με αυτόν τον τρόπο, δημιουργείται η έννοια του αγωγού (pipeline) που συνδυάζει αρχιτεκτονικές, βήματα επεξεργασίας και ρυθμίσεις εκπαίδευσης, καθορίζοντας τη μορφολογία κάθε μοντέλου ταξινόμησης. Συνολικά, κατασκευάζονται έξι αγωγοί, και για κάθε έναν από αυτούς αναπτύσσονται πέντε μοντέλα αυξανόμενου μεγέθους.

Για την εκπαίδευση των μοντέλων χρησιμοποιείται το δημόσια προσβάσιμο σύνολο δεδομένων CIC-IDS-2017. Καθώς το σύνολο αυτό φέρει ετικέτες, οι ταξινομητές εκπαιδεύονται υπό επίβλεψη. Το σύνολο δεδομένων περιέχει εκατομμύρια δείγματα δικτυακής κίνησης, με το καθένα να περιλαμβάνει δεκάδες χαρακτηριστικά. Μέσω της εκπαίδευσης, τα μοντέλα εντοπίζουν και εσωτερικεύουν υποκείμενες σχέσεις και μοτίβα στα δεδομένα. Στη συνέχεια, πραγματοποιείται αξιολόγηση των μοντέλων, με χρήση πλήθους μετρικών αξιολόγησης της απόδοσης.

Όλοι οι ταξινομητές επιτυγχάνουν κορυφαίες επιδόσεις (State of the Art). Παράλληλα, εξετάζεται το αντιστάθμισμα μεταξύ ακρίβειας ταξινόμησης και κατανάλωσης μνήμης, όσον αφορά τον αριθμό παραμέτρων. Η εργασία ολοκληρώνεται με τη σύγκριση των μοντέλων και την εξαγωγή συμπερασμάτων.

Λέξεις Κλειδιά

Ασφάλεια Πληροφοριακών Συστημάτων, Ανίχνευση Εισβολών, Βαθιά Μάθηση, Επιβλεπόμενη Μάθηση, Perceptrons Πολλών Επιπέδων, Συνελικτικά Νευρωνικά Δίκτυα, Μετασχηματιστές, Κυβερνοασφάλεια, CIC-IDS-2017.

Abstract

This study examines deep learning architectures for intrusion detection in the field of cybersecurity. The objective is to develop a mechanism for classifying network traffic, which, by receiving network flow data, will determine whether the flow is benign or malicious. Additionally, variations are developed for the precise identification of the type of cyber threat.

Intrusion detection is generally achieved through classifiers. The classifiers under study include three deep learning architectures: Multi-Layer Perceptrons, Transformers, and Convolutional Neural Networks. These architectures are presented and analyzed in terms of their structure and functionality and implemented in either binary or multi-class form, suitably adapted for classification. At the same time, the data is preprocessed in various ways that are not common to all architectures. Techniques such as SMOTE, feature selection with the XGBoost algorithm, and data scaling with the Quantile Transformer are included. This approach creates the concept of a pipeline, combining architectures, processing steps, and training settings, determining the morphology of each classifier model. In total, six pipelines are constructed, and for each, five models of increasing size are developed.

For the model training, the publicly accessible CIC-IDS-2017 dataset is used. Since this dataset is labeled, the classifiers are trained under supervision. The dataset contains millions of network traffic samples, each with dozens of features. Through training, the models detect and internalize underlying relationships and patterns in the data. Subsequently, the models are evaluated using a plethora of performance evaluation metrics.

All classifiers achieve state-of-the-art performance. Additionally, the trade-off between classification accuracy and memory consumption, in terms of the number of parameters, is examined. The study concludes with the comparison of the models and the derivation of conclusions.

Keywords

Information Systems Security, Intrusion Detection, Deep Learning, Supervised Learning, Multilayer Perceptrons, Convolutional Neural Networks, Transformers, Cybersecurity, CIC-IDS-2017

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Λίστα Αλγορίθμων

Κεφάλαιο 1

Εισαγωγή

1.1 Ασφάλεια Δικτύων

Η πληροφορική δικτύων (networking) αποτελεί έναν από τους θεμέλιους λίθους της σύγχρονης πληροφορικής, επιτρέποντας τη διασύνδεση υπολογιστικών συστημάτων και την ανταλλαγή δεδομένων. Βασιζόμενη σε διάφορα πρωτόκολλα και τεχνολογίες εξασφαλίζει την αποτελεσματική και αξιόπιστη μεταφορά πληροφοριών σε τοπικά και ευρύτερα δίκτυα. Η ανάπτυξη και διαχείριση δικτύων απαιτεί την κατανόηση των αρχών δρομολόγησης, μεταγωγής και διαχείρισης κυκλοφορίας για τη διασφάλιση της απόδοσης και της διαθεσιμότητας των υπηρεσιών.

Η ασφάλεια δικτύων αποτελεί ζήτημα καίριας σημασίας και στοχεύει στην προστασία των δικτυακών υποδομών και των υπό μετάδοση δεδομένων από κακόβουλες ενέργειες και απειλές. Περιλαμβάνει την εφαρμογή πολιτικών ασφαλείας, τη χρήση τειχών προστασίας, συστημάτων ανίχνευσης εισβολών και την κρυπτογράφηση της επικοινωνίας. Η ευρύτερη έννοια της κυβερνοασφάλειας (cybersecurity) αναφέρεται στην αντιμετώπιση συνεχώς εξελισσόμενων απειλών και τη διασφάλιση της εμπιστευτικότητας, της ακεραιότητας και της διαθεσιμότητας των πληροφοριών.

Τα συστήματα ανίχνευσης εισβολών (intrusion detection system - IDS) αποτελούν εργαλεία ασφάλειας δικτύων. Τα συστήματα αυτά παρακολουθούν την κυκλοφορία του δικτύου και τα συμβάντα του συστήματος σε πραγματικό χρόνο, αναζητώντας ανωμαλίες και δραστηριότητες που αντιστοιχούν σε γνωστά μοτίβα επιθέσεων. Τα IDS χρησιμοποιούν τεχνικές ανάλυσης της δικτυακής ροής και ανίχνευσης ανωμαλιών για την αναγνώριση πιθανών εισβολών, επιτρέποντας την άμεση απόκριση και τον περιορισμό των κυβερνοαπειλών (cyberthreats).

1.2 Αυτοματοποιημένη Ανίχνευση

Βαθιά Μάθηση Η βαθιά μάθηση (deep learning), μια υποκατηγορία της μηχανικής μάθησης και της τεχνητής νοημοσύνης, εστιάζει στη χρήση πολυεπίπεδων νευρωνικών δικτύων και «μεγάλων» ή «βαθιών» μοντέλων πρόβλεψης για την ανάλυση και την εκμάθηση από σύνθετα σύνολα δεδομένων. Η πρακτική αυτή επιτρέπει την αυτόματη εκμάθηση αναπαραστάσεων των δεδομένων σε πολλαπλά επίπεδα αφαίρεσης τα οποία είναι ικανά να συλλάβουν πολύπλοκα μοτίβα και σχέσεις στα δεδομένα, καθιστώντας τη βαθιά μάθηση εξαιρετικά αποδοτική για μια ποικιλία εφαρμογών όπως η αναγνώριση εικόνας, η επεξεργασία φυσικής γλώσσας, η αυτόνομη οδήγηση, αλλά και η ανίχνευση εισβολών. Οι αρχιτεκτονικές, δηλαδή οι μορφολογίες των μοντέλων, εκπαιδεύονται μέσω μεγάλων συνόλων δεδομένων και αλγορίθμων βελτιστοποίησης, προκειμένου να επιτευχθεί υψηλή ακρίβεια και απόδοση στις διεργασίες πρόβλεψης.

Ταξινόμηση Η αυτοματοποιημένη ανίχνευση εισβολών ανάγεται σε μία διεργασία ταξινόμησης (classification). Ειδικότερα, κάθε δείγμα δικτυακής ροής θα πρέπει να ταξινομηθεί είτε ως καλοήθης είτε ως κακόβουλη δραστηριότητα. Η έννοια της εκπαίδευσης περιγράφει την εξαγωγή ενός προτύπου ταξινόμησης από τα δεδομένα, με χρήση επαναληπτικών μεθοδολογιών. Συγκεκριμένα, η παρούσα εργασία πραγματεύεται την έννοια της επιβλεπόμενης μάθησης (supervised learning), κατά την οποία τα δεδομένα φέρουν ετικέτες (labels). Οι ετικέτες υποδεικνύουν την κατηγορία στην οποία ανήκει κάθε στοιχείο του συνόλου δεδομένων, ενώ η ταξινόμηση αναφέρεται στην αυτοματοποιημένη πρόβλεψη της ετικέτας ενός στοιχείου. Διακρίνονται δύο διατάξεις (configurations) ταξινόμησης, ανάλογα με το πλήθος των κατηγοριών:

- **Δυαδική διάταξη** (binary configuration): Τα δεδομένα διακρίνονται σε δύο κλάσεις. Είθισται οι κατηγορίες αυτές να ονομάζονται κλάση 0 και κλάση 1 ή πλειοψηφούσα και μειοψηφούσα (majority & minority) κατηγορία. Στο πλαίσιο της εργασίας αυτής, η δυαδική διάταξη των δεδομένων αναφέρεται στην καλοήθη κλάση (benign class) και την κακόβουλη κλάση (malicious class).
- **Πολυκατηγορική διάταξη** (multiclass configuration): Τα δεδομένα διακρίνονται σε περισσότερες από δύο κλάσεις, στην πλειοψηφική και στις μειοψηφικές κατηγορίες. Στο πλαίσιο της ανίχνευσης εισβολών, η πολυκλασική διάταξη των δεδομένων αναφέρεται στην καλοήθη κλάση και σε συγκεκριμένους τύπους κυβερνοεπιθέσεων, π.χ. DoS attack, DDoS, Heartbleed, Brute Force κλπ.

1.3 NIDS vs. HIDS

Ορισμός Ο όρος «σύστημα ανίχνευσης εισβολών» αναφέρεται σε λογισμικό παρακολούθησης δικτυακών δραστηριοτήτων και πληροφοριών του συστήματος, αποσκοπώντας στον εντοπισμό ύποπτης ή κακόβουλης δραστηριότητας, η οποία μπορεί να αποτελεί ένδειξη εισβολής. Χρησιμοποιούνται τεχνικές όπως η ανάλυση δικτυακής ροής και η ανίχνευση ανωμαλιών για τη σύγκριση των δεδομένων με γνωστά πρότυπα επιθέσεων ή την παρατήρηση αποκλίσεων από την κανονική συμπεριφορά. Σκοπός του λογισμικού αυτού είναι η ενίσχυση της ασφάλειας των πληροφοριακών συστημάτων, επιτρέποντας την άμεση απόκριση σε πιθανές απειλές.

Τα συστήματα ανίχνευσης εισβολών διακρίνονται σε δύο κύριες κατηγορίες: τα συστήματα ανίχνευσης εισβολών βασισμένα σε δίκτυα (Network-based Intrusion Detection Systems - NIDS) και τα συστήματα ανίχνευσης εισβολών βασισμένα σε hosts (Host-based Intrusion Detection Systems - HIDS). Τα NIDSs παρακολουθούν και αναλύουν την κυκλοφορία δικτύου για ύποπτες δραστηριότητες και επιθέσεις, ενώ τα HIDS παρακολουθούν τη δραστηριότητα ενός μεμονωμένου συστήματος ή συσκευής (host). Η επιλογή μεταξύ NIDS και HIDS εξαρτάται από τις συγκεκριμένες ανάγκες και τον τύπο του περιβάλλοντος που πρέπει να προστατευθεί. Στην παρούσα εργασία επιλέγεται η ανάπτυξη ενός NIDS, το οποίο θα βασιστεί στο σύνολο δεδομένων CIC-IDS-2017 (Canadian Institute of Cybersecurity - Intrusion Detection System - 2017), όπως περιγράφεται στο κεφάλαιο 4. Το εν λόγω σύνολο δεδομένων είναι δημόσια προσβάσιμο και ειδικά σχεδιασμένο για τη διεργασία ανίχνευσης εισβολών.

1.4 Δομή της Εργασίας

Η παρούσα εργασία είναι οργανωμένη ως εξής: Στο εισαγωγικό κεφάλαιο παρουσιάζεται το πλαίσιο και η σημασία της ανίχνευσης εισβολών, ο στόχος της εργασίας, καθώς και βασικές έννοιες της τεχνητής νοημοσύνης και της βαθιάς μάθησης. Στη συνέχεια, στο κεφάλαιο 2 αναλύονται αρχιτεκτονικές μοντέλων βαθιάς μάθησης και συγκεκριμένα οι perceptrons πολλαπλών

επιπέδων, τα συνελικτικά νευρωνικά δίκτυα και οι μετασχηματιστές. Κατόπιν, στο κεφάλαιο 3 παρουσιάζονται διάφορες μετρικές αξιολόγησης των ταξινομητών και στο κεφάλαιο 4 περιγράφεται το σύνολο δεδομένων CIC-IDS-2017. Επακολουθούν τα βήματα προεπεξεργασίας στο κεφάλαιο 5, οι διαδικασίες υλοποίησης στο κεφάλαιο 6 και τέλος στο κεφάλαιο 7 παρουσιάζονται τα κύρια ευρήματα της έρευνας και τα συμπεράσματα που προέκυψαν, ενώ προτείνονται κατευθύνσεις για μελλοντική έρευνα και βελτιώσεις.

Με αυτήν τη δομή, η εργασία επιδιώκει να παρέχει μια ολοκληρωμένη εικόνα της διαδικασίας ανάπτυξης ενός συστήματος ανίχνευσης εισβολών με τη χρήση τεχνικών βαθιάς μάθησης και να αναδείξει την αποτελεσματικότητα και τις δυνατότητες αυτής της προσέγγισης.

Κεφάλαιο 2

Αρχιτεκτονική Εξερεύνηση

2.1 Perceptrons Πολλαπών Επιπέδων

Σε αυτήν την ενότητα, η θεωρία και οι βασικές έννοιες των perceptrons[1] *πολλαπλών επιπέδων παρουσιάζονται εκτενώς, αντλώντας κυρίως από το έργο του Charu C. Aggarwal [1, 2].*

2.1.1 Εισαγωγή

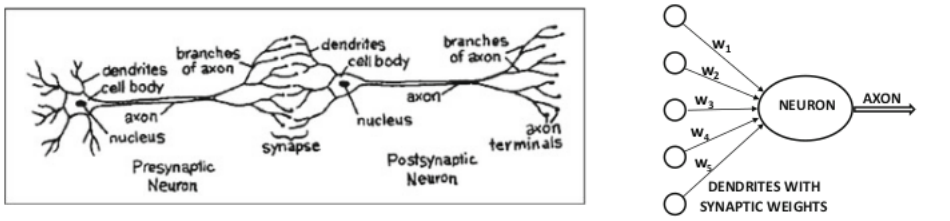
Όπως έξοχα γράφει ο Aggarwal στην εισαγωγή του βιβλίου του "Neural Networks and Deep Learning", [1]:

> "Τα τεχνητά νευρωνικά δίκτυα αποτελούν μία δημοφιλή τεχνική μηχανικής μάθησης, η οποία προσομοιώνει τον νευροφυσιολογικό μηχανισμό μάθησης των βιολογικών οργανισμών. Ειδικότερα, το ανθρώπινο νευρικό σύστημα περιέχει διάφορα κύτταρα, εκ των οποίων κυριαρχούν οι **νευρώνες** (neurons). Οι νευρώνες συνδέονται μεταξύ τους με **άξονες** (axons) και **δενδρίτες** (dendrites), και οι συνδετικές περιοχές μεταξύ αξόνων και δενδριτών αναφέρονται ως **συνάψεις** (synapses). [...] Η μορφολογία των συναπτικών απολήξεων συχνά μεταβάλλεται συναρτήσει εξωτερικών ερεθισμάτων. Η διαδικασία αυτή είναι ο τρόπος με τον οποίο λαμβάνει χώρα η νευροφυσιολογική μάθηση.
>
> Ο εν λόγω βιολογικός μηχανισμός προσομοιώνεται με τα τεχνητά νευρωνικά δίκτυα, τα οποία περιέχουν υπολογιστικές μονάδες, επονομαζόμενες ως νευρώνες (neurons). [...] Οι μονάδες αυτές συνδέονται μεταξύ τους μέσω βαρών, τα οποία εξομοιώνουν τις συναπτικές συνδέσεις των βιολογικών οργανισμών. Κάθε είσοδος σε έναν νευρώνα κλιμακώνεται από ένα βάρος, το οποίο συναποφασίζει την υπολογιστική λειτουργία της μονάδας. Επομένως, ένα τεχνητό νευρωνικό δίκτυο εκτιμά μια συνάρτηση των εισόδων διαδίδοντας αλλεπάλληλες τιμές υπολογισμών από τους νευρώνες εισόδου προς τους νευρώνες εξόδου, ενώ τα βάρη χρησιμοποιούνται ως ενδιάμεσες παράμετροι. Η διαδικασία της μάθησης συντελείται κατά τον επαναπροσδιορισμό των βαρών των νευρωνικών συνδέσεων.
>
> Κατά αναλογία με τη μάθηση των βιολογικών οργανισμών μέσω εξωτερικών ερεθισμάτων, τα αντίστοιχα ερεθίσματα στα τεχνητά νευρωνικά δίκτυα παρέχονται από

[1] Οι perceptrons αποτελούν την κύρια έννοια αυτού του κεφαλαίου, ωστόσο ο όρος αυτός θα παραμείνει αμετάφραστος. Το ζήτημα τίθεται κυρίως στη μετάφραση του όρου perceptron, ο οποίος εσφαλμένα αποδίδεται ως νευρώνας και συγχέεται με την έννοια neuron, η οποία θα οριστεί αναλυτικά αργότερα στο κεφάλαιο και θα εισαχθεί ως μαθηματική γενίκευση του perceptron

τα δεδομένα εκπαίδευσης σε ζεύγη δειγμάτων εισόδου-εξόδου της προς μάθηση συνάρτησης. Ως εκ τούτου, ο στόχος του επαναπροσδιορισμού των βαρών είναι η τροποποίηση της προς υπολογισμό συνάρτησης, ώστε να επιτευχθεί βελτίωση των προβλέψεων σε μελλοντικές επαναλήψεις. Επομένως, τα βάρη ανανεώνονται υπό μία μαθηματικώς διαρθρωμένη τροπολογία, αποσκοπώντας στη μείωση του σφάλματος απόκλισης. Κατά τη διαδοχική προσαρμογή των βαρών για πλήθος ζευγών εισόδου-εξόδου, η προσέγγιση της προς υπολογισμό συνάρτησης βελτιώνεται με την πάροδο του χρόνου, οδηγώντας σε ακριβέστερες προβλέψεις. Αυτή η ικανότητα ακριβούς προσδιορισμού αγνώστων συναρτήσεων, ύστερα από εκπαίδευση επί πεπερασμένου συνόλου ζευγών εισόδου-εξόδου αναφέρεται ως ικανότητα **γενίκευσης** του μοντέλου (model generalization). Συνεπακολούθως, η κύρια χρησιμότητα των μοντέλων μηχανικής μάθησης προκύπτει από την ικανότητα να γενικεύουν τη μάθησή τους από τα δεδομένα εκπαίδευσης σε άγνωστα παραδείγματα."

Σχήμα 2.1: Οι συναπτικές συνδέσεις μεταξύ των νευρώνων



(a) Biological neural network (b) Artificial neural network

Ένα βιολογικό-υπολογιστικό ανάλογο νευρωνικών δικτύων. Πηγή: [1].

Ιστορικό και Βιολογικό Υπόβαθρο Η έννοια των τεχνητών νευρωνικών δικτύων, κατάγεται από το πρωτοποριακό έργο των Warren McCulloch και Walter Pitts του 1943 [3]. Οι ίδιοι ανέπτυξαν ένα μαθηματικό μοντέλο για την περιγραφή της λειτουργίας των βιολογικών νευρώνων, θέτοντας τα θεμέλια για την επερχόμενη θεωρία των νευρωνικών δικτύων. Πρότειναν ένα απλοποιημένο μοντέλο νευρώνα, γνωστό και ως McCulloch-Pitts neuron, υπό το οποίο οι περίπλοκες λειτουργίες των πραγματικών νευρώνων προτυποποιούνται σε μία λογική μονάδα δυαδικού κατωφλίου, έννοια αναφερόμενη σε ηλεκτρικά κυκλώματα. Αυτό το αφαιρετικό μοντέλο κατέδειξε ότι τέτοια δίκτυα νευρώνων μπορούσαν, αρχικά να εκτελέσουν οποιονδήποτε λογικό ή αριθμητικό υπολογισμό, δεδομένου επαρκούς πλήθους νευρώνων και κατάλληλων συνδέσεων. Το έργο τους παρείχε το πρώτο θεωρητικό πλαίσιο για την κατανόηση του νευρωνικού υπολογισμού, επηρεάζοντας τις μετέπειτα εξελίξεις στην τεχνητή νοημοσύνη.

Η εξέλιξη των νευρωνικών δικτύων προήχθη αισθητά με τη δημιουργία του μηχανήματος Mark I Perceptron από τον Frank Rosenblatt το έτος 1958, [4], το οποίο κατασκευάστηκε στα πλαίσια υλοποίησης του σχετικού αλγορίθμου. Ο αλγόριθμος αυτός σχεδιάστηκε από τον Rosenblatt, για την αναγνώριση προτύπων και τη μάθηση από δεδομένα μέσω επαναληπτικών προσαρμογών των συναπτικών βαρών. Το μοντέλο perceptron αποτελείτο από ένα επίπεδο εισόδου συνδεδεμένο σε έναν νευρώνα εξόδου μέσω ρυθμιζόμενων βαρών, καταδεικνύοντας την ικανότητα των συστημάτων μάθησης να αναπροσαρμόζονται με βάση την εμπειρία. Παρά την αρχική του επιτυχία, το perceptron περιορίστηκε από την αδυναμία επίλυσης μη γραμμικώς διαχωρίσιμων προβλημάτων, όπως ευφυώς επισήμαναν οι Minsky και Papert το 1969, [5]. Ωστόσο, οι αρχές του perceptron έθεσαν τα θεμέλια για πιο πολύπλοκες αρχιτεκτονικές, οδηγώντας στην ανάπτυξη των perceptrons πολλαπλών επιπέδων.

2.1.2 Το Μοντέλο Perceptron

Σε αυτήν την ενότητα, εξετάζεται το μοντέλο **perceptron**, το οποίο αποτελεί τη θεμελιώδη συνιστώσα των perceptrons πολλαπλών επιπέδων (multi-layer). Εν γένει, οι perceptrons απεικονίζουν ένα σύνολο εισόδων σε μια έξοδο.

Ορισμός και Δομή Ο perceptron είναι το απλούστερο νευρωνικό δίκτυο, αποτελούμενο από ένα μόνο επίπεδο εισόδου και έναν κόμβο εξόδου. Στο πλαίσιο της δυαδικής ταξινόμησης, κάθε ζεύγος εκπαίδευσης αναπαριστάται ως $\left(\bar{X}, y\right)$, όπου το διάνυσμα $\bar{X} = [x_1, x_2, \ldots, x_d]$ περιέχει d μεταβλητές **χαρακτηριστικών** (features), ενώ $y \in \{-1, +1\}$ είναι η δυαδική **ετικέτα** κλάσης. Εν γένει, ο στόχος είναι η πρόβλεψη της ετικέτας κλάσης για τις περιπτώσεις όπου δεν είναι γνωστή. Το επίπεδο εισόδου του perceptron αναμεταδίδει τις d μεταβλητές χαρακτηριστικών $\bar{X}$ μέσω ακμών με βάρη $\bar{W} = [w_1, w_2, \ldots, w_d]$ προς έναν κόμβο εξόδου. Η γραμμική συνάρτηση $\bar{W} \cdot \bar{X}$ υπολογίζεται στο σημείο εξόδου και το πρόσημο αυτής της τιμής χρησιμοποιείται για την πρόβλεψη της μεταβλητής κλάσης $\hat{y}$. Μαθηματικά, η πρόβλεψη θα είναι:

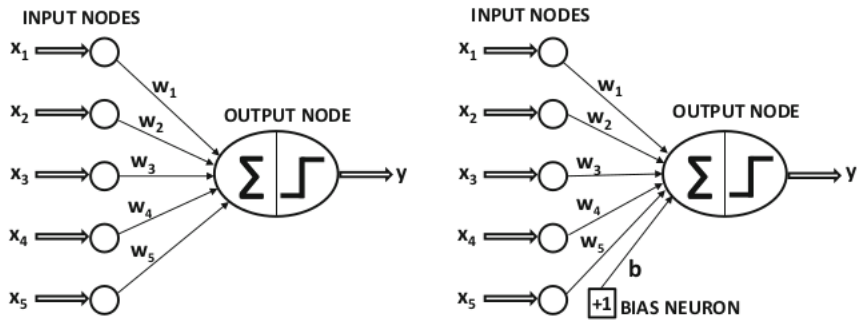
$$\hat{y} = \text{sign}\left(\bar{W} \cdot \bar{X}\right) = \text{sign}\left(\sum_{i=1}^{d} w_i \cdot x_i\right), \quad \text{sign}\left(y\right) = \begin{cases} +1, & y \geq 0 \\ -1, & y < 0 \end{cases} \tag{2.1}$$

Η συνάρτηση προσήμου $\text{sign}\left(\cdot\right)$ αντιστοιχεί[2] το εσωτερικό γινομένο είτε σε $+1$ είτε σε -1, τιμές κατάλληλες για τη διεργασία της δυαδικής ταξινόμησης. Το σχετικό σφάλμα πρόβλεψης ορίζεται ως:

$$\left[\mathbb{R}^d \ni \bar{X} \overset{\text{Error}(\cdot)}{\longrightarrow} \{-2, 0, 2\}\right], \quad \left[\text{Error}\left(\bar{X}\right) \overset{\text{def}}{=} y - \hat{y}\right] \tag{2.2}$$

Για μη μηδενικό σφάλμα, τα βάρη ενημερώνονται προς την κατεύθυνση της κλίσης του σφάλματος (error gradient). Εδώ, η συνάρτηση προσήμου λειτουργεί ως συνάρτηση ενεργοποίησης.

Σχήμα 2.2: Η βασική αρχιτεκτονική του Perceptron



(a) Perceptron without bias (b) Perceptron with bias

Οι κατευθυνόμενες ακμές (από την είσοδο προς την έξοδο) φέρουν τα βάρη $w_1, w_2, \ldots, w_d$ με τα οποία θα πολλαπλασιαστούν οι μεταβλητές χαρακτηριστικών, ενώ στη συνέχεια τα d γινόμενα προστίθενται στον κόμβο εξόδου. Πηγή: [1].

Η περίπτωση της μεροληψίας Σε περιπτώσεις όπου η δυαδική κατανομή των κλάσεων είναι ανισομερής, εισάγεται ένας όρος μεροληψίας b (bias) για την εξισορρόπηση της πρόβλεψης. Επακολούθως, ο τύπος της πρόβλεψης προσαρμόζεται ως εξής:

[2]Γενικά, η συνάρτηση προσήμου έχει τρεις εξόδους $-1, 0, +1$, όπου: $\text{sign}\left(0\right) = 0$. Όμως, στα πλαίσια δημιουργίας δυαδικών ετικετών, η συνάρτηση παραλλάσεται και οι μη αρνητικές τιμές συνενώνονται σε μία τιμή.

$$\hat{y} = \text{sign}\left(\bar{W} \cdot \bar{X} + b\right) = \text{sign}\left(\sum_{i=1}^{d} w_i \cdot x_i + b\right) \tag{2.3}$$

Η μεροληψία αντιμετωπίζεται ως το βάρος μιας επιπλέον ακμής προερχόμενης από έναν μεροληπτικό νευρώνα, ο οποίος πάντα θα μεταδίδει την τιμή 1 στον κόμβο εξόδου. Επομένως, αυτή η τροποποίηση δεν αλλάζει τον αλγόριθμο εκπαίδευσης, καθώς ο νευρώνας μεροληψίας αντιμετωπίζεται όπως οποιοσδήποτε άλλος νευρώνας με σταθερή τιμή ενεργοποίησης 1.

Ο αλγόριθμος Perceptron Ο αλγόριθμος perceptron είναι μια απλή αλλά ισχυρή μέθοδος εκπαίδευσης του ιδίου μοντέλου. Ρυθμίζει επαναληπτικά τα βάρη και τη μεταβλητή μεροληψίας, με στόχο την ελαχιστοποίηση του σφάλματος πρόβλεψης, δηλαδή τον αριθμό των εσφαλμένων ταξινομήσεων.

Αλγόριθμος 2.1 Ο αλγόριθμος Perceptron

Αρχικοποίηση: Το διάνυσμα βαρών $\bar{W}$ αρχικοποιείται με μηδενικές ή μικρές τιμές, ο όρος μεροληψίας b στο μηδέν και ο ρυθμός μάθησης α σε μια μικρή θετική τιμή.
Επανάληψη: Η διαδικασία επαναλαμβάνεται για έναν προκαθορισμένο αριθμό από εποχές (epochs) ή μέχρι την επίτευξη σύγκλισης:

1. **Πρόβλεψη**: Για το i-οστό δείγμα $\left(\bar{X}_i, y_i\right)$ υπολογίζεται η προβλεπόμενη έξοδος:

$$\hat{y}_i = \text{sign}\left(\bar{W} \cdot \bar{X}_i + b\right)$$

2. **Ενημέρωση**: Εφόσον η προβλεπόμενη ετικέτα $\hat{y}_i$ δεν αντιστοιχεί στην πραγματική ετικέτα y_i, τα βάρη και η μεροληψία ενημερώνονται:

$$\bar{W} \leftarrow \bar{W} + \alpha \cdot (y_i - \hat{y}_i) \cdot \bar{X}_i$$

$$b \leftarrow b + \alpha \cdot (y_i - \hat{y}_i)$$

Ο αλγόριθμος του perceptron στοχεύει στην ελαχιστοποίηση του σφάλματος πρόβλεψης, το οποίο μπορεί να εκφραστεί μέσω μιας συνάρτησης απώλειας. Συνεπώς, σκοπός είναι να ελαχιστοποιηθεί ο αριθμός των λανθασμένων ταξινομήσεων στο σύνολο εκπαίδευσης $\mathcal{D}$, το οποίο εμπεριέχει ζεύγη χαρακτηριστικών-ετικετών $\left(\bar{X}, y\right)$. Η συνάρτηση απώλειας για τον perceptron σε μορφή ελαχίστων τετραγώνων ορίζεται ως εξής:

$$\text{Minimize}_{\bar{W}} L = \frac{1}{2} \sum_{\left(\bar{X}, y\right) \in \mathcal{D}} (y - \hat{y})^2 = \frac{1}{2} \sum_{\left(\bar{X}, y\right) \in \mathcal{D}} \left(y - \text{sign}\left\{\bar{W} \cdot \bar{X} + b\right\}\right)^2$$

Ο αλγόριθμος του perceptron χρησιμοποιεί έμμεσα μια ομαλή προσέγγιση της κλίσης της συνάρτησης απώλειας, εκπεφρασμένη ως προς κάθε δείγμα εκπαίδευσης:

$$\nabla L_{\text{smooth}} = \sum_{\left(\bar{X}, y\right) \in \mathcal{D}} (y - \hat{y}) \cdot \bar{X}$$

Αυτή η προσέγγιση **καθόδου κλίσης** (gradient descent) ενημερώνει το διάνυσμα βαρών $\bar{W}$ για κάθε σημείο δεδομένων $\bar{X}_i$ ως εξής :

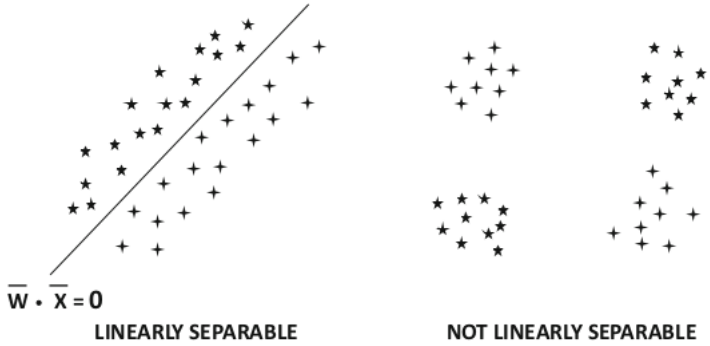
$$\bar{W} \leftarrow \bar{W} + \alpha \cdot (y - \hat{y}) \cdot \bar{X}_i \tag{2.4}$$

Ο βασικός αλγόριθμος του perceptron μπορεί να λάβει μορφή **στοχαστικής κλίσης καθόδου** (stochastic gradient descent), η οποία ελαχιστοποιεί το τετραγωνικό σφάλμα πρόβλεψης ενημερώνοντας τα βάρη βάσει τυχαία επιλεγμένων σημείων εκπαίδευσης. Χρησιμοποιώντας την εν λόγω μέθοδο, οι ενημερώσεις πραγματοποιούνται σε ένα υποσύνολο των σημείων εκπαίδευσης S (mini-batch):

$$\bar{W} \leftarrow \bar{W} + \alpha \cdot \sum_{\bar{X} \in S} \text{Error}\left(\bar{X}\right) \cdot \bar{X}$$

Περιορισμοί του Μοντέλου Perceptron Το γραμμικό μοντέλο perceptron ορίζει ένα **γραμμικό υπερεπίπεδο** (linear hyperplane), μέσω του εσωτερικού γινομένου $\bar{W} \cdot \bar{X}$. Εδώ, το $\bar{W} = [w_1, w_2, \ldots, w_d]$ είναι ένα d-διάστατο διάνυσμα, κάθετο στο υπερεπίπεδο. Εν γένει, το μοντέλο είναι αποτελεσματικό όταν αντιμετωπίζει **γραμμικώς διαχωρίσιμα** (linearly separable) δεδομένα. Πραγματοποιεί ταξινομήσεις καθορίζοντας αν το γινόμενο $\bar{W} \cdot \bar{X}$ είναι θετικό ή αρνητικό. Ωστόσο, η απόδοσή του υποβαθμίζεται σημαντικά σε μη γραμμικώς διαχωρίσιμα δεδομένα.

Σχήμα 2.3: Παραδείγματα δύο κλάσεων γραμμικώς διαχωρίσιμων και μη διαχωρίσιμων δεδομένων



Το σχήμα απεικονίζει την απόδοση του perceptron σε γραμμικώς διαχωρίσιμα και μη διαχωρίσιμα σύνολα δεδομένων, αναδεικνύοντας τους περιορισμούς του. Στα αριστερά, ένα γραμμικώς διαχωρίσιμο σύνολο δεδομένων ταξινομείται σωστά, ενώ στα δεξιά, ένα μη γραμμικώς διαχωρίσιμο σύνολο δεδομένων, γνωστό ως πρόβλημα XOR, καταδεικνύει την ανεπάρκεια του perceptron. Ο εν λόγω περιορισμός απαιτεί την υιοθέτηση πιο σύνθετων αρχιτεκτονικών νευρωνικών δικτύων για την αντιμετώπιση της εγγενούς μη γραμμικότητας στα δεδομένα. Πηγή: [1].

Ενώ ο αλγόριθμος του perceptron εγγυάται τη σύγκλιση υπό μηδενικό σφάλμα για γραμμικώς διαχωρίσιμα δεδομένα, αποτυγχάνει για την περίπτωση μη γραμμικώς διαχωρίσιμων δεδομένων. Σε τέτοιες περιπτώσεις, οι επαναλήψεις του αλγορίθμου μπορεί να οδηγήσουν σε ανεπαρκείς λύσεις, αυξάνοντας τον αριθμό των εσφαλμένων ταξινομήσεων. Αυτό το πρόβλημα επιδεινώνεται περαιτέρω από την ευαισθησία της συνάρτησης απώλειας στο μέγεθος του διανύσματος βαρών, αποσταθεροποιώντας το **σύνορο ταξινόμησης** (classification boundary).

Για να αντιμετωπιστούν αυτοί οι περιορισμοί, έχουν προταθεί διάφορες παραλλαγές του perceptron. Μέθοδοι όπως οι αλγόριθμοι Pocket και SVM προσφέρουν βελτιώσεις, όμως οι perceptrons πολλαπλών επιπέδων υπερβαίνουν ουσιαστικά τις προκλήσεις αυτές.

Ο Perceptron ως Θεμελιώδης Συνιστώσα των MLPs Ο perceptron, ως ένα νευρωνικό δίκτυο ενός επιπέδου, αποτελεί το θεμελιώδες δομικό στοιχείο των MLPs. Σε ένα MLP

δίκτυο, οι νευρώνες οργανώνονται σε διάφορα επίπεδα: ένα επίπεδο εισόδου, ένα ή περισσότερα κρυφά επίπεδα και ένα επίπεδο εξόδου. Κάθε νευρώνας σε αυτά τα επίπεδα λειτουργεί ως perceptron, επεξεργαζόμενος τις αντίστοιχες εισόδους σε εξόδους μέσω σταθμισμένων αθροισμάτων και συναρτήσεων ενεργοποίησης. Η εν λόγω αρχιτεκτονική ανά επίπεδο, γνωστή και ως **δίκτυο εμπρόσθιας τροφοδοσίας** (feed forward network), καταλύει την αποτελεσματική εξαγωγή και προτυποποίηση σύνθετων, μη γραμμικων σχέσεων στα δεδομένα.

Η κατανόηση των αρχών λειτουργίας των MLPs συνάγεται από τη λειτουργικότητα του perceptron. Ειδικότερα, η επαναληπτική διαδικασία εκπαίδευσης σε έναν perceptron επεκτείνεται και στα επίπεδα των MLPs, με όρους προσαρμογής βαρών και μεροληψίας. Με αυτόν τον τρόπο ενισχύεται η ικανότητά τους να μαθαίνουν από ποικίλα και περίπλοκα σύνολα δεδομένων. Επιπλέον, η εισαγωγή **μη γραμμικότητας** (non-linearity) μέσω κατάλληλων συναρτήσεων ενεργοποίησης επιτρέπει στα δίκτυα αυτά να συλλάβουν πιο σύνθετα μοτίβα. Έτσι, η απλή αλλά ισχυρή έννοια του perceptron θέτει τα θεμέλια για την ανάπτυξη και κατανόηση πιο εξελιγμένων αρχιτεκτονικών νευρωνικών δικτύων.

2.1.3 Συναρτήσεις Ενεργοποίησης

Η επιλογή συνάρτησης ενεργοποίησης είναι καθοριστική στον σχεδιασμό των νευρωνικών δικτύων. Για τον perceptron, η συνάρτηση ενεργοποίησης πρόσημου $\text{sign}(\cdot)$ χρησιμοποιείται για την πρόβλεψη ετικετών δυαδικών κλάσεων. Ωστόσο, για διαφορετικές μεταβλητές στόχου επιλέγονται άλλες συναρτήσεις ενεργοποίησης. Για παράδειγμα, είθισται η χρήση της ταυτοτικής συνάρτησης για προβλέψεις πραγματικών τιμών, καθιστώντας τη διαδικασία ισοδύναμη με τη γραμμική παλινδρόμηση ή η χρήση της σιγμοειδούς συνάρτησης για την πρόβλεψη δυαδικών κλάσεων, όπου η έξοδος $\hat{y}$ θα υποδηλώνει πιθανότητα.

Γενικεύοντας τον Perceptron: η Έννοια του Νευρώνα Οι μη γραμμικές συναρτήσεις ενεργοποίησης γίνονται απαραίτητες κατά τη μετάβαση από την έννοια του perceptron σε πολυεπίπεδες αρχιτεκτονικές. Διάφορες μη γραμμικές συναρτήσεις, όπως η σιγμοειδής και η υπερβολική εφαπτομένη, μπορούν να χρησιμοποιηθούν σε διαφορετικά επίπεδα. Πλέον, η γενική μορφή της εξόδου ενός **νευρώνα** (neuron) μπορεί να αναπαρασταθεί ως εξής:

$$\hat{y} = \Phi\left(\bar{W} \cdot \bar{X} + b\right) \tag{2.5}$$

όπου με $\Phi(\cdot)$ σημειώνεται η γενικευμένη συνάρτηση ενεργοποίησης. Σε αντίθεση με τον perceptron, ο οποίος ορίζεται αποκλειστικά με την προαναφερθείσα συνάρτηση προσήμου $\text{sign}(\cdot)$, οι νευρώνες μεταχειρίζονται ποικίλες μη γραμμικές συναρτήσεις ενεργοποίησης.

Η τιμή που υπολογίζεται πριν από την εφαρμογή της συνάρτησης ενεργοποίησης είναι η τιμή προ-ενεργοποίησης (**pre-activation**), ενώ η τιμή μετά την εφαρμογή της είναι η **μετα-ενεργοποίηση** (post-activation). Η έξοδος ενός νευρώνα είναι πάντα η μετα-ενεργοποίηση, ενώ οι τιμές προ-ενεργοποίησης εμφανίζονται κατά την οπίσθια διάδοση.

Κοινές Συναρτήσεις Ενεργοποίησης Η ταυτοτική απεικόνιση μπορεί να αποτελέσει συνάρτηση ενεργοποίησης, όμως δεν παρέχει μη-γραμμικότητα:

$$\Phi(x) = x$$

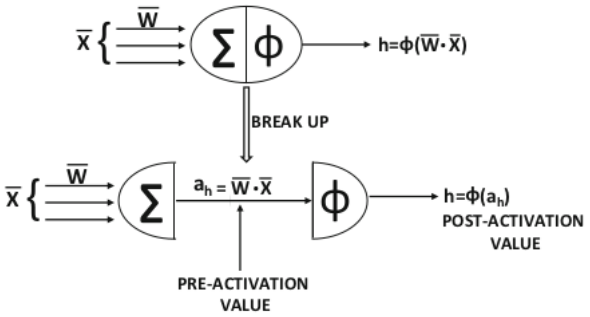
Η εν λόγω γραμμική συνάρτηση ενεργοποίησης χρησιμοποιείται συχνά στον νευρώνα εξόδου, όταν ο στόχος είναι μια πραγματική τιμή, στα πλαίσια παλινδρόμησης. Κλασσικές συναρτήσεις ενεργοποίησης, εμφανιζόμενες νωρίς στην ανάπτυξη των νευρωνικών δικτύων υπήρξαν η συνάρτηση προσήμου $\text{sign}(\cdot)$, η σιγμοειδής και η υπερβολική εφαπτομένη:

$$(\text{sigmoid}) : \Phi(x) \stackrel{\text{def}}{=} \frac{1}{1+e^{-x}} \tag{2.6}$$

$$(\tanh) : \Phi(x) \stackrel{\text{def}}{=} \frac{e^{2x}-1}{e^{2x}+1} \tag{2.7}$$

Ενώ η ενεργοποίηση προσήμου μπορεί να παράξει δυαδικές εξόδους κατά την πρόβλεψη, η μη διαφορισιμότητά της εμποδίζει την παρουσία της σε συναρτήσεις απώλειας. Παράλληλα, το πεδίο τιμών της παραγωγίσιμης σιγμοειδούς εντοπίζεται στο $(0, 1)$, επιτρέποντας μία πιθανοτική ερμηνεία αποτελεσμάτων.

Σχήμα 2.4: Αποσύνθεση του νευρωνικού υπολογισμού



Στο σχήμα απεικονίζεται η αποσύνθεση των υπολογισμών έκαστου νευρώνα σε δύο συναρτήσεις, η μία εκ των οποίων αναπαρίσταται από το σύμβολο άθροισης $\sum$ και η δεύτερη με το σύμβολο ενεργοποίησης Φ. Ακόμη, ένας όρος μεροληψίας υπονοείται ως μέρος της προενεργοποίησης, αν και δεν εμφανίζεται εδώ. Πηγή: [1].

Επιπροσθέτως, η σιγμοειδής, πέρα από την παραγωγή πιθανοτικών εξόδων κατασκευάζει συναρτήσεις απώλειας προερχόμενες από μοντέλα μεγίστης πιθανοφάνειας. Η υπερβολική εφαπτομένη έχει σχήμα παρόμοιο με αυτό της σιγμοειδούς, αλλά διαφορετικό εύρος τιμών $(-1, 1)$, γεγονός που την καθιστά προτιμότερη όταν απαιτείται οι έξοδοι των υπολογισμών να λαμβάνουν τόσο θετικές όσο και αρνητικές τιμές. Ιστορικά, οι δύο αυτές συναρτήσεις υπήρξαν οι κυρίαρχες επιλογές για την εισαγωγή μη γραμμικότητας σε ένα νευρωνικό δίκτυο. Τα τελευταία χρόνια ωστόσο, δύο τμηματικά γραμμικές συναρτήσεις ενεργοποίησης έχουν γίνει πιο δημοφιλείς. Ειδικότερα, η ανορθωμένη γραμμική μονάδα (Rectified Linear Unit - ReLU)

$$\text{ReLU}(x) \stackrel{\text{def}}{=} \max\{x, 0\} \tag{2.8}$$

και η «σκληρή» υπερβολική εφαπτομένη (hard tanh):

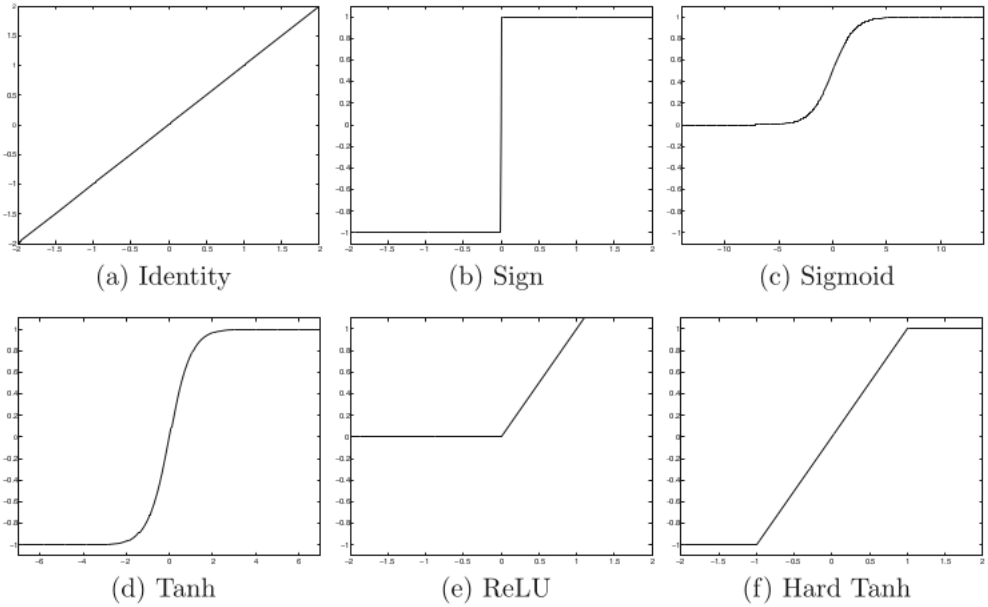
$$[\text{hard } \tanh](x) \stackrel{\text{def}}{=} \max\{\min[x, 1], -1\} \tag{2.9}$$

έχουν αντικαταστήσει σε μεγάλο βαθμό τη σιγμοειδή και τη συμβατική υπερβολική εφαπτομένη στα σύγχρονα νευρωνικά δίκτυα, λόγω της μείωσης του υπολογιστικού φορτίου που επιφέρουν.

2.1.4 Διάταξη Ταξινόμησης

Ο σχεδιασμός του επιπέδου εξόδου ενός νευρωνικού δικτύου καθορίζει τη λειτουργικότητα διεργασιών ταξινόμησης. Αυτή η ενότητα εξερευνά την τυπική επιλογή αρχιτεκτονικών εξόδου, αλλά και τις συναρτήσεις απώλειας για προβλήματα δυαδικής όσο και πολυκατηγορικής ταξινόμησης.

Σχήμα 2.5: Διάφορες Συναρτήσεις Ενεργοποίησης



Απεικονίσεις των προαναφερθεισών συναρτήσεων ενεργοποίησης. Αξιοσημείωτα, όλες είναι μονοτονικές. Πηγή: [1].

2.1.4.1 Σχεδιάζοντας το Επίπεδο Εξόδου

Ο σχεδιασμός του επιπέδου εξόδου καθορίζει τη λειτουργικότητα του νευρωνικού δικτύου και χαρακτηρίζεται από την επιλογή του πλήθους των νευρώνων, καθώς και των αντιστοίχων συναρτήσεων ενεργοποίησης.

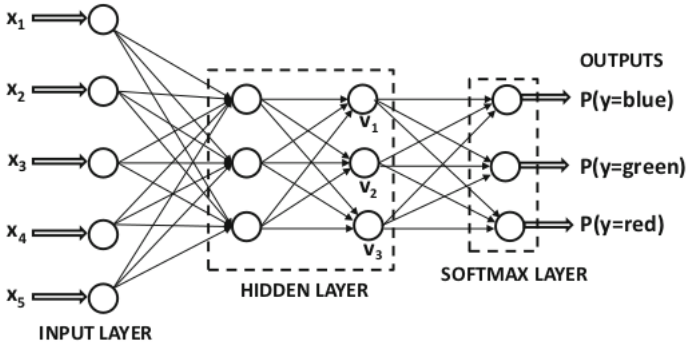
Δυαδική Ταξινόμηση Για εργασίες δυαδικής ταξινόμησης, είθισται το επίπεδο εξόδου να αποτελείται από έναν μοναδικό νευρώνα με μια σιγμοειδή συνάρτηση ενεργοποίησης, η οποία παράγει εξόδους με εύρος τιμών στο $(0, 1)$, ερμηνευόμενες ως πιθανότητες της θετικής κλάσης. Αυτή η διαμόρφωση είναι η τυπική προσέγγιση για την πρόβλεψη δυαδικών αποτελεσμάτων, όπου η έξοδος αντιπροσωπεύει την πιθανότητα της μίας από τις δύο κλάσεις.

Πολυκατηγορική Ταξινόμηση Σε προβλήματα πολυκατηγορικής ταξινόμησης, είθισται το επίπεδο εξόδου να περιέχει k νευρώνες, καθένας εκ των οποίων αντιστοιχεί σε μία από τις k κλάσεις. Η συνάρτηση ενεργοποίησης softmax εφαρμόζεται στις k αυτές εξόδους, και τις μετατρέπει σε πιθανότητες που αθροίζονται στη μονάδα. Η συνάρτηση softmax για την i-οστή έξοδο ορίζεται ως εξής:

$$\forall i \in \{1, \ldots, k\} : \Phi(v_i) = \frac{\exp(v_i)}{\sum_{j=1}^{k} \exp(v_j)} \tag{2.10}$$

Συνιστάται η θεώρηση των k αυτών τιμών ως μετα-ενεργοποιήσεις των τελικών k νευρώνων, όπου οι αντίστοιχες είσοδοι είναι $v_1, \ldots, v_k$. Ένα παράδειγμα της συνάρτησης softmax με τρεις εξόδους απεικονίζεται στο σχήμα 2.6. Σημειώνεται ότι οι τρεις έξοδοι αντιστοιχούν στις πιθανότητες των τριών κατηγορικών κλάσεων και επομένως μετατρέπουν τις αντίστοιχες εξόδους του τελικού κρυφού επιπέδου σε πιθανότητες μέσω της softmax.

Σχήμα 2.6: Πολλαπλές έξοδοι υπό την εφαρμογή ενός επιπέδου softmax



Στην προκειμένη περίπτωση πολυκατηγορικής ταξινόμησης αντιστοιχούν τρεις νευρώνες εξόδου για τις τρεις κατηγορίες, συγκεκριμένα: μπλε, πράσινο και κόκκινο. Οι τιμές του επιπέδου softmax θα αντιπροσωπεύουν πιθανότητες ταξινόμησης και το άθροισμά τους είναι 1. Πηγή: [1].

2.1.4.2 Επιλογή Συνάρτησης Απώλειας

Η επιλογή μιας κατάλληλης **συνάρτησης απώλειας** (loss function) είναι καθοριστική για την αποτελεσματική εκπαίδευση. Η συνάρτηση απώλειας προσμετρά τη διαφορά μεταξύ των προβλεπόμενων εξόδων και των πραγματικών τιμών του στόχου, καθοδηγώντας τη διαδικασία βελτιστοποίησης για τη βελτίωση της απόδοσης του μοντέλου. Ιδιαίτερα στις εργασίες ταξινόμησης, χρησιμοποιούνται διαφορετικές συναρτήσεις απώλειας, ανάλογα με τον τύπο του προβλήματος: δυαδικό ή πολυκατηγορικό.

Δυαδική Εντροπική Συνάρτηση Απώλειας Για διεργασίες δυαδικής ταξινόμησης, είθισται η χρήση της δυαδικής εντροπικής συνάρτησης απώλειας (binary cross-entropy loss), επίσης γνωστή και ως λογαριθμική απώλεια (log loss). Αυτή η συνάρτηση αξιολογεί την απόδοση ενός μοντέλου του οποίου η έξοδος είναι μια πιθανότητα με εύρος τιμών μεταξύ 0 και 1. Η εν λόγω απώλεια ορίζεται ως εξής:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \cdot \log\left(\hat{y}\right) + (1-y_i)\cdot\log\left(1-y_i\right)\right] \tag{2.11}$$

όπου με $\log(\cdot)$ συμβολίζεται ο νεπέριος λογάριθμος, N είναι το πλήθος των δειγμάτων, y_i είναι η πραγματική ετικέτα του i-οστού δείγματος και $\hat{y}_i$ είναι η προβλεπόμενη πιθανότητα για το i-οστό δείγμα. Σε όρους λειτουργικότητας, η συνάρτηση ποινικοποιεί το μοντέλο όταν η προβλεπόμενη πιθανότητα αποκλίνει από την πραγματική ετικέτα, προωθώντας την εξαγωγή πιθανοτήτων πλησιέστερων στις αληθινές τιμές των δεδομένων εκπαίδευσης.

Αραιή Κατηγορική Συνάρτηση Εντροπικής Απώλειας Για διεργασίες πολυκατηγορικής ταξινόμησης, είθισται η χρήση της αραιής κατηγορικής συνάρτησης εντροπικής απώλειας (sparse categorical cross-entropy). Λειτουργεί παρόμοια με τη δυαδική εντροπία, προσαρμοσμένη σε πολλαπλές κλάσεις και ορίζεται ως εξής:

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=0}^{k-1}\log\left(\mathbb{P}\left[y_i = c\right]\right) \tag{2.12}$$

όπου N είναι ο αριθμός των δειγμάτων, c είναι η ετικέτα των κλάσεων, δηλαδή ένας ακέραιος αριθμός από 0 έως $k-1$ για τη γενική περίπτωση των k κλάσεων, y_i είναι η πραγματική

ετικέτα κλάσης του i-οστού δείγματος και είθισται η πιθανότητα να συμβολίζεται με $\mathbb{P}[\cdot]$, άρα $\mathbb{P}[y_i = c]$ είναι η προβλεπόμενη πιθανότητα της πραγματικής κλάσης y_i για το i-οστό δείγμα να ανήκει στην κλάση c. Η εν λόγω συνάρτηση είναι ιδιαίτερα αποδοτική, καθώς μεταχειρίζεται απευθείας τις ακέραιες ετικέτες κλάσης δίχως περαιτέρω επεξεργασία (μετατροπή σε μορφή one-hot encoding), μειώνοντας έτσι το υπολογιστικό κόστος.

> **Σημείωση**: *Όλη η ορολογία που αναπτύσσεται σε αυτήν την ενότητα, στο πλαίσιο των MLPs, εφαρμόζεται επίσης στα Συνελικτικά Νευρωνικά Δίκτυα (CNNs). Αυτό οφείλεται στο γεγονός ότι τόσο τα MLPs όσο και τα CNNs μοιράζονται την ίδια αρχιτεκτονική για το επίπεδο εξόδου και τη συνάρτηση απώλειας. Η περίπτωση του TabNet διαφέρει μόνο στη δυαδική περίπτωση, όπου χρησιμοποιούνται δύο νευρώνες υπό την ενεργοποίηση softmax. Ωστόσο, εύκολα αποδεικνύεται ότι η αρχιτεκτονική αυτή είναι μαθηματικά ισοδύναμη με έναν συμβατικό σιγμοειδή νευρώνα εξόδου.*

2.1.5 Δίκτυα Εμπρόσθιας Τροφοδοσίας

> Τα MLPs ξεπερνούν τους περιορισμούς του perceptron με την ενσωμάτωση κρυφών επιπέδων, συμπράττοντας στην προτυποποίηση μη γραμμικών σχέσεων και βελτιώνοντας σημαντικά τις δυνατότητες των νευρωνικών δικτύων σε διάφορες εφαρμογές.

Τα MLPs περιέχουν πολλαπλά υπολογιστικά επίπεδα, σε αντίθεση με απλούστερα μοντέλα όπως το perceptron, το οποίο περιλαμβάνει μόνο ένα επίπεδο εισόδου και ένα επίπεδο εξόδου. Σε ένα perceptron, το επίπεδο εξόδου είναι το μοναδικό υπολογιστικό επίπεδο, καθώς το επίπεδο εισόδου απλώς μεταδίδει τα δεδομένα. Επακολούθως, όλοι οι υπολογισμοί σε ένα perceptron είναι πλήρως εμφανείς στον χρήστη. Αντιθέτως, τα πολυεπίπεδα νευρωνικά δίκτυα έχουν επιπλέον ενδιάμεσα επίπεδα, γνωστά ως κρυφά επίπεδα (**hidden layers**), στα οποία οι αντίστοιχοι υπολογισμοί δεν είναι άμεσα προσβάσιμοι στον χρήστη. Η εν λόγω αρχιτεκτονική των πολυεπίπεδων νευρωνικών δικτύων αναφέρεται τυπικά ως δίκτυα εμπρόσθιας τροφοδοσίας (**feed-forward networks**), αφού τα διαδοχικά επίπεδα τροφοδοτούν αλλεπάλληλα τα επακόλουθα τους με κατεύθυνση από την είσοδο προς την έξοδο.

$$\boldsymbol{h}_1 = \Phi\left(W_1^T \cdot \boldsymbol{x}\right)$$

$$\forall 1 \leq p \leq k-1 : \boldsymbol{h}_{p+1} = \Phi\left(W_{p+1}^T \cdot \boldsymbol{h_p}\right)$$

$$\boldsymbol{o} = \Phi\left(W_{k+1}^T \cdot \boldsymbol{h}_k\right)$$

Για ένα υποθετικό δίκτυο $k-1$ κρυφών επιπέδων, με $\boldsymbol{x}$ συμβολίζεται η αρχική είσοδος (με προέλευση από το σύνολο εκπαίδευσης), ως $\boldsymbol{h}_j$ αναγράφεται η ενεργοποίηση του j–οστού επιπέδου και με $\boldsymbol{o}$ σημειώνεται η έξοδος (output) του δικτύου. Οι παραπάνω ενεργοποιήσεις αντιστοιχούν στις μεταβάσεις:

$$[\text{Input} \longrightarrow \text{ Hidden Layer}]$$

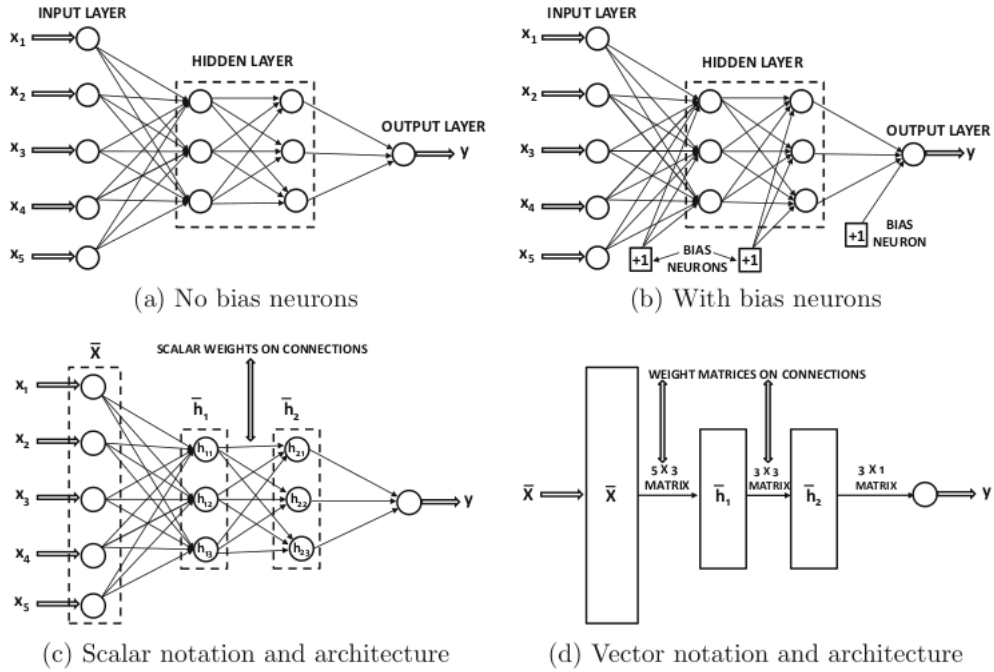
$$[\text{Hidden} \longrightarrow \text{Hidden Layer}]$$

$$[\text{Hidden} \longrightarrow \text{Output Layer}]$$

Σε ένα δίκτυο εμπρόσθιας τροφοδοσίας, κάθε νευρώνας σε ένα επίπεδο συνδέεται με κάθε νευρώνα στο επόμενο επίπεδο, μια δομή γνωστή ως **πλήρως συνδεδεμένη** αρχιτεκτονική (fully connected, FC). Επομένως, η αρχιτεκτονική του νευρωνικού δικτύου καθορίζεται σχεδόν πλήρως, εφόσον αποφασιστεί ο αριθμός των επιπέδων και ο αριθμός των νευρώνων σε κάθε επίπεδο. Η τελευταία λεπτομέρεια θα είναι η συνάρτηση απώλειας που βελτιστοποιείται στο επίπεδο εξόδου, η οποία συζητήθηκε στην προηγούμενη ενότητα.

Παρόμοια με το μοντέλο perceptron, τα πολυεπίπεδα δίκτυα προώθησης ενσωματώνουν νευρώνες μεροληψίας στα κρυφά επίπεδα, αλλά δυνητικά και στο επίπεδο εξόδου. Σημειώνεται ότι το επίπεδο εισόδου συνήθως δεν προσμετράται στο γενικότερο μέγεθος του δικτύου, επειδή απλώς μεταδίδει δεδομένα (δεν πραγματοποιείται κανένας υπολογισμός σε αυτό το επίπεδο). Αν ένα νευρωνικό δίκτυο περιέχει $p_1, \ldots, p_k$ υπολογιστικές μονάδες σε καθένα από τα k επίπεδα του, τότε ο αριθμός των κόμβων σε κάθε επίπεδο p_i αναφέρεται ως η διαστατικότητα του i-οστού επιπέδου.

Σχήμα 2.7: Δίκτυα εμπρόσθιας τροφοδοσίας



(a) No bias neurons (b) With bias neurons

(c) Scalar notation and architecture (d) Vector notation and architecture

Η πλήρως συνδεδεμένη αρχιτεκτονική ενός δικτύου εμπρόσθιας τροφοδοσίας με δύο κρυφά επίπεδα και ένα ενιαίο επίπεδο εξόδου, υπό βαθμωτή και διανυσματική απεικόνιση. Η προσθήκη μεροληπτικών νευρώνων αυξάνει σημαντικά το πλήθος των παραμέτρων. Πηγή: [1].

2.1.6 Οπίσθια Διάδοση

Θα παρουσιαστεί μια σύντομη επισκόπηση του αλγορίθμου οπίσθιας διάδοσης, χωρίς εκτεταμένες μαθηματικές λεπτομέρειες.

Σε ένα μοντέλο perceptron η εκπαίδευση είναι απλή, καθώς το σφάλμα (αλλά και η συνάρτηση απώλειας) είναι μία συνάρτηση των βαρών, γεγονός που συνεπάγεται τον άμεσο υπολογισμό της παραγώγου. Στα δίκτυα πολλαπλών επιπέδων, η απώλεια αποτελείται από μία σύνθεση συναρτήσεων των βαρών του κάθε επιπέδου. Η εκπαίδευση μέσω **οπίσθιας διάδοσης** (backpropagation) αντιμετωπίζει αυτό το ζήτημα χρησιμοποιώντας τον κανόνα της αλυσίδας, προκειμένου να υπολογίσει τις παραγώγους του σφάλματος. Ο αλγόριθμος περιλαμβάνει δύο κύριες φάσεις:

τη **φάση προώθησης** (forward phase) και τη **φάση οπίσθιας διάδοσης** (backward phase).

1. **Φάση Προώθησης**: Οι είσοδοι τροφοδοτούνται στο νευρωνικό δίκτυο, διαδιδόμενες μέσω των επιπέδων. Χρησιμοποιώντας το τρέχον σύνολο βαρών παράγεται ένα αποτέλεσμα - έξοδος, το οποίο συγκρίνεται με τον στόχο. Στη συνέχεια, υπολογίζεται η παράγωγος της συνάρτησης απώλειας.

2. **Φάση Οπίσθιας Διάδοσης**: Με χρήση του κανόνα της αλυσίδας, ο αλγόριθμος υπολογίζει τις παραγώγους της συνάρτησης απώλειας ως προς τα βάρη, ξεκινώντας από την έξοδο και κινούμενος προς την κατεύθυνση της εισόδου. Αυτές οι παράγωγοι χρησιμοποιούνται στη συνέχεια για την ενημέρωση των βαρών, ελαχιστοποιώντας την συνάρτηση απώλειας.

Ο αλγόριθμος οπισθοδιάδοσης, η πλέον βασική μέθοδος εκπαίδευσης πολυεπίπεδων νευρωνικών δικτύων, επιτρέπει την αποτελεσματική μάθηση σύνθετων προτύπων από τα δεδομένα εκπαίδευσης.

2.1.7 Αρχικοποίηση Βαρών

2.1.7.1 Εισαγωγή

Η αρχικοποίηση των βαρών σε perceptrons πολλαπλών επιπέδων αποτελεί κρίσιμο παράγοντα που επηρεάζει τη δυναμική της εκπαίδευσης και την τελική απόδοση του δικτύου. Δίχως προσεκτική αρχικοποίηση, τα νευρωνικά δίκτυα ίσως αντιμετωπίσουν προκλήσεις όπως αργή σύγκλιση, εξαφάνιση παραγώγων (vanishing gradients) ή ασταθείς συμπεριφορές μάθησης. Ως εκ τούτου, η ανάπτυξη αποτελεσματικών μεθόδων αρχικοποίησης βαρών είναι απαραίτητη για την εξασφάλιση επιτυχούς εκπαίδευσης βαθιών νευρωνικών δικτύων.

Δύο κύριες προσεγγίσεις, η αρχικοποίηση Glorot και η αρχικοποίηση He, έχουν προταθεί για την αντιμετώπιση ζητημάτων αρχικοποίησης βαρών σε MLPs. Η αρχικοποίηση Glorot, αποσκοπεί στη διατήρηση της διακύμανσης των σημάτων που διασχίζουν το δίκτυο, πραγματοποιώντας δειγματοληψίες βαρών από μια ομοιόμορφη (ή κανονική) κατανομή που παραμετροποιείται σύμφωνα με τον αριθμό των εισόδων και εξόδων. Αντίθετα, η αρχικοποίηση He στοχεύει σε συναρτήσεις ενεργοποίησης όπως η ReLU και οι παραλλαγές της. Αυτή η μέθοδος αρχικοποιεί τα βάρη από μια κανονική ή μία ομοιόμορφη κατανομή με μηδενική μέση τιμή και με διακύμανση εξαρτώμενη από τον αριθμό των σχετικών εισόδων. Οι εν λόγω καινοτόμες τεχνικές αρχικοποίησης επηρεάζουν σημαντικά τη σταθερότητα και την αποτελεσματικότητα της εκπαίδευσης βαθιών νευρωνικών δικτύων.

2.1.7.2 Αρχικοποίηση Glorot

Η αρχικοποίηση Glorot, επίσης γνωστή ως αρχικοποίηση Xavier, είναι μια μέθοδος αρχικοποίησης βαρών των νευρωνικών δικτύων, που προτάθηκε από τους Xavier Glorot και Yoshua Bengio, [6]. Έχει σχεδιαστεί για να αποτρέπει την εξαφάνιση ή την έκρηξη των παραγώγων κατά την εκπαίδευση, δηλαδή των μηδενισμό ή την απόκλιση των τιμών στο άπειρο. Η αρχικοποίηση Glorot επιλέγει βάρη από μια ομοιόμορφη κατανομή γύρω από το μηδέν, με το εύρος να καθορίζεται από τον αριθμό των εισόδων και εξόδων.

Η εν λόγω τεχνική είναι ιδιαίτερα αποτελεσματική για συναρτήσεις ενεργοποίησης όπως η σιγμοειδής και η υπερβολική εφαπτομένη, καθώς διατηρεί τη διακύμανση των σημάτων που διέρχονται στο δίκτυο, προτρέποντας τη φραγή των τιμών των βαρών. Τυπικά, τα βάρη W μπορούν να ληφθούν από μια ομοιόμορφη κατανομή εντός του εύρους:

$$W \sim U\left(-\sqrt{\frac{6}{n_i + n_{i+1}}}, \sqrt{\frac{6}{n_i + n_{i+1}}}\right) \tag{2.13}$$

Εναλλακτικά, τα βάρη W ακολουθούν μια κανονική κατανομή με παραμέτρους:

$$W \sim \mathcal{N}\left(0, \frac{2}{n_i + n_{i+1}}\right) \tag{2.14}$$

όπου n_i είναι το πλήθος των εισόδων και n_{i+1} είναι το πλήθος των εξόδων του i-οστού επιπέδου.

2.1.7.3 Αρχικοποίηση He

Η ανορθωμένη γραμμική μονάδα (Rectified Linear Unit - ReLU) είναι μια από τις συνηθέστερες συναρτήσεις ενεργοποίησης στη βαθιά μάθηση λόγω πλήθους πλεονεκτημάτων. Είναι υπολογιστικά αποδοτική, καθώς η παράγωγος υπολογίζεται εύκολα κατά την οπίσθια διάδοση, ούσα είτε 0 είτε 1. Επιπλέον, η ReLU περιορίζει το πρόβλημα της εξαφάνισης των παραγώγων, επιτρέποντας μόνο μη αρνητικές εξόδους, γεγονός που συνεπάγεται ταχύτερη εκπαίδευση και δυνατότητα για βαθύτερα δίκτυα. Αυτή η συνάρτηση είναι ιδιαίτερα διαδεδομένη στα συνελικτικά νευρωνικά δίκτυα (CNNs), τα οποία συχνά έχουν μεγάλες διαστάσεις εισόδου και βαθιές δομές.

Στο άρθρο "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification" από τους He et al. (2015) [7], οι συγγραφείς προτείνουν μια μεθοδολογία για τη βέλτιστη αρχικοποίηση των επιπέδων ReLU ενεργοποιήσεων. Αυτή η τεχνική, γνωστή ως αρχικοποίηση He, διασφαλίζει ότι η διακύμανση των εισόδων και εξόδων είναι συνεπής τόσο κατά τη φάση της προώθησης όσο και κατά την οπίσθια διάδοση, επάγοντας βελτιωμένη σταθερότητα και ταχύτητα εκπαίδευσης.

Η αρχικοποίηση He είναι παρόμοια με την αρχικοποίηση Glorot αλλά είναι ειδικά προσαρμοσμένη για τη ReLU και τις παραλλαγές της. Αντιμετωπίζει το ζήτημα της εξαφάνισης των παραγώγων παραμετροποιώντας τη διακύμανση της κατανομής των βαρών ανάλογα με τον αριθμό των εισόδων.

Τυπικά, στην αρχικοποίηση He, τα βάρη W λαμβάνονται από μια κανονική κατανομή με μηδενική μέση τιμή και διακύμανση $\frac{2}{n_{\text{in}}}$:

$$W \sim \mathcal{N}\left(0, \frac{2}{n_{\text{in}}}\right) \tag{2.15}$$

Εναλλακτικά, τα βάρη μπορούν επίσης να επιλεχθούν από μια ομοιόμορφη κατανομή εντός του εύρους:

$$W \sim U\left(-\sqrt{\frac{6}{n_{\text{in}}}}, \sqrt{\frac{6}{n_{\text{in}}}}\right) \tag{2.16}$$

όπου n_{in} είναι ο αριθμός των εισόδων. Και οι δύο προσεγγίσεις διασφαλίζουν ότι τα βάρη διατηρούν τις επιθυμητές ιδιότητες διακύμανσης, βελτιώνοντας έτσι την απόδοση των βαθιών νευρωνικών δικτύων που χρησιμοποιούν ReLU συναρτήσεις ενεργοποίησης.

2.1.7.4 Κανονική vs. Ομοιόμορφη Κατανομή

Καθώς αμφότερες κατανομές είναι διαδεδομένες επιλογές για την αρχικοποίηση βαρών, η βέλτιστη απόφαση παραμένει ένα ενεργό πεδίο έρευνας. Αν και οι Glorot και Bengio εισήγαγαν την έννοια της διατήρησης της διακύμανσης με μια ομοιόμορφη κατανομή, το θεωρητικό

πλεονέκτημά της έναντι μιας κανονικής κατανομής με την ίδια διακύμανση δεν έχει αποδειχθεί οριστικά.

Το βιβλίο "Deep Learning" από τους Goodfellow et al. [8], στο κεφάλαιο 8.4 αναγνωρίζει αυτό το ζήτημα, υποδεικνύοντας ότι η επιλογή της κατανομής μπορεί να έχει ελάχιστο αντίκτυπο στην απόδοση. Ωστόσο, το μέγεθος των αρχικών βαρών παίζει κρίσιμο ρόλο στη βελτιστοποίηση και τη γενίκευση:

> *"Σχεδόν πάντα αρχικοποιούμε όλα τα βάρη στο μοντέλο με τιμές που προέρχονται τυχαία από μια γκουσσιανή ή μία ομοιόμορφη κατανομή. Η επιλογή της κανονικής ή της ομοιόμορφης κατανομής δεν φαίνεται να έχει μεγάλη σημασία, αλλά δεν έχει μελετηθεί εξαντλητικά. Ωστόσο, το εύρος της αρχικής κατανομής έχει μεγάλη επίδραση τόσο στην έκβαση της διαδικασίας βελτιστοποίησης όσο και στην ικανότητα του δικτύου να γενικεύει."*

2.2 Συνελικτικά Νευρωνικά Δίκτυα

2.2.1 Εισαγωγή

Τα Συνελικτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks - CNNs) είναι μια θεμελιώδης έννοια της Βαθιάς Μάθησης και η βάση της σύγχρονης όρασης υπολογιστών. Χρησιμοποιούνται σε εύρος εφαρμογών, όπως στην επεξεργασία εικόνας και βίντεο, την ανάλυση ιατρικών εικόνων και διάφορους άλλους τομείς. Αυτή η ενότητα παρέχει μια συνοπτική επισκόπηση της αρχιτεκτονικής και της εκπαίδευσης των CNNs, αντλώντας κυρίως από τα έργα του Charu C. Aggarwal: [1, 2].

Όπως σημειώνει ο Aggarwal στο Κεφάλαιο 8 του βιβλίου "An Introduction to Artificial Intelligence", [2]:

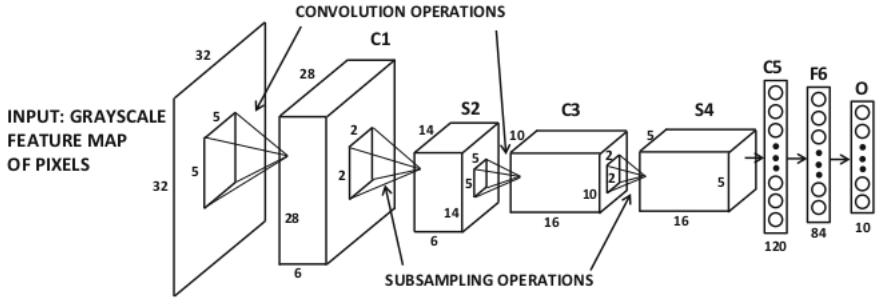
> "Τα συνελικτικά νευρωνικά δίκτυα είναι σχεδιασμένα για εισόδους τανυστικής δομής. Προφανές παράδειγμα δεδομένων τανυστικού χαρακτήρα αποτελεί η δισδιάστατη εικόνα. Αυτός ο τύπος δεδομένων εμφανίζει επίσης χωρικές εξαρτήσεις, καθώς οι γειτονικές περιοχές μια εικόνας συχνά έχουν παρόμοιες τιμές χρώματος των μεμονωμένων εικονοστοιχείων (pixels). Μια επιπλέον διάσταση καταγράφει τα διάφορα χρώματα, δημιουργώντας έναν τρισδιάστατο χώρο εισόδου.
>
> Άλλες μορφές διαδοχικών δεδομένων όπως το κείμενο, οι χρονοσειρές και οι ακολουθίες μπορούν επίσης να θεωρηθούν ειδικές περιπτώσεις τανυστικών δεδομένων με διάφορους τύπους εξάρτησης μεταξύ γειτονικών στοιχείων. Καθώς ένα σύνολο διαδοχικών δεδομένων ή χρονοσειρών μπορεί να θεωρηθεί ως ένα μονοδιάστατο σύνολο δεδομένων με γειτονικές (χρονικές) εξαρτήσεις, ενώ ένα σύνολο δεδομένων εικόνας μπορεί να θεωρηθεί ως ένα δισδιάστατο σύνολο δεδομένων με γειτονικές (χωρικές) εξαρτήσεις, οι ισχυρές σχέσεις μεταξύ γειτονικών τιμών καθιστούν δυνατή τη χρήση συνελικτικών νευρωνικών δικτύων και στις δύο περιπτώσεις. Η συντριπτική πλειονότητα των εφαρμογών των συνελικτικών νευρωνικών δικτύων επικεντρώνεται στα δεδομένα εικόνας, αν και μπορεί κανείς να χρησιμοποιήσει αυτά τα δίκτυα για όλους τους τύπους χρονικών, χωρικών και χωροχρονικών δεδομένων.
>
> Ένα σημαντικό χαρακτηριστικό που καθορίζει τα συνελικτικά νευρωνικά δίκτυα είναι η συνέλιξη. Η συνέλιξη είναι μια πράξη εσωτερικού γινομένου μεταξύ ενός τανυστή βαρών και ενός τανυστή που προέρχεται από τον χώρο εισόδου. Αυτή η πράξη είναι χρήσιμη για δεδομένα χωρικού τύπου, όπως τα δεδομένα εικόνας. Επομένως, τα συνελικτικά νευρωνικά δίκτυα ορίζονται ως δίκτυα που χρησιμοποιούν τη συνέλιξη σε τουλάχιστον ένα επίπεδο (layer), αν και τα περισσότερα συνελικτικά νευρωνικά δίκτυα χρησιμοποιούν αυτήν τη λειτουργία σε πολλαπλά επίπεδα."

Ιστορική Εξέλιξη & Βιολογικό Υπόβαθρο Η ανάπτυξη των CNNs έχει επηρεαστεί βαθιά απο το σύστημα όρασης των θηλαστικών, και συγκεκριμένα απο την ιεραρχική και πολυεπίπεδη νευρική επεξεργασία που ανακαλύφθηκε από τους νευροεπιστήμονες Hubel και Wiesel τη δεκαετία του 1960, [9]. Η εργασία τους αποκάλυψε ότι οι νευρώνες στον οπτικό φλοιό αντιστοιχούν σε συγκεκριμένα χωρικά ερεθίσματα, γεγονός που καταδεικνύει ότι πολύπλοκη οπτική αντίληψη ανακατασκευάζεται από απλούστερα χαρακτηριστικά μέσω διαδοχικών επιπέδων επεξεργασίας.

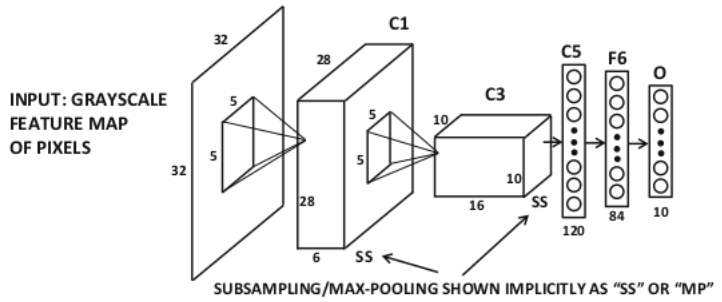
Επεκτείνοντας την παραπάνω αντίληψη, ο Kunihiko Fukushima ανέπτυξε το neocognitron τη δεκαετία του 1980 [10], ένα μοντέλο σχεδιασμένο για να αναγνωρίζει οπτικά μοτίβα μέσω μίας ιεραρχικής, πολυστρωματικής δομής. Αν και καινοτόμο, το neocognitron στερούνταν ορισμένα βασικά χαρακτηριστικά των σύγχρονων CNNs. Η μεγάλη πρόοδος ήρθε με το μοντέλο LeNet-5 από τον Yann LeCun και τους συναδέλφους του στα τέλη της δεκαετίας του 1980 και στις

αρχές της δεκαετίας του 1990, [11]. Το LeNet-5 χρησιμοποίησε αποτελεσματικά συνελικτικά επίπεδα για εργασίες όπως η αναγνώριση χειρόγραφων ψηφίων, εφαρμόζοντας φίλτρα (filters) σε διάφορα μέρη της εισόδου.

Σχήμα 2.8: Απεικόνιση της Αρχιτεκτονικής LeNet



(a) Detailed architectural representation



(b) Concise architectural representation

LeNet-5: Ένα από τα πρώιμα Συνελικτικά Νευρωνικά Δίκτυα. Πηγές: [2], [11].

Η επαναφορά των CNNs τη δεκαετία του 2010, υπό την παρουσία προηγμένων τεχνικών εκπαίδευσης, μεγάλων συνόλων δεδομένων με ετικέτες και ισχυρών GPUs, οδήγησε σε σημαντικά επιτεύγματα όπως το AlexNet, [12]. Κερδίζοντας τον διαγωνισμό ImageNet το 2012, το μοντέλο AlexNet εισήγαγε καινοτομίες όπως την συνάρτηση ΡεΛΥ, την κανονικοποίηση με dropout και την τεχνική της δημιουργίας συνθετικών δεδομένων (data augmentation). Αυτές οι εξελίξεις καθιέρωσαν νέα πρότυπα για την ταξινόμηση εικόνων και έκαναν τη βαθιά μάθηση δημοφιλή σε διάφορους τομείς.

Ευρύτερες Παρατηρήσεις Τα συνελικτικά νευρωνικά δίκτυα έχουν αρκετά χαρακτηριστικά που τα διακρίνουν από τα παραδοσιακά νευρωνικά δίκτυα:

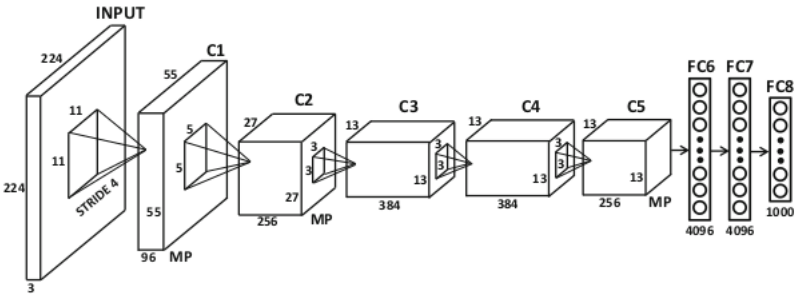
1. **Τοπική Συνεκτικότητα** (Local Connectivity): Τα CNNs χαρακτηρίζονται από τοπική συνεκτικότητα, ήτοι κάθε νευρώνας σε ένα συνελικτικό επίπεδο συνδέεται με μια μικρή, τοπική περιοχή της σχετικής εισόδου, γνωστή ως πεδίο υποδοχής (receptive field). Αυτό μειώνει σημαντικά τον αριθμό των παραμέτρων, καθιστώντας τα CNNs υπολογιστικώς αποδοτικά και αποτελεσματικά στην καταγραφή χωρικών ιεραρχιών.

2. **Κοινά βάρη** (Shared Weights): Τα CNNs χρησιμοποιούν κοινά βάρη (φίλτρα - filters) σε διαφορετικές χωρικές τοποθεσίες της εικόνας, επιτρέποντας στο δίκτυο να ανιχνεύει το ίδιο χαρακτηριστικό, όπως μια ακμή (edge) ή μια υφή (texture), ανεξάρτητα από τη θέση του χαρακτηριστικού στην εικόνα εισόδου. Αυτό ενισχύει την ικανότητα του δικτύου για γενίκευση.

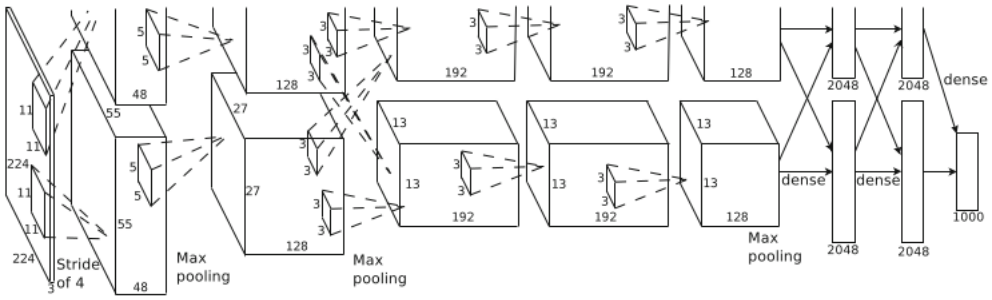
3. **Αναλλοίωτο ως προς τις Μετατοπίσεις** (Translation Invariance): Οι τελεστές pooling στα CNNs χαρακτηρίζονται (σε ένα βαθμό) από την ιδιότητα του αναλλοίωτου ως

προς τις μετατοπίσεις. Συγκεντρώνοντας πληροφορία σε μικρές περιοχές, διασφαλίζουν ότι μικρές μετατοπίσεις στο χώρο της εισόδου δεν επηρεάζουν σημαντικά την έξοδο. Αυτή η ιδιότητα είναι κρίσιμη για διεργασίες όπως η αναγνώριση αντικειμένων, όπου τα αντικείμενα μπορεί να εμφανίζονται σε διάφορες θέσεις μέσα σε μια εικόνα.

Σχήμα 2.9: Απεικόνιση της Αρχιτεκτονικής AlexNet



(a) Without GPU partitioning



(b) With GPU partitioning (original architecture)

Η αρχιτεκτονική AlexNet. Πηγές: [2], [12]. Οι συναρτήσεις ενεργοποίησης τύπου ReLU επακολουθούν κάθε συνελικτικό επίπεδο, και ως εκ τούτου δεν καταδεικνύονται επακριβώς. Να σημειωθεί ότι τα στρώματα max-pooling επισημαίνονται ως MP και απεικονίζονται να επακολουθούν μόνο ένα υποσύνολο των επιπέδων συνέλιξης - ReLU. Το διάγραμμα της αρχιτεκτονικής στο (b) προέρχεται από [12]: [A. Krizhevsky, I. Sutskever, and G. Hinton. Imagenet classification with deep convolutional neural networks. *NIPS Conference*, pp. 1097–1105. 2012.] © 2012 A. Krizhevsky, I. Sutskever, and G. Hinton.

4. **Ιεραρχική Εξαγωγή Χαρακτηριστικών** (Hierarchical Feature Extraction): Τα CNNs διαπρέπουν στην ιεραρχική εξαγωγή χαρακτηριστικών, με τα αρχικά επίπεδα να καταγράφουν low-level χαρακτηριστικά όπως ακμές και υφές, και τα βαθύτερα επίπεδα να εξάγουν high-level χαρακτηριστικά όπως σχήματα και αντικείμενα. Αυτή η ιεραρχική αναπαράσταση αντικατοπτρίζει τη διαδικασία επεξεργασίας της όρασης στον ανθρώπινο εγκέφαλο.

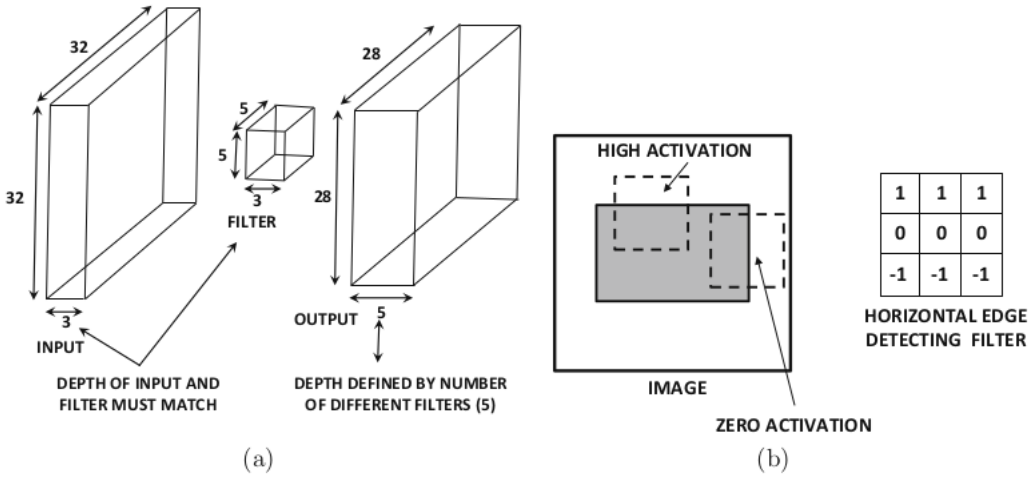
Αυτές οι ιδιότητες καθιστούν τα CNNs ιδιαίτερα κατάλληλα για εργασίες που σχετίζονται με εικόνες, όπου οι χωρικές ιεραρχίες και τα τοπικά χαρακτηριστικά κυριαρχούν. Με την πάροδο των ετών, τα CNNs έχουν εξελιχθεί για να χειρίζονται πιο σύνθετες διεργασίες, να ενσωματώνουν προηγμένες τεχνικές και να επιτυγχάνουν κορυφαίες επιδόσεις (state-of-the-art performance) σε ένα ευρύ φάσμα εφαρμογών.

2.2.2 Η Βασική Δομή ενός Συνελικτικού Δικτύου

Αυτή η ενότητα εξερευνά τα θεμελιώδη δομικά στοιχεία των CNNs, εξηγώντας πώς κάθε συνιστώσα συμβάλλει στη συνολική λειτουργία του δικτύου.

Στα CNNs, οι καταστάσεις επεξεργασίας σε κάθε επίπεδο είναι διατεταγμένες σε ένα χωρικό πλέγμα, διατηρώντας τις χωρικές σχέσεις από το ένα επίπεδο στο επόμενο. Κάθε επίπεδο σε ένα CNN είναι ένας τρισδιάστατος τανυστής (πλέγμα) που χαρακτηρίζεται από ύψος, πλάτος και βάθος.

Σχήμα 2.10: Απεικόνιση του Μηχανισμού της Συνέλιξης



Πηγή: [2]. (a) Η συνέλιξη μεταξύ ενός επιπέδου εισόδου διάστασης $32 \times 32 \times 3$ και ενός φίλτρου διάστασης $5 \times 5 \times 3$ παράγει ένα επίπεδο εξόδου με χωρικές διαστάσεις 28×28. Ως εκ τούτου, το βάθος της εξόδου δεν εξαρτάται από τις χωρικές διαστάσεις του επιπέδου εισόδου ή των φίλτρων. (b) Η μετακίνηση ενός φίλτρου στην εικόνα αποσκοπεί στην ανίχνευση ενός συγκεκριμένου χαρακτηριστικού, όπως τις οριζόντιες ακμές.

Το βάθος αναφέρεται στον αριθμό των καναλιών, όπως τα κανάλια χρώματος σε μια εικόνα ή τους χάρτες χαρακτηριστικών (feature maps) στα κρυφά επίπεδα. Τα CNNs λειτουργούν παρόμοια με τα παραδοσιακά νευρωνικά δίκτυα εμπρόσθιας τροφοδότησης, αλλά με χωρική οργάνωση και αραιές, προσεκτικά σχεδιασμένες συνδέσεις. Περιλαμβάνουν συνήθως συνελικτικά, pooling και ReLU επίπεδα, αλλά και πλήρως συνδεδεμένα επίπεδα που αντιστοιχούν στους κόμβους εξόδου. Η είσοδος σε ένα CNN είναι οργανωμένη ως ένα δισδιάστατο πλέγμα εικονοστοιχείων, με κάθε εικονοστοιχείο να αντιπροσωπεύεται από πολλαπλές τιμές που αντιστοιχούν σε διαφορετικά κανάλια χρώματος (π.χ. ο τύπος εικόνας RGB αντιστοιχεί σε τρία κανάλια χρώματος: κόκκινο - Red, πράσινο - Green και μπλε - Blue). Το βάθος του επιπέδου εισόδου καθορίζεται από αυτά τα κανάλια, ενώ τα κρυφά επίπεδα έχουν ένα βάθος που αντιπροσωπεύει τον αριθμό των χαρτών χαρακτηριστικών.

Τα συνελικτικά επίπεδα χρησιμοποιούν φίλτρα (πυρήνες - kernels) για να σαρώσουν την είσοδο, εκτελώντας μία πράξη εσωτερικού γινομένου σε κάθε θέση ώστε να παράξουν ως έξοδο έναν χάρτη χαρακτηριστικών. Το μέγεθος του φίλτρου είναι συνήθως πολύ μικρότερο από το επίπεδο εισόδου ως προς τις χωρικές διαστάσεις, αλλά αντιστοιχεί στο βάθος του, ενώ ο αριθμός των πιθανών θέσεων για την τοποθέτηση του φίλτρου καθορίζει τις χωρικές διαστάσεις του επιπέδου εξόδου. Για παράδειγμα, μια είσοδος 32×32 με ένα φίλτρο 5×5 παράγει μια έξοδο διάστασης 28×28. Το βάθος του επιπέδου εξόδου καθορίζεται από τον αριθμό των φίλτρων που χρησιμοποιούνται κατά την συνέλιξη. Πολλαπλά φίλτρα παράγουν πολλαπλούς χάρτες χαρακτηριστικών, αυξάνοντας το βάθος του επόμενου επιπέδου. Αυτό το βάθος είναι ανεξάρτητο από το βάθος του επιπέδου εισόδου.

Ο τελεστής της συνέλιξης επιδεικνύει την ιδιότητα του αναλλοίωτο ως προς τις μετατοπίσεις, που σημαίνει ότι η μετατόπιση της εισόδου έχει ως αποτέλεσμα μια αντίστοιχη μετατόπιση στον χάρτη χαρακτηριστικών. Αυτή η ιδιότητα, σε συνδυασμό με την ιεραρχική εξαγωγή χαρακτηριστικών, επιτρέπει στα CNNs να καταγράφουν και να επεξεργάζονται αποτελεσματικά χωρικά

μοτίβα στα δεδομένα.

Ο Μηχανισμός της Συνέλιξης Ο μηχανισμός της συνέλιξης αποτελεί τη θεμελιώδη συνιστώσα των Συνελικτικών Νευρωνικών Δικτύων, συμβάλλοντας στην αυτόματη και προσαρμοστική μάθηση χωρικών ιεραρχιών επί των χαρακτηριστικών των δεδομένων εισόδου. Τα συνελικτικά επίπεδα εφαρμόζουν ένα σύνολο φίλτρων στα δεδομένα εισόδου, όπου κάθε φίλτρο χρησιμοποιείται για να εξάγει διαφορετικά χαρακτηριστικά από την είσοδο.

Μαθηματική Περιγραφή της Συνέλιξης Ο τελεστής της συνέλιξης περιλαμβάνει την ολίσθηση ενός φίλτρου (γραμμικού πυρήνα) επί των δεδομένων εισόδου και τον υπολογισμό του εσωτερικού γινομένου μεταξύ του φίλτρου και της σχετικής υποπεριοχής που καλύπτει. Σε μαθηματική γραφή, για ένα μόνο χρωματικό κανάλι εισόδου (greyscale images), ο τελεστής της συνέλιξης εκφράζεται ως:

$$I * K\left(i,j\right) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I\left(i+m, j+n\right) \cdot K\left(m,n\right) \tag{2.17}$$

όπου: $I\left(\cdot,\cdot\right)$ είναι ο πίνακας εισόδου (input matrix), δηλαδή μία εικόνα, $K\left(\cdot,\cdot\right)$ είναι ο πίνακας του πυρήνα (kernel matrix), (i,j) είναι οι συντεταγμένες του πίνακα εξόδου και M, N είναι οι διαστάσεις του πυρήνα. Για εισόδους πολλαπλών χρωματικών καναλιών, όπως οι εικόνες RGB, η πράξη της συνέλιξης επεκτείνεται ως εξής:

$$I * K\left(i,j\right) = \sum_{c=0}^{C-1} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} I_c\left(i+m, j+n\right) \cdot K_c\left(m,n\right) \tag{2.18}$$

όπου: C ο αριθμός των καναλιών εισόδου, $I_c\left(\cdot,\cdot\right)$ το c-στό κανάλι εισόδου και K_c το c-στο κανάλι του πυρήνα.

2.2.2.1 Συνέλιξη: Padding και Strides

Συνέλιξη: Padding Η έννοια padding αναφέρεται σε μία βασική έννοια της συνέλιξης, η οποία χρησιμοποιείται στα CNNs για τον έλεγχο των χωρικών διαστάσεων των χαρτών χαρακτηριστικών εξόδου και μεταφράζεται ως προσθήκη μηδενικών στο σύνορο. Προσθέτοντας επιπλέον εικονοστοιχεία με μηδενικά γύρω από την εικόνα εισόδου, το padding διατηρεί τις πρωταρχικές διαστάσεις κατόπιν της συνέλιξης, γεγονός ιδιαίτερα ωφέλιμο για τη διατήρηση της πληροφορίας στα άκρα της εικόνας. Οι κύριοι τύποι padding είναι οι εξής:

1. **Valid Padding**: Επίσης γνωστό ως "no padding", αυτή η μέθοδος περιγράφει την εκτέλεση της συνέλιξης χωρίς την επιπλέον προσθήκη εικονοστοιχείων. Ως αποτέλεσμα, οι διαστάσεις εξόδου είναι μικρότερες από τις διαστάσεις εισόδου.

   $$\text{[διάσταση εισόδου: } (n \times n)] \wedge \text{[διάσταση φίλτρου: } (f \times f)]$$

   $$\xrightarrow{\text{Valid Padding}}$$

   $$\text{διάσταση εξόδου: } (n-f+1) \times (n-f+1)$$

   Αυτή η τεχνική αποφαίνεται χρήσιμη όταν ο στόχος είναι η μείωση των διαστάσεων των χαρτών χαρακτηριστικών. Ωστόσο, αυτή η προσέγγιση μπορεί να οδηγήσει σε απώλεια πληροφορίας στα άκρα της εικόνας εισόδου.

2. **Same Padding**: Αυτή η μέθοδος "συμπληρώνει" την είσοδο έτσι ώστε οι διαστάσεις εξόδου να είναι ίδιες με τις διαστάσεις εισόδου. Η συμπλήρωση με μηδενικά διασφαλίζει ότι ο τελεστής της συνέλιξης δεν μειώνει το χωρικό μέγεθος των χαρτών χαρακτηριστικών και καθιστά εφικτή τη στοίβαξη πολλαπλών επιπέδων χωρίς απώλεια χωρικών πληροφοριών. Για φίλτρο διάστασης $f \times f$, το εύρος του padding p υπολογίζεται ως εξής:

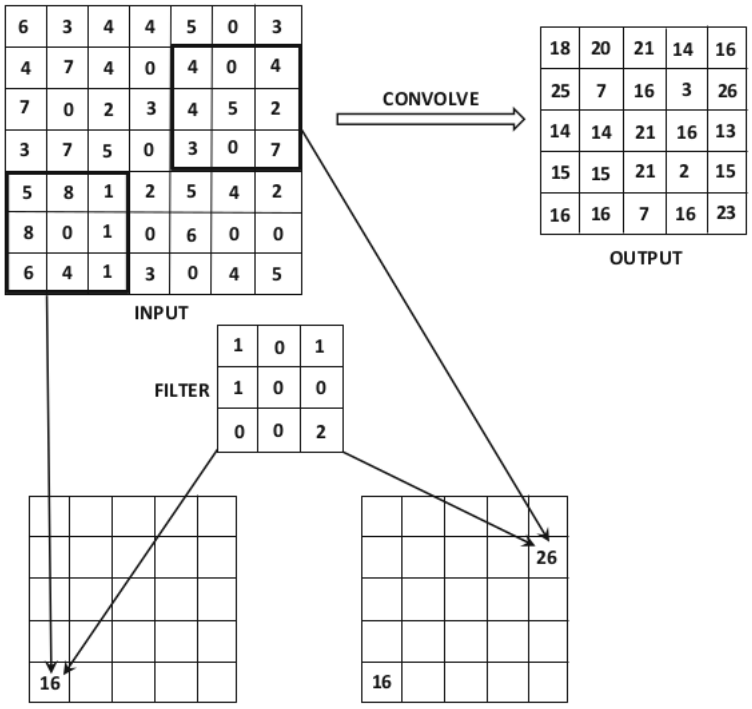
$$(\text{Same Padding}) : p = \left\lfloor \frac{f-1}{2} \right\rfloor$$

Με αυτόν τον τρόπο διασφαλίζεται ότι η διάσταση της εξόδου αντιστοιχεί σε αυτήν της συνελικτικής εισόδου.

3. **Full Padding**: Η μέθοδος "Full padding" είναι μια παραλλαγή της μεθόδου "same padding", η οποία πρωτίστως διατηρεί την πληροφορία στα άκρα της εικόνας, προσθέτοντας ακόμη περισσότερα εικονοστοιχεία σε σύγκριση με το same padding, με αποτέλεσμα έναν χάρτη χαρακτηριστικών εξόδου που έχει μεγαλύτερες διαστάσεις από την αρχική είσοδο. Με αυτόν τον τρόπο, εξασφαλίζεται ότι κάθε εικονοστοιχείο στην εικόνα εισόδου, συμπεριλαμβανομένων εκείνων στα άκρα της εικόνας, περιλαμβάνεται σε μια πράξη συνέλιξης τουλάχιστον μία φορά.

Συνέλιξη: Strides Τα strides (μετάφραση: βηματισμός) καθορίζουν το μέγεθος του βήματος με το οποίο το συνελικτικό φίλτρο μετακινείται κατά μήκος της εικόνας εισόδου. Ένα βήμα ίσο με ένα σημαίνει ότι το φίλτρο μετακινείται κατά ένα εικονοστοιχείο κάθε φορά, ενώ ένα βήμα μεγαλύτερο από ένα έχει ως αποτέλεσμα βηματισμό μεγαλύτερου μεγέθους.

Σχήμα 2.11: Απεικόνιση του Μηχανισμού της Συνέλιξης: Stride ίσο με 1



Ένα παράδειγμα συνέλιξης μεταξύ μίας εισόδου διάστασης $7 \times 7 \times 1$ και ενός $3 \times 3 \times 1$ φίλτρου με stride ίσο με 1. Για το συγκεκριμένο παράδειγμα έχει επιλεγεί είσοδος και φίλτρο με βάθος μεγέθους 1. Για βάθος μεγαλύτερο από 1, οι συνεισφορές κάθε χάρτη χαρακτηριστικών εισόδου θα προστεθούν ώστε να δημιουργηθεί μία μοναδική τιμή στον χάρτη χαρακτηριστικών. Πηγή: [2].

Η επιλογή του stride επηρεάζει τις χωρικές διαστάσεις του χάρτη χαρακτηριστικών εξόδου:

- **Stride ίσο με 1**: Παράγει χάρτες χαρακτηριστικών ύψιστης ανάλυσης (resolution), διατηρώντας τις περισσότερες λεπτομέρειες. Αποτελεί την τυπική προεπιλογή βηματισμού.
- **Stride μεγαλύτερο από 1**: Οδηγεί σε υποδειγματοληψία (down-sampling), μειώνοντας τις χωρικές διαστάσεις του χάρτη χαρακτηριστικών, το οποίο συνεπάγεται ταχύτερους υπολογισμούς, με κόστος τη μειωθείσα ανάλυση εικόνας.

2.2.2.2 Το επίπεδο ReLU

Η συνάρτηση ενεργοποίησης ανορθωμένης γραμμικής μονάδας[3] (rectified linear unit - ReLU) είναι ένας μη-γραμμικός τελεστής, που εφαρμόζεται ύστερα από τη συνέλιξη ώστε να εισάγει μη γραμμικότητα στο δίκτυο, και ορίζεται ως:

$$f(x) = \max(0, x)$$

Η ReLU αντιμετωπίζει αποτελεσματικά το ζήτημα του μηδενισμού της κλίσης κατά την οπίσθια διάδοση (vanishing gradient problem), το οποίο παρατηρείται υπό τις συνήθεις συναρτήσεις ενεργοποίησης, όπως τη σιγμοειδή και την υπερβολική εφαπτομένη. Θέτοντας όλες τις αρνητικές τιμές σε μηδέν, η ReLU υποστηρίζει το δίκτυο στην αποτελεσματική μάθηση σύνθετων μοτίβων και χαρακτηριστικών. Η απλότητα και η αποτελεσματικότητά της, την καθιστούν τυπική επιλογή συνάρτησης ενεργοποίησης σε πολλές αρχιτεκτονικές CNN. Όπως δηλώνει ο Aggarwal, [2]:

> "Είναι αξιοσημείωτο ότι η χρήση της συνάρτησης ενεργοποίησης ReLU είναι μια πρόσφατη εξέλιξη στον σχεδιασμό νευρωνικών δικτύων. Παλαιότερα, συναρτήσεις ενεργοποίησης όπως η σιγμοειδής (sigmoid) και η υπερβολική εφαπτομένη (tanh) χρησιμοποιούνταν ευρέως. Ωστόσο, όπως καταδείχθηκε [12], η χρήση της ReLU παρουσιάζει τεράστια πλεονεκτήματα έναντι αυτών των συναρτήσεων ενεργοποίησης, σε επίπεδο ταχύτητας και ακρίβειας. Η επαυξημένη ταχύτητα συνδέεται επίσης με την ακρίβεια, καθώς επιτρέπει τη χρήση βαθύτερων μοντέλων και την εκτεταμένη εκπαίδευση τους. Τα τελευταία χρόνια, η χρήση της ReLU έχει αντικαταστήσει τις άλλες συναρτήσεις ενεργοποίησης στον σχεδιασμό συνελικτικών νευρωνικών δικτύων, [...]."

Συμπερασματικά, τα πλεονεκτήματα της συνάρτησης αυτής, όπως η αποτροπή του μηδενισμού της κλίσης, η επιτάχυνση της διαδικασίας εκπαίδευσης και η ενίσχυση της λειτουργικής απόδοσης, έχουν εδραιώσει τη ReLU ως την τυπική συνάρτηση ενεργοποίησης στα σύγχρονα CNNs.

2.2.2.3 Pooling

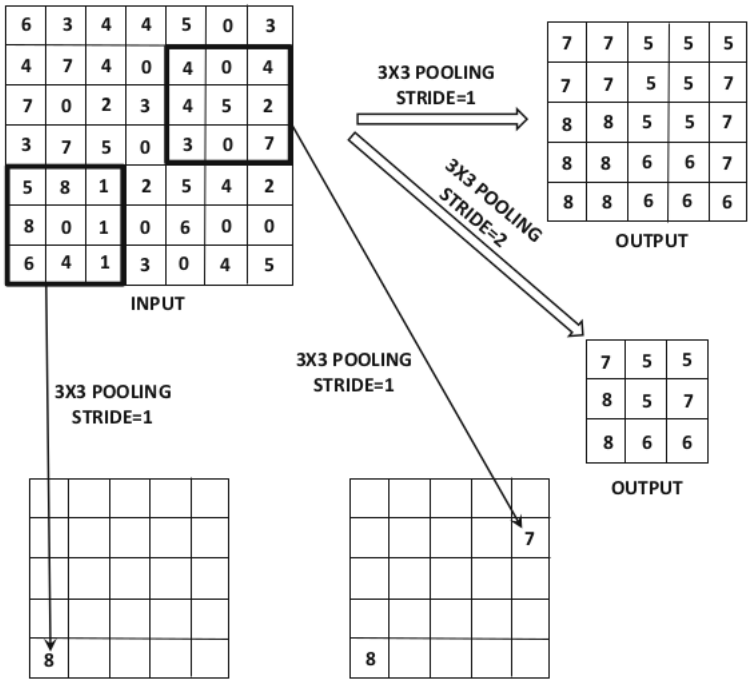
Ο τελεστής pooling (μετάφραση: τελεστής συγκέντρωσης), μία σημαντική συνιστώσα των CNNs, μειώνει τις χωρικές διαστάσεις των χαρτών ενεργοποίησης διατηρώντας το βάθος τους. Σε αντίθεση με τη συνέλιξη, το pooling επιδρά ανεξάρτητα σε κάθε χάρτη ενεργοποίησης, διατηρώντας τον αριθμό τους, αλλά μειώνοντας τις χωρικές τους διαστάσεις.

Η πλέον κύριαρχη μορφή pooling ανάμεσα σε άλλες είναι το max-pooling (διαφορετικές επιλογές: average pooling, global-pooling, stohastic-pooling, κ.α.). Ο τελεστής pooling με stride ίσο με 2 και ένα 2×2 φίλτρο αποτελεί συνήθη επιλογή για την επίτευξη αναλλοίωτου

[3]Για λόγους συνέπειας, θα αναγράφεται ως ReLU.

μετατοπίσεων, συμβάλλοντας σε διεργασίες όπως η ταξινόμηση αντικειμένων, ανεξάρτητα από τη θέση τους στην εικόνα. Επιπροσθέτως, η παρουσία του pooling αυξάνει το μέγεθος του πεδίου υποδοχής, επιτρέποντας στο δίκτυο να καταγράφει μεγαλύτερες περιοχές της εικόνας στα επόμενα επίπεδα.

Σχήμα 2.12: Απεικόνιση του τελεστή pooling



Απεικονίζονται δύο παραδείγματα από χάρτες ενεργοποίησης τύπου max-pooling, διάστασης 7×7 με αντίστοιχα strides 1 και 2. Στην περίπτωση όπου το stride ισούται με 1, δημιουργείται ένας 5×5 πίνακας ενεργοποίησης (activation map) με έντονα επαναλαμβανόμενα στοιχεία λόγω της στρατηγικής μεγιστοποίησης σε επικαλυπτόμενες περιοχές. Όταν το stride είναι ίσον με 2, δημιουργείται ένας 3×3 χάρτης ενεργοποίησης με λιγότερη επικάλυψη. Σε αντίθεση με τη συνέλιξη, κάθε χάρτης ενεργοποίησης παράγεται ανεξάρτητα, και επομένως ο αριθμός των χαρτών εξόδου είναι ακριβώς ίσος με τον αριθμό των χαρτών εισόδου. Πηγή: [2].

2.2.2.4 Πλήρως Συνδεδεμένο επίπεδο

Σε ένα συνελικτικό νευρωνικό δίκτυο, κάθε χαρακτηριστικό στο τελικό χωρικό επίπεδο συνδέεται με κάθε κρυφή κατάσταση (hidden state) στο πρώτο πλήρως συνδεδεμένο επίπεδο (fully connected layer), λειτουργώντας παρόμοια με ένα παραδοσιακό δίκτυο εμπρόσθιας τροφοδοσίας. Συχνά, χρησιμοποιούνται πολλαπλά πλήρως συνδεδεμένα επίπεδα για την επαύξηση της υπολογιστικής ισχύος, δομημένα όπως τα παραδοσιακά δίκτυα. Αυτά τα επίπεδα είναι πυκνά συνδεδεμένα, περιέχοντας το μεγαλύτερο μέρος των παραμέτρων του δικτύου. Όπως εξηγεί ο Aggarwal, [2]:

> "Εφόσον τα πλήρως συνδεδεμένα επίπεδα είναι πυκνά συνδεδεμένα, η συντριπτική πλειονότητα των παραμέτρων εντοπίζεται στα πλήρως συνδεδεμένα επίπεδα. [...] Ομοίως, οι συνδέσεις από το τελευταίο χωρικό επίπεδο στο πρώτο πλήρως συνδεδεμένο επίπεδο θα έχουν μεγάλο αριθμό παραμέτρων. Παρόλο που τα συνελικτικά επίπεδα παρουσιάζουν μεγαλύτερο αριθμό ενεργοποιήσεων (και άρα μεγαλύτερο αποτύπωμα μνήμης - memory footprint), τα πλήρως συνδεδεμένα επίπεδα συχνά έχουν μεγαλύτερο αριθμό συνδέσεων (και αριθμό παραμέτρων). Ο λόγος που οι ενεργοποιήσεις συμβάλλουν σημαντικά στο αποτύπωμα μνήμης είναι ότι ο αριθμός

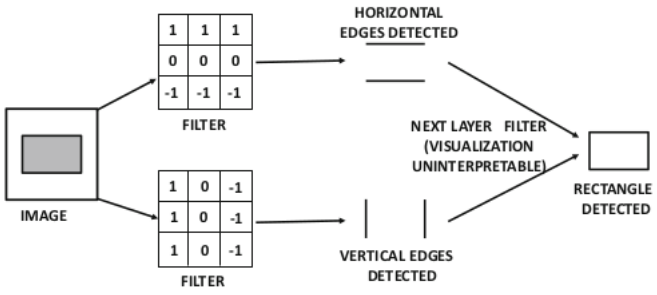
των ενεργοποιήσεων πολλαπλασιάζεται με το μέγεθος του mini-batch για την παρακολούθηση μεταβλητών κατά τις εμπρός και πίσω φάσεις της οπίσθιας διάδοσης (backpropagation)."

Αυτές οι παρατηρήσεις είναι κρίσιμες κατά τον σχεδιασμό νευρωνικών δικτύων βάσει περιορισμών πόρων όπως τα δεδομένα και η μνήμη. Η φύση των πλήρως συνδεδεμένων επιπέδων διαφέρει ανά εφαρμογή. Γενικά, σε εργασίες ταξινόμησης, το επίπεδο εξόδου του δικτύου είναι πλήρως συνδεδεμένο με το προτελευταίο επίπεδο και κάθε σύνδεση έχει ένα σχετικό βάρος. Η συνάρτηση ενεργοποίησης που χρησιμοποιείται, όπως η sigmoid ή η softmax, εξαρτάται από τον τύπο της ταξινόμησης - δυαδική ή πολυκατηγορική.

2.2.2.5 Ιεραρχική Κατασκευή Χαρακτηριστικών

Η ιεραρχική κατασκευή χαρακτηριστικών (hierarchical feature engineering) στα CNNs περιλαμβάνει την κατανόηση του τρόπου με τον οποίο τα διαφορετικά επίπεδα του δικτύου ανιχνεύουν και συναρμολογούν χαρακτηριστικά από τις εικόνες εισόδου. Στα αρχικά επίπεδα, τα φίλτρα των CNNs ανιχνεύουν low-level χαρακτηριστικά, όπως ακμές. Για παράδειγμα, ένα φίλτρο μπορεί να ανιχνεύσει οριζόντιες ακμές αναγνωρίζοντας διαφορές στις τιμές των εικονοστοιχείων κατά μήκος αυτής της ακμής.

Σχήμα 2.13: Αναγνώριση Ακμών



Τα φίλτρα αναγνωρίζουν τις ακμές, οι οποίες συνδιάζονται ώστε να δημιουργηθεί ένα ορθογώνιο. Πηγή: [2].

Στα μεσαία επίπεδα, το δίκτυο συνδυάζει αυτά τα low-level χαρακτηριστικά για να σχηματίσει πιο σύνθετα σχήματα. Για παράδειγμα, ένα mid-level χαρακτηριστικό μπορεί να κατασκευάσει ένα εξάγωνο από ακμές, ενώ τα higher-level επίπεδα μπορούν να ανακατασκευάσουν αυτά τα εξάγωνα σε πιο πολύπλοκες δομές, όπως μια κυψέλη.

Το βάθος του δικτύου επηρεάζει σημαντικά την ιεραρχική εξαγωγή χαρακτηριστικών. Αρχικά, ένα συνελικτικό επίπεδο μπορεί να ανιχνεύσει τοπικά χαρακτηριστικά, μόνο εντός του φίλτρου του. Ωστόσο, τα επόμενα επίπεδα μπορούν να συνδυάσουν αυτές τις μικρές επιφάνειες για να αναγνωρίσουν μεγαλύτερα και πιο γενικά μοτίβα. Τα χαρακτηριστικά που μαθαίνονται στα αρχικά επίπεδα ενσωματώνονται με σημασιολογικά ουσιαστικούς τρόπους για να καταγράψουν σύνθετα οπτικά στοιχεία.

Η αποτελεσματικότητα των CNNs στην αναγνώριση εικόνας, όπως καταδεικνύεται στους πρόσφατους διαγωνισμούς ImageNet, εξαρτάται σε μεγάλο βαθμό από το βάθος του δικτύου. Τα βαθύτερα δίκτυα μπορούν να μάθουν περισσότερες ιεραρχικές σχέσεις στις εικόνες, ενισχύοντας την ικανότητά τους να αναγνωρίζουν σημασιολογικά συναφείς οντότητες. Η φύση των χαρακτηριστικών που μαθαίνονται επηρεάζεται επίσης από το σύνολο δεδομένων που χρησιμοποιείται για την εκπαίδευση.

2.2.3 Επισκόπηση της Εκπαίδευσης

Εκπαίδευοντας τα CNNs Η εκπαίδευση ενός CNN περιλαμβάνει αρκετά σημαντικά βήματα, αποσκοπώντας στη βελτιστοποίηση της απόδοσης του μοντέλου για μια δεδομένη διεργασία, όπως η ταξινόμηση εικόνων ή η ανίχνευση αντικειμένων. Η διαδικασία ξεκινά με την αρχικοποίηση των βαρών, χρησιμοποιώντας συχνά τεχνικές όπως η αρχικοποίηση Glorot ή He, ώστε να διασφαλιστεί η σταθερότητα και η ταχύτητα σύγκλισης. Στη συνέχεια, τα δεδομένα εκπαίδευσης τροφοδοτούνται στο δίκτυο σε παρτίδες (batches), πρακτική γνωστή ως «mini-batching», η οποία συμβάλλει στην περαιτέρω σταθεροποίηση και επιτάχυνση της διαδικασίας εκπαίδευσης.

> *Η εκπαίδευση των CNNs περιλαμβάνει δύο κύριες φάσεις διάδοσης: τη φάση προώθησης (forward propagation) και τη φάση οπίσθιας διάδοσης (backward propagation), όπως και στα MLPs.*

Φάση Προώθησης Κατά την προώθηση, τα δεδομένα εισόδου διακινούνται επίπεδο προς επίπεδο μέσα στο δίκτυο. Κάθε εικόνα εισόδου διέρχεται από μια σειρά συνελικτικών επιπέδων, όπου διάφορα φίλτρα ενεργοποίησης ολισθαίνουν επί της εισόδου προκειμένου να παράξουν χάρτες χαρακτηριστικών που αποτυπώνουν διάφορες πτυχές της εικόνας, όπως ακμές ή υφές. Στη συνέχεια, οι χάρτες χαρακτηριστικών διέρχονται από συναρτήσεις ενεργοποίησης τύπου ReLU για την εισαγωγή μη γραμμικότητας. Τα επίπεδα pooling μειώνουν τη διάσταση των χαρτών χαρακτηριστικών, διατηρώντας τις πλέον σημαντικές πληροφορίες, ενώ περιορίζουν την υπολογιστική πολυπλοκότητα. Αυτή η διαδικασία συνεχίζεται μέχρι τα πλήρως συνδεδεμένα επίπεδα, τα οποία ανασυνθέτουν τα επεξεργασμένα χαρακτηριστικά και παράγουν την τελική έξοδο, δηλαδή πιθανότητες κατηγοριών στην περίπτωση διεργασιών ταξινόμησης. Η προβλεπόμενη έξοδος συγκρίνεται με τις πραγματικές ετικέτες χρησιμοποιώντας μια συνάρτηση απώλειας, και συγκεκριμένα της εντροπικής απώλειας για εργασίες ταξινόμησης, η οποία ποσοτικοποιεί τη διαφορά μεταξύ των προβλεπόμενων και των πραγματικών εξόδων.

Φάση Οπίσθιας Διάδοσης Η οπίσθια διάδοση (backward propagation ή backpropagation), είναι η διαδικασία ρύθμισης των βαρών του δικτύου για την ελαχιστοποίηση της συνάρτησης απώλειας. Κατόπιν μίας φάσης προώθησης, υπολογίζεται η παράγωγος της συνάρτησης απώλειας ως προς κάθε μεταβλητή βάρους, με εφαρμογή του κανόνα της αλυσίδας. Στη συνέχεια, οι εξαχθείσες παράγωγοι χρησιμοποιούνται για την ενημέρωση των βαρών μέσω ενός αλγόριθμου βελτιστοποίησης όπως η στοχαστική κάθοδος κλίσης (stochastic gradient descent) ή ο Adam. Η εν λόγω επαναληπτική διαδικασία συνεχίζεται μέχρι να πραγματοποιηθεί σύγκλιση σε ένα σύνολο βαρών τα οποία ελαχιστοποιούν τη συνάρτηση απώλειας, οδηγώντας σε ένα ιδανικό μοντέλο που γενικεύει επιτυχώς σε νέα, άγνωστα δεδομένα.

> Η συνέργεια των δύο φάσεων επάγει την εκμάθηση πολύπλοκων μοτίβων και χαρακτηριστικών από τα δεδομένα, καθιστώντας τα CNNs ισχυρά εργαλεία για ένα ευρύ φάσμα εργασιών όρασης υπολογιστών.

2.3 Μετασχηματιστές

Οι μετασχηματιστές (transformers) προτάθηκαν από ερευνητές της Google το 2017 στο άρθρο "Attention Is All You Need" [13] και αποτελούν μια σημαντική εξέλιξη στη βαθιά μάθηση, ιδιαίτερα στον τομέα επεξεργασίας φυσικής γλώσσας (natural language processing - NLP). Εδώ, οι μετασχηματιστές χρησιμοποιούνται για τη συμβατική διεργασία της ταξινόμησης και συγκεκριμένα, επιλέγεται το μοντέλο TabNet [14].

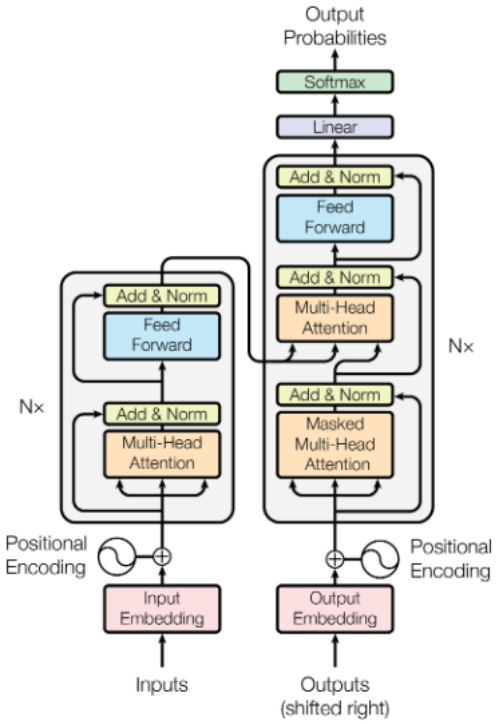
Ο μετασχηματιστής αντικαθιστά τα παραδοσιακά επαναλαμβανόμενα (recurrent) ή συνελικτικά (convolutional) νευρωνικά δίκτυα με έναν μηχανισμό αυτοπροσοχής (self-attention), υποστηρίζοντας επαυξημένη παράλληλη επεξεργασία και αποδοτικότητα. Αυτή η καινοτομία έχει επιφέρει επαναστατική πρόοδο σε NLP διεργασίες όπως η αυτόματη μετάφραση (machine translation), η ανάλυση συναισθημάτων (sentiment analysis) και απάντηση ερωτήσεων (question answering), αλλά έχει επίσης εφαρμοστεί και σε άλλους τομείς, όπως η όραση υπολογιστών.

2.3.1 Εισαγωγή

Στον πυρήνα της αρχιτεκτονικής των μετασχηματιστών βρίσκεται ο μηχανισμός **αυτοπροσοχής** (self-attention), ο οποίος επιτρέπει στο μοντέλο να σταθμίζει τη σχετική σημασία διαφορετικών λέξεων ή στοιχείων σε μια ακολουθία. Αυτή η προσέγγιση επιτρέπει στο μοντέλο να καταγράφει εξαρτήσεις και σύνθετες σχέσεις μέσα σε δεδομένα αποτελεσματικότερα από τους προκατόχους του.

Η πρωταρχική αρχιτεκτονική του μετασχηματιστή συνίσταται από μια δομή **κωδικοποιητή-αποκωδικοποιητή** (encoder-decoder), όπως απεικονίζεται στο σχήμα 2.14. Οι δύο αυτές συνιστώσες αποτελούνται από στοιβαγμένα επίπεδα αυτοπροσοχής και νευρωνικών δικτύων εμπρόσθιας τροφοδοσίας. Ο κωδικοποιητής επεξεργάζεται τα δεδομένα εισόδου παράγοντας ένα σύνολο συνεχών αναπαραστάσεων, ενώ ο αποκωδικοποιητής χρησιμοποιεί αυτές τις αναπαραστάσεις για να παράγει την επιθυμητή έξοδο, όπως μια μεταφρασμένη πρόταση.

Σχήμα 2.14: Η πρωταρχική αρχιτεκτονική μετασχηματιστή



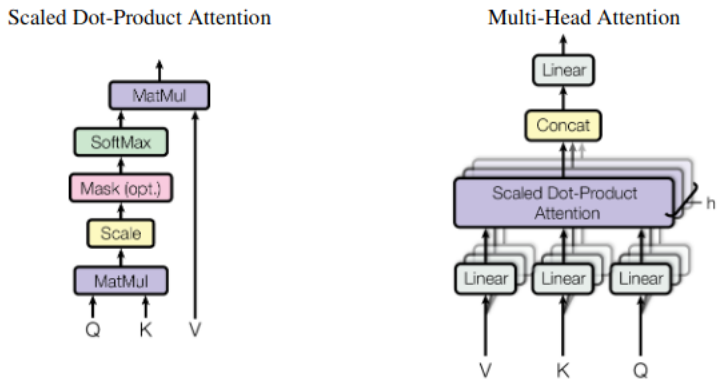
Ο κωδικοποιητής στην αριστερή πλευρά και ο αποκωδικοποιητής στη δεξιά. Πηγή: [13].

Η επιτυχία των μετασχηματιστών στο NLP ανάγεται στην ανάπτυξη διαφόρων μοντέλων, όπως το BERT (Bidirectional Encoder Representations from Transformers), το GPT (Generative Pre-trained Transformer), και το T5 (Text-To-Text Transfer Transformer). Αυτά τα μοντέλα έχουν θέσει νέα πρότυπα απόδοσης σε πολυάριθμες διεργασίες NLP, αποδεικνύοντας την ευελιξία και την ισχύ της εν λόγω αρχιτεκτονικής.

2.3.2 Attention Is All You Need

Το κομβικό άρθρο "Attention Is All You Need" από τους Vaswani et al. (2017) εισάγει την έννοια του μετασχηματιστή και αναδιαρθρώνει πλήρως τον τρόπο με τον οποίο τα διαδοχικά δεδομένα επεξεργάζονται στο πλαίσιο της βαθιάς μάθησης. Το άρθρο αυτό παρουσιάζει τον μηχανισμό αυτοπροσοχής ως βασική συνιστώσα της εν λόγω αρχιτεκτονικής, ο οποίος επιτρέπει στο μοντέλο να εστιάζει σε διαφορετικά μέρη της ακολουθίας εισόδου κατά την πρόβλεψη, καταγράφοντας έτσι περίπλοκα μοτίβα και εξαρτήσεις. Ακολουθεί μια περιγραφή των κύριων συνεισφορών του καθοριστικού αυτού άρθρου.

Σχήμα 2.15: Απεικόνιση του μηχανισμού προσοχής



(αριστερά): Προσοχή κλιμακωμένου εσωτερικού γινομένου (scaled dot-product attention) *(δεξιά):* Η προσοχή πολλαπλών κεφαλών (multi-head attention) αποτελείται από διάφορα επίπεδα προσοχής τα οποία λειτουργούν παράλληλα. MatMul: *Matrix Multiplication - Πολλαπλασιασμός Πινάκων,* Concat: *Concatanate - Συνένωση.*

Ο μηχανισμός αυτοπροσοχής Ο μηχανισμός αυτοπροσοχής επιτρέπει σε κάθε στοιχείο της ακολουθίας εισόδου να αλληλεπιδρά με κάθε άλλο στοιχείο, υπολογίζοντας έτσι ένα σταθμισμένο άθροισμα που καθορίζει τη σχετική σημασία των χαρακτηριστικών ως προς τα υπόλοιπα. Μαθηματικά, η έννοια αυτή υπολογίζεται με χρήση τελεστών γραμμικής άλγεβρας και οι τιμές της καθορίζονται μέσω μιας συνάρτησης softmax που εφαρμόζεται στο εσωτερικό γινόμενο των διανυσμάτων του **ερωτήματος** (query), του **κλειδιού** (key) και της τιμής (value), [13]:

$$\text{Attention}\,(Q, K, V) = \text{softmax}\left(\frac{Q \cdot K^T}{\sqrt{d_k}}\right) \cdot V$$

Εδώ, τα Q , K , και V είναι προβολές της ακολουθίας εισόδου και αντιπροσωπεύουν τους αντίστοιχους πίνακες ερωτήματος, κλειδιού και τιμής, ενώ η παράμετρος d_k υποδηλώνει τη διαστατικότητα των διανυσμάτων κλειδιού. Όπως φαίνεται και στο σχήμα 2.15, η πράξη $\cdot$ αντιστοιχεί στον πολλαπλασιασμό πινάκων.

Κωδικοποίηση Θέσης Προκειμένου να ενσωματωθεί κατάλληλα η σειρά της ακολουθίας εισόδου, η αρχιτεκτονική μετασχηματιστή προσθέτει **κωδικοποιήσεις θέσης** (positional encoding) στα embeddings[4] εισόδου. Αυτές οι κωδικοποιήσεις παρέχουν μοναδικές πληροφορίες θέσης σε κάθε στοιχείο της ακολουθίας, έτσι ώστε να αναπτυχθεί η δυνατότητα διάκρισης διαφορετικών θέσεων. Αυτή η ιδιότητα είναι απαραίτητη ειδικά για διεργασίες που μεταχειρίζονται κείμενο, στο οποίο η σχετική θέση των λέξεων επηρεάζει ουσιαστικά το νόημα των προτάσεων.

Προσοχή Πολλαπλών Κεφαλών Το μοντέλο χρησιμοποιεί πολλαπλούς μηχανισμούς αυτοπροσοχής, γνωστούς ως multi-head attention, προκειμένου να καταγράψει τις διαφορετικές πτυχές των σχέσεων μεταξύ στοιχείων της υπό επεξεργασία ακολουθίας. Κάθε επιμέρους συνιστώσα υπολογίζει ανεξάρτητα βαθμολογίες προσοχής και οι έξοδοί τους συνενώνονται και μετασχηματίζονται γραμμικά, όπως απεικονίζεται στο σχήμα 2.15.

Αρχιτεκτονική Κωδικοποιητή-Αποκωδικοποιητή Ο μετασχηματιστής αποτελείται από μια αρχιτεκτονική κωδικοποιητή-αποκωδικοποιητή, όπου και τα δύο συστατικά αποτελούνται από πολλά πανομοιότυπα επίπεδα. Κάθε επίπεδο περιλαμβάνει έναν μηχανισμό προσοχής πολλαπλών κεφαλών και ένα νευρωνικό δίκτυο εμπρόσθιας τροφοδότησης (position-wise feed-forward network). Συνοπτικά, ο κωδικοποιητής επεξεργάζεται την ακολουθία εισόδου και παράγει μια συνεχή αναπαράσταση, την οποία ο αποκωδικοποιητής χρησιμοποιεί, μαζί με τις προηγουμένως παραγόμενες εξόδους, για να παράγει την τελική ακολουθία.

Κανονικοποίηση Επιπέδου Κάθε υποεπίπεδο στον κωδικοποιητή και τον αποκωδικοποιητή ακολουθείται από **κανονικοποίηση επιπέδου** (layer normalization).

2.3.3 Η Αρχιτεκτονική TabNet

Η αρχιτεκτονική TabNet είναι ένας τύπος μετασχηματιστή προσανατολισμένος για χρήση δεδομένων σε μορφή πίνακα.

2.3.3.1 Εισαγωγή στην Αρχικτεκτονική TabNet για Δυαδική Ταξινόμηση

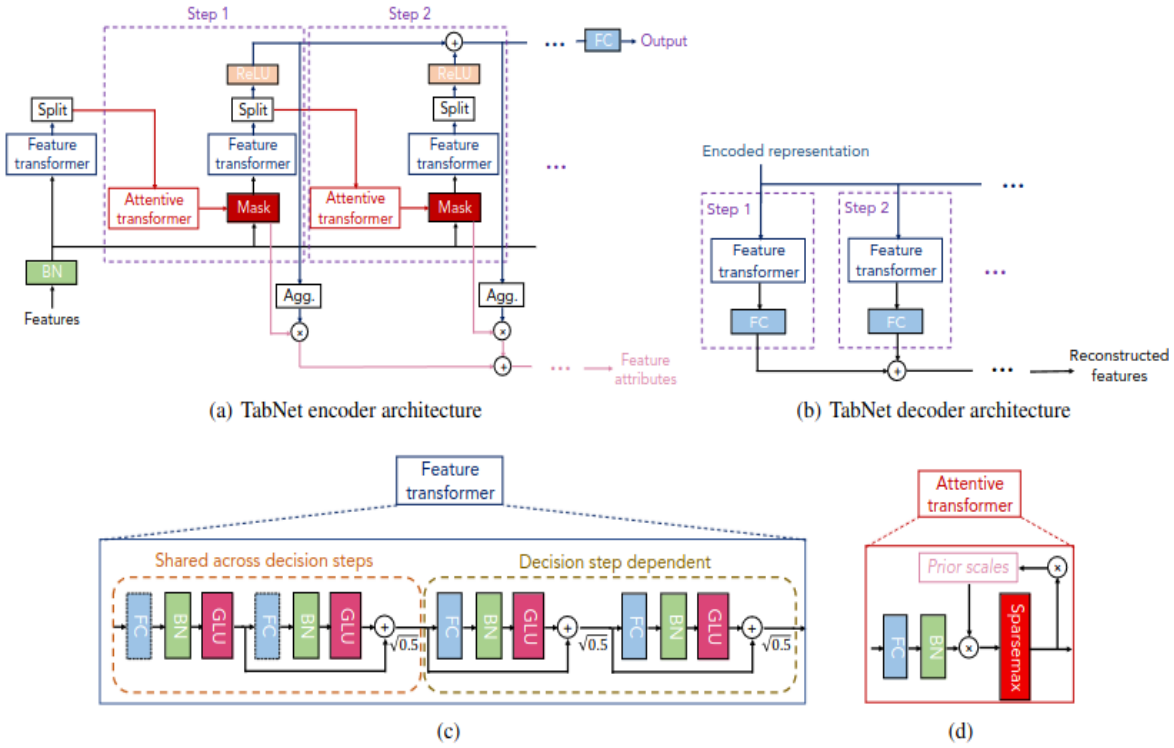
Για τη διεργασία ταξινόμησης στα πλαίσια της ανίχνευσης εισβολών, επιλέγεται το μοντέλο TabNet, σύντμηση του Tabular Network, μια αρχιτεκτονική βαθιάς μάθησης που έχει σχεδιαστεί ειδικά για δεδομένα σε **μορφή πινάκων** (tabular data), όπως το σύνολο δεδομένων CIC-IDS-2017. Η αρχιτεκτονική του TabNet προέρχεται από την οικογένεια των μετασχηματιστών, γνωστοί για την αποτελεσματικότητά τους στην αποτύπωση σύνθετων σχέσεων μέσα σε δεδομένα υψηλών διαστάσεων (high-dimensional data) μέσω του διάσημου μηχανισμού **προσοχής** (attention). Αυτό το μοντέλο παρουσιάστηκε το 2019 από ερευνητές της Google, [14].

Εκτός από την αυτοπροσοχή, το TabNet εισάγει νέες λειτουργίες όπως η αραιή (sparse) επιλογή χαρακτηριστικών και η διαδοχική απόκρυψη χαρακτηριστικών (feature masking), βελτιώνοντας περαιτέρω την απόδοσή του στην ταξινόμηση. Αυτές οι καινοτομίες ενισχύουν την απόδοσή του στην ταξινόμηση δεδομένων πίνακα, επιτρέποντας στο TabNet να επικεντρώνεται

[4]Τα embeddings (μεταφράση: ενσωματώσεις) είναι πυκνές αναπαραστάσεις δεδομένων (όπως λέξεις, εικόνες ή αντικείμενα) σε διανύσματα, όπου παρόμοιες οντότητες έχουν παρόμοιες αναπαραστάσεις. Στην επεξεργασία φυσικής γλώσσας, τα embeddings λέξεων χαρτογραφούν λέξεις ή φράσεις σε συνεχείς διανυσματικούς χώρους, καταγράφοντας σημασιολογικές έννοιες και σχέσεις. Αυτό επιτρέπει στα μοντέλα να επεξεργάζονται και να κατανοούν το κείμενο πιο αποτελεσματικά, μετατρέποντας τις διακριτές λέξεις σε μια αριθμητική μορφή που διατηρεί το συμφραζόμενο νόημά τους. Δημοφιλείς τεχνικές ενσωμάτωσης περιλαμβάνουν τα Word2Vec, GloVe και BERT.

επιλεκτικά στα σχετικά μέρη των δεδομένων, βελτιώνοντας τόσο την ερμηνευσιμότητα όσο και την ακρίβεια.

Σχήμα 2.16: Απεικόνιση της εσωτερικής αρχιτεκτονικής του TabNet



(a) TabNet encoder architecture (b) TabNet decoder architecture

(c) (d)

Figure 4: (a) TabNet encoder, composed of a feature transformer, an attentive transformer and feature masking. A split block divides the processed representation to be used by the attentive transformer of the subsequent step as well as for the overall output. For each step, the feature selection mask provides interpretable information about the model's functionality, and the masks can be aggregated to obtain global feature important attribution. (b) TabNet decoder, composed of a feature transformer block at each step. (c) A feature transformer block example – 4-layer network is shown, where 2 are shared across all decision steps and 2 are decision step-dependent. Each layer is composed of a fully-connected (FC) layer, BN and GLU nonlinearity. (d) An attentive transformer block example – a single layer mapping is modulated with a prior scale information which aggregates how much each feature has been used before the current decision step. sparsemax (Martins and Astudillo 2016) is used for normalization of the coefficients, resulting in sparse selection of the salient features.

Προσοχή: Ο αποκωδικοποιητής του σχήματος (b) απευθύνεται αποκλειστικά στην διεργασία της μη επιβλεπόμενης προ-εκπαίδευσης (unsupervised pre-training) και δεν αποτελεί θεμελιώδη συνιστώσα της υλοποίησης του μοντέλου. Πηγή: το πρωτότυπο TabNet άρθρο, [14].

2.3.3.2 Επισκόπηση της Αρχικτεκτονικής

Θεμελιώδης Δομή Το μοντέλο TabNet χρησιμοποιεί μια αρχιτεκτονική **αποκλειστικού κωδικοποιητή** (encoder-only architecture). Αυτό σημαίνει ότι εστιάζει αποκλειστικά στην εξαγωγή κατατοπιστικών χαρακτηριστικών από τα δεδομένα και δεν απαιτεί ξεχωριστό αποκωδικοποιητή για την παραγωγή εξόδων. Η αρχιτεκτονική αποκλειστικού κωδικοποιητή βασίζεται σε μια σειρά διαδοχικών βημάτων - μετασχηματισμών για την εξαγωγή διαφωτιστικών χαρακτηριστικών από τα δεδομένα, μαθαίνοντας προοδευτικά εξελισσόμενες αναπαραστάσεις των δεδομένων εισόδου. Αρχικά, τα ακατέργαστα χαρακτηριστικά μετασχηματίζονται χρησιμοποιώντας τεχνικές **κλιμάκωσης** (scaling) και **κανονικοποίησης** (normalization). Στη συνέχεια, το TabNet αποτελείται από δύο βασικές συνιστώσες: έναν μετασχηματιστή χαρακτηριστικών (feature transformer) και έναν μετασχηματιστή προσοχής (attention transformer). Ο μετασχηματιστής χαρακτηριστικών χρησιμοποιεί έναν αραιό μηχανισμό προσοχής για να εντοπίσει τα πιο σχετικά χαρακτηριστικά για κάθε σημείο δεδομένων. Εν συνεχεία, ο μετασχη-

ματιστής προσοχής χρησιμοποιεί τον μηχανισμό αυτοπροσοχής για να καταγράψει σύνθετες σχέσεις μεταξύ αυτών των επιλεγμένων χαρακτηριστικών.

Αυτοπροσοχή Στον πυρήνα του TabNet βρίσκεται ο μηχανισμός αυτοπροσοχής, ένα ισχυρό εργαλείο για την καταγραφή σχέσεων μέσα σε δεδομένα υψηλών διαστάσεων, όπως περιγράφεται στην προηγούμενη παράγραφο. Αυτός ο μηχανισμός επιτρέπει στο TabNet να προσαρμόζεται δυναμικά σε διαφορετικά μέρη του χώρου εισόδου, καταγράφοντας αποτελεσματικά σύνθετα μοτίβα και εξαρτήσεις μέσα στα δεδομένα. Αξιοποιώντας την αυτοπροσοχή, το TabNet μπορεί να δίνει προτεραιότητα στα πιο κατατοπιστικά χαρακτηριστικά, βελτιώνοντας τόσο την ερμηνευσιμότητα όσο και την ακρίβεια των προβλέψεών του. Η ενσωμάτωση της αραιής επιλογής και της διαδοχικής απόκρυψης χαρακτηριστικών βελτιώνει περαιτέρω τη διαδικασία, διασφαλίζοντας ότι το μοντέλο εστιάζει στις πιο σχετικές πτυχές των δεδομένων για κάθε πρόβλεψη.

2.3.3.3 Επιλογή Χαρακτηριστικών

Αραιή Επιλογή Χαρακτηριστικών Το TabNet χρησιμοποιεί μια εξελιγμένη προσέγγιση για την επιλογή χαρακτηριστικών, η οποία ενισχύει τόσο την ερμηνευσιμότητα όσο και την αποδοτικότητα του μοντέλου. Αυτή η προσέγγιση επιτρέπει στο μοντέλο να επιλέγει δυναμικά ένα υποσύνολο σχετικών χαρακτηριστικών για κάθε σημείο δεδομένων, αντί να χρησιμοποιεί το σύνολο όλων των χαρακτηριστικών, μειώνοντας έτσι την πολυπλοκότητα και τον κίνδυνο **υπερεκπαίδευσης** (overfitting).

Μηχανισμός Αραιής Επιλογής Χαρακτηριστικών Η αραιή επιλογή χαρακτηριστικών στο TabNet επιτυγχάνεται κυρίως μέσω της χρήσης μασκών (masks) και τεχνικών αραιής κανονικοποίησης, [15]. Ακολουθεί μια λεπτομερής εξήγηση της μαθηματικής περιγραφής:

1. **Μάσκες Επιλογής Χαρακτηριστικών** (Feature Selection Masks): Το TabNet χρησιμοποιεί έναν διαδοχικό μηχανισμό προσοχής για τη δημιουργία μασκών επιλογής χαρακτηριστικών. Αυτές οι μάσκες καθορίζουν σε ποια χαρακτηριστικά θα επικεντρωθεί το μοντέλο σε κάθε βήμα απόφασης. Δεδομένου ενός διανύσματος χαρακτηριστικών εισόδου x, η μάσκα M δημιουργείται ως εξής:

$$M\left(x\right) = \text{sparsemax}\left(U\left(x\right)\right)$$

   Εδώ, η $U\left(x\right)$ είναι ένας γραμμικός μετασχηματισμός των χαρακτηριστικών εισόδου, προερχόμενος από μάθηση και η συνάρτηση Sparsemax είναι μια συνάρτηση κανονικοποίησης που προωθεί την αραιότητα εξαναγκάζοντας πολλές από τις εξόδους της στις μηδενικές τιμές.

2. **Συνάρτηση Sparsemax**: Η συνάρτηση sparsemax, η οποία προτάθηκε από τους Martins & Astudillo [15], είναι μία γενίκευση της softmax που επιφέρει **αραιότητα** (sparsity). Για ένα διάνυσμα $z \in \mathbb{R}^d$, η sparsemax προβάλλει το z επί του probability simplex:

$$\text{sparsemax}\left(z\right) = \underset{p \in \Delta^{d-1}}{\arg\min} \left\|p - z\right\|^2$$

   όπου Δ^{d-1} είναι το $(d-1)$-διάστατο probability simplex (μία γεωμετρική αναπαράσταση του συνόλου όλων των δυνατών κατανομών πιθανότητας επί ενός πεπερασμένου συνόλου αποτελεσμάτων):

$$\Delta^{d-1} = \left\{ p \in \mathbb{R}^d \mid (\forall 1 \leq i \leq d)) : [p_i \geq 0] \wedge \left[\sum_{i=1}^{d} p_i = 1 \right] \right\}$$

Αυτό έχει ως αποτέλεσμα μια αραιή κατανομή πιθανότητας επί των χαρακτηριστικών, επιλέγοντας ουσιαστικά ένα υποσύνολο αυτών.

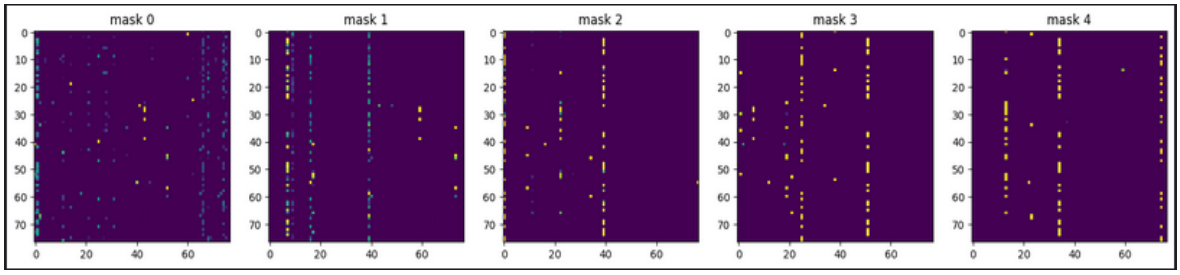
3. **Αραιή Κανονικοποίηση** (Sparse Regularization): Προκειμένου να ενισχυθεί περαιτέρω η αραιότητα στην επιλογή χαρακτηριστικών, το TabNet ενσωματώνει έναν όρο ποινής στη συνάρτηση απώλειας. Αυτή η ποινή είναι συνήθως μια μορφή κανονικοποίησης L_1 που εφαρμόζεται στις μάσκες επιλογής χαρακτηριστικών:

$$\mathcal{L}_{\text{sparse}} = \lambda \sum_{i=1}^{T} \|M_i\|_1$$

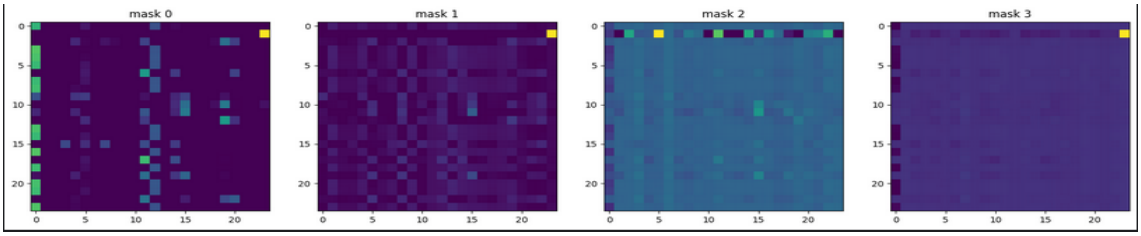
όπου λ είναι μια παράμετρος κανονικοποίησης, T είναι ο αριθμός των βημάτων απόφασης, και η M_i είναι η εκάστοτε μάσκα στο βήμα i. Αυτή η ποινή αναγκάζει πολλά στοιχεία της M_i να είναι μηδενικά, επιλέγοντας έτσι μόνο λίγα σημαντικά χαρακτηριστικά.

Σχήμα 2.17: Μάσκες απόκρυψης χαρακτηριστικών

Στο TabNet, κάθε μάσκα αντιστοιχεί σε έναν μετασχηματιστή χαρακτηριστικών.



Επάνω: Μάσκες για έναν ταξινομητή με 5 μετασχηματιστές χαρακτηριστικών, ο καθένας εκ των οποίων έχει διαστατικότητα: $n_a = 77$. Κάτω: Μάσκες για έναν ταξινομητή με 4 μετασχηματιστές χαρακτηριστικών, ο καθένας εκ των οποίων έχει διαστατικότητα: $n_a = 24$.



Ο μηχανισμός προσοχής στο TabNet δημιουργεί μάσκες που καθορίζουν ποια χαρακτηριστικά επιλέγονται και παρακολουθούνται σε κάθε βήμα απόφασης. Επιπλέον, η ποικιλία των χρωμάτων αντικατοπτρίζει την **ομαλότητα** (όχι δυαδικότητα) της διαδικασίας.

Διαστατικότητα των Μασκών (Dimensionality of Masks): Οι μάσκες δημιουργούνται από τον μηχανισμό προσοχής, ο οποίος χρησιμοποιεί embeddings προσοχής μεγέθους n_a. Επομένως, η διάσταση του επιπέδου προσοχής (n_a) επηρεάζει άμεσα την πολυπλοκότητα και την ικανότητα της διαδικασίας δημιουργίας μασκών. Αυτές οι μάσκες στη συνέχεια εφαρμόζονται στα μετασχηματισμένα χαρακτηριστικά, ελέγχοντας ποια χαρακτηριστικά περνούν στο επόμενο επίπεδο (μετασχηματιστής χαρακτηριστικών ή βήμα απόφασης).

Ο μηχανισμός επιλογής χαρακτηριστικών του TabNet ενισχύει την ερμηνευσιμότητα επικεντρώνοντας σε μικρό αριθμό χαρακτηριστικών και καθιστώντας πιο εύκολη την κατανόηση των

χαρακτηριστικών που οδηγούν στις προβλέψεις. Αυτή η αραιή επιλογή βελτιώνει επίσης την αποδοτικότητα μειώνοντας το υπολογιστικό φορτίο κατά την εκπαίδευση. Επιπλέον, αγνοώντας τα άσχετα ή πλεονάζοντα χαρακτηριστικά, το μοντέλο μετριάζει την υπερεκπαίδευση, οδηγώντας σε καλύτερα μοντέλα.

Η Ομαλή Διαδικασία Επιλογής Χαρακτηριστικών Η ομαλή διαδικασία επιλογής χαρακτηριστικών του TabNet έχει σχεδιαστεί για να εξασφαλίζει μια σταδιακή και συνεκτική μετάβαση μεταξύ των επιλεγμένων χαρακτηριστικών, η οποία επιτυγχάνεται μέσω συνεχών βαρών προσοχής αντί για δυαδικές αποφάσεις. Ειδικότερα:

1. **Ομαλές Μάσκες** (Smooth Masks) : Αντί για "απότομες" δυαδικές μάσκες, το TabNet χρησιμοποιεί ομαλές μάσκες με συνεχείς τιμές που επιτρέπουν διάφορους βαθμούς σημαντικότητας χαρακτηριστικών. Αυτός ο ομαλός μηχανισμός προσοχής αναπαρίσταται ως:

$$M_t(x) = \text{softmax}(U_t(x))$$

   όπου $U_t(x)$ είναι ένας μετασχηματισμός κατά το βήμα t.

2. **Σταδιακή Σημαντικότητα Χαρακτηριστικών** (Gradual Feature Importance): Οι συνεχείς τιμές στις ομαλές μάσκες επιτρέπουν στο μοντέλο να προσαρμόζει τη σημαντικότητα των χαρακτηριστικών σταδιακά με την πάροδο του χρόνου, αποτρέποντας απότομες αλλαγές που θα μπορούσαν να αποσταθεροποιήσουν τη διαδικασία μάθησης.

Σχήμα 2.18: Απεικόνιση της διαδικασίας επιλογής χαρακτηριστικών

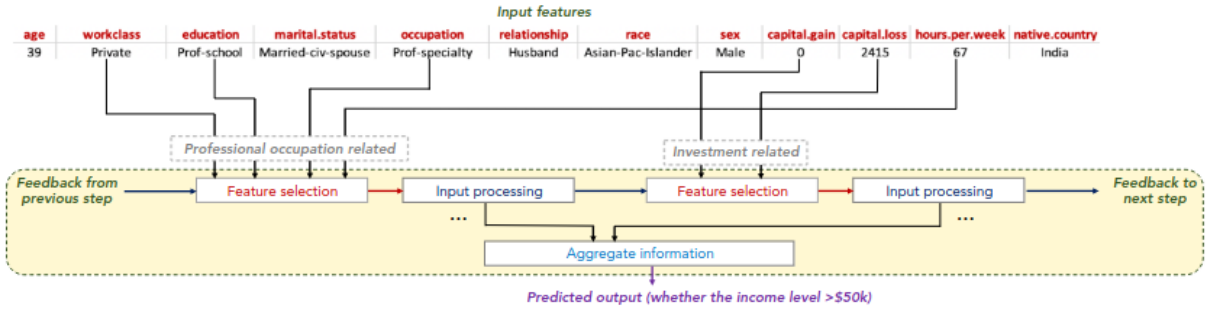


Figure 1: TabNet's sparse feature selection exemplified for Adult Census Income prediction (Dua and Graff 2017). Sparse feature selection enables interpretability and better learning as the capacity is used for the most salient features. TabNet employs multiple decision blocks that focus on processing a subset of input features for reasoning. Two decision blocks shown as examples process features that are related to professional occupation and investments, respectively, in order to predict the income level.

Η προσέγγιση του TabNet για την επιλογή χαρακτηριστικών συνδυάζει αραιές και ομαλές διαδικασίες επιλογής. Αξιοποιούνται προχωρημένοι μηχανισμοί κανονικοποίησης και προσοχής ώστε να ενισχυθεί η ερμηνευσιμότητα και η απόδοση του μοντέλου. Ο συνδυασμός της sparsemax για την πρόκληση αραιότητας και της softmax για ομαλές μεταβάσεις εξασφαλίζει ότι το TabNet μπορεί δυναμικά και αποτελεσματικά να επιλέγει τα πιο σχετικά χαρακτηριστικά για κάθε πρόβλεψη, οδηγώντας σε robust και κατανοητά μοντέλα. Πηγή: [14].

Η ομαλή διαδικασία επιλογής χαρακτηριστικών στο TabNet βελτιώνει τη σταθερότητα αποτρέποντας ξαφνικές αλλαγές στη σημαντικότητα των χαρακτηριστικών, εξασφαλίζοντας συνεπή μάθηση. Επιτρέπει επίσης μια λεπτομερή εξέταση του χώρου των χαρακτηριστικών, καταγράφοντας μοτίβα με συνεχείς τιμές προσοχής. Αυτή η προσέγγιση ενισχύει την ερμηνευσιμότητα παρέχοντας μια σαφή εικόνα για τον τρόπο αλλαγής της εστίασης του μοντέλου με την πάροδο του χρόνου.

2.3.3.4 Διαδοχική Προσοχή

Το TabNet χρησιμοποιεί **διαδοχική προσοχή** (sequential attention) ώστε να βελτιώνει την εστίασή του σε σχετικά χαρακτηριστικά μέσω μιας σειράς βημάτων απόφασης. Με αυτόν τον τρόπο επιτυγχάνεται δυναμική προσαρμογή της προσοχής σε διαφορετικά χαρακτηριστικά, αντί για μία μόνο προσπέλαση των δεδομένων. Επιμέρους ανάλυση της διαδοχικής προσοχής:

1. **Αρχικός Μετασχηματισμός**: Η διαδικασία ξεκινά με τα ακατέργαστα χαρακτηριστικά εισόδου x να μετασχηματίζονται σε έναν χώρο υψηλότερης διάστασης χρησιμοποιώντας έναν μετασχηματιστή χαρακτηριστικών.

2. **Βήματα Προσοχής** (Attention Steps): Σε κάθε βήμα απόφασης t, το μοντέλο παράγει μία βαθμολογία προσοχής (attention score) για κάθε χαρακτηριστικό και δημιουργεί μια νέα μάσκα επιλογής χαρακτηριστικών (ήτοι έναν γραμμικό μετασχηματισμό), ως προϊόν μάθησης.

3. **Βελτίωση Χαρακτηριστικών** (Feature Refinement): Στη συνέχεια, τα επιλεγμένα χαρακτηριστικά επεξεργάζονται από τον μετασχηματιστή χαρακτηριστικών, παράγοντας μια βελτιωμένη αναπαράσταση. Το μοντέλο επαναλαμβάνει τη διαδικασία μέσω πολλών βημάτων απόφασης, κάθε φορά ενημερώνοντας τη μάσκα επιλογής και βελτιώνοντας την τρέχουσα αναπαράσταση, εφαρμόζοντας τη μάσκα M_t στην αναπαράσταση των χαρακτηριστικών x_t:

$$x_{t+1} = \text{FeatureTransformer}\left(M_t \odot x_t\right)$$

όπου $\odot$ υποδηλώνει πολλαπλασιασμό πινάκων κατά στοιχείο (element-wise).

Τα Πλεονεκτήματα της Διαδοχικής Προσοχής Η διαδοχική προσοχή επιτρέπει στο TabNet να προσαρμόζει δυναμικά τη σημαντικότητα των χαρακτηριστικών σε κάθε βήμα, αυξάνοντας την ευελιξία και την ουσιαστικότητα των διεργασιών πρόβλεψης. Αυτή η προσέγγιση βελτιώνει την ερμηνευσιμότητα καθιστώντας τη διαδικασία λήψης αποφάσεων του μοντέλου πιο διαφανή (transparent), αυξάνει τη σταθερότητα (robustness) βελτιώνοντας την επιλογή χαρακτηριστικών και αντιμετωπίζοντας τον θόρυβο, και τέλος βελτιώνει την απόδοση εστιάζοντας τους υπολογιστικούς πόρους στα πλέον κατατοπιστικά χαρακτηριστικά - κάτι που είναι ιδιαίτερα ωφέλιμο για σύνολα δεδομένων υψηλών διαστάσεων.

Αυτή η καινοτόμος προσέγγιση επιτρέπει στο TabNet να χειρίζεται αποτελεσματικά σύνθετα δεδομένα σε μορφή πινάκων, καθιστώντας το ένα ισχυρό εργαλείο για διάφορα καθήκοντα μηχανικής και βαθιάς μάθησης.

Κεφάλαιο 3

Επισκόπηση Μετρικών Ταξινόμησης

Μετά την επιτυχή ανάπτυξη και εκπαίδευση ταξινομητών, είναι απαραίτητη η αξιολόγησή τους. Συνεπώς, θα πρέπει να οριστούν μετρικές αξιολόγησης της απόδοσης και να προσαρμοστούν σε δυαδικούς (binary) ή πολυκατηγορικούς (multiclass) ταξινομητές. Η κεντρική έννοια της αξιολόγησης είναι ο πίνακας σύγχυσης (confusion matrix), από τον οποίο εξάγονται όλες οι μετρικές. Σε αυτό το κεφάλαιο, αρχικά παρουσιάζεται ο πίνακας σύγχυσης και στη συνέχεια εξάγονται διάφορες μετρικές, όπως accuracy, precision, recall και F_1-score, οι οποίες προσαρμόζονται κατάλληλα ως προς τη μορφή του προβλήματος ταξινόμησης. Ειδικά για την πολυκατηγορική ταξινόμηση παρουσιάζονται διάφορες παραλλαγές των μετρικών: micro, macro και weighted averaging. Τέλος, θα συζητηθούν προχωρημένες μετρικές, ειδικά προσαρμοσμένες για την διεργασία της ανίχνευσης εισβολών στο πλαίσιο ενός Συστήματος Ανίχνευσης Εισβολών.

3.1 Ο Πίνακας Σύγχυσης

Αρχικά, εισάγεται η βασική ορολογία του πίνακα σύγχυσης. Ο πίνακας σύγχυσης παρέχει μια συνολική εικόνα απόδοσης της ταξινόμησης, επιτρέποντας την λεπτομερή ανάλυση της απόδοσης του μοντέλου και την ανάδειξη προτύπων εσφαλμένων ταξινομήσεων. Για πρακτική κατανόηση των πινάκων σύγχυσης, συνιστάται η μελέτη των πηγών [16, 17], ενώ κατά την προγραμματιστική υλοποίηση έχει υιοθετηθεί η προσέγγιση scikit-learn [18].

3.1.1 Ο Πίνακας Σύγχυσης υπό Δυαδική Διάταξη

Σε μια δυαδική διάταξη, ο πίνακας σύγχυσης κατασκευάζεται ως εξής:

Πίνακας 3.1: Ο Πίνακας Σύγχυσης υπό Δυαδική Διάταξη

	Predicted Condition	
True/Actual Condition	Predicted Positive [PP]	Predicted Negative [PN]
Actual Positive (Class 0) [AP]	True Positive [TP]	False Negative [FN]
Actual Negative (Class 1) [AN]	False Positive [FP]	True Negative [TN]

Επεξήγηση του πίνακα σύγχυσης:

- **Predicted Positive** [PP]: Υπό πρόβλεψη θετικά δείγματα.
- **Predicted Negative** [PN]: Υπό πρόβλεψη αρνητικά δείγματα.
- **Actual Positive** [AP]: Πραγματικά θετικά δείγματα.

- **Actual Negative** [AN]: Πραγματικά αρνητικά δείγματα.

Εάν όλος ο πληθυσμός έχει μέγεθος N τότε θα πρέπει να ισχύει:

$$|[PP] \cup [PN]| = |[AP] \cup [AN]| = N \tag{3.1}$$

- **True Positive** [TP] $\stackrel{\text{def}}{=}$ [AP] $\cap$ [PP]: Ο αριθμός των ορθώς προβλεφθέντων θετικών δειγμάτων. Στο πλαίσιο του πίνακα σύγχυσης, η παράμετρος TP αναπαριστά τον αριθμό των δειγμάτων που ανήκουν σε μια συγκεκριμένη κατηγορία και ταξινομήθηκαν ορθά ως ανήκοντα σε αυτήν την κατηγορία.

- **False Positive** [FP] $\stackrel{\text{def}}{=}$ [AN] $\cap$ [PP]: Ο αριθμός των εσφαλμένα προβλεφθέντων θετικών δειγμάτων. Στο πλαίσιο του πίνακα σύγχυσης, η παράμετρος FP αναπαριστά τον αριθμό των δειγμάτων που δεν ανήκουν σε μια συγκεκριμένη κατηγορία, αλλά ταξινομήθηκαν εσφαλμένα ως ανήκοντα σε αυτήν την κατηγορία.

- **True Negative** [TN] $\stackrel{\text{def}}{=}$ [AN] $\cap$ [PN]: Ο αριθμός των ορθώς προβλεφθέντων αρνητικών δειγμάτων. Στο πλαίσιο του πίνακα σύγχυσης, η παράμετρος TN αναπαριστά τον αριθμό των δειγμάτων που δεν ανήκουν σε μια συγκεκριμένη κατηγορία και ταξινομήθηκαν ορθά ως μη ανήκοντα σε αυτήν την κατηγορία.

- **False Negative** [FN] $\stackrel{\text{def}}{=}$ [AP] $\cap$ [PN]: Ο αριθμός των εσφαλμένα προβλεφθέντων αρνητικών δειγμάτων. Στο πλαίσιο του πίνακα σύχυσης, η παράμετρος FN αναπαριστά τον αριθμό των δειγμάτων που ανήκουν σε μια συγκεκριμένη κατηγορία, αλλά ταξινομήθηκαν εσφαλμένα ως μη ανήκοντα σε αυτήν την κατηγορία.

Αυτή η ορολογία μπορεί να γενικευτεί για την πολυταξική διάταξη n κατηγοριών:

3.1.2 Ο Πίνακας Σύγχυσης υπό Πολυκατηγορική Διάταξη

Υπό πολυκατηγορική διάταξη, ο πίνακας σύγχυσης κατασκευάζεται ως εξής:

Πίνακας 3.2: Ο Πίνακας Σύγχυσης υπό Πολυκατηγορική Διάταξη

Multiclass Configuration	Predicted Class 1	Predicted Class 2	...	Predicted Class n
True Class 1	TP_{11}	FP_{12}	...	FP_{1n}
True Class 2	FP_{21}	TP_{22}	...	FP_{2n}
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
True Class n	FP_{n1}	FP_{n2}	...	TP_{nn}

Επεξήγηση του πίνακα σύγχυσης:

- **True Positive** [TP_{ii}]: Ο αριθμός των ορθώς προβλεφθέντων δειγμάτων της κατηγορίας i. Η παράμετρος TP_{ii} αναπαριστά τον αριθμό των δειγμάτων της κατηγορίας i, τα οποία ορθώς ταξινομήθηκαν ως ανήκοντα στην κατηγορία i.

- **False Positive** $[\text{FP}_{ij}], i \neq j$: Ο αριθμός των εσφαλμένα προβλεφθέντων δειγμάτων της κατηγορίας j, όταν η πραγματική κατηγορία προέλευσης είναι η i. Η παράμετρος FP_{ij} αναπαριστά τον αριθμό των δειγμάτων που προέρχονται από την κατηγορία i, αλλά εσφαλμένα ταξινομηθήκαν στην κατηγορία j.

Σημείωση: Οι όροι *True Negative & False Negative* δεν ορίζονται στην πολυκατηγορική διάταξη, καθώς χρησιμοποιούνται αποκλειστικά στη δυαδική περίπτωση. Ειδικότερα, τα false negatives δείγματα για μία δεδομένη κατηγορία i εκπροσωπούνται έμμεσα μέσω των παραμέτρων false positives FP_{ij} άλλων κατηγοριών j, όπου $j \neq i$.

3.2 Τυπικές Δυαδικές Μετρικές Αξιολόγησης

Σε αυτήν την ενότητα, παρουσιάζονται οι τυπικές μετρικές αξιολόγησης για τη δυαδική ταξινόμηση: accuracy, precision, recall και F_1-score. Αυτές οι μετρικές είναι απαραίτητες για την αξιολόγηση της απόδοσης των δυαδικών ταξινομητών, [19, 20].

3.2.1 Accuracy

Η accuracy (ακρίβεια) είναι η πλέον κοινή μετρική για την απόδοση της ταξινόμησης. Ορίζεται ως ο λόγος των ορθώς ταξινομημένων δειγμάτων προς τον συνολικό αριθμό των δειγμάτων, και μπορεί να εκφραστεί μαθηματικά ως εξής:

$$\text{Accuracy} \stackrel{\text{def}}{=} \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{3.2}$$

όπου τα TP (True Positives), TN (True Negatives), FP (False Positives) και FN (False Negatives) είναι τα αντίστοιχα στοιχεία του πίνακα σύγχυσης.

Η accuracy, μια ευρέως αναγνωρισμένη μετρική, παρουσιάζει σημαντικούς περιορισμούς σε μη ισορροπημένα σύνολα δεδομένων. Για παράδειγμα, στην περίπτωση όπου ο αριθμός των δειγμάτων της αρνητικής κλάσης αυξηθεί κατά έναν παράγοντα α, οι τιμές TN και FP θα αυξηθούν αντίστοιχα: $\alpha \cdot \text{TN}$ και $\alpha \cdot$ FP. Τότε, η προσαρμοσμένη accuracy θα είναι:

$$(\text{Accuracy})_\alpha = \frac{\text{TP} + \alpha \cdot \text{TN}}{\text{TP} + \alpha \cdot \text{TN} + \alpha \cdot \text{FP} + \text{FN}} \neq \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} = \text{Accuracy}$$

Αυτή η προσαρμοσμένη μετρική μπορεί να διαφέρει σημαντικά από την αρχική, αναδεικνύοντας την αναποτελεσματικότητα της accuracy σε μη ισορροπημένα σύνολα δεδομένων.

Ένας άλλος περιορισμός της accuracy είναι ότι δεν διαφοροποιεί τις ορθές και εσφαλμένες ταξινομήσεις. Δύο ταξινομητές μπορεί να έχουν την ίδια accuracy, αλλά να αποδίδουν διαφορετικά σε όρους precision και recall, [21]. Συνεπώς, είναι απαραίτητες εναλλακτικές μετρικές για μια πιο ολοκληρωμένη αξιολόγηση της απόδοσης των ταξινομητών.

3.2.2 Precision

Η Positive Prediction Value [PPV], ή precision, αντιπροσωπεύει την αναλογία των θετικών δειγμάτων που ταξινομήθηκαν ορθώς, ως προς το συνολικό αριθμό των υπό πρόβλεψη θετικών δειγμάτων. Υπολογίζεται ως εξής:

$$\text{Precision} = \text{PPV} \stackrel{\text{def}}{=} \frac{\text{TP}}{\text{TP} + \text{FP}} \tag{3.3}$$

Η precision αποδίδει αξιόπιστα σε προβλήματα μη ισορροπημένων κατηγοριών, καθώς υποδεικνύει την ακρίβεια του μοντέλου στην αναγνώριση της κατηγορίας-στόχου (target class). Είναι ιδιαίτερα χρήσιμη όταν το κόστος των false positives είναι υψηλό, καθώς τονίζει τη σημασία της ορθής αναγνώρισης του στόχου, ακόμα και αν αυτό συνεπάγεται την εσφαλμένη ταξινόμηση ορισμένων οριακών περιπτώσεων. Ωστόσο, η μετρική αυτή δεν λαμβάνει υπόψη τα false negatives, δηλαδή δεν εξετάζει τις περιπτώσεις όπου η ζητούμενη κατηγορία δεν αναγνωρίζεται.

3.2.3 Recall

Οι όροι sensitivity, ή True Positive Rate [TPR], ή hit rate, ή recall ενός ταξινομητή, αντιπροσωπεύουν την αναλογία των ορθώς ταξινομημένων θετικών δειγμάτων προς το συνολικό αριθμό των πραγματικά θετικών δειγμάτων. Υπολογίζεται ως εξής:

$$\text{Recall} = \text{TPR} \stackrel{\text{def}}{=} \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{3.4}$$

Γενικά, η recall μπορεί να θεωρηθεί ως μια παραλλαγή της accuracy για τα θετικά δείγματα και εξαρτάται από τα [TP] και [FN], τα οποία βρίσκονται στην ίδια στήλη του πίνακα σύγχυσης. Ακόμη, η sensitivity είναι ιδιαίτερα χρήσιμη για την αξιολόγηση της απόδοσης ταξινόμησης σε μη ισορροπημένα δεδομένα, διότι επικεντρώνεται στην ικανότητα του ταξινομητή να αναγνωρίζει θετικά δείγματα, [21]:

> "Η precision και η recall είναι δύο θεμελιώδεις μετρικές που χρησιμοποιούνται για την αξιολόγηση της απόδοσης μοντέλων ταξινόμησης, ιδιαίτερα σε περιπτώσεις μη ισορροπημένων δεδομένων. Ενώ η precision δίνει έμφαση στην ικανότητα του μοντέλου να αναγνωρίζει ορθώς τα θετικά δείγματα, η recall εστιάζει στην ικανότητά του να αναγνωρίζει όλες τις θετικές περιπτώσεις. Αυτή η δυαδικότητα συχνά παρουσιάζει ένα trade-off: η αύξηση της precision συνήθως μειώνει την recall, και το αντίστροφο. Η επίτευξη υψηλής precision συχνά περιλαμβάνει τον καθορισμό ενός υψηλότερου κατωφλίου για την ταξινόμηση, μειώνοντας τον αριθμό των false positives αλλά πιθανώς αυξάνοντας τα false negatives. Αντιθέτως, η έμφαση στη recall μπορεί να μειώσει το κατώφλι, αναγνωρίζοντας περισσότερες θετικές περιπτώσεις αλλά πιθανώς αυξάνοντας τα false positives. Η εξισορρόπηση της precision και της recall είναι κρίσιμη και συχνά εξαρτάται από τις συγκεκριμένες απαιτήσεις του προβλήματος. Για αυτούς τους λόγους, εισάγεται το F_1-score, *μια μετρική σχεδιασμένη να ισορροπεί αυτόν τον συμβιβασμό, ως αρμονικός μέσος.*"

3.2.4 F_1-score

Το F_1-score, επίσης γνωστό και ως F-measure, αντιπροσωπεύει τον αρμονικό μέσο όρο της precision [PPV] και της recall [TPR]. Κυμαίνεται από το μηδέν έως το ένα, με τις υψηλότερες τιμές να υποδεικνύουν καλύτερη απόδοση ταξινόμησης. Το F_1-score υπολογίζεται ως εξής:

$$F\text{-measure} = F_1\text{-score} \stackrel{\text{def}}{=} 2 \cdot \frac{\text{PPV} \times \text{TPR}}{\text{PPV} + \text{TPR}} = \cdots = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}} \tag{3.5}$$

Αυτή η μετρική εκφράζεται και μέσω μίας άλλης παραλλαγής που ονομάζεται F_β-measure, η οποία αντιπροσωπεύει τον **weighted** (σταθμισμένο) αρμονικό μέσο μεταξύ της precision και της sensitivity:

$$F_\beta\text{-measure} \stackrel{\text{def}}{=} \left(1 + \beta^2\right) \cdot \frac{\text{PPV} \times \text{TPR}}{\beta^2 \cdot \text{PPV} + \text{TPR}} = \cdots = \frac{(1 + \beta^2) \cdot \text{TP}}{(1 + \beta^2) \cdot \text{TP} + \beta^2 \cdot \text{FP} + \text{FN}}$$

Γενικότερα, αυτή η μετρική είναι ευαίσθητη σε αλλαγές στις κατανομές των δεδομένων. Για παράδειγμα, αν ο αριθμός των δειγμάτων της αρνητικής κλάσης αυξηθεί κατά έναν παράγοντα a, το F-measure υπολογίζεται ως εξής:

$$\left(F\text{-measure}\right)_\alpha \stackrel{\text{def}}{=} \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \alpha \cdot \text{FP} + \alpha \cdot \text{FN}}$$

Παρά την ευαισθησία αυτή, το F_1-score παραμένει μια αξιόπιστη μετρική για μη ισορροπημένα σύνολα δεδομένων, επειδή δεν λαμβάνει υπόψη τα true negatives, τα οποία συχνά αφθονίζουν σε τέτοια σύνολα. Με αυτόν τον τρόπο, το F_1-score επηρεάζεται λιγότερο από την πλειοψηφούσα κλάση (majority class) και επικεντρώνεται περισσότερο στην απόδοση της μειοψηφούσας κλάσης (minority class).

3.2.5 Άλλες Μετρικές

Εκτός από τις τυπικές μετρικές αξιολόγησης, υπάρχουν αρκετές άλλες μετρικές που μπορούν να χρησιμοποιηθούν για τη δυαδική ταξινόμηση:

> *Specificity ή True negative rate (TNR) ή Inverse Recall, False Positive Rate (FPR) γνωστό επίσης ως False Alarm Rate (FAR) ή Fallout, False Negative Rate (FNR) ή Miss Rate, Negative Predictive Value (NPV) ή Inverse Precision ή True Negative Accuracy (TNA), False Discovery Rate (FDR) και False Omission Rate (FOR), Positive likelihood (LR_+) και Negative likelihood (LR_-), Diagnostic Odds Ratio (DOR), Youden's Index (YI) ή Bookmaker Informedness (BM), Matthews Correlation Coefficient (MCC), Discriminant Power (DP), Adjusted F-measure (AGF), Markedness (MK), Balanced Classification Rate ή Balanced Accuracy (BCR), Geometric Mean (GM) και Adjusted Geometric Mean (AGM), Optimization Precision (OP) και τέλος η Jaccard's Metric.*

Οι παραπάνω μετρικές αναφέρονται για εγκυκλοπαιδικούς σκοπούς. Τέλος, άλλες κοινές μεθοδολογίες για την αξιολόγηση της απόδοσης είναι οι **Receiver Operating Characteristics (ROC) και Area Under the Curve (AUC)**, οι οποίες δεν θα παρουσιαστούν λεπτομερώς εδώ, ούτε θα χρησιμοποιηθούν αργότερα.

3.3 Τυπικές Πολυκατηγορικές Μετρικές Αξιολόγησης

Σε αυτήν την ενότητα, επεκτείνεται η μελέτη των μετρικών αξιολόγησης σε διεργασίες πολυκατηγορικής ταξινόμησης Εισάγονται έννοιες όπως micro-averaging, macro-averaging και weighted averaging, μαζί με άλλες καθιερωμένες μετρικές, ειδικά προσαρμοσμένες για πολυκατηγορικά προβλήματα.

3.3.1 Accuracy

Όπως και στη δυαδική ταξινόμηση, η accuracy θα προσμετρά την αναλογία των ορθώς ταξινομημένων δειγμάτων σε σχέση με τον συνολικό αριθμό των δειγμάτων. Ωστόσο, σε αυτό το πλαίσιο, λαμβάνονται υπόψιν όλες οι κατηγορίες και υπολογίζεται σύμφωνα με τον παρακάτω τύπο, [22]:

$$\text{Accuracy} \stackrel{\text{def}}{=} \frac{1}{N} \cdot \sum_{i=1}^{n} \text{TP}_i \tag{3.6}$$

όπου οι παράμετροι TP_i αντιπροσωπεύουν τον αριθμό των true positives για την κατηγορία i, n είναι το συνολικό πλήθος των κατηγοριών και N είναι το άθροισμα όλων των στοιχείων του πίνακα σύγχυσης.

Η accuracy προσφέρει μια γενική εικόνα της απόδοσης του ταξινομητή. Ωστόσο, ενδέχεται να μην καταγράφει επαρκώς ανισομερείς κατανομές στόχου. Σε αυτές τις περιπτώσεις μπορεί να είναι απατηλή, καθώς η υψηλή ακρίβεια μπορεί να προκύψει κυρίως από την ορθή ταξινόμηση

των δειγμάτων της πλειοψηφούσας κατηγορίας, παραβλέποντας την απόδοση στις μειοψηφούσες κατηγορίες.

3.3.2 Δυαδικός Μετασχηματισμός του Πολυκατηγορικού Πίνακα Σύγχυσης

Προκειμένου να οριστούν οι μετρικές $(\text{Precision})_i$, $(\text{Recall})_i$ και $(F_1\text{-score})_i$ για την εκάστοτε κατηγορία i, υιοθετείται η προσέγγιση one-vs-rest, κατά την οποία κάθε κατηγορία i θεωρείται η θετική κατηγορία (ή κατηγορία 0 ή πλειοψηφούσα κατηγορία) και όλες οι υπολοιπες νοούνται ως αρνητική κατηγορία (μειοψηφούσα ή κατηγορία 1). Αυτός ο δυαδικός μετασχηματισμός επεκτείνει τις μετρικές δυαδικής ταξινόμησης σε κάθε κατηγορία. Παρατίθενται οι ορισμοί:

$$\left[(\text{Precision})_i \stackrel{\text{def}}{=} \frac{\text{TP}_{ii}}{\text{TP}_{ii} + \sum_{j=1}^{n} [\text{FP}_{ij}]_{j\neq i}}\right] \wedge \left[(\text{Recall})_i \stackrel{\text{def}}{=} \frac{\text{TP}_{ii}}{\text{TP}_{ii} + \sum_{j=1}^{n} [\text{FN}_{ij}]_{j\neq i}}\right] \tag{3.7}$$

και:

$$(F_1\text{-score})_i \stackrel{\text{def}}{=} 2 \cdot \frac{(\text{Precision})_i \cdot (\text{Recall})_i}{(\text{Precision})_i + (\text{Recall})_i} \tag{3.8}$$

Σε αυτό το πλαίσιο:

- TP_{ii} είναι ο αριθμός των true positives για την κατηγορία i.
- $[\text{FP}_{ij}]_{j\neq i}$ είναι ο αριθμός των false positives για την κλάση j, όταν η πραγματική κατηγορία είναι η i.
- $[\text{FN}_{ij}]_{j\neq i}$ είναι ο αριθμός των false negatives για την κλάση j, όταν η πραγματική κατηγορία είναι η i.

3.3.3 Precision, Recall και F_1-score

Κατά την πολυκατηγορική ταξινόμηση, οι μετρικές Precision, Recall και F_1-score μπορούν να υπολογιστούν χρησιμοποιώντας μεθόδους micro (μ), macro (M), ή weighted averaging. Προς διευκρίνηση, ορίζονται ρητά αυτοί οι όροι και παρατίθονται οι σχετικοί τύποι, προσαρμοσμένοι από τις πηγές [23, 24, 25].

3.3.3.1 Micro-Averaging

Η τεχνική micro-averaging συγκεντρώνει τις συνεισφορές όλων των κατηγοριών για τον υπολογισμό μιας ενιαίας μετρικής, δίνοντας ίσο βάρος σε κάθε δείγμα και παρέχοντας μια συνολική εικόνα. Μπορεί να εκφραστεί ως εξής:

1. **Micro-Precision** (μ-Precision): Συγκεντρώνει τις συνεισφορές όλων των κατηγοριών για τον υπολογισμό μιας ενιαίας μετρικής, δίνοντας ίσο βάρος σε κάθε δείγμα. Υπολογίζεται ως:

$$\mu\text{-Precision} \stackrel{\text{def}}{=} \frac{\sum_{i=1}^{n} \text{TP}_{ii}}{\sum_{i=1}^{n} \left(\text{TP}_{ii} + \sum_{j=1}^{n} \text{FP}_{ij}\right)} = \frac{\sum_{i=1}^{n} \text{TP}_{ii}}{N}$$

Εδώ, ο αριθμητής είναι ο συνολικός αριθμός των true positives σε όλες τις κατηγορίες, και ο παρανομαστής είναι ο συνολικός αριθμός των υπό πρόβλεψη δειγμάτων σε όλες τις κατηγορίες (N).

2. **Micro-Recall** (μ-Recall): Παρομοίως συγκεντρώνει τις συνεισφορές όλων των κατηγοριών. Υπολογίζεται ως:

$$\text{μ-Recall} \stackrel{\text{def}}{=} \frac{\sum_{i=1}^{n} \text{TP}_{ii}}{\sum_{i=1}^{n} \left(\text{TP}_{ii} + \sum_{j=1}^{n} \text{FN}_{ij} \right)} = \text{μ-Precision}$$

Προφανώς, η micro-precision και η micro-recall είναι ταυτόσημες, δεδομένου ότι δεν εμφανίζονται τα false negatives κατά τον ορισμό. Συγκεντρώνουν τις συνεισφορές όλων των κατηγοριών για τον υπολογισμό της μέσης τιμής, κάτι που είναι ιδιαίτερα χρήσιμο για μη ισορροπημένα σύνολα δεδομένων.

3. **Micro-F_1-score** (μ-F_1-score): Συνδυάζει τη μ-Precision και τη μ-Recall ως τον αρμονικό μέσο τους:

$$\text{μ-F1-score} \stackrel{\text{def}}{=} 2 \cdot \frac{\text{μ-Precision} \cdot \text{μ-Recall}}{\text{μ-Precision} + \text{μ-Recall}} \stackrel{\text{μ-Recall=μ-Precision}}{\Longrightarrow} \text{μ-F1-score} = \text{μ-Recall} = \text{μ-Precision}$$

3.3.3.2 Macro-Averaging

Η τεχνική macro-averaging εξάγει τη μετρική ανεξάρτητα για κάθε κατηγορία και στη συνέχεια υπολογίζει τον αντίστοιχο μέσο όρο, δίνοντας ίσο βάρος σε κάθε κατηγορία. Υπολογίζεται ως εξής:

1. **Macro-Precision** (M-Precision): Υπολογίζει τον μέσο όρο όλων των κατηγοριών για να προκύψει μια ενιαία μετρική, δίνοντας ίσο βάρος σε κάθε κατηγορία. Υπολογίζεται ως:

$$\text{M-Precision} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} (\text{Precision})_i$$

2. **Macro-Recall** (M-Recall): Παρομοίως, υπολογίζει τον μέσο όρο όλων των κατηγοριών. Υπολογίζεται ως:

$$\text{M-Recall} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} (\text{Recall})_i$$

3. **Macro-F_1-score** (M-F_1-score): Εξάγει το μέσο όρο των F_1-scores για όλες τις κλάσεις:

$$\text{M-F1-score} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} (F_1\text{-score})_i$$

3.3.3.3 Weighted-Averaging

Η τεχνική weighted-averaging υπολογίζει τη μετρική για κάθε κατηγορία, σταθμισμένη ανάλογα με το πλήθος των αντίστοιχων πραγματικών περιπτώσεων σε κάθε κατηγορία, και στη συνέχεια εξάγει τον μέσο όρο. Λαμβάνοντας υπόψη τον αριθμό των περιπτώσεων ανά κατηγορία,

παρέχει μια ισορροπημένη εικόνα που αντανακλά τη συμβολή κάθε κατηγορίας ανάλογα με το μέγεθός της. Τα βάρη (weights) της κάθε κατηγορίας υπολογίζονται ως εξής:

$$w_i \stackrel{\text{def}}{=} \frac{1}{N} \sum_{j=1}^{n} C_{ij} \tag{3.9}$$

όπου C_{ij} είναι το στοιχείο του πολυκατηγορικού πίνακα σύγχυσης στη θέση (i, j) και N είναι ο συνολικός αριθμός περιπτώσεων σε όλες τις κατηγορίες. Εναλλακτικά:

$$w_i \stackrel{\text{def}}{=} \frac{N_i}{N} \tag{3.10}$$

όπου N_i είναι ο αριθμός των πραγματικών περιπτώσεων στην κατηγορία i.

Παρατίθενται οι ορισμοί των μετρικών:

1. **Weighted-Precision**: Υπολογίζει τη μετρική για κάθε κατηγορία, σταθμισμένη ανάλογα με τον αριθμό των πραγματικών περιπτώσεων σε κάθε κατηγορία, και στη συνέχεια εξάγει τον μέσο όρο:

$$\text{Weighted-Precision} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} w_i \cdot \text{Precision}_i \tag{3.11}$$

2. **Weighted-Recall**: Παρομοίως υπολογίζει τις σταθμισμένες συνεισφορές όλων των κατηγοριών:

$$\text{Weighted-Recall} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} w_i \cdot \text{Recall}_i \tag{3.12}$$

3. **Weighted-F_1-score**: Υπολογίζει τον σταθμισμένο μέσο όρο του F_1-score όλων των κατηγοριών:

$$\text{Weighted-F1-score} \stackrel{\text{def}}{=} \frac{1}{n} \sum_{i=1}^{n} w_i \cdot F_1\text{-score}_i \tag{3.13}$$

Καθώς η απόδοση ενός μοντέλου ταξινόμησης μπορεί να υπολογιστεί μέσω πολλών πρακτικών, με την τεχνική micro-averaging να δίνει έμφαση στην συνολική απόδοση, την πρακτική macro-averaging να παρέχει πληροφορίες για μια ισορροπημένη εικόνα της απόδοσης και την μέθοδο weighted averaging να λαμβάνει υπόψη την ανισορροπία των κατηγοριών, οι ποικίλες εκφράσεις αυτών των μετρικών είναι αξιοσημείωτες σε εφαρμογές όπου κάποιες κλάσεις μπορεί να είναι πιο σημαντικές από άλλες ή η ανισομέρεια της κατανομής των κλάσεων αναδεικνύεται ως καίριο ζήτημα.

3.3.4 Άλλες Μετρικές

Εκτός από τις τυπικές μετρικές αξιολόγησης όπως οι πίνακες σύγχυσης, accuracy, class-wise precision, recall, και F_1-score, κατά την πολυκατηγορική ταξινόμηση χρησιμοποιούνται επίσης και άλλες μετρικές προς ανάλυση της συμπεριφοράς του μοντέλου, όπως:

Balanced Accuracy & **Weighted Balanced Accuracy**: Αυτές οι μετρικές προσφέρουν μια ισορροπημένη εικόνα της απόδοσης του ταξινομητή, λαμβάνοντας υπόψη οποιαδήποτε ανισορροπία κατηγοριών του στόχου.

Cohen's Kappa & **Mathew's Correlation Coefficient** (πολυκατηγορικές παραλλαγές): Αυτές οι μετρικές αξιολογούν τη συμφωνία μεταξύ των προβλεπόμενων και των πραγματικών κατηγοριών, λαμβάνοντας υπόψη την πιθανότητα τυχαίας συμφωνίας.

Αν και δεν συζητούνται εκτενώς εδώ, άλλες μεθοδολογίες όπως η **Receiver Operating Characteristics (ROC) Curve** και η **Area Under the Curve (AUC)** παραμένουν εφαρμόσιμες για την αξιολόγηση πολυκατηγορικών ταξινομητών, προσφέροντας πολύτιμες πληροφορίες.

3.4 Προηγμένες Μετρικές Αξιολόγησης για Πολυκατηγορική Ανίχνευση Εισβολών

Αυτή η ενότητα εξετάζει προηγμένες μετρικές αξιολόγησης, ειδικά προσαρμοσμένες για την ανίχνευση εισβολών στο πλαίσιο ενός IDS.

3.4.1 Error Rate per Class

Μαθηματική Περιγραφή: Το Error Rate per Class (ποσοστό σφάλματος ανά κατηγορία) υπολογίζει το ποσοστό των false positives δειγμάτων αθροίζοντας τις εσφαλμένες ταξινομήσεις για την εν λόγω κλάση, και διαιρώντας με τον συνολικό αριθμό των δειγμάτων που ανήκουν στην κλάση. Μαθηματικά, εκφράζεται ως:

$$\text{Error Rate per Class} = \frac{\text{Total False Positives for the Class}}{\text{Total True Samples for the Class}} \tag{3.14}$$

Στατιστική Ερμηνεία Το Error Rate per Class παρέχει πληροφορίες σχετικά με την ακρίβεια των προβλέψεων για μεμονωμένες κατηγορίες. Ποσοτικοποιεί το ποσοστό των δειγμάτων από κάθε κατηγορία που ταξινομήθηκαν εσφαλμένα από το μοντέλο, υποδεικνύοντας την απόδοση του ταξινομητή σε κάθε κλάση.

Εξαγωγή από τον Πίνακα Σύγχυσης Το Error Rate per Class προκύπτει από τον πίνακα σύγχυσης προσθέτοντας τις false positives περιπτώσεις για κάθε κατηγορία. Διαιρώντας αυτό το άθροισμα με τον συνολικό αριθμό των πραγματικών δειγμάτων για την κλάση, προκύπτει το ποσοστό σφάλματος ανά κατηγορία. Με άλλα λόγια, υπολογίζεται ως το άθροισμα της γραμμής της κατηγορίας (που αντιπροσωπεύει όλες τις προβλέψεις για αυτήν την κατηγορία) του πίνακα σύγχυσης, μείον το διαγώνιο στοιχείο (true positives), διαιρεμένο με τα συνολικά πραγματικά δείγματα ανά κατηγορία.

Για τον υποθετικό πίνακα σύγχυσης παραπάνω, το Error Rate per Class της κλάσης i θα γραφτεί ως $(\text{ERpC})_i$:

$$(3.14) \implies \text{ERpC}_i = \frac{\sum_{j=1}^{n} [\text{FP}_{ij}]_{j \neq i}}{\text{TP}_{ii} + \sum_{j=1}^{n} [\text{FP}_{ij}]_{j \neq i}}$$

όπου οι παράμετροι FP_{ij} καταδεικνύουν τα false positives της κλάσης i όταν προβλέφθηκαν ως κατηγορία j, και οι παράμετροι TP_{ii} αναπαριστούν τα true positives της κατηγορίας i.

3.4.2 Class-Wise Misclassification Rate

Μαθηματική Περιγραφή Το Class-Wise Misclassification Rate (ποσοστό λανθασμένης ταξινόμησης ανά κατηγορία) μετρά το συνολικό ποσοστό εσφαλμένων ταξινομήσεων για κάθε μεμονωμένη κατηγορία. Μαθηματικά, εκφράζεται ως εξής:

$$\text{Class-Wise Misclassification Rate} = \frac{\text{Total Misclassifications for the Predicted Class}}{\text{Total True Samples that are Not the Class}}$$

$$= \frac{\text{Total "False Negatives" the Predicted Class}}{\text{Total "Negatives" for the Class}} \tag{3.15}$$

Προσοχή: *Όπως επισημάνθηκε προηγουμένως, ο όρος true ή false "negative" δεν μπορεί να οριστεί για μια πολυκατηγορική διάταξη. Επομένως, υπονοείται ένας δυαδικός μετασχηματισμός του προβλήματος, με την υπό εξέταση κατηγορία ως την πλειοψηφούσα και την συνολοθεωρητική ένωση των υπόλοιπων κατηγοριών ως τη μειοψηφούσα.*

Στατιστική Ερμηνεία Το Class-Wise Misclassification Rate παρέχει πληροφορίες σχετικά με την ακρίβεια των προβλέψεων για κάθε μεμονωμένη κατηγορία. Ποσοτικοποιεί το ποσοστό των δειγμάτων μιας υπό πρόβλεψη κατηγορίας, τα οποία ταξινομήθηκαν εσφαλμένα από το μοντέλο. Αυτή η μετρική είναι πολύτιμη για τον εντοπισμό των κατηγοριών που είναι πλέον επιρρεπείς σε εσφαλμένες ταξινομήσεις.

Εξαγωγή από τον Πίνακα Σύγχυσης Το Class-Wise Misclassification Rate προκύπτει από τον πίνακα σύγχυσης λαμβάνοντας υπόψη τον συνολικό αριθμό εσφαλμένων ταξινομήσεων για κάθε υπό πρόβλεψη κατηγορία. Αθροίζοντας τις ψευδείς προβλέψεις για μια συγκεκριμένη κατηγορία, προκύπτει το συνολικό ποσοστό λανθασμένων ταξινομήσεων. Διαιρώντας αυτήν την τιμή με τον συνολικό αριθμό των πραγματικών δειγμάτων που δεν ανήκουν στην κατηγορία, προκύπτει το ποσοστό λανθασμένης ταξινόμησης ανά κατηγορία. Με άλλα λόγια, υπολογίζεται ως το άθροισμα της στήλης της προβλεπόμενης κατηγορίας (που αντιπροσωπεύει όλες τις προβλέψεις για αυτήν την κατηγορία) μείον το διαγώνιο στοιχείο (true positives), διαιρούμενο με το άθροισμα κάθε γραμμής που δεν ανήκει στην πραγματική κατηγορία.

Για τον υποθετικό πίνακα σύγχυσης παραπάνω, το Class-Wise Misclassification Rate για την υπό πρόβλεψη κατηγορία j_0 θα γραφτεί ως CWMR_{j_0}:

$$(3.15) \Longrightarrow \text{CWMR}_{j_0} = \frac{\sum_{i=1}^{n} [\text{FP}_{ij}]_{i \neq j_0}}{\sum_{i=1}^{n} \left[\text{TP}_{ii} + \sum_{j=1}^{n} [\text{FP}_{ij}]_{j \neq i} \right]_{i \neq j_0}}$$

όπου FP_{ij_0} υποδεικνύει τα false positives δείγματα της κλάσης i όταν προβλέφθηκαν ως κλάση j_0, και η παράμετρος TP_{ii} αντιπροσωπεύει τα true positives της κλάσης i. Σε δυαδική μορφή, όπου η κατηγορία j_0 μετατρέπεται σε κατηγορία 0 και η ένωση $\cup_{i \neq j_0} [\text{class}]_i$ μετατρέπεται στην κατηγορία 1:

$$\left[\text{FN} = \sum_{j=1}^{n} [\text{FP}_{ij}]_{j \neq j_0} \right] \wedge \left[\text{FP} = \sum_{i=1}^{n} [\text{FP}_{ij_0}]_{i \neq j_0} \right] \wedge [\text{TP}_{j_0 j_0} = \text{TP}] \wedge \left[\text{TN} = \left(\sum_{i=1}^{n} \sum_{j=1}^{n} [\text{FP}_{ij}] \right) - \text{TP} \right]$$

Τώρα, ο υπολογισμός του CWMR_{j_0} της κλάσης j_0, γράφεται:

$$\text{CWMR}_{j_0} = \frac{FP}{FP + TN}$$

3.4.3 Ελαχιστοποίηση των Μετρικών

Ο πειραματισμός με χρήση παραδειγμάτων, καταδεικνύει πώς οι προτεινόμενες μετρικές υπολογίζονται από τον πίνακα σύγχυσης. Μικρότερες τιμές τόσο για το Error Rate per Class όσο και για το Class-Wise Misclassification Rate υποδεικνύουν καλύτερη απόδοση ταξινόμησης. Ωστόσο, η επίτευξη τέλειας απόδοσης (μηδενικό σφάλμα) μπορεί να μην είναι πάντα εφικτή σε πραγματικά σενάρια λόγω διαφόρων παραγόντων:

Data Complexity: Τα δεδομένα του πραγματικού κόσμου παρουσιάζουν εγγενή πολυπλοκότητα με θόρυβο ή επικαλυπτόμενα χαρακτηριστικά μεταξύ κατηγοριών. Συνεπώς, η επίτευξη μηδενικού σφάλματος μπορεί να μην είναι ρεαλιστική σε τέτοιες καταστάσεις.

Class Imbalance: Τα μη ισορροπημένα σύνολα δεδομένων, όπου ορισμένες κατηγορίες έχουν σημαντικά λιγότερα δείγματα από άλλες, μπορούν να επηρεάσουν τις μετρικές προς την πλειοψηφούσα κατηγορία. Οι μειοψηφούσες κατηγορίες μπορεί να παρουσιάζουν υψηλότερα ποσοστά σφάλματος απλά λόγω της ασύμμετρης κατανομής των δειγμάτων. Σε τέτοιες περιπτώσεις, είναι σημαντικό να λαμβάνονται υπόψη πρόσθετες μετρικές όπως το F_1-score, το οποίο παρέχει μια ισορροπημένη στάθμιση της precision και της recall για κάθε κατηγορία.

Στην πράξη, οι τιμές κατωφλίου (threshold values), που εξάγονται από p-values, χρησιμοποιούνται για να καθορίσουν την αποδοχή ή την απόρριψη του ταξινομητή. Αυτές οι τιμές κατωφλίου εξαρτώνται από τις συγκεκριμένες απαιτήσεις της εργασίας ταξινόμησης και μπορεί να διαφέρουν με βάση παράγοντες όπως η σχετική σημασία των false positives σε σχέση με τα false negatives.

Επακολούθως, το κόστος της εσφαλμένης ταξινόμησης διαφέρει ανάμεσα σε διάφορες κατηγορίες. Για παράδειγμα, σε ένα σύστημα ιατρικής διάγνωσης, ένα false positive για μια κρίσιμη ασθένεια μπορεί να έχει σοβαρότερες συνέπειες από ένα false negative για μια ήπια κατάσταση. Επομένως, είναι κρίσιμο να ενσωματωθούν οι γνώσεις του υπό μελέτη τομέα εφαρμογής και να ληφθούν υπόψη οι συνέπειες της λανθασμένης ταξινόμησης κατά τον καθορισμό των τιμών κατωφλίου.

Εναλλακτικά, οι p-values μπορούν να προκύψουν χρησιμοποιώντας παραδοσιακές στατιστικές μεθοδολογίες αριθμητικών πειραμάτων.

3.4.4 Εξειδίκευση: Σύστημα Ανίχνευσης Εισβολών

> Σε αυτό το τελευταίο υποκεφάλαιο, οι μετρικές που εξηγήθηκαν παραπάνω θα εξειδικευθούν στην περίπτωση ενός Συστήματος Ανίχνευσης Εισβολών.

Υποτεθίστω ότι το IDS υλοποιείται μέσω ενός μοντέλου Μηχανικής Μάθησης ή Βαθιάς Μάθησης, δηλαδή ενός ταξινομητή. Τότε, η αξιολόγηση του IDS βασίζεται στην αξιολόγηση αυτού του ταξινομητή. Τα εν λόγω μοντέλα συνήθως εκπαιδεύονται σε μη ισορροπημένα σύνολα δεδομένων (για παράδειγμα: το σύνολο δεδομένων CIC-IDS-2017), με την πλειοψηφούσα κατηγορία να είναι η κατηγορία της "καλοήθους" δραστηριότητας και τη μειοψηφούσα κατηγορία να είναι η "κακόβουλη" ή οι κακόβουλες κατηγορίες.

> *Εδώ, η έννοια "False Positive" στο πλαίσιο του IDS αναφέρεται στην καλοήθη κίνηση που εσφαλμένα ταξινομήθηκε ως κακόβουλη, και "False Negative" αναφέρεται στις πραγματικές επιθέσεις που το σύστημα δεν εντόπισε (για ένα IDS ανεπτυγμένο υπό δυαδική διάταξη).*

Δεδομένων των χαρακτηριστικών και των απαιτήσεων ενός IDS, πώς προσαρμόζουμε τις μετρικές που συζητήθηκαν προηγουμένως για να αξιολογήσουμε την απόδοσή του·

Προτείνονται οι ακόλουθες μετρικές, [26]:

1. **False Alarm Ratio (FAR)**: Ο λόγος ψευδών συναγερμών ή False alarm ratio, επίσης γνωστός και ως False Positive Rate, προσμετρά το ποσοστό των καλοήθων περιπτώσεων

που ταξινομήθηκαν εσφαλμένα ως επιθέσεις. Υπολογίζεται ως ο αριθμός των false positives διαιρούμενος με τον συνολικό αριθμό των καλοήθων περιπτώσεων. Ένας χαμηλότερος FAR υποδεικνύει λιγότερους ψευδείς συναγερμούς, συμβάλλοντας στη διατήρηση της αξιοπιστίας του συστήματος, αφού μειώνονται οι περιττές ειδοποιήσεις.

Η μετρική FAR αποτελεί μια ειδική περίπτωση της μετρικής Error Rate per Class της προηγούμενης υποενότητας και υπολογίζεται μόνο για την καλοήθη κατηγορία. Μαθηματικά, μπορεί να εκφραστεί ως:

$$(3.14) \implies \text{FAR} = \frac{\text{False Positives for the Benign Class}}{\text{Number of Total Benign Class Instances}} \tag{3.16}$$

2. **Attack Miss Ratio (AMR)**: Ο λόγος αποτυχίας εντοπισμού επίθεσης ή Attack Miss Ratio, προσμετρά το ποσοστό των επιθέσεων που δεν εντοπίστηκαν σωστά από το σύστημα ανίχνευσης εισβολών. Υπολογίζεται ως ο αριθμός των false negatives διαιρούμενος με τον συνολικό αριθμό των επιθέσεων. Ένας χαμηλότερος AMR υποδεικνύει καλύτερη απόδοση όσον αφορά την ανίχνευση επιθέσεων. Στην ανίχνευση εισβολών, η ελαχιστοποίηση του AMR είναι κρίσιμη για να διασφαλιστεί η δικτυακή ασφάλεια.

Το AMR αποτελεί μια ειδική περίπτωση της μετρικής Classwise Misclassification Rate της προηγούμενης υποενότητας και υπολογίζεται μόνο για την ένωση των κακόβουλων κατηγοριών, άρα επάγεται δυαδικος μετασχηματισμός, αφού όλες οι κακόβουλες κατηγορίες θα ενωθούν σε μία, τη μειοψηφούσα κατηγορία. Μαθηματικά, το AMR για την κατηγορία επίθεσης μπορεί να εκφραστεί ως:

$$(3.15) \implies \text{AMR} = \frac{\text{False Positives for the Union of Attack Classes}}{\text{Total Number of Instances for all Attack Classes}} \tag{3.17}$$

Κεφάλαιο 4

Το Σύνολο Δεδομένων CIC-IDS-2017

Σε αυτήν την ενότητα παρουσιάζεται το σύνολο δεδομένων CIC-IDS-2017, καθώς και τα χαρακτηριστικά του.

4.1 Εισαγωγή

Το CIC-IDS-2017 αποτελεί ένα ευρέως διαδεδομένο σύνολο δεδομένων στον τομέα της κυβερνοασφάλειας. Αρχικά παρουσιάστηκε από τον Sharafaldin et al. το έτος 2018 [27], και αναφέρεται σε πολλά ερευνητικά άρθρα λόγω της δημόσιας διαθεσιμότητάς του και της αντιπροσωπευτικής φύσης του, ιδιαίτερα ως προς την ανίχνευση εισβολών στα πλαίσια δικτυακής κυκλοφορίας. Το σύνολο δεδομένων περιέχει μια εκτενή συλλογή δεδομένων δικτυακής ροής, καταγεγραμμένης από πραγματικά σενάρια. Περιλαμβάνονται διάφοροι τύποι κυβερνοαπειλών και επιθέσεων. Πρόκειται για ένα σύνολο δεδομένων με ετικέτες, με δείγματα ταξινομημένα είτε ως καλοήθη (benign) είτε ως κακόβουλη (malicious) κυκλοφορία, παρέχοντας μια βάση για την εκπαίδευση και την αξιολόγηση συστημάτων ανίχνευσης εισβολών, υπό όρους εποπτευόμενης μάθησης.

Το εν λόγω σύνολο δεδομένων περιέχει χαρακτηριστικά δικτυακής κυκλοφορίας όπως: τύπους πρωτοκόλλων, μεγέθη πακέτων, διευθύνσεις IP προέλευσης και προορισμού, και άλλα. Αυτά τα χαρακτηριστικά περιγράφουν ενδελεχώς την επικείμενη δικτυακή ροή, επιτρέποντας την ανάπτυξη προηγμένων μοντέλων για την ανίχνευση ανώμαλης ή κακόβουλης συμπεριφοράς. Επιπλέον, το σύνολο δεδομένων είναι αρκετά μεγάλο, περιέχοντας εκατομμύρια δείγματα, γεγονός που εξασφαλίζει την ικανότητα γενίκευσης των μοντέλων.

4.1.1 Προέλευση - CIC

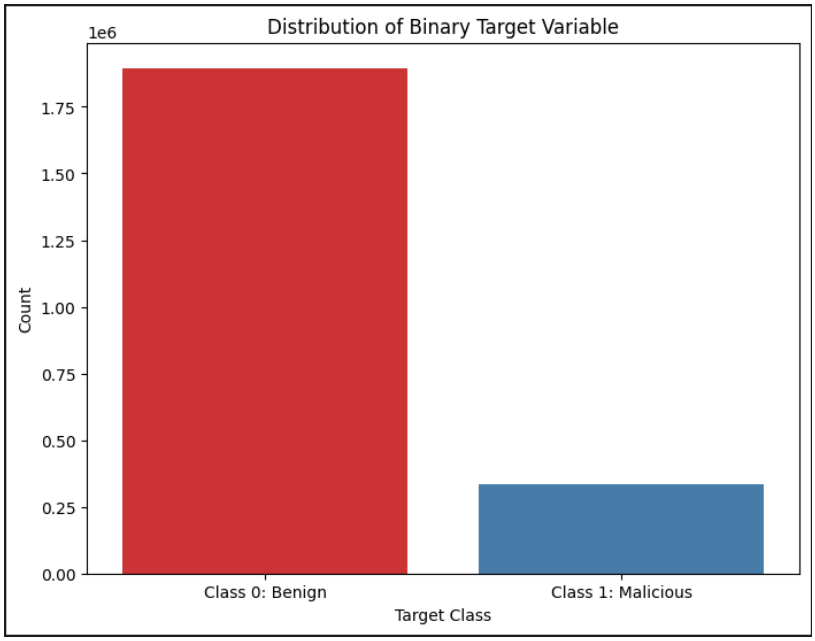
Το υπό μελέτη σύνολο δεδομένων προέρχεται από το Καναδικό Ινστιτούτο Κυβερνοασφάλειας - Canadian Institute of Cybersecurity (CIC), έναν παγκοσμίως αναγνωρισμένο οργανισμό που εστιάζει στην προαγωγή της έρευνας και της εκπαίδευσης στον τομέα της κυβερνοασφάλειας. Βρίσκεται στο Πανεπιστήμιο του New Brunswick στον Καναδά και ιδρύθηκε με στόχο να αντιμετωπίσει τις αυξανόμενες προκλήσεις στον χώρο της ασφάλειας πληροφοριών και της προστασίας των δεδομένων. Η ερευνητική δραστηριότητα του Ινστιτούτου καλύπτει ευρύ φάσμα θεμάτων, όπως η ανίχνευση εισβολών, η ανάλυση κακόβουλου λογισμικού, η ασφάλεια δικτύων και η προστασία της ιδιωτικότητας.

4.1.2 Δυαδική Ανισομέρεια

Το σύνολο δεδομένων CIC-IDS-2017 είναι heavily imbalanced, δηλαδή ανισομερές ως προς την κατανομή της μεταβλητής στόχου. Συγκεκριμένα, το πλήθος των κακόβουλων δειγμάτων

είναι πολύ μικρότερο από το μέγεθος της καλοήθους ροής. Η εν λόγω ανισότητα αποτελεί μια υπολογιστική πρόκληση[1] στα πλαίσια της Βαθιάς Μάθησης, και διαχειρίζεται με χρήση συγκεκριμένων μεθοδολογιών που θα συζητηθούν στο επόμενο κεφάλαιο.

Σχήμα 4.1: Απεικόνιση της Ανισομέρειας των δυαδικών Δεδομένων



Ελάχιστη Ακρίβεια Η κατανόηση της κατανομής του στόχου είναι απαραίτητη για την αξιολόγηση της απόδοσης των μοντέλων μηχανικής μάθησης ή βαθιάς μάθησης. Στο εν λόγω σύνολο δεδομένων:

- η πλειοψηφούσα κατηγορία (Κατηγορία 0) της benign traffic αποτελεί περίπου το 84,92% των δειγμάτων
- η μειοψηφούσα κατηγορία (Κατηγορία 1) των malicious samples συνιστά το 15,08% της κατανομής

Συνεπάγεται ότι η ελάχιστη αναμενόμενη ακρίβεια ταξινόμησης θα είναι 84,92%, αφού ένα αφελές (naive) μοντέλο που πάντα προβλέπει την πλειοψηφούσα κατηγορία θα επιτύχει μια ελάχιστη βασική ακρίβεια 84,92%. Η ίδια απαίτηση θα ισχύει και για την πολυκατηγορική διάταξη, καθώς η ένωση όλων των κατηγοριών επίθεσης μπορεί να θεωρηθεί ως η μειοψηφική κλάση.

Επομένως, οποιοδήποτε μοντέλο μάθησης αναπτυχθεί για αυτό το σύνολο δεδομένων, και στις δύο διατάξεις, θα πρέπει να υπερβεί αυτήν τη βασική (baseline) ακρίβεια για να αποδείξει ουσιαστική απόδοση πρόβλεψης.

4.2 Attack Scenarios

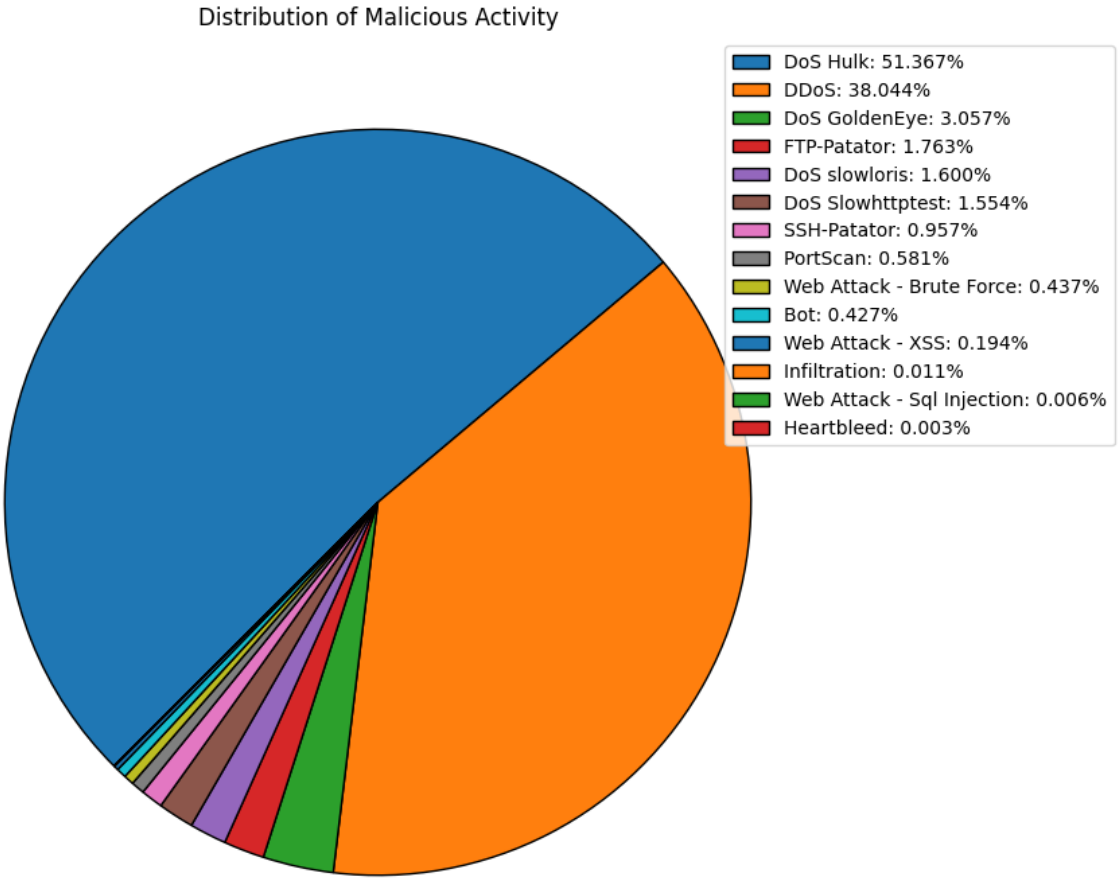
Στην παρούσα ενότητα παρουσιάζονται οι 14 επιμέρους κλάσεις της κακοήθους δικτυακής κίνησης, οι οποίες συνιστούν την πολυκατηγορική διάταξη. Οι κλάσεις αυτές ομαδοποιούνται κατάλληλα σε 7 τύπους κυβερνοαπειλών (cyberthreats): Scanning and Reconnaissance, Denial of Service, Botnet, Authentication and Access Control, Web Application Attacks, Data

[1] Όμως, στο πλαίσιο της συλλογής δεδομένων η παρατηρούμενη ανισομέρεια είναι τουλάχιστον αναμενόμενη, αφού η καλοήθης δραστηριότητα είναι εν γένει συνηθέστερη των επιθέσεων.

Exfiltration and Infiltration & Vulnerability Exploitation. Αξίζει να σημειωθεί ότι παρατηρείται ανισομέρεια και υπό την πολυκατηγορική διάταξη, δηλαδή οι επιμέρους κακόβουλες κλάσεις δεν έχουν ίση πληθικότητα. Στο διάγραμμα απεικονίζεται η παρακάτω κατανομή των δειγμάτων:

Benign: 1895314, *DoS Hulk*: 172846, *DDoS*: 128014, *DoS GoldenEye*: 10286, *FTP-Patator*: 5931, *DoS slowloris*: 5385, *DoS Slowhttptest*: 5228, *SSH-Patator*: 3219, *PortScan* 1956, *Web Attack - Brute Force*: 1470, *Bot*: 1437, *Web Attack - XSS*: 652, *Infiltration*: 36, *Web Attack - Sql Injection*: 21, *Heartbleed*: 11

Σχήμα 4.2: Απεικόνιση της Πολυκατηγορικής Ανισομέρειας



Καθώς η πλειοψηφική κλάση έχει εξαιρεθεί, όλα τα ποσοστά αντιστοιχούν στην κακόβουλη δραστηριότητα.

Ακολουθεί η λίστα των κλάσεων. Αξίζει να σημειωθεί ότι στο σύνολο δεδομένων έχει τη μορφολοφία ενός χαρακτηριστικού, ήτοι αποτελεί άλλη μία στήλη πίνακα δεδομένων υπό τον τίτλο 'Label'.

Scanning and Reconnaissance

Οι επιθέσεις σάρωσης και αναγνώρισης συγκεντρώνουν πληροφορίες για ένα δίκτυο ή σύστημα με σκοπό τον εντοπισμό ευπαθειών.

1. **PortScan**: Η τεχική port scanning χρησιμοποιείται για τον εντοπισμό προσβάσιμων θυρών και υπηρεσιών σε ένα σύστημα στόχο, για τη χαρτογράφηση δικτύων και για την εύρεση ευάλωτων σημείων, [28]. Αποτελεί ένα προκαταρκτικό βήμα σε ευρύτερες στρατηγικές επίθεσης, συγκεντρώνοντας βασικές πληροφορίες για την αρχιτεκτονική και την άμυνα ενός συστήματος

   Botnet

   Οι επιθέσεις αυτές διεξάγονται από ένα δίκτυο παραβιασμένων συσκευών (botnet), που χρησιμοποιούνται για κακόβουλες δραστηριότητες.

2. **Bot**: Μια bot attack περιλαμβάνει τη χρήση παραβιασμένων συσκευών, γνωστών ως bots, για την εκτέλεση κακόβουλων δραστηριοτήτων υπό τον έλεγχο ενός διαχειριστή, [29, 30]. Αυτά τα δίκτυα μολυσμένων συσκευών, γνωστά ως botnets, εκτελούν εργασίες όπως αποστολή ανεπιθύμητων μηνυμάτων (spam), επιθέσεις DDoS, κλοπή ευαίσθητων πληροφοριών και συμμετοχή σε απάτες click fraud, καθιστώντας τα μια σημαντική απειλή κυβερνοασφάλειας.

   Denial of Service (DoS)

   Οι επιθέσεις Denial of Service (Άρνησης Εξυπηρέτησης) αποτελούν κακόβουλες απόπειρες διακοπής της διαθεσιμότητας υπηρεσιών. Πραγματοποιούνται υπερφορτώνοντας ένα σύστημα με υπερβολική κίνηση ή αιτήσεις, καθιστώντας το μη ανταποκρινόμενο. Οι εν λόγω επιθέσεις εκμεταλλεύονται τους περιορισμούς των πόρων ενός server[2], όπως το bandwidth[3], τη μνήμη και την επεξεργαστική ισχύ (CPU), επιβραδύνοντας σημαντικά ή διακόπτοντας πλήρως την παροχή υπηρεσιών. Η ευρύτερη κατηγορία επιθέσεων DoS περιλαμβάνει πληθώρα μεθόδων, καθεμία εκ των οποίων μεταχειρίζεται μοναδικές τακτικές για την επίτευξη διακοπής εξυπηρέτησης.

3. **DoS Hulk**: Οι επιθέσεις DoS Hulk[4] στέλνουν έναν τεράστιο αριθμό αιτήσεων HTTP GET σε σύντομο χρονικό διάστημα, παρακάμπτοντας τους μηχανισμούς αποθήκευσης στην cache μνήμη, αποσκοπώντας στην κατανάλωση πόρων του εξυπηρετητή [31, 32]. Επακολούθως, ο υψηλός όγκος αιτήσεων προκαλεί την κατάρρευση (crash) ή την αδυναμία απόκρισης του server λόγω της εξάντλησης υπολογιστικών πόρων.

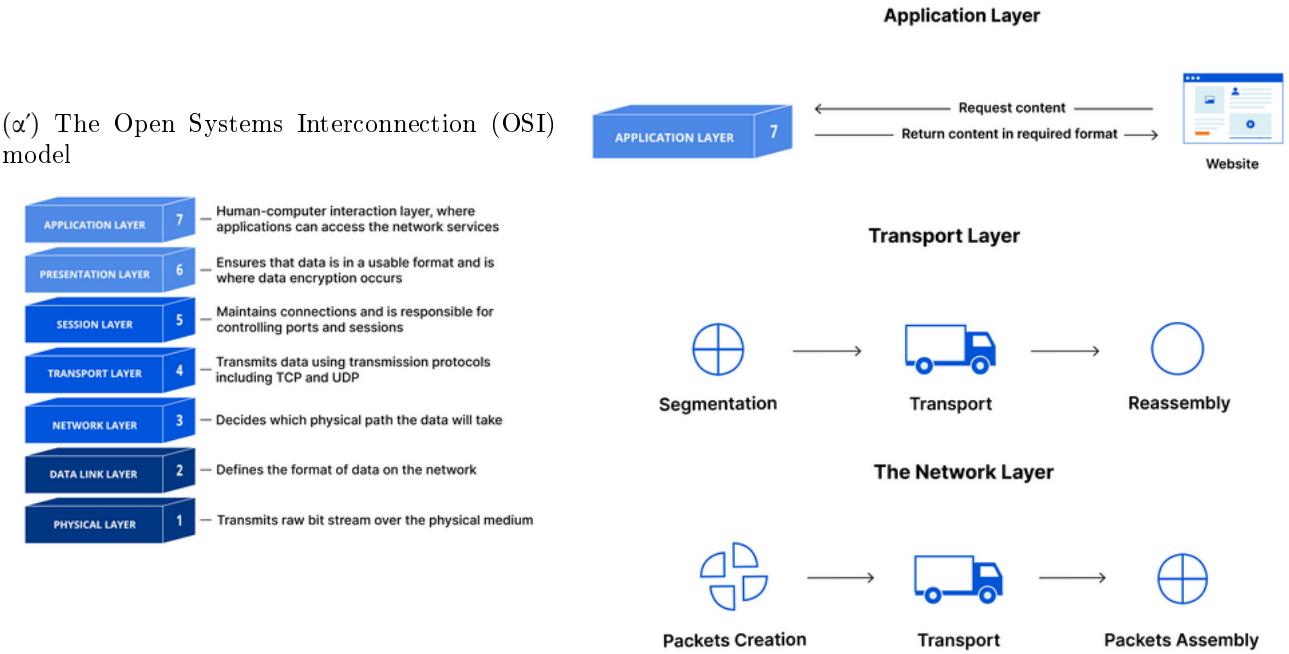
4. **DoS GoldenEye**: Οι εν λόγω επιθέσεις, παρόμοια με τις HTTP floods, υπερφορτώνουν έναν εξυπηρετητή στέλνοντας πολυάριθμες ημιτελείς αιτήσεις HTTP GET, αφήνοντας τις συνδέσεις ανοιχτές και παρεμποδίζοντας την απελευθέρωση πόρων. Αυτή η μέθοδος στοχεύει στην εξάντληση δικτυακών πόρων και διαταράσσει την ικανότητα του εξυπηρετητή να επεξεργάζεται legitimate requests, [31, 33].

5. **DDoS**: Οι επιθέσεις Distributed Denial of Service (Κατανεμημένης Άρνησης Εξυπηρέτησης) εκμεταλλεύονται πολλαπλά παραβιασμένα συστήματα, συχνά σχηματίζοντας ένα botnet, προκειμένου να κατακλύσουν έναν στόχο με κακόβουλη traffic. Η καταστολή τους θεωρείται σημαντικά πιο δύσκολη σε σύγκριση με τις παραδοσιακές επιθέσεις DoS. Ελεγχόμενα από έναν μοναδικό attacker, τα botnets αποτελούνται από πολυάριθμες μολυσμένες μηχανές, και η κατανεμημένη φύση της επίθεσης καθιστά τις απλές στρατηγικές αποκλεισμού αναποτελεσματικές. Οι επιθέσεις DDoS στοχεύουν διάφορα επίπεδα της δικτυακής αρχιτεκτονικής OSI (επίπεδα 3, 4 και 7): περιλαμβάνουν volumetric attacks

[2]server: εξυπηρετητής
[3]bandwidth: εύρος ζώνης
[4]HULK: HTTP-Unbearable-Load-King.

προς κατανάλωση εύρους ζώνης, ενώ παράλληλα εκμεταλλεύονται αδυναμίες δικτυακών πρωτοκόλλων, αλλά και επιτίθενται στο επίπεδο εφαρμογής προς εξάντληση πόρων[5], [34].

Σχήμα 4.3: DDos - Vulnerable OSI layers: Application, Transport & Network Layer



(α′) The Open Systems Interconnection (OSI) model

(α′). The open systems interconnection (OSI) model is a conceptual model created by the International Organization for Standardization which enables diverse communication systems to communicate using standard protocols, i.e. the OSI provides a standard for different computer systems to be able to communicate with each other in terms of computer networking. It is based on the concept of splitting up a communication system into seven abstract layers, each one stacked upon the last.

Application Layer (Layer 7): *The Application Layer of the OSI model is the topmost layer that provides the interface for end-user communication and data transfer, supporting application services such as email, file transfer, and web browsing. It enables applications to interact with the network by providing protocols and services like HTTP, FTP, and SMTP.*

Transport Layer (Layer 4): *The Transport Layer of the OSI model ensures reliable data transfer between devices by providing end-to-end communication control and error-checking. It manages data flow, error recovery, and retransmission through protocols like TCP and UDP.*

Network Layer (Layer 3): *The Network Layer of the OSI model is responsible for routing, forwarding, and addressing data packets across interconnected networks. It determines the best path for data transmission using protocols like IP.*

6. **DoS Slowhttptest**: Η εν λόγω επίθεση διατηρεί πολλαπλές HTTP συνδέσεις ανοιχτές και μεταδίδει δεδομένα με αργούς ρυθμούς. Αυτή η τεχνική μονοπωλεί τους πόρους του εξυπηρετητή, καθιστώντας δύσκολο τον χειρισμό των legitimate requests. Είναι ιδιαίτερα

[5] **Σημείωση**: Οι εξηγήσεις και τα σχόλια που συνοδεύουν τις εικόνες είναι φορτωμένα από τεχνική ορολογία. Παρότι η υιοθέτηση της αγγλικής γλώσσας είναι ασυνεπής, εν τέλει δεν επιλέγεται η μετάφραση στα ελληνικά, για λόγους ευκρίνειας και αναγνωσιμότητας.

επικίνδυνη λόγω των χαμηλών απαιτήσεων σε εύρος ζώνης και της ικανότητάς της να προσομειώνει legitimate traffic patterns, [35].

7. **DoS slowloris**: Η εν λόγω επίθεση, παρόμοια με την προηγούμενη, διατηρεί πολλές συνδέσεις ανοιχτές για παρατεταμένες χρονικές περιόδους, αποστέλοντας partial HTTP requests, [36]. Αυτή η επίθεση επίσης καταναλώνει τους πόρους του εξυπηρετητή χωρίς να απαιτεί υψηλό εύρος ζώνης, γεγονός που την καθιστά δύσκολα ανιχνέυσιμη.

Authentication and Access Control

Οι επιθέσεις ταυτοποίησης και ελέγχου πρόσβασης στοχεύουν στους μηχανισμούς προστασίας των συστημάτων και των δεδομένων, προσπαθώντας να παρακάμψουν ή να πραγματοποιήσουν επίθεση brute force στις δομές ταυτοποίησης, με στόχο την απόκτηση μη εξουσιοδοτημένης πρόσβασης, [37]. Αυτές οι επιθέσεις εκμεταλλεύονται αδύναμους κωδικούς πρόσβασης, ακατάλληλες ρυθμίσεις και την απουσία προστατευτικών μέτρων, π.χ. κλείδωμα λογαριασμών. Οι επιτιθέμενοι χρησιμοποιούν αυτοματοποιημένα εργαλεία για τη συστηματική δοκιμή διαφορετικών συνδυασμών ονομάτων χρήστη και κωδικού πρόσβασης, με στόχο την παραβίαση του συστήματος στόχου.

8. **FTP-Patator**: Αποτελεί εργαλείο επίθεσης brute force, σχεδιασμένο για την απόκτηση μη εξουσιοδοτημένης πρόσβασης σε FTP servers, δοκιμάζοντας συστηματικά συνδυασμούς ονόματος χρήστη και κωδικού πρόσβασης. Το εργαλείο αυτό αποπειράται μεγάλο όγκο προσπαθειών σύνδεσης, εκμεταλλευόμενο αδύναμους κωδικούς πρόσβασης ή ανεπαρκώς ρυθμισμένους servers που δεν διαθέτουν μηχανισμούς κλειδώματος λογαριασμών. Οι επιτιθέμενοι, χρησιμοποιώντας το FTP-Patator, αποκτούν πρόσβαση σε ευαίσθητα δεδομένα, και επακολούθως δύνανται να κλέψουν πληροφορίες ή να χρησιμοποιήσουν τον παραβιασμένο διακομιστή για περαιτέρω κακόβουλες δραστηριότητες.

9. **SSH-Patator**: Το SSH-Patator λειτουργεί παρομοίως, στοχεύοντας σε διακομιστές SSH για την απόκτηση μη εξουσιοδοτημένης πρόσβασης. Αναλόγως, δοκιμάζει συστηματικά συνδυασμούς ονόματος χρήστη και κωδικού πρόσβασης. Το SSH, ένα καίριο πρωτόκολλο για την ασφαλή απομακρυσμένη διαχείριση διακομιστών, καθίσταται πρωταρχικός στόχος για τους επιτιθέμενους. Το SSH-Patator αυτοματοποιεί τη διαδικασία brute force, αυξάνοντας την πιθανότητα εύρεσης έγκυρων credentials στην περίπτωση χρήσης αδύναμων κωδικών πρόσβασης. Μόλις αποκτηθεί πρόσβαση, οι επιτιθέμενοι μπορούν να εκτελέσουν διάφορες κακόβουλες δραστηριότητες, όπως κλοπή δεδομένων και εγκατάσταση κακόβουλου λογισμικού.

Vulnerability Exploitation

Οι επιθέσεις εκμετάλλευσης ευάλωτων στοιχείων στοχεύουν σε συγκεκριμένες ευπαθείς λειτουργίες λογισμικού, για την απόκτηση μη εξουσιοδοτημένης πρόσβασης, την κλοπή δεδομένων ή τη διατάραξη υπηρεσιών και την ευρύτερη παραβίαση της ασφάλειας του συστήματος.

10. **Heartbleed**: Με τον όρο Heartbleed εννοείται μία ζωτικής σημασίας ευπάθεια (vulnerability) στη βιβλιοθήκη κρυπτογράφησης OpenSSL, επιτρέπουσα την ανάγνωση μνήμης συστημάτων με αποτέλεσμα την έκθεση ευαίσθητων πληροφοριών, [38]. Η εν λόγω αδυναμία ανακαλύφθηκε στην επέκταση heartbeat του TLS/DTLS το έτος 2014 [39], και δίνει τη δυνατότητα εξαγωγής ιδιωτικών κλειδιών, κωδικών πρόσβασης, session tokens και άλλων εμπιστευτικων δεδομένων. Ο μηχανισμός Heartbleed ανάγεται σε ανεπαρκείς ελέγχους εισόδου στην επέκταση heartbeat του OpenSSL. Ειδικότερα, οι επιτιθέμενοι στέλνουν ένα αίτημα heartbeat με μήκος φορτίου μεγαλύτερο από το πραγματικό φορτίο,

προκαλώντας τον εξυπηρετητή να απαντήσει με πρόσθετα δεδομένα από τη μνήμη του. Αυτά τα δεδομένα μπορεί να περιλαμβάνουν ευαίσθητες πληροφορίες και η επανειλημμένη εκμετάλλευση παραβιάζει την ασφάλεια του εξυπηρετητή.

Web Application Attacks

Οι επιθέσεις ιστού εκμεταλλεύονται ευάλωτα σημεία σε εφαρμογές ιστού για την απόκτηση μη εξουσιοδοτημένης πρόσβασης, την αλλοτρίωση δεδομένων ή τη διατάραξη υπηρεσιών, [40]. Στοχεύουντας σε αδυναμίες του κώδικα της εφαρμογής ή στους μηχανισμούς ταυτοποίησης, αποτελούν σημαντικές απειλές για την εμπιστευτικότητα, την ακεραιότητα και τη διαθεσιμότητα των δεδομένων και των υπηρεσιών. Κοινές μορφές επιθέσεων σε εφαρμογές ιστού περιλαμβάνουν επιθέσεις Brute Force, Cross-Site Scripting (XSS) και SQL Injection.

11. **Web Attack - Brute Force**: Η εν λόγω επίθεση στοχεύει εφαρμογές ιστού δοκιμάζοντας συστηματικά διάφορους συνδυασμούς ονομάτων χρήστη και κωδικών πρόσβασης αποσκοπώντας στη μη εξουσιοδοτημένη πρόσβαση, [41]. Οι επιτιθέμενοι χρησιμοποιούν αυτοματοποιημένα εργαλεία για τη δημιουργία μεγάλου όγκου προσπαθειών σύνδεσης, εκμεταλλευόμενοι αδύναμες πολιτικές κωδικών πρόσβασης, αλλά και την απουσία μηχανισμών κλειδώματος λογαριασμών. Επιτυχημένες επιθέσεις brute force συνεπάγονται παραβιάσεις συστηματων, κλοπές δεδομένων και περαιτέρω εκμετάλλευση του παραβιασμένου συστήματος. Στρατηγικές πρόληψης περιλαμβάνουν την επιβολή ισχυρών πολιτικών κωδικών πρόσβασης, την εφαρμογή μηχανισμών κλειδώματος λογαριασμών, τη χρήση CAPTCHA και πολυπαραγοντικής ταυτοποίησης (MFA) και τον περιορισμό του ρυθμού προσπαθειών σύνδεσης.

12. **Web Attack - XSS**: Ο όρος Cross-Site Scripting (XSS) αποτελεί μια επίθεση όπου κακόβουλα scripts διαχέονται σε αξιόπιστους ιστότοπους, εκμεταλλευόμενα vulnerabilities εφαρμογών ιστού, [42, 43, 44]. Οι επιθέσεις XSS μπορούν να κλέψουν cookies, session tokens (διακριτικά συνεδρίας) ή άλλες ευαίσθητες πληροφορίες, να χειριστούν το περιεχόμενο της ιστοσελίδας και να εκτελέσουν ενέργειες εκ μέρους των χρηστών. Υπάρχουν τρεις κύριοι τύποι επιθέσεων XSS:

    (a) **Stored XSS**: Επίσης γνωστή ως persistent - επίμονη XSS, αυτή η επίθεση συμβαίνει όταν το κακόβουλο σενάριο αποθηκεύεται μόνιμα στον εξυπηρετητή στόχο, όπως σε μια βάση δεδομένων. Το script εκτελείται στους browsers (περιηγητές) των χρηστών κατά την ανάκτηση αποθηκευμένων δεδομένων, με αποτέλεσμα οι λογαριασμοί τους να τεθούν σε κίνδυνο και κατά συνέπεια τη δυνητική κλοπή ευαίσθητων πληροφοριών.

    (b) **Reflected XSS**: Αυτή η επίθεση συμβαίνει όταν το κακόβουλο script ανακλάται από έναν εξυπηρετητή ιστού και εκτελείται αμέσως στον περιηγητή του χρήστη. Οι ανακλασμένες επιθέσεις XSS συχνά παραδίδονται μέσω phishing emails ή κακόβουλων συνδέσμων.

    (c) **DOM-based XSS**: Αυτή η επίθεση διεξάγεται όταν το ευάλωτο σημείο βρίσκεται στον κώδικα της πλευράς του client και όχι στον κώδικα της πλευράς του server. Η επίθεση εκμεταλλεύεται το περιβάλλον Document Object Model (DOM) στον περιηγητή του χρήστη. Όταν ο χρήστης αλληλεπιδρά με την ιστοσελίδα, το σενάριο τροποποιεί το DOM, οδηγώντας στην εκτέλεση του κακόβουλου κώδικα.

13. **Web Attack - Sql Injection**: Η SQL Injection (SQLi) είναι μια επίθεση κατά την οποία κακόβουλες SQL queries εισάγονται παράνομα προς εκτέλεση, στοχεύοντας το

επίπεδο βάσης δεδομένων μιας εφαρμογής. Αυτή η επίθεση εκμεταλλεύεται την ανεπαρκή προστασία των εισόδων εφαρμογών, επιτρέποντας στους επιτιθέμενους την απόκτηση πρόσβασης σε ευαίσθητα δεδομένα, την τροποποίηση ή τη διαγραφή εγγραφών και την ευρύτερη δυνατότητα διοικητικού ελέγχου της βάσης δεδομένων.

Data Exfiltration and Infiltration

Η διείσδυση αναφέρεται σε έναν εξελιγμένο τύπο κυβερνοεπίθεσης, κατά τον οποίο παραβιάζεται η περιμετρική ασφάλεια ενός δικτύου και αποκτάται μη εξουσιοδοτημένη πρόσβαση σε εσωτερικά συστήματα.

14. **Infiltration**: Κατά τις επιθέσεις διείσδυσης παραβιάζεται η ασφάλεια του δικτύου για την απόκτηση μη εξουσιοδοτημένης πρόσβασης, με αποτέλεσμα την κλοπή δεδομένων, την παραβίαση συστημάτων και την εκμετάλλευση πόρων. Αυτές οι επιθέσεις συχνά παραμένουν μη ανιχνεύσιμες για παρατεταμένες περιόδους, επιτρέποντας στους επιτιθέμενους να διεξάγουν κατασκοπεία, κλοπή δεδομένων, διακοπή λειτουργίας και εγκατάσταση ransomware. Οι επιθέσεις διείσδυσης συνήθως περιλαμβάνουν πολλαπλά στάδια: Αρχικά πραγματοποιείται αναγνώριση, όπου συλλέγονται πληροφορίες για το δίκτυο - στόχο. Στη συνέχεια, ακολουθεί η αρχική παραβίαση, και επακολούθως εγκαθίσταται κακόβουλο λογισμικό ή backdoors για διατήρηση της πρόσβασης. Έπειτα, ακολουθεί μία προσπάθεια κλιμάκωσης προνομίων, για την απόκτηση πρόσβασης υψηλότερου επιπέδου. Τέλος, πραγματοποιείται η εξαγωγή δεδομένων, δηλαδή η κλοπή και μεταφορά ευαίσθητων δεδομένων.

4.3 Περιγραφή των Χαρακτηριστικών

4.3.1 Συλλογή Δεδομένων

Η διαδικασία συλλογής των δεδομένων του CIC-IDS-2017 έχει σχεδιαστεί επακριβώς προκειμένου να αντικατοπτρίζει real-world network environments, καθώς και διάφορα σενάρια επιθέσεων. Σε αυτήν την ενότητα παρουσιάζεται το μεθοδολογικό πλαίσιο εξαγωγής των δεδομένων και των χαρακτηριστικών, δίνοντας έμφαση στα βήματα εξασφάλισης της συνάφειας, της ακρίβειας και της πληρότητας των δεδομένων.

Δημιουργία Δικτυακής Κίνησης Το σύνολο δεδομένων CIC-IDS-2017 προσομοιώνει πραγματικές συνθήκες [45], καταγράφοντας τη δικτυακή κίνηση σε ένα ελεγχόμενο περιβάλλον. Αυτό το περιβάλλον περιλαμβάνει τόσο benign δραστηριότητες, όσο και διάφορα σενάρια κυβερνοεπιθέσεων, με κύριο στόχο τη δημιουργία ρεαλιστικής ροής. Το σύστημα B-Profile, που προτάθηκε από τον Sharafaldin et al. (2016), χρησιμοποιήθηκε για την καταγραφή των ανθρώπινων αλληλεπιδράσεων και τη δημιουργία φυσιολογικής κανονικής κίνησης. Το σύνολο δεδομένων καλύπτει δραστηριότητες 25 χρηστών που χρησιμοποιούν πρωτόκολλα όπως HTTP, HTTPS, FTP, SSH και email.

> *Η περίοδος καταγραφής δεδομένων διήρκεσε πέντε ημέρες, ξεκινώντας στις 9 π.μ. τη Δευτέρα 3 Ιουλίου 2017 και λήγοντας στις 5 μ.μ. την Παρασκευή 7 Ιουλίου 2017. Κάθε ημέρα περιλάμβανε συγκεκριμένους τύπους δραστηριοτήτων και επιθέσεων.*

Καταγραφή Δεδομένων Σύμφωνα με την ιστοσελίδα του CIC [45], η δικτυακή κίνηση καταγράφηκε χρησιμοποιώντας packet sniffers για την δημιουρία PCAP (packet capture)

files. Τα καταγεγραμμένα δεδομένα περιλαμβάνουν τόσο κανονική όσο και κακόβουλη κίνηση, προσομοιώνοντας πραγματικές καταστάσεις.

Τα αρχεία καταγραφής πακέτων (PCAP files) επεξεργάστηκαν χρησιμοποιώντας το CICFlowMeter, προκειμένου να δημιουργηθούν δεδομένα με ροές και ετικέτες. Το CICFlowMeter συγκεντρώνει τα πακέτα σε ροές με βάση τη χρονική σήμανση, τις διευθύνσεις IP πηγής και προορισμού, τις θύρες πηγής και προορισμού, αλλά και τα πρωτόκολλα. Το εργαλείο παράγει αρχεία "Σ" που περιέχουν τα εξαχθέντα χαρακτηριστικά και τις ετικέτες, οι οποίες καταδεικνύουν αν η ροή είναι καλοήθης ή μέρος μιας επίθεσης. Επιπροσθέτως, το σύνολο δεδομένων παρέχει ορισμούς και περιγραφές για αυτά τα χαρακτηριστικά, εξασφαλίζοντας σαφήνεια και συνέπεια στην ερμηνεία των δεδομένων.

Οι τύποι επιθέσεων περιλαμβάνουν Brute Force, DoS, DDoS, Heartbleed, Web επιθέσεις, Infiltration και δραστηριότητες Botnet. Αυτές οι επιθέσεις εκτελέστηκαν σε συγκεκριμένα χρονικά διαστήματα καθ' όλη τη διάρκεια της εβδομάδας. Επιπλέον, τα network configurations παρείχαν τόσο attacker όσο και victim networks, με συγκεκριμένες διευθύνσεις IP να αποδίδονται σε διάφορες συσκευές και διακομιστές (servers). Για παράδειγμα, οι επιτιθέμενοι χρησιμοποιούσαν τις IPs 205.174.165.69, 70 και 71, ενώ τα θύματα βρίσκονταν εντός του υποδικτύου 192.168.10.x.

Πίνακας 4.1: Χρονοδιάγραμμα της Συλλογής Δεδομένων

Ημερομηνία		Θύμα	IP Θύματος
Δευτέρα, 3 Ιουλίου, 2017	**Benign Activity**		
Τρίτη, 4 Ιουλίου, 2017	**Brute Force**: *FTP-Patator & SSH-Patator*	WebServer Ubuntu	205.174.165.68
Τετάρτη, 5 Ιουλίου, 2017	**DoS/DDoS**: *Slowhttptest, slowloris, Hulk & GoldenEye*	WebServer Ubuntu	205.174.165.68
		Ubuntu12 server	192.168.10.25
	Heartbleed	Ubuntu12 server	205.174.165.66
Πέμπτη, 6 Ιουλίου, 2017	**Web Attacks Infiltration**	WebServer Ubuntu	205.174.165.68
		Windows Vista	192.168.10.8
		MAC	192.168.10.25
Παρασκευή, 7 Ιουλίου, 2017	**Botnet ARES**	Windows machines	πολλαπλές
	DDoS LOIT	Ubuntu16 server	205.174.165.68

Σχεδόν όλες οι επιθέσεις εξαπολύθηκαν από Kali attacker machine (IP: 205.174.165.73), με την εξαίρεση της DDoS LOIT η οποία περιλάμβανε τρεις Windows 8.1 machines (IPs: 205.174.165.69 - 71).

Γενικότερα, η καταγραφή δεδομένων έλαβε χώρα σε μια αυστηρά ελεγχόμενη χρονική περίοδο, εξασφαλίζοντας τη σύλληψη ποικίλων τύπων δραστηριότητας και επιθέσεων, παρέχοντας έτσι μια ολοκληρωμένη ανάλυση και αξιολόγηση των δυνατοτήτων ανίχνευσης εισβολών συστημάτων μηχανικής και βαθιάς μάθησης.

4.3.2 Περιγραφή Χαρακτηριστικών

Σε αυτήν την ενότητα, παρουσιάζονται τα εξαχθέντα 83 χαρακτηριστικά του συνόλου δεδομένων CIC-IDS-2017, [27, 46], στα οποία βασίζεται το εγχείρημα της ανίχνευσης και κατανόησης μοτίβων δικτυακής εισβολής. Αυτά τα χαρακτηριστικά περιγράφουν μια σειρά από ιδιότητες δικτυακής κίνησης, αποτυπώνοντας διάφορες πτυχές των packet flows, όπως τον χρόνο, το μέγεθος, τη συχνότητα και τα πρότυπα αλληλεπίδρασης. Παρακάτω παρατίθεται μια λεπτομερής ανάλυση κάθε χαρακτηριστικού, οργανωμένη σε κατηγορίες για λόγους σαφήνειας.

Basic Features

Τα βασικά χαρακτηριστικά παρέχουν θεμελιώδεις πληροφορίες για τη δικτυακή ροή, αποτυπώνοντας βασικά στοιχεία περιγραφής όπως τον τύπο της σύνδεσης και τη διάρκεια της επικοινωνίας. Αυτά τα χαρακτηριστικά είναι κρίσιμα για την αρχική ανάλυση και ταξινόμηση της δικτυακής κίνησης, προσφέροντας πληροφορίες για τα θεμελιώδη στοιχεία κάθε ροής.

1. **Protocol**: Το πρωτόκολλο που χρησιμοποιείται στη ροή δικτύου.

   Το πρωτόκολλο υποδεικνύει το ευρύτερο σύνολο κανόνων ή συμβάσεων επικοινωνίας μεταξύ δικτυακών συσκευών. Κοινά πρωτόκολλα περιλαμβάνουν το **TCP** (Transmission Control Protocol), **UDP** (User Datagram Protocol) και **ICMP** (Internet Control Message Protocol). Η κατανόηση του πρωτοκόλλου είναι κρίσιμη για την αναγνώριση του τύπου της επικοινωνίας που πραγματοποιείται, καθώς διαφορετικά πρωτόκολλα εξυπηρετούν διαφορετικούς σκοπούς (π.χ., το TCP χρησιμοποιείται για αξιόπιστη μεταφορά δεδομένων, ενώ το UDP χρησιμοποιείται για low-latency επικοινωνία). Η ανάλυση του πρωτοκόλλου βοηθά στην ταξινόμηση της δικτυακής κίνησης και στον εντοπισμό πιθανών security threats που συνδέονται με συγκεκριμένα πρωτόκολλα.

2. **Flow Duration**: Η διάρκεια της ροής, μετρημένη σε μικροδευτερόλεπτα.

   Η διάρκεια της ροής προσμετρά το συνολικό χρονικό διάστημα μιας δικτυακής ροής, από την έναρξη του πρώτου πακέτου μέχρι τη λήξη του τελευταίου πακέτου. Αποτελεί ένα βασικό χαρακτηριστικό για την κατανόηση της διάρκειας και της σταθερότητας των δικτυακών συνεδριών. Οι μακροχρόνιες ροές μπορεί να υποδεικνύουν διαρκείς συνδέσεις όπως είναι το streaming ή οι μεταφορές αρχείων, ενώ οι βραχυχρόνιες ροές μπορεί να αντιπροσωπεύουν γρήγορες συναλλαγές ή προσπάθειες ανίχνευσης του δικτύου. Η ανάλυση της διάρκειας της ροής βοηθά στην αναγνώριση της φύσης της επικοινωνίας και στον εντοπισμό ανωμαλιών, όπως ασυνήθιστα μεγάλες ή μικρές συνεδρίες που μπορεί να υποδεικνύουν πιθανές απειλές ασφαλείας.

Packet & Length Features

Τα χαρακτηριστικά πακέτων και μήκους παρέχουν λεπτομερείς πληροφορίες για το μέγεθος και το πλήθος των πακέτων που μεταδίδονται προς τις δύο κατευθύνσεις (προς τα εμπρός και προς τα πίσω) μέσα σε μια ροή δικτύου. Αυτά τα χαρακτηριστικά είναι απαραίτητα για την κατανόηση της μεταφοράς δεδομένων, συμπεριλαμβανομένου του όγκου, της κατανομής και της μεταβλητότητας των μεγεθών των πακέτων, τα οποία συνδράμουν στον εντοπισμό ανωμαλιών και στη βελτιστοποίηση των επιδόσεων του δικτύου.

Forward Packets: Packets transmitted from the source to the destination within a flow.
Backward Packets: Packets transmitted from the destination back to the source.

3. **Total Fwd Packets**: Συνολικός αριθμός forward πακέτων.

   Το εν λόγω χαρακτηριστικό προσμετρά όλα τα πακέτα που μεταδίδονται από την πηγή προς τον προορισμό μέσα σε μια ροή. Βοηθά στην κατανόηση της έντασης και του όγκου της μετάδοσης δεδομένων προς τα εμπρός. Ένας μεγάλος αριθμός forward πακέτων μπορεί να υποδεικνύει bulk data transfers (μαζικές μεταφορές δεδομένων), συχνή επικοινωνία ή δραστηριότητες streaming.

4. **Total Backward Packets**: Συνολικός αριθμός backward πακέτων.

   Το εν λόγω χαρακτηριστικό προσμετρά όλα τα πακέτα που μεταδίδονται από τον προορισμό πίσω στην πηγή. Παρέχει πληροφορίες για τον όγκο της επιστρεφόμενης κίνησης. Ένας μεγάλος αριθμός πακέτων προς τα πίσω μπορεί να υποδεικνύει συχνές επιβεβαιώσεις παραλαβής, αποκρίσεις ή δεδομένα που αποστέλλονται πίσω στην πηγή.

5. **Total Length of Fwd Packets**: Συνολικό μέγεθος (σε bytes) όλων των forward πακέτων.

   Το εν λόγω χαρακτηριστικό αθροίζει τα μεγέθη όλων των fowrard πακέτων, παρέχοντας ένα μέτρο του συνολικού όγκου δεδομένων που αποστέλλονται προς τα εμπρός. Βοηθά στην κατανόηση του φορτίου και των χαρακτηριστικών μεταφοράς forward δεδομένων, τα οποία μπορούν να υποδείξουν τη φύση της εφαρμογής ή της υπηρεσίας που δημιουργεί τη δικτυακή κίνηση.

6. **Total Length of Bwd Packets**: Συνολικό μέγεθος (σε bytes) όλων των backward πακέτων.

   Το εν λόγω χαρακτηριστικό αθροίζει τα μεγέθη όλων των backward πακέτων, παρέχοντας ένα μέτρο του συνολικού όγκου δεδομένων που αποστέλλονται προς τα πίσω. Βοηθά στην κατανόηση του φορτίου απόκρισης και των χαρακτηριστικών μεταφοράς δεδομένων από τον προορισμό.

7. **Fwd Packet Length Max**: Μέγιστο μέγεθος (σε bytes) forward πακέτων.

   Το εν λόγω χαρακτηριστικό καταγράφει το μέγεθος του μεγαλύτερου forward πακέτου. Τα μεγάλα μεγέθη πακέτων μπορεί να υποδεικνύουν μαζικές μεταφορές δεδομένων ή την παρουσία bulk (μεγάλων) φορτίων, και επακολούθως μεταφορές αρχείων ή υπηρεσίες streaming (συνεχούς ροής).

8. **Fwd Packet Length Min**: Ελάχιστο μέγεθος (σε bytes) ενός forward πακέτου.

   Το εν λόγω χαρακτηριστικό καταγράφει το μέγεθος του μικρότερου forward πακέτου. Τα μικρά μεγέθη πακέτων συχνά υποδεικνύουν μηνύματα ελέγχου, acknowledgments ή άλλες low-data επικοινωνίες.

9. **Fwd Packet Length Mean**: Μέσο μέγεθος (σε bytes) των forward πακέτων.

   Το εν λόγω χαρακτηριστικό υπολογίζει το μέσο μέγεθος των forward πακέτων. Αυτή η μέση τιμή βοηθά στην κατανόηση του τυπικού μεγέθους πακέτων για τη ροή δεδομένων, παρέχοντας πληροφορίες για τη φύση της επικοινωνίας και της συμπεριφοράς της εφαρμογής.

10. **Fwd Packet Length Std**: Τυπική απόκλιση μεγεθών των forward πακέτων.

    Το εν λόγω χαρακτηριστικό μετρά τη μεταβλητότητα στα μεγέθη των forward πακέτων. Μια υψηλή τυπική απόκλιση υποδεικνύει ένα ευρύ φάσμα μεγεθών, υποδεικνύοντας διαφορετικούς τύπους κίνησης. Μια χαμηλή τυπική απόκλιση υποδηλώνει πιο ομοιόμορφα μεγέθη πακέτων.

11. **Bwd Packet Length Max**: Μέγιστο μέγεθος (σε bytes) ενός backward πακέτου στην κατεύθυνση επιστροφής.

Το εν λόγω χαρακτηριστικό καταγράφει το μέγεθος του μεγαλύτερου backward πακέτου. Τα μεγάλα πακέτα στην κατεύθυνση απόκρισης μπορούν να υποδεικνύουν σημαντικά δεδομένα που αποστέλλονται πίσω στην πηγή, γεγονός σύνηθες στις απαντήσεις με μεγάλο όγκο δεδομένων.

12. **Bwd Packet Length Min**: Ελάχιστο μέγεθος (σε bytes) ενός backward πακέτου.

    Το εν λόγω χαρακτηριστικό καταγράφει το μέγεθος του μικρότερου backward πακέτου, όπως μηνύματα ελέγχου ή επιβεβαιώσεις (acknowledgments).

13. **Bwd Packet Length Mean**: Μέσο μέγεθος (σε bytes) backward πακέτων στην κατεύθυνση επιστροφής.

    Το εν λόγω χαρακτηριστικό υπολογίζει το μέσο μέγεθος των backward πακέτων. Αυτή η μέση τιμή παρέχει πληροφορίες για το τυπικό μέγεθος των πακέτων απόκρισης.

14. **Bwd Packet Length Std**: Τυπική απόκλιση των μεγεθών των backward πακέτων.

    Το εν λόγω χαρακτηριστικό μετρά τη μεταβλητότητα στα μεγέθη των backward πακέτων. Μια υψηλή τυπική απόκλιση υποδεικνύει ένα ευρύ φάσμα μεγεθών, ενώ μια χαμηλή τυπική απόκλιση υποδηλώνει πιο ομοιόμορφα μεγέθη.

Flow-based Features

Τα flow-based χαρακτηριστικά παρέχουν μια συνολική εικόνα της μετάδοσης δεδομένων σε μια δικτυακή ροή. Αυτά τα χαρακτηριστικά περιλαμβάνουν τόσο τον ρυθμό μεταφοράς δεδομένων όσο και τα χρονικά διαστήματα αναμεταξύ των πακέτων, προσφέροντας πληροφορίες για τη συνολική συμπεριφορά και απόδοση της δικτυακής κίνησης.

15. **Flow Bytes/s**: Αριθμός bytes ανά δευτερόλεπτο στη ροή.

    Το εν λόγω χαρακτηριστικό προσμετρά τον ρυθμό μεταφοράς δεδομένων ροής σε bytes ανά δευτερόλεπτο. Παρέχει πληροφορίες για το throughput της δικτυακής κίνησης, υποδεικνύοντας πόσα δεδομένα μεταδίδονται με την πάροδο του χρόνου. Υψηλές τιμές μπορεί να υποδεικνύουν μεγάλες μεταφορές δεδομένων, streaming ή λήψεις, ενώ χαμηλές τιμές μπορεί να υποδεικνύουν κίνηση ελέγχου ή περιόδους αδράνειας.

16. **Flow Packets/s**: Αριθμός πακέτων ανά δευτερόλεπτο στη ροή.

    Το εν λόγω χαρακτηριστικό υπολογίζει τον ρυθμό μετάδοσης πακέτων ροής ανά δευτερόλεπτο. Βοηθά στην κατανόηση της συχνότητας ανταλλαγής πακέτων, παρέχοντας πληροφορίες για την ένταση της επικοινωνίας. Υψηλές τιμές μπορεί να υποδεικνύουν γρήγορες ανταλλαγές δεδομένων ή εκρήξεις δραστηριότητας, ενώ χαμηλές τιμές μπορεί να αντικατοπτρίζουν σποραδική επικοινωνία.

Το ακρωνύμιο IAT: Inter-Arrival Time μεταφράζεται ως Χρόνος Μεταξύ Αφίξεων, ο οποίος είναι το χρονικό διάστημα μεταξύ διαδοχικών πακέτων που φθάνουν σε ένα σημείο του δικτύου.

17. **Flow IAT Mean**: Μέσο χρονικό διάστημα μεταξύ πακέτων στη ροή.

Ο μέσος Inter-Arrival Time (IAT) προσμετρά το μέσο χρονικό διάστημα μεταξύ διαδοχικών πακέτων εντός της ροής. Παρέχει μια αίσθηση της κανονικότητας και του ρυθμού μετάδοσης πακέτων. Σταθερές τιμές του μέσου υποδεικνύουν κανονικές ροές δεδομένων, ενώ μεταβαλλόμενες τιμές μπορεί να υποδηλώνουν ανώμαλα πρότυπα κίνησης.

18. **Flow IAT Std**: Τυπική απόκλιση του χρονικού διαστήματος μεταξύ πακέτων στη ροή.

    Η τυπική απόκλιση του IAT μετρά τη μεταβλητότητα στα χρονικά διαστήματα μεταξύ πακέτων. Υψηλές τιμές τυπικής απόκλισης υποδεικνύουν σημαντικές διακυμάνσεις στους χρόνους άφιξης των πακέτων, υποδηλώνοντας εκρηκτική ή ακανόνιστη κίνηση. Χαμηλές τιμές υποδεικνύουν συνεπή χρονικά διαστήματα μετάδοσης πακέτων.

19. **Flow IAT Max**: Μέγιστο χρονικό διάστημα μεταξύ πακέτων στη ροή.

    Το εν λόγω χαρακτηριστικό καταγράφει το μεγαλύτερο χρονικό διάστημα που παρατηρείται μεταξύ διαδοχικών πακέτων εντός της ροής, υπογραμμίζοντας περιόδους αδράνειας ή καθυστερήσεις στη μετάδοση δεδομένων. Υψηλές τιμές μέγιστου IAT μπορεί να υποδεικνύουν πιθανή συμφόρηση δικτύου, καθυστερήσεις ή προβλήματα σε application-level.

20. **Flow IAT Min**: Ελάχιστο χρονικό διάστημα μεταξύ πακέτων στη ροή.

    Το εν λόγω χαρακτηριστικό καταγράφει το μικρότερο χρονικό διάστημα που παρατηρείται μεταξύ διαδοχικών πακέτων εντός της ροής. Υπογραμμίζει το ταχύτερο δυνατό χρονικό διάστημα μετάδοσης πακέτων, υποδεικνύοντας άμεση επικοινωνία ή μετάδοση δεδομένων υψηλής ταχύτητας.

Forward Flow IAT Features

Τα χαρακτηριστικά IAT (Inter-Arrival Time) forward ροής εστιάζουν ειδικά στα χρονικά χαρακτηριστικά των πακέτων που αποστέλλονται από την πηγή στον προορισμό. Αυτά τα χαρακτηριστικά παρέχουν λεπτομερείς πληροφορίες για τη δυναμική της μετάδοσης δεδομένων στην κατεύθυνση προώθησης, η οποία είναι κρίσιμη για την κατανόηση της χρονικής απόκρισης και της απόδοσης της δικτυακής κίνησης.

21. **Fwd IAT Total**: Συνολικό χρονικό διάστημα (σε μικροδευτερόλεπτα) forward πακέτων.

    Το εν λόγω χαρακτηριστικό προσμετρά το συνολικό χρονικό διάστημα μεταξύ όλων των διαδοχικών πακέτων ροής που αποστέλλονται από την πηγή προς τον προορισμό. Παρέχει μια αθροιστική εικόνα της συνολικής διάρκειας και του ρυθμού της ροής δεδομένων στην κατεύθυνση προώθησης.

22. **Fwd IAT Mean**: Μέσο χρονικό διάστημα μεταξύ forward πακέτων.

    Ο μέσος IAT υπολογίζει το μέσο χρονικό διάστημα μεταξύ διαδοχικών forward πακέτων και βοηθά στην κατανόηση της τάξης και της συχνότητας της μετάδοσης πακέτων. Σταθερές μέσες τιμές υποδεικνύουν τακτικά χρονικά διαστήματα πακέτων, ενώ μεταβαλλόμενες μέσες τιμές υποδηλώνουν μεταβλητά πρότυπα μετάδοσης.

23. **Fwd IAT Std**: Τυπική απόκλιση του χρονικού διαστήματος μεταξύ forward πακέτων.

    Το εν λόγω χαρακτηριστικό προσμετρά τη μεταβλητότητα στα χρονικά διαστήματα μεταξύ forward πακέτων. Μια υψηλή τυπική απόκλιση υποδεικνύει σημαντικές διακυμάνσεις

στους χρόνους μετάδοσης, υποδηλώνοντας εκρηκτική ή ακανόνιστη κίνηση, ενώ μια χαμηλή τυπική απόκλιση υποδηλώνει συνεπή και σταθερά χρονικά διαστήματα πακέτων.

24. **Fwd IAT Max**: Μέγιστο χρονικό διάστημα μεταξύ forward πακέτων.

    Η μέγιστη τιμή IAT καταγράφει το μεγαλύτερο παρατηρούμενο χρονικό διάστημα μεταξύ forward διαδοχικών πακέτων, υπογραμμίζοντας περιόδους αδράνειας ή καθυστερήσεις της ροής δεδομένων. Υψηλές τιμές μέγιστου IAT μπορεί να υποδεικνύουν δικτυακή συμφόρηση, καθυστερήσεις σε επίπεδο εφαρμογής ή άλλα ζητήματα μετάδοσης.

25. **Fwd IAT Min**: Ελάχιστο χρονικό διάστημα μεταξύ forward πακέτων.

    Η ελάχιστη τιμή IAT καταγράφει το μικρότερο παρατηρούμενο χρονικό διάστημα μεταξύ διαδοχικών forward πακέτων, υπογραμμίζοντας τον ταχύτερο δυνατό ρυθμό μετάδοσης της ροής. Χαμηλές τιμές ελάχιστου IAT μπορεί να υποδεικνύουν περιόδους έντονης δραστηριότητας ή ταχείες ανταλλαγές δεδομένων.

Backward Flow IAT Features

Τα χαρακτηριστικά IAT (Inter-Arrival Time) backward ροής εστιάζουν στα χρονικά χαρακτηριστικά των πακέτων που αποστέλλονται από τον προορισμό πίσω στην πηγή. Αυτά τα χαρακτηριστικά είναι κρίσιμα για την κατανόηση της δυναμικής απόκρισης και της απόδοσης της δικτυακής κίνησης στην αντίστροφη κατεύθυνση.

26. **Bwd IAT Total**: Συνολικό χρονικό διάστημα (σε μικροδευτερόλεπτα) backward πακέτων.

    Το εν λόγω χαρακτηριστικό προσμετρά το συνολικό χρονικό διάστημα μεταξύ όλων των διαδοχικών πακέτων ροής που αποστέλλονται από τον προορισμό πίσω στην πηγή. Παρέχει μια αθροιστική εικόνα του συνολικού χρόνου μετάδοσης στην κατεύθυνση επιστροφής, βοηθώντας στην κατανόηση της διάρκειας και του ρυθμού της ροής απόκρισης δεδομένων.

27. **Bwd IAT Mean**: Μέσο χρονικό διάστημα μεταξύ backward πακέτων.

    Ο μέσος IAT υπολογίζει το μέσο χρονικό διάστημα μεταξύ διαδοχικών backward πακέτων και βοηθά στην κατανόηση της τάξης και της συχνότητας της μετάδοσης πακέτων απόκρισης. Σταθερές μέσες τιμές υποδεικνύουν τακτικά χρονικά διαστήματα, ενώ μεταβαλλόμενες μέσες τιμές υποδηλώνουν μεταβλητά πρότυπα απόκρισης.

28. **Bwd IAT Std**: Τυπική απόκλιση του χρονικού διαστήματος μεταξύ backward πακέτων.

    Το εν λόγω χαρακτηριστικό προσμετρά τη μεταβλητότητα στα χρονικά διαστήματα μεταξύ backward πακέτων. Μια υψηλή τυπική απόκλιση υποδεικνύει σημαντικές διακυμάνσεις στους χρόνους απόκρισης, υποδηλώνοντας ακανόνιστη κίνηση ή μεταβαλλόμενους χρόνους απόκρισης του διακομιστή, ενώ μια χαμηλή τυπική απόκλιση υποδηλώνει συνεπή και σταθερά χρονικά διαστήματα απόκρισης.

29. **Bwd IAT Max**: Μέγιστο χρονικό διάστημα μεταξύ backward πακέτων.

    Η μέγιστη τιμή IAT καταγράφει το μεγαλύτερο παρατηρούμενο χρονικό διάστημα μεταξύ backward διαδοχικών πακέτων, υπογραμμίζοντας περιόδους αδράνειας ή καθυστερήσεις

της ροής απόκρισης. Υψηλές τιμές μέγιστου IAT μπορεί να υποδεικνύουν συμφόρηση δικτύου, καθυστερήσεις στην επεξεργασία του διακομιστή ή άλλα ζητήματα.

30. **Bwd IAT Min**: Ελάχιστο χρονικό διάστημα μεταξύ backward πακέτων.

    Η ελάχιστη τιμή IAT καταγράφει το μικρότερο παρατηρούμενο χρονικό διάστημα μεταξύ διαδοχικών backward πακέτων, υπογραμμίζοντας τον ταχύτερο δυνατό ρυθμό μετάδοσης της ροής και υποδεικνύοντας άμεση επικοινωνία ή μετάδοση δεδομένων υψηλής ταχύτητας. Χαμηλές τιμές ελάχιστου IAT μπορεί να υποδεικνύουν περιόδους έντονης δραστηριότητας ή ταχείες ανταλλαγές δεδομένων.

Packet-based Features

Τα packet-based χαρακτηριστικά παρέχουν πληροφορίες για συγκεκριμένες ιδιότητας και συμπεριφορές των πακέτων δικτυακής ροής, επικεντρώνοντας ιδιαίτερα στη χρήση συγκεκριμένων **TCP flags**, καθώς και στις επικεφαλίδες των πακέτων. Είναι απαραίτητα για την κατανόηση των μηχανισμών ελέγχου, καταστάσεων επείγουσας ανάγκης και της δομής της μετάδοσης δεδομένων στο δίκτυο.

31. **Fwd PSH Flags**: Πλήθος forward πακέτων με ενεργοποιημένη PSH flag.

    Η **PSH** (**Push**) flag δίνει εντολή στο σύστημα λήψης να μεταβιβάσει αμέσως τα δεδομένα στην εφαρμογή, αντί να περιμένει μέχρις ότου γεμίσει η προσωρινή μνήμη. Ένας υψηλός αριθμός από forward PSH flags υποδηλώνει συχνές άμεσες μεταφορές δεδομένων από την πηγή, γεγονός τυπικό σε διαδραστικές εφαρμογές όπου τα δεδομένα υποβάλλονται σε επεξεργασία αφότου φτάσουν.

32. **Bwd PSH Flags**: Πλήθος backward πακέτων με ενεργοποιημένη PSH flag

    Παρόμοια με την κατεύθυνση προώθησης, ένας υψηλός αριθμός backward PSH flags υποδηλώνει ότι ο προορισμός συχνά χρειάζεται τα δεδομένα να υποβάλλονται σε άμεση επεξεργασία, ως δυνητική απάντηση σε επείγοντα αιτήματα από την πηγή ή λόγω εφαρμογών πραγματικού χρόνου.

33. **Fwd URG Flags**: Πλήθος forward πακέτων με ενεργοποιημένη URG flag

    Η **URG** (**Urgent**) flag υποδεικνύει ότι πρέπει να δοθεί προτεραιότητατα στα δεδομένα του πακέτου και ως εκ τούτου να υποβληθούν σε άμεση επεξεργασία. Ένας υψηλός αριθμός από forward URG flags υποδηλώνει ότι η πηγή συχνά αποστέλλει δεδομένα υψηλής προτεραιότητας, τα οποία μπορεί να είναι κρίσιμες πληροφορίες ή μηνύματα ελέγχου που απαιτούν άμεση προσοχή.

34. **Bwd URG Flags**: Πλήθος backward πακέτων με ενεργοποιημένη URG flag.

    Στην κατεύθυνση επιστροφής, ένας υψηλός αριθμός από URG flags υποδηλώνει ότι ο προορισμός συχνά χρειάζεται να στέλνει δεδομένα υψηλής προτεραιότητας πίσω στην πηγή, υποδεικνύοντας μηνύματα ελέγχου, αναφορές σφαλμάτων ή άλλες κρίσιμες επικοινωνίες που χρειάζονται άμεση επεξεργασία.

35. **Fwd Header Length**: Συνολικά bytes επικεφαλίδων σε forward πακέτα.

    Το εν λόγω χαρακτηριστικό προσμετρά το συνολικό μέγεθος των επικεφαλίδων σε όλα τα

forward πακέτα. Παρέχει πληροφορίες για την επιβάρυνση που σχετίζεται με τις επικεφαλίδες πακέτων, περιλαμβάνοντας πληροφορίες απαραίτητες για δρομολόγηση και έλεγχο σφαλμάτων. Μεγαλύτερο μήκος επικεφαλίδας συνήθως υποδεικνύει σύνθετα πρωτόκολλα επικοινωνίας ή υψηλότερη επιβάρυνση ελέγχου.

36. **Bwd Header Length**: Συνολικά bytes επικεφαλίδων σε backward πακέτα.

    Παρόμοια με την κατεύθυνση προώθησης, το εν λόγω χαρακτηριστικό προσμετρά το συνολικό μέγεθος των επικεφαλίδων σε backward πακέτα. Η ανάλυση του μήκους της επικεφαλίδας βοηθά στην κατανόηση της επιβάρυνσης ελέγχου και δρομολόγησης απαντήσεων, παρέχοντας πληροφορίες για την πολυπλοκότητα και την αποδοτικότητα της επικοινωνίας.

Packet Rate Features

Τα εν λόγω χαρακτηριστικά επικεντρώνονται στον ρυθμό μετάδοσης των πακέτων σε κάθε κατεύθυνση δικτυακής ροής και είναι καθοριστικά για την κατανόηση της έντασης και της συμπεριφοράς της μετάδοσης δεδομένων, παρέχοντας πολύτιμες πληροφορίες για την απόδοση του δικτύου, τα μοτίβα κίνησης και πιθανές ανωμαλίες.

37. **Fwd Packets/s**: Αριθμός forward πακέτων ανά δευτερόλεπτο.

    Το εν λόγω χαρακτηριστικό προσμετρά το ρυθμό αποστολής forward πακέτων ανά δευτερόλεπτο. Ένας υψηλός ρυθμός πακέτων προώθησης μπορεί να υποδεικνύει έντονη μετάδοση δεδομένων από την πηγή, γεγονός τυπικό σε εφαρμογές όπως οι μεταφορές αρχείων ή το streaming.

38. **Bwd Packets/s**: Αριθμός backward πακέτων ανά δευτερόλεπτο.

    Το εν λόγω χαρακτηριστικό προσμετρά τον ρυθμό αποστολής backward πακέτων ανά δευτερόλεπτο. Ένας υψηλός ρυθμός πακέτων επιστροφής συχνά υποδεικνύει υψηλό επίπεδο δραστηριότητας επιβεβαίωσης ή απόκρισης από τον προορισμό, γεγονός τυπικό σε εφαρμογές που απαιτούν συχνές ενημερώσεις κατάστασης ή επιβεβαιώσεις. Η ανάλυση του ρυθμού αυτού είναι σημαντική για την κατανόηση της ανταπόκρισης του προορισμού και τον εντοπισμό πιθανών προβλημάτων ή επιθέσεων στο δίκτυο.

Packet Size Features

Τα packet size features παρέχουν ουσιαστικές πληροφορίες για την κατανομή και τη μεταβλητότητα του μήκους των πακέτων και συνδράμουν στην κατανόηση του τρόπου μετάδοσης δεδομένων, στην ανίχνευση ανωμαλιών και στη βελτιστοποίηση της απόδοσης του δικτύου.

39. **Min Packet Length**: Το ελάχιστο παρατηρηθέν μήκος πακέτου.

    Το εν λόγω χαρακτηριστικό καταγράφει το ελάχιστο παρατηρηθέν μέγεθος πακέτου, συμβάλλοντας στον εντοπισμό της παρουσίας πολύ μικρών πακέτων, τα οποία μπορεί να υποδηλώνουν πακέτα ελέγχου, επιβεβαιώσεις (acknowledgments) ή ενδεχομένως κατακερματισμένα δεδομένα (fragmented data).

40. **Max Packet Length**: Το μέγιστο παρατηρηθέν μήκος πακέτου.

    Το εν λόγω χαρακτηριστικό καταγράφει το μεγαλύτερο παρατηρηθέν μέγεθος πακέτου,

υποδεικνύοντας μαζικές μεταφορές δεδομένων ή δεδομένα συνεχούς ροής και παρέχοντας πληροφορίες για τη φύση των εφαρμογών και των υπηρεσιών που δημιουργούν την κίνηση.

41. **Packet Length Mean**: Μέσο παρατηρηθέν μήκος πακέτου.

    Το εν λόγω χαρακτηριστικό καταγράφει το μέσο μέγεθος των πακέτων της ροής, μέτρηση χρήσιμη για την κατανόηση του τυπικού μεγέθους των πακέτων μετάδοσης, στα πλαίσια εντοπισμού του τύπου της κίνησης.

42. **Packet Length Std**: Τυπική απόκλιση παρατηρηθέντων μηκών των πακέτων.

    Το εν λόγω χαρακτηριστικό προσμετρά τη μεταβλητότητα των μεγεθών των πακέτων εντός της ροής, υπολογίζοντας την τυπική απόκλιση. Μια υψηλή τυπική απόκλιση υποδεικνύει ένα ευρύ φάσμα μεγεθών πακέτων, υποδηλώνοντας διάφορους τύπους κίνησης, ενώ μια χαμηλή τυπική απόκλιση υποδηλώνει περισσότερη ομοιομορφία στα μεγέθη πακέτων.

43. **Packet Length Variance**: Διακύμανση του παρατηρηθέντος μήκους πακέτων ροής.

    Το εν λόγω χαρακτηριστικό υπολογίζει την διακύμανση στα μεγέθη των πακέτων ροής, παρέχοντας ένα ακόμη μέτρο της διασποράς του μήκους των πακέτων. Η διακύμανση, όπως και η τυπική απόκλιση, υποδεικνύει συνέπεια ή μεταβλητότητα στο μέγεθος των πακέτων.

Flag-based Features

Τα εν λόγω flags είναι καθοριστικά για την κατανόηση της κατάστασης και της συμπεριφοράς των συνδέσεων TCP, συμβάλλοντας στη διαχείριση της ροής δεδομένων, την εγκαθίδρυση και τον τερματισμό συνδέσεων και τον έλεγχο της συμφόρησης στο δίκτυο.

44. **FIN Flag Count**: Αριθμός πακέτων με ενεργοποιημένη την FIN flag.

    FIN (Finish): Η FIN flag χρησιμοποιείται στο TCP (Transmission Control Protocol) για να δηλώσει ότι ο αποστολέας έχει ολοκληρώσει την αποστολή δεδομένων και θέλει να τερματίσει τη σύνδεση. Η εν λόγω flag είναι μέρος του TCP three-way handshake που χρησιμοποιείται για τον τερματισμό μιας σύνδεσης.

45. **SYN Flag Count**: Αριθμός πακέτων με ενεργοποιημένη την SYN flag.

    SYN (Synchronize): Η SYN flag χρησιμοποιείται για την έναρξη μιας TCP σύνδεσης. Όταν πραγματοποιείται ένα αίτημα σύνδεσης, η SYN flag ανατίθεται στο πρώτο πακέτο. Ομοίως, αποτελεί μέρος του TCP three-way handshake που χρησιμοποιείται για την εγκαθίδρυση μιας σύνδεσης.

46. **RST Flag Count**: Αριθμός πακέτων με ενεργοποιημένη την RST flag.

    RST (Reset): Η RST flag χρησιμοποιείται για την απότομη διακοπή μιας σύνδεσης και μπορεί να αποσταλεί από οποιαδήποτε πλευρά προς άμεση επαναρύθμιση της σύνδεσης.

47. **PSH Flag Count**: Αριθμός πακέτων με ενεργοποιημένη την PSH flag.

PSH (Push): Η PSH flag δίνει εντολή στο σύστημα λήψης να μεταβιβάσει αμέσως τα δεδομένα στην εφαρμογή και χρησιμοποιείται για να διασφαλίσει ότι τα δεδομένα μεταδίδονται άμεσα, δίχως αναμονή προς κατάκλυση της προσωρινής μνήμης (buffer).

48. **ACK Flag Count**: Αριθμός πακέτων με ενεργοποιημένη την ACK flag.

ACK (Acknowledgment): Η ACK flag χρησιμοποιείται για την επιβεβαίωση της λήψης ενός πακέτου. Κάθε byte δεδομένων που αποστέλλεται αριθμείται και ο αριθμός επιβεβαίωσης χρησιμοποιείται για να υποδείξει το επόμενο byte που αναμένεται.

49. **URG Flag Count**: Αριθμός πακέτων με ενεργοποιημένη την URG flag.

URG (Urgent): Η URG flag υποδεικνύει ότι τα δεδομένα που περιέχονται στο πακέτο πρέπει να υποβληθούν σε άμεση επεξεργασία, ενώ χρησιμοποιείται σε συνδυασμό με το Urgent Pointer field για να δώσει προτεραιότητα στη διαχείριση συγκεκριμένων δεδομένων.

50. **CWR Flag Count**: Αριθμός πακέτων με ενεργοποιημένη την CWR flag.

CWR (Congestion Window Reduced): Η CWR flag χρησιμοποιείται στην Explixit Congestion Notification (ECN) ώστε να σηματοδοτήσει ότι το παράθυρο συμφόρησης του αποστολέα έχει μειωθεί για τη διαχείριση της δικτυακής συμφόρησης.

51. **ECE Flag Count**: Αριθμός πακέτων με ενεργοποιημένη την ECE flag.

ECE (ECN-Echo): Η σημαία ECE (ECN-Echo) είναι μέρος του μηχανισμού Explicit Congestion Notification (ECN) και υποδεικνύει ότι ο TCP peer, ως ECN-capable, έχει λάβει ένα πακέτο με την Congestion Experienced (CE) flag ενεργοποιημένη, το οποίο ενημερώνει τον αποστολέα για τη συμφόρηση στο δίκτυο.

Header-based Features

Τα header-based features παρέχουν πληροφορίες για τη σύνθεση και τη συμπεριφορά των πακέτων. Αυτά τα χαρακτηριστικά βοηθούν στην κατανόηση της δομής, του μεγέθους και των γνωρισμάτων της δικτυακής ροής, δηλαδή των προτύπων και των συμπεριφορών της δικτυακής κίνησης.

52. **Down/Up Ratio**: Αναλογία Download / Upload.

Το εν λόγω χαρακτηριστικό προσμετρά την αναλογία των downloaded bytes προς τα uploaded bytes της δικτυακής ροής και συμβάλλει στην κατανόηση της ισορροπίας μεταξύ εισερχόμενης και εξερχόμενης κίνησης, υποδεικνύοντας διαφορετικούς τύπους δικτυακών δραστηριοτήτων όπως heavy downloads, streaming ή εξαγωγή δεδομένων.

53. **Avg Packet Size**: Μέσο μέγεθος πακέτου ροής.

Υπολογίζοντας το μέσο μέγεθος των πακέτων ροής, παρέχονται πληροφορίες για τον τύπο της εφαρμογής ή της υπηρεσίας που χρησιμοποιείται στα πλαίσια ταξινόμησης της κίνησης, καθώς διαφορετικές εφαρμογές έχουν συχνά διαφορετικά πρότυπα μεγέθους πακέτων.

54. **Fwd Segment Size Avg**: Μέσο μέγεθος των forward segments.

Το εν λόγω χαρακτηριστικό προσμετρά το τυπικό μέγεθος των forward segments. Είναι χρήσιμο για την κατανόηση της συμπεριφοράς της πηγής σε όρους data segmentation, υποδεικνύοντας τη γενικότερη φύση της κίνησης, όπως μικρά πακέτα ελέγχου έναντι μεγάλων μεταφορών δεδομένων.

55. **Bwd Segment Size Avg**: Μέσο μέγεθος των backward segments.

    Ο υπολογισμός του μέσου μέγεθους των backward segments συμβάλλει στην ανάλυση του προτύπου απόκρισης του προορισμού, αποκαλύπτοντας πληροφορίες για τον τύπο της υπηρεσίας ή της εφαρμογής που χρησιμοποιείται και ειδικότερα για τον τρόπο χειρισμού αποκρίσεων.

56. **Fwd Bytes/Bulk Avg**: Μέσος όγκος των forward bulk bytes.

    *Τα **bulk data** αναφέρονται σε μεγάλα τμήματα δεδομένων που αποστέλλονται ταυτόχρονα, αντί για μικρότερα, αποκλειστικά πακέτα.*

    Το εν λόγω χαρακτηριστικό προσμετρά τον μέσο όγκο forward bulk δεδομένων, υποδεικνύοντας την παρουσία μεταφορών μεγάλων δεδομένων ή δραστηριοτήτων συνεχούς ροής.

57. **Fwd Packet/Bulk Avg**: Μέσος αριθμός των forward bulk packets.

    *Τα **bulk packets** αναφέρονται σε πολλαπλά πακέτα, αποστελλόμενα σε συνεχή ροή.*

    Υπολογίζοντας τον μέσο αριθμό forward bulk πακέτων, το εν λόγω χαρακτηριστικό επεξηγεί την packetization στρατηγική της πηγής και παρέχει πληροφορίες για την αντίστοιχη απόδοση αποστολής δεδομένων.

58. **Fwd Bulk Rate Avg**: Μέσος forward bulk ρυθμός.

    Αυτό το χαρακτηριστικό μετρά τη μέση ταχύτητα με την οποία αποστέλλονται μαζικά δεδομένα στην κατεύθυνση προώθησης. Είναι χρήσιμο για την ανάλυση των ταχυτήτων μεταφοράς δεδομένων και τον εντοπισμό πιθανών προβλημάτων απόδοσης ή συμφόρησης στο δίκτυο που επηρεάζουν τη μετάδοση δεδομένων.

59. **Bwd Bytes/Bulk Avg**: Μέσος όγκος των backward bulk bytes.

    Το εν λόγω χαρακτηριστικό προσμετρά τον μέσο όγκο backward bulk δεδομένων, παρέχοντας πληροφορίες για τη συμπεριφορά απόκρισης του προορισμού και συμβάλλοντας στον εντοπισμό προτύπων που σχετίζονται με συγκεκριμένες εφαρμογές ή υπηρεσίες.

60. **Bwd Packet/Bulk Avg**: Μέσος αριθμός των backward bulk packets.

    Μετρώντας τον μέσο αριθμό backward bulk πακέτων, αυτό το χαρακτηριστικό συνδράμει στην κατανόηση της απόδοσης απόκρισης, αποκαλύπτοντας πώς ο προορισμός χειρίζεται τη στρατηγική packetization και τη μετάδοση δεδομένων πίσω στην πηγή.

61. **Bwd Bulk Rate Avg**: Μέσος backward bulk ρυθμός.

    Αυτό το χαρακτηριστικό υπολογίζει τη μέση ταχύτητα αποστολής bulk δεδομένων από τον προορισμό πίσω στην πηγή και είναι ουσιαστικό για την ανάλυση των ταχυτήτων απόκρισης, αλλά και τον εντοπισμό προβλημάτων που σχετίζονται με την απόδοση του

δικτύου ή τη συμφόρηση στη backward διαδρομή.

Sub-flow Features

Τα sub-flow features παρέχουν μια συγκεκριμένη θεώρηση της δικτυακής κίνησης, διασπώντας τις ροές σε μικρότερα τμήματα ή υπο-ροές. Αυτά τα χαρακτηριστικά βοηθούν στην ανάλυση της συμπεριφοράς μεμονωμένων τμημάτων εντός μιας ευρύτερης ροής, παρέχοντας λεπτομερείς πληροφορίες για τα πρότυπα επικοινωνίας και τα γνωρίσματα μεταφοράς δεδομένων.

62. **Subflow Fwd Packets**: Αριθμός forward πακέτων στην υπο-ροή.

    Το εν λόγω χαρακτηριστικό προσμετρά τον αριθμό των forward πακέτων υπο-ροής. Η ανάλυση του αριθμού των forward πακέτων συμβάλλει στον προσδιορισμό της έντασης και της συχνότητας μετάδοσης δεδομένων, σε κάθε τμήμα της συνολικής ροής.

63. **Subflow Fwd Bytes**: Αριθμός forward bytes στην υπο-ροή.

    Το εν λόγω χαρακτηριστικό προσμετρά τον συνολικό όγκο forward δεδομένων υπο-ροής, σε bytes. Παρέχει πληροφορίες για την ποσότητα δεδομένων που μεταφέρονται σε κάθε τμήμα, στα πλαίσια κατανόησης του φόρτου και των προτύπων μεταφοράς δεδομένων της πηγής.

64. **Subflow Bwd Packets**: Αριθμός backward πακέτων στην υπο-ροή.

    Το εν λόγω χαρακτηριστικό προσμετρά τον αριθμό των backward πακέτων υπο-ροής. Με την ανάλυση του αριθμού αυτού, είναι δυνατή η κατανόηση της συμπεριφοράς απόκρισης και των προτύπων επικοινωνίας του προορισμού.

65. **Subflow Bwd Bytes**: Αριθμός backward bytes στην υπο-ροή.

    Το εν λόγω χαρακτηριστικό προσμετρά τον συνολικό όγκο backward δεδομένων υποροής, σε bytes. Συμβάλλει στην κατανόηση της ποσότητας των backward δεδομένων, παρέχοντας μια λεπτομερή εικόνα της ανταλλαγής δεδομένων σε κάθε υπο-ροή.

Initial TCP Handshake Features

Τα initial TCP handshake features καταγράφουν μετρικές σχετικές με την εγκαθίδρυση των συνδέσεων TCP. Τα εν λόγω χαρακτηριστικά παρέχουν πληροφορίες για την αποτελεσματικότητα, την ταχύτητα και την αξιοπιστία της αρχικής ανταλλαγής δεδομένων μεταξύ των επικοινωνούντων μερών, συμβάλλοντας στην κατανόηση της ρύθμισης και της απόδοσης των δικτυακών συνδέσεων.

66. **Init_Win_bytes_forward**: Συνολικά bytes forward αποστολής στο initial window.

    Το εν λόγω χαρακτηριστικό προσμετρά τον συνολικό όγκο, σε bytes, forward δεδομένων - απεσταλμένων κατά τη διάρκεια του initial TCP window. Το initial window είναι το πρώτο τμήμα δεδομένων που ανταλλάσσεται κατά την εγκαθίδρυση της σύνδεσης, και αυτή η μέτρηση βοηθά στην αξιολόγηση του όγκου δεδομένων που αποστέλλεται κατά την αρχικοποίηση της σύνδεσης.

67. **Init_Win_bytes_backward**: Συνολικά bytes backward αποστολής στο initial window.

Παρόμοια με την forward κατεύθυνση, το εν λόγω χαρακτηριστικό προσμετρά τον συνολικό όγκο σε bytes, backward δεδομένων κατά τη διάρκεια του initial TCP window. Παρέχει πληροφορίες για την αρχική απόκριση από τον προορισμό, συμβάλλοντας στην κατανόηση της αμφίδρομης ανταλλαγής δεδομένων κατά τη διάρκεια της εγκαθίδρυσης της σύνδεσης.

68. **act_data_pkt_fwd**: Αριθμός ενεργών πακέτων δεδομένων στη forward κατεύθυνση.

    Το εν λόγω χαρακτηριστικό προσμετρά τον αριθμό των forward πακέτων δεδομένων που μεταδίδονται **ενεργά**. *Ενεργά πακέτα δεδομένων είναι αυτά που περιέχουν πραγματικά φορτία δεδομένων, σε αντίθεση με τα πακέτα ελέγχου.* Αυτή η μετρική εξυπηρετεί την κατανόηση του όγκου των ουσιαστικών forward δεδομένων, απεσταλμένων κατά τα αρχικά στάδια της σύνδεσης.

69. **min_seg_size_forward**: Ελάχιστο μέγεθος τμήματος (segment) στη forward κατεύθυνση.

    Το εν λόγω χαρακτηριστικό καταγράφει το μικρότερο μέγεθος τμήματος σε bytes παρατηρηθέν στη forward κατεύθυνση κατά την initial TCP handshake. Η κατανόηση της ποσότητας αυτής, συμβάλλει στον προσδιορισμό της αποτελεσματικότητας της μετάδοσης δεδομένων κατά τη διάρκεια της αρχικής ρύθμισης της σύνδεσης.

Additonal Features - Πρόσθετα Χαρακτηριστικά

Τα πρόσθετα χαρακτηριστικά καταγράφουν διάφορες πτυχές της συμπεριφοράς της δικτυακής ροής και των χρονικών μεταβολών, εστιάζοντας στις ενεργές και αδρανείς καταστάσεις των ροών. Οι κάτωθι μετρικές περιγράφουν τη διάρκεια και τη μεταβλητότητα της δραστηριότητας ή της αδράνειας των δικτυακών επικοινωνιών, οι οποίες είναι καθοριστικές για την ανάλυση απόδοσης και την ανίχνευση ανωμαλιών.

Στο πλαίσιο των δικτύων, η έννοια idle - αδρανής αναφέρεται σε μια κατάσταση κατά την οποία μια σύνδεση ή συσκευή δικτύου είναι ενεργή αλλά δεν μεταδίδει ή λαμβάνει δεδομένα. Επιπροσθέτως:

(a) **Idle Connection** (Αδρανής Σύνδεση): Μια *αδρανής δικτυακή σύνδεση* περιγράφει μία σύνδεση καθιερωμένη και διαθέσιμη, στην οποία δεν αποστέλλονται ή λαμβάνονται πακέτα δεδομένων τη δεδομένη στιγμή.

(b) **Idle Time** (Χρόνος Αδράνειας): Αναφέρεται στην περίοδο κατά την οποία μια σύνδεση δικτύου παραμένει ανοιχτή αλλά αχρησιμοποίητη, και αποτελεί παράγοντα που επηρεάζει την απόδοση και την αποδοτικότητα του δικτύου (performance & efficiency), καθώς οι αδρανείς συνδέσεις μπορεί να εξακολουθούν να καταναλώνουν πόρους.

(c) **Network Devices** (Δικτυακές Συσκευές): Για δικτυακές συσκευές όπως δρομολογητές (routers), διακόπτες (switches) ή υπολογιστές (computers), η έννοια "αδρανής" περιγράφει μια κατάσταση κατα την οποία η συσκευή είναι ενεργή και συνδεδεμένη στο δίκτυο, αλλά δεν επεξεργάζεται ενεργά κίνηση δικτύου.

(d) **Idle Detection** (Ανίχνευση Αδράνειας): Τα πρωτόκολλα και τα συστήματα δικτύου συχνά διαθέτουν μηχανισμούς για την ανίχνευση αδρανών συνδέσεων, με απώτερο σκοπό τη βελτίωση της διαχείρισης των πόρων, όπως για παράδειγμα τον τερματισμό αδρανών συνδέσεων μετά από μια προκαθορισμένη περίοδο αδράνειας, για την απελευθέρωση εύρους ζώνης (bandwidth) και πόρων.

70. **Active Mean**: Μέσο χρονικό διάστημα κατά το οποίο μια ροή υπήρξε ενεργή, προτού μεταβεί σε αδράνεια.

    Το εν λόγω χαρακτηριστικό υπολογίζει τη μέση διάρκεια κατά την οποία μια ροή δικτύου παρέμεινε ενεργή (μεταδίδοντας δεδομένα), προτού μεταβεί σε αδρανή κατάσταση και συμβάλλει ευρύτερα στον προσδιορισμό της τυπικής διάρκειας δραστηριότητας των ροών, υποδεικνύοντας φυσιολογικά ή ανώμαλα πρότυπα συμπεριφοράς.

71. **Active Std**: Τυπική απόκλιση του χρονικού διαστήματος κατά το οποίο μια ροή υπήρξε ενεργή, προτού μεταβεί σε αδράνεια.

    Το εν λόγω χαρακτηριστικό προσμετρά τη μεταβλητότητα της διάρκειας των ενεργών καταστάσεων, προτού μεταβούν σε καταστάσεις αδράνειας. Μια υψηλή τυπική απόκλιση υποδεικνύει σημαντικές διακυμάνσεις κατά τη διάρκεια της δικτυακής δραστηριότητας, υποδηλώνοντας ασυνεπή συμπεριφορά δικτύου ή πιθανά προβλήματα στη μετάδοση δεδομένων.

72. **Active Max**: Μέγιστο χρονικό διάστημα κατά το οποίο μια ροή υπήρξε ενεργή, προτού μεταβεί σε αδράνεια.

    Το εν λόγω χαρακτηριστικό καταγράφει τη μεγαλύτερη διάρκεια κατά την οποία μια ροή παρέμεινε ενεργή, προτού μεταβεί σε αδράνεια και συμβάλλει στην αναγνώριση των κορυφαίων περιόδων δραστηριότητας δικτυακής ροής, κατατοπίζοντας την ανίχνευση περιόδων υψηλής μετάδοσης δεδομένων ή παρατεταμένων συνδέσεων.

73. **Active Min**: Ελάχιστο χρόνικό διάστημα κατά το οποίο μια ροή υπήρξε ενεργή, προτού μεταβεί σε αδράνεια.

Το εν λόγω χαρακτηριστικό καταγράφει την ελάχιστη διάρκεια κατά την οποία μια ροή υπήρξε ενεργή, προτού μεταβεί σε αδρανή κατάσταση. Περιγράφει τα minimum activity bursts δικτυακής δραστηριότητας, προσδιορίζοντας τα κατώτερα όρια του flow activity.

74. **Idle Mean**: Μέσο χρονικό διάστημα κατά το οποίο μια ροή υπήρξε αδρανής, προτού μεταβεί σε ενεργή κατάσταση.

    Το εν λόγω χαρακτηριστικό υπολογίζει τη μέση διάρκεια κατά την οποία μια ροή δικτύου παρέμεινε αδρανής, προτού μεταβεί σε ενεργή κατάσταση, υποδεικνύοντας τη γενική διάρκεια αδράνειας εντός των ροών.

75. **Idle Std**: Τυπική απόκλιση του χρονικού διαστήματος κατά το οποίο μια ροή υπήρξε αδρανής, προτού μεταβεί σε ενεργή κατάσταση.

    Το εν λόγω χαρακτηριστικό προσμετρά τη μεταβλητότητα της διάρκειας των αδρανών καταστάσεων, προτού μεταβούν σε ενεργή κατάσταση. Μια υψηλή τυπική απόκλιση υποδηλώνει σημαντικές διακυμάνσεις στις περιόδους αδράνειας, καταδεικνύουντας ακανόνιστα ή σποραδικά πρότυπα μετάδοσης δεδομένων.

76. **Idle Max**: Μέγιστο χρονικό διάστημα κατά το οποίο μια ροή υπήρξε αδρανής, προτού μεταβεί σε ενεργή κατάσταση.

    Το εν λόγω χαρακτηριστικό καταγράφει το μεγαλύτερο χρονικό διάστημα κατά το οποίο μια ροή παρέμεινε αδρανής προτού μεταβεί σε ενεργή κατάσταση, συμβάλλοντας στην αναγνώριση των κορυφαίων περιόδων αδράνειας της δικτυακής ροής, στα πλαίσια ανίχνευσης παρατεταμένης αδράνειας ή πιθανών διακοπών στην επικοινωνία.

77. **Idle Min**: Ελάχιστο χρονικό διάστημα κατά το οποίο μια ροή υπήρξε αδρανής, προτού μεταβεί σε ενεργή κατάσταση.

    Το εν λόγω χαρακτηριστικό καταγράφει το ελάχιστο χρονικό διάστημα κατά το οποίο μια ροή παρέμεινε αδρανής, προτού μεταβεί σε ενεργή κατάσταση, προσδιορίζοντας τα κατώτερα χρονικά όρια της αδράνειας των ροών.

Μεταδεδομένα

Τα μεταδεδομένα (metadata) παρέχουν πρόσθετες πληροφορίες για τη δικτυακή ροή, καταγράφοντας βασικές λεπτομέρειες σχετικά με τα end points και το timing της επικοινωνίας. Τα εν λόγω χαρακτηριστικά έχουν εξέχουσα σημασία για την παρακολούθηση, ανάλυση και διαχείριση της δικτυακής κίνησης, προσφέροντας πληροφορίες για τις πηγές (sources) και τους προορισμούς (destinations). Ωστόσο, συχνά αφαιρούνται πριν από την εκπαίδευση μοντέλων μηχανικής μάθησης για να αποφευχθεί το ενδεχόμενο του data leakage.

78. **Flow ID**: Ένας μοναδικός αναγνωριστικός αριθμός (identifier), που αποδίδεται σε κάθε δικτυακή ροή.

    Ο flow identifier διακρίνει μεταξύ διαφορετικών ροών, επιτρέποντας την παρακολούθηση και ανάλυση μεμονωμένων συνεδριών επικοινωνίας (communication sessions). Το εν λόγω χαρακτηριστικό είναι απαραίτητο για τη διαχείριση και οργάνωση των δεδομένων δικτύου, αλλά δεν συμβάλλει στην ανάλυση της συμπεριφοράς της ίδιας της ροής.

79. **Source IP**: Η διεύθυνση IP της συσκευής από την οποία ξεκίνησε τη ροή.

    Η IP της πηγής αναγνωρίζει την προέλευση της κίνησης δικτύου, και κατ΄ επέκταση το σημείο προέλευσης της ροής. Αυτό το χαρακτηριστικό εξακριβώνει την πηγή των πακέτων δεδομένων και συμβάλλει στην αναγνώριση πιθανών κακόβουλων δραστών (agents), αλλά και στην παρακολούθηση της ευρύτερης δικτυακής δραστηριότητας.

80. **Source Port**: Ο αριθμός port[6] που χρησιμοποιεί η συσκευή πηγής για την αποστολή των πακέτων.

    Η Source Port, σε συνδυασμό με την IP πηγής, παρέχει έναν αναγνωριστικό αριθμό για το αρχικό άκρο της επικοινωνίας. Αυτό το χαρακτηριστικό χρησιμοποιείται για τον προσδιορισμό της συγκεκριμένης εφαρμογής ή υπηρεσίας στη συσκευή πηγής όπου ξεκινά η ροή.

81. **Destination IP**: Η διεύθυνση IP της συσκευής προορισμού (destination).

    Η IP προορισμού αναγνωρίζει τον στόχο της κίνησης δικτύου και είναι απαραίτητη για την κατανόηση του σημείου αποστολής των πακέτων δεδομένων. Το εν λόγω χαρακτηριστικό εξακριβώνει τον προορισμό των πακέτων δεδομένων και συμβάλλει στην αναγνώριση πιθανών στόχων κακόβουλων επιθέσεων και στην παρακολούθηση της ευρύτερης δικτυακής δραστηριότητας.

82. **Destination Port**: Ο αριθμός port στο τελικό άκρο της επικοινωνίας.

    Η destination port, ένα καίριο αναγνωριστικό του προορισμού, προσδιορίζει τη συγκεκριμένη υπηρεσία ή εφαρμογή στην οποία στοχεύει η κίνηση δικτύου. Κάθε port number αντιστοιχεί σε διαφορετική υπηρεσία ή πρωτόκολλο εφαρμογής (π.χ. το πρωτόκολλο HTTP χρησιμοποιεί την port 80, ενώ το HTTPS χρησιμοποιεί την port 443). Η ανάλυση των destination ports βοηθά στον εντοπισμό του τύπου της υπηρεσίας που χρησιμοποιείται, στην ανίχνευση μη εξουσιοδοτημένων προσπαθειών πρόσβασης και στην κατανόηση της ευρύτερης συμπεριφοράς της κίνησης στο επίπεδο της εφαρμογής.

83. **Timestamp**: Οι χρονικές στιγμές κατά τις οποίες καταγράφηκε η ροή.

    Η χρονική σήμανση παρέχει το χρονικό πλαίσιο της δικτυακής ροής, υποδεικνύοντας την εξέλιξη της επικοινωνίας. Το εν λόγω χαρακτηριστικό είναι απαραίτητο για χρονικές αναλύσεις δικτυακής δραστηριότητας, για την αναγνώριση χρονικών προτύπων και για τη συσχέτιση γεγονότων μεταξύ διαφορετικών ροών.

4.3.3 Η Επίδραση των Μεταδεδομένων

Στο πλαίσιο της ανίχνευσης εισβολών, η παρουσία χαρακτηριστικών μεταδεδομένων (metadata) στα μοντέλα μηχανικής ή βαθιάς μάθησης μπορεί να οδηγήσει σε ένα φαινόμενο γνωστό ως shortcut learning [47] ή data leakage. Το shortcut learning περιγράφει ένα μοντέλο το οποίο μαθαίνει να προβλέπει με βάση άσχετα ή εσφαλμένα patterns στα δεδομένα εκπαίδευσης, αντί για τις υποκείμενες σχέσεις που αντιπροσωπεύουν πραγματικά την μεταβλητή στόχου. Η έννοια data leakage (διαρροή δεδομένων) αναφέρεται ειδικά στην περίπτωση όπου πληροφορίες εκτός του συνόλου δεδομένων εκπαίδευσης χρησιμοποιούνται για την ανάπτυξη του μοντέλου,

[6] Ο όρος port μεταφράζεται ως θύρα.

οδηγώντας σε υπερβολικά αισιόδοξες εκτιμήσεις απόδοσης κατά την εκπαίδευση, αλλά κακή γενίκευση σε νέα, αόρατα δεδομένα.

Η μελέτη του Laurens D'hooge et al., [47], εξετάζει αυτό το ζήτημα σε διάφορα σύνολα δεδομένων, συμπεριλαμβανομένου του CIC-IDS-2017. Οι ερευνητές αναγνωρίζουν αρκετά χαρακτηριστικά αυτού του συνόλου δεδομένων ως μεταδεδομένα, ήτοι:

Flow ID, Source IP, Source Port, Destination IP, Destination Port & Timestamp

Τα εν λόγω χαρακτηριστικά, ενώ είναι χρήσιμα για την καταγραφή (logging) και την ταυτοποίηση της ροής, δεν συμβάλλουν στην ανίχνευση κακόβουλων δραστηριοτήτων με βάση τη συμπεριφορά της δικτυακής κίνησης. Αντ΄ αυτού, μπορούν a posteriori να αναδείξουν ισχυρές συσχετίσεις για τη διάκριση μεταξύ διαφορετικών κατηγοριών, οδηγώντας σε data leakage.

Η συμπερίληψη αυτών των χαρακτηριστικών μεταδεδομένων επηρεάζει σημαντικά τα αποτελέσματα των μοντέλων μάθησης. Για παράδειγμα, οι διευθύνσεις IP και οι port numbers μπορεί να έχουν υψηλή συσχέτιση με τύπους κίνησης ή δραστηριότητες στα πλαίσια ενός συγκεκριμένου συνόλου δεδομένων, όμως δεν καθίσταται δυνατή η αποτελεσματική γενίκευση σε διαφορετικά σύνολα δεδομένων ή πραγματικά σενάρια όπου οι παρατηρηθείσες συσχετίσεις ενδέχεται να μην ισχύουν. Ομοίως, οι χρονικές σημάνσεις εισάγουν χρονικές προκαταλήψεις, εξαναγκάζοντας το μοντέλο να βασίζεται στον χρόνο συλλογής των δεδομένων αντί στα πραγματικά χαρακτηριστικά της δικτυακής κίνησης.

Προκειμένου να προληφθούν οι κίνδυνοι διαρροής δεδομένων, η μελέτη συνιστά την εξάλειψη των μεταδεδομένων από το dataset για κάθε περίπτωση μοντέλων μάθησης. Αφαιρώντας τα χαρακτηριστικά: Flow ID, Source IP, Source Port, Destination IP, Destination Port και Timestamp, τα μοντέλα εν τέλει εκπαιδεύονται από το περιεχόμενο και τη συμπεριφορά των πακέτων, οδηγώντας σε πιο robust και γενικεύσιμα συστήματα ανίχνευσης εισβολών.

Κεφάλαιο 5

Μεθοδολογίες Προεπεξεργασίας Δεδομένων

Σε αυτό το κεφάλαιο, παρουσιάζονται διάφορες μεθοδολογίες προεπεξεργασίας που θα εφαρμοστούν στους αγωγούς των Συστημάτων Ανίχνευσης Εισβολών. Εν γένει, η προεπεξεργασία δεδομένων αποτελεί απαραίτητο βήμα στα πλαίσια της Μηχανικής και της Βαθιάς Μάθησης και συντελείται με απώτερο στόχο τον μετασχηματισμό των δεδομένων σε μορφή συμβατή και κατάλληλη για είσοδο στους υπό μελέτη ταξινομητές.

5.1 Feature Selection με XGBoost

Αυτό το βήμα προεπεξεργασίας εφαρμόζεται αποκλειστικά στους αγωγούς των MLPs.

Σε αυτήν την ενότητα, εξετάζεται η επιλογή χαρακτηριστικών (feature selection) υπό τη χρήση του αλγόριθμου XGBoost. Ο XGBoost, ένας tree-based ensemble learning αλγόριθμος, εκτελεί εγγενώς την επιλογή χαρακτηριστικών κατά τη διαδικασία εκπαίδευσής του. Αυτή η embedded μέθοδος επιλογής χαρακτηριστικών εντοπίζει τα πλέον κατατοπιστικά χαρακτηριστικά, μειώνοντας έτσι τη διαστατικότητα του χώρου εισόδου ενώ ταυτόχρονα διατηρεί ή βελτιώνει την προβλεπτική απόδοση του μοντέλου.

5.1.1 Ο Αλγόριθμος XGBoost

Ο XGBoost (eXtreme Gradient Boosting) αλγόριθμος, που παρουσιάστηκε από τους Chen et al. [48], αποτελεί μια υλοποίηση του gradient boosting. Αυτή η machine learning ensemble μέθοδος, λόγω της ισχυρής απόδοσης και της υψηλής ακρίβειας της, χρησιμοποιείται ευρέως σε διαγωνισμούς και εφαρμογές. Ο XGBoost βελτιώνει τις παραδοσιακές gradient boosting μεθοδολογίες ενσωματώνοντας αλγοριθμικές βελτιστοποιήσεις.

> *Ο αλγόριθμος XGBoost είναι μέρος της ευρύτερης κατηγορίας machine learning ensemble methods.*

Επισκόπηση: Ensemble Learning Οι ensemble methods είναι τεχνικές μηχανικής μάθησης που συνδυάζουν πολλαπλά μοντέλα για να βελτιώσουν τη συνολική απόδοση, [49]. Η κεντρική ιδέα πίσω από τις ensemble μεθόδους είναι ότι συνδυάζοντας τις προβλέψεις πολλών μοντέλων, μειώνεται ο κίνδυνος υπερεκπαίδευσης και ενισχύεται η ακρίβεια. Αυτοί είναι οι κύριοι τύποι ensemble μεθόδων:

- **Bagging (Bootstrap Aggregating)**: Αυτή η μέθοδος περιλαμβάνει την ανεξάρτητη εκπαίδευση πολλών μοντέλων σε διαφορετικά υποσύνολα των δεδομένων και στη συνέχεια

την εξαγωγή του μέσου όρου των προβλέψεών τους. Η μεθοδολογία των Random Forests είναι ένα κοινό παράδειγμα ensemble μεθόδου τύπου bagging.

- **Stacking**: Ένα μοντέλο Stacking είναι ένα meta-model που αξιοποιεί τις εξόδους από μια συλλογή πολλών συνήθως σημαντικά διαφορετικών μοντέλων, ως χαρακτηριστικά εισόδου, προκειμένου να μειώσει την υπερεκπαίδευση και να βελτιώσει την ακρίβεια.

- **Boosting**: Αυτή η μέθοδος περιλαμβάνει την διαδοχική εκπαίδευση μοντέλων, όπου κάθε νέο μοντέλο προσπαθεί να διορθώσει τα σφάλματα των προηγούμενων. Το Boosting στοχεύει στη μείωση του bias και της variance στο τελικό μοντέλο. Το XGBoost είναι ένα παράδειγμα ensemble μεθόδου τύπου boosting.

Παρουσιάζεται μια σύντομη επισκόπηση των μεθόδων Gradient Boosting, ως υποκατηγορία των ensemble μεθόδων τύπου Boosting.

Η μέθοδος Gradient Boosting Η μέθοδος Gradient Boosting είναι μια ensemble τεχνική που κατασκευάζει μοντέλα διαδοχικά, [50]. Κάθε νέο μοντέλο εκπαιδεύεται για να διορθώσει τα σφάλματα των προηγούμενων μοντέλων. Οι κύριες συνιστώσες του gradient boosting περιλαμβάνουν:

1. **Loss Function**: Μια διαφορίσιμη συνάρτηση απώλειας προς ελαχιστοποίηση.

2. **Weak Learner**: Ένα απλό μοντέλο, όπως ένα decision tree, που αποδίδει ελαφρώς καλύτερα από random guessing.

3. **Additive Model**: Τα μοντέλα προστίθενται διαδοχικά προς ελαχιστοποίηση της συνάρτησης απώλειας.

Η θεμελιώδης έννοια του gradient boosting περιλαμβάνει τη χρήση της κλίσης (gradient) της συνάρτησης απώλειας για την προοδευτική βελτίωση των προβλέψεων του μοντέλου. Κάθε επανάληψη, ή γύρος boosting, εστιάζει στην αντιμετώπιση των residuals (υπολειπόμενων σφαλμάτων) των προηγούμενων μοντέλων.

Επισκόπηση του Αλγόριθμου XGBoost Ο XGBoost είναι ένας ισχυρός αλγόριθμος μηχανικής μάθησης που κατασκευάζει μοντέλα σε στάδια. Πρωταρχικά, ξεκινάει με μια αρχική σταθερή τιμή και προσθέτει weak learners προκειμένου να διορθώσει τα σφάλματα πρόβλεψης, ενσωματώνοντας τεχνικές κανονικοποίησης και βελτιστοποίησης για την παραγωγή ιδιαίτερα robust μοντέλων υψηλής ακρίβειας.

Περιλαμβάνει διαδικασίες κανονικοποίησης για την αποτροπή overfitting (υπερεκπαίδευσης) και τη βελτίωση της γενίκευσης του μοντέλου, και χρησιμοποιεί weighted quantile sketching ώστε να διαχειρίζεται αποδοτικά τα βάρη δεδομένων κατά την εύρεση σημείων διαχωρισμού στα δένδρα απόφασης. Με την ευαισθησία του στην αραιότητα (sparsity awareness), μεταχειρίζεται αραιά δεδομένα μαθαίνοντας αυτόματα την αντιμετώπιση ελλείπουσων τιμών (missing values). Επιπλέον, ο XGBoost περιλαμβάνει ενσωματωμένη cross-validation για τη ρύθμιση και την επιλογή του υπό κατασκευή μοντέλου, αξιοποιεί την παράλληλη επεξεργασία για ταχύτερη κατασκευή δέντρων και χρησιμοποιεί μια παράμετρο "μέγιστου βάθους" - max depth για την περικοπή δέντρων και τη μείωση της πολυπλοκότητας, αποτρέποντας έτσι την υπερεκπαίδευση. Επιπλέον, υποστηρίζει παράλληλη και κατανεμημένη υπολογιστική επεξεργασία, επιτρέποντας ταχύτερη εκπαίδευση σε μεγάλα σύνολα δεδομένων. Συνδυάζοντας αυτά τα χαρακτηριστικά, το μοντέλο XGBoost παρέχει μια robust και scalable λύση για ένα ευρύ φάσμα προβλημάτων πρόβλεψης, καθιστώντας το μια δημοφιλή επιλογή τόσο για καθήκοντα παλινδρόμησης, όσο και ταξινόμησης.

Αλγόριθμος 5.1 Ο αλγόριθμος XGBoost

1. **Initialization** (Αρχικοποίηση): Ο XGBoost ξεκινά με ένα αρχικό μοντέλο ως baseline prediction, το οποίο είναι μια σταθερή τιμή. Συνήθως επιλέγεται ο μέσος όρος της μεταβλητής στόχου (target variable) για εργασίες παλινδρόμησης ή τα log-odds για δυαδική ταξινόμηση.

2. **Iterative Additions**: Το XGBoost κατασκευάζει το μοντέλο σε στάδια προσθέτοντας έναν weak learner κάθε φορά (συνήθως ένα desicion tree). Κάθε weak learner εκπαιδεύεται για να διορθώσει τα σφάλματα του υπάρχοντος μοντέλου.

   Για κάθε boosting round:

   (a) **Calculate Residuals**: Υπολογίζονται τα residuals, που είναι οι διαφορές μεταξύ των παρατηρούμενων τιμών (y_i) και των προβλεπόμενων τιμών ($\hat{y}_i$) από το τρέχον μοντέλο.

   $$\text{Residual} = y_i - \hat{y}_i$$

   Τα residuals αντιπροσωπεύουν τα σφάλματα που πραγματοποιεί το τρέχον μοντέλο σε κάθε boosting round.

   (b) **Fit Weak Learner**: Εκπαίδευση ενός νέου weak learner, π.χ. desicion tree - δένδρο απόφασης, για την πρόβλεψη των residuals. Αυτό το δένδρο επικεντρώνεται στην καταγραφή υποκείμενων μοτίβων στα σφάλματα του τρέχοντος μοντέλου.

   (c) **Update the model**: Ενημέρωση του μοντέλου με τις προβλέψεις του weak learner, scaled από ένα ρυθμό μάθησης.

   Οι προβλέψεις του νέου δένδρου πολλαπλασιάζονται από ένα ρυθμό μάθησης η (μία υπερπαράμετρος που ελέγχει τη συμβολή κάθε δένδρου) προτού προστεθούν στο τρέχον μοντέλο.

   $$\hat{y_i}^{(\text{new})} = \hat{y}_i + \eta \cdot h_i(x)$$

   όπου $h_i(x)$ είναι η πρόβλεψη από το νέο δένδρο.

   (d) **Apply regularization**: Εφαρμογή κανονικοποίησης για την αποτροπή υπερεκπαίδευσης.

   Η κανονικοποίηση εφαρμόζεται για τον έλεγχο της πολυπλοκότητας του μοντέλου και την αποτροπή της υπερεκπαίδευσης. Το XGBoost χρησιμοποιεί τόσο την L_1 (Lasso) όσο και την L_2 (Ridge) κανονικοποίηση στα δενδρικά βάρη.

3. **Final Model**: Το τελικό μοντέλο είναι ένα ensemble (σύνολο) όλων των weak learners (δένδρων απόφασης) που προστέθηκαν σε κάθε γύρο boosting. Η συνολική πρόβλεψη θα είναι το άθροισμα της αρχικής πρόβλεψης και των συνεισφορών από όλα τα δένδρα.

   $$\hat{y_i}^{(\text{final})} = \hat{y_i}^{(\text{initial})} + \sum_{m=1}^{M} \eta \cdot h_i^{(m)}(x)$$

   όπου M είναι ο συνολικός αριθμός των boosting rounds.

Output: Οι τελικές προβλέψεις γίνονται χρησιμοποιώντας το ensemble μοντέλο, το οποίο συγκεντρώνει τις εξόδους από το σύνολο των weak learners. Αυτή η ensemble προσέγγιση επιτρέπει στον XGBoost να κατασκευάζει ιδιαίτερα accurate και robust μοντέλα.

Ο XGBoost έχει εμπνεύσει αρκετές δημοφιλείς παραλλαγές που ενισχύουν την τεχνική gradient boosting για συγκεκριμένες ανάγκες. Σημαντικές μεταξύ αυτών είναι οι CatBoost, [51] και LightGBM, [52]. Το μοντέλο CatBoost, που αναπτύχθηκε από την εταιρεία Yandex, είναι σχεδιασμένο για να χειρίζεται πιο αποτελεσματικά κατηγορικά δεδομένα και να μειώνει την υπερεκπαίδευση εφαρμόζοντας τεχνικές όπως το ordered boosting. Ο αλγόριθμος LightGBM, που δημιουργήθηκε από τη Microsoft, είναι βελτιστοποιημένος για ταχύτητα και αποδοτικότητα με μεγάλα σύνολα δεδομένων μέσω τεχνικών όπως histogram-based decision tree learning και leaf-wise tree growth. Αυτές οι παραλλαγές προσφέρουν μοναδικά πλεονεκτήματα και χρησιμοποιούνται ευρέως σε εργασίες μηχανικής μάθησης παράλληλα με τον XGBoost.

5.1.2 Επιλογή Χαρακτηριστικών με τον αλγόριθμο XGBoost

Η επιλογή χαρακτηριστικών αποτελεί ένα κρίσιμο βήμα για την κατασκευή αποτελεσματικών μοντέλων Μηχανικής και Βαθιάς Μάθησης. Βοηθά στον εντοπισμό των πλέον κατατοπιστικών χαρακτηριστικών στα πλαίσια της πρόβλεψης, βελτιώνοντας έτσι την απόδοση του μοντέλου και μειώνοντας τη διαστατικότητα του χώρου εισόδου.

Αλγόριθμος 5.2 Επιλογή Χαρακτηριστικών με τον αλγόριθμο XGBoost

1. **Train the Model**: Εκπαίδευση ενός μοντέλου XGBoost στο σύνολο δεδομένων. Το μοντέλο θα χρησιμοποιήσει όλα τα διαθέσιμα χαρακτηριστικά για να μάθει τις σχέσεις και τα μοτίβα στα δεδομένα.

2. **Calculate Feature Importance** (Υπολογισμός Σημασίας Χαρακτηριστικών): Ο αλγόριθμος XGBoost υπολογίζει τις βαθμολογίες σημασίας χαρακτηριστικών, οι οποίες αντικατοπτρίζουν τη στατιστική συνεισφορά κάθε χαρακτηριστικού στο μοντέλο.

3. **Rank Features** (Κατάταξη Χαρακτηριστικών): Με βάση τα ήδη υπολογισμένα attention scores, πραγματοποιείται κατάταξη των χαρακτηριστικών σε φθίνουσα σειρά. Αυτή η κατάταξη εντοπίζει τα χαρακτηριστικά μεγαλύτερης προγνωστικής επίδρασης.

4. **Select Top Features** (Επιλογή Κορυφαίων Χαρακτηριστικών): Επιλέγεται ένα υποσύνολο των χαρακτηριστικών κορυφαίας σημασίας. Ο αριθμός των χαρακτηριστικών που θα επιλεγούν μπορεί να καθοριστεί με βάση domain knowledge, την απόδοση του μοντέλου ή ένα threshold (κατώφλι) στις βαθμολογίες σημασίας.

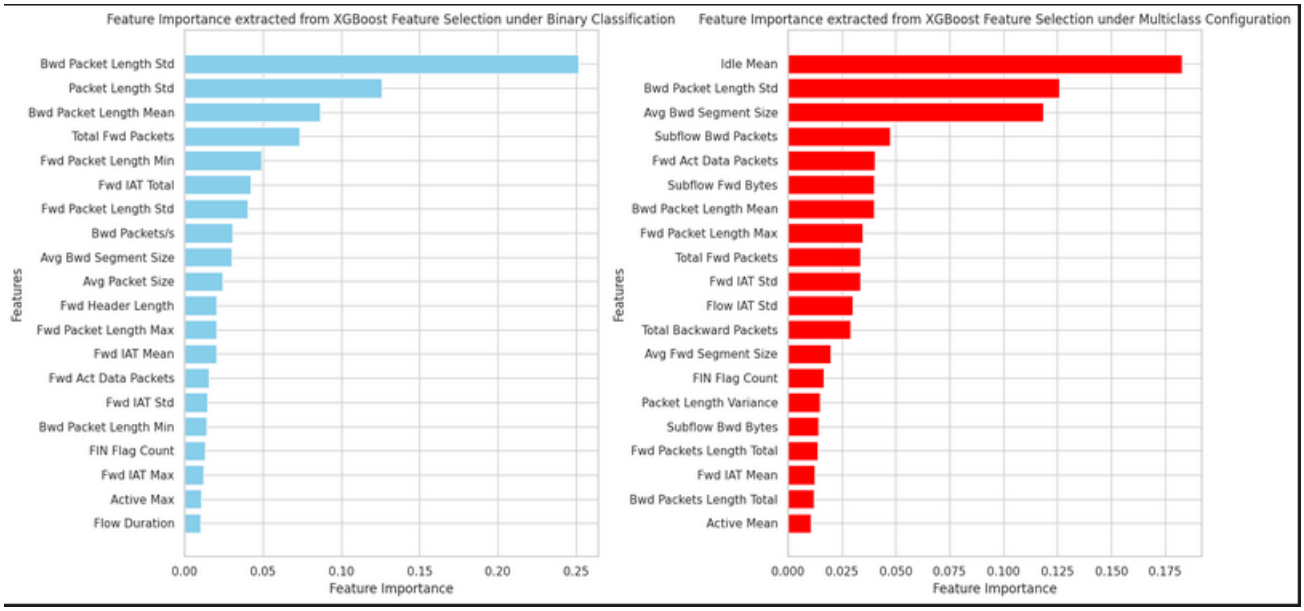
5. **Retrain the Model**: Επανεκπαίδευση του μοντέλου με χρήση μόνο των επιλεγμένων χαρακτηριστικών. Αυτό το βήμα επαληθεύει ότι τα επιλεγμένα χαρακτηριστικά είναι όντως επαρκή για να διατηρήσουν ή ακόμα και να βελτιώσουν την απόδοση του μοντέλου, ενώ ταυτόχρονα μειώνεται η πολυπλοκότητα.

6. **Evaluate Model Performance**: Αξιολόγηση απόδοσης του μοντέλου και σύγκριση της απόδοσης υπό την εκπαίδευση με τα επιλεγμένα χαρακτηριστικά ως προς το αρχικό μοντέλο. Μετρικές απόδοσης ταξινόμησης όπως accuracy, precision, recall, και F_1-score θα πρέπει να χρησιμοποιηθούν για να αξιολογηθεί αν η επιλογή χαρακτηριστικών έχει οδηγήσει σε βελτίωση ή διατήρηση ενός ικανοποιητικού επιπέδου απόδοσης.

Ένα από τα κύρια πλεονεκτήματα του XGBoost είναι η δυνατότητά του να εκτελεί αυτόματα την επιλογή χαρακτηριστικών κατά τη διάρκεια της διαδικασίας εκπαίδευσης. Ο XGBoost

προσφέρει εγγενείς μεθόδους επιλογής χαρακτηριστικών λόγω της αρχιτεκτονικής του, η οποία βασίζεται σε δέντρα απόφασης και παρέχει importance scores (βαθμολογίες σημασίας), οι οποίες υποδεικνύουν τη συνεισφορά κάθε χαρακτηριστικού στην προγνωστική απόδοση του μοντέλου. Οι εν λόγω βαθμολογίες μπορούν να εξαχθούν[1] με διάφορους τρόπους, καθιστώντας τον υπολογισμό των feature importances ένα ζωτικό μέρος του αλγορίθμου. Δεδομένης της εκτεταμένης μαθηματικής οντολογίας που αναπτύσσεται, για περαιτέρω λεπτομέρειες ο αναγνώστης παραπέμπεται στις πηγές: [48, 53].

Σχήμα 5.1: Αποτελέσματα της Επιλογής Χαρακτηριστικών με τον Αλγόριθμο XGBoost



Η βαθμολογία των χαρακτηριστικών παρουσιάζει τιμές φραγμένες στο διάστημα $[0, 1]$, οι οποίες αθροίζονται στη μονάδα (μορφολογία συνάρτησης μάζας πιθανότητας). Από τα αρχικά 77 χαρακτηριστικά, μόνο 20 δείχνουν στατιστική σημασία που υπερβαίνει το threshold value 1%.

Υλοποίηση Η επιλογή χαρακτηριστικών με χρήση του αλγόριθμου XGBoost έχει υλοποιηθεί ως βήμα προεπεξεργασίας τόσο σε δυαδικές όσο και σε πολυκατηγορικές διατάξεις. Και στις δύο περιπτώσεις, τέθηκε ως selection threshold (κατώφλι επιλογής) η τιμή 1%. Παρόλο που ο αριθμός των χαρακτηριστικών που υπερβαίνουν αυτό το κατώφλι είναι ο ίδιος και για τις δύο διατάξεις, τα επιλεγμένα χαρακτηριστικά δεν είναι ταυτόσημα. Ακολουθούν μοναδικά χαρακτηριστικά, τα οποία εντοπίζονται είτε στην δυαδική είτε στην πολυκατηγορική διάταξη:

> **Δυαδική Διάταξη**: "*Packet Length Std*", "*Avg Packet Size*", "*Bwd Packets/s*", "*Fwd Header Length*", "*Flow Duration*", "*Fwd IAT Total*", "*Bwd Packet Length Min*", "*Fwd Packet Length Min*", "*Fwd Packet Length Std*", "*Fwd IAT Max*", "*Active Max*"

[1] Αυτά τα importance scores μπορούν να εξαχθούν ως προς την βελτιστοποίηση των κάτωθι εννοιών:

- **Gain**: Η μέση μείωση της απώλειας που επιτυγχάνεται όταν χρησιμοποιείται ένα χαρακτηριστικό για διαχωρισμό.
- **Frequency**: Το πλήθος των χρήσεων ενός χαρακτηριστικού για τον διαχωρισμό των δεδομένων σε όλα τα δέντρα.
- **Cover**: Το πλήθος των χρήσεων ενός χαρακτηριστικού για τον διαχωρισμό των δεδομένων σε όλα τα δέντρα, σταθμισμένο από τα σημεία των δεδομένων εκπαίδευσης.

Πολυκατηγορική Διάταξη: "*Bwd Packets Length Total*", "*Idle Mean*", "*Total Backward Packets*", "*Packet Length Variance*", "*Fwd Packets Length Total*", "*Flow IAT Std*", "*Subflow Bwd Bytes*", "*Subflow Fwd Bytes*", "*Subflow Bwd Packets*", "*Avg Fwd Segment Size*", "*Active Mean*"

5.2 Feature Engineering

Αυτή η διαδικασία προεπεξεργασίας εφαρμόζεται αποκλειστικά στην δυαδική διάταξη των MLPs.

5.2.1 Feature Engineering υπό Δυαδική Διάταξη

Η διαδικασία feature engineering είναι ένα κρίσιμο βήμα στην ευρύτερη ροή προεπεξεργασίας δεδομένων, το οποίο στοχεύει στη βελτίωση της προγνωστικής ικανότητας των μοντέλων Μηχανικής και Βαθιάς μάθησης. Περιλαμβάνει τη μετατροπή των ακατέργαστων (raw) δεδομένων σε σημαντικά και κατατοπιστικά χαρακτηριστικά που μπορούν να ενισχύσουν την ικανότητα του μοντέλου να μαθαίνει μοτίβα και να διεξάγει ακριβείς προβλέψεις. Το αποτελεσματικό feature engineering μπορεί να επηρεάσει σημαντικά την απόδοση ενός μοντέλου, μειώνοντας τον θόρυβο, αναδεικνύοντας σημαντικές σχέσεις και εξασφαλίζοντας ότι τα δεδομένα παρουσιάζονται στην πλέον κατάλληλη μορφή για τον αλγόριθμο μάθησης.

> *Η επιλογή χαρακτηριστικών προηγείται της διαδικασίας feature engineering και περιορίζει τα αρχικά 77 χαρακτηριστικά σε αυτά με στατιστική σημασία που υπερβαίνει το κατώφλι του 1%. Στη συνέχεια, πραγματοποιείται feature engineering σε αυτά τα επιλεγμένα χαρακτηριστικά. Επακολούθως, πραγματοποιείται μία επαναληπτική feature selection, προκειμένου να αξιολογηθεί η αποτελεσματικότητα των επαγόμενων χαρακτηριστικών. Η σημασία τους θα αξιολογηθεί με χρήση του ίδιου κριτηρίου (attention score μεγαλύτερο του 1%) υπό τον αλγόριθμο ταξινόμησης XGBoost. Αυτό το βήμα διασφαλίζει ότι διατηρούνται μόνο τα απαραίτητα χαρακτηριστικά για την τελική εκπαίδευση του μοντέλου.*

Αυτή η διαδικασία feature engineering προορίζεται για τη δυαδική περίπτωση, καθώς δεν επιδεικνύει βελτίωση στην απόδοση ταξινόμησης υπό την πολυκατηγορική διάταξη.

5.2.2 Δημιουργία Νέων Χαρακτηριστικών

Αυτή η ενότητα περιγράφει τα νέα χαρακτηριστικά που δημιουργήθηκαν για να βελτιώσουν την αποτελεσματικότητα του μοντέλου στην ανίχνευση και ταξινόμηση δικτυακών επιθέσεων. Αυτά τα χαρακτηριστικά χωρίζονται σε τρεις κατηγορίες με βάση τον τρόπο δημιουργίας τους: Weighted Feature Score, Feature Differences και Interaction Feature.

5.2.2.1 Weighted Feature Score

Το Weighted Feature Score αποδίδει βάρη στα χαρακτηριστικά με βάση τα importance scores τους, πολλαπλασιάζοντας τα με τις κανονικοποιημένες τιμές αυτών των χαρακτηριστικών. Με άλλα λόγια, κάθε χαρακτηριστικό, αφότου κανονικοποηθεί, θα πολλαπλασιαστεί με τη βαθμολογία σημασίας του, προκύπτουσα από τη γνωστή διαδικασία επιλογής χαρακτηριστικών XGBoost. Αυτή η προσέγγιση στοχεύει στη σύλληψη της συνολικής σημασίας πολλαπλών χαρακτηριστικών σε μια μετρική, αναδεικνύοντας κρίσιμα μοτίβα που υποδηλώνουν επιθέσεις.

$$(\text{Combined Importance Score})_i = X_i \odot \text{importance} \tag{5.1}$$

Εδώ, ο τελεστής $\odot$ υποδηλώνει πολλαπλασιασμό κατά στοιχείο (element-wise) των κανονικοποιημένων χαρακτηριστικών X_i με τα importance scores. Η κανονικοποίηση εξασφαλίζει ότι κάθε διάνυσμα χαρακτηριστικών κλιμακώνεται σε μήκος μονάδας. Το μοναδιαίο διάνυσμα $\overrightarrow{e}_{\vec{v}}$ στη διεύθυνση του διανύσματος $\overrightarrow{v} = (v_1, v_2, \ldots, v_n)$, ήτοι η κανονικοποίηση του $\overrightarrow{v}$ ορίζεται ως:

$$\overrightarrow{e}_{\vec{v}} = \frac{\overrightarrow{v}}{\|\overrightarrow{v}\|_2} = \frac{(v_1, v_2, \ldots, v_n)}{\sqrt{v_1^2 + v_2^2 + \ldots + v_n^2}}$$

Επομένως, κάθε γραμμή (σημείο στον χώρο χαρακτηριστικών) θα πρέπει να κανονικοποιηθεί. Το διάνυσμα των importances έχει προφανώς την ίδια διάσταση με το κανονικοποιημένο δείγμα X_i.

5.2.2.2 Feature Differences

Υπολογίζεται η διαφορά μεταξύ χαρακτηριστικών όπως *"Bwd Packet Length Std"* και *"Packet Length Std"*, ή *"Total Fwd Packets"* με *"Total Backward Packets"*.

$$\text{Packet Length Std Diff} = \text{Bwd Packet Length Std} - \text{Packet Length Std} \tag{5.2}$$

και

$$\text{Total Packet Diff} = \text{Total Fwd Packets} - \text{Total Backward Packets} \tag{5.3}$$

Ακολουθεί μια διαισθητική προσέγγιση των εν λόγω διαφορών:

> *Σημαντικές διαφορές στα μήκη ή στους αριθμούς πακέτων μεταξύ των forward και backward κατευθύνσεων δυνητικά σηματοδοτούν ανώμαλη δραστηριότητα, στο πλαίσιο της ανίχνευσης εισβολών.*

5.2.2.3 Interaction Feature / Γινόμενο Χαρακτηριστικών

Το Interaction Feature παράγεται με τον πολλαπλασιασμό επιμέρους χαρακτηριστικών, προκειμένου να συλληφθούν πολύπλοκες σχέσεις, μη εντοπιζόμενες από μεμονωμένα χαρακτηριστικά. Για παράδειγμα, το προϊόν των *"Bwd Packets/s"* και *"Flow Duration"* αποκαλύπτει την αλληλεπίδραση μεταξύ του ρυθμού των backward πακέτων ανά δευτερόλεπτο και της διάρκειας ροής, παρέχοντας πληροφορίες για τη δυναμική της κυκλοφορίας δικτύου.

$$\text{Bwd Fwd Product} = \text{Bwd Packets/s} \odot \text{Flow Duration}$$

Εδώ, ο τελεστής $\odot$ υποδηλώνει πολλαπλασιασμό κατά στοιχείο (element-wise). Αυτός ο όρος αλληλεπίδρασης σχεδιάσθηκε για τον εντοπισμό ασυνήθιστων μοτίβων, όπως υψηλό ρυθμό backward πακέτων σε συνδυασμό με υψηλές διάρκειες ροής, που δυνητικά υποδηλώνει ανώμαλη κίνηση.

> *Η διαδικασία feature engineering ολοκληρώνεται με τη δημιουργία 4 νέων χαρακτηριστικών, ενώ περαιτέρω χαρακτηριστικά θα μπορούσαν να δημιουργηθούν με εναλλακτικές μεθοδολογίες. Στη συνέχεια, απαιτείται εκτίμηση της αποτελεσματικότητας των νέων χαρακτηριστικών, προκειμένου να διασφαλιστεί η θετική συμβολή τους στην απόδοση του μοντέλου.*

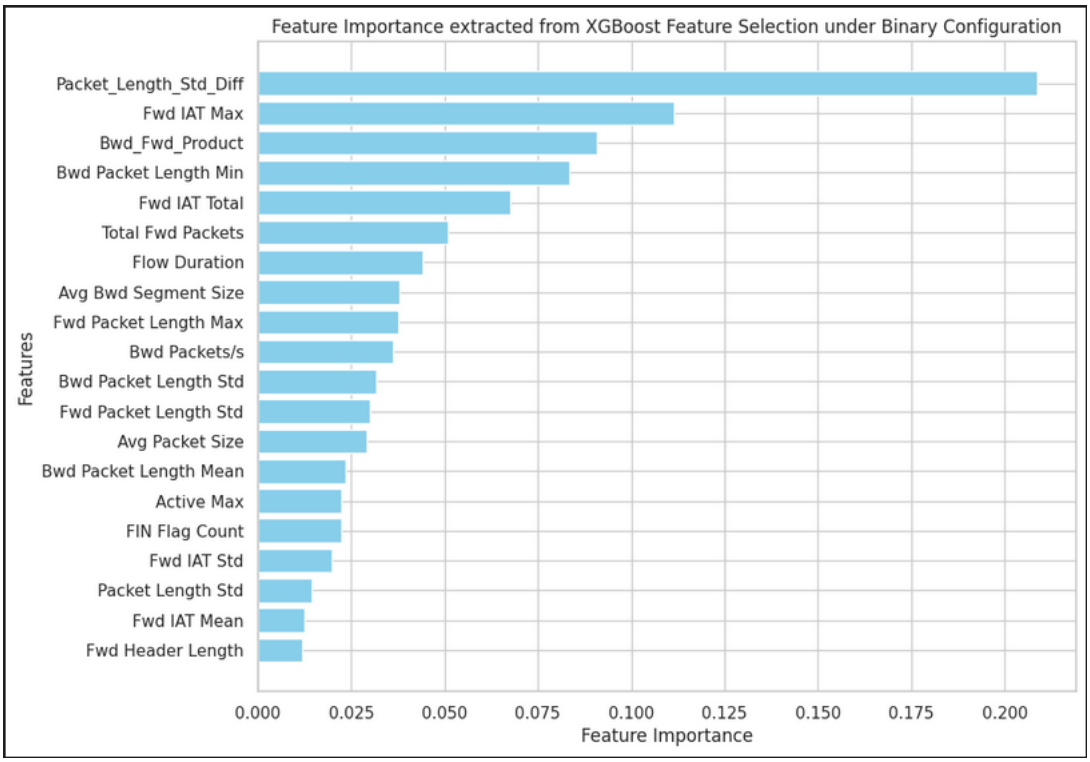
5.2.3 Επαναληπτική Αξιολόγηση

Κατά τη διάρκεια αυτής της φάσης, διεξάγεται μια δευτερεύουσα αξιολόγηση της επίδρασης των πρόσφατα δημιουργημένων χαρακτηριστικών. Χρησιμοποιείται η ίδια μέθοδος επιλογής χαρακτηριστικών με χρήση XGBoost.

Αναδεικνύονται 20 επιλεγμένα χαρακτηριστικά, με την αντίστοιχη σημασία τους:

Packet Length Std Diff: 20.87%, *Fwd IAT Max*: 11.13%, *Bwd Fwd Product*: 9.08%, *Bwd Packet Length Min*: 8.35%, *Fwd IAT Total*: 6.75%, *Total Fwd Packets*: 5.10%, *Flow Duration*: 4.42%, *Avg Bwd Segment Size*: 3.80%, *Fwd Packet Length Max*: 3.77%, *Bwd Packets/s*: 3.64%, *Bwd Packet Length Std*: 3.18%, *Fwd Packet Length Std*: 2.99%, *Avg Packet Size*: 2.92%, *Bwd Packet Length Mean*: 2.35%, *Active Max*: 2.25%, *FIN Flag Count*: 2.23%, *Fwd IAT Std*: 1.99%, *Packet Length Std*: 1.44%, *Fwd IAT Mean*: 1.24%, *Fwd Header Length*: 1.19%.

Σχήμα 5.2: Απεικόνιση Αποτελεσμάτων της Επαναληπτικής Αξιολόγησης



Η επαναληπτική αξιολόγηση πραγματοποιείται υπό τη μεθοδολογία της προηγούμενης παραγράφου, ήτοι τον αλγόριθμο XGBoost.

Συμπέρασμα Συνοψίζοντας, η διαδικασία κατασκευής χαρακτηριστικών ολοκληρώνεται επιτυχώς, καθώς δύο από τα τέσσερα δημιουργηθέντα χαρακτηριστικά επιλέγονται από τον αλγόριθμο XGBoost. Επιπλέον, σχεδόν όλα τα αρχικά επιλεγμένα χαρακτηριστικά διατηρούνται, με εξαίρεση τα *"Fwd Act Data Packets"* και *"Fwd Packet Length Min"*, τα οποία αντικαθίστανται από τα *"Bwd Fwd Product"* και *"Packet Length Std Diff"*. Ως αποτέλεσμα, επιλέγονται συνολικά 20 χαρακτηριστικά υπό τη δυαδική διάταξη.

5.3 SMOTE για μη Ισορροπημένα Σύνολα Δεδομένων

Τα μη ισορροπημένα σύνολα δεδομένων παρουσιάζουν σημαντικές προκλήσεις στους τομείς της μηχανικής και βαθιάς μάθησης, συχνά οδηγώντας σε προκατειλημμένα μοντέλα που δεν αποδίδουν επιθυμητά στη μειοψηφική κατηγορία. Η Τεχνική Συνθετικής Υπερδειγματοληψίας της Μειοψηφίας (Synthetic Minority Oversampling Technique - SMOTE) είναι μια ευρέως χρησιμοποιούμενη μέθοδος προεπεξεργασίας που έχει σχεδιαστεί για να αντιμετωπίσει αυτό το ζήτημα, μεταβάλλοντας την κατανομή του συνόλου δεδομένων. Σε αυτήν την ενότητα, θα μελετηθεί η θεωρία πίσω από την SMOTE, εστιάζοντας στη χρήση της στη δυαδική ταξινόμηση. Η παρούσα συζήτησή εκτείνεται κυρίως στην περίπτωση δυαδικής διάταξης, όμως οι ίδιες αρχές μπορούν να εφαρμοστούν και σε πολυκατηγορικές διατάξεις. Ωστόσο, είναι σημαντικό να σημειωθεί ότι στα πειράματά υπό πολυκατηγορική ταξινόμηση, η SMOTE δεν βελτίωσε την απόδοση για κανένα μοντέλο και ως εκ τούτου, δεν επιλέχθη ως βήμα προεπεξεργασίας σε αυτές τις περιπτώσεις.

5.3.1 Εισαγωγή

Το σύνολο δεδομένων CIC-IDS-2017 υποφέρει από ανισομέρεια κλάσεων, δηλαδή υπάρχουν σημαντικά περισσότερα δείγματα κανονικής κίνησης (benign traffic) σε σύγκριση με κακόβουλες επιθέσεις (malicious attacks).

- Ποσοστό Κλάσης 0: 84.92% - Πλειοψηφική κλάση / Benign Activity
- Ποσοστό Κλάσης 1: 15.08% - Μειοψηφική κλάση / Malicious Activity

Η ανισορροπία αυτή εκτρέπει τα μοντέλα Μηχανικής και Βαθιάς Μάθησης προς την πλειοψηφική κατηγορία, δηλαδή δεν αποφέρουν αξιόλογα αποτελέσματα στην ανίχνευση σπάνιων τύπων επιθέσεων. Για την αρχιτεκτονική TabNet, όλα τα βήματα προεπεξεργασίας διαχειρίζονται εσωτερικά, συμπεριλαμβανομένης της διαχείρισης της ασυμμετρίας της κατανομής - στόχου. Ωστόσο, για τα MLPs και τα CNNs, η ανισορροπία του συνόλου δεδομένων επηρεάζει αρνητικά την απόδοση της ταξινόμησης.

Για την αντιμετώπιση του εν λόγω ζητήματος, προτείνεται η χρήση της SMOTE, ήτοι η εξισορρόπηση της κατανομής - στόχου, ως βήμα προεπεξεργασίας, [54]. Η SMOTE δημιουργεί τεχνητά δείγματα της κλάσης μειοψηφίας, με σκοπό να αυξήσει την παρουσία τους (υπερδειγματοληψία). Στη δυαδική ταξινόμηση, αυτή η μέθοδος χρησιμοποιεί στατιστική δειγματοληψία από την ομοιόμορφη κατανομή και τον αλγόριθμο k-Nearest Neighbor (k-NN). Η εφαρμογή της SMOTE μπορεί να βελτιώσει την απόδοση της ταξινόμησης κατά 1-5% ως προς την ακρίβεια επί του συνόλου δεδομένων CIC-IDS-2017, καθιστώντας την μία ισχυρά προτεινόμενη τεχνική προεπεξεργασίας.

5.3.2 Μαθηματική Περιγραφή

Η μέθοδος SMOTE (Synthetic Minority Oversampling Technique) αντιμετωπίζει την ανισορροπία των κλάσεων με την τεχνητή δημιουργία νέων δεδομένων της μειοψηφικής κλάσης. Η μεθοδολογία μπορεί να παρουσιαστεί ως αλγόριθμος δημιουργίας τεχνητών σημείων στον χώρο των χαρακτηριστικών, με τις συνθήκες τερματισμού να καθορίζονται εξωτερικά από τον χρήστη.

> *Ο παραδοσιακός αλγόριθμος k-NN εφαρμόζεται εδώ, με $k = 5$, όπως περιγράφεται στο documentation του imbalanced-learn, [55].*

Αλγόριθμος 5.3 SMOTE

1. **Identifying Minority Class**: Αρχικά, αναγνωρίζεται η μειοψηφική κλάση στο σύνολο δεδομένων.

2. **Selecting a Data Point**: Έπειτα, ένα τυχαίο σημείο δεδομένων επιλέγεται από τη μειοψηφική κλάση.

3. **Finding Nearest Neighbors**: Στη συνέχεια, η SMOTE εντοπίζει τους k κοντινότερους γείτονες του επιλεγμένου σημείου δεδομένων. Οι εν λόγω γείτονες θα ανήκουν στην ίδια κλάση.

4. **Synthetic Data Generation**: Η SMOTE επιλέγει τυχαία έναν από τους k κοντινότερους γείτονες. Στη συνέχεια, δημιουργεί ένα νέο συνθετικό σημείο δεδομένων υπολογίζοντας τη διαφορά μεταξύ του επιλεγμένου σημείου και του γείτονά του. Αυτό το διάνυσμα διαφοράς πολλαπλασιάζεται με έναν τυχαίο αριθμό μεταξύ 0 και 1 και προστίθεται στο αρχικό σημείο, με αποτέλεσμα ένα νέο συνθετικό σημείο δεδομένων εντός του χώρου χαρακτηριστικών, ανήκον στη μειοψηφούσα κλάση.

Μαθηματική Περιγραφή Έστω $\vec{x} \in \mathbb{R}^n$ το επιλεγμένο σημείο δεδομένων της μειοψηφικής κλάσης και k ο αριθμός των κοντινότερων γειτόνων. Έστω $y \in \mathbb{R}^n$ κάποιος από τους k κοντινότερους γείτονες, τυχαία επιλεγμένος. Το νέο συνθετικό σημείο δεδομένων, s, θα προκύψει από την ακόλουθη εξίσωση:

$$s = x + \lambda \cdot (y - x) \tag{5.4}$$

όπου λ ένας τυχαίος αριθμός μεταξύ 0 και 1 προερχόμενος από μία ομοιόμορφη κατανομή.

> *Οι αναγνώστες ενθαρρύνονται να επισκεφτούν το documentation,* [55]*, καθώς και την πρωταρχική εργασία των Chawla et al. (2002),* [54]*, για μια εις βάθος κατανόηση της SMOTE.*

5.3.3 Συνέπειες

Με τη δημιουργία συνθετικών σημείων δεδομένων μειοψηφικής κλάσης, η SMOTE εξισορροπεί την κατανομή των κλάσεων. Με αυτόν τον τρόπο, τα μοντέλα μάθησης μαθαίνουν από ένα πιο αντιπροσωπευτικό δείγμα. Ως αποτέλεσμα, βελτιώνεται η ανίχνευση ανωμαλιών, όπως οι κυβερνοεπιθέσεις στην περίπτωση του συνόλου δεδομένων CIC-IDS-2017.

Ωστόσο, αν και η SMOTE είναι μια δημοφιλής τεχνική, δεν είναι πανάκεια και έχει σημαντικούς περιορισμούς. Εν γένει, η απλή δημιουργία νέων σημείων δεδομένων μπορεί να μην αποτυπώνει την πραγματική πολυπλοκότητα της κατανομής της μειοψηφικής κλάσης.

Αποτροπή Data Leakage Η διεργασία της SMOTE θα πρέπει να περιορίζεται μόνο στο σύνολο εκπαίδευσης. Η εφαρμογή της SMOTE στα validation και test sets εισάγει συνθετικά δείγματα από την κλάση μειοψηφίας, οδηγώντας σε data leakage (διαρροή δεδομένων - πληροφοριών) και μεροληψία στην αξιολόγηση του μοντέλου. Τα validation και test sets πρέπει να αντικατοπτρίζουν την πραγματική κατανομή των δεδομένων και να παραμένουν αμετάβλητα για την ακριβή αξιολόγηση της γενίκευσης του μοντέλου. Τελικά, με τον περιορισμό της SMOTE μόνο στο σύνολο εκπαίδευσης, εξασφαλίζεται αμερόληπτη αξιολόγηση και αξιόπιστη εκτίμηση της απόδοσης του μοντέλου σε άγνωστα δεδομένα.

5.3.4 Σύντομη επισκόπηση του αλγορίθμου k-NN

Για λόγους πληρότητας, παρουσιάζεται μια σύντομη επισκόπηση του αλγορίθμου k-NN, [56].

Ο αλγόριθμος k-Nearest Neighbors (k-NN) είναι μια απλή και αποτελεσματική μέθοδος που χρησιμοποιείται για εργασίες ταξινόμησης . Η βασική ιδέα του αλγορίθμου είναι η διεξαγωγή προβλέψεων με βάση τα k πλησιέστερα δείγματα εκπαίδευσης στον χώρο χαρακτηριστικών.

Αλγόριθμος 5.4 Περιγραφή του Αλγόριθμου Ταξινόμησης k-NN

1. **Choose the number of neighbors** (k): Αρχικά, καθορίζεται ο αριθμός των πλησιέστερων γειτόνων προς εξέταση. Η εν λόγω παράμετρος k ορίζεται από τον χρήστη.
2. **Compute Distances**: Για δεδομένο σημείο από το test set, υπολογίζεται η απόσταση μεταξύ αυτού του σημείου και όλων των σημείων στο σύνολο εκπαίδευσης. Η απόσταση εξάγεται από κάποια νόρμα, π.χ. την Ευκλείδεια νόρμα, τις p-norms ή την max-νόρμα.
3. **Identify Neighbors**: Επιλέγονται k σημεία δεδομένων από το σύνολο εκπαίδευσης, τα πλέον πλησιέστερα στο σημείο δοκιμής.
4. **Make Predictions**: Η ετικέτα κατηγορίας του εν λόγω σημείου καθορίζεται από την πλειοψηφία ψήφων (majority vote) των k πλησιέστερων γειτόνων του, δηλαδή κάθε γείτονας "ψηφίζει" για την κατηγορία του, και η κατηγορία με τις περισσότερες ψήφους ανατίθεται στο σημείο δοκιμής.

Ο αλγόριθμος k-NN προσφέρει απλότητα και αποτελεσματικότητα κατά την εκπαίδευση, αλλά χαρακτηρίζεται από trade-offs, όπως αργοί χρόνοι πρόβλεψης για μεγάλα σύνολα δεδομένων, υψηλή χρήση μνήμης λόγω της αποθήκευσης ολόκληρου του συνόλου εκπαίδευσης και ευαισθησία σε outliers ή noisy features που μπορούν να επηρεάσουν αρνητικά την απόδοση.

Η εξήγηση των βασικών αρχών του k-NN συμβάλλει στη διαισθητική κατανόηση της δημιουργίας συνθετικών δειγμάτων της SMOTE. Για παράδειγμα, οι αργοί χρόνοι πρόβλεψης που είναι εγγενείς στον k-NN επεξηγούν τον παρατεταμένο χρόνο επεξεργασίας που απαιτείται από τη SMOTE.

5.4 Data Scaling υπό τον Quantile Transformer

Στο πλαίσιο της προεπεξεργασίας δεδομένων, ως Data Scaling αναφέρεται η διαδικασία μετασχηματισμού των τιμών των χαρακτηριστικών σε μια τυποποιημένη κλίμακα, προς διευθέτηση ζητημάτων εκπαίδευσης. Μεταξύ διαφόρων μεθοδολογιών, ο Quantile Transformer επιλέγεται για όλα τα μοντέλα. Σε αυτήν την ενότητα, παρέχεται μια ολοκληρωμένη παρουσίαση αυτής της μεθόδου μετασχηματισμού, η οποία υλοποιείται στο scikit-learn, [18].

Εναλλακτικές Μέθοδοι Δεδομένων των ασύμμετρων, μη γκαουσιανών κατανομών που παρατηρούνται στο σύνολο δεδομένων, είναι απαραίτητος ο πειραματισμός με μεθόδους data scaling που αντιμετωπίζουν τα έντονα ασύμμετρα features και τη σημαντική παρουσία σημείων - outliers. Για τους λόγους αυτούς, ο Quantile Transformer επιλέγεται μεταξύ άλλων τεχνικών scaling [18], συγκεκριμένα: *Unit Vector Scaling (Normalization), Min-Max Scaling (ή Min-Max Normalization), Standardization (ή Z-Score Normalization), Robust Scaling, Max Abs Scaling, Log Transformation, Exponential Transformation, Box-Cox Transformation και Yeo-Johnson Transformation.*

Μαθηματική Περιγραφή Η μαθηματική διαίσθηση του εν λόγω μετασχηματισμού, συνίσταται στις quantile (ποσοστιαίες) συναρτήσεις, [57]. Για μια συνεχή τυχαία μεταβλητή με αθροιστική συνάρτηση κατανομής (cumulative distribution function - CDF) F, η quantile συνάρτηση $Q : [0, 1] \mapsto \mathbb{R}$ ορίζεται ως η αντίστροφη συνάρτηση της CDF:

$$Q(p) = F^{-1}(p) \tag{5.5}$$

Ο Quantile Transformer, ως τεχνική προεπεξεργασίας, μετασχηματίζει τα χαρακτηριστικά ώστε να ακολουθούν μια επιθυμητή κατανομή, συχνά ομοιόμορφη ή κανονική. Έστω X πίνακας χαρακτηριστικών με n δείγματα και p το πλήθος χαρακτηριστικά, τότε ο μετασχηματισμός για κάθε χαρακτηριστικό $j \in \{1, \ldots, p\}$ θα πραγματοποιηθεί ως εξής:

1. **Estimate Empirical Cumulative Distribution Function (ECDF)**: Συντελείται υπολογισμός της ECDF, συμβολιζόμενη ως $\hat{F}_j(x)$, για κάθε χαρακτηριστικό j, η οποία αντιπροσωπεύει την αναλογία των δειγμάτων στο X_j που είναι μικρότερα ή ίσα με την τιμή x.

2. **Quantile Mapping to Uniform Distribution**: Απεικόνιση κάθε σημείου δεδομένων $x_i^{(j)} \in X_j$ στην αντίστοιχη ποσοστιαία (quantile) τιμή $u_i^{(j)}$ μεταξύ 0 και 1, με χρήση της αντίστροφης ECDF:

   $$u_i^{(j)} = \hat{F}_j\left(x_i^{(j)}\right) \tag{5.6}$$

   Με αυτόν τον τρόπο, το feature X_j μετασχηματίζεται σε μια ομοιόμορφη κατανομή.

3. **Quantile Mapping to Target Distribution**: Τέλος, τα ομοιόμορφα πλέον quantiles $u_i^{(j)}$ απεικονίζονται στην κατανομή - στόχο (π.χ., μία τυπική κανονική με μέση τιμή μ και τυπική απόκλιση σ) χρησιμοποιώντας την quantile συνάρτησή, ήτοι την αντίστροφη CDF, που συμβολίζεται ως $Q_T(u)$:

   $$x_i^{(j')} = Q_T\left(u_i^{(j)}\right)$$

   Με αυτόν τον τρόπο, τα δεδομένα μετασχηματίζονται ώστε να έχουν την επιθυμητή κατανομή.

*Για την τρέχουσα υλοποίηση του μετασχηματισμού, η **ομοιόμορφη κατανομή** υιοθετείται ως η κατανομή - στόχος, επομένως το τελευταίο βήμα δεν είναι απαραίτητο.*

Ειδικότερα, η ECDF - Εμπειρική Αθροιστική Συνάρτηση Κατανομής (με τον όρο *εμπειρική*, αναφερόμαστε στην εμπειρική πιθανότητα ή πιθανότητα κατά Laplace ή διαφορετικά τη συχνότητα εμφάνισης) ενός δείγματος $X_1, X_2, \ldots, X_n$ πλήθους n είναι μια βηματική συνάρτηση που αυξάνεται κατά $\frac{1}{n}$ σε κάθε ένα από τα n σημεία δεδομένων. Μαθηματικά, η ECDF $F_n(x)$ σε ένα σημείο x ορίζεται ως:

$$F_n(x) = \frac{1}{n} \sum_{i=1}^{n} 1_{\{X_i \leq x\}} \tag{5.7}$$

και η δείκτρια συνάρτηση ορίζεται ως: $1_{\{X_i \leq x\}} = \begin{cases} 1, & X_i \leq x \\ 0, & X_i > x \end{cases}$.

Πρόληψη Data Leakage Ο Quantile Transformer έχει υλοποιηθεί για όλα τα μοντέλα, με στόχο τη διασφάλιση ομοιόμορφης κατανομής των αντίστοιχων συνόλων χαρακτηριστικών. Ο μετασχηματισμός εφαρμόζεται στα training, validation και test sets. Ωστόσο, οι μέθοδοι εφαρμογής διαφέρουν για να αποτρέψουν ενδεχόμενο data leakage.

Το data leakage είναι ένα κρίσιμο ζήτημα στη μηχανική μάθηση, αναφερόμενο στην ακούσια έκθεση πληροφοριών από τα δεδομένα εκπαίδευσης στα σύνολα επικύρωσης και αξιολόγησης. Επακολούθως, πραγματοποιείται υπερεκτίμηση της απόδοσης, καθώς το μοντέλο έχει τη δυνατότητα να μάθει από πληροφορίες στις οποίες δεν θα είχε πρόσβαση υπό πραγματικές συνθήκες. Προκειμένου να αποτραπεί αυτός ο κίνδυνος, ο transformer προσαρμόζεται στα δεδομένα εκπαίδευσης και στη συνέχεια εφαρμόζεται ο μετασχηματισμός. Για τα σύνολα validation και test, χρησιμοποιείται μόνο η μέθοδος μετασχηματισμού, εφαρμόζουσα τις scaling παραμέτρους προερχόμενες από το συνόλο εκπαίδευσης, δίχως την επαναπροσαρμογή του μετασχηματιστή. Με αυτόν τον τρόπο, εξασφαλίζεται ότι η διαδικασία data scaling παραμένει συνεπής σε διαφορετικά σύνολα δεδομένων και διατηρεί την ακεραιότητα της αξιολόγησης του μοντέλου αποτρέποντας την εισαγωγή μεροληψίας ή ανακριβειών. Ως εκ τούτου, αυτή η μέθοδος υποστηρίζει μια αξιόπιστη αξιολόγηση της απόδοσης του μοντέλου σε άγνωστα δεδομένα, παρέχοντας μια αμερόληπτη εκτίμηση της γενίκευσής του.

5.5 Ο Μετασχηματισμός Tab2Img

Σε αυτήν την ενότητα, παρουσιάζεται η διαδικασία μετασχηματισμού των tabular data σε εικόνες, ως απαραίτητο βήμα προεπεξεργασίας για τις CNN-pipelines. Ο μετασχηματισμός αυτός εφαρμόζεται τόσο σε δυαδικές όσο και σε πολυκατηγορικές διατάξεις.

5.5.1 Εισαγωγή

Η βιβλιοθήκη Tab2Img αναφέρεται στην καινοτόμο προσέγγιση μετασχηματισμού των tabular data (δεδομένων μορφής πίνακα) σε εικόνες [58], επιτρέποντας τη χρήση ισχυρών εργαλείων ταξινόμησης εικόνων βαθιάς μάθησης. Αυτή η μέθοδος στοχεύει στην υπέρβαση ορισμένων εγγενών περιορισμών των παραδοσιακών τεχνικών μηχανικής μάθησης που εφαρμόζονται σε πίνακες δεδομένων, αξιοποιώντας τη χωρική δομή και τις δυνατότητες εξαγωγής χαρακτηριστικών των CNNs.

Υπόβαθρο Οι παραδοσιακές μέθοδοι μάθησης μπορεί να μην καταγράφουν αποτελεσματικά σύνθετα μοτίβα. Η μετατροπή των πινάκων δεδομένων σε εικόνες μέσω της βιβλιοθήκης Tab2Img παρουσιάζει μια εναλλακτική λύση, απεικονίζοντας τα δεδομένα σε μια οπτική δομή, η οποία μπορεί να επεξεργαστεί καλύτερα από συγκεκριμένα μοντέλα βαθιάς μάθησης. Ειδικότερα, τα CNNs είναι αποδοτικά σε εργασίες ταξινόμησης εικόνων λόγω της ικανότητάς τους να μαθαίνουν ιεραρχικές αναπαραστάσεις χαρακτηριστικών. Με τη μετατροπή των πινάκων δεδομένων σε εικόνες αξιοποιούνται αυτές οι δυνατότητες, βελτιώνοντας ενδεχομένως την απόδοση της ταξινόμησης.

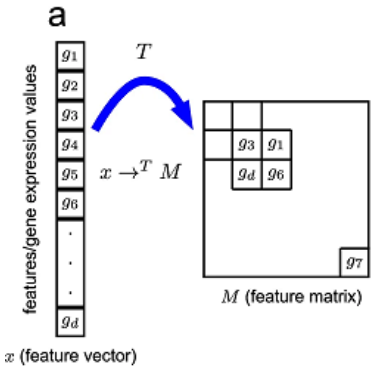
5.5.2 Μαθηματική Περιγραφή

> *Η παρακάτω επεξήγηση της βιβλιοθήκης Tab2Img προέρχεται απευθείας από το documentation, [59].*

Στο paper "DeepInsight: A methodology to transform a non-image data to an image for convolution neural network architecture", [58] οι συγγραφείς προτείνουν μια μέθοδο μετατρο-

πής των πινάκων δεδομένων σε εικόνες, με σκοπό την αξιοποίηση της ισχύος των συνελικτικών νευρωνικών δικτύων.

Σχήμα 5.3: Ο DeepInsight προταθείς μετασχηματισμός



Η εικόνα απεικονίζει την κύρια ιδέα: δεδομένου ενός συνόλου εκπαίδευσης με $X \in \mathbb{R}^{m \times n}$ με m δείγματα και n χαρακτηριστικά, απαιτείται η εύρεση συνάρτησης $M : \mathbb{R}^{m \times n} \to \mathbb{R}^{m \times d \times d}$, όπου $d = \lceil \sqrt{n} \rceil$. Υπάρχουν πολυάριθμοι τρόποι επιλογής της M. Πηγή: [59].

Στην παρούσα υλοποίηση, τα χαρακτηριστικά οργανώνονται ως προς το διάνυσμα συσχέτισης $\rho(X, Y)$, όπου Y είναι η κατανομή - στόχος. Δεδομένων των X και Y ως εξής:

$$X = \begin{pmatrix} x_1^{(1)} & \dots & x_n^{(1)} \\ \vdots & \ddots & \vdots \\ x_1^{(m)} & \dots & x_n^{(m)} \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

τότε, το διάνυσμα $\rho(X, Y) = (\rho_1, \dots, \rho_n)$ εκφράζει τον συντελεστή συσχέτισης **Pearson**:

$$\rho = \frac{\mathrm{Cov}(x, y)}{\sigma(x) \cdot \sigma(y)} \tag{5.8}$$

όπου:

$$\rho_i = \rho(X_i, Y), \quad X_i = \begin{pmatrix} x_i^{(1)} \\ \vdots \\ x_i^{(m)} \end{pmatrix}$$

Σε αυτό το σημείο, το $\rho(X, Y)$ ταξινομείται από το μεγαλύτερο προς το μικρότερο στοιχείο, δημιουργώντας ένα διάνυσμα δεικτών:

$$\boldsymbol{J} = (J_k : \rho(X_j) \geq \rho(X_{j-1}), k \in [1, \dots, n]) \tag{5.9}$$

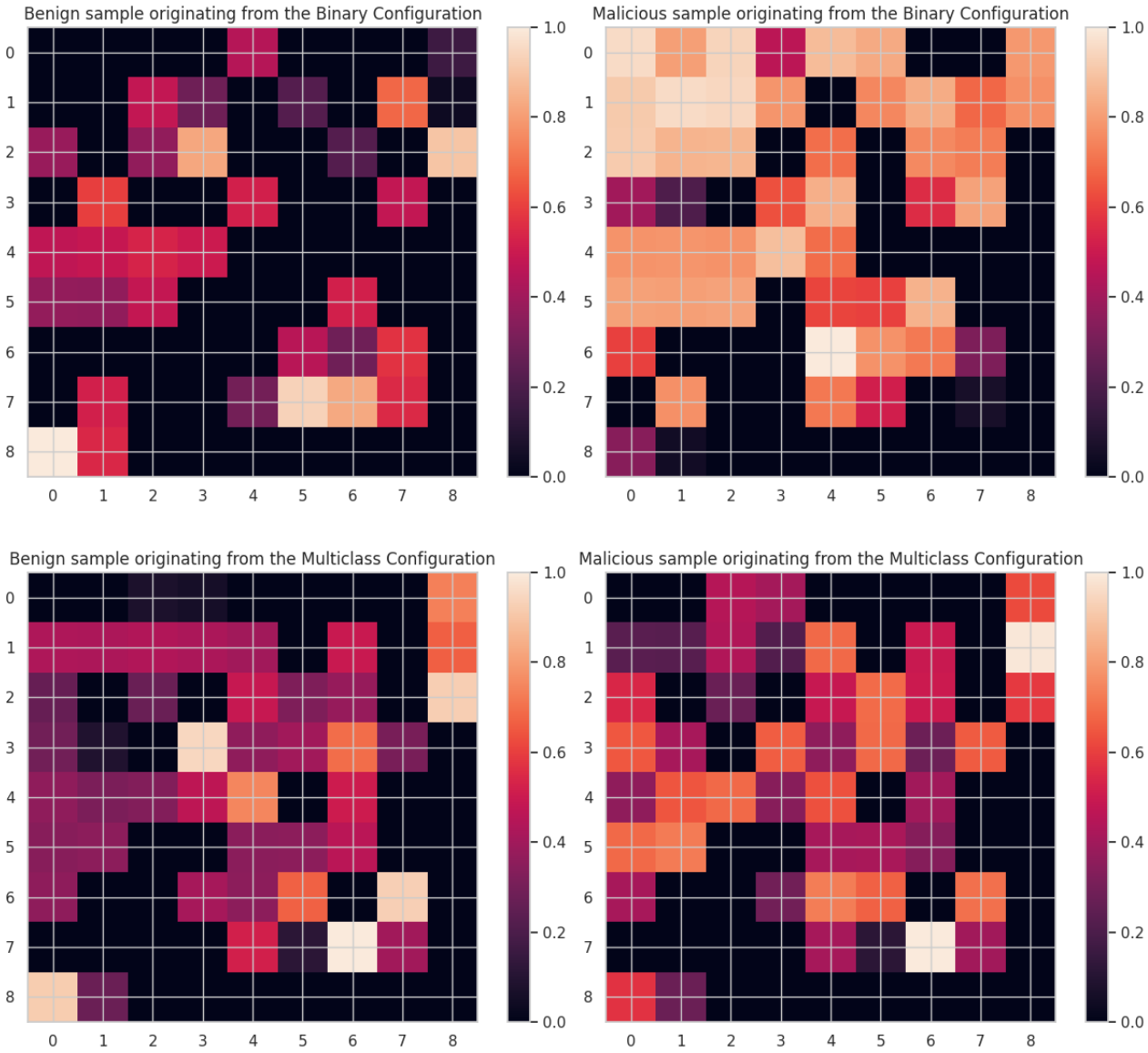
Τελικά, ο τανυστής M θα είναι:

$$M = \begin{pmatrix} X_{J_1} & X_{J_2} & X_{J_5} & \dots \\ X_{J_3} & X_{J_4} & X_{J_7} & \dots \\ X_{J_6} & X_{J_8} & X_{J_9} & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \tag{5.10}$$

Η συνάρτηση αντιστοίχισης του στοιχείου J_k στην κατάλληλη γραμμή και στήλη $(r, c)_k$ του M ορίζεται ως:

$$(r,c)_k = \begin{cases} \left(\sqrt{k}, \sqrt{k}\right) & \text{if } \sqrt{k} \in \mathbb{N} \\ \left(\left\lceil \sqrt{k} \right\rceil, \left\lceil \sqrt{k} \right\rceil - \frac{1}{2}\left(\left\lceil \sqrt{k} \right\rceil^2 - k\right)\right) & \text{if } \sqrt{k} \notin \mathbb{N} \text{ and } \left\lceil \sqrt{k} \right\rceil^2 - k = 0 \bmod 2 \\ \left(\left\lceil \sqrt{k} \right\rceil - \frac{1}{2}\left(\left\lceil \sqrt{k} \right\rceil^2 - k\right), \left\lceil \sqrt{k} \right\rceil\right) & \text{if } \sqrt{k} \notin \mathbb{N} \text{ and } \left\lceil \sqrt{k} \right\rceil^2 - k \neq 0 \bmod 2 \end{cases} \tag{5.11}$$

Σχήμα 5.4: Ο Μετασχηματισμός Tab2Img



Παρουσίαση δειγμάτων του μετασχηματισμού Tab2Img. Αυτές οι τέσσερις εικόνες αντιστοιχούν σε τέσσερις γραμμές πινάκων δεδομένων, προερχόμενες από τα σύνολα εκπαίδευσης (δυαδικά και πολυκατηγορικά). Συγκεκριμένα, τα καλοήθη δείγματα αντιστοιχούν στις πρώτες γραμμές των συνόλων, ενώ οι "κακόβουλες εικόνες" αντιστοιχούν στις τελευταίες γραμμές.

Σε αυτό το πλαίσιο, ο συντελεστής συσχέτισης Pearson για ένα δείγμα X υλοποιείται ως εξής:

$$\rho(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2} \cdot \sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Αυτή η μέθοδος μετασχηματισμού διασφαλίζει ότι τα δεδομένα σε πίνακες απεικονίζονται αποτελεσματικά σε μορφή εικόνας, διατηρώντας και αναδεικνύοντας τις υποκείμενες σχέσεις μεταξύ των χαρακτηριστικών για βελτιωμένη απόδοση ταξινόμησης υπό την χρήση CNNs.

5.5.3 Υλοποίηση

Ο μετασχηματισμός Tab2Img εφαρμόζεται στο σύνολο δεδομένων CIC-IDS-2017. Δεδομένου ότι η προεπεξεργασμένη μορφή του συνόλου δεδομένων περιλαμβάνει 77 χαρακτηριστικά, οι διαστάσεις της εικόνας υπολογίζονται ως $d = \lceil \sqrt{n} \rceil \overset{n=77}{\Longrightarrow} d = 9$. Κατά συνέπεια, κάθε γραμμή από το σύνολο δεδομένων μετατρέπεται σε μια 9×9 greyscale εικόνα. Παρά τις πολύχρωμες εικόνες που παρουσιάζονται στα σχήματα, οι παραγόμενες εικόνες είναι στην πραγματικότητα ασπρόμαυρες, καθώς έχουν μία μόνο διάσταση χρώματος: είναι εικόνες $9 \times 9 \times 1$, υπό διαστάσεις ύψους, πλάτους και χρώματος. Μια άλλη κοινή μορφή είναι ο τύπος εικόνας RGB: εικόνες $9 \times 9 \times 3$, με τρία κανάλια χρώματος: κόκκινο, πράσινο και μπλε.

Πρακτικά, ο στόχος του μετασχηματισμού είναι η επιτυχής προσέγγιση του καναλιού χρώματος. Σύμφωνα με το σχήμα, το κανάλι χρώματος περιορίζεται σε τιμές που κυμαίνονται από 0 έως 1, γεγονός που αντιτίθεται στην αντιστοιχία κάθε pixel με έναν συντελεστή Pearson, ο οποίος παίρνει τιμές από -1 έως 1. Η εξήγηση αυτού του φαινομένου έγγυται στο γεγονός ότι η είσοδος του μετασχηματισμού έχει υποστεί προεπεξεργασία και συγκεκριμένα, τη διαδικασία του Quantile Transformation. Καθώς τα δεδομένα ακολουθούν μια συγκεκριμένη ομοιόμορφη κατανομή, το εύρος των τιμών κανονικοποιείται εντός του διαστήματος $[0, 1]$.

Πρόληψη Data Leakage Η υλοποίηση του μετασχηματισμού Tab2Img θα πρέπει να αποτρέπει δυνητικό data leakage και να διασφαλίζει την ακεραιότητα της αξιολόγησης του μοντέλου. Η εν λόγω πρακτική έχει σχεδιαστεί με τρόπο ώστε οι σχέσεις και τα μοτίβα που εξάγονται από το σύνολο εκπαίδευσης να χρησιμοποιούνται για τον μετασχηματισμό των validation και test sets. Συγκεκριμένα, ο μετασχηματισμός εφαρμόζεται στο σύνολο εκπαίδευσης, καταγράφοντας απαραίτητες συσχετίσεις ή μοτίβα και παράγοντας παραμετρικές τιμές, όπως το σύνολο δεικτών $\boldsymbol{J}$, που προβάλλονται στη συνέχεια στα σύνολα επικύρωσης και αξιολόγησης δίχως επανεκπαίδευση. Αυτή η μέθοδος διασφαλίζει ότι καμμία πληροφορία από τα σύνολα validation ή test δεν επηρεάζει τη διαδικασία εκπαίδευσης, αποτρέποντας έτσι το ενδεχόμενο data leakage. *Όπως συζητήθηκε και προηγουμένως,* η διατήρηση μιας σαφούς διάκριση μεταξύ των δεδομένων εκπαίδευσης και των υπόλοιπων δεδομένων, επιτυγχάνει μια αξιόπιστη και αμερόληπτη αξιολόγηση της απόδοσης του μοντέλου σε άγνωστα δεδομένα.

Κεφάλαιο 6

Αναπτύσσοντας ένα IDS

6.1 Αγωγοί IDS

6.1.1 Η Έννοια του Αγωγού

Στο πλαίσιο της μηχανικής και της βαθιάς μάθησης, η έννοια του **αγωγού** (pipeline) αναφέρεται σε μια σειρά βημάτων επεξεργασίας δεδομένων, στα πλαίσια της μετατροπής ακατέργαστων δεδομένων σε μία μορφή αποτελεσματικά αξιοποιήσιμη από ένα μοντέλο, [60]. Οι αγωγοί αυτοματοποιούν τις επαναλαμβανόμενες εργασίες προεπεξεργασίας δεδομένων, διασφαλίζοντας τη συνέπεια και την αποδοτικότητα σε κάθε βήμα. Οι κύριες συνιστώσες ενός τυπικού αγωγού περιλαμβάνουν τον καθαρισμό των δεδομένων, τις διαδικασίες επιλογής και κατασκευής χαρακτηριστικών, την κλιμάκωση δεδομένων και τέλος την εκπαίδευση του μοντέλου. Κάθε ένα από αυτά τα βήματα, διατεταγμένα σε σειριακή ακολουθία, μπορεί να αναπτυχθεί, να δοκιμαστεί και να επαναχρησιμοποιηθεί ανεξάρτητα, προάγοντας τις δυνατότητες για modularity και maintainability. Η εν λόγω δομημένη προσέγγιση αποσκοπεί στη βελτιστοποίηση της ροής εργασίας αλλά και στη μείωση των ανθρώπινων σφαλμάτων, εξασφαλίζοντας ότι τα βήματα επεξεργασίας δεδομένων εφαρμόζονται ομοιόμορφα για κάθε σύνολο δεδομένων ή μοντέλο. Τελικά, η έννοια του αγωγού προσδιορίζει τη φάση της προεπεξεργασίας, αναπτύσσουντας μια τυποποιημένη και επαναλαμβανόμενη μεθοδολογία για τη δημιουργία μοντέλων.

6.1.2 Αγωγοί για ένα IDS

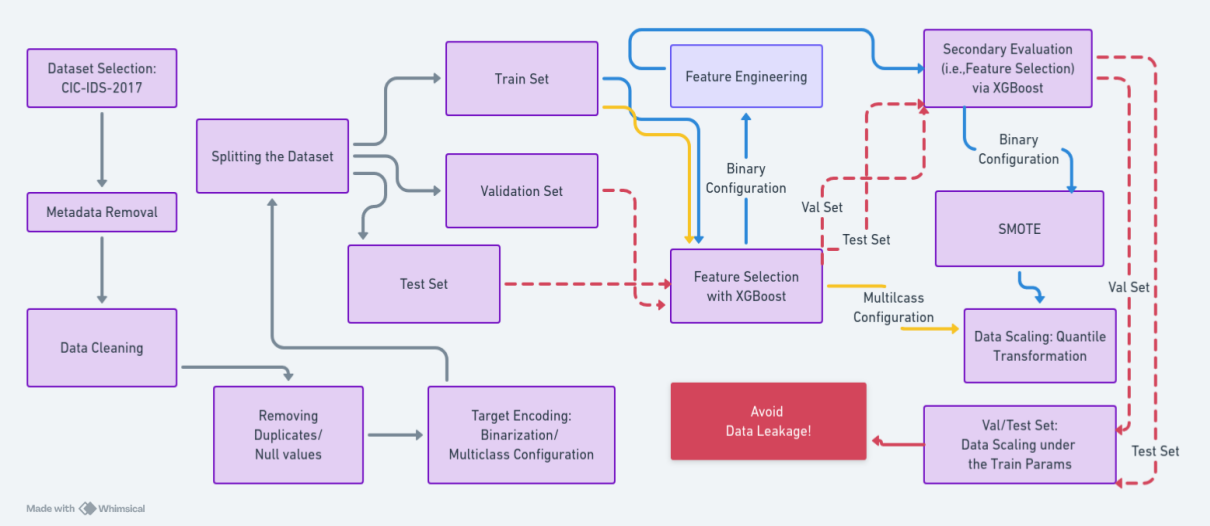
Στο πλαίσιο της υλοποίησης ενός συστήματος ανίχνευσης εισβολών επί του συνόλου δεδομένων CIC-IDS-2017, ζητούμενη είναι η ανάπτυξη ταξινομητών δικτυακής κίνησης. Συνεπώς απαιτείται ο σχεδιασμός, η υλοποίηση και η εκπαίδευση ταξινομητών, αλλά και η κατάλληλη προεπεξεργασία των δεδομένων. Επομένως, η ανάπτυξη ενός IDS ανάγεται στην ανάπτυξη ενός αντίστοιχου αγωγού. Δεδομένου ότι η παρούσα εργασία δεν περιλαμβάνει φάσεις συλλογής δεδομένων ή deployment phases, οι κύριες δραστηριότητες επικεντρώνονται στην προεπεξεργασία των ακατέργαστων δεδομένων και την εκπαίδευση των μοντέλων. Μέσω της δημιουργίας ενός αγωγού για κάθε μοντέλο, τα δεδομένα θα προεπεξεργαστούν αποτελεσματικά, και τα μοντέλα θα εκπαιδευτούν με συστηματικό και επαληθεύσιμο τρόπο. Με αυτόν τον τρόπο διασφαλίζεται η αξιοπιστία και η συνέπεια κατά την αναπτυξη ενός IDS.

6.1.3 Περιγραφή των Αγωγών

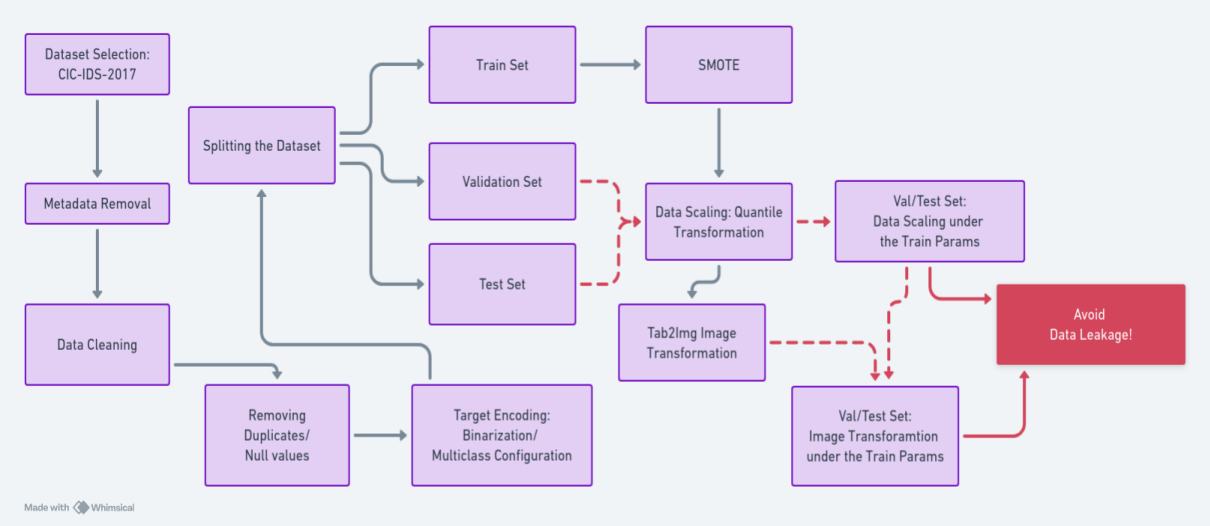
Λαμβάνοντας υπόψη τα παραπάνω, η ανάπτυξη των μοντέλων θα πρέπει να οργανωθεί κατάλληλα. Αρχικά, η υλοποίηση των ταξινομητών θα βασιστεί στις αρχιτεκτονικές που μελετήθηκαν στις εισαγωγικές ενότητες, ήτοι: MLPs, CNNs και Transformers.

Σχήμα 6.1: Αγωγοί προεπεξεργασίας - δυαδικές & πολυκατηγορικές διατάξεις

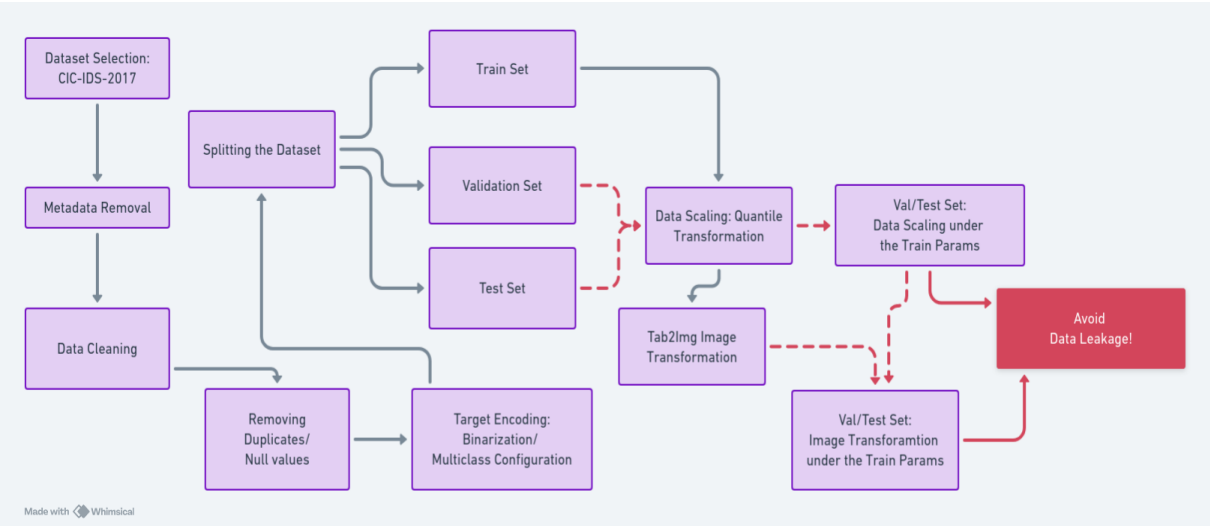
(α′) Αγωγοί των MLPs



(β′) Αγωγοί των CNNs



(γ′) Αγωγοί των TabNets



Παράλληλα, η διάταξη των δεδομένων εισόδου (δυαδική - binary ή πολυκατηγορική - multiclass) θα διαφοροποιήσει τους αγωγούς περαιτέρω, τόσο σε επίπεδο προεπεξεργασίας δεδομένων, όσο και ως προς τα εγγενή αρχιτεκτονικά γνωρίσματα των μοντέλων, δηλαδή τη συνάρτηση απώλειας και τη μορφολογία του επιπέδου εξόδου, με έμφαση στο είδος της συνάρτησης ενεργοποίησης και στο πλήθος των υπολογιστικών μονάδων του επιπέδου εξόδου. Ως εκ τούτου, επάγεται η ανάγκη για έξι αγωγούς:

$$3 \text{ αρχιτεκτονικές} \times 2 \text{ διατάξεις} = 6 \text{ αγωγοί}$$

Πέρα από τις αρχιτεκτονικές παραλλαγές, αξίζει να σημειωθεί ότι η διαδικασία προεπεξεργασίας μπορεί να προσαρμοστεί σύμφωνα με τη διάταξη, γεγονός που ίσως αντιτίθεται στη διαίσθηση του αναγνώστη. Ωστόσο δεν τίθεται κάποιος περιορισμός, και στα πλαίσια βελτιστοποίησης της απόδοσης ταξινόμησης μία τέτοια διαφοροποίηση ενδείκνυται. Συγκεκριμένα, ο αγωγός των MLPs παρουσιάζει αυτήν την απόκλιση, καθώς ύστερα από σχετική αξιολόγηση, παρατηρήθηκε ότι η απουσία της κατασκευής χαρακτηριστικών και της SMOTE, βελτιώνει την απόδοση ταξινόμησης. Η συγκεκριμένη παρατήρηση προκύπτει από τη δυνατότητα αξιολόγησης και επαλήθευσης κάθε μεμονωμένου βήματος, υπογραμμίζοντας την αξία των αγωγών.

$$6 \text{ αγωγοί} \times 5 \text{ μεγέθη} = 30 \text{ μοντέλα}$$

Για κάθε αγωγό απαιτείται η ανάπτυξη 5 μοντέλων διαφορετικών μεγεθών, άρα συνολικά τα ζητούμενα μοντέλα θα είναι 30. Στόχος είναι η μελέτη του trade-off απόδοσης ταξινόμησης (σε όρους accuracy), ως προς κατανάλωση μνήμης (memory usage). Σε κάθε κατηγορία των πέντε, το αύξον μέγεθος, προσμετρόμενο σε αριθμό παραμέτρων, θα συσχετιστεί με μία διαφορετική ακρίβεια, και έτσι θα προκύψει το υπό μελέτη trade-off. Εδώ, η έννοια του μεγέθους (size) μοντέλου, της κατανάλωσης μνήμης (memory usage) και του αριθμού παραμέτρων (number of parameters) ταυτίζονται. Στο σχήμα που ακολουθεί, απεικονίζονται οι αγωγοί προεπεξεργασίας των IDSs. Όπως αναφέρθηκε παραπάνω, η διαφοροποίηση στα MLPs αναπαρίσταται με διαφορετικό χρώμα (μπλε - κίτρινο), ενώ στις περιπτώσεις των CNNs και των Transformers η διαφοροποίηση λαμβάνει χώρα κατά την αριθμητική κωδικοποίηση του στόχου (target encoding).

6.2 Υλοποίηση

6.2.1 Σύνολα Εκπαίδευσης, Επικύρωσης και Αξιολόγησης

Στα πλαίσια της Βαθιάς Μάθησης, η ορθή διαχείριση των δεδομένων μέσω της χρήσης των συνόλων εκπαίδευσης, επικύρωσης και αξιολόγησης (train, validation και test sets) αποτελεί τυπική προσέγγιση για την εκπαίδευση, την επικύρωση και την τελική αξιολόγηση ενός μοντέλου. Σε κάθε περίπτωση, υιοθετήθηκε η ίδια ποσόστωση διαμέρισης του συνόλου δεδομένων:

$$\text{train set: } 75\%, \text{ validation set: } 10\%, \text{ test set: } 15\%$$

Το σύνολο εκπαίδευσης χρησιμοποιήθηκε για την προσαρμογή των παραμέτρων του μοντέλου. Τα δεδομένα αυτού του συνόλου τροφοδοτήθηκαν στα υπό ανάπτυξη μοντέλα, αποσκοπώντας στην εξαγωγή σχέσεων και προτύπων από τα δεδομένα. Ως πρωταρχικός στόχος του train set, τίθεται η μείωση του σφάλματος μεταξύ των προβλέψεων του μοντέλου και των πραγματικών τιμών (targets), μέσω της συνεχούς προσαρμογής των βαρών και των παραμέτρων μεροληψίας.

Το σύνολο επικύρωσης εξυπηρετεί την αξιολόγηση της απόδοσης του μοντέλου κατά τη διάρκεια της εκπαίδευσης. Προσμετρά την απόδοση του μοντέλου και προάγει πολιτικές αποτροπής του φαινομένου της υπερεκπαίδευσης (overfitting). Μέσω του validation set, διεξάγεται

η διαδικασία της ρύθμισης των υπερπαραμέτρων (hyperparameter tuning) και εφαρμόζεται η τεχνική του early stopping, κατά την οποία η εκπαίδευση διακόπτεται, όταν δεν παρατηρείται βελτίωση της απόδοσης.

Το σύνολο αξιολόγησης χρησιμοποιείται για την τελική εξέταση του μοντέλου μετά το πέρας της εκπαίδευσης, και κατόπιν της επιλογής βέλτιστων υπερπαραμέτρων. Απολύτως ανεξάρτητα από τα σύνολα εκπαίδευσης και επικύρωσης, παρέχει μια αντικειμενική εκτίμηση της απόδοσης σε άγνωστα δεδομένα, εξασφαλίζοντας την ικανότητα γενίκευσης του μοντέλου.

6.2.2 Αποτροπή Data Leakage

Όπως έχει πολλάκις αναφερθεί κατά τις ενότητες προεπεξεργασίας, είναι απαραίτητη η αποφυγή του φαινομένου data leakage, δηλαδή της ακούσιας έκθεσης δεδομένων εκπαίδευσης στις διαδικασίες επικύρωσης και αξιολόγησης. Συνεπακολούθως, τα βήματα προεπεξεργασίας προσαρμόζονται κατάλληλα, όπως εξηγήθηκε διεξοδικά κατά τις σχετικές υποενότητες του κεφαλαίου 8. Η λεπτομέρεια αυτή είναι ζωτικής σημασίας και για τον λόγο αυτόν σημειώνεται στα αντίστοιχα διαγράμματα με έντονο κόκκινο χρώμα.

6.2.3 Ρύθμιση Υπερπαραμέτρων

Η ρύθμιση των υπερπαραμέτρων, ή αλλιώς hyperparameter tuning, αποτελεί ένα στάδιο στη διαδικασία ανάπτυξης μοντέλων βαθιάς μάθησης. Οι υπερπαράμετροι είναι παράμετροι των αλγορίθμων μάθησης που δεν ενημερώνονται κατά τη διάρκεια της εκπαίδευσης του μοντέλου, αλλά καθορίζονται πριν ξεκινήσει η διαδικασία εκμάθησης. Οι υπερπαράμετροι περιλαμβάνουν παραμέτρους όπως ο ρυθμός εκμάθησης (learning rate), ο αριθμός των επιπέδων στο νευρωνικό δίκτυο, ο αριθμός των νευρώνων σε κάθε επίπεδο, το μέγεθος των batches (batch size), καθώς και άλλα μεγέθη, όπως οι παράμετροι κανονικοποίησης και το είδος των συναρτήσεων ενεργοποίησης. Η διαδικασία αυτή έχει εξέχουσα σημασία στα πλαίσια της απόδοσης ταξινόμησης, όμως είναι χρονοβόρα λόγω του υπολογιστικού φορτίου.

Ιδιαίτερη μνεία αξίζει στην περίπτωση του TabNet. Όπως θα επεξηγηθεί παρακάτω, η αρχιτεκτονική αυτή πάσχει από την κατάρα της διαστατικότητας - curse of dimensionality, καθώς παραμετροποιείται από μεγάλο αριθμό μεταβλητών: 14 υπερπαράμετροι για τον ορισμό και άλλες 6 για την εκπαίδευση του μοντέλου. Η ήδη χρονοβόρα διαδικασία της ρύθμισης γίνεται σχεδόν αδύνατη, για τον τεράστιο αριθμό των 20 παραμέτρων.

> Σε κάθε αγωγό, τα πέντε μοντέλα χαρακτηρίζονται από ταυτόσημα βήματα προεπεξεργασίας και υπερπαραμέτρους εκπαίδευσης. Η διαφορά έγκειται στο μέγεθος των μοντέλων. Με τον τρόπο αυτόν διασφαλίζεται η συνέπεια της αξιολόγησης απόδοσης στα πλαίσια της μελέτης του accuracy - memory usage trade-off.

Κεφάλαιο 7

Αποτελέσματα & Εξαγωγή Συμπερασμάτων

7.1 Αξιολόγηση Αποτελεσμάτων Ταξινόμησης

Εν συντομία, επεξηγούνται τα στοιχεία των πινάκων αξιολόγησης. Αρχικά, όλα τα δυικά μοντέλα φέρουν το γράμμα ***b*** (binary), ενώ τα πολυκατηγορικά το γράμμα ***m*** (multiclass). Η ονομασία τους συμπληρώνεται από έναν αύξοντα αριθμό, καθώς και από την αρχιτεκτονική τους προέλευση. Οργανώνονται σε έξι πίνακες, έναν για κάθε αγωγό. Παράλληλα, αξιολογούνται ως προς διάφορες μετρικές, οι οποίες προσαρμόζονται ανάλογα με τη διάταξη. Υπενθυμίζεται ότι υπό την πολυκατηγορική διάταξη, υιοθετείται η τροπολογία weighted-averaging για τις μετρικές precision, recall και F_1-score, αλλά και ότι οι μετρικές FAR & AMR παρουσιάζουν επίσης παραλλαγές διάταξης. Ακόμη, η απώλεια αναφέρεται στην εντροπική συνάρτηση απώλειας, η οποία εδώ νοείται ως μέτρο αξιολόγησης και κατά τα γνωστά προσαρμόζεται επίσης ως προς τη διάταξη.

Η αξιολόγηση των ταξινομητών καταδεικνύει αποτελέσματα **κορυφαίας απόδοσης** (State of the Art - SoA), όπως καταγράφεται στον πίνακα 7.1. Γενικότερα, όταν η ακρίβεια υπερβαίνει το κατώφλι του 98.5%, η απόδοση των ταξινομητών θεωρείται SoA. Ειδικότερα, 29 ταξινομητές επιδεικνύουν accuracy υπερβαίνουσα το κατώφλι του 99%, με την εξαίρεση του πρώτου TabNet δυικού ταξινομητή, ο οποίος έχει accuracy 98.84% και άρα ανήκει επίσης στην κλάση SoA.

Επιθεωρώντας τα αποτελέσματα, παρατηρείται ότι τα μοντέλα έχουν διαταχθεί κατά αύξουσα ως προς το μέγεθος σειρά. Συγκεκριμένα, όπως φαίνεται και στον πίνακα καταγραφής του αριθμού παραμέτρων, τα μοντέλα παρουσιάζουν μη γραμμική αύξηση του πλήθους παραμέτρων, και άρα της κατανάλωσης μνήμης. Το πρώτο μοντέλο του κάθε αγωγού επονομάζεται ως «ελάχιστο» (minimal), ενώ το τελευταίο θα είναι το «μέγιστο» (maximal). Στην πλειονότητα των περιπτώσεων, η εν λόγω αύξουσα σειρά παραμέτρων αναπαράγεται και στις μετρικές απόδοσης, δηλαδή το ελάχιστο μοντέλο παρουσιάζει την ελάχιστη απόδοση, η οποία σειριακά αυξάνεται μέχρι το μέγιστο μοντέλο, μέγιστης απόδοσης. Στο σημείο αυτό, υπενθυμίζεται ότι η για τις μετρικές accuracy, precision, recall και F_1-score ζητούμενη είναι η μεγιστοποίηση τους, ενώ αντίθετα απαιτείται η ελαχιστοποίηση των loss, FAR και AMR.

Η αύξηση της απόδοσης εντοπίζεται σε ένα φραγμένο εύρος τιμών. Καθολικά, η μέγιστη προσαύξηση της accuracy είναι μικρότερη από το κατώφλι του 0.7%, όπως διαπιστώνεται από την περίπτωση του δυικού TabNet, ενώ στις περισσότερες περιπτώσεις το εύρος τιμών δεν παρουσιάζει διακύμανση μεγαλύτερη του 0.3%. Παράλληλα, ο αριθμός παραμέτρων των μοντέλων αυξάνεται δραματικά. Σύμφωνα με τον πίνακα 7.2, ο αριθμός των παραμέτρων αυξάνεται από μερικές εκατοντάδες εώς και δεκάδες χιλιάδες, ενώ στην ειδική περίπτωση του TabNet από δεκάδες χιλιάδες σε εκατοντάδες χιλιάδες παραμέτρους. Επομένως, η σχέση της απόδοσης ταξινόμησης είναι εξαιρετικά δυσανάλογη με το μέγεθος των μοντέλων.

Πίνακας 7.1: Αξιολόγηση της απόδοσης ταξινόμησης

Binary Models	Loss	Accuracy	Precision	Recall	F_1-score	FAR	AMR
clf_1_b_mlp	0.0341	99.07%	96.59%	97.25%	96.92%	0.0061	0.0275
clf_2_b_mlp	0.0225	99.48%	98.36%	98.16%	98.26%	0.0029	0.0184
clf_3_b_mlp	0.0191	99.53%	98.44%	98.42%	98.43%	0.0028	0.0158
clf_4_b_mlp	0.0152	99.56%	98.37%	98.74%	98.55%	0.0029	0.0126
clf_5_b_mlp	0.0119	99.64%	98.59%	99.03%	98.81%	0.0025	0.0097

Multiclass Models	Loss	Accuracy	Precision	Recall	F_1-score	FAR	AMR
clf_1_m_mlp	0.0263	99.30%	99.28%	99.30%	99.19%	0.0020	0.0324
clf_2_m_mlp	0.0227	99.43%	99.43%	99.43%	99.34%	0.0018	0.0263
clf_3_m_mlp	0.0203	99.48%	99.48%	99.48%	99.43%	0.0016	0.0241
clf_4_m_mlp	0.0187	99.51%	99.50%	99.51%	99.45%	0.0016	0.0222
clf_5_m_mlp	0.0146	99.59%	99.59%	99.59%	99.54%	0.0011	0.0200

Binary Models	Loss	Accuracy	Precision	Recall	F_1-score	FAR	AMR
clf_1_b_cnn	0.0122	99.64%	98.56%	99.05%	98.80	0.0026	0.0095
clf_2_b_cnn	0.0106	99.71%	98.94%	99.17%	99.06%	0.0019	0.0083
clf_3_b_cnn	0.0070	99.77%	99.12%	99.39%	99.25%	0.0016	0.0061
clf_4_b_cnn	0.0065	99.79%	99.04%	99.57%	99.31%	0.0017	0.0043
clf_5_b_cnn	0.0055	99.81%	99.15%	99.60%	99.37%	0.0015	0.0040

Multiclass Models	Loss	Accuracy	Precision	Recall	F_1-score	FAR	AMR
clf_1_m_cnn	0.0212	99.51%	99.51%	99.51%	99.45%	0.0016	0.0204
clf_2_m_cnn	0.0138	99.64%	99.62%	99.64%	99.62%	0.0014	0.0127
clf_3_m_cnn	0.0108	99.72%	99.73%	99.72%	99.70%	0.0011	0.0098
clf_4_m_cnn	0.0079	99.77%	99.78%	99.77%	99.75%	0.0008	0.0084
clf_5_m_cnn	0.0064	99.80%	99.81%	99.80%	99.78%	0.0006	0.0076

Binary Models	Loss	Accuracy	Precision	Recall	F_1-score	FAR	AMR
clf_1_b_TNet	0.03854	98.84%	97.50%	94.73%	96.09%	0.0043	0.0527
clf_2_b_TNet	0.02111	99.35%	98.67%	97.00%	97.83%	0.0023	0.0300
clf_3_b_TNet	0.02044	99.36%	98.84%	96.88%	97.85%	0.0020	0.0312
clf_4_b_TNet	0.01817	99.42%	98.96%	97.20%	98.07%	0.0018	0.0280
clf_5_b_TNet	0.01431	99.53%	98.97%	97.92%	98.45%	0.0018	0.0208

Multiclass Models	Loss	Accuracy	Precision	Recall	F_1-score	FAR	AMR
clf_1_m_TNet	0.02700	99.30%	99.30%	99.30%	99.23%	0.0030	0.0264
clf_2_m_TNet	0.02070	99.45%	99.39%	99.45%	99.39%	0.0018	0.0240
clf_3_m_TNet	0.01926	99.46%	99.45%	99.46%	99.41%	0.0019	0.0223
clf_4_m_TNet	0.01935	99.47%	99.48%	99.47%	99.45%	0.0022	0.0194
clf_5_m_TNet	0.01605	99.58%	99.56%	99.58%	99.56%	0.0017	0.0153

Παρουσιάζονται τα αποτελέσματα της απόδοσης ταξινόμησης. Υπενθυμίζεται, ότι όλα τα αποτελέσματα λαμβάνονται επί του **συνόλου αξιολόγησης** (test set) και ότι η κάθε μία από τις επτά παρούσες μετρικές αξιολόγησης προσαρμόζεται κατάλληλα ως προς την διάταξη του ταξινομητή (δυαδική ή πολυκατηγορική μορφή).

Πίνακας 7.2: Αριθμοί παραμέτρων μοντέλων

Total Parameters	Model-1	Model-2	Model-3	Model-4	Model-5
MLPs-Binary	217	521	1,385	7,529	49,321
MLPs-Multiclass	863	1,135	1,407	1,679	5,391
CNNs-Binary	673	1,265	4,961	9,601	74,561
CNNs-Multiclass	799	1,503	5,199	10,063	39,247
TabNet-Binary	60,418	159,523	305,428	572,768	685,993
TabNet-Multiclass	88,578	121,678	232,074	262,542	294,882

Παρουσιάζεται ο αντίστοιχος αριθμός παραμέτρων όλων των μοντέλων, προκειμένου να μελετηθεί η δυνατότητα κλιμάκωσης των ταξινομητών.

Έπειτα τίθεται η ανάγκη σύγκρισης των μοντέλων. Η επιλογή του βέλτιστου μοντέλου θα πρέπει να πραγματοποιηθεί ως προς δύο άξονες, την απόδοση ταξινόμησης και το μέγεθος. Ως προς την απόδοση η συζήτηση θα εστιάσει στην μετρική accuracy, η οποία εξ υποθέσεως θα θεωρηθεί η αντιπροσωπευτική μετρική αξιολόγησης, παρά την ανισομέρεια του συνόλου δεδομένων. Αφετέρου, το μέγεθος αναφέρεται στον αριθμό παραμέτρων, δηλαδή την κατανάλωση μνήμης. Επομένως, τίθεται το ερώτημα:

> Δεδομένων των περιορισμών μνήμης, ποιο μοντέλο συνιστάται για την ανάπτυξη ενός IDS·

Επομένως, τίθεται η ανάγκη επιπρόσθετης μελέτης του **συμβιβασμού** (trade-off) μεταξύ της accuracy και του αριθμού παραμέτρων, στα πλαίσια σύγκρισης των ταξινομητών. Παράλληλα, εξετάζονται εγγενείς **νόμοι κλιμάκωσης** της απόδοσης (scaling laws) για κάθε αγωγό ταξινομητών, αντικείμενο μελέτης σύνηθες στη βαθιά μάθηση και ιδιαίτερα στον τομέα του NLP, [61, 62].

7.2 Συμπερασματολογία

Η ανάπτυξη ταξινομητών κορυφαίας απόδοσης, συνεπάγεται την επιτυχή ανίχνευση εισβολών.

> Ως εκ τούτου, το αίτημα ανάπτυξης ταξινομητή στα πλαίσια ενός συστήματος ανίχνευσης εισβολών **έχει επιτευχθεί**.

Αυτό είναι το βασικό συμπέρασμα της παρούσας εργασίας. Ακολουθούν δευτερεύοντα συμπεράσματα, καθώς και μία μελέτη της κλιμάκωσης των μοντέλων.

7.2.1 Συμπεράσματα - Α΄ Μέρος

Από την αξιολόγηση των αποτελεσμάτων ταξινόμησης, εξάγεται μια σειρά από συμπεράσματα. Πρώτον, όπως υπογραμμίστηκε και παραπάνω, το πρωταρχικό αίτημα της ανάπτυξης ταξινομητή στα πλαίσια ενός Συστήματος Ανίχνευσης Εισβολών έχει επιτευχθεί. Επιπροσθέτως:

- Δυνητικά, **όλα τα μοντέλα θα μπορούσαν να αναλάβουν το έργο της ταξινόμησης** δικτυακής κίνησης, αποτελώντας βασικό πυλώνα του IDS.

Συνεπώς, άγονται τα συμπεράσματα:

- Πρώτον, **αναδεικνύεται η ισχύς και η απόδοση του Quantile Transformer**, ο οποίος απαντάται σε όλους τους αγωγούς προεπεξεργασίας και αποτελεί την βέλτιστη τακτική διαχείρισης των δεδομένων δικτυακής κίνησης, τα οποία προέρχονται από εξαιρετικά ασύμμετρες κατανομές και πάσχουν από την παρουσία πληθώρας σημείων-**outliers**.

- Δεύτερον, στα πλαίσια της ανάπτυξης των MLPs, **καταδεικνύεται η κυριαρχία του αλγόριθμου XGBoost**, μέσω του οποίου πραγματοποιείται επιλογή χαρακτηριστικών και **η διαστατικότητα του χώρου εισόδου μειώνεται σημαντικά**, γεγονός που συντελεί στην ανάπτυξη μοντέλων ελάχιστης μνήμης, ενώ παράλληλα διατηρείται η SoA απόδοση.

- Η **αύξηση του μεγέθους συνεπάγεται εξαιρετικά μικρή βελτίωση της απόδοσης**, καθώς η διακύμανση δεκάδων χιλιάδων παραμέτρων αντιστοιχεί σε εύρος accuracy 0.7%.

- Γενικεύοντας, συμπεραίνεται η σημασία της κατάλληλης προεπεξεργασίας στα πλαίσια της βαθιάς μάθησης. Ο σχεδιασμός και η υλοποίηση των αγωγών προεπεξεργασίας υπήρξε αδιαμφισβήτητα το πλεόν χρονοβόρο και απαιτητικό σκέλος κατά την εκπόνηση αυτής της εργασίας. **Η επίτευξη των SoA αποτελεσμάτων ανάγεται στην ανάπτυξη κατάλληλων διαδικασιών προεπεξεργασίας**.

Τέλος, **διακηρύσσεται η ευρύτερη επιτυχία της βαθιάς μάθησης στη διεργασία ανίχνευσης εισβολών**. Για άλλη μία φορά, αποδεικνύεται η αποτελεσματικότητα της βαθιάς μάθησης στην αυτοματοποίηση εργασιών και στη διαχείριση μαζικών δεδομένων. Στο πλαίσιο της πληροφορικής δικτύων, οι αυξανόμενες και διαρκώς αναπτυσσόμενες εφαρμογές της τεχνητής νοημοσύνης αναμένεται να συνεχίσουν να επηρεάζουν θετικά την απόδοση και την ασφάλεια των συστημάτων.

7.2.2 Σύγκριση μοντέλων

Ποιο μοντέλο θα πρέπει να προτιμηθεί για τη υλοποίηση ενός IDS;

Η απάντηση προέρχεται από την αρχιτεκτονική των CNNs, η οποία εμφανίζει την βέλτιστη απόδοση ταξινόμησης. Ακόμη και αν οι αριθμοί παραμέτρων υπερβαίνουν την περίπτωση των MLPs, της αρχιτεκτονικής με την ελάχιστη κατανάλωση μνήμης, η διαφορά δεν είναι σημαντική. Ειδικότερα, ανάμεσα στα CNNs, συνιστάται η πολυκατηγορική μορφολογία των ταξινομητών. Ενώ οι δυαδικοί ταξινομητές εμφανίζουν την υψηλότερη accuracy και ταυτόχρονα τον μικρότερο αριθμό παραμέτρων, η διαφορά ακρίβειας εύρους δεκαδικών ψηφίων δεν υπερβαίνει τα ουσιαστικά πλεονεκτήματα των multiclass μοντέλων.

Συγκεκριμένα, η δυνατότητα ταυτοποίησης της κλάσης επίθεσης έχει εξέχουσα σημασία στα πλαίσια της κυβερνοασφάλειας, διότι δίνει τη δυνατότητα λήψης αποφάσεων και εκτέλεσης ενεργειών καταστολής, ειδικά προσαρμοσμένων στον τύπο επίθεσης. Συγκεκριμένα, το IDS μπορεί να αναλάβει περαιτέρω λειτουργικότητες πέρα από την έκδοση warnings πιθανής κακόβουλης δραστηριότητας, όπως την ενεργοποίηση εξεζητημένων αμυντικών μηχανισμών που θα στοχεύουν προς την αντιμετώπιση του συγκεκριμένου τύπου κακόβουλης δραστηριότητας. Επομένως, οι πολυκατηγορικοί ταξινομητές μπορούν να επεκτείνουν το IDS σε ένα αποτελεσματικό IPS (Intrusion Prevention System), με τη δυνατότητα προσαρμοστικής αντιμετώπισης. Ωστόσο, πέρα από τη σύνδεση με μηχανισμούς καταστολής, η γνώση του τύπου επίθεσης έχει καίρια σημασία στην ανάλυση, στη συλλογή πληροφοριών και στον σχεδιασμό στρατηγικών πρόληψης κυβερνοασφάλειας.

Επομένως, συστήνεται ο αγωγός των πολυκατηγορικών CNNs και απομένει ο προσδιορισμός του τελικού ταξινομητή. Προφανώς, βέλτιστη επιλογή θα είναι το maximal μοντέλο

μέγιστης κατανάλωσης μνήμης, ακομή και αν επάγεται η επαύξηση του υπολογιστικού φορτίου. Εντούτις, η τελική επιλογή επαφίεται στην ομάδα μηχανικών λογισμικού, η οποία θα αναλάβει την υλοποίηση του IDS. Βασική συνιστώσα του λογισμικού θα είναι ο ταξινομητής, η επιλογή του οποίου ίσως πραγματοποιηθεί με βάση δευτερεύοντες παράγοντες, όπως για παράδειγμα τη διαθεσιμότητα υπολογιστικών πόρων και μνήμης, ή τις εξειδικευμένες ανάγκες υλοποίησης. Παρ΄ όλα αυτά, επαναδιατυπώνεται η πρόταση του maximal μοντέλου, καθώς η μέγιστη απόδοση έχει ζωτική σημασία στην κυβερνοασφάλεια.

Περα από τη σχετική σύγκριση των μοντέλων, μελετάται και το trade-off μεταξύ ακρίβειας και μεγέθους, που αποτελεί συμπληρωματικό στόχο της παρούσας εργασίας. Ιδιαίτερα, αναδεικνύεται μία επαναλαμβανόμενη συμπεριφορά (pattern) στους 6 αγωγούς. Τα μοντέλα καταδεικνύουν αύξουσα accuracy κατά την επαύξηση των παραμέτρων τους. Η αύξηση αυτή όμως φαίνεται να παρουσιάζει φθίνοντα ρυθμό. Εικάζεται ότι η παρατηρούμενη καμπύλη συμπεριφοράς προτυποποιείται από την παρακάτω παραβολική εξίσωση:

$$(C) : -\alpha t^2 + \beta t - \gamma = 0, \text{ όπου: } a, \beta, \gamma > 0$$

Τότε, το μέγιστο σημείο απόδοσης θα προσδιορίζεται στο σημείο $\frac{\beta}{2\alpha}$. Η προτυποποίηση δεν δύναται να αποτελέσει κριτήριο σύγκρισης των αγωγών, καθώς η δειγματοληψία του μεγέθους δεν πραγματοποιήθηκε ως προς ίσα διαστήματα. Επιπροσθέτως, τα εγγενή χαρακτηριστικά και οι περιορισμοί των μοντέλων δεν επιτρέπουν την άνευ όρων δημιουργία μοντέλων διαφορετικής αρχιτεκτονικής με ίσο αριθμών παραμέτρων. Συνεπώς, η ανάλυση του trade-off θα πραγματοποιηθεί ποιοτικώς. Αρχικά, η υπέρβαση του σημείου μεγίστου ερμηνεύεται ως υπέρβαση ενός κατωφλίου υπερεκπαίδευσης, έπειτα από το οποίο η προσαύξηση παραμέτρων όχι μόνο δεν συμβάλλει σε αύξηση της απόδοσης, αλλά μάλιστα οδηγεί και σε μείωση υπό φθίνοντα ρυθμό.

Ακολουθεί μία σύντομη σύνοψη των θεωρητικών αιτίων που καθορίζουν την υπό μελέτη καμπύλη απόδοσης, καθώς και μία περίληψη της κατάρας της διαστατικότητας.

7.2.2.1 Θεωρητικό Υπόβαθρο

Η προαναφερθείσα δυναμική αντιστάθμισης υποκρύπτει μια θεωρητικώς αναμενόμενη καμπύλη, η οποία πρωταρχικά ανάγεται στον θεωρητικό **συμβιβασμό μεταξύ μεροληψίας και διακύμανσης** (bias-variance trade-off), [2, 8], μία θεμελιώδη σχέση της στατιστικής μάθησης [22], όπου η μεροληψία μειώνεται και η απόκλιση αυξάνεται κατά την πολυπλοκότητα του μοντέλου, επηρεάζοντας την ακρίβεια. Αφετέρου, πρόσφατες μελέτες από το πεδίο του NLP [61, 62] αναδεικνύουν εμπειρικούς νόμους κλιμάκωσης των μοντέλων, ποσοτικοποιώντας την κλιμάκωση της απόδοσης ως προς το μέγεθος του μοντέλου. Προφανώς, υποκείμενοι νόμοι κλιμάκωσης περιγράφουν και τις διεργασίας ταξινόμησης. Τέλος, το φαινόμενο έχει εξερευνηθεί και επικυρωθεί μέσω αριθμητικών πειραμάτων, [63].

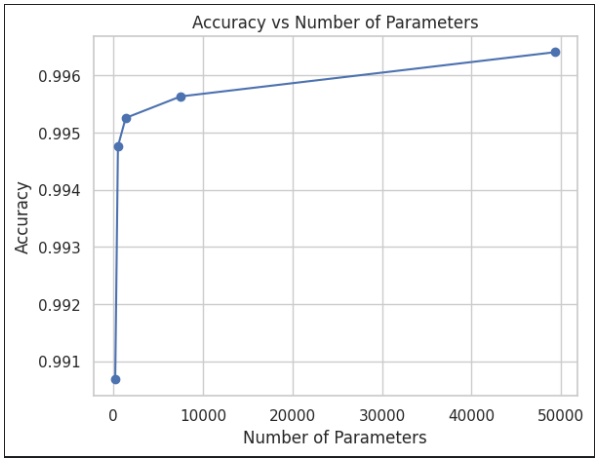
7.2.2.2 Κατάρα της Διαστατικότητας

Η κατάρα της διαστατικότητας (curse of dimensionality) αναφέρεται στις διάφορες δυσκολίες και προκλήσεις που προκύπτουν κατά την ανάλυση δεδομένων σε χώρους υψηλών διαστάσεων. Καθώς ο αριθμός των διαστάσεων αυξάνεται, ο όγκος του χώρου των χαρακτηριστικών αυξάνεται εκθετικά, με αποτέλεσμα τα δεδομένα να γίνονται αραιά. Συνεπακολούθως, οι αποστάσεις μεταξύ των σημείων αυξάνονται, η ομοιότητα μεταξύ των δεδομένων προσδιορίζεται δυσκολότερα και η αποτελεσματικότητα των μοντέλων μειώνεται. Το φαινόμενο επηρεάζει επίσης την ικανότητα γενίκευσης των μοντέλων, οδηγώντας σε υπερπροσαρμογή, καθώς αφομοιώνεται ο θόρυβος των δεδομένων εκπαίδευσης αντί για τις υποκείμενες σχέσεις των χαρακτηριστικών. Παράλληλα, τα μοντέλα πολλών υπερπαραμέτρων απαιτούν μεγαλύτερα σύνολα δεδομένων για να εκπαιδευτούν επαρκώς. Σε αυτήν την κατηγορία εμπίπτει η περίπτωση του TabNet, το οποίο

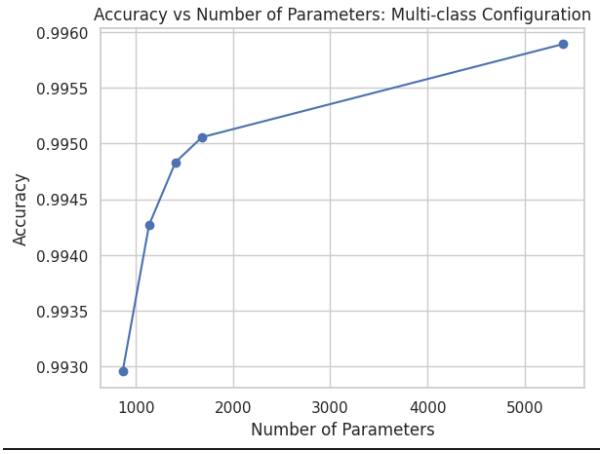
καθορίζεται από 13 υπερπαραμέτρους αρχιτεκτονικής και επιπλέον 5 εκπαίδευσης. Εν τέλει, το πλήθος των υπερπαραμέτρων απαιτεί επαυξημένο σύνολο εκπαίδευσης προκειμένου να επιτευχθεί αποτελεσματική προσαρμογή και εκπαίδευση. Καθώς το σύνολο εκπαίδευσης δεν είναι δυνατόν να αυξηθεί, θα παρατηρηθούν έντονα φαινόμενα υπερπροσαρμογής.

Σχήμα 7.1: Trade-off μεταξύ accuracy και αριθμού παραμέτρων

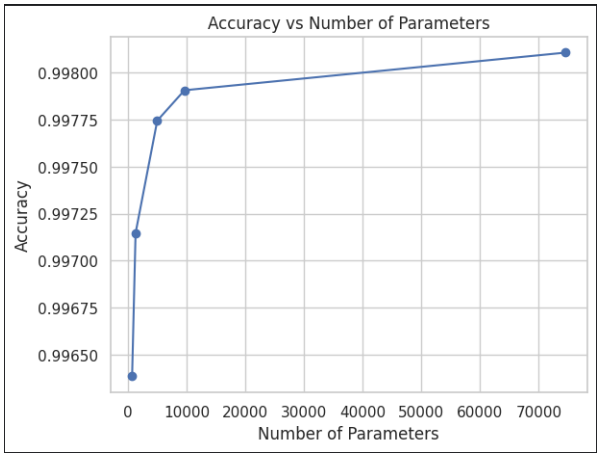
(α΄) δυικά MLPs: Trade-off



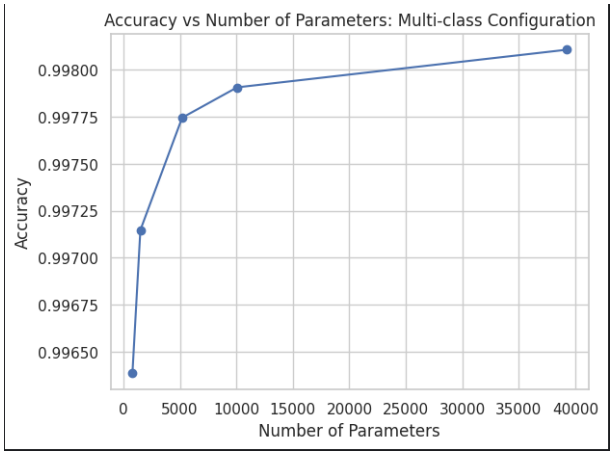
(β΄) πολυκατηγορικά MLPs: Trade-off



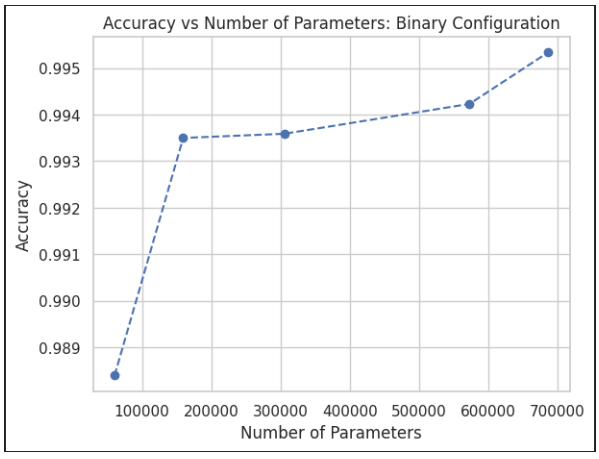
(γ΄) δυικά CNNs: Trade-off



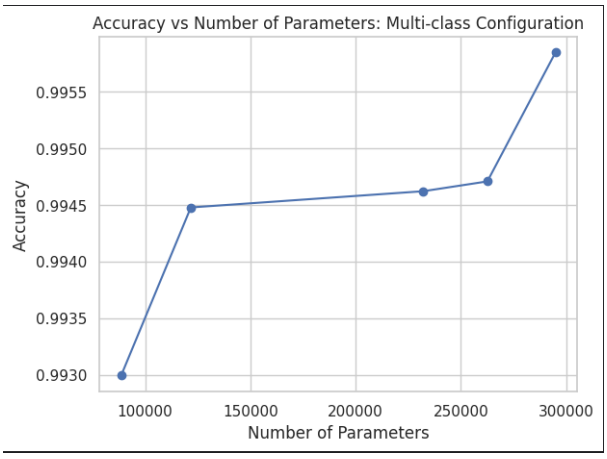
(δ΄) πολυκατηγορικά CNNs: Trade-off



(ε΄) δυικά TabNets: Trade-off



(ϛ΄) πολυκατηγορικά TabNets: Trade-off



7.2.3 Συμπεράσματα - Β΄ Μέρος

Σε σχέση με τη διερεύνηση του συμβιβασμού μεταξύ της ακρίβειας και της κατανάλωσης μνήμης, άγονται τα εξής συμπεράσματα:

- Υπό το κριτήριο της βέλτιστης ακρίβειας, **επιλέγονται τα CNNs**, ενώ για την ελάχιστη χρήση μνήμης προτιμούνται τα MLPs.
- Απορρίπτεται η περίπτωση του TabNet, καθώς παρουσιάζει το χειρότερο trade-off.
- Ακόμη, οι τέσσερις αγωγοί των MLPs και των CNNs **παρουσιάζουν τη θεωρητικώς αναμενόμενη συμπεριφορά κλιμάκωσης απόδοσης**.
- Τέλος, και στις δύο περιπτώσεις της αρχιτεκτονικής TabNet, παρατηρείται **απόκλιση από τη θεωρητικώς αναμενόμενη συμπεριφορά**, η οποία αποδίδεται στην **κατάρα της διαστατικότητας**.

Σημειώνεται ότι η συμπεριφορά κλιμάκωσης του TabNet αναγνωρίζεται στη βιβλιογραφία [63] και ότι το παρατηρηθέν ποιοτικό πρότυπο του συμβιβασμού έχει καταγραφεί σε μοντέλα NLP.

7.3 Μελλοντικές Κατευθύνσεις

Στο σημείο αυτό, ολοκληρώνεται επιτυχώς η παρούσα εργασία. Φυσικά, το ερευνητικό αντικείμενο θα πρέπει να μελετηθεί περαιτέρω. Προτείνεται ο πειραματισμός με άλλες τεχνικές βαθιάς μάθησης, προκειμένου να αξιολογηθεί η απόδοσή τους σε σχέση με τις παρούσες αρχιτεκτονικές. Επιπλέον σε επίπεδο λογισμικού, τα ευρήματα της έρευνας θα μπορούσαν να ενσωματωθούν σε ολοκληρωμένα συστήματα, κατάλληλα για εμπορική χρήση. Τέλος, για λόγους εγκυρότητας των αποτελεσμάτων ενδείκνυται ο πειραματισμός με άλλα σύνολα δεδομένων πέρα από το CIC-IDS-2017. Οι εν λόγω κατευθύνσεις θα μπορούσαν να προσφέρουν πολύτιμες γνώσεις και να οδηγήσουν σε περαιτέρω βελτιώσεις και καινοτομίες στην ανάπτυξη συστημάτων ανίχνευσης εισβολών με χρήση βαθιάς μάθησης.

Επίσης, η έρευνα μπορεί να συνεχιστεί σε μελλοντικό χρόνο, με στόχο τη δημιουργία ενός υβριδικού IDS, στα πλαίσια προστασίας ενός IoT περιβάλλοντος (Internet of Things), π.χ. ένα smart home. Προτείνεται η δημιουργία ενός υβριδικού IDS, το οποίο θα περιλαμβάνει μια κεντρική δικτυακή συνιστώσα NIDS, καθώς και επί μέρους HIDSs προσαρμοσμένα κατάλληλα σε κάθε συσκευή. Ζητούμενη είναι η συνέργεια NIDS-HIDSs για την αποτελεσματικότερη ανίχνευση εισβολών: Ένα HIDS μπορεί να εκτελεί δυαδική ταξινόμηση της δικτυακής ροής σε επίπεδο συσκευής, ενώ παράλληλα ένας κεντρικός κόμβος (hub) να αποφαίνεται τον πολυκατηγορικό τύπο κυβερνοαπειλής, σε περίπτωση κακόβουλης δραστηριότητας. Για τον λόγο αυτό, θα πρέπει να αναπτυχθούν HIDSs τα οποία θα είναι συμβατά με το παρόν NIDS.

Παράρτημα Α΄

Υλοποίηση των MLPs & CNNs

Παρατίθενται οι αρχιτεκτονικές των μοντέλων MLPs & CNNs, διατεταγμένες σε αύξουσα κατά μέγεθος σειρά. Υιοθετείται αυτούσια η python μορφή του κώδικα, όπως γράφτηκε υπό τις βιβλιοθήκες keras και tensorflow.

Η συνάρτηση ενεργοποίησης SELU Για λόγους πληρότητας, παρατίθεται η συνάρτηση ενεργοποίησης SELU (Scaled Exponential Linear Unit), [64]. Αποτελεί μια παραλλαγή της ReLU και χρησιμοποιείται σε βαθιά νευρωνικά δίκτυα, προσφέροντας σταθερότητα και ταχύτερη εκπαίδευση. Η SELU έχει self - normalizing ιδιότητες (αυτοκανονικοποίηση), δηλαδή διατηρεί τις τιμές των ενεργοποιήσεων σε ένα συγκεκριμένο εύρος, αποτρέποντας προβλημάτα εξαφάνισης και έκρηξης των παραγώγων. Η συνάρτηση ορίζεται ως:

$$\text{SELU}\,(x) = \lambda \begin{cases} x, & x > 0 \\ \alpha \cdot e^x - \alpha, & x \leq 0 \end{cases}$$

όπου $\lambda \approx 1.0507$ και $\alpha \approx 1.6733$. Οι τιμές αυτές προέρχονται από τη θεωρητική ανάλυση και τα πειραματικά αποτελέσματα που παρουσιάστηκαν στο ερευνητικό άρθρο "Self-Normalizing Neural Networks", 2017 από τον Klaumbauer et al., [64]. Οι σταθερές τιμές επιλέχθηκαν έτσι ώστε η μέση τιμή και η διασπορά των ενεργοποιήσεων να παραμένουν σταθερές κατά τη διάρκεια της προώθησης του σήματος στα δικτυακά επίπεδα, διασφαλίζοντας την αυτοκανονικοποίηση, δηλαδή τη διατήρηση της μέσης τιμής κοντά στο μηδέν και της διασποράς κοντά στη μονάδα. Εδώ, η επιλογή της selu πραγματοποιείται υπό όρους hyperparameter tuning και τελικά διαλέγεται, καθώς επιδεικνύει ανώτερη απόδοση ταξινόμησης.

Α΄.1 Δυικά MLPs

Υπενθυμίζεται ότι το πλήθος των επιπέδων, καθώς και ο αντίστοιχος αριθμός νευρώνων αποτελούν υπερπαραμέτρους της αρχιτεκτονικής. Σημειώνεται επίσης ότι η αύξηση των μεταβλητών του μεγέθους συντελείται με όρους γεωμετρικής προόδου. Τέλος, είναι καταφανής η χρήση αρχικοποίησης βαρών. Επιλέγεται η Glorot αρχικοποίηση για τη σιγμοειδή και την υπερβολική εφαπτομένη, ενώ επιλέγεται η αρχικοποίηση He για τη SELU, καθώς αυτή θεωρείται παραλλαγή της ReLU.

Επεξήγηση Κώδικα Η έννοια Dense αντιστοιχεί σε πλήρως συνδεδεμένα επίπεδα, ενώ ο αριθμός που ακολουθεί ισούται με τον αριθμό των νευρώνων του εν λόγω επιπέδου. Η μεταβλητή input_size αντιστοιχεί στη διαστατικότητα του χώρου εισόδου των προεπεξεργασμένων

χαρακτηριστικών, και εν προκειμένω λαμβάνει την τιμή 20. Η παραλλαγή των μοντέλων τελείται με όρους συνέπειας ως προς τα στοιχεία της αρχιτεκτονικής, με εξαίρεση τις κλιμακούμενες υπερπαραμέτρους μεγέθους.

```
model_1_b_mlp = Sequential([
  Input(shape=(input_shape,)),
  Dense(8, activation='tanh', kernel_initializer=GlorotUniform()),
  Dense(4, activation='selu', kernel_initializer=HeUniform()),
  Dense(2, activation='selu', kernel_initializer=HeUniform()),
  Dense(1, activation='sigmoid', kernel_initializer=GlorotUniform())])

model_2_b_mlp = Sequential([
  Input(shape=(input_shape,)),
  Dense(16, activation='tanh', kernel_initializer=GlorotUniform()),
  Dense(8, activation='selu', kernel_initializer=HeUniform()),
  Dense(4, activation='selu', kernel_initializer=HeUniform()),
  Dense(2, activation='selu', kernel_initializer=HeUniform()),
  Dense(1, activation='sigmoid', kernel_initializer=GlorotUniform())])

model_3_b_mlp = Sequential([
  Input(shape=(input_shape,)),
  Dense(32, activation='tanh', kernel_initializer=GlorotUniform()),
  Dense(16, activation='selu', kernel_initializer=HeUniform()),
  Dense(8,  activation='selu', kernel_initializer=HeUniform()),
  Dense(4,  activation='selu', kernel_initializer=HeUniform()),
  Dense(2,  activation='selu', kernel_initializer=HeUniform()),
  Dense(1,  activation='sigmoid', kernel_initializer=GlorotUniform())])

model_4_b_mlp = Sequential([
  Input(shape=(input_shape,)),
  Dense(64, activation='tanh', kernel_initializer=GlorotUniform()),
  Dense(32, activation='selu', kernel_initializer=HeUniform()),
  Dense(16, activation='selu', kernel_initializer=HeUniform()),
  Dense(8, activation='selu', kernel_initializer=HeUniform()),
  Dense(4, activation='selu', kernel_initializer=HeUniform()),
  Dense(2, activation='selu', kernel_initializer=HeUniform()),
  Dense(1, activation='sigmoid', kernel_initializer=GlorotUniform())])

model_5_b_mlp = Sequential([
  Input(shape=(input_shape,)),
  Dense(128, activation='tanh', kernel_initializer=GlorotUniform()),
  Dense(64, activation='selu', kernel_initializer=HeUniform()),
  Dense(32, activation='selu', kernel_initializer=HeUniform()),
  Dense(16, activation='selu', kernel_initializer=HeUniform()),
  Dense(8, activation='selu', kernel_initializer=HeUniform()),
  Dense(4, activation='selu', kernel_initializer=HeUniform()),
  Dense(2, activation='selu', kernel_initializer=HeUniform()),
  Dense(1, activation='sigmoid', kernel_initializer=GlorotUniform())])
```

Α΄.2 Πολυκατηγορικά MLPs

Επεξήγηση Κώδικα Παρατίθενται οι αρχιτεκτονικές των πέντε μοντέλων MLPs,

διατεταγμένες σε αύξουσα κατά μέγεθος σειρά. Η αύξηση πραγματοποιείται με όρους αριθμητικής προόδου. Όλα τα προηγούμενα σχόλια θα ισχύουν και για την πολυκατηγορική διάταξη, αφού η διάσταση εισόδου θα είναι και πάλι 20. Παρατηρείται η θεωρητικώς αναμενόμενη διαφοροποίηση στο επίπεδο εξόδου, και συγκεκριμένα η χρήση της softmax αλλά και οι πολλαπλοί νευρώνες. Συγκεκριμένα, θα είναι 15 το πλήθος (num_classes=15), αφού υπάρχουν 14 attack scenarios, συν η benign traffic.

```
model_1_m_mlp = Sequential([
  Input(shape=(input_shape,)),
  Dense(16, activation='tanh', kernel_initializer=GlorotUniform()),
  Dense(16, activation='selu', kernel_initializer=HeUniform()),
  Dense(num_classes, activation='softmax', kernel_initializer=GlorotUniform())])

model_2_m_mlp = Sequential([
  Input(shape=(input_shape,)),
  Dense(16, activation='tanh', kernel_initializer=GlorotUniform()),
  Dense(16, activation='selu', kernel_initializer=HeUniform()),
  Dense(16, activation='selu', kernel_initializer=HeUniform()),
  Dense(num_classes, activation='softmax', kernel_initializer=GlorotUniform())])

model_3_m_mlp = Sequential([
  Input(shape=(input_shape,)),
  Dense(16, activation='tanh', kernel_initializer=GlorotUniform()),
  Dense(16, activation='selu', kernel_initializer=HeUniform()),
  Dense(16, activation='selu', kernel_initializer=HeUniform()),
  Dense(16, activation='selu', kernel_initializer=HeUniform()),
  Dense(num_classes, activation='softmax', kernel_initializer=GlorotUniform())])

model_4_m_mlp = Sequential([
  Input(shape=(input_shape,)),
  Dense(16, activation='tanh', kernel_initializer=GlorotUniform()),
  Dense(16, activation='selu', kernel_initializer=HeUniform()),
  Dense(16, activation='selu', kernel_initializer=HeUniform()),
  Dense(16, activation='selu', kernel_initializer=HeUniform()),
  Dense(16, activation='selu', kernel_initializer=HeUniform()),
  Dense(num_classes, activation='softmax', kernel_initializer=GlorotUniform())])

model_5_m_mlp = Sequential([
  Input(shape=(input_shape,)),
  Dense(32, activation='tanh', kernel_initializer=GlorotUniform()),
  Dense(32, activation='selu', kernel_initializer=HeUniform()),
  Dense(32, activation='selu', kernel_initializer=HeUniform()),
  Dense(32, activation='selu', kernel_initializer=HeUniform()),
  Dense(32, activation='selu', kernel_initializer=HeUniform()),
  Dense(num_classes, activation='softmax', kernel_initializer=GlorotUniform())])
```

Α′.3 Δυικά CNNs

```
model_1_b_cnn = Sequential([
  Input(shape=(9, 9, 1)),
  Conv2D(8, kernel_size=(3, 3), activation='relu', kernel_initializer=HeNormal()),
```

```
  MaxPooling2D(pool_size=(2, 2)),
  Flatten(),
  Dense(8, activation = 'relu', kernel_initializer = HeUniform(),
  Dense(1, activation = 'sigmoid', kernel_initializer = GlorotUniform())
])

model_2_b_cnn = Sequential([
  Input(shape=(9, 9, 1)),
  Conv2D(8, kernel_size=(3, 3), activation='relu', kernel_initializer=HeNormal()),
  MaxPooling2D(pool_size=(2, 2)),
  Flatten(),
  Dense(16, activation = 'relu', kernel_initializer = HeUniform(),
  Dense(1, activation = 'sigmoid', kernel_initializer = GlorotUniform())
])

model_3_b_cnn = Sequential([
  Input(shape=(9, 9, 1)),
  Conv2D(32, kernel_size=(3, 3), activation='relu', kernel_initializer=HeNormal()),
  MaxPooling2D(pool_size=(2, 2)),
  Flatten(),
  Dense(16, activation = 'relu', kernel_initializer = HeUniform(),
  Dense(1, activation = 'sigmoid', kernel_initializer = GlorotUniform())
])

model_4_b_cnn = Sequential([
  Input(shape=(9, 9, 1)),
  Conv2D(32, kernel_size=(3, 3), activation='relu', kernel_initializer=HeNormal()),
  MaxPooling2D(pool_size=(2, 2)),
  Flatten(),
  Dense(32, activation = 'relu', kernel_initializer = HeUniform(),
  Dense(1, activation = 'sigmoid', kernel_initializer = GlorotUniform())
])

model_5_b_cnn = Sequential([
  Input(shape=(9, 9, 1)),
  Conv2D(32, kernel_size=(3, 3), activation='relu', kernel_initializer=HeNormal()),
  MaxPooling2D(pool_size=(2, 2)),
  Flatten(),
  Dense(256, activation = 'relu', kernel_initializer = HeUniform(),
  Dense(1, activation = 'sigmoid', kernel_initializer = GlorotUniform())
])
```

Επεξήγηση Κώδικα Ως γνωστόν, η είσοδος έχει τη μορφή μίας $9 \times 9 \times 1$ greyscale εικόνας. Με Conv2D γράφονται τα συνελικτικά επίπεδα, ενώ ο ακέραιος αριθμός του ορίσματος εισόδου καθορίζει τον αριθμό των φίλτρων στο συνελικτικό επίπεδο. Κάθε φίλτρο είναι υπεύθυνο για την ανίχνευση διαφορετικών χαρακτηριστικών από τα εισερχόμενα δεδομένα (π.χ. άκρες, υφές). Για παράδειγμα, στο συνελικτικό επίπεδο του πρώτου μοντέλου υπάρχουν 8 φίλτρα. Με άλλα λόγια, το επίπεδο Conv2D θα παράγει 8 χάρτες χαρακτηριστικών ως έξοδο. Η παράμετρος kernel_size καθορίζει το μέγεθος του φίλτρου, εν προκειμένω 3×3. Αυτό σημαίνει ότι κάθε φίλτρο θα είναι ένας πίνακας 3×3 που θα σαρώσει την εικόνα για να εξαγάγει χαρακτηριστικά. Ακολουθεί το επίπεδο MaxPooling2D, προκειμένου να μειώσει τις διαστάσεις των χαρτών που έχουν εξαχθεί από το συνελικτικό στρώμα και χαρακτηρίζεται από την παράμετρο pool_size,

η οποία καθορίζει τη διάσταση του pooling map ή pooling window, εν προκειμένω 3×3. Τέλος, το επίπεδο Flatten μετατρέπει τους πολυδιάστατους χάρτες χαρακτηριστικών σε ένα μονοδιάστατο διάνυσμα, προετοιμάζοντας τα δεδομένα για τα πλήρως συνδεδεμένα επίπεδα.

Α΄.4 Πολυκατηγορικά CNNs

Επεξήγηση Κώδικα Παρατίθενται οι αρχιτεκτονικές των πολυκατηγορικών CNNs. Όπως και στην περίπτωση των MLPs, εν τέλει διαφέρουν ως προς το επίπεδο εξόδου, αφού χρησιμοποείται η softmax επί 15 υπολογιστικών μονάδων.

```
model_1_m_cnn = Sequential([
  Input(shape=(9, 9, 1)),
  Conv2D(8, kernel_size=(3, 3), activation='relu', kernel_initializer=HeNormal()),
  MaxPooling2D(pool_size=(2, 2)),
  Flatten(),
  Dense(8, activation = 'relu', kernel_initializer = HeUniform(),
  Dense(num_classes, activation='softmax', kernel_initializer = GlorotUniform())])

model_2_m_cnn = Sequential([
  Input(shape=(9, 9, 1)),
  Conv2D(8, kernel_size=(3, 3), activation='relu', kernel_initializer=HeNormal()),
  MaxPooling2D(pool_size=(2, 2)),
  Flatten(),
  Dense(16, activation = 'relu', kernel_initializer = HeUniform(),
  Dense(num_classes, activation='softmax', kernel_initializer = GlorotUniform())])

model_3_m_cnn = Sequential([
  Input(shape=(9, 9, 1)),
  Conv2D(32, kernel_size=(3, 3), activation='relu', kernel_initializer=HeNormal()),
  MaxPooling2D(pool_size=(2, 2)),
  Flatten(),
  Dense(16, activation = 'relu', kernel_initializer = HeUniform(),
  Dense(num_classes, activation='softmax', kernel_initializer = GlorotUniform())])

model_4_m_cnn = Sequential([
  Input(shape=(9, 9, 1)),
  Conv2D(32, kernel_size=(3, 3), activation='relu', kernel_initializer=HeNormal()),
  MaxPooling2D(pool_size=(2, 2)),
  Flatten(),
  Dense(32, activation = 'relu', kernel_initializer = HeUniform(),
  Dense(num_classes, activation='softmax', kernel_initializer = GlorotUniform())])

model_5_m_cnn = Sequential([
  Input(shape=(9, 9, 1)),
  Conv2D(32, kernel_size=(3, 3), activation='relu', kernel_initializer=HeNormal()),
  MaxPooling2D(pool_size=(2, 2)),
  Flatten(),
  Dense(128, activation = 'relu', kernel_initializer = HeUniform(),
  Dense(num_classes, activation='softmax', kernel_initializer = GlorotUniform())])
```

Α΄.5 Υπερπαράμετροι Εκπαίδευσης

Στο σημείο αυτό είναι απαραίτητη η παρουσίαση των υπεπαραμέτρων εκπαίδευσης. Οι τέσσερις αγωγοί μοιράζονται κοινές υπερπαραμέτρους εκπαίδευσης.

```
#General Callbacks
early_stopping = EarlyStopping(monitor='val_accuracy',
    patience=10, restore_best_weights=True)
reduce_lr = ReduceLROnPlateau( monitor='val_accuracy',
    patience= 5, min_lr= 1e-07, factor= 0.1)

#Binary Configuration
compilation_params_b = dict(optimizer='adam', loss='binary_crossentropy',
    metrics=['accuracy'])
fitting_params_b = dict(x=X_train_b, y=y_train_b,
    validation_data=(X_val_b, y_val_b),epochs=250, batch_size = 1024,
    callbacks=[early_stopping, reduce_lr])

#Multiclass Configuration
compilation_params_m = dict(optimizer='adam',
    loss='sparse_categorical_crossentropy', metrics=['accuracy'])
fitting_params_b = dict(x=X_train_m, y=y_train_m,
    validation_data=(X_val_m, y_val_m), epochs=250, batch_size = 1024,
    callbacks=[early_stopping, reduce_lr])
```

1. **Πρόωρη Διακοπή**. Το callback EarlyStopping είναι ένα ισχυρό εργαλείο για την αποτροπή της υπερεκπαίδευσης και την εξοικονόμηση υπολογιστικών πόρων. Παρακολουθεί μια συγκεκριμένη μετρική - στην προκειμένη περίπτωση, την ακρίβεια επικύρωσης (val_accuracy) και διακόπτει τη διαδικασία εκπαίδευσης εάν η μετρική δεν βελτιωθεί για ένα προκαθορισμένο αριθμό εποχών, ορισμένο από την παράμετρο υπομονής (patience). Η παράμετρος restore_best_weights (επαναφορά καλύτερων βαρών) εξασφαλίζει ότι τα βάρη του μοντέλου κατά την εποχή με τη μέγιστη ακρίβεια επικύρωσης επαναφέρονται στο τέλος της εκπαίδευσης, προς επίτευξη της βέλτιστης δυνατής απόδοσης του μοντέλου με βάση το σύνολο επικύρωσης.

2. **Μείωση Ρυθμού Μάθησης**. Το callback ReduceLROnPlateau προσαρμόζει δυναμικά το ρυθμό μάθησης (LR: learning rate) όταν η υπό παρακολούθηση μετρική σταθεροποιείται, βοηθώντας το μοντέλο να συγκλίνει αποτελεσματικά. Όπως και στο EarlyStopping, παρακολουθείται η ακρίβεια επικύρωσης, ενώ η υπομονή ορίζει τον μέγιστο αριθμό των εποχών χωρίς βελτίωση, μέχρις ότου ενεργοποιηθεί η μείωση του ρυθμού μάθησης. Ο min_lr (ελάχιστος ρυθμός μάθησης) θέτει την ελάχιστη επιτρεπτή τιμή του ρυθμού και τέλος, ο factor (συντελεστής) καθορίζει το ποσοστό με το οποίο θα μειωθεί ο ρυθμός μάθησης.

3. **Compilation Παράμετροι**. Για τον πλήρη ορισμό του μοντέλου, είναι απαραίτητη η επιλογή του optimizer (μέθοδος βελτιστοποίησης), της συνάρτησης απώλειας και της υπό παρακολούθηση μετρικής. Σε όλες τις περιπτώσεις διαλέγεται ο αλγόριθμος βελτιστοποίησης Adam, μια δημοφιλής επιλογή λόγω της δυνατότητας προσαρμογής του ρυθμού μάθησης. Όπως έχει εξηγηθεί στα εισαγωγικά κεφάλαια, για δυαδική ταξινόμηση εφαρμόζεται η συνάρτηση απώλειας binary_crossentropy, ενώ στην πολυκατηγορική ταξινόμηση η συνάρτηση sparse_categorical_crossentropy. Τέλος, ως μετρικές αξιολόγη-

σης της απόδοσης του μοντέλου επιλέγονται οι προσαρμοσμένες κατά διάταξη μορφές της accuracy.

4. **Παράμετροι Εκπαίδευσης**. Οι παράμετροι εκπαίδευσης, πέρα από τα δεδομένα εκπαίδευσης και επικύρωσης περιλαμβάνουν σημαντικές υπερπαραμέτρους όπως, το πλήθος των epochs (εποχές), το batch size και τα προαναφερθέντα callbacks (πρόωρης διακοπής και μείωσης ρυθμού εκπαίδευσης). Το πλήθος των epochs καθορίζει τον μέγιστο αριθμό επαναλήψεων επί του συνόλου εκπαίδευσης. Βέβαια, η εκπαίδευση σταματάει νωρίτερα αν πληρούνται τα κριτήρια πρόωρης διακοπής. Το batch size (μέγεθος παρτίδας) καθορίζει τον αριθμό των δειγμάτων που επεξεργάζονται πριν ενημερωθούν οι εσωτερικές παράμετροι του μοντέλου και σε όλες τις περιπτώσεις είναι 1024.

Παράρτημα Β΄

Υλοποίηση των μοντέλων TabNet

Παρατίθονται οι αρχιτεκτονικές των μοντέλων TabNet, διατεταγμένες σε αύξουσα κατά μέγεθος σειρά. Υιοθετείται αυτούσια η python μορφή του κώδικα, όπως γράφτηκε υπό τη βιβλιοθήκη pytorch.

```
clf_1_params_b = dict( n_d = 20, n_a = 20, n_steps = 5, n_shared = 2,
  n_independent = 2, momentum = 0.02, lambda_sparse = 0.001, clip_value = 4.5,
  gamma = 1.85, optimizer_fn=adam, optimizer_params=dict(lr=1e-2),
  verbose = 0, seed = seed_pi, device_name = device, mask_type="sparsemax",
  scheduler_params = scheduler_params_ElR_loss, scheduler_fn = scheduler_ElR)

clf_2_params_b = dict( n_d = 35, n_a = 35, n_steps = 5, n_shared = 2,
  n_independent = 2, momentum = 0.02, lambda_sparse = 0.001, clip_value = 4.5,
  gamma = 1.85, optimizer_fn=adam, optimizer_params=dict(lr=1e-2),
  verbose = 0, seed = seed_pi, device_name = device, mask_type="sparsemax",
  scheduler_params = scheduler_params_ElR_loss, scheduler_fn = scheduler_ElR)

clf_3_params_b = dict( n_d = 50, n_a = 50, n_steps = 5, n_shared = 2,
  n_independent = 2, momentum = 0.02, lambda_sparse = 0.001, clip_value = 4.5,
  gamma = 1.85, optimizer_fn=adam, optimizer_params=dict(lr=1e-2),
  verbose = 0, seed = seed_pi, device_name = device, mask_type="sparsemax",
  scheduler_params = scheduler_params_ElR_loss, scheduler_fn = scheduler_ElR)

clf_4_params_b = dict( n_d = 70, n_a = 70, n_steps = 5, n_shared = 2,
  n_independent = 2, momentum = 0.02, lambda_sparse = 0.001, clip_value = 4.5,
  gamma = 1.85, optimizer_fn=adam, optimizer_params=dict(lr=1e-2),
  verbose = 0, seed = seed_pi, device_name = device, mask_type="sparsemax",
  scheduler_params = scheduler_params_ElR_loss, scheduler_fn = scheduler_ElR)

clf_5_params_b = dict( n_d = 77, n_a = 77, n_steps = 5, n_shared = 2,
  n_independent = 2, momentum = 0.02, lambda_sparse = 0.001, clip_value = 4.5,
  gamma = 1.85, optimizer_fn=adam, optimizer_params=dict(lr=1e-2),
  verbose = 0, seed = seed_pi, device_name = device, mask_type="sparsemax",
  scheduler_params = scheduler_params_ElR_loss, scheduler_fn = scheduler_ElR)

fit_params_b = dict( X_train = X_train_b, y_train = y_train_b.values,
  eval_set = [(X_val_b, y_val_b.values)], patience = 5, max_epochs = 200,
  eval_metric = ["logloss"], eval_name = ["Validation"], drop_last = False,
  batch_size = 16384, virtual_batch_size = 1024, num_workers = 2,
  pin_memory = True, warm_start = False, compute_importance = True)
```

```
clf_1_params_m = dict( n_d = 25, n_a = 25, n_steps = 5, n_shared = 2,
  n_independent = 2, momentum = 0.02, lambda_sparse = 0.001, clip_value = 4.5,
  gamma = 1.85, optimizer_fn=adam, optimizer_params=dict(lr=1e-2),
  verbose = 0, seed = seed_pi, device_name = device, mask_type="sparsemax",
  scheduler_params = scheduler_params_ElR_loss, scheduler_fn = scheduler_ElR)

clf_2_params_m = dict( n_d = 30, n_a = 30, n_steps = 5, n_shared = 2,
  n_independent = 2, momentum = 0.02, lambda_sparse = 0.001, clip_value = 4.5,
  gamma = 1.85, optimizer_fn=adam, optimizer_params=dict(lr=1e-2),
  verbose = 0, seed = seed_pi, device_name = device, mask_type="sparsemax",
  scheduler_params = scheduler_params_ElR_loss, scheduler_fn = scheduler_ElR)

clf_3_params_m = dict( n_d = 43, n_a = 43, n_steps = 5, n_shared = 2,
  n_independent = 2, momentum = 0.02, lambda_sparse = 0.001, clip_value = 4.5,
  gamma = 1.85, optimizer_fn=adam, optimizer_params=dict(lr=1e-2),
  verbose = 0, seed = seed_pi, device_name = device, mask_type="sparsemax",
  scheduler_params = scheduler_params_ElR_loss, scheduler_fn = scheduler_ElR)

clf_4_params_m = dict( n_d = 46, n_a = 46, n_steps = 5, n_shared = 2,
  n_independent = 2, momentum = 0.02, lambda_sparse = 0.001, clip_value = 4.5,
  gamma = 1.85, optimizer_fn=adam, optimizer_params=dict(lr=1e-2),
  verbose = 0, seed = seed_pi, device_name = device, mask_type="sparsemax",
  scheduler_params = scheduler_params_ElR_loss, scheduler_fn = scheduler_ElR)

clf_5_params_m = dict( n_d = 49, n_a = 49, n_steps = 5, n_shared = 2,
  n_independent = 2, momentum = 0.02, lambda_sparse = 0.001, clip_value = 4.5,
  gamma = 1.85, optimizer_fn=adam, optimizer_params=dict(lr=1e-2),
  verbose = 0, seed = seed_pi, device_name = device, mask_type="sparsemax",
  scheduler_params = scheduler_params_ElR_loss, scheduler_fn = scheduler_ElR)

fit_params_m = dict( X_train = X_train_m, y_train = y_train_m.values,
  eval_set = [(X_val_m, y_val_m.values)], patience = 5, max_epochs = 200,
  eval_metric = ["logloss"], eval_name = ["Validation"], drop_last = False,
  batch_size = 16384, virtual_batch_size = 1024, num_workers = 2,
  pin_memory = True, warm_start = False, compute_importance = True)
```

Ακολουθεί μία παρουσίαση των υπερπαραμέτρων του TabNet, το οποίο υλοποιείται μέσω pytorch. Η εν *λόγω* παρουσίαση θα είναι κοινή και για τις δύο διατάξεις. Αρχικά παρουσιάζονται οι δικτυακές παράμετροι αρχιτεκτονικής:

- **n_d** Το πλάτος του επιπέδου απόφασης του κωδικοποιητή, στο οποίο διεξάγονται οι προβλέψεις βάσει των χαρακτηριστικών που έχουν επιλεγεί από τον μηχανισμό προσοχής. Μεγαλύτερες τιμές της παραμέτρου προσδίδουν μεγαλύτερη προβλεπτική ικανότητα, αλλά ελλοχεύουν τον κίνδυνο της υπερεκπαίδευσης. Συνήθεις τιμές κυμαίνονται από 8 έως 64.
- **n_a** Το πλάτος του attention embedding για κάθε μάσκα. Σύμφωνα με τη βιβλιογραφία [14], είθισται η ρύθμιση n_d=n_a.
- **n_steps** Ο αριθμός των βημάτων απόφασης.
- **gamma** Συντελεστής επιλογής χαρακτηριστικών στις μάσκες.
- **n_independent** Ο αριθμός των ανεξάρτητων Gated Linear Units επιπέδων σε κάθε βήμα.

- **n_shared** Ο αριθμός των κοινών Gated Linear Units επιπέδων σε κάθε βήμα.

Ακολουθούν οι παράμετροι κανονικοποίησης και βελτιστοποίησης.

- **momentum** Η παράμετρος ορμής για το batch normalization.
- **clip_value** Άνω φράγμα των τιμών.
- **lambda_sparse** Αυτός είναι ο συντελεστής κανονικοποίησης της σπανιότητας όπως προτείνεται στο αρχικό άρθρο. Μεγαλύτερες τιμές αυτού του συντελεστή κάνουν το μοντέλο πιο αραιό όσον αφορά την επιλογή χαρακτηριστικών.

Τέλος, παρουσιάζονται οι παράμετροι βελτιστοποίησης και χρονικού προγραμματισμού (scheduler).

- **optimizer_fn** Η μέθοδος βελτιστοποίησης για την εκπαίδευση του μοντέλου. Στην προκειμένη περίπτωση, χρησιμοποιείται ο αλγόριθμος Adam.
- **optimizer_params** Οι παράμετροι βελτιστοποιήσης, όπου lr είναι ο αρχικός ρυθμός μάθησης.
- **scheduler_fn** Ο scheduler που προσαρμόζει τον ρυθμό μάθησης κατά τη διάρκεια της εκπαίδευσης.
- **scheduler_params** Οι παράμετροι χρονικής προσαρμογής

Οι παραπάνω υπερπαράμετροι συλλογικά καθορίζουν τη δομή και τη διαδικασία εκπαίδευσης του TabNet μοντέλου, επηρεάζοντας την απόδοση, τη σταθερότητα και τη γενίκευση του μοντέλου σε δεδομένα εκτός δείγματος. Οι έννοιες verbose, seed και device έχουν αμιγώς προγραμματιστικό περιεχόμενο και δεν αποτελούν υπερπαραμέτρους. Ακολουθεί μία σύντομη ανάλυση των υπερπαραμέτρων εκπαίδευσης (οι υπόλοιπες παράμετροι επίσης παραλείπονται). Παρουσιάζεται η σημασία τους και ο ρόλος τους στη διαδικασία εκπαίδευσης.

- **max_epochs** Ο μέγιστος αριθμός epochs - εποχών εκπαίδευσης. Εάν δεν ενεργοποιηθεί ο πρόωρος τερματισμός - early stopping, η εκπαίδευση θα συνεχιστεί μέχρι αυτόν τον αριθμό εποχών.
- **patience** Ο μέγιστος αριθμός των epochs (εποχές) εκπαίδευσης, στις οποίες εάν δεν υπάρξει βελτίωση στη μετρική αξιολόγησης, η διαδικασία εκπαίδευσης θα τερματιστεί (early stopping).
- **eval_metric** Η υπό παρατήρηση μετρική αξιολόγησης της απόδοσης του μοντέλου κατά τη διάρκεια της εκπαίδευσης. Αναφέρεται στη συνάρτηση απώλειας.
- **batch_size** Το μέγεθος του batch (παρτίδας) που χρησιμοποιείται κατά την εκπαίδευση. Καθορίζει τον αριθμό των δειγμάτων που επεξεργάζονται πριν ενημερωθούν οι εσωτερικές παράμετροι του μοντέλου.
- **virtual_batch_size** Το μέγεθος της εικονικής παρτίδας που χρησιμοποιείται κατά την επεξεργασία των δεδομένων.

Αυτές οι υπερπαράμετροι συλλογικά καθορίζουν τη διαδικασία εκπαίδευσης του TabNet μοντέλου, επηρεάζοντας την αποδοτικότητα, την ταχύτητα και την ποιότητα της εκπαίδευσης.

Bibliography

[1] Charu C Aggarwal et al. *Neural networks and deep learning*, volume 10. Springer, 2018.

[2] Charu C Aggarwal and Charu C Aggarwal. An introduction to artificial intelligence. *Artificial Intelligence: A Textbook*, pages 1–34, 2021.

[3] Warren S McCullough and Walter Pitts. A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5(4):115–127, 1943.

[4] Frank Rosenblatt. The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386, 1958.

[5] Jan Mycielski. Marvin minsky and seymour papert, perceptrons, an introduction to computational geometry. 1972.

[6] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010.

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[8] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[9] David H Hubel and Torsten N Wiesel. Receptive fields of single neurones in the cat's striate cortex. *The Journal of physiology*, 148(3):574, 1959.

[10] Kunihiko Fukushima. Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4):193–202, 1980.

[11] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60(6):84–90, 2017.

[13] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[14] Sercan Ö Arik and Tomas Pfister. Tabnet: Attentive interpretable tabular learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 6679–6687, 2021.

[15] Andre Martins and Ramon Astudillo. From softmax to sparsemax: A sparse model of attention and multi-label classification. In *International conference on machine learning*, pages 1614–1623. PMLR, 2016.

[16] V7 Labs. Confusion matrix: How to use it & interpret results [examples], 2023.

[17] Evidently AI. Classification metrics guide, 2023.

[18] Fabian Pedregosa, G. Varoquaux, Alexandre Gramfort, V. Michel, B. Thirion, O. Grisel, Mathieu Blondel, Gilles Louppe, P. Prettenhofer, Ron Weiss, Ron J. Weiss, J. Vanderplas, Alexandre Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in python. *ArXiv*, abs/1201.0490, 2011.

[19] Alaa Tharwat. Classification assessment methods. *Applied computing and informatics*, 17(1):168–192, 2020.

[20] David MW Powers. Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation. *arXiv preprint arXiv:2010.16061*, 2020.

[21] Vicente García, Ramon A Mollineda, and J Salvador Sánchez. Theoretical analysis of a performance measure for imbalanced data. In *2010 20th International Conference on Pattern Recognition*, pages 617–620. IEEE, 2010.

[22] Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, et al. *An introduction to statistical learning*, volume 112. Springer, 2013.

[23] Marina Sokolova and Guy Lapalme. A systematic analysis of performance measures for classification tasks. *Information processing & management*, 45(4):427–437, 2009.

[24] Juri Opitz and Sebastian Burst. Macro f1 and macro f1. *arXiv preprint arXiv:1911.03347*, 2019.

[25] Margherita Grandini, Enrico Bagli, and Giorgio Visani. Metrics for multi-class classification: an overview. *arXiv preprint arXiv:2008.05756*, 2020.

[26] Wisam Elmasry, Akhan Akbulut, and Abdul Halim Zaim. Empirical study on multiclass classification-based network intrusion detection. *Computational Intelligence*, 35(4):919–954, 2019.

[27] Iman Sharafaldin, Arash Habibi Lashkari, Ali A Ghorbani, et al. Toward generating a new intrusion detection dataset and intrusion traffic characterization. *ICISSp*, 1:108–116, 2018.

[28] Gordon Fyodor Lyon. *Nmap network scanning: The official Nmap project guide to network discovery and security scanning*. Insecure, 2009.

[29] Sérgio SC Silva, Rodrigo MP Silva, Raquel CG Pinto, and Ronaldo M Salles. Botnets: A survey. *Computer Networks*, 57(2):378–403, 2013.

[30] Moheeb Abu Rajab, Jay Zarfoss, Fabian Monrose, and Andreas Terzis. A multifaceted approach to understanding the botnet phenomenon. In *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, pages 41–52, 2006.

[31] Iuri A Mundstock, Yuri Santo, Thiago LT da Silveira, Roger Immich, André Riker, and Bruno L Dalmazo. On feature selection techniques for detecting dos attacks with a multi-class classifier. 2024.

[32] Sefat Mahjabin. Implementation of dos and ddos attacks on cloud servers. *Periodicals of Engineering and Natural Sciences*, 6(2):148–158, 2018.

[33] Jeremy Charlier, Aman Singh, Gaston Ormazabal, Radu State, and Henning Schulzrinne. Syngan: Towards generating synthetic network attacks using gans. *arXiv preprint arXiv:1908.09899*, 2019.

[34] Sajal Bhatia, Sunny Behal, and Irfan Ahmed. Distributed denial of service attacks and defense mechanisms: current landscape and future directions. *Versatile Cybersecurity*, pages 55–97, 2018.

[35] Shekyan. Github - shekyan/slowhttptest: Application layer dos attack simulator. Accessed: 2024-06-26.

[36] Suroto Suroto. A review of defense against slow http attack. *JOIV: International Journal on Informatics Visualization*, 1(4):127–134, 2017.

[37] Md Delwar Hossain, Hideya Ochiai, Fall Doudou, and Youki Kadobayashi. Ssh and ftp brute-force attacks detection in computer networks: Lstm and machine learning approaches. In *2020 5th international conference on computer and communication systems (ICCCS)*, pages 491–497. IEEE, 2020.

[38] Zakir Durumeric, Frank Li, James Kasten, Johanna Amann, Jethro Beekman, Mathias Payer, Nicolas Weaver, David Adrian, Vern Paxson, Michael Bailey, et al. The matter of heartbleed. In *Proceedings of the 2014 conference on internet measurement conference*, pages 475–488, 2014.

[39] Aalok Thakkar. Heartbleed: A formal methods perspective. *GitHub. Accessed: Feb*, 24, 2023.

[40] Andrew Hoffman. *Web application security.* " O'Reilly Media, Inc.", 2024.

[41] Jan Vykopal. A flow-level taxonomy and prevalence of brute force attacks. In *International Conference on Advances in Computing and Communications*, pages 666–675. Springer, 2011.

[42] Jeremiah Grossman. *XSS attacks: cross site scripting exploits and defense.* Syngress, 2007.

[43] Shashank Gupta and Brij Bhooshan Gupta. Cross-site scripting (xss) attacks and defense mechanisms: classification and state-of-the-art. *International Journal of System Assurance Engineering and Management*, 8:512–530, 2017.

[44] Gary Wassermann and Zhendong Su. Static detection of cross-site scripting vulnerabilities. In *Proceedings of the 30th international conference on Software engineering*, pages 171–180, 2008.

[45] Canadian Institute for Cybersecurity. CICIDS 2017 Dataset, 2017. Accessed: 2024-06-21.

[46] Naveen Bindra and Manu Sood. Evaluating the impact of feature selection methods on the performance of the machine learning models in detecting ddos attacks. *Rom. J. Inf. Sci. Technol*, 23(3):250–261, 2020.

[47] Laurens D'hooge, Miel Verkerken, Bruno Volckaert, Tim Wauters, and Filip De Turck. Establishing the contaminating effect of metadata feature inclusion in machine-learned network intrusion detection models. In *International Conference on Detection of Intrusions and Malware, and Vulnerability Assessment*, pages 23–41. Springer, 2022.

[48] Tianqi Chen and Carlos Guestrin. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pages 785–794, 2016.

[49] Trevor Hastie, Robert Tibshirani, Jerome H Friedman, and Jerome H Friedman. *The elements of statistical learning: data mining, inference, and prediction*, volume 2. Springer, 2009.

[50] Jerome H Friedman. Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.

[51] Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, and Andrey Gulin. Catboost: unbiased boosting with categorical features. *Advances in neural information processing systems*, 31, 2018.

[52] Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. Lightgbm: A highly efficient gradient boosting decision tree. *Advances in neural information processing systems*, 30, 2017.

[53] Tom Maguire, Lennard Manuel, RA Smedinga, and M Biehl. A review of feature selection and ranking methods. *19th SC@ RUG 2021-2022*, page 15, 2022.

[54] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16:321–357, 2002.

[55] Guillaume Lemaître, Fernando Nogueira, and Christos K. Aridas. *imbalanced-learn: A Python Toolbox to Tackle the Curse of Imbalanced Datasets in Machine Learning*, 2024. Version 0.12.3.

[56] Thomas Cover and Peter Hart. Nearest neighbor pattern classification. *IEEE transactions on information theory*, 13(1):21–27, 1967.

[57] Roger Koenker. Quantile regression: 40 years on. *Annual review of economics*, 9:155–176, 2017.

[58] Alok Sharma, Edwin Vans, Daichi Shigemizu, Keith A Boroevich, and Tatsuhiko Tsunoda. Deepinsight: A methodology to transform a non-image data to an image for convolution neural network architecture. *Scientific reports*, 9(1):11399, 2019.

[59] Nicola Mignoni. tab2img 0.0.2: Convert tabular data into images for deep learning models. Available at PyPI: https://pypi.org/project/tab2img/ and GitHub: https://github.com/nicomignoni/tab2img, 2021.

[60] Charu C Aggarwal et al. *Data mining: the textbook*, volume 1. Springer, 2015.

[61] Jared Kaplan, Sam McCandlish, Tom Henighan, Tom B Brown, Benjamin Chess, Rewon Child, Scott Gray, Alec Radford, Jeffrey Wu, and Dario Amodei. Scaling laws for neural language models. *arXiv preprint arXiv:2001.08361*, 2020.

[62] Jordan Hoffmann, Sebastian Borgeaud, Arthur Mensch, Elena Buchatskaya, Trevor Cai, Eliza Rutherford, Diego de Las Casas, Lisa Anne Hendricks, Johannes Welbl, Aidan Clark, et al. Training compute-optimal large language models. *arXiv preprint arXiv:2203.15556*, 2022.

[63] Mikhail Belkin, Daniel Hsu, Siyuan Ma, and Soumik Mandal. Reconciling modern machine-learning practice and the classical bias–variance trade-off. *Proceedings of the National Academy of Sciences*, 116(32):15849–15854, 2019.

[64] Günter Klambauer, Thomas Unterthiner, Andreas Mayr, and Sepp Hochreiter. Self-normalizing neural networks. *Advances in neural information processing systems*, 30, 2017.