Εθνικο Μετσοβιο Πολυτεχνειο
Σχολη Ηλεκτρολογων Μηχανικων και Μηχανικων Υπολογιστων
Τομεας Τεχνολογιας Πληροφορικης και Υπολογιστων

Μη επιβλεπόμενη ανίχνευση ανωμαλιών με χρήση βαθιών νευρωνικών δικτύων και εφαρμογή στη ναυτιλία

Διπλωματικη Εργασια

του

Ανδρέα Αθανασόπουλου

Επιβλέπων: Α-Γ Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.
Συνεπιβλέπων: Γ. Σιόλας
Ε.ΔΙ.Π. ΕΜΠ

Εργαστηριο Ευφυων Συστηματων
Αθήνα, Ιούλιος 2019

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών
Εργαστήριο Ευφυών Συστημάτων

Μη επιβλεπόμενη ανίχνευση ανωμαλιών με χρήση βαθιών νευρωνικών δικτύων και εφαρμογή στη ναυτιλία

Διπλωματικη Εργασια

του

Ανδρέα Αθανασόπουλου

Επιβλέπων: Α-Γ Σταφυλοπάτης
Καθηγητής Ε.Μ.Π.
Συνεπιβλέπων: Γ. Σιόλας
Ε.ΔΙ.Π. ΕΜΠ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 29η Ιουλίου 2019.

(Υπογραφή)	*(Υπογραφή)*	*(Υπογραφή)*
.....	.....	.....
Α-Γ Σταφυλοπάτης	Γιώργος Στάμου	Μαρία Λάμπρου
Καθηγητής Ε.Μ.Π.	Καθηγητής Ε.Μ.Π.	Καθηγήτρια Πανεπιστημίου Αιγαίου

Αθήνα, Ιούλιος 2019

(Υπογραφή)

.....

Ανδρεας Αθανασοπουλος
Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Τεχνολογίας Πληροφορικής και Υπολογιστών
Εργαστήριο Ευφυών Συστημάτων

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τους επιβλέποντες καθηγητές μου. Επίσης ευχαριστώ ιδιαίτερα τον συμφοιτητή μου Ευστάθιο Ανδριανόπουλο για την πολύτιμη βοήθεια του στην διάρκεια της φοίτησης μας. Τέλος, θα ήθελα να ευχαριστήσω την οικογένειά μου για την στήριξη τους.

Περίληψη

Η παρούσα διπλωματική εργασία ασχολείται με τις μεθόδους ανίχνευσης ανωμαλιών και την εφαρμογή τους σε δεδομένα που σχετίζονται με το μηχανολογικό σύστημα ενός εμπορικού πλοίου. Στόχος είναι η δημιουργία ενός συστήματος προβλεπτικής συντήρησης ανιχνεύοντας ανώμαλα δεδομένα στο σύστημα της μηχανής του πλοίου και συσχετίζοντας τα με ιστορικές βλάβες του πλοίου. Γίνεται χρήση τριών διαφορετικών βαθιών νευρωνικών δικτύων αυτών του αυτοκωδικοποιητή, ενός ανταγωνιστικού αυτοκωδικοποιητή και ενός ανατροφοδοτούμενου νευρωνικού δικτύου. Αρχίσαμε με την υλοποίηση του αυτοκωδικοποιητή ο οποίος ανιχνεύει ανώμαλα δεδομένα σε περίπτωση που το σφάλμα ανακατασκευής περάσει ένα εμπειρικώς ορισμένο κατώφλι απόφασης, αφού στην περίπτωση των στατιστικώς ορισμένων παρατηρήσαμε υψηλό αριθμό ανωμαλιών. Στην συνέχεια, υλοποιήσαμε το ανατροφοδοτούμενο νευρωνικό δίκτυο του οποίου συγκρίναμε τα αποτελέσματα με αυτά του παραπάνω αποκωδικοποιητή. Θέλοντας να χρησιμοποιήσουμε τον ενδιάμεσο χώρο προβολής υλοποιήσαμε έναν ανταγωνιστικό αυτοκωδικοποιητή για την ανίχνευση των ανωμαλιών. Τέλος ερευνήσαμε την ικανότητα του ανταγωνιστικού αυτοκωδικοποιητή στην ανίχνευση ανωμαλιών μέσω μιας αρχιτεκτονικής συσταδοποίησης δεδομένων περνώντας ιδιαίτερα ικανοποιητικά αποτελέσματα παρατηρώντας όμως αστάθεια των αποτελεσμάτων σε περίπτωση δυσανάλογων πιθανοτήτων των κλάσεων των συστάδων.

Λέξεις Κλειδιά

Ανίχνευση ανωμαλιών, Προβλεπτική συντήρηση, Αυτοκωδικοποιητής, Ανταγωνιστικός αυτοκωδικοποιητής, Ανατροφοδοτούμενα νευρωνικά δίκτυα, Βαθιά νευρωνικά δίκτυα, Μηχανική μάθηση.

Abstract

This thesis explores with outlier detection techniques and their application in the real world problem of identifying anomalous data, in relation to the mechanical system of a merchant boat. The aim is to develop a predictive maintenance system by detecting anomalous behaviour and assess the correlation to historical engine damages. Three different deep neural network architectures were developed for the aforementioned purpose: an autoencoder, an adversarial autoencoder and a recurrent neural network that were compared to their ability to detect anomalies. We start by utilizing the autoencoder to identify anomalies when the reconstruction error exceeds a empirically defined threshold. Due to issues created by the high number of anomalous data, it was rendered difficult to create a statistically defined threshold. Furthermore we implemented the recurrent neural network, which served to compare it with the results of the autoencoder. In order to utilize the latent dimension on anomaly detection we integrated an adversarial autoencoder. Finally, we investigated the ability of adversarial autoencoder to identify anomalies in a clustering set-up with a categorical imposed probability and we concluded that besides the positive results, the training process was unstable when we imposed an imbalanced categorical distribution.

Keywords

Anomaly detection, Outlier detection, Predictive maintenance, Autoencoder, Adversarial autoencoder, Recurrent neural network, Deep neural network, Machine learning

Περιεχόμενα

Κατάλογος Σχημάτων

Κατάλογος Πινάκων

Κεφάλαιο 1

Εισαγωγή

Ο γενικότερος στόχος της επιστήμης είναι να κατανοήσουμε τον κόσμο που ζούμε ενώ παράλληλα προσφέρει λύσεις σε προβλήματα που έχει ο άνθρωπος ανά τους αιώνες. Σήμερα, η τεχνητή νοημοσύνη βρίσκεται στο επίκεντρο του ενδιαφέροντος της επιστήμης καθώς παρέχει μια νέα οπτική στην αντίληψη του κόσμου μέσω των υπολογιστών. Τα τελευταία χρόνια παρατηρείται ραγδαία ανάπτυξη πληροφοριακών συστημάτων και εφαρμογών που συλλέγουν και επεξεργάζονται δεδομένα που επηρεάζουν έμμεσα ή άμεσα την καθημερινότητά όλων μας. Τα δεδομένα συλλέγονται από διάφορες πηγές όπως το διαδίκτυο, αισθητήρες με συνεχώς αυξανόμενο ρυθμό και η επεξεργασία τους εξυπηρετεί ένα ευρύ φάσμα εφαρμογών καθώς και επιστημονικές μελέτες. Έτσι, η τεχνητή νοημοσύνη καλείται να παρέχει πληροφορία και γνώση μέσω των δεδομένων και υπόσχεται να δώσει λύσεις σε πολλά από τα προβλήματα του ανθρώπου.

Στην καρδιά της τεχνητής νοημοσύνης σήμερα βρίσκονται οι τεχνικές μηχανικής μάθησης όπου γίνεται χρήση μοντέλων ικανών να μαθαίνουν από τα δεδομένα χωρίς να έχει γίνει κάποιος ρητός προγραμματισμός. Σήμερα, ένα μεγάλο κομμάτι της έρευνας πάνω στην μηχανική μάθηση εστιάζει στα βαθιά νευρωνικά δίκτυα. Ο λόγος που τα βαθιά νευρωνικά δίκτυα βρίσκονται στο επίκεντρο του ενδιαφέροντος σήμερα είναι ότι έχουν δώσει επαναστατικά αποτελέσματα σε πολλά προβλήματα που έχουν σχέση με πεδία όπως η όραση υπολογιστών [12], η επεξεργασία φωνής και φυσικής γλώσσας [10, 3] καθώς και της υπολογιστικής βιολογίας [2]. Η απόδοση των συγκεκριμένων δικτύων έγκειται στην ικανότητα τους να επεξεργάζονται μεγάλο όγκο δεδομένων σε αρχιτεκτονικές με χιλιάδες παραμέτρους σε σχέση με τον όγκο που μπορούσαν να επεξεργάζονται αποτελεσματικά προηγούμενοι αλγόριθμοι μηχανικής μάθησης.

Ο μεγάλος όγκος των δεδομένων καθιστά δύσκολη την διαδικασία επισκόπησης και επικύρωσης της ορθότητας τους. Το ζήτημα αυτό είναι ιδιαίτερα σημαντικό καθώς οι αλγόριθμοι επεξεργασίας δεδομένων συμπεριφέρονται απρόβλεπτα όταν καλούνται να διαχειριστούν πληροφορία η οποία δεν είναι έγκυρη. Ένα χαρακτηριστικό παράδειγμα είναι τα νευρωνικά δίκτυα των οποίων η συμπεριφορά εξαρτάται σε μεγάλο βαθμό από την ποιότητα των δεδομένων εισόδου. Επίσης πολλές φορές εμφάνιση ανωμαλιών στα δεδομένα μπορούν να οδηγήσουν σε πολύ σημαντικά συμπεράσματα όπως για παράδειγμα σε εφαρμογές στην οικονομία, μηχανολογικά συστήματα και την υγεία.

1.1 Αντικείμενο της διπλωματικής

Η παρούσα διπλωματική εργασία ασχολείται με την ανίχνευση ανωμαλιών δηλαδή μη κανονικών καταστάσεων και την κατηγοριοποίηση τους σε προβλήματα μεγάλου όγκου δεδομένων με χρήση βαθιών νευρωνικών δικτύων. Η έρευνα επικεντρώνεται στην ανάλυση πολυδιάστατων χρονοσειρών που αφορούν δεδομένα πλοίων. Στόχος της εργασίας είναι η παρουσίαση τεχνικών εντοπισμού ανωμαλιών και η εφαρμογή τους στα διαθέσιμα δεδομένα. Πιο συγκεκριμένα στην διάθεση μας έχουμε χρονοσείρες που αφορούν τα μηχανολογικά στοιχεία ενός πλοίου και ενδιαφερόμαστε να εντοπίσουμε ανωμαλίες οι οποίες μπορούν να βοηθήσουν μηχανικούς στην επίβλεψη και την συντήρηση του πλοίου. Ένα πρόβλημα που ανακύπτει είναι ότι πέρα από ορισμένες καταγραμμένες βλάβες το τι μπορεί να θεωρηθεί σαν ανωμαλία στα δεδομένα δεν μπορεί να ορισθεί με ακρίβεια. Το παραπάνω προκύπτει από το γεγονός ότι ένα τέτοιο μηχανολογικό σύστημα έχει μεγάλη πολυπλοκότητα όσον αναφορά τις συσχετίσεις μεταξύ των χαρακτηριστικών και είναι σχεδόν αδύνατη η δημιουργία ενός συστήματος κανόνων (rule based system) για την ανίχνευση ασυνήθιστων συμπεριφορών. Τα παρόντα συστήματα επικεντρώνονται στις ακραίες τιμές στα διάφορα χαρακτηριστικά ή χωρίζουν το σύστημα σε υπό-ομάδες με σαφή συσχέτιση με σκοπό κυρίως την διαφύλαξη της ασφάλειας του συστήματος. Η χρήση βαθιών νευρωνικών δικτύων έχει το πλεονέκτημα να μοντελοποιεί κρυφές συσχετίσεις που υπάρχουν στα δεδομένα μέσω της διαδικασίας μάθησης. Η προσέγγιση του προβλήματος έγινε με χρήση μεθόδων μη επιβλεπόμενης μάθησης καθώς τα δεδομένα που έχουμε στην διάθεσή μας είναι μη επισημειωμένα. Συγκεκριμένα υλοποιήσαμε έναν αυτοκωδικοποιητή έναν ανταγωνιστικό αυτοκωδικοποιητή και ένα ανατροφοδοτούμενο νευρωνικό δίκτυο τα οποία εμπλουτίσαμε με τεχνικές ώστε να ανιχνεύσουμε απροσδόκητες συμπεριφορές. Τέλος, γίνεται η προσπάθεια συσχέτισης ιστορικών βλαβών με τις περιοχές τις οποίες επισημαίνει το σύστημα ως ανώμαλες.

1.1.1 Συνεισφορά

Η συνεισφορά της διπλωματικής συνοψίζεται ως εξής:

1. Παρουσίαση εφαρμογών που σχετίζονται με την ανίχνευση ανωμαλιών σε δεδομένα
2. Οργάνωση μεθόδων και συστηματική παρουσίαση του προβλήματος ανίχνευσης ανωμαλιών σε δεδομένα
3. Ανάλυση μεθόδων συντήρησης μηχανικών συστημάτων και συσχέτιση με το πρόβλημα ανίχνευσης ανωμαλιών
4. Συστηματική μελέτη και υλοποίηση αλγορίθμων βαθιάς μηχανικής μάθησης για την ανίχνευση ανωμαλιών σε δεδομένα
5. Σύγκριση μοντέλων βαθιάς μηχανικής μάθησης
6. Υλοποίηση συστήματος ανίχνευσης ανωμαλιών σε μη επισημειωμένα δεδομένα και τρόποι αντιμετώπισης του προβλήματος

7. Υλοποίηση ανταγωνιστικού αυτοκωδικοποιητή και έρευνα στην ανίχνευση ανωμαλιών μέσω συσταδοποίησης με δυσανάλογες κλάσεις.

1.2 Οργάνωση του τόμου

Στο υπόλοιπο του τόμου αρχικά παρουσιάζουμε συγγενικές εργασίες οι οποίες εστιάζουν σε εφαρμογές ανίχνευσης ανωμαλιών δίνοντας ιδιαίτερη έμφαση σε εφαρμογές βιομηχανικού περιβάλλοντος. Στο τρίτο κεφάλαιο, παρουσιάσαμε το πρόβλημα ανίχνευσης ανωμαλιών αναλύοντας τον τρόπο προσέγγισης του προβλήματος καθώς επίσης αλγορίθμους και δίκτυα μηχανικής και βαθιάς μηχανικής μάθησης. Στην συνέχεια παρουσιάσαμε τα δεδομένα του προβλήματος εφαρμογής τα οποία έχουν να κάνουν με τα δεδομένα πλοίων. Στο πέμπτο κεφάλαιο έγινε υλοποίηση πειραμάτων των δικτύων του αυτοκωδικοποιητή και του ανατροφοδοτούμενου νευρωνικού δικτύου. Στην διάρκεια του συγκεκριμένου κεφαλαίου έγινε περιγραφή του τρόπου προσέγγισης από τις επιλογές που παρουσιάσαμε στο κεφάλαιο θεωρητικού υποβάθρου καθώς επίσης εφαρμογή τεχνικών για την βελτίωση των αποτελεσμάτων. Στο έκτο κεφάλαιο υλοποιήσαμε ένα ανταγωνιστικό αυτοκωδικοποιητή στον οποίο ερευνήσαμε την ικανότητα του να ανιχνεύει ανωμαλίες μέσω τριών τεχνικών : συσταδοποίηση δεδομένων, χρήση ενδιάμεσου χώρου προβολής και του σφάλματος ανακατασκευής. Τέλος συνοψίσαμε τα αποτελέσματα της εργασίας και προτείνουμε μελλοντικές επεκτάσεις της συγκεκριμένης διπλωματικής.

Κεφάλαιο 2

Συγγενικές εργασίες και εφαρμογές ανίχνευσης ανωμαλιών σε δεδομένα

2.1 Εισαγωγή

Η ανίχνευση ανωμαλιών αποτελεί ένα σημαντικό κομμάτι στην έρευνα που αφορά την μηχανική μάθηση και τα βαθιά νευρωνικά δίκτυα με ιδιαίτερα χρήσιμες εφαρμογές. Σε αυτό το κεφάλαιο, θα παρουσιάσουμε μερικές από τις εφαρμογές που σχετίζονται με την ανίχνευση ανωμαλιών. Η παρουσίαση επικεντρώνεται στις εφαρμογές και όχι στην αναλυτική περιγραφή μοντέλων και τεχνικών οι οποίες αναλύονται σε επόμενο κεφάλαιο. Μια περισσότερο αναλυτική περιγραφή των παρακάτω μπορεί να βρεθεί στα παρακάτω άρθρα [6, 7]. Η παράγραφος που εστιάζουμε περισσότερο είναι εκείνη των εφαρμογών σε βιομηχανικό περιβάλλον και ιδιαίτερα στον τομέα της προβλεπτικής συντήρησης συστημάτων καθώς ο απώτερος στόχος της εργασίας είναι η κατασκευή ενός συστήματος για την επίβλεψη μηχανικών συστημάτων.

2.2 Ανίχνευση εισβολής σε υπολογιστικά συστήματα

Η ανίχνευση εισβολής (Intrusion Detection) αναφέρεται στον εντοπισμό κακόβουλης δραστηριότητας σε υπολογιστικά συστήματα. Η ανίχνευση εισβολής μπορεί να εφαρμοστεί είτε σ΄ έναν μεμονωμένο υπολογιστή (Host-Based Intrusion Detection Systems) είτε σ΄ ένα δίκτυο υπολογιστών (Network Intrusion Detection Systems) για την ανίχνευση κακόβουλης δραστηριότητας. Επίσης τα συστήματα εντοπισμού εισβολής μπορεί να είναι ενεργά ή παθητικά υπό την έννοια ότι τα πρώτα μπορούν να επεμβαίνουν εκτός από το να ειδοποιούν το σύστημα. Πρακτικά, για την εκπαίδευση των μοντέλων χρησιμοποιούνται log files είτε της δραστηριότητας του δικτύου είτε των system calls του συστήματος. Μια ακόμα ιδιαίτερα ενδιαφέρουσα εφαρμογή αφορά την ανίχνευση ιών στο λογισμικό υπολογιστών όπως περιγράφεται στο [28].

2.3 Ανίχνευση απάτης

Η ανίχνευση απάτης fraud detection καθώς είναι ένας πολύ γενικός όρος αναφέρεται σε πολλά διαφορετικά πεδία όπως τραπεζικές απάτες, απάτες στις τηλεπικοινωνίες, πλαστογραφίες κλπ. Σε όλα τα παραπάνω, τα νευρωνικά δίκτυα έχουν επαναστατικά αποτελέσματα και υπόσχονται ακόμα καλύτερες επιδόσεις καθώς συλλέγονται μεγαλύτερα και πιο πλούσια σύνολα από δεδομένα εκπαίδευσης. Σε ότι αφορά στις τραπεζικές απάτες οι εφαρμογές εστιάζουν στις απάτες μέσω τραπεζικών καρτών [24] ενώ για τις απάτες στις τηλεπικοινωνίες αναφέρεται το παρακάτω άρθρο [1]. Μια εξαιρετικά ενδιαφέρουσα εφαρμογή αποτελεί και η ανίχνευση πλαστογραφίας αφού κοστίζει κάθε χρόνο αρκετά δισεκατομμύρια σε τράπεζες και οργανισμούς. Έτσι γίνεται χρήση συνεκτικών νευρωνικών δικτύων CNN για την κατηγοριοποίηση της υπογραφής [19].

2.4 Ανίχνευση ανωμαλιών στην ιατρική

Ο τομέας της υγείας αποτελεί ίσως τον πιο ενδιαφέροντα τομέα για εφαρμογή τεχνητής νοημοσύνης [4] και θα μπορούσε να αλλάξει ριζικά το βιοτικό επίπεδο των ανθρώπων. Παράδειγμα αποτελούν ρομποτικά συστήματα τα οποία πραγματοποιούν επεμβάσεις σε ασθενείς άλλα και πολλά τεστ τα οποία γίνονται μέσω μοντέλων τεχνικής νοημοσύνης. Είναι γεγονός ότι πρέπει να είμαστε επιφυλακτικοί χωρίς όμως να εθελοτυφλούμε στο γεγονός ότι ένα μεγάλο ποσοστό παγκοσμίως στερείται ιατρικής περίθαλψης και πρόληψης. Σε ότι αφορά την ανίχνευση ανωμαλιών στην ιατρική, εφαρμογή βρίσκουν μοντέλα που επεξεργάζονται ιατρικές εικόνες [11] ή σήματα από ηλεκτροεγκεφαλογράφηματα (EEG) [27] με στόχο την διάγνωση κάποιας ασθένειας ή την πρόληψη σε διάφορες ιατρικές συνθήκες.

2.5 Παρακολούθηση βίντεο

Η ανίχνευση ανωμαλιών μέσω βίντεο είναι ιδιαίτερα χρήσιμη σε εφαρμογές που έχουν σχέση με θέματα ασφάλειας. Ένα πρόβλημα που παρουσιάζεται σε εφαρμογές βίντεο είναι ότι δεν είναι δυνατό να προσδιοριστεί με ακρίβεια τι μπορεί να θεωρηθεί ως ανωμαλία και έτσι είναι δύσκολο να εφαρμοστούν τεχνικές επιβλεπόμενης μάθησης. Σε αντίθεση τεχνικές μη επιλεγόμενης μάθησης με στόχο την ανίχνευση ανωμαλιών είναι εφικτές και κυριαρχούν στις εφαρμογές. Ενδιαφέρουσες παράλληλα είναι και εφαρμογές ελέγχου κυκλοφορίας [26] και παρακολούθησης γραμμών παραγωγής.

2.6 Ανίχνευση ανωμαλιών στην βιομηχανία

Βιομηχανικά συστήματα όπως εργοστάσια παραγωγής ηλεκτρικής ενέργειας, μεγάλα οχήματα όπως αεροσκάφη και πλοία αλλά και ανεμογεννήτριες υφίσταται μεγάλη καταπόνηση κατά την διάρκεια της λειτουργίας τους. Βλάβες οι οποίες μπορούν να συμβούν σε τέτοια συστήματα μπορούν να έχουν καταστροφικές συνέπειες για το περιβάλλον αλλά και για τον

άνθρωπο. Παράλληλα προβλήματα που μπορούν να θέσουν τέτοιου είδους συστήματα εκτός λειτουργίας για ένα ορισμένο χρονικό διάστημα και να προκαλέσουν μεγάλη οικονομική ζημιά στις εταιρείες και οργανισμούς που τα διαχειρίζονται. Πολλοί ερευνητές έχουν εστιάσει στην ανίχνευση ανωμαλιών σε βιομηχανικά συστήματα όπως στο άρθρο [18] όπου οι συγγραφείς εφαρμόζουν μοντέλα μηχανικής μάθησης και συγκεκριμένα SVM για την ανίχνευση ανωμαλιών σε χρονοσειρές που σχετίζονται με στροβιλομηχανές. Στο βιβλίο [22] γίνεται αναλυτική περιγραφή του εντοπισμού ανωμαλιών σε μηχανικά συστήματα και προσέγγιση μέσω μπεϋζιανών στατιστικών μοντέλων σε συνδυασμό με την φυσική του εκάστοτε προβλήματος.

2.6.1 Στρατηγικές συντήρησης συστημάτων

Στο πλαίσιο της ανίχνευσης ανωμαλιών στην βιομηχανία και ιδιαίτερα σε συστήματα που εμπλέκονται κινούμενα μηχανικά μέρη, ιδιαίτερο ενδιαφέρον κατέχει το πρόβλημα της προβλεπτικής συντήρησης. Όπως αναφέραμε και προηγουμένως μια επερχόμενη βλάβη μπορεί να κοστίσει πολλά σε εταιρείες που εκμεταλλεύονται τέτοια μηχανικά συστήματα για την παραγωγή αγαθών και υπηρεσιών. Έτσι γεννάται η ανάγκη για μια συστηματική συντήρηση τέτοιων συστημάτων με σκοπό των προγραμματισμό τις συντήρηση, την ελαχιστοποίηση του χρόνου συντήρησης των μηχανημάτων καθώς την ελαχιστοποίηση των βλαβών.

Ο τρόπος με τον οποίον συντηρείται ένα μηχανικό σύστημα είναι ιδιαίτερα σημαντικός και εξαρτάται σε μεγάλο βαθμό από την χρήση και τις προδιαγραφές του συστήματος. Παρακάτω φαίνονται οι προσεγγίσεις συντήρησης σε τέτοια συστήματα.

- Διορθωτική συντήρηση (Corrective Maintenance): όπου πρακτικά το εκάστοτε μηχανικό μέρος επισκευάζεται αφού πρώτα πραγματοποιηθεί μια βλάβη. Σε τέτοιου είδους συντήρηση το σύστημα βρίσκεται εκτός λειτουργίας για αρκετά μεγάλο χρονικό διάστημα μέχρις ότου αντικατασταθεί το μέρος το οποίο υπέστη την βλάβη και απαιτείται πολύ καλή διαχείριση και διαθεσιμότητα των ανταλλακτικών για την άμεση επιδιόρθωση της βλάβης.

- Προληπτική συντήρησή (Preventive Maintenance): όπου αφορά την πιο διαδεδομένη στρατηγική διόρθωσης όπου συνήθως ο κατασκευαστής έχει μελετήσει εκτενώς τις ανάγκες του συστήματος για συντήρηση και έχει προγραμματίσει εκ των πρότερων συγκεκριμένα χρονικά διαστήματα συντήρησης. Η συγκεκριμένη στρατηγική προστατεύει αρκετά καλά το μεγαλύτερο μέρος του συστήματος από απροσδόκητες βλάβες.

- Προβλεπτική συντήρησή (Predictive Maintenance): η προβλεπτική συντήρηση αποτελεί την πλέον σύγχρονη προσέγγιση συντήρησης όπου με την βοήθεια ενός συστήματος παρακολουθήσης του μηχανικού συστήματος εντοπίζεται η ανάγκη για συντήρηση του συστήματος. Επιπροσθέτως με την χρήση προβλεπτικής συντήρησης μπορούν να αποφευχθούν καταστροφικές βλάβες που μπορεί να παρουσιαστούν σε ένα σύστημα αφού ουσιαστικά εντοπίζεται το αίτιο πριν την παρουσίαση της βλάβης.

Η σωστή συντήρηση συστημάτων απαιτεί τον συνδυασμό όλων τον παραπάνω τεχνικών. Η επιλογή τεχνικής εξαρτάται από το μέρος του συστήματος που πρόκειται να συντηρηθεί

και είναι συνάρτηση της αναγκαιότητας του συγκεκριμένου μέρους αλλά των επιπτώσεων σε περίπτωση βλάβης του. Έτσι, η σωστή επιλογή τεχνικών συντήρησης βρίσκεται στην καρδιά της αποτελεσματικότητας και της ασφάλειας σε βιομηχανικά συστήματα.

Η μηχανική μάθηση έρχεται να βοηθήσει στην προβλεπτική συντήρηση καθώς η δυσκολία στην επιτήρηση πολύπλοκων συστημάτων αυξάνεται με μη γραμμικό τρόπο με την αύξηση των μερών τα οποία υφίστανται παρακολούθηση. Οι τρόποι με τους οποίους προσεγγίζεται το πρόβλημα φαίνονται παρακάτω.

- Ταξινόμηση (Classification) : όπου γίνεται ταξινόμηση της εισόδου στις εκάστοτε κατηγορίες (συνήθως δυαδική ταξινόμηση σε περίπτωσή βλάβης ή όχι) για την ανίχνευση βλάβης στα μέρη του συστήματος. Η συγκεκριμένη προσέγγιση προϋποθέτει μεγάλο όγκο ετικετών εκπαίδευσης στα δεδομένα που στην πράξη είναι ιδιαίτερα δύσκολο.

- Παλινδρόμηση (Regression) : στην συγκεκριμένη προσέγγιση υπάρχουν ορισμένες βλάβες στο σύστημα και το μοντέλο προσπαθεί να προβλέψει τον υπολειπόμενο χρόνο ζωής (remaining useful lifetime (RUL)) του συγκεκριμένου μέρους. Η τεχνική αυτή πάσχει από το γεγονός ότι σε δυο διαφορετικά χρονικά διαστήματα με ίδιες συνθήκες εισόδου το μοντέλο δεν μπορεί να δώσει διαφορετικό χρόνο υπολειπομένου χρόνου ζωής οπότε και δεν μπορεί να συγκλίνει σε μια καλή λύση. Σε αντίθεση, σε περιπτώσεις όπου ορισμένα από τα χαρακτηριστικά εισόδου υφίστανται μετατόπιση στην κατανομή τους (concept drift) η τεχνική αυτή μπορεί να είναι ιδιαίτερα πετυχημένη [23]. Όπως γίνεται κατανοητό, η συγκεκριμένη τεχνική γίνεται αποτελεσματικότερη με την αύξηση των δεδομένων εκπαίδευσης και ιδιαίτερα με την αύξηση των βλάβες στο σύνολο των δεδομένων.

- Επισήμανση απροσδόκητης συμπεριφοράς (Flagging anomalous behaviour): η συγκεκριμένη τεχνική έχει ως στόχο την επισήμανση περιοχών όπου τα δεδομένα δεν ακολουθούν την κατανομή από την οποία παράγονται τα δεδομένα με σκοπό την επίβλεψη του συστήματος. Περιοχές με συνεχόμενες ανωμαλίες μπορεί να προμηνύουν κάποια επερχόμενη βλάβη. Στην συνέχεια, είναι εφικτό να συγκρίνουμε τις συγκεκριμένες καταστάσεις με ιστορικές βλάβες του συστήματος και έτσι να γίνει μια πρόωρη ανίχνευση. Στην συγκεκριμένη τεχνική είναι απαραίτητη η παρακολούθηση και η διάγνωση της κατάστασης απο κάποιον ειδικό και το πλεονέκτημα της είναι η αυτόματη επισήμανση των περιοχών [20]. Τέλος, η συγκεκριμένη τεχνική είναι εφαρμόσιμη σε περιπτώσεις όπου το πλήθος των βλαβών στο σύνολο δεδομένων εκπαίδευσης είναι μικρό ή δεν περιγράφει πλήρως τις πιθανώς απροσδόκητες συμπεριφορές στο σύστημα.

Κεφάλαιο 3

Θεωρητικό υπόβαθρο

3.1 Εισαγωγή

Ανίχνευση ανωμαλιών καλείται η αναγνώριση προτύπων από ένα σύνολο δεδομένων που εμφανίζουν διαφορετική συμπεριφορά από την προσδοκώμενη. Άλλες ονομασίες που συναντούνται στην βιβλιογραφία για τα πρότυπα είναι ακραίες τιμές, ασύμφωνες παρατηρήσεις, εξαιρέσεις, παρεκκλίσεις, ιδιαιτερότητες και προσμίξεις σε διαφορετικούς τομείς εφαρμογών [7].

Αν θεωρήσουμε ένα σύνολο δεδομένο X η κατανομή των δεδομένων συντελείται από καθαρά δεδομένα και ανωμαλίες όπως φαίνεται παρακάτω:

$$p_{full}(x, y) \sim p(y = 1)p(x|y = 1) + p(y = 0)p(x|y = 0)$$

$$p_{normal}(x) \sim p(x|y = 0)$$

$$p_{abnormal}(x) \sim p(x|y = 1)$$

Το πρόβλημα της ανίχνευσης ανωμαλιών αναφέρεται ουσιαστικά στην εκτίμηση των δυο κατανομών $p_{normal}(x), p_{abnormal}(x)$.

Οι λόγοι που μπορούν να δημιουργηθούν ανωμαλίες στα δεδομένα είναι πολλοί καθώς τα σύγχρονα συστήματα συλλογής είναι πολύπλοκα και αποτελούνται από πολλά ανεξάρτητα μέρη. Κάθε ένας από τους λόγους έχει διαφορετική σημασία και η ανίχνευση και η κατηγοριοποίηση παίζουν σημαντικό ρόλο στην επίτευξη ενός σωστού πληροφοριακού συστήματος.

Μερικοί από τους πιο συχνούς λόγους φαίνονται παρακάτω:

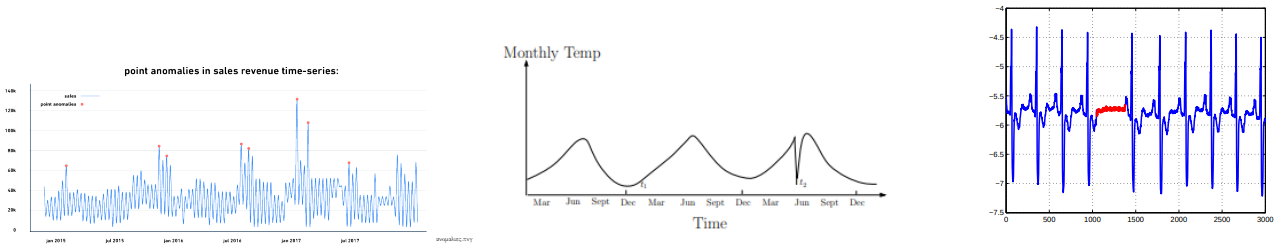
- Φυσικά αίτια , υπό την έννοια ότι οι ανωμαλίες όντως υπάρχουν στο δείγμα.
- Ανθρώπινα λάθη, κυρίως όταν τα δεδομένα υποβάλλονται από ανθρώπους
- Σφάλμα αισθητήρων.
- Σφάλματα λόγω λογισμικού συλλογής δεδομένων.

3.2 Κατηγοριοποίηση Ανωμαλιών σε δεδομένα

Οι Ανωμαλίες στα δεδομένα μπορούν να κατηγοριοποιηθούν στις παρακάτω τρεις βασικές κατηγορίες:

- Στιγμιαίες ανωμαλίες (Point Anomalies): Εδώ τα δεδομένα διαφέρουν διακριτά από την την κατανομή των δεδομένων.
- Ανωμαλίες υπό συνθήκες (Contextual Anomalies): Αφορά δεδομένα που διαφέρουν διακριτά από τα γειτονικά τους ή θα έπρεπε να είχαν άλλη τιμή δεδομένων των συνθηκών.
- Συλλογικές Ανωμαλίες (Collective Anomalies): Εκεί που μια συλλογή από δεδομένα εμφανίζουν απροσδόκητη συμπεριφορά.

Οι διαφορετικές κατηγορίες ανωμαλιών στα δεδομένα παρουσιάζονται γραφικά παρακάτω.



(α′) Στιγμιαίες ανωμαλίες (β′) Ανωμαλίες υπό συνθήκες (γ′) Συλλογικός Ανωμαλίες

Σχήμα 3.1: Κατηγοριοποίηση Ανωμαλιών

3.3 Μέθοδοι μάθησης για τον εντοπισμό ανωμαλιών σε δεδομένα

Όπως και σε οποιοδήποτε πρόβλημα μηχανικής μάθησης οι μέθοδοι μάθησης για την προσέγγιση του προβλήματος εντοπισμού ανωμαλιών σε δεδομένα εξαρτάται σε μεγάλο βαθμό από τον τύπο των δεδομένων που καλούμαστε να επεξεργαστούμε. Ένα από τα πιο σημαντικά χαρακτηριστικά που πρέπει να λαμβάνονται υπόψιν πριν να αρχίσουμε να προσεγγίζουμε το πρόβλημα είναι το αν υπάρχει κάποιου είδους επισημείωση στα δεδομένα που θα χρησιμοποιηθούν για την εκπαίδευση του μοντέλου.

Έτσι από την σκοπιά των μεθόδων μάθησης έχουμε τις εξής 3 κατηγορίες.

- Επιβλεπόμενη μάθηση (Supervised learning).
- Ήμι-Επιβλεπόμενη μάθηση (Semi-supervised learning).
- Μη Επιβλεπόμενη μάθηση (Unsupervised learning).

3.3.1 Επιβλεπόμενη μάθηση

Η Επιβλεπόμενη μάθηση είναι μια κατηγορία μηχανικής μάθησης, στόχος της οποίας είναι ο χαρακτηρισμός δεδομένων με βάση κάποια δεδομένα εκπαίδευσης. Σε αυτήν την κατηγορία το σύνολο των δεδομένων εκπαίδευσης αποτελείται από το σύνολο εισόδου X μαζί με το σύνολο εξόδου Y. Στόχος είναι η αντιστοίχηση μέσω μιας απεικόνισης από σύνολο εισόδου στο σύνολο εξόδου. Γνωρίζοντας τα δεδομένα εξόδου Y το πρόβλημα εντοπισμού ανωμαλιών στα δεδομένα αποτελεί ένα πρόβλημα κατηγοριοποίησης μέσω μιας διαχωριστικής επιφάνειας στον χώρο εισόδου έτσι ώστε τα δεδομένα να διαχωρίζονται σε προσδοκώμενα και μη δεδομένα. Έτσι τα διαθέσιμα σύνολα $X_{train,test}$ περιέχουν και τα δυο τις κατανομές $p_{normal}(x)$, $p_{abnormal}(x)$.
Το μεγαλύτερο πλεονέκτημα σε αυτήν την κατηγορία μάθησης είναι ότι μπορούμε να αξιολογήσουμε τα αποτελέσματα των μοντέλων καθώς επίσης και να εφαρμόσουμε οποιαδήποτε μέθοδο μάθησης. Τα προβλήματα τα οποία προκύπτουν είναι ότι είναι δύσκολο συχνά να βρούμε επισημειωμένα δεδομένα η να τα επισημάνουμε μόνοι μας.

3.3.2 Μη Επιβλεπόμενη μάθηση

Στο πρόβλημα της μη επιβλεπόμενης μάθησης δεν γνωρίζουμε το οτιδήποτε για τα δεδομένα εκπαίδευσης. Στόχος συνήθως είναι η κατασκευή ενός μοντέλου που μαθαίνει την κατανομή των δεδομένων εισόδου επιλύοντας ένα δεύτερο πρόβλημα. Με αυτόν τον τρόπο στο πρόβλημα ανίχνευσης ανωμαλιών στιγμιότυπα εισόδου που απέχουν από την κατανομή εισόδου του μοντέλου χαρακτηρίζονται ως ανωμαλίες. Ένα βασικό σημείο είναι ότι δεν γνωρίζουμε εκ των προτέρων πια από τα δεδομένα εισόδου είναι τα αναμενόμενα, οπότε γίνεται η υπόθεση ότι πιθανές ανωμαλίες στα δεδομένα εισόδου υστερούν αριθμητικά έναντι των καθαρών δεδομένων οπότε το σύστημα δεν είναι ικανό να τα μοντελοποιήσει. Αντιθέτως, το μοντέλο θα αναγνωρίζει πρότυπα όπου έχουν μεγαλύτερη συχνότητα. Η μη επιβλεπόμενη μάθηση αποτελεί ίσως την πλέον ενδιαφέρουσα κατηγορία μάθησης αφού στα περισσότερα προβλήματα του πραγματικού κόσμου τα δεδομένα είναι μη επισημειωμένα.
Περισσότερο φορμαλιστικά σε πρόβλημα μη επιβλεπόμενης μάθησης έχουμε στο σύνολο εκπαίδευσης ισχύει ότι $X_{train} \sim p_{full}$ ενώ το σύνολο στο έλεγχου $X_{test} \sim p_{full}$.

3.3.3 Ήμι-Επιβλεπόμενη μάθηση

Προβλήματα ήμι-επιβλεπόμενης μάθησης είναι εκείνα όπου τα μοντέλα κάνουν χρήση επισημειωμένων και μη δεδομένων. Στην πληθώρα των περιπτώσεων έχουμε μεγαλύτερο όγκο μη επισημειωμένων δεδομένων και άρα δεν μπορούμε να εφαρμόσουμε τεχνικές επιβλεπόμενης μάθησης. Στο πρόβλημα του εντοπισμού ανωμαλιών συνήθως γνωρίζουμε την κατανομή των προσδοκώμενων δεδομένων οπότε και εφαρμόζουμε μοντέλα που προσπαθούν να μάθουν αυτήν την κατανομή. Όταν κάποιο στιγμιότυπο δεν ανήκει στην κατανομή των δεδομένων εισόδου τότε χαρακτηρίζεται σαν ανωμαλία.

Έτσι, στο συγκεκριμένο τρόπο μάθησης έχουμε στην διάθεσή μας ένα σύνολο $X_{train} \sim$

p_{normal} και ένα σύνολο έλεγχου $X_{test} \sim p_{full}$.

Το πλεονέκτημα αυτής της μεθόδου είναι ότι δεν χρειάζεται να κάνουμε υποθέσεις για την κατανομή εισόδου οπότε στην γενική περίπτωση δουλεύουν καλύτερα από μοντέλα μη επιβλεπόμενης μάθησης. Ένα από τα μεγαλύτερα προβλήματα αυτής της προσέγγισης είναι ότι υπάρχει περίπτωση η κατανομή των προσδοκώμενων δεδομένων να μην είναι πλήρης. Έτσι συχνά οδηγούμαστε σε λάθος ταξινόμηση κανονικών δεδομένων επειδή δεν υπήρχαν στην κατανομή εισόδου.

	Μέθοδοι Μηχανικής μάθησης	Normal Label	Anomaly Labels	Χαρακτηριστικά μεθόδου
1	Επιβλεπόμενη Μάθηση	✓	✓	• Κατηγοριοποίηση • Δυσανάλογες κλάσεις
2	Ημι Επιβλεπόμενη Μάθηση	✓	✗	• Εκπαίδευση σε κανονικά δεδομένα • Δύσκολη ερμηνεία αποτελεσμάτων
3	Μη επιβλεπόμενη Μάθηση	✗	✗	• Υπόθεση: Μεγάλος αριθμός κανονικών δεδομένων • Δύσκολη ερμηνεία

Σχήμα 3.2: Μέθοδοι μάθησης για τον εντοπισμό ανωμαλιών σε δεδομένα

3.4 Τεχνικές εντοπισμού ανωμαλιών στα δεδομένα

Σ' αυτήν την ενότητα θα παρουσιάσουμε τις βασικές τεχνικές μηχανικής μάθησης για την ανίχνευση ανωμαλιών. Η επιλογή των κατηγοριών είναι αρκετά γενική και υπάρχουν υποκατηγορίες τεχνικών ενώ μπορεί ορισμένοι αλγόριθμοι να κάνουν χρήση ιδεών από περισσότερες από μια τεχνικές. Επίσης, μια πιο λεπτομερής ανάλυση έγινε στις στατιστικές μεθόδους, γιατί αυτή αποτελεί την βάση για τους περισσοτέρους αλγορίθμους που υλοποιήθηκαν κατά την εξέλιξη της διπλωματικής εργασίας.

3.4.1 Κατηγοριοποίηση

Η κατηγοριοποίηση (Classification) είναι μια από τις πιο γνωστές τεχνικές επιβλεπόμενης μάθησης. Αφορά ουσιαστικά την διαδικασία κατασκευής ενός μοντέλου ικανού να διαχωρίζει τα δεδομένα εισόδου σε κατηγορίες κάνοντας χρήση των επισημασμένων δεδομένων εκπαίδευσης. Η υπόθεση στην οποία βασίζεται η τεχνική της κατηγοριοποίησης είναι ότι ανώμαλα και κανονικά δεδομένα είναι γραμμικώς διαχωρίσιμα σε κάποιον χώρο. Το μοντέλο μπορεί να έχει δυαδική έξοδο στην περίπτωση που έχουμε μια κλάση εξόδου η ένα διάνυσμα από δυαδικές τιμές στην περίπτωση των πολλαπλών κλάσεων.

Ένα πρόβλημα που προκύπτει είναι, ότι τις περισσότερες φορές η αναλογία μεταξύ των κατηγοριών είναι δυσανάλογη και πρέπει να προβούμε σε διαδικασίες εξομάλυνσης των δεδομένων εισόδου. Έτσι, είναι σύνηθες σε προβλήματα εντοπισμού ανωμαλιών να καταλήγουμε σε ποσοστά 99 προς 1.

3.4.2 Συσταδοποίηση

Η τεχνική της συσταδοποίησης (Clustering) αφορά την δημιουργία ομάδων στο σύνολο δεδομένων. Η υπόθεση της συσταδοποίησης στο πρόβλημα της ανίχνευσης ανωμαλιών είναι ότι κανονικά δεδομένα βρίσκονται κοντά στον χώρο δημιουργώντας καλοσχηματισμένες συστάδες ενώ τα ανώμαλα δεδομένα είτε είναι εντελώς απομονωμένα είτε ανήκουν σε περισσότερο αραιές συστάδες η μικρές ομάδες. Η συγκεκριμένη τεχνική κάνει χρήση αλγορίθμων μη επιβλεπόμενης μάθησης και έχει πολλές παραλλαγές στον τρόπο προσέγγισης για την επίλυση του προβλήματος.

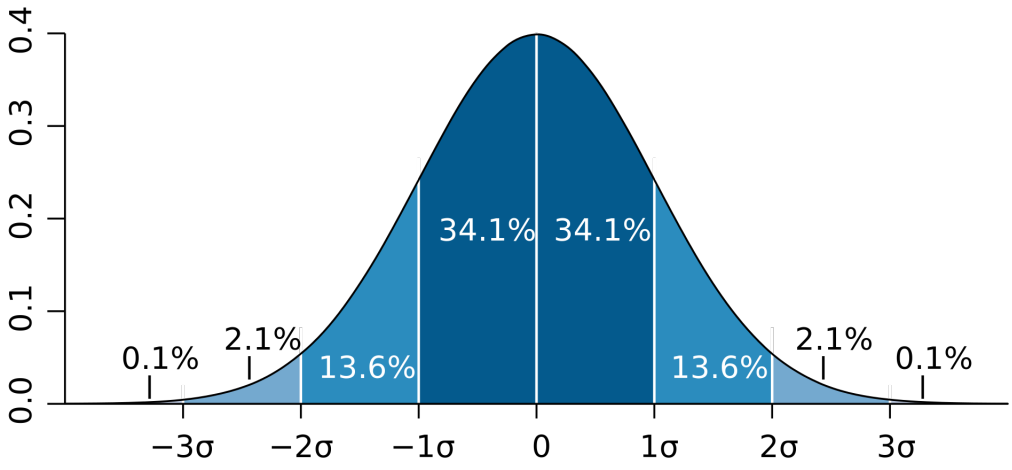
Ένα πρόβλημα που υπάρχει σε αυτήν την κατηγορία είναι ότι πολλές φορές ανωμαλίες από δεδομένα μπορούν να υπάρχουν ανάμεσα σε μια πυκνή συστάδα από σημεία και έτσι να είναι αδύνατο να εντοπισθούν. Επίσης, η τεχνική αυτή κληρονομεί όλα τα πλεονεκτήματα και τα μειονεκτήματα της μη επιβλεπόμενης μάθησης.

3.4.3 Στατιστικές μέθοδοι

Οι στατιστικές μέθοδοι ανίχνευσης ανωμαλιών κάνουν χρήση ενός στατιστικού κριτηρίου για την ανίχνευση ανωμαλιών. Η πιο συχνή προσέγγιση είναι η μοντελοποίηση της κατανομής των κανονικών δεδομένων και στην συνέχεια η εφαρμογή ενός τεστ ελέγχου υποθέσεων για την κατηγοριοποίηση των δεδομένων. Η συγκεκριμένη κατηγορία είναι πολύ γενική υπό την έννοια ότι η μοντελοποίηση της κατανομής των δεδομένων μπορεί να γίνει από πολλών ειδών μοντέλα με τις ανάλογες υποθέσεις κάθε φορά. Επίσης είναι πολύ συχνό να εφαρμόζεται σε κάποιο ενδιάμεσο στάδιο ενός αλγορίθμου ανίχνευσης ανωμαλιών ή για παράδειγμα στις εξόδους ενός μοντέλου παλινδρόμησης. Τα μοντέλα που χρησιμοποιεί η συγκεκριμένη τεχνική μπορεί να είναι παραμετρικά ή μη.

Η πιο απλή στατιστική μέθοδος είναι αυτή της θεώρησης ότι τα δεδομένα ακολουθούν κανονική κατανομή με παραμέτρους που υπολογίζονται με την μέθοδο της μέγιστης πιθανοφάνειας. Στην συνέχεια μια τιμή που δείχνει ποσό πιθανό είναι ένα δείγμα να είναι εκτός κατανομής, είναι η απόσταση του δείγματος από την μέση τιμή κανονικοποιημένη με την τυπική απόκλιση του συνόλου των δεδομένων. Το σχήμα 3.3 δείχνει τα ποσοστά για τις αντίστοιχες αποστάσεις από την μέση τιμή. Άλλες παρόμοιες μέθοδοι με την θεώρηση της μονοδιάστατης κανονικής κατανομής είναι το Grubb's test για τον εντοπισμό ενός μόνο δείγματος και Tietjen-Moore test, generalized extreme studentized deviate test στον εντοπισμό πολλών δειγμάτων με συγκεκριμένο αριθμό ανωμαλιών ή ένα άνω φράγμα από αριθμό ανωμαλιών στο σύνολο των δεδομένων αντίστοιχα.

Στην περίπτωση της πολυδιάστατης κανονικής κατανομής γίνεται χρήση της Mahanobis απόστασης που άφορα την απόσταση ενός σημείου $\vec{x} = (x_1, x_2, x_3, \ldots, x_N)^T$ και μιας πολυ-



Σχήμα 3.3: Κανονική κατανομή

διάστατης κατανομής D με μέση τιμή $\vec{\mu} = (\mu_1, \mu_2, \mu_3, \ldots, \mu_N)^T$ και πίνακα συνδιακυμάνσεων S.

$$D_M(\vec{x}) = \sqrt{(\vec{x} - \vec{\mu})^T S^{-1} (\vec{x} - \vec{\mu})}.$$

Στην συνέχεια μπορεί να υπολογιστεί με παρόμοιο τρόπο μια τιμή είτε εμπειρική από το σύνολο των δειγμάτων είτε θεωρητική για την ανίχνευση ανωμαλιών στο σύνολο των δεδομένων. Η θεωρητική τιμή προέρχεται από το γεγονός ότι οι Mahanobis αποστάσεις των σημείων μιας κανονικής κατανομής ακολουθεί κατανομή $\chi 2$ οπότε και οι ανάλογες κρίσιμες τιμές μπορούν να βρεθούν εύκολα.

3.5 Αλγόριθμοι μηχανικής μάθησης για τον εντοπισμό ανωμαλιών στα δεδομένα

Μερικοί από τους πιο διαδεδομένους αλγορίθμους και μοντέλα για ανίχνευση ανωμαλιών φαίνονται στην παρακάτω λίστα.

- Grubbs Test - ESD Test.
- Bayesian Outlier Detection.
- Local Outlier Factor.
- Isolation Forest.
- SVM.
- Autoencoders.
- VAE.

- Adversarial Autoencoders.
- LSTM.
- Gans.

Από τους παραπάνω τρόπους ανίχνευσης ανωμαλιών υλοποιήθηκαν οι Local Outlier Factor, Isolation Forest, Autoencoders, LSTM, Adversarial Autoencoders. Τα αποτελέσματα από τους αλγορίθμους Local Outlier Factor, Isolation Forest ήταν λιγότερο ικανοποιητικά και προέκυπτε αδυναμία ερμηνείας των αποτελεσμάτων τους. Σε αντίθεση, τα μοντέλα Autoencoders, LSTM, Adversarial Autoencoders παρουσιάζονται στην πειραματική ανάλυση του κεφαλαίου 6 καθώς τα αποτελέσματα τους ήταν περισσότερο ικανοποιητικά.

3.5.1 Τοπικός παράγοντας απόκλισης LOF

Ο τοπικός παράγοντας απόκλισης (LOF) [5] είναι ένας αλγόριθμος μη επιβλεπόμενης μάθησης ο οποίος κάνει χρήση της τοπικής πυκνότητας των δεδομένων για την ανίχνευση ανωμαλιών. Η βασική ιδέα του αλγορίθμου είναι ότι όσο μικρότερη είναι η πυκνότητα ενός σημείου σχετικά με τους k-γείτονες του τόσο πιο πιθανό το συγκεκριμένο σημείο να αποτελεί ανωμαλία.

Ο αλγόριθμος ορίζει την απόσταση προσβασιμότητας (reachability distance) ενός σημείου A σε ενα σημείο B ως την απόσταση :

$$\text{reachability-distance}_k(A, B) = \max\{\kappa\text{-distance}(B), d(A, B)\}$$

Όπου:

κ-distance(B): Η απόσταση του σημείου B από τον κ μακρινότερο γείτονα του.
$d(A, B)$: Η απόσταση μεταξύ των σημείων A, B

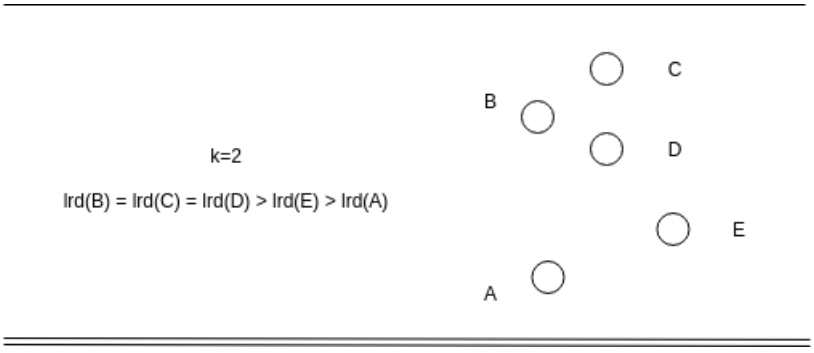
Η παραπάνω απόσταση θεωρεί ίση απόσταση μεταξύ ενός σημείου A που ανήκει στο νέφος κ κοντινότερων γειτόνων του B και ως εκ τούτου δεν αποτελεί απόσταση από μαθηματικής άποψης. Στην συνέχεια ορίζεται η τοπική πυκνότητα προσβασιμότητας (local reachability density) :

$$\text{lrd}_k(A) := 1/\left(\frac{\sum_{B \in N_k(A)} \text{reachability-distance}_k(A, B)}{|N_k(A)|}\right)$$

Όπου ουσιαστικά σημεία που ανήκουν σε ένα πυκνό νέφος μεγέθους και έχουν την ίδια πυκνότητα ενώ ένα σημείο A που δεν ανήκει στο συγκεκριμένο νέφος αλλά έχει γείτονες σημεία του νέφους έχει μικρότερη πυκνότητα όπως φαίνεται στο σχήμα 3.4.

Τέλος, ο τοπικός παράγοντας απόκλισης (LOF) ενός σημείου A υπολογίζεται συγκρίνοντας την τοπική πυκνότητα προσβασιμότητας του σημείου A με αυτές των γειτόνων του $N_k(A)$:

$$\text{LOF}_k(A) := \frac{\sum_{B \in N_k(A)} \frac{\text{lrd}(B)}{\text{lrd}(A)}}{|N_k(A)|} = \frac{\sum_{B \in N_k(A)} \text{lrd}(B)}{|N_k(A)|} / \text{lrd}(A)$$



Σχήμα 3.4: Παράδειγμα οπτικοποίησης τοπικής πυκνότητας προσβασιμότητας

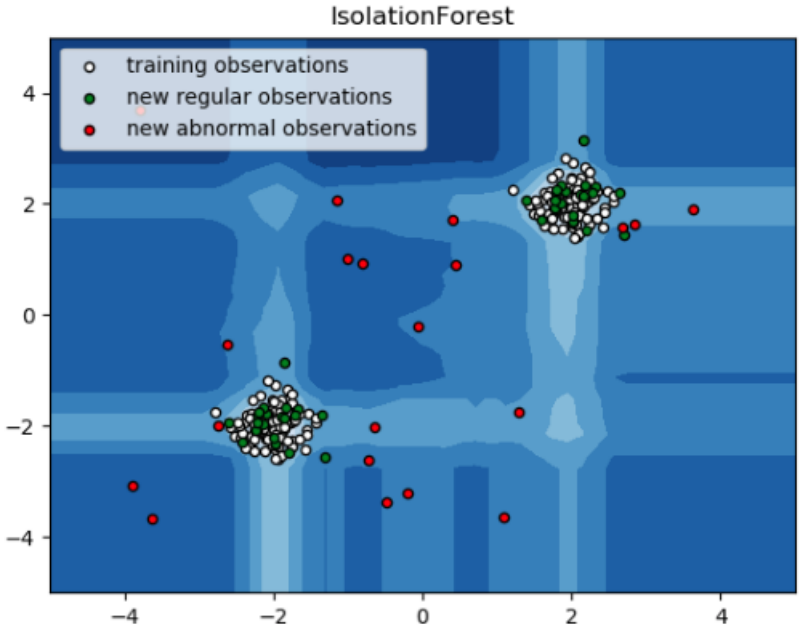
Όσο πιο κοντά στην μονάδα είναι ο παράγοντας LOF υποδηλώνει ότι το συγκεκριμένο σημείο βρίσκεται κοντά σε κάποιο πυκνό νέφος από σημεία ενώ όσο μεγαλώνει ο παράγοντας το σημείο είναι όλο και πιο απομονωμένο. Για Παράδειγμα τα σημεία B,C,D στο σχήμα 3.4 έχουν LOF = 1 ενώ τα σημεία A,E θα έχουν σαφώς μεγαλύτερες τιμές.

Στο συγκεκριμένο αλγόριθμο η παράμετρος κ είναι ιδιαίτερα σημαντική για την ανίχνευση ανωμαλιών.Όσο πιο μικρή είναι η συγκεκριμένη παράμετρος ο αλγόριθμος εστιάζει σε μικρά νέφη ενώ όσο μεγαλύτερη είναι η παράμετρος ο αλγόριθμος επικεντρώνεται σε μεγαλυτέρου όγκου περιοχές. Μια ακόμα παράμετρος c που προκύπτει έμμεσα από το πρόβλημα ανίχνευσης ανωμαλιών αφορά την πρότερη πεποίθησή μας για το ποσοστό των ανώμαλων δεδομένων μέσα στο σύνολο δεδομένων ούτως ώστε να ορίσουμε ένα κατώφλι για την ανίχνευση ανωμαλιών στο σύνολο εκπαίδευσης.

3.5.2 Δάσος Απομόνωσης Isolation Forest

Ο αλγόριθμος Isolation Forest [15] είναι ένας ιδιαίτερα αποδοτικός αλγόριθμος εντοπισμού ανωμαλιών σε δεδομένα. Η βασική ιδέα του αλγορίθμου είναι ότι ανώμαλα σημεία στο σύνολο των δεδομένων διαφέρουν από τα κανονικώς παραγόμενα σημεία και έτσι θα είναι πιο απομονωμένα στο χώρο. Ο αλγόριθμος πρακτικά τέμνει τον χώρο αναδρομικά διαλέγοντας τυχαία ένα χαρακτηριστικό και στην συνέχεια μια τυχαία τιμή διαχωρισμού. Η αναδρομική δομή του αλγορίθμου δημιουργεί μια δεντρική δομή. Στην συνέχεια όσο πιο γρήγορα ένα δείγμα απομονώθηκε από τα υπόλοιπα στο σύνολο δεδομένων τόσο πιο πιθανό είναι να αποτελεί ανωμαλία.

Περισσότερο αναλυτικά ο αλγόριθμος δημιουργεί t δέντρα iT, ώστε να ορίσει ένα δάσος απομόνωσης IF, τα οποία τέμνουν τυχαία τον χώρο. Κάθε ένα από τα δέντρα iT εκπαιδεύεται (δημιουργείται) σε ένα υποσυνόλου ψ του συνόλου δεδομένων $\mathcal{D}$ διαμερίζοντας το. Συγκεκριμένα, επιλέγεται τυχαία ένα χαρακτηριστικό q του συνόλου $\mathcal{D}$ και στην συνέχεια μια τυχαία διαμέριση p τέτοια ώστε $p \in (max_q(\psi), min_q(\psi))$. Αυτή η τυχαία διαμέριση δημιουργεί δυο σύνολα $\mathcal{D}_l$ και $\mathcal{D}_r$ με τα σημεία που είναι μικρότερα και μεγαλύτερα από την τιμή διαμέρισης p αντίστοιχά. Αν κάποιο από τα δυο σύνολα είναι αδιαίρετο δημιουργείται ένας τελικός κόμβος. Στην συνέχεια το δέντρο χρησιμοποιεί τα συγκεκριμένα σύνολα για να δημιουργήσει τους υπολοίπους κόμβους αναδρομικά μέχρι να φτάσει το μέγιστο ύψος.



Σχήμα 3.5: Παράδειγμα οπτικοποίησης δάσους απομόνωσης

Στο στάδιο της αποτίμησης κάθε ένα από τα δεδομένα του συνόλου έλεγχου $\mathcal{D}_{test}$ μεγέθους n διατρέχουν τα δέντρα iT και υπολογίζεται το μήκος $h(x)$ της διαδρομής μέχρις ότου φτάσουν σε τερματικό κόμβο. Έπειτα υπολογίζεται ένα σκορ s για το πόσο πιθανό είναι ένα δείγμα να είναι ανώμαλο:

$$s(x,n) = 2^{\frac{-E(h(x))}{c(n)}}$$

Όπου:

$E(h(x))$: Το μέσο μήκος διαδρομής των t δέντρων .

$c(n)$: Το μέσο ύψος ενός δέντρου δυαδικής αναζήτησης BST με n δείγματα ως παράγοντας κανονικοποίησης.

Η διαδικασία της εκπαίδευσης έχει πολυπλοκότητα $O(t\psi \log \psi)$ ενώ αποτίμησης έχει πολυπλοκότητα $O(nt \log \psi)$.

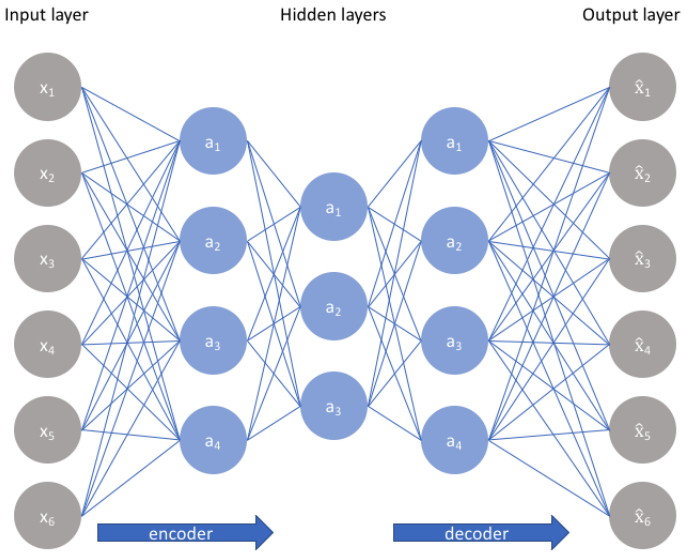
Για την ανίχνευση ανωμαλιών μια επιλογή είναι να ταξινομήσουμε τα σκορ s των δειγμάτων μας σε φθίνουσα σειρά και να επιλέξουμε τα m μεγαλύτερα. Μια δεύτερη επιλογή είναι να κάνουμε χρήση των δειγμάτων εκπαίδευσης ώστε να ορίσουμε ένα κατώφλι thr, ανάλογα με την πρότερη πεποίθησή μας για ανώμαλα δεδομένα, το όποιο θα χρησιμοποιήσουμε στο σύνολο δοκιμών.

3.6 Αλγόριθμοι για τον εντοπισμό ανωμαλιών με χρήση βαθιών νευρωνικών δικτύων

Τα νευρωνικά δίκτυα την τελευταία δεκαετία συγκεντρώνουν όλο και περισσότερο το ενδιαφέρον ερευνητών καθώς όλο ένα και περισσότερες εφαρμογές πετυχαίνουν συναρπαστικά αποτελέσματα. Σε ότι αφορά το πρόβλημα της ανίχνευσης ανωμαλιών τα νευρωνικά δίκτυα έχουν ιδιαίτερο ενδιαφέρον καθώς οι περισσότερες από τις νέες έρευνες επικεντρώνονται στην επίλυση του προβλήματος της μη επιβλεπόμενης ανίχνευσης ανωμαλιών. Τις περισσότερες φορές τα νευρωνικά δίκτυα χρησιμοποιούν στατιστικές μεθόδους όπως αυτές περιγράφηκαν στην παράγραφο 3.4.3 στα σφάλματα του διανύσματος εισόδου, ενώ παράλληλα τεχνικές μη επιβλέπομενης συσταδοποίησης είναι εφικτές και έχουν ιδιαίτερο ενδιαφέρον. Τα πιο γνωστά μοντέλα τα οποία χρησιμοποιούνταν για την ανίχνευση ανωμαλιών είναι τα Autoencoder, VAE, Adversarial Autoencoder, Gan, LSTM. Από τα παραπάνω παρουσιάζονται τα Autoencoder, LSTM, Adversarial Autoencoder.

3.6.1 Αυτοκωδικοποιητής -Autoencoder

Ο Αυτοκωδικοποιητής Autoencoder [8] είναι ένα νευρωνικό δίκτυο το οποίο προσπαθεί να αναπαράγει την είσοδο $x \in X = \mathbb{R}^n$ στην έξοδο $x' \in X$. Ο αυτοκωδικοποιητής αποτελείται από δυο δίκτυα έναν κωδικοποιητή $E : X \to Z$ ο οποίος προβάλλει την είσοδο σε έναν χώρο προβολής Z, και έναν αποκωδικοποιητή $D : Z \to X$ ο οποίος προσπαθεί να ανακατασκευάσει την είσοδο. Κάθε έναν από τα δυο δίκτυα αποτελεί ένα νευρωνικό δίκτυο με m κρυφά επίπεδα συνήθως ίδια στον αριθμό, ενώ ο κωδικοποιητής προβάλλει την είσοδο σε ένα χώρο Z συνήθως μικρότερης διάστασης από αυτήν της εισόδου X όπως φαίνεται στο σχήμα 3.6. Ο αυτοκωδικοποιητής εκπαιδεύεται προσπαθώντας να ελαχιστοποιήσει το λάθος ανακατασκευής $L(x, D(E(x)))$ όπου συνηθίζεται να είναι το μέσο τετραγωνικό σφάλμα.



Σχήμα 3.6: Διάγραμμα Αποκωδικοποιητή

Ο αυτοκωδικοποιητής πρέπει να σχεδιάζεται ώστε να μην μπορεί να ανακατασκευάζει ακριβώς την είσοδο στην έξοδο στο σύνολο των δεδομένων εκπαίδευσης. Έτσι για να μην επιτρέπεται στον αυτοκωδικοποιητή να αναπαράγει την είσοδο είναι σημαντικό ο χώρος προβολής Z του κωδικοποιητή να έχει μικρότερη διάσταση όπως προαναφέρθηκε. Με αυτόν τον τρόπο στην προσπάθειά του το δίκτυο να ανακατασκευάσει την είσοδο μαθαίνει χρήσιμες σχέσεις μεταξύ των χαρακτηριστικών του συνόλου των δεδομένων. Ορισμένες από τις εφαρμογές στις οποίες είναι χρήσιμα τέτοιου είδους δίκτυα είναι η αποθορυβοποίηση σημάτων, ανακατασκευή αλλοιωμένων εικόνων ενώ πολύ συχνά αποτελούν την βάση για την δημιουργία περισσότερο συνθέτων δικτύων κυρίως σε εφαρμογές μην επιβλεπόμενης μάθησης.

Λόγω της ικανότητάς τους να ανακαλύπτουν συσχετίσεις μεταξύ των δεδομένων οι αυτοκωδικοποιητές είναι ιδιαίτερα χρήσιμοι στην ανίχνευση ανωμαλιών. Η συνήθης πρακτική που χρησιμοποιείται είναι η εκπαίδευση του αυτοκωδικοποιητή σε καθαρά δεδομένα ώστε να μοντελοποιηθεί η κατανομή των δεδομένων. Ύστερα μπορεί να οριστεί ένα κατώφλι απόφασης T_{thr} με βάση την κατανομή των καθαρών δεδομένων έτσι ώστε τα δεδομένα που πρόκειται να ελεγχθούν από το δίκτυο να κατηγοριοποιούνται σαν ανώμαλα αν έχουν σφάλμα μεγαλύτερο από το συγκεκριμένο κατώφλι T_{thr}.

Προβλήματα του αυτοκωδικοποιητή

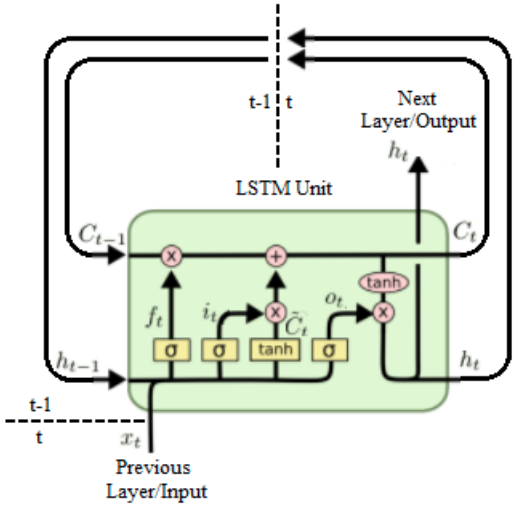
Το μεγαλύτερο πρόβλημα που αντιμετωπίζουν τέτοιου είδους μοντέλα είναι ότι σε περίπτωση που δεν έχουμε στην διάθεσή μας την γνώση για το πια δεδομένα είναι καθαρά ο αυτοκωδικοποιητής θα πρέπει να εκπαιδεύεται σε όλη την κατανομή των δεδομένων $p_{full}(x, y)$ με συνέπεια να υπερεκπαιδεύεται σε ανώμαλα δεδομένα. Η υπόθεσή είναι ότι το δίκτυο θα μάθει να ανακατασκευάζει καλύτερα τα δείγματα με βάση την συχνότητα. Έτσι, είναι αναγκαίο να εφαρμόζονται τεχνικές ώστε ο αυτοκωδικοποιητής να μην υπερεκπαιδεύεται όπως μείωση της χωρητικότητας του μοντέλου. Ένα δεύτερο πρόβλημα είναι, ότι ο χώρος προβολής του αυτοκωδικοποιητή δεν είναι κάλως σχηματισμένος με αποτέλεσμα να μην μπορεί να γίνει ερμηνεία όσον αφορά την ανίχνευση ανωμαλιών.

Για την λύση του συγκεκριμένου προβλήματος έρευνες προσεγγίζουν από διαφορετική οπτική σκοπιά. Μερικές από τις πιο ελπιδοφόρες προσεγγίσεις είναι αυτές της ενεργής μάθησης και της μη επιβλεπόμενης συσταδοποίησης των δεδομένων μέσω των ανταγωνιστικών αυτοκωδικοποιητών οι οποίες αναλύονται στην συνέχεια.

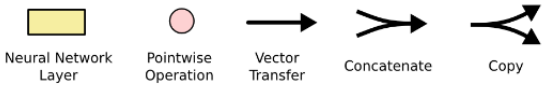
3.6.2 Ανατροφοδοτούμενα νευρωνικά δίκτυα LSTM

Τα ανατροφοδοτούμενα νευρωνικά δίκτυα είναι δίκτυα τα οποία προσπαθούν να μοντελοποιήσουν ακολουθίες δεδομένων. Τα συγκεκριμένα δίκτυα έχουν ιδιαίτερη επιτυχία στην μοντελοποίηση κειμένων όπου η ακολουθία αποτελεί προτάσεις ή φράσεις λέξεων καθώς ακόμα και στην μοντελοποίηση οποιουδήποτε είδους χρονοσειράς. Ένα είδος ανατροφοδοτούμενου νευρωνικού δικτύου αποτελεί το $LSTM$ το οποίο φαίνεται στο παρακάτω σχήμα.

Περισσότερο αναλυτικά το $LSTM$ αποτελείται από μονάδες οι οποίες ανατροφοδοτούνται με τα αποτελέσματα των προηγουμένων μονάδων καθώς επίσης από τις μεμονωμένες εισόδους



Σχήμα 3.7: Διάγραμμα $LSTM$



Σχήμα 3.8: Διάγραμμα τελεστών $LSTM$ 2

της ακολουθίας. Η πληροφορία σχετικά με τις προηγούμενες εισόδους περνά στις μεταγενέστερες μονάδες μέσω του διανύσματος C_t. Αναλυτικότερα κάθε μονάδα αποτελείται από τις πύλες f_t, i_t, o_t. Η πρώτη πύλη f_t ονομάζεται πύλη απώλειας μνήμης και είναι υπεύθυνη για την πληροφορία η οποία θα διατηρηθεί από την προηγουμένη μονάδα στην παρούσα. Αυτό γίνεται, πολλαπλασιάζοντας την έξοδο της συγκεκριμένης πύλης με το διάνυσμα C_{t-1}. Στην συνέχεια, την πληροφορία που πρόκειται να περάσει στην επομένη μονάδα έρχεται να συμπληρώσει η πύλη εισόδου i_t. Τέλος, η πύλη εξόδου o_t είναι υπεύθυνη για την τιμή του διανύσματος h_t το οποίο αποτελεί ουσιαστικά και την πρόβλεψη ολόκληρης της μονάδας και η οποία χρησιμοποιείται στην είσοδο των πυλών τις επομένης μονάδας. Περισσότερο φορμαλιστικά οι εξισώσεις κάθε μονάδας φαίνονται παρακάτω:

$$f_t = \sigma_g(W_f x_t + U_f h_{t-1} + b_f)$$

$$i_t = \sigma_g(W_i x_t + U_i h_{t-1} + b_i)$$

$$o_t = \sigma_g(W_o x_t + U_o h_{t-1} + b_o)$$

$$c_t = f_t \circ c_{t-1} + i_t \circ \sigma_c(W_c x_t + U_c h_{t-1} + b_c)$$

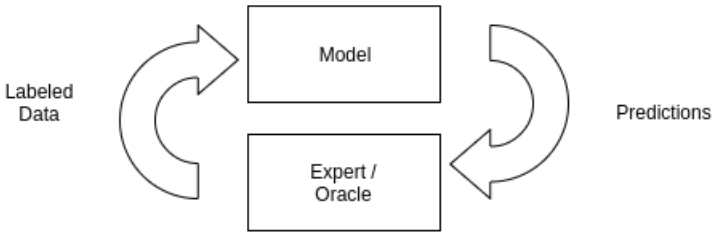
$$h_t = o_t \circ \sigma_h(c_t)$$

Ανίχνευση ανωμαλιών με χρήση Ανατροφοδοτούμενων νευρωνικών δικτύων

Για την ανίχνευση ανωμαλιών με την χρήση $LSTM$ [17] προσπαθούμε να προβλέψουμε την τιμή του διανύσματος $X_o = \{x_t, x_{t1}, x_{t+2}, \ldots, x_{t+n}\}$ για τις n χρονικές στιγμές από t μέχρι $t+n$ με χρήση d χρονικών στιγμών στο παρελθόν $X_i = \{x_{t-d}, x_{t-d+1}, x_{t-d+2}, \ldots, x_{t-1}\}$. Στην συνέχεια μπορεί να ακολουθηθεί μια διαδικασία παρόμοια με αυτές που περιγράφηκαν παραπάνω για την περίπτωση του αυτοκωδικοποιητή ώστε μέσω του λάθους ανακατασκευής του δικτύου να γίνεται ανίχνευση ανωμαλιών. Το συγκεκριμένο δίκτυο είναι ιδιαίτερα χρήσιμο καθώς λόγω της αρχιτεκτονικής του μας επιτρέπει να εντοπίζουμε ανωμαλίες υπό συνθήκη.

3.6.3 Ενεργή μάθηση - Active learning

Ενεργή μάθηση (Active learning) είναι μια εναλλακτική διαδικασία μάθησης στην περίπτωση όπου δεν έχουμε στην διάθεσή μας ένα πλήρες επισημειωμένο σύνολο δεδομένων. Η διαδικασία της μάθησης χωρίζεται σε βήματά όπου το μοντέλο αλληλεπιδρά με ένα σύστημα εξωτερικής γνώσης για την βελτίωση των αποτελεσμάτων του. Το σύστημα εξωτερικής γνώσης μπορεί να είναι άνθρωπος ικανός να επισημειώνει τα αποτέλεσμα του μοντέλου η οποιοδήποτε σύστημα που μπορεί να κρίνει τα αποτελέσματα, όπως για παράδειγμα ένα δεύτερο μοντέλο. Έπειτα, το δίκτυο συνεχίζει την εκπαίδευση με την πρόσθετη γνώση και η διαδικασία συνεχίζεται για ορισμένα βήματα.



Σχήμα 3.9: Διάγραμμα Ενεργής μάθησης

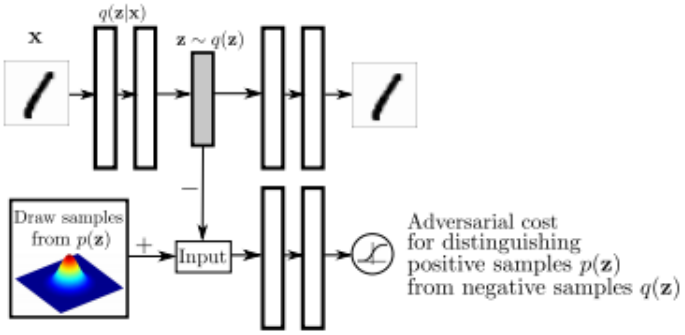
Περισσότερο φορμαλιστικά η ενεργή μάθηση εστιάζει στο ποια από τα δεδομένα πρέπει να επισημειωθούν ώστε να επιτευχθούν τα προσδοκώμενα αποτελέσματα το συντομότερο δυνατό και χωρίς να έχουμε επισημειώσει όλο το σύνολο των δεδομένων. Ο αλγόριθμος ξεκινά εκπαιδεύοντας σε ένα μικρο σύνολο $L_t \subset D$ από επισημειωμένα δεδομένα (εφόσον είναι αναγκαίο) και στην συνέχεια κάνει τις προβλέψεις για το εναπομένον σύνολο $U \subset D$ με το t να συμβολίζει το βήμα στην διαδικασία μάθησης. Στην συνέχεια βάση των αποτελέσματων ένα υποσύνολο $C \subset U$ δίνεται στο σύστημα εξωτερικής γνώσης για επισημείωση. Τέλος, ορίζεται το σύνολο $L_{t+1} = L_t \bigcup C$ και συνεχίζει η διαδικασία στο επόμενο βήμα. Υπάρχουν πολλοί

τρόποι με τους όποιους επιλέγεται το υποσύνολο C και σκοπός τους είναι η επιλογή των σημείων που θα μεγιστοποιήσει την ακρίβεια του μοντέλου στο τέλος του επομένου βήματος, όπως για παράδειγμα σημεία με μεγάλη εντροπία ή μεγάλη αβεβαιότητα.

Στο πλαίσιο της ανίχνευσης ανωμαλιών τον ρόλο του εξωτερικού συστήματος γνώσης μπορεί να παίζει ο άνθρωπος, ειδικοί οι οποίοι μπορούν να αποφανθούν για την ποιότητα των δεδομένων. Το κίνητρο της συγκεκριμένης διαδικασίας είναι ότι είναι δύσκολο να έχουμε στην διάθεση μας μεγάλο όγκο επισημειωμενών δεδομένων καθώς επίσης το τι αποτελεί ανωμαλία είναι συνήθως δυσδιάκριτο.

3.6.4 Ανταγωνιστικός αυτοκωδικοποιητής - Adversarial Autoencoder

Μερικά από τα προβλήματα που αναλύθηκαν στην παράγραφο 3.6.1 έρχεται να λύσει το μοντέλο του ανταγωνιστικού αυτοκωδικοποιητή [16]. Το συγκεκριμένο δίκτυο αφορά έναν πιθανοτικό αυτοκωδικοποιητή που εκπαιδεύεται μέσω μιας ανταγωνιστικής διαδικασίας μάθησης, όπως περιγράφεται στο [9], με σκοπό ο χώρος προβολής του κωδικοποιητή να ακολουθεί μια επιβαλλόμενη κατανομή $p(z)$. Έτσι, στον ανταγωνιστικό αυτοκωδικοποιητή τον ρόλο του γενετικού δικτύου G (Generator) έχει ο κωδικοποιητής ενώ υπάρχει ένα ακόμα δίκτυο διαχωρισμού D (Discriminator) όπως φαίνεται στο σχήμα 3.10. Ο διαχωριστής D έχει σκοπό να αναγνωρίζει αν τα δείγματα στην είσοδο του είναι δειγματοληπτημένα από την επιβαλλόμενη κατανομή $p(z)$ η από την κατανομή $q(z|x)$ της εξόδου του κωδικοποιητή. Έτσι, η ανταγωνιστική εκπαίδευση των δυο δικτύων έχει ως αποτέλεσμα η έξοδος του κωδικοποιητή να προσεγγίζει την επιβαλλομένη κατανομή $q(z) \sim p(z)$.



Σχήμα 3.10: Ανταγωνιστικός αυτοκωδικοποιητής

Η διαδικασία της εκπαίδευσης γίνεται μέσω του αλγορίθμου SGD και χωρίζεται στα στάδια της ανακατασκευής και της ομαλοποίησης. Στο στάδιο της ανακατασκευής ο αυτοκωδικοποιητής εκπαιδεύεται ώστε να ανακατασκευάζει την είσοδο στην έξοδο. Στο στάδιο ομαλοποίησης αρχικά ο διαχωριστής εκπαιδεύεται ώστε να διαχωρίζει τα δείγματα όπως αναφέρθηκε και προηγουμένως ενώ ο αυτοκωδικοποιητής ανανεώνει τις παραμέτρους του ώστε να μειώσει την ικανότητα του διαχωριστή D. Η παραπάνω διαδικασία μοντελοποιείται ως ένα παίγνιο μεγιστοποίησης ελαχιστοποίησης με τα δυο δίκτυα να ανταγωνίζονται όπου η λύση φαίνεται

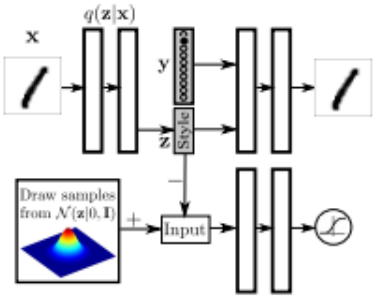
παρακάτω:

$$min_G max_D(E_{x\sim pdata}[logD(x)] + E_{z\sim p(z)}[log(1 - D(G(z))])$$

Το συγκεκριμένο δίκτυο μπορεί να υποστεί αρκετές μετατροπές όπως αναφέρεται στο [9] και ορισμένες από αυτές παρουσιάζονται παρακάτω.

Επιβλεπόμενος Ανταγωνιστικός αυτοκωδικοποιητής

Οι συγγραφείς του άρθρου [9] προτείνουν την χρήση επιπρόσθετης πληροφορίας όπου αφορά τις m διαφορετικές κλάσεις των δειγμάτων εκπαίδευσης με σκοπό την βελτίωση της απόδοσης του δικτύου στην παραγωγή δεδομένων. Συγκεκριμένα προτείνουν την χρήση μιας πρόσθετης εισόδου στον αποκωδικοποιητή με την μορφή ενός δυαδικού διανύσματος $C \in R^m$ όπου λαμβάνει μηδενικές τιμές σε όλη την διάσταση του εκτός από το στοιχείο c_i που αντιστοιχεί στην κλάση i της εισόδου. Η υπόθεση στο συγκεκριμένο δίκτυο είναι ότι αφού έχουμε την πληροφορία σχετικά με τις κλάσεις των δειγμάτων ο κωδικοποιητής θα κωδικοποιεί τις διακυμάνσεις της εκάστοτε κλάσης.



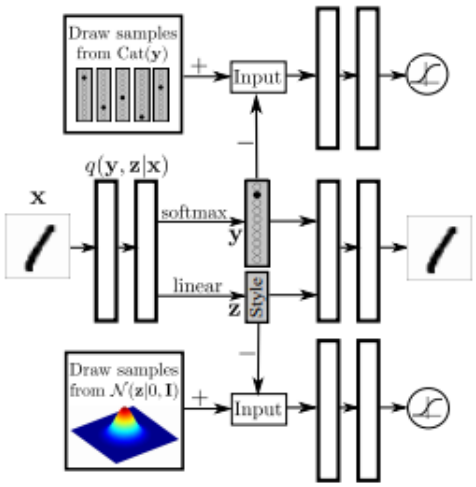
Σχήμα 3.11: Επιβλεπόμενος Ανταγωνιστικός αυτοκωδικοποιητής

Μια ενδιαφέρουσα εφαρμογή που υλοποιήθηκε στο συγκεκριμένο άρθρο είναι αυτή της παραγωγής χειρογράφων ψηφίων. Στο σχήμα 3.12 οι στήλες έχουν συγκεκριμένο διάνυσμα C που αφορά την κλάση του χειρογράφου ψηφίου ενώ έχουμε διαφορετικά δείγματα του χώρου προβολής της κατανομής που υποδεικνύουν την τεχνοτροπία του χαρακτήρα.

Ήμι-επιβλεπόμενος ανταγωνιστικός αυτοκωδικοποιητή

Η δεύτερη αρχιτεκτονική που προτείνεται είναι αυτή της ήμι επιβλεπόμενης μάθησης του αυτοκωδικοποιητή με σκοπό την κατηγοριοποίηση των μη επισημειωμένων δεδομένων. Πιο συγκεκριμένα το δίκτυο του κωδικοποιητή παράγει δυο διανύσματα εξόδου. Το πρώτο αφορά τον χώρο προβολής του όπως συνήθως, ενώ το δεύτερο αφορά το διάνυσμα $C \in R^m$ όπως περιγράφεται στην προηγουμένη ενότητα. Τα δυο αυτά διανύσματα χρησιμοποιούνται από τον αυτοκωδικοποιητή για την ανακατασκευή του σήματος εισόδου. Το δεύτερο διάνυσμα εκπαιδεύεται με ένα δεύτερο ανταγωνιστικό κριτήριο στόχος του οποίου είναι το διάνυσμα C να ακολουθεί μια νέα επιβάλλουσα κατηγορική κατανομή $p(y) = Cat(y)$. Ολόκληρο το δίκτυο φαίνεται στην εικόνα 3.13

Σχήμα 3.12: Παραγωγή χειρογράφων ψηφίων μέσω ανταγωνιστικού αυτοκωδικοποιητή



Σχήμα 3.13: Ήμι-επιβλεπόμενος Ανταγωνιστικός αυτοκωδικοποιητής

Η διαδικασία της εκπαίδευσης χωρίζεται στα στάδια της ανακατασκευής, της ομαλοποίησης και της ήμι-επιβλεπόμενης διόρθωσης του διανύσματος C. Έτσι, στο πρώτο στάδιο εκπαιδεύουμε τον αυτοκωδικοποιητή ώστε να ελαχιστοποιεί το λάθος ανακατασκευής. Στο στάδιο της ομαλοποίησης αρχικά εκπαιδεύονται οι διαχωριστές των δυο διανυσμάτων και στην συνέχεια ανανεώνονται τα γενετικά τους δίκτυα. Στο τελευταίο στάδιο, ανανεώνεται ο διαχωριστής του κατηγορικού διανύσματος έτσι ώστε να αναγνωρίζει τις ετικέτες του ενός υποσυνόλου των επισημειωμένων δεδομένων.

Συσταδοποίηση μέσω Ανταγωνιστικού Αυτοκωδικοποιητή

Σαν συνέχεια της προηγουμένης αρχιτεκτονικής, προτείνεται επίσης η παράλειψη του τελευταίου σταδίου της ήμι-επιβλεπόμενης διόρθωσης έτσι ώστε ο ανταγωνιστικός αυτοκωδικοποιητής να δημιουργεί συστάδες στα δεδομένα εισόδου. Ο αριθμός των συστάδων είναι όσος και το μήκος του διανύσματος C ενώ το ποσοστό τους μπορεί να ορισθεί μέσω της κατηγορικής κατανομής που υποχρεούται να ακολουθεί ο αυτοκωδικοποιητής.

Ανίχνευση ανωμαλιών με χρήση Ανταγωνιστικού Αυτοκωδικοποιητή

Η ανίχνευση ανωμαλιών με χρήση του ανταγωνιστικού αυτοκωδικοποιητή [4] μας δίνει την δυνατότητα να χρησιμοποιήσουμε τον χώρο προβολής $E : X \to Z$ που ορίζεται μέσω των παραμέτρων του κωδικοποιητή E. Ο τρόπος που εκπαιδεύεται το δίκτυο οδηγεί την έξοδο του κωδικοποιητή να ακολουθεί μια ορισμένη κατανομή, έτσι ο χώρος είναι καλύτερα σχηματισμένος χωρίς ασυνέχειες, επιτρέποντάς μας να εφαρμόσουμε περισσότερο φορμαλιστικά, ένα οποιοδήποτε τεστ υποθέσεων για την ανίχνευση ανωμαλιών καθώς η κατανομή θεωρείται πλέον γνωστή. Επίσης, είναι δυνατόν η κατανομή που ορίζουμε να αποτελεί μίξη κατανομών [13] οπότε και η προβολή ενός δείγματος εισόδου στην εκάστοτε κατανομή να αποτελεί νέο κριτήριο για ανίχνευση ανωμαλιών. Εκτός από την εκμετάλλευση του συνεχούς διανύσματος προβολής είναι δυνατή και η ανίχνευση ανωμαλιών μέσω του σφάλματος ανακατασκευής ή και ένα κριτήριο που αφορά και τα δυο μέτρα. Στην παρούσα διπλωματική εργασία γίνεται ακόμα έρευνα πάνω στην ανίχνευση ανωμαλιών μέσω της τεχνικής της συσταδοποίησης η οποία παρουσιάστηκε στην προηγουμένη παράγραφο. Συγκεκριμένα, ερευνήσαμε αν είναι δυνατόν να ορίσουμε μια κλάση μικρότερης πιθανότητας στην επιβαλλομένη διακριτή κατανομή που θα αντιστοιχεί στα ανώμαλα δεδομένα σύμφωνα με την πεποίθηση μας για το ποσοστό τους.

Κεφάλαιο 4

Περιγραφή Δεδομένων

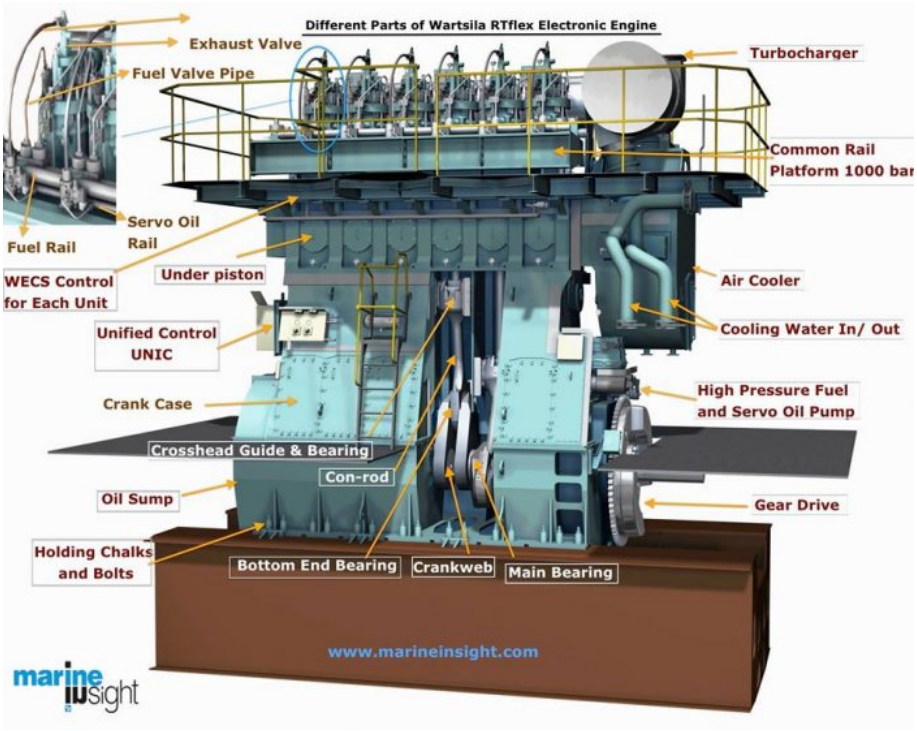
4.1 Εισαγωγή

Στην παρούσα διπλωματική εργασία έγινε ανάλυση και εφαρμογή διάφορων αλγορίθμων και τεχνικών όπως αυτές παρουσιάστηκαν στο τρίτο κεφάλαιο της εργασίας. Τα δεδομένα που χρησιμοποιήθηκαν για την ανάλυση αφορούν δεδομένα πλοίων. Πιο συγκεκριμένα, το σύνολο των δεδομένων περιγράφουν στοιχεία που συλλέγονται από διάφορα σημεία της κύριας μηχανής του πλοίου όπως θερμοκρασίες και πιέσεις, καθώς επίσης έγινε χρήση χαρακτηριστικών που περιγράφουν την απόδοση του εκάστοτε πλοίου σαν ενιαίο σύστημα. Ο στόχος της εργασίας είναι να μπορέσουμε να εντοπίσουμε ασυνήθιστες καταστάσεις στα δεδομένα του πλοίου και να τις συσχετίσουμε με τις ιστορικές ζημιές που έχουν καταγραφεί ώστε να φτιάξουμε ένα σύστημα επίβλεψης του πλοίου το οποίο θα είναι ικανό να εντοπίζει πρόωρα πιθανές βλάβες ή καταστάσεις οι οποίες επηρεάζουν την απόδοση του κινητήρα.

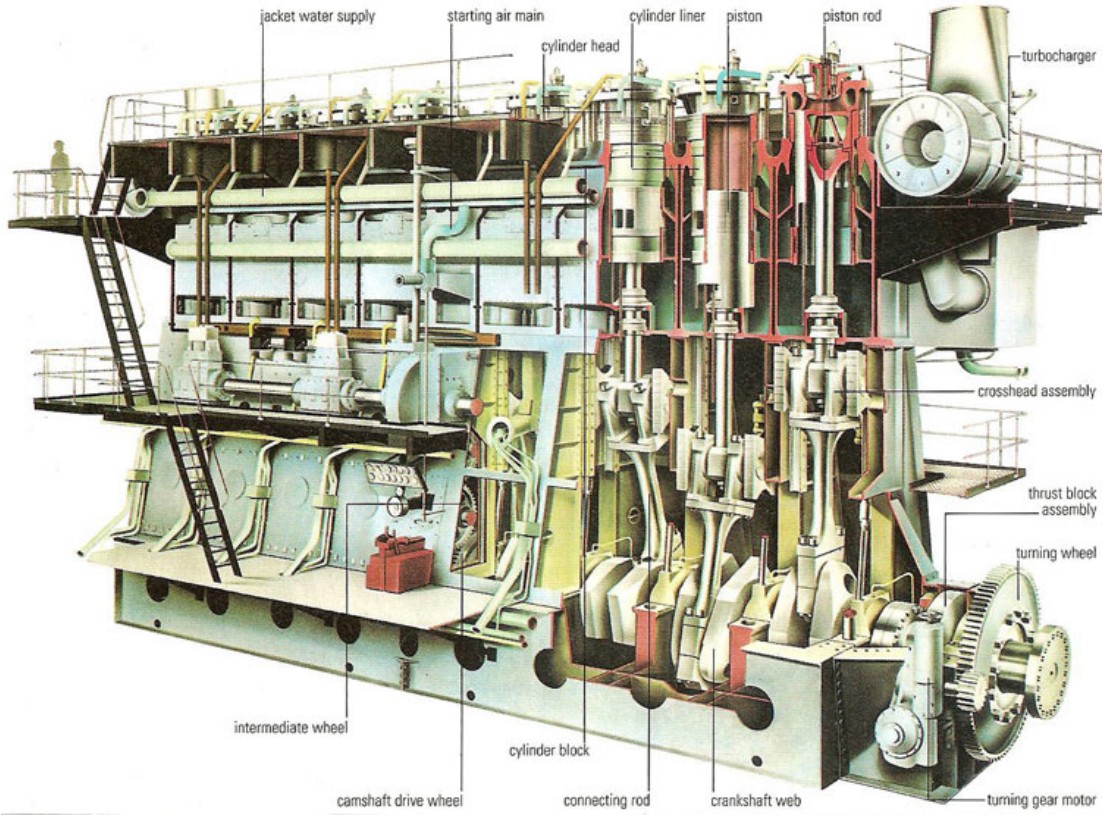
4.2 Περιγραφή Δεδομένων

Τα δεδομένα που έχουμε στην διάθεση μας αφορούν ιστορικά δεδομένα του πλοίου από το 2016 μέχρι το 2018 τα οποία συλλέγονται με συχνότητά 1 λεπτού και στέλνονται μέσω δορυφορικής σύνδεσης στην βάση δεδομένων. Τα περισσότερα από τα δεδομένα αφορούν την κύρια μηχανή του πλοίου όπως αυτά παρουσιάζονται στον πίνακά 4.1.

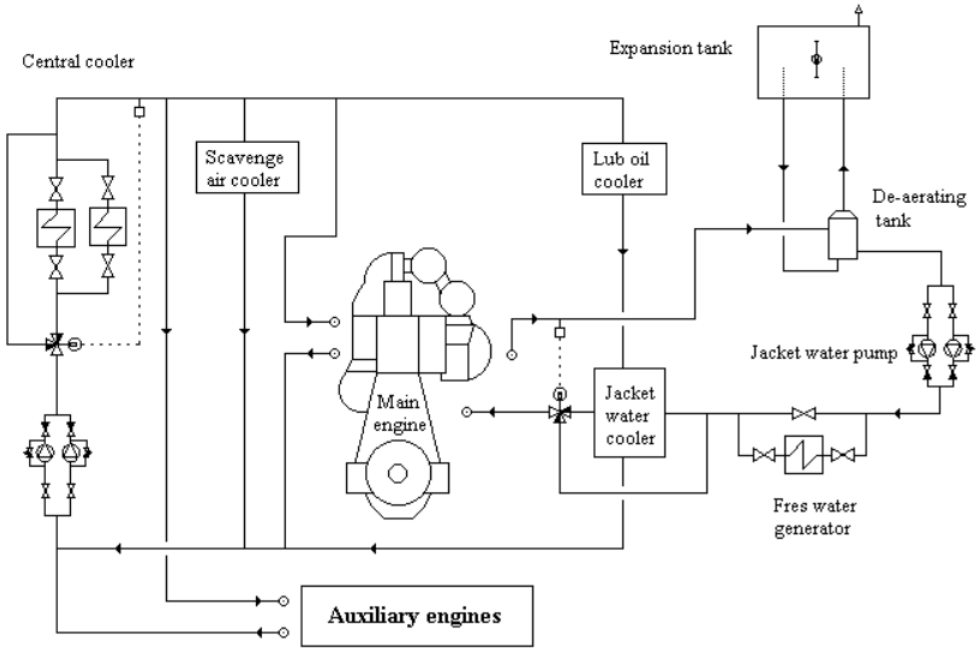
Σε γενικές γραμμές οι χρονοσειρές αφορούν θερμοκρασίες και πιέσεις σε διάφορα μέρη της κύριας μηχανής. Παράδειγμα αποτελούν θερμοκρασίες και πιέσεις των καυσαερίων εξόδου των κυλίνδρων της κύριας μηχανής ή των συστημάτων ψύξης ολοκλήρου του κινητήρα. Επιπροσθέτως έχουμε πρόσβαση και στην καταναλισκόμενη ισχύ πάνω στον άξονα του κινητήρα καθώς επίσης στην ροπή και τις στροφές του άξονα. Τέλος έχουμε στην διάθεση μας δεδομένα από ιστορικές βλάβες του πλοίου τις οποίες θα προσπαθήσουμε να συσχετίσουμε με τα δεδομένα. Μια γραφική απεικόνιση του συστήματος της κύριας μηχανής βρίσκεται στις εικόνες 4.1,4.2 ενώ στην εικόνα 4.3 φαίνεται το διάγραμμα του συστήματος ψύξης.



Σχήμα 4.1: Παράδειγμα κύριας μηχανής 1



Σχήμα 4.2: Παράδειγμα κύριας μηχανής 2



Σχήμα 4.3: Σύστημα ψύξης κύριας μηχανής

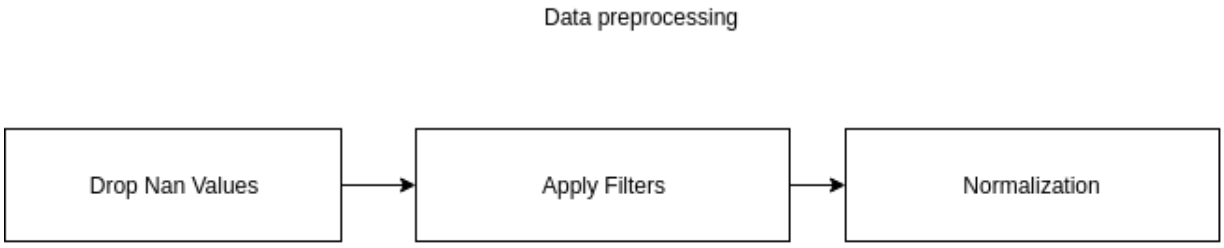
4.3 Παρουσίαση δεδομένων

Σ΄ αυτήν την ενότητα θα παρουσιάσουμε γραφικά τα δεδομένα προτού γίνει η προ-επεξεργασία τους ώστε να κατανοήσουμε την φύση των δεδομένων, διαδικασία που είναι αναπόσπαστη με την ανάλυση των αποτελεσμάτων των μοντέλων σε οποιαδήποτε πρόβλημα μηχανικής μάθησης. Αξίζει να σημειωθεί ότι λόγω του μεγάλου όγκου δεδομένων τόσο στο πλήθος όσο και στο αριθμό τους είναι αδύνατο να παρουσιάσουμε το σύνολο τους σε κάθε επιμέρους ανάλυση που θα γίνεται στην συνέχεια. Γι΄ αυτό τον λόγο θα παρουσιάζεται ένα υποσύνολο όπου θα είναι αντιπροσωπευτικό για την αξιολόγηση των αποτελεσμάτων. Στα διάγραμμα 4.4,4.5,4.6 φαίνονται οι χρονοσειρές σε διάρκεια 7 μηνών.

Από τα προαναφερόμενα διαγράμματα παρατηρούμε ότι μετά τον 09-2018 είναι εμφανής μια απροσδόκητη συμπεριφορά των δεδομένων. Τα δεδομένα αυτά προφανώς αντιστοιχούν σε κάποιο σφάλμα στην μετάδοση ή στο λογισμικό του αλλά δεν θα απορριφθούν από το σύνολο των δεδομένων αφού σκοπός μας εκτός από τον εντοπισμό βλαβών είναι η γενική επίβλεψη της ροής των δεδομένων. Επιπροσθέτως είναι ένας καλός τρόπος να εξετάσουμε την ορθότητά των αλγορίθμων καθώς όπως και θα αναφερθεί στην συνέχεια δεν έχουμε ετικέτες εκπαίδευσης για σύνολο των δεδομένων πράγμα που καθιστά την αποτίμηση των αποτελεσμάτων ιδιαίτερα δύσκολη.

4.4 Προ-επεξεργασία δεδομένων

Σαν πρώτο στάδιο της προ-επεξεργασίας των δεδομένων θα διώχνουμε τις χρονικές στιγμές του συνόλου δεδομένων όπου παρουσιάζουν έλλειψη σε οποιοδήποτε χαρακτηριστικό. Ο λόγος που δεν προβαίνουμε σε κάποια εκτίμηση του χαρακτηριστικού που λείπει είναι ώστε πρώτον να μην επηρεάσουμε το μοντέλο που πρόκειται να εκπαιδεύσουμε και δεύτερον ο όγκος των δεδομένων είναι αρκετός ώστε να εκπαιδευτούν τα μοντέλα που θα παρουσιάσουμε στο επόμενο κεφάλαιο. Το δεύτερο στάδιο είναι να φιλτράρουμε τις χρονικές στιγμές όπου το καράβι δεν βρίσκεται σε πλεύση αφού μας αφορά η κανονική λειτουργία του και όχι τυχόν απροσδόκητες συμπεριφορές που μπορεί να εμφανιστούν στην διάρκεια όπου βρίσκεται κοντά σε κάποιο λιμάνι όπως απότομες στροφές και μεταβολές της ισχύος του πλοίου. Το δεύτερο το επιτυγχάνουμε μέσω των τηλεγραφημάτων του πλοίου, στα οποία υπάρχει η σχετική πληροφορία με την αναχώρηση και την προσέλευση του πλοίου στα λιμάνια.



Σχήμα 4.7: Διάγραμμα ροής προ επεξεργασίας δεδομένων.

Επειδή τα δεδομένα που έχουμε στην διάθεση μας δεν έχουν περάσει από προ-επεξεργασία και προέρχονται από μια σύνθετη διαδικασία συλλογής όπου συμμετέχουν πολλά επιμέρους συστήματα συλλογής και διανομής πολλές φορές παρουσιάζονται τιμές εντελώς εκτός από το πεδίο τιμών που ορίζει η φυσική στην οποία υπόκεινται. Είναι λοιπόν αναγκαίο να εφαρμόσουμε φίλτρα συμβατά με την φυσική του προβλήματος ώστε να καθαρίσουμε τα δεδομένα. Τα φίλτρα που εφαρμόσαμε αντιστοιχούν σε ομάδες χαρακτηριστικών και τα αποτελέσματα της επεξεργασίας φαίνονται παρακάτω. Εδώ, αξίζει να αναφέρουμε ότι τα εν λόγω φίλτρα δεν θα επηρεάσουν την ανάλυση μας και οι τιμές τους προσδιορίστηκαν με βάση τα τεχνικά χαρακτηρικά των αισθητήρων.

Το τελευταίο στάδιο της προ-επεξεργασίας των δεδομένων είναι η κανονικοποίηση των δεδομένων. Ο λόγος που κανονικοποιούμε τα δεδομένα πριν από κάθε πρόβλημα μηχανικής μάθησης είναι το γεγονός ότι διαφορετικά δεδομένα έχουν τελείως διαφορετικό εύρος τιμών. Πρακτικά, τα πλεονεκτήματα της κανονικοποίησης διαφέρουν αναλόγως με τον αλγόριθμο που εφαρμόζουμε. Ένας διαισθητικός λόγος είναι ότι με την κανονικοποίηση όλα τα χαρακτηριστικά έρχονται στο ίδιο εύρος και έτσι ο χώρος στον οποίο μπορούν να κινηθούν τα βάρη ώστε να συγκλίνει ο αλγόριθμος είναι πολύ μικρότερος. Συνέπεια του προηγουμένου είναι ότι με

την κανονικοποίηση οι αλγόριθμοι είναι πιο πιθανό να συγκλίνουν σε μια καλύτερη λύση ενώ παράλληλα ο απαιτούμενος χρόνος σύγκλισης είναι πολύ μικρότερος. Έτσι, για κάθε ένα από τα χαρακτηριστικά του συνόλου των δεδομένων κανονικοποιούμε τα δεδομένα όπως φαίνεται παρακάτω:

$$z = \frac{x - \mu}{\sigma}$$

όπου:

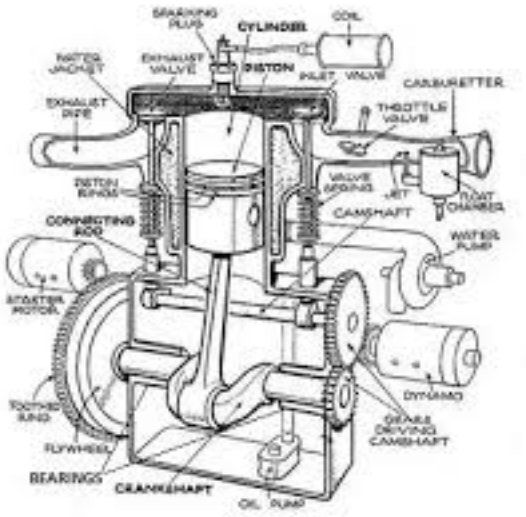
x: Το στοιχείο προς κανονικοποίηση

μ: Η μέση τιμή του χαρακτηριστικού

σ : Η τυπική απόκλιση του χαρακτηριστικού

4.5 Παρουσίαση βλαβών

Πριν προβούμε στην παρουσίαση και στην ανάλυση των πειραμάτων πρέπει να παρουσιάσουμε τις βλάβες που έχουμε στην διάθεση μας. Οι βλάβες που έχουμε στην διάθεση μας αφορούν τα έδρανα της κύριας μηχανής. Τα έδρανα είναι στοιχεία στήριξης μηχανικών εξαρτημάτων της μηχανής. Τα συγκεκριμένα έδρανα αφορούν τα έδρανα των κυλίνδρων της κύριας μηχανής. Ένα παράδειγμα ενός μηχανικού συστήματος φαίνεται στο σχήμα 4.8. Οι καταγραμμένες βλάβες στη διάρκεια του χρόνου φαίνονται στα προαναφερόμενα διαγράμματα 4.4, 4.5, 4.6 με τις κάθετες κόκκινες ευθείες.



Σχήμα 4.8: Παράδειγμα εδράνων κύριας μηχανής

4.6 Περιορισμοί στα δεδομένα εκπαίδευσης

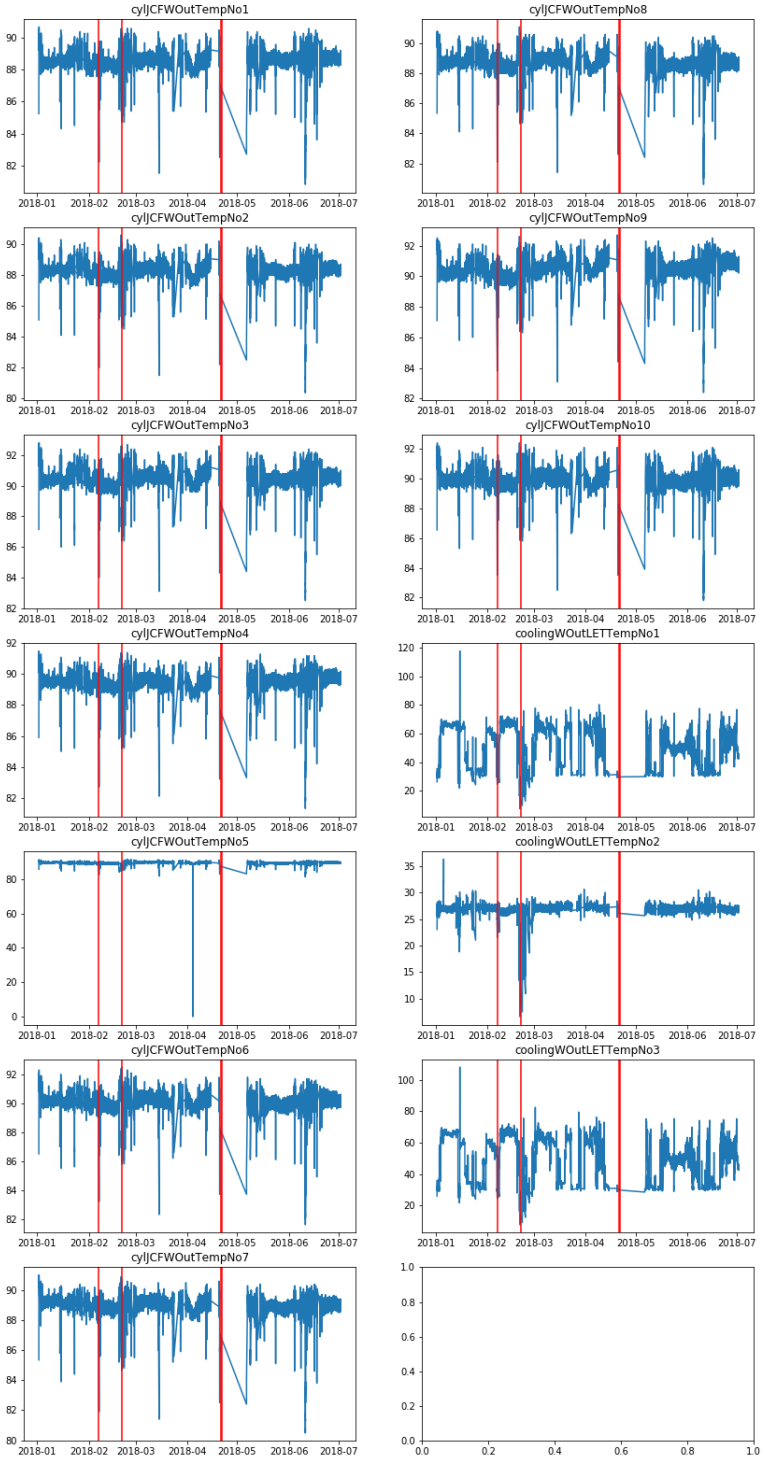
Σε αυτό το σημείο αξίζει να αναφέρουμε ότι οι καταγραμμένες βλάβες αντιστοιχούν σε ημερομηνία επισκευής και όχι ανίχνευσης της βλάβης. Αυτό σε συνδυασμό με το γεγονός ότι ο αριθμός τους είναι ο μικρός καθιστούν δύσκολη την ανίχνευση τους με μοντέλα μηχανικής μάθησης. Ένας άλλος περιορισμός προκύπτει από την φύση του περιβάλλοντος του προβλήματος στο οποίο τα πλοία υπόκεινται. Πιο συγκεκριμένα, στο πλαίσιο της ναυτιλίας τα πλοία πρέπει να βρίσκονται όσο το δυνατόν λιγότερο εκτός λειτουργίας αφού τα έξοδα συντήρησης τους είναι αρκετά μεγάλα ενώ το περιθώριο κέρδους λειτουργίας αρκετά μεγάλο. Αυτό σημαίνει ότι μια ημερομηνία επισκευής δεν αντιστοιχεί σε ημερομηνία βλάβης αφού θα μπορούσε ο προγραμματισμός των εργασιών να επηρεάσει την ημερομηνία επισκευής εφόσον η βλάβη δεν είναι απαγορευτική στην πλεύση και απλά επηρεάζει την απόδοση του πλοίου με κάποια επερχόμενη καταπόνηση στα μηχανικά μέρη του κινητήρα. Επίσης όπως παρουσιάστηκε και στην παράγραφο 2.6.1 οι εν λόγω επισκευές μπορεί να είναι προϊόν μιας προληπτικής στρατηγικής συντήρησης του πλοίου. Έτσι, σε περίπτωση που δεν υπάρχει εκτενής πληροφόρηση σχετικά με τις βλάβες των πλοίων είναι δύσκολο να ανιχνευτεί η αιτία της βλάβης μέσω των δεδομένων. Έπειτα τα σήματα των διαγραμμάτων 4.4, 4.5, 4.6 είναι εξαιρετικά θορυβώδη και έτσι είναι δύσκολο να τα κατηγοριοποιήσουμε σαν ανώμαλα ή μη.

Σε ότι αφορά παρόμοιες προσπάθειες για ανίχνευση βλαβών που αφορούν μηχανικά μέρη της μηχανής η πλειοψηφία της βιβλιογραφίας προσεγγίζει το πρόβλημα έχοντας στην διάθεση της σήματα από τις δονήσεις της μηχανής. Στο παρόν σύνολο δεδομένων δεν έχουμε στην διάθεση μας παρόμοια σήματα.

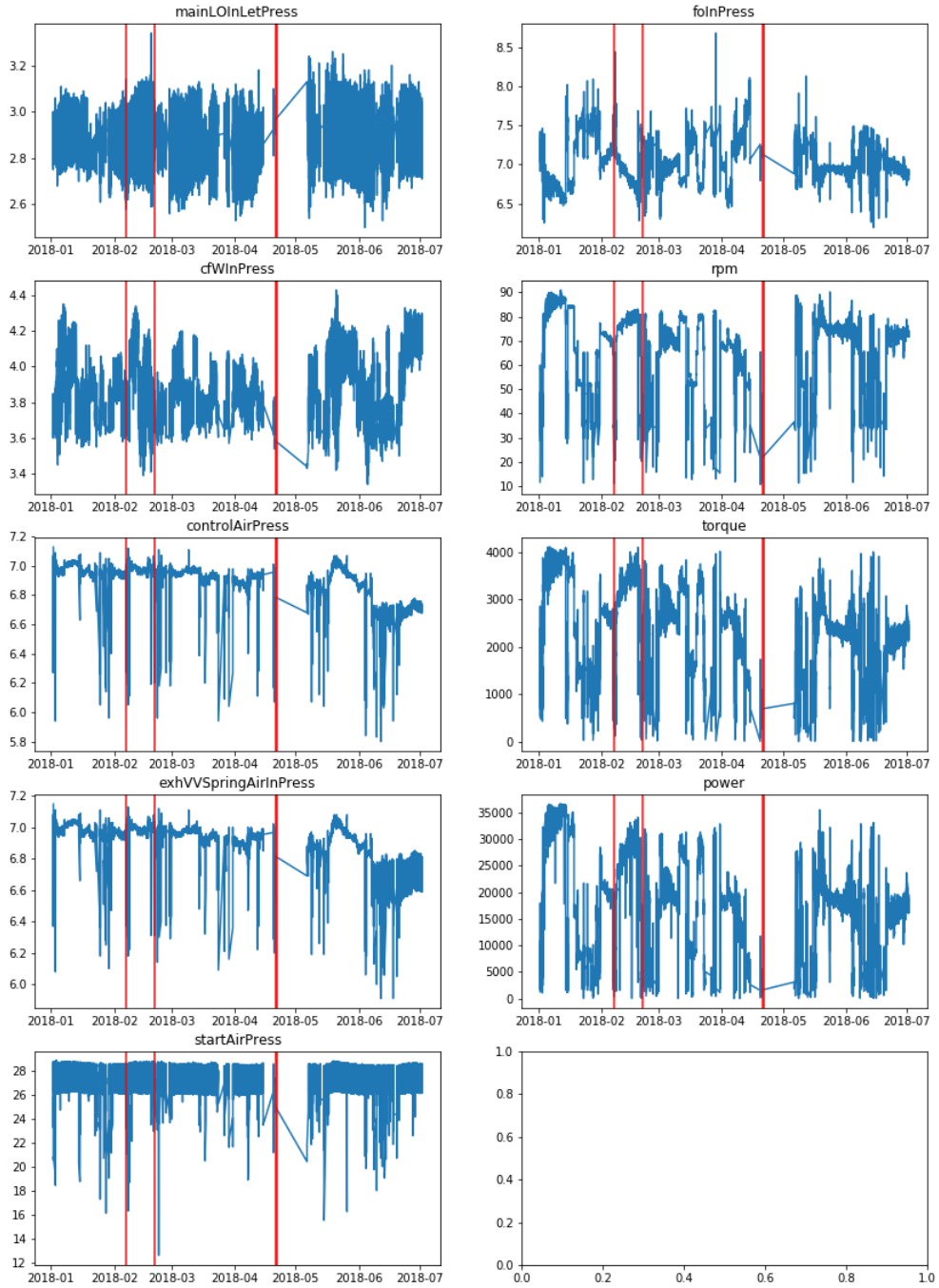
Συνοψίζοντας ο αριθμός των αναγραφομένων βλαβών είναι μικρός και σε συνδυασμό με την έλλειψη πληροφόρησης για κάθε βλάβη καθιστά δύσκολη την εφαρμογή μοντέλων μηχανικής μάθησης. Επίσης η παρούσα εργασία δεν επικεντρώνεται μόνο στο να εντοπίσει τις συγκεκριμένες βλάβες αλλά ένα υπερσύνολο από απροσδόκητες καταστάσεις στην πλεύση του πλοίου. Στο επόμενο κεφάλαιο θα αναλυθεί ο τρόπος προσέγγισης του προβλήματος.

Αντιστοίχηση ελληνικής ονοματολογίας χαρακτηριστικών εισόδου	
ID χαρακτικών	Ελληνική ονοματολογία
1. Rpm	Στροφές ανά λεπτό.
2. Torque	Ροπή.
3. Power	Ισχύς.
4. Main Lube Oil Inlet Persure	Πίεση στο κύριο σύστημα λίπανσης της μηχανής.
5. Cooling Fresh Water Inlet Pressure	Πίεση στο σύστημα ψύξης καθαρού νερού.
6. Control Air Pressure	Πίεση στο σύστημα ελέγχου αέρα.
7. Exhaust Valve Spring Air Inlet Pressure	Πίεση στην είσοδο της βαλβίδας εξαέρωσης.
8. Starting Air Pressure	Πίεση στο σύστημα αέρα εκκίνησης.
9. Fuel Oil Inlet Pressure	Πίεση του καυσίμου.
10. Scavenge Air Receiver Temperature	Πίεση στο σύστημα διοχέτευσης καθαρού αέρα.
11. Fuel Oil Temperature	Θερμοκρασία καυσίμου.
12. Fuel Oil Inlet Temperature	Θερμοκρασία καυσίμου στην είσοδο του κινητήρα.
13. Jacket Cooling Fresh Water Inlet Temperature Low	Θερμοκρασία καθαρού νερού στην είσοδο του συστήματος μανδύα ψύξης του κινητήρα.
14. Jacket Cooling Fresh Water Outlet Temperature 1	Θερμοκρασία του μανδύα στο πιστόνι 1.
15. Jacket Cooling Fresh Water Outlet Temperature 2	Θερμοκρασία του μανδύα στο πιστόνι 2.
16. Jacket Cooling Fresh Water Outlet Temperature 3	Θερμοκρασία του μανδύα στο πιστόνι 3.
17. Jacket Cooling Fresh Water Outlet Temperature 4	Θερμοκρασία του μανδύα στο πιστόνι 4.
18. Jacket Cooling Fresh Water Outlet Temperature 5	Θερμοκρασία του μανδύα στο πιστόνι 5.
19. Jacket Cooling Fresh Water Outlet Temperature 6	Θερμοκρασία του μανδύα στο πιστόνι 6.
20. Jacket Cooling Fresh Water Outlet Temperature 7	Θερμοκρασία του μανδύα στο πιστόνι 7.
21. Jacket Cooling Fresh Water Outlet Temperature 8	Θερμοκρασία του μανδύα στο πιστόνι 8.
22. Jacket Cooling Fresh Water Outlet Temperature 9	Θερμοκρασία του μανδύα στο πιστόνι 9.
23. Jacket Cooling Fresh Water Outlet Temperature 10	Θερμοκρασία του μανδύα στο πιστόνι 10.
24. Turbo-Charge Air Cooler Cooling Water Outlet Temperature 2	Θερμοκρασία νερού στο σύστημα ψύξης του υπερσυμπιεστή 1.
25. Turbo-Charge Air Cooler Cooling Water Outlet Temperature 3	Θερμοκρασία νερού στο σύστημα ψύξης του υπερσυμπιεστή 2.
26. Turbo-Charge Air Cooler Cooling Water Outlet Temperature 3	Θερμοκρασία νερού στο σύστημα ψύξης του υπερσυμπιεστή 3.

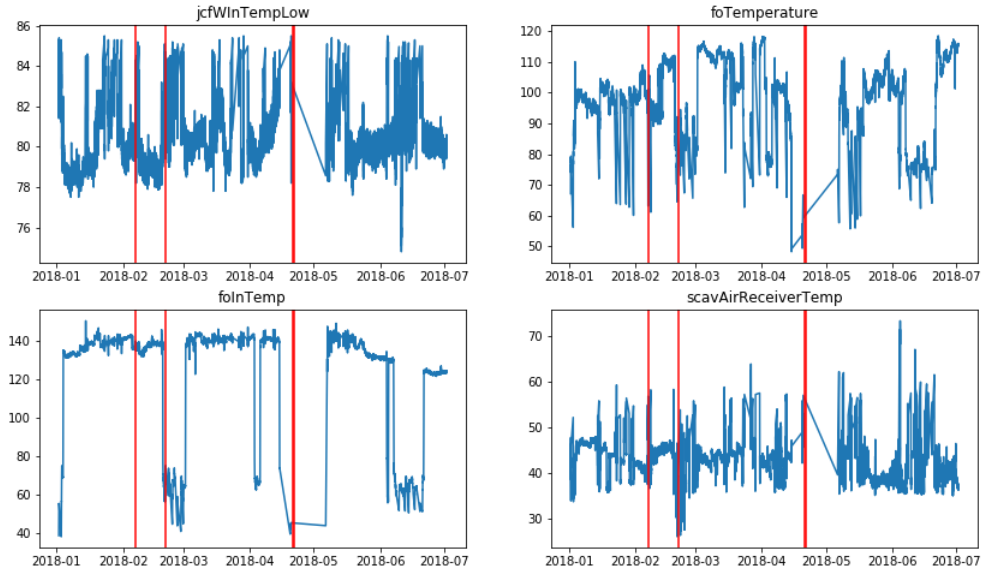
Πίνακας 4.1: Πίνακας διαθέσιμων χαρακτηριστικών



Σχήμα 4.4: Παραπάνω φαίνονται τα διαγράμματα των χαρακτηριστικών για 7 μήνες της πρώτης ομάδας χαρακτηριστικών.Οι κάθετες ευθείες υποδηλώνουν καταγραμμένες βλάβες των πλοίων



Σχήμα 4.5: Παραπάνω φαίνονται τα διαγράμματα των χαρακτηριστικών για 7 μήνες της δεύτερης ομάδας χαρακτηριστικών.Οι κάθετες ευθείες υποδηλώνουν καταγραμμένες βλάβες των πλοίων



Σχήμα 4.6: Παραπάνω φαίνονται τα διαγράμματα των χαρακτηριστικών για 7 μήνες της τρίτης ομάδας χαρακτηριστικών.Οι κάθετες ευθείες υποδηλώνουν καταγραμμένες βλάβες των πλοίων

Κεφάλαιο 5

Πειραματική ανάλυση

5.1 Εισαγωγή

Σε αυτό το κεφάλαιο θα γίνει παρουσίαση του τρόπου προσέγγισης του προβλήματος καθώς και ανάλυση των αποτελεσμάτων των πειραμάτων. Συνοπτικά προσεγγίσαμε το πρόβλημα σαν ένα πρόβλημα μη επιβλεπόμενης μάθησης και πειραματιστήκαμε με πολλά μοντέλα και τεχνικές για την ανίχνευση ανωμαλιών στα δεδομένα. Εφαρμόσαμε έναν αυτοκωδικοποιητή και τον συγκρίναμε με ένα ανατροφοδοτούμενο δίκτυο αναλύοντας παράλληλα τα κύρια σημεία κάθε τεχνικής ανίχνευσης ανωμαλιών. Αξίζει να αναφέρουμε ότι στην καρδιά της ανίχνευσης ανωμαλιών βρίσκονται τα κατώφλια απόφασης για τα οποία δοκιμάσαμε διάφορες τεχνικές ώστε να οριστούν συμπεριλαμβανομένων και κατωφλιών σύμφωνα με την φυσική του προβλήματος. Τέλος, προσομοιώσαμε την πραγματική ροή των δεδομένων επανεκπαιδεύοντας το μοντέλο στο πέρασμα του χρόνου ώστε να έχουμε μια πιο ρεαλιστική εικόνα της απόδοσης του συστήματος και την συγκρίναμε με την προσέγγιση της στατικής εκπαίδευσης.

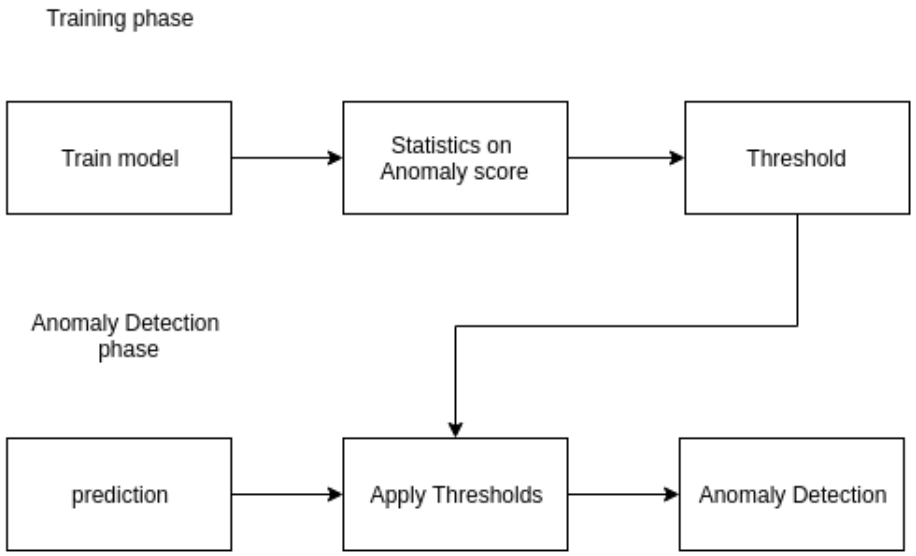
5.2 Προσέγγιση του Προβλήματος

Όπως αναφέρθηκε και στην παράγραφο 4.6 όπου έγινε ανάλυση των περιορισμών του συνόλου δεδομένων, ο αριθμός των αναφερομένων βλαβών είναι αρκετά μικρός περιέχοντας παράλληλα μεγάλο βαθμό αβεβαιότητας και έτσι είναι σχεδόν αδύνατον να εφαρμόσουμε τεχνικές επιβλεπόμενης μάθησης για την λύση του προβλήματος. Επιπροσθέτως, τεχνικές ήμιεπιβλεπόμενης μάθησης όπου θα μπορούσαμε να εκπαιδεύσουμε το μοντέλο σε περιοχές όπου γνωρίζουμε ότι δεν έχουν παρουσιάσει βλάβη, γεννά τον περιορισμό της ανίχνευσης βλαβών σε πλοία χωρίς ιστορικό βλαβών καθώς παράλληλα αποκλίνει από τον στόχο της ανίχνευσης ανωμαλιών στα δεδομένα, ανεξάρτητα από την ύπαρξη κάποιας βλάβης. Προσεγγίσαμε λοιπόν το πρόβλημα σαν ένα πρόβλημα μη επιβλεπόμενης μάθησης.

Σαν ένα πρόβλημα μη επιβλεπόμενης μάθησης και όπως αναφέρθηκε στην παράγραφο 3.3 η υπόθεση είναι ότι κανονικές λειτουργίες υπερτερούν αριθμητικά έναντι των ανωμαλιών στο σύνολο των δεδομένων εκπαίδευσης και έτσι τα μοντέλα προσπαθώντας να ελαχιστοποιήσουν το σφάλμα τους μαθαίνουν να αναπαράγουν την πιο πιθανή κατάσταση δεδομένων των συν-

θηκών συσχετίζοντας τα δεδομένα εισόδου.

Περιληπτικά, η τεχνική ανίχνευσης φαίνεται στο παρακάτω διάγραμμα 5.1 όπου αρχικά εκπαιδεύουμε ένα μοντέλο σε ένα σύνολο εκπαίδευσης και στην συνέχεια μέσω κάποιου στατιστικού κριτηρίου ορίζουμε κατώφλια απόφασης για την ανίχνευση ανωμαλιών στο σύνολο ελέγχου.



Σχήμα 5.1: Διάγραμμα ροής τεχνικής ανίχνευσης ανωμαλιών.

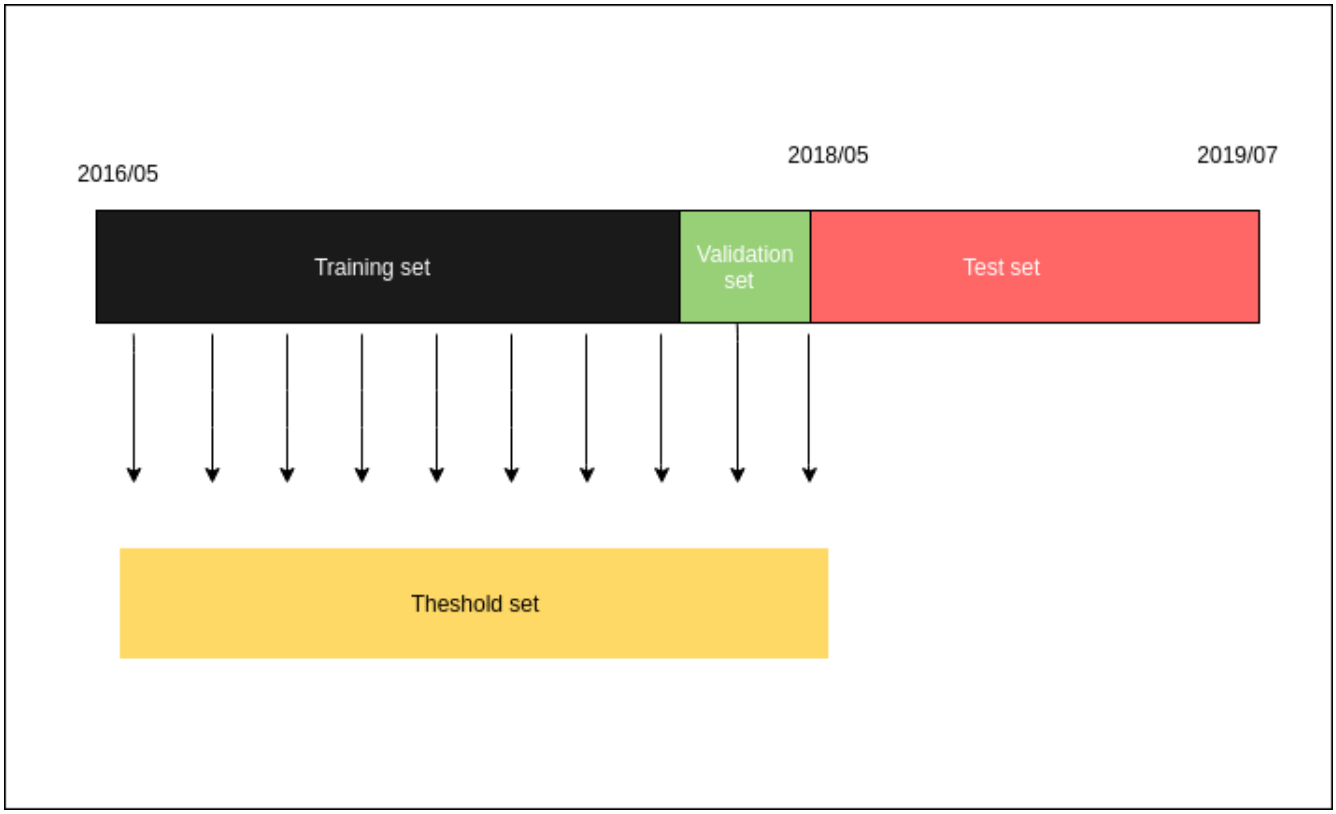
Στην συνέχεια, θα γίνει αναλυτική περιγραφή των βημάτων ανίχνευσης της παραπάνω τεχνικής εμπλουτίζοντας την με τεχνικές λεπτομέρειες. Αρχικά, θα παρουσιαστεί το μοντέλο του αυτοκωδικοποίητη και σταδιακά θα γίνεται σύγκριση με ένα μοντέλο ανατροφοδοτούμενου νευρωνικού δικτύου.

5.3 Διαχωρισμός του συνόλου δεδομένων

Για την εκπαίδευση και την επικύρωση των αποτελεσμάτων του μοντέλου χωρίσαμε το σύνολο των δεδομένων X με τον ακόλουθο τρόπο. Αρχικά έγινε ένας διαχωρισμός του συνόλου ώστε να έχουμε στην διάθεσή μας ένα σύνολο εκπαίδευσης X_{model} και ένα σύνολο ελέγχου X_{test}. Ο διαχωρισμός έγινε χρονολογικά και με γνώμονα ότι ο όγκος των δεδομένων είναι αρκετά μεγάλος και στα δυο σύνολα. Στην συνέχεια, από το σύνολο εκπαίδευσης X_{model} αποσπάσαμε ένα ποσοστό της τάξεως του 10 % με τυχαία δειγματοληψία χωρίς επανατοποθέτηση ώστε να δημιουργήσουμε ένα σύνολο $X_{threshold}$ για τον ορισμό των κατωφλιών μειώνοντας τον αριθμό του συνόλου X_{model}. Στην συνέχεια χωρίσαμε εκ νέου χρονολογικά το σύνολο X_{model} ώστε να αποσπάσουμε τα τελικά σύνολα εκπαίδευσης X_{train} και επικύρωσης

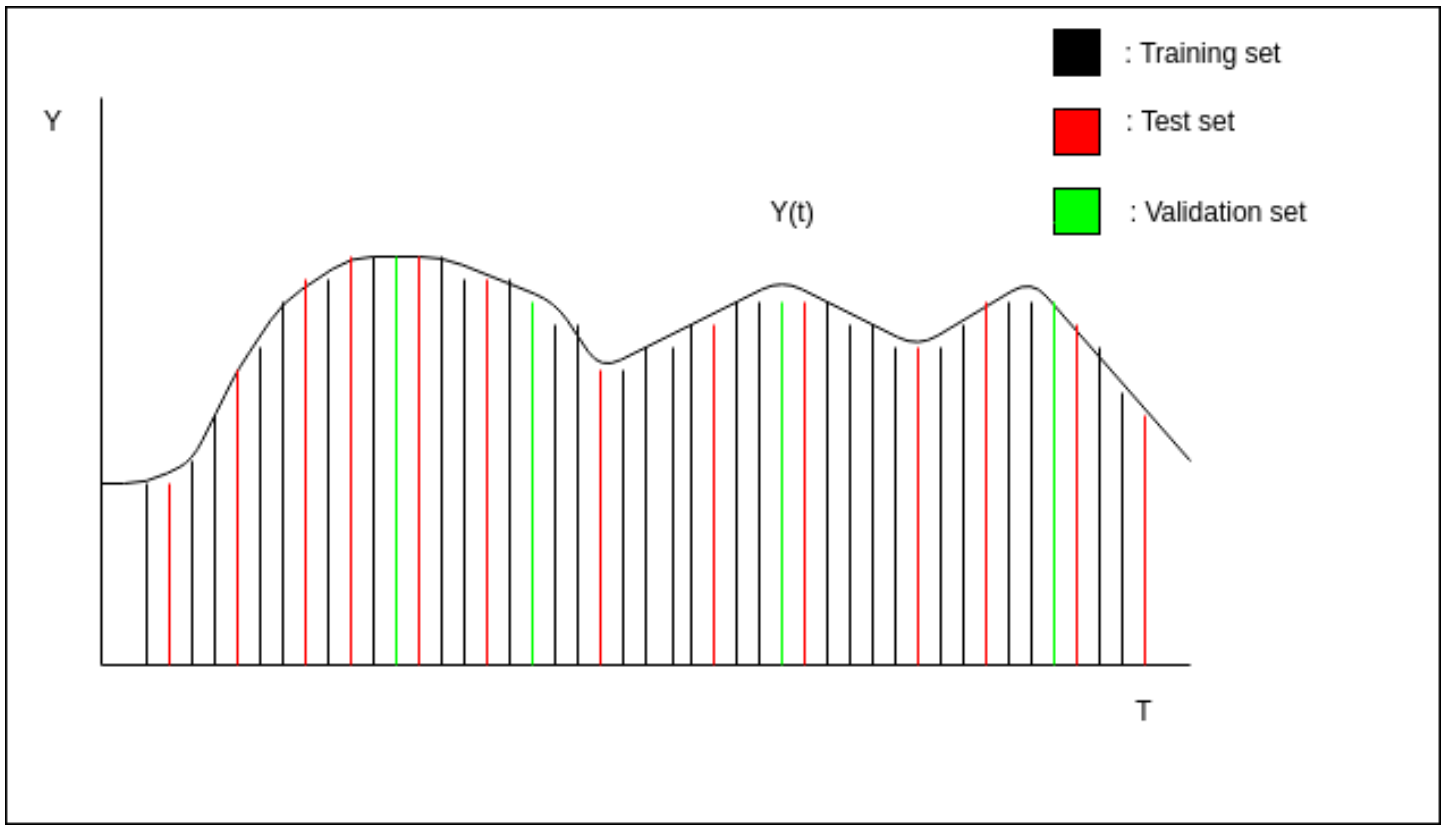
X_{val}.

Η ανίχνευση ανωμαλιών που παρουσιάζεται στην συνέχεια γίνεται πάνω σε ολόκληρο το σύνολο δεδομένων X. Ο λόγος που γίνεται αυτό είναι για να σιγουρευτούμε ότι το μοντέλο δουλεύει εξίσου καλά και στα δυο σύνολα αφού θα μπορούσε το ποσοστό των ανωμαλιών στο σύνολο ελέγχου X_{test} να είναι αρκετά μεγαλύτερο. Από μια δεύτερη οπτική σκοπιά, αλγόριθμοι ανίχνευσης ανωμαλιών εστιάζουν στο πρόβλημα του διαχωρισμού μιας κατανομής p_{full} στις p_{normal}, $p_{abnormal}$ κάνοντας χρήση ολόκληρης της κατανομής των δεδομένων και όχι της πρόβλεψης για το αν ένα δείγμα είναι απροσδόκητο δεδομένης μιας κατανομής p.



Σχήμα 5.2: Διαχωρισμός συνόλου δεδομένων

Είναι σημαντικό, σε δεδομένα που αφορούν χρονοσειρές τα σύνολα εκπαίδευσης X_{train} επικύρωσης X_{val} και ελέγχου X_{test} να είναι χωρισμένα χρονολογικά ώστε να προσομοιώνεται η πραγματική λειτουργία του συστήματος. Εκτός αυτού, αν χωρίζαμε τυχαία τα παραπάνω σύνολα θα είχαμε μια τυχαία δειγματοληψία του σήματος στην οποία το σύνολο X_{train} περιέχει ένα μεγάλο ποσοστό του συνόλου της πληροφορίας για το αρχικό σύνολο X και έτσι το αποτέλεσμα δεν θα είναι αντικειμενικό όπως φαίνεται παρακάτω:

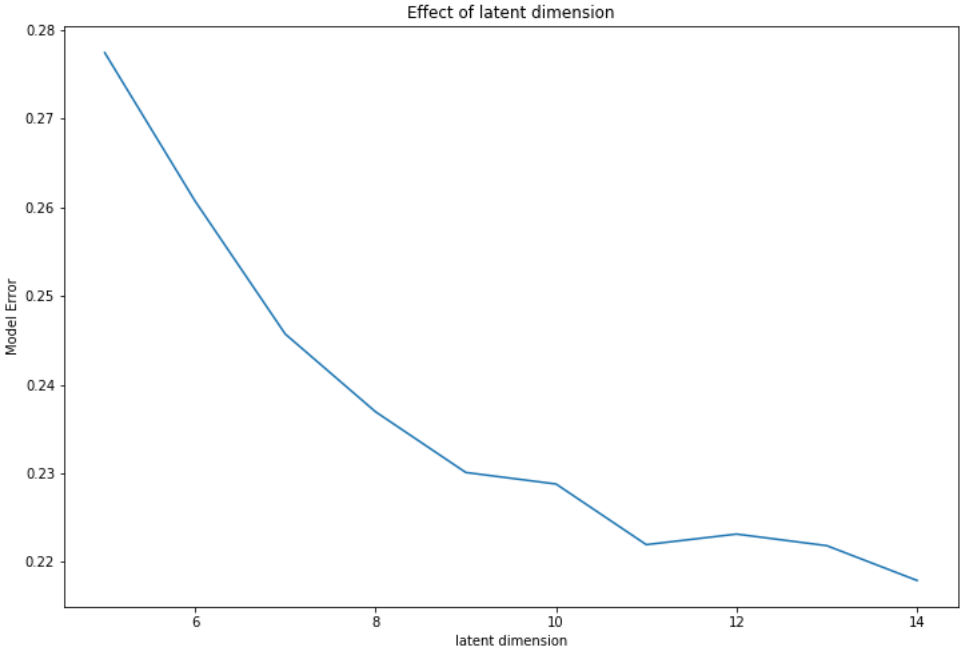


Σχήμα 5.3: Οπτικοποίηση προβλήματος ανάμιξης συνόλων στα δεδομένα

5.4 Εκπαίδευση αυτοκωδικοποιητή

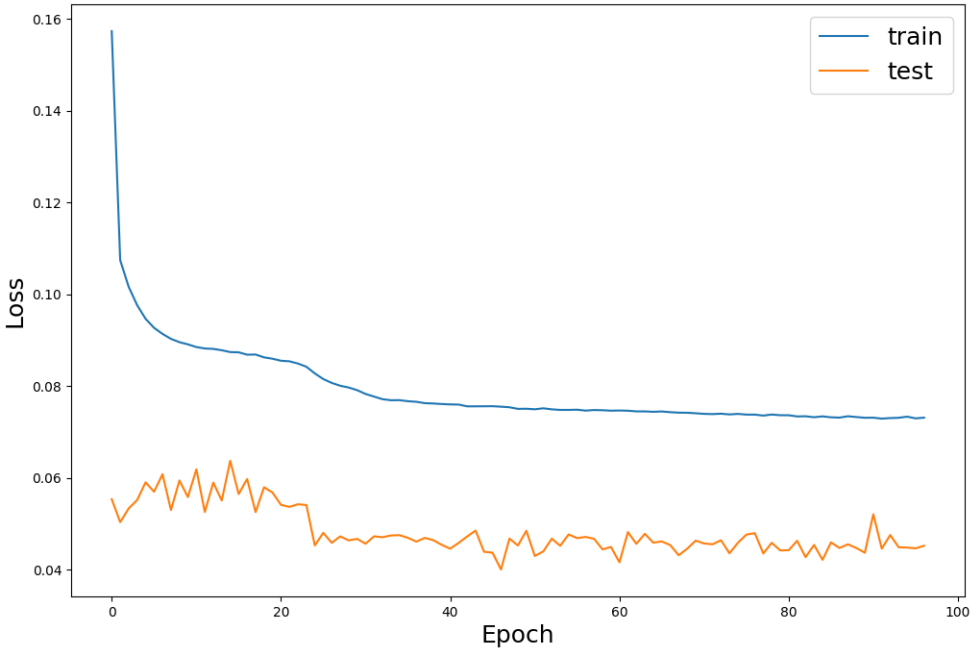
Το πρώτο μοντέλο το οποίο υλοποιήσαμε είναι το μοντέλο του αυτοκωδικοποιητή. Στόχος του μοντέλου όπως αναφέρθηκε και στο κεφάλαιο 3 είναι η μείωση της διαστατικότητας του διανύσματος εκπαίδευσης σε ένα χώρο μικρότερης διάστασης και η ταυτόχρονη ανακατασκευή των διανυσμάτων εισόδου. Στην πράξη είναι αρκετά δύσκολη η εκπαίδευση και ερμηνεία ενός τέτοιου μοντέλου αφού πρέπει να υπάρχει ισορροπία μεταξύ της ικανότητας του αυτοκωδικοποιητή να ανασκευάζει την είσοδο στην έξοδο καθώς μια απλή αντιγραφή της εισόδου στην έξοδο καθιστά τον χώρο προβολής του αυτοκωδικοποιητή άνευ περαιτέρω χρησιμότητας.

Το μοντέλο που υλοποιήσαμε αφορά έναν βαθύ αυτοκωδικοποιητή με τρία επίπεδα τόσο στον κωδικοποιητή όσο και στον αποκωδικοποιητή. Στην προσπάθεια μας να δημιουργήσουμε ένα πραγματικά χρήσιμο μοντέλο αυτοκωδικοποιητή έπρεπε να ορίσουμε την διάστασης του χώρου προβολής του διανύσματος εισόδου. Για τον σκοπό αυτό τρέξαμε πειράματα με διαφορετικές διαστάσεις προβολής από 5 έως 14. Για να αποβάλλουμε την στοχαστικότητα που διακατέχει τα μοντέλα στην διάρκεια της εκπαίδευσης και να έχουμε μια αρκετά ικανοποιητική καμπύλη τρέξαμε από 10 φορές το πείραμα για κάθε διάσταση και ορίσαμε το μέσο όρο σφάλματος των δοκιμών σαν την τιμή σφάλματος του μοντέλου. Όπως φαίνεται και στο παρακάτω διάγραμμα, το δίκτυο με διάσταση προβολής 10 είναι κοντά στο σημείο όπου στην καμπύλη αρχίζει να εμφανίζει κορεσμό, πράγμα που καθιστά το εν λόγω σημείο ικανοποιητική επιλογή.



Σχήμα 5.4: Αποτελέσματα πειράματος εκπαίδευσης του μοντέλου με διαφορετικές διαστάσεις στο επίπεδο κωδικοποίησης του. Ο χ-άξονας υποδηλώνει τις διαφορετικές διαστάσεις ενώ ο ψ-άξονας το μέσο λάθος καθενός μοντέλου σε 10 διαφορετικές εκφάνσεις της εκπαίδευσης. Η διάρκεια εκπαίδευσης κάθε πειράματος ήταν 10 εποχές.

Το επόμενο βήμα στην ρύθμιση υπέρ παραμέτρων του μοντέλου είναι να δημιουργήσουμε ένα μοντέλο όπου γενικεύει στο σύνολο επικύρωσης X_{val}. Το συγκεκριμένο είναι ιδιαίτερα σημαντικό καθώς μια κατάσταση στην οποία το μοντέλο υπερεκπαιδεύεται στο σύνολο εκπαίδευση X_{train}, καθιστά το μοντέλο ανίκανο να αναπαράγει το σύνολο επικύρωσης X_{val} και ως εκ τούτου θα γεννά μεγάλο αριθμό από ανωμαλίες στην έξοδο του συστήματος. Για τον λόγο αυτό εφαρμόσαμε την τεχνική *dropout* [25] στο πρώτο κρυφό επίπεδο του κωδικοποιητή καθώς και ένα επιπρόσθετο όρο ποινής στο τελευταίο επίπεδο της κωδικοποίησης. Σαν συναρτήσεις ενεργοποίησης του δικτύου χρησιμοποιήθηκε η $tanh(x)$ εκτός από τα τελευταία επίπεδα κωδικοποίησης και αποκωδικοποίησης που χρησιμοποιήθηκε η γραμμική συνάρτηση ενεργοποίησης. Παρακάτω φαίνεται το διάγραμμα εκπαίδευσης του μοντέλου.



Σχήμα 5.5: Διάγραμμα εκπαίδευσης αυτοκωδικοποιητή.

5.5 Ορισμός κατωφλιών απόφασης

Στην καρδία της ανίχνευσης ανωμαλιών είναι ο ορισμός των κατωφλιών απόφασης. Η συγκεκριμένη διαδικασία είναι αυτή που θα κρίνει το αποτέλεσμα του συστήματος και ως εκ τούτου χρειάζεται ιδιαίτερη προσοχή. Η διαδικασία συνοψίζεται στην εξαγωγή στατιστικών κριτηρίων από το σύνολο των σφαλμάτων του συνόλου $X_{threshold}$, όπως αυτό ορίστικε στην παράγραφό 5.3, με σκοπό τον ορισμό κατωφλιών απόφασης. Πιο συγκεκριμένα, ορίσαμε ένα κατώφλι για το σύνολο των χαρακτηριστικών $Threshold_{global}$ και ένα για κάθε ένα από τα χαρακτηριστικά μεμονωμένα $Threshold_i$. Ο τρόπος με τον οποίο ορίσαμε τα κατώφλια φαίνεται παρακάτω:

$$Threshold_i = \mu_i + scale * \sigma_i$$

$$Threshold_{global} = \mu_{global} + scale * \sigma_{global}$$

Όπου:

$Threshold_i$: Το κατώφλι απόφασης του χαρακτηριστικού ι

μ_i: Η μέση τιμή του σφάλματος του χαρακτηριστικού ι

σ_i : Η τυπική απόκλιση του σφάλματος του χαρακτηριστικού

$scale$:Η παράμετρος μεγένθυσης

Ενώ:

$Threshold_{global}$: Το κατώφλι απόφασης για το σύνολο των χαρακτηριστικών

μ_{global}: Η μέση τιμή του μέσου σφάλματος των χαρακτηριστικών
σ_{global} : Η τυπική απόκλιση του μέσου σφάλματος των χαρακτηριστικών
$scale$: Η παράμετρος μεγέθυνσης

Στα παραπάνω αξίζει να αναφέρουμε ότι επειδή το σφάλμα ανακατασκευής είναι θετικά ορισμένο, στα κατώφλια απόφασης συμπεριλαμβάνεται μόνο ένας προσθετικός όρος που σχετίζεται με τις τυπικές αποκλίσεις των σφαλμάτων.

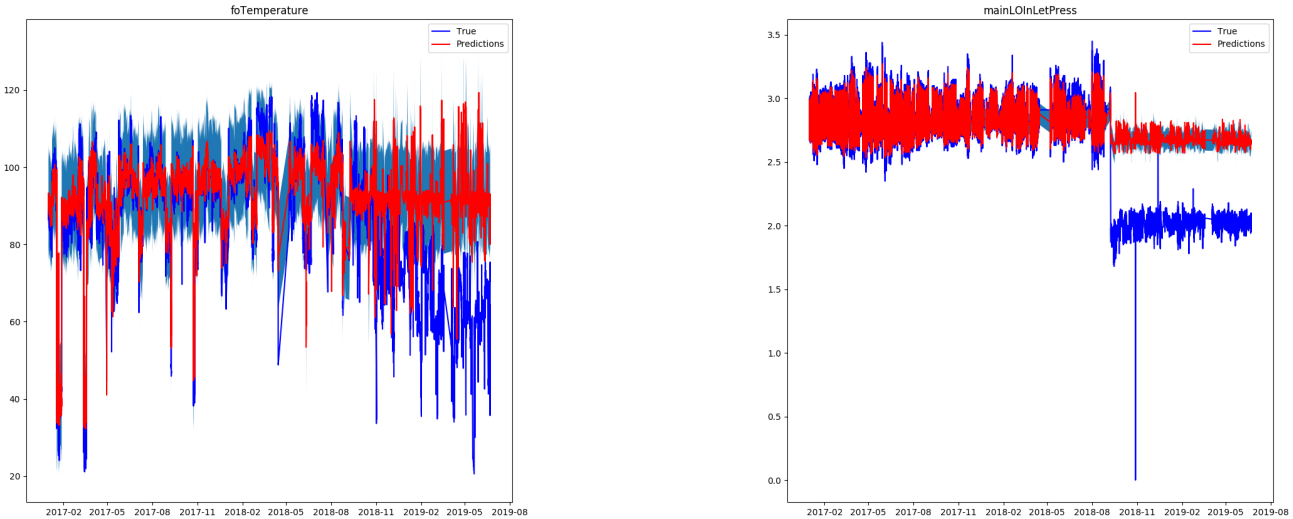
Ένα πρόβλημα, το οποίο προκύπτει σχετικά με τον ορισμό των κατωφλιών για το σύνολο των χαρακτικών $Threshold_{global}$ είναι ότι διαφορετικά χαρακτηριστικά έχουν διαφορετικό εύρος σφάλματος και έτσι η επιλογή της μέσης τιμής του σφάλματος είναι ίσως ένας ατυχής τρόπος προσέγγισης αφού το κατώφλι θα είναι περισσότερο ευαίσθητο σε χαρακτηριστικά με μεγαλύτερο σφάλμα. Έτσι, ένα νέο κατώφλι υπολογίζεται ως η μέση τιμή της $mahalanobis$ απόστασης των σημείων του συνόλου X_{thr} πολλαπλασιασμένη με έναν παράγοντά μεγέθυνσης όπως φαίνεται παρακάτω.

$$Threshold_{global} = scale * \mu_{MD}$$

Όπου:

$Threshold_{global}$: Το νέο κατώφλι απόφασης για το σύνολο των χαρακτηριστικών
μ_{MD}: Η μέση τιμή της $mahalanobis$ απόστασης του σφάλματος
$scale$: Η παράμετρος μεγένθυσης

Παρακάτω φαίνονται οι γραφικές αναπαραστάσεις των αποτελεσμάτων. Με μπλε χρώμα παρουσιάζονται τα πραγματικά σήματα ενώ με κόκκινο τα σήματα που προέκυψαν σαν έξοδος από το μοντέλο του αυτοκωδικοποιητή και με γαλάζιο χρώμα φαίνονται τα περιθώρια απόφασης.



Σχήμα 5.6: Διαγράμματα σύγκρισης προβλέψεων και πραγματικών τιμών

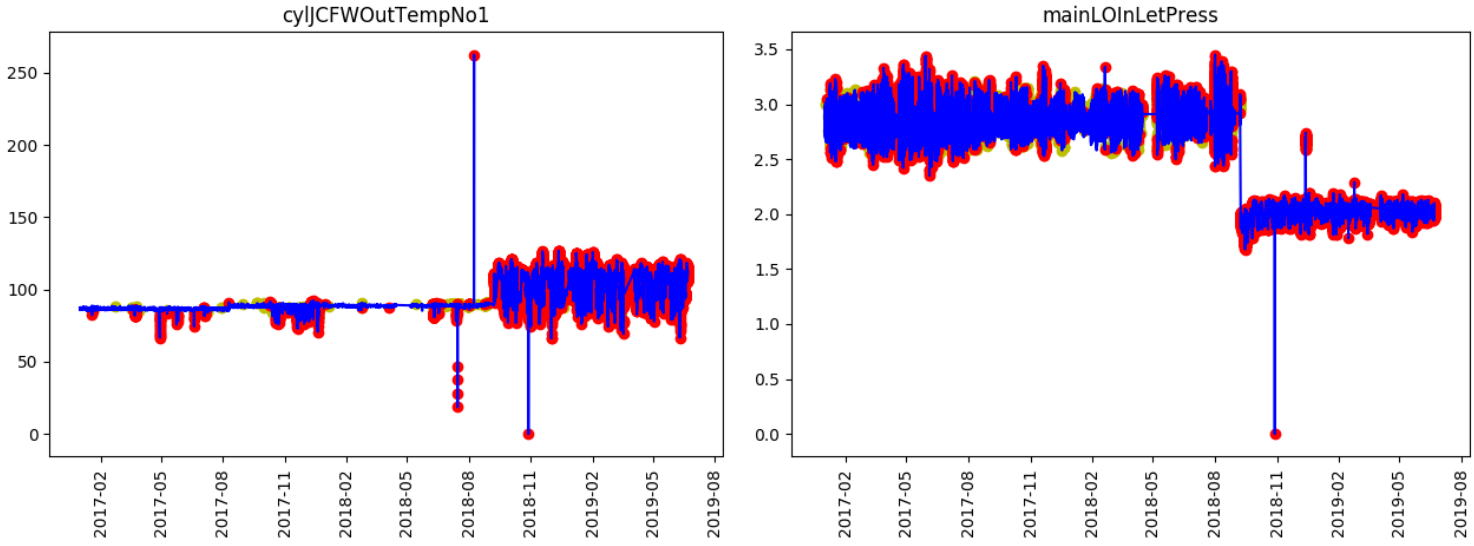
Από τα παραπάνω συμπεραίνουμε ότι αναλόγως με την ικανότητα του αυτοκωδικοποιητή να ανακατασκευάζει την είσοδο στην έξοδο έχουμε και διαφορετικό εύρος στα περιθώρια απόφασης. Με άλλα λόγια, αν ο αυτοκωδικοποιητής έχει την ικανότητα να αναπαράγει καλά το σήμα εισόδου κατά την διάρκεια της εκπαίδευσης η μέση τιμή και η τυπική απόκλιση του σφάλματος για ένα συγκεκριμένο χαρακτηριστικό είναι αρκετά μικρές με συνέπεια το σύστημα ανίχνευσης ανωμαλιών να είναι εξαιρετικά ευαίσθητο. Αυτό μας ωθεί στο γεγονός να ορίσουμε εμπειρικώς τα κατώφλια για κάθε ένα από τα χαρακτηριστικά ανεξάρτητα με το σφάλμα του μοντέλου.

5.6 Παρουσίαση αποτελεσμάτων

Η παρουσίαση αποτελεσμάτων είναι ιδιαίτερα σημαντική τόσο ώστε να κατανοήσουμε τις αδυναμίες του συστήματος όσο και για την επικύρωση των αποτελεσμάτων καθώς το σύνολο δεδομένων δεν είναι επισημειωμένο. Σε αυτήν την ενότητα, θα παρουσιαστούν τα αποτελέσματα της ανίχνευσης για τους δυο τρόπους με τους οποίους ορίστηκαν τα περιθώρια απόφασης. Τα αποτελέσματα θα παρουσιαστούν για ορισμένα από τα χαρακτηριστά του συνόλου δεδομένων καθώς επίσης σε δυο διαφορετικές κλίμακες ώστε να έχουμε καλύτερη εποπτεία των αποτελεσμάτων.

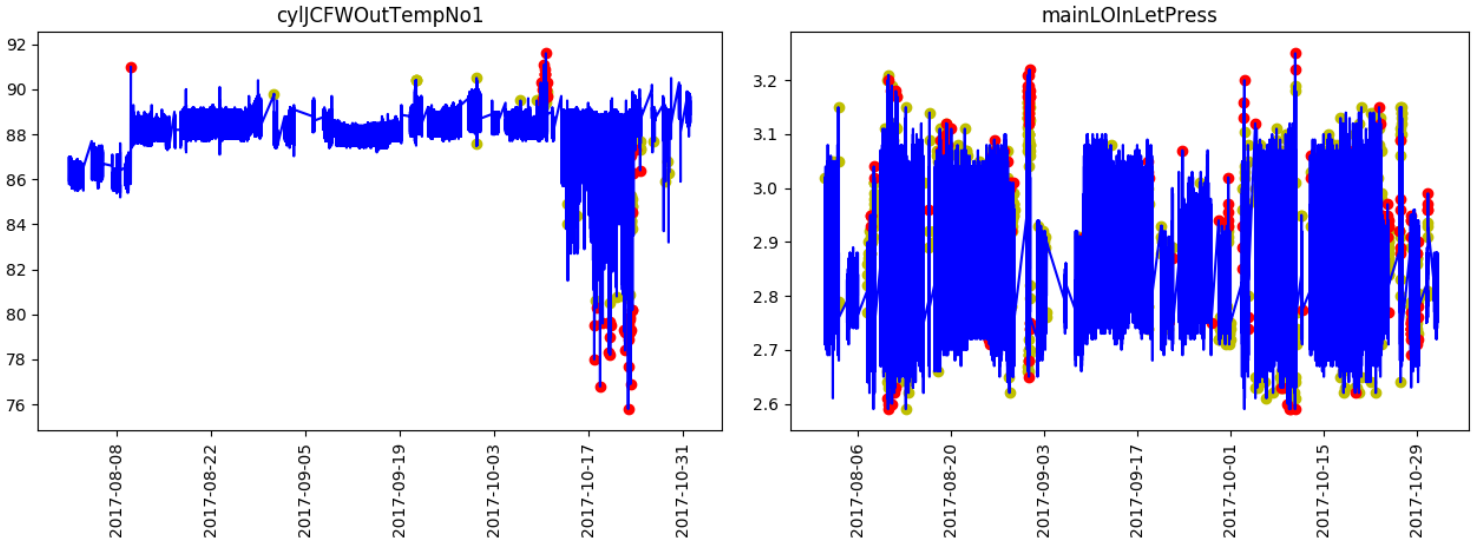
5.6.1 Αποτελέσματα στατιστικού κριτηρίου απόφασης

Παρακάτω φαίνονται τα αποτελέσματα σε όλη την διάρκεια του συστήματος.



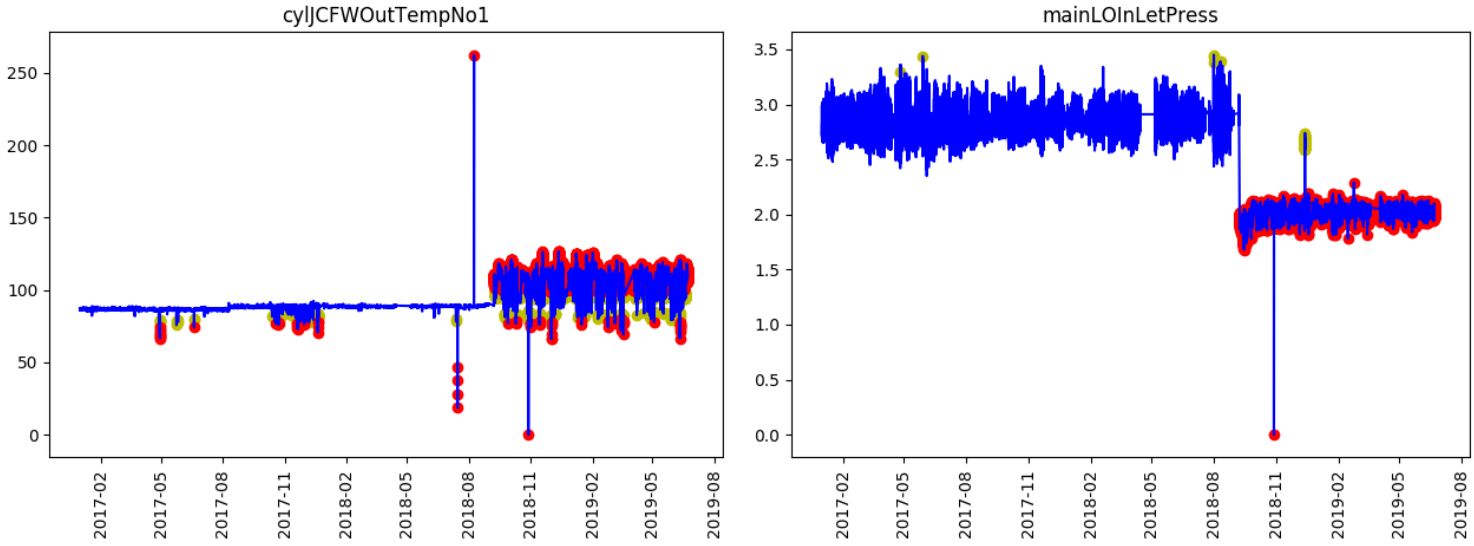
Σχήμα 5.7: Διάγραμμα ανίχνευσης ανωμαλιών αυτοκωδικοποιητή με εφαρμογή στατιστικού κριτηρίου στο σύνολο των δεδομένων.

Παρακάτω φαίνονται τα αποτελέσματα του συστήματος σε κλίμακα τριών μηνών.



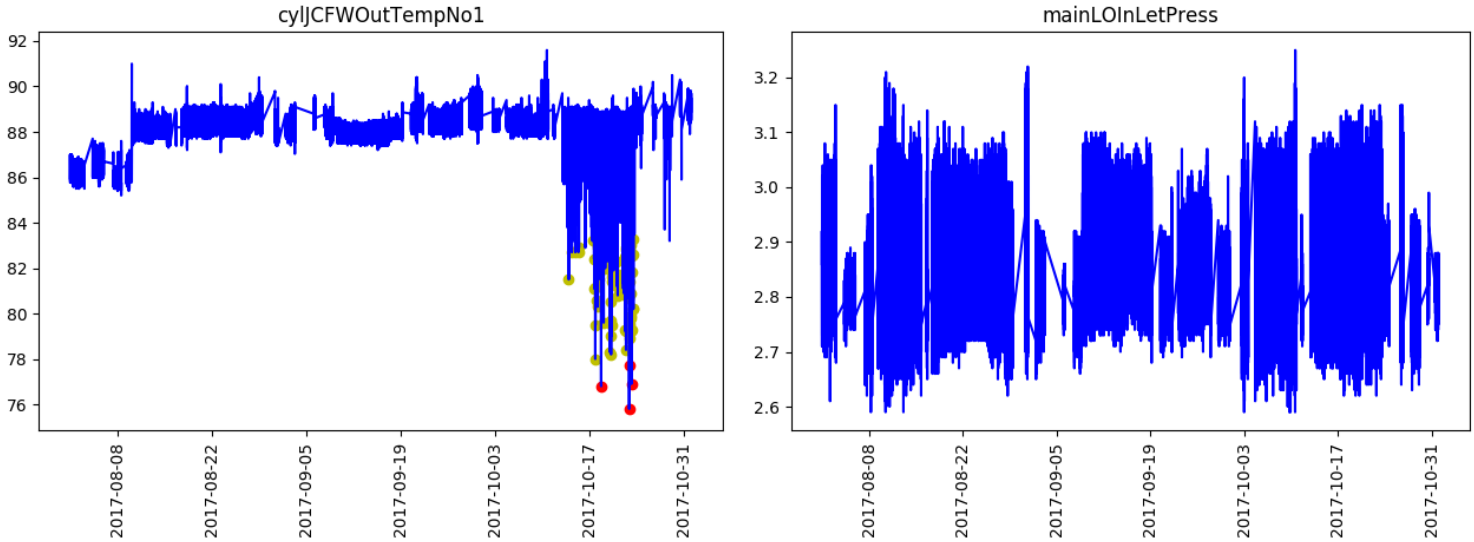
Σχήμα 5.8: Διάγραμμα ανίχνευσης ανωμαλιών αυτοκωδικοποιητή με εφαρμογή στατιστικού κριτηρίου σε διάρκεια 3 μηνών

5.6.2 Αποτελέσματα εμπειρικού κριτηρίου απόφασης



Σχήμα 5.9: Διάγραμμα ανίχνευσης ανωμαλιών αυτοκωδικοποιητή με εφαρμογή εμπειρικού κριτηρίου στο σύνολο των δεδομένων.

Παρακάτω φαίνονται τα αποτελέσματα του συστήματος σε κλίμακα τριών μηνών.



Σχήμα 5.10: Διάγραμμα ανίχνευσης ανωμαλιών αυτοκωδικοποιητή με εφαρμογή εμπειρικού κριτηρίου σε διάρκεια 3 μηνών

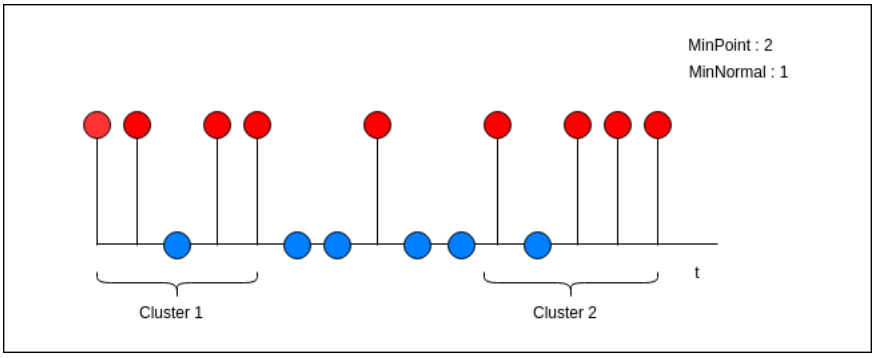
5.6.3 Σύγκριση αποτελεσμάτων χρήσης διαφορετικών κατωφλιών απόφασης

Όπως φαίνεται και από τα παραπάνω διαγράμματα, η χρήση στατιστικού κριτηρίου απόφασης παρουσιάζει μεγαλύτερο αριθμό από ανωμαλίες πράγμα που γίνεται ιδιαίτερα αντιληπτό στην περίπτωση του διαγράμματος 5.8 όπου το δεύτερο χαρακτηριστικό παρουσιάζει ανωμαλίες πολύ κοντά στις συνθήκες τις συνήθους λειτουργίας. Σε γενικές γραμμές είναι εξαιρετικά δύσκολο να βρούμε ένα κατώφλι απόφασης που να δίνει καλά αποτελέσματα ειδικά στην περίπτωση μη επιβλεπόμενης μάθησης. Πολλές φορές, στην βιβλιογραφία το κατώφλι απόφασης ορίζεται ελαχιστοποιώντας κάποιο κριτήριο που βασίζεται στην ακρίβεια των αποτελεσμάτων του συστήματος κάνοντας χρήση των ετικετών πράγμα που καθιστά την διαδικασία επιβλεπόμενη ανεξάρτητα με τον αν το μοντέλο κάνει χρήση των διαθέσιμων ετικετών. Σε αντίθεση, το εμπειρικό κριτήριο επιλέχτηκε μελετώντας τις διακυμάνσεις κάθε χαρακτηριστικού ξεχωριστά και όχι ερευνώντας τα αποτελέσματα της ανίχνευσης. Στην συνέχεια, γίνεται χρήση του στατιστικού κριτηρίου απόφασης και τεχνικές οι οποίες δίνουν καλύτερα αποτελέσματα με μια περισσότερο φορμαλιστική σκοπιά.

5.7 Διάγνωση ανωμαλιών

Όπως αναφέραμε και στην παράγραφο 3.2 υπάρχουν τρεις γενικές κατηγορίες στις οποίες κατηγοριοποιούνται οι ανωμαλίες των δεδομένων, στιγμιαίες ανωμαλίες, ανωμαλίες υπό συνθήκη και ομάδες από συνεχόμενες ανωμαλίες. Υποθέτοντας ότι ο αυτοκωδικοποιητής είναι ικανός να ανιχνεύει τις στιγμιαίες ανωμαλίες το επόμενο βήμα είναι να ορίσουμε έναν συστηματικό τρόπο για την ανίχνευση ομάδων με μεγάλο ποσοστό από συνεχόμενες ανωμαλίες. Το παραπάνω είναι ιδιαίτερα σημαντικό διότι τα δεδομένα που έχουμε στην διάθεσή μας είναι ιδιαίτερα θορυβώδη και έτσι ο αυτοκωδικοποιητής επισημειώνει ως ανωμαλίες ένα μεγάλο ποσοστό του συνόλου των δεδομένων. Επίσης για να είναι το σύστημα ανίχνευσης ανωμαλιών πραγματικά χρήσιμο στην ανάλυση και την επίβλεψη του πλοίου θα πρέπει να παρουσιάζει συστάδες από ανωμαλίες με σχετικά μεγάλη διάρκεια.

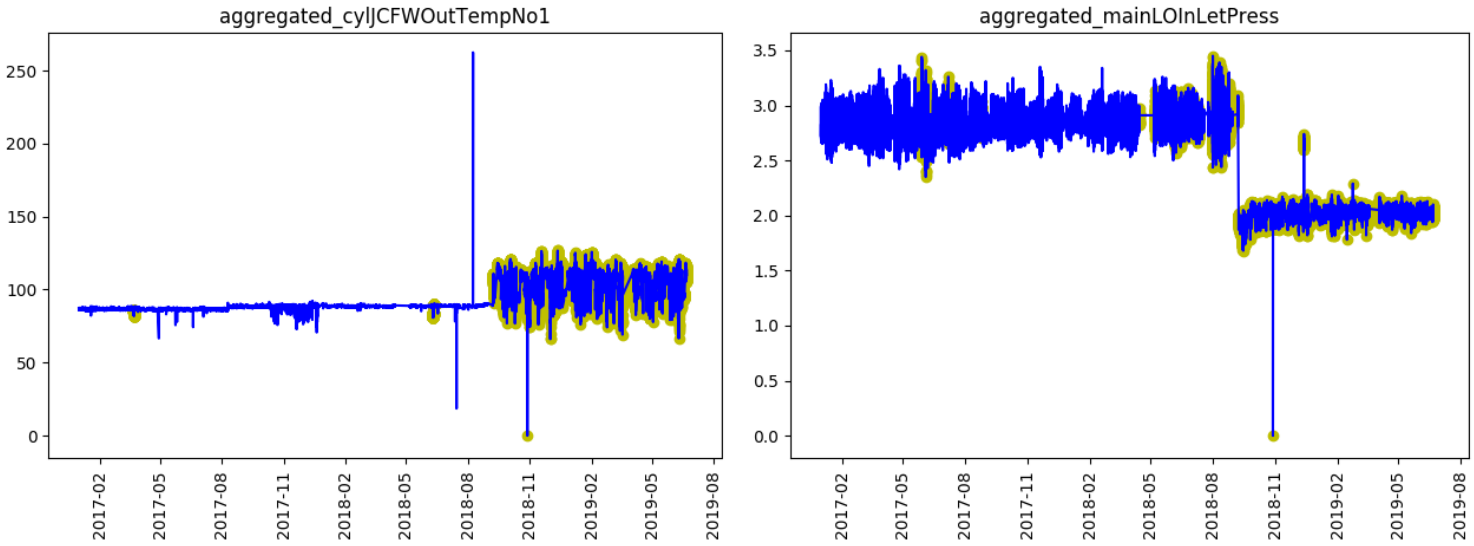
Για την ανίχνευση ομάδων από ανωμαλίες εφαρμόσαμε έναν απλό αλγόριθμο που ομαδοποιεί τις ανωμαλίες όπως περιγράφεται παρακάτω. Ο αλγόριθμος έχει σαν παραμέτρους τον ελάχιστο αριθμό $MinPoints$ από σημεία ώστε να οριστεί μια ομάδα καθώς και τον αριθμό από σημεία κανονικής λειτουργίας $NumNormal$ που επιτρέπονται ανάμεσα στα μη προσδοκώμενα δεδομένα. Στην συνέχεια διατρέχουμε τα αποτελέσματα της ανίχνευσης του μοντέλου του αυτοκωδικοποιητή, στον χρόνο, μετρώντας τον αριθμό των ανωμάλων δεδομένων μέχρι να βρούμε συνεχόμενα $NumNormal$, σημεία κανονικής λειτουργίας, οπότε αν ο αριθμός των ανώμαλων δεδομένων ξεπερνά των $MinPoints$ συντελείται μια συστάδα. Ένα παράδειγμα του παραπάνω αλγορίθμου φαίνεται παρακάτω.



Σχήμα 5.11: Παράδειγμα αλγορίθμου ομαδοποίησης ανωμαλιών

Έπειτα από πειραματική μελέτη ορίσαμε τα $MinPoints = 60$ και $NumNormal = 10$ και τα αποτελέσματα φαίνονται παρακάτω.

5.7.1 Αποτελέσματα διάγνωσης ανωμαλιών



Σχήμα 5.12: Αποτελέσματα διάγνωσης ανωμαλιών

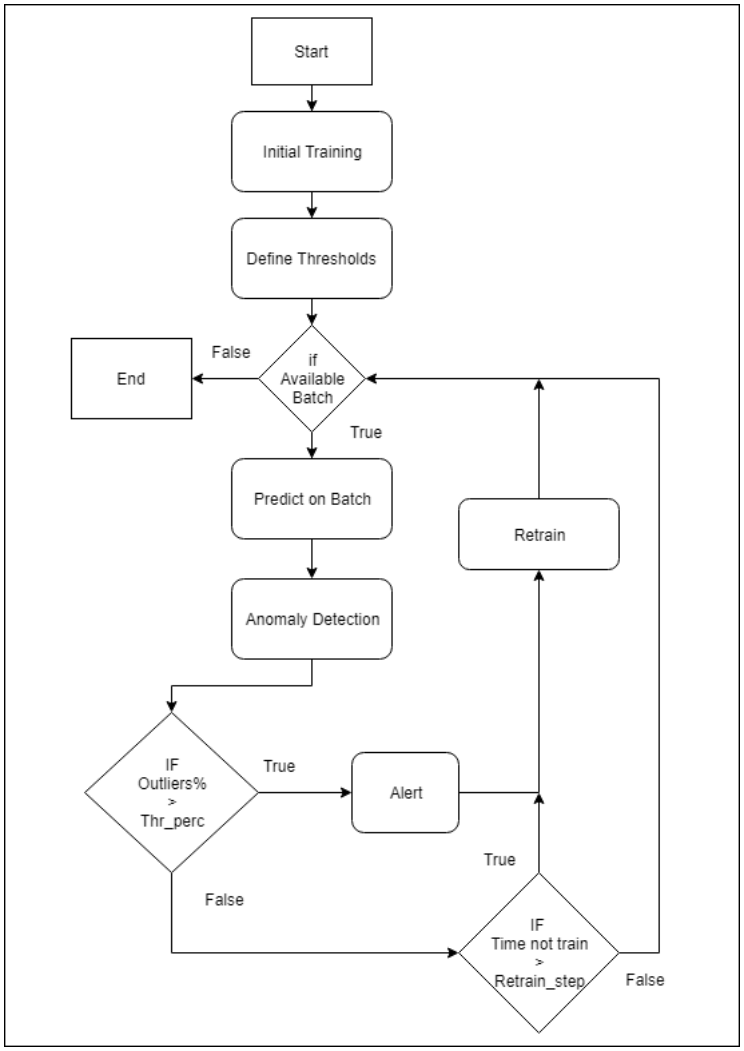
Παραπάνω παρατηρούμε ότι ο αλγόριθμος κατάφερε να δημιουργήσει συστάδες ανωμαλιών περιορίζοντας τον αριθμό των στιγμιαίων ανωμαλιών. Ένα αρνητικό στα παραπάνω γραφήματα είναι ότι στην περίπτωση του πρώτου χαρακτηριστικού η χρονική περίοδος 2017-10 - 2018-01 παύει να θεωρείται ανώμαλη ενώ παράλληλα μετά την συγκεκριμένη χρονική περίοδο εμφανίζεται βλάβη. Το συγκεκριμένο γεγονός μας υποδεικνύει ότι βλάβες μπορεί να ανιχνεύονται από περιοδικώς εμφανιζόμενες στιγμιαίες ανωμαλίες, επίσης γίνεται σαφές ότι η προεπεξεργασία των δεδομένων είναι ιδιαίτερα σημαντική, αφού η χρήση τεχνικών ομαλοποίησης θα είχε παρόμοιο αντίκτυπο στα αποτελέσματα του μοντέλου. Έτσι, στον συγκεκριμένο αλγόριθμο πρέπει να μελετηθούν διεξοδικά οι υπερπαραμέτροι ώστε να μην χάνουμε πληροφορία στην διαδικασία της ομαδοποίησης.

5.8 Ανίχνευση ανωμαλιών στην διάρκεια του χρόνου

Ένα από τα προβλήματα των παραπάνω μεθόδων είναι ότι αδυνατούν να προσομοιώσουν την πραγματικότητα υπό την έννοια ότι η εκπαίδευση και η πρόβλεψη των αποτελεσμάτων του μοντέλου γίνεται από μια φορά και σε έναν αρκετά μεγάλο όγκο δεδομένων. Από τα γραφήματα του κεφαλαίου 4 γίνεται σαφές επίσης ότι τα δεδομένα μπορεί να αλλάξουν κατανομή στην διάρκεια του χρόνου. Το παραπάνω δεν σημαίνει απαραίτητα ότι η αλλαγή στην κατανομή των δεδομένων αποτελεί βλάβη άλλα είναι ιδιαίτερα χρήσιμη πληροφορία για ένα σύστημα επίβλεψης.

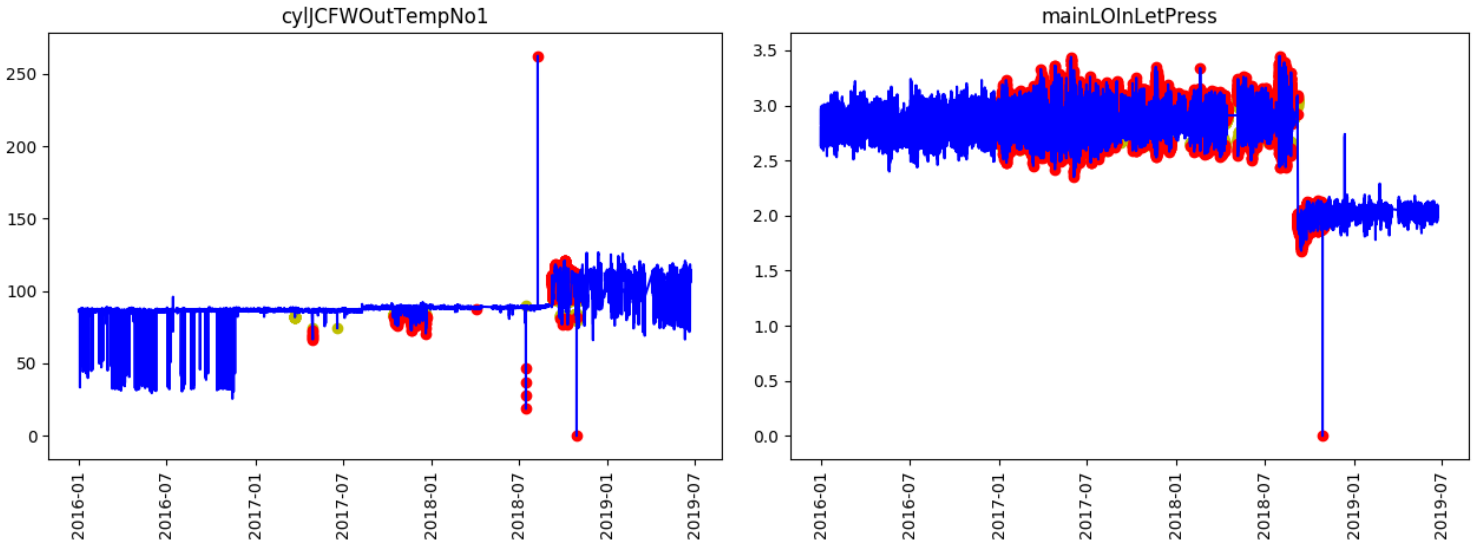
Με στόχο λοιπόν να προσομοιώσουμε την πραγματικότητα και να κάνουμε το σύστημα περισσότερο ρεαλιστικό εκπαιδεύσαμε τον αυτοκωδικοποιητή στο πέρασμα του χρόνου σύμφωνα με την διαδικασία που περιγράφεται παρακάτω. Αρχικά ορίσαμε χρονολογικά ένα αρχικό

διάστημα X_{init} στο οποίο εκπαιδεύσαμε τον αυτοκωδικοποιητή. Στην συνέχεια χωρίσαμε το εναπομένον σύνολο δεδομένων σε διαστήματα (πχ.δυο μηνών) $X_{batch,i}$ τα οποία σταδιακά εφαρμόζουμε στο μοντέλο και εντοπίζουμε τις ανωμαλίες. Σε περίπτωση που στο συγκεκριμένο διάστημα το ποσοστό των ανωμαλιών ξεπερνά ένα συγκεκριμένο ποσοστό $Threshold_{percent}$ τότε θεωρούμε ότι η κατανομή των δεδομένων μας έχει αλλάξει οπότε αναφέρουμε την περιοχή και στην συνέχεια επανεκπαιδεύουμε το μοντέλο. Σε περίπτωση που m συνεχόμενα διαστήματα $X_{batch,i}$ δεν παρουσίασαν ποσοστό ανωμαλιών μεγαλύτερο από $Threshold_{percent}$, τότε επανεκπαιδεύουμε το μοντέλο ώστε να είναι περισσότερο ακριβές σε περαιτέρω προβλέψεις. Όλη η διαδικασία περιγράφεται σχηματικά στο διάγραμμα ροής 5.13.



Σχήμα 5.13: Διάγραμμα ροής συστήματος ανίχνευσης ανωμαλιών στην διάρκεια του χρόνου

5.8.1 Αποτελέσματα ανίχνευσης στην διάρκεια του χρόνου



Σχήμα 5.14: Αποτελέσματα εκπαίδευσης στον χρόνο

Το συγκεκριμένο αποτέλεσμα δεν είναι τόσο ικανοποιητικό καθώς στο δεύτερο χαρακτηριστικό του παραπάνω σχήματος εμφανίζεται μεγάλος αριθμός από ανωμαλίες. Το πρόβλημα έγκειται στο κριτήριο απόφασης όπως έχουμε προαναφέρει. Σε αντίθεση το πρώτο χαρακτηριστικό έχει καλύτερα αποτελέσματα καθώς στις τελευταίες στιγμές της χρονοσειράς τα δεδομένα σταματούν να είναι ανώμαλα αφού στο δίκτυο έγινε επανεκπαίδευση θεωρώντας ότι μετά από συγκεκριμένο χρόνο οι ανωμαλίες αφορούν αλλαγές στην κατανομή. Παράλληλα η υπόλοιπη χρονοσειρά παρουσιάζει ενδιαφέρουσα αποτελέσματα ενώ πρέπει να σημειωθεί ότι πριν το 2017 δεν έχει εφαρμοστεί κάποιος αλγόριθμος ανίχνευσης ανωμαλιών.

5.9 Εκπαίδευση ανατροφοδοτούμενου δικτύου

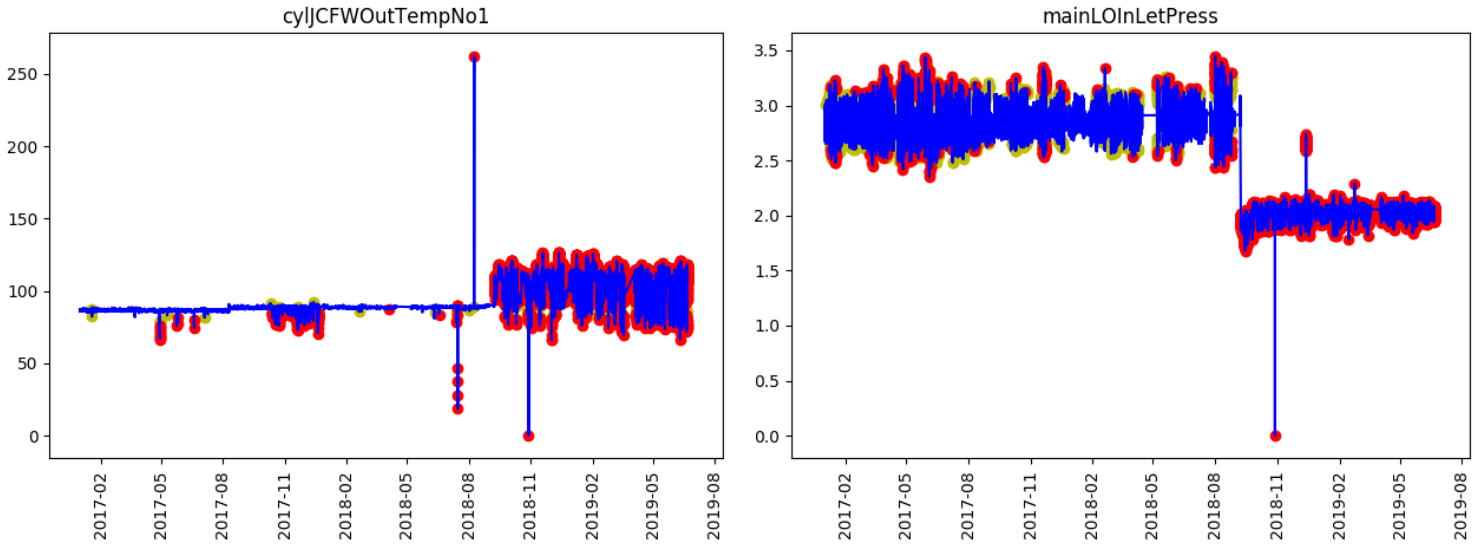
Ένα από τα μεγαλύτερα προβλήματα που αντιμετωπίζουμε στην παρούσα διπλωματική εργασία είναι ότι το σύνολο δεδομένων μας δεν είναι επισημειωμένο και συνεπώς δεν μπορούμε να αξιολογήσουμε τα αποτέλεσμα των αλγορίθμων. Γίνεται λοιπόν επιτακτική η ανάγκη στο να εφαρμόσουμε διαφορετικά είδη μοντέλων ώστε να συγκρίνουμε να αποτελέσματα μας, αφού η χρήση κάθε μοντέλου προδιαθέτει έμμεσα τα αποτελέσματα. Έπειτα, ο αυτοκωδικοποιητής δεν έχει τρόπο να μοντελοποιήσει τις ανωμαλίες υπό συνθήκη καθώς δεν έχει πληροφορία σχετικά με τα προηγούμενα δείγματα στο πέρασμα του χρόνου. Έτσι, ως δεύτερο μοντέλο η επιλογή μας είναι ένα ανατροφοδοτούμενο νευρωνικό δίκτυο. Η επιλογή έγινε με βάση ότι τα ανατροφοδοτούμενα νευρωνικά δίκτυα και συγκεκριμένα το $LSTM$ είναι ικανά να κωδικοποιήσουν τις εξαρτήσεις μιας ακολουθίας εισόδου καθώς διατηρούν ένα είδος μνήμης στην αρχιτεκτονική τους όπως περιγράφεται αναλυτικότερα στο 3.6.2.

Από τεχνικής σκοπιάς, ορίσαμε ένα $LSTM$ με σταθερό μέγεθος ακολουθίας 10 χρονικών

στιγμών με στόχο την πρόβλεψη της επόμενης χρονικής στιγμής. Ο λόγος που επιλέχτηκε το συγκεκριμένο μέγεθος ακολουθίας είναι επειδή τα συγκεκριμένα δίκτυα δεν μπορούν να κωδικοποιήσουν εξαρτήσεις μακράς διάρκειας λόγω του προβλήματος εξαφάνισης της παραγώγου (*vanishing gradient*) στην διάρκεια της εκπαίδευσης, ενώ παράλληλα το συγκεκριμένο μέγεθος είναι σύμφωνο και με την φυσική του προβλήματος. Έτσι, σε περίπτωση που το μοντέλο μας αδυνατεί να αναπαράγει τις τιμές της επόμενης χρονικής στιγμής τότε θεωρούμε την συγκεκριμένη τιμή ως απροσδόκητη. Η διαδικασία ορισμού των κατωφλιών απόφασης για την ανίχνευση μη προσδοκώμενων τιμών είναι παρόμοια με αυτήν που περιγράφτηκε στην ενότητα 5.2.

5.9.1 Αποτελέσματα ανατροφοδοτούμενου δικτύου

Παρακάτω φαίνονται τα αποτελέσματα του ανατροφοδοτούμενου δικτύου. Τα αποτελέσματα του σχήματος 5.15 είναι αρκετά ικανοποιητικά όσον αφορά το πρώτο χαρακτηριστικό καθώς χωρίς ιδιαίτερη δυσκολία στην ρύθμιση των υπερπαραμέτρων έχουμε λογικό αριθμό από ανώμαλα δεδομένα. Όσον αφορά το δεύτερο χαρακτηριστικό τα αποτελέσματα εκ πρώτης όψεως δεν έχουν μεγάλη διαφορά από αυτά του απλού αυτοκωδικοποιητή της ενότητας 5.6.1, πάρα ταύτα ο όγκος τους είναι 5 % λιγότερος σε ποσοστό ανωμάλων δεδομένων.



Σχήμα 5.15: αποτελέσματα ανατροφοδοτούμενου δικτύου

Κεφάλαιο 6

Πειραματική ανάλυση Ανταγωνιστικού αυτοκωδικοποιητή

6.1 Εισαγωγή

Στο προηγούμενο κεφάλαιο αναλύθηκε ο αυτοκωδικοποιητής και το ανατροφοδοτούμενο νευρωνικό δίκτυο για την ανίχνευση ανωμαλιών. Σε αυτό το κεφάλαιο θα γίνει πειραματική ανάλυση ενός Ανταγωνιστικού αυτοκωδικοποιητή στην προσπάθεια μας να βελτιώσουμε τα αποτελέσματα. Όπως παρουσιάστηκε και στα προηγούμενα κεφάλαια τα προβλήματα του αυτοκωδικοποιητή είναι ότι πρώτον ο χώρος προβολής είναι κακώς σχηματισμένος με αποτέλεσμα ανώμαλα δεδομένα να προβάλλονται γειτονικά με τα καλώς ορισμένα δεδομένα και κατά συνέπεια η ανίχνευση ανωμαλιών γίνεται αποκλειστικά μέσω του σφάλματος ανακατασκευής. Το δεύτερο πρόβλημα που υπάρχει είναι ότι τα δεδομένα εισόδου περιέχουν ένα ποσοστό ανωμαλιών που πιθανώς να μοντελοποιείται κατά την διάρκεια της εκπαίδευσης. Στην περίπτωση του ανταγωνιστικού αυτοκωδικοποιητή ο χώρος προβολής είναι καλύτερα σχηματισμένος από την άποψη ότι γειτονικά σημεία στην είσοδο προβάλλονται κοντά στην προσπάθεια του αυτοκωδικοποιητή να ακολουθήσει την επιβάλλουσα κατανομή και κατά συνέπεια μπορεί να χρησιμοποιηθεί για την ανίχνευση ανωμαλιών.

Υλοποιήσαμε λοιπόν 2 πειράματα που αφορούν τον ανταγωνιστικό αυτοκωδικοποιητή. Στο πρώτο πείραμα, έγινε προσπάθεια ανίχνευσης ανωμαλιών μέσω συσταδοποίησης όπως αυτή παρουσιάστηκε στο κεφάλαιο 3 ενώ στο δεύτερο πείραμα, γίνονται προσπάθειες για ανίχνευση ανωμαλιών με χρήση του σφάλματος ανακατασκευής και του χώρου προβολής του αυτοκωδικοποιητή.

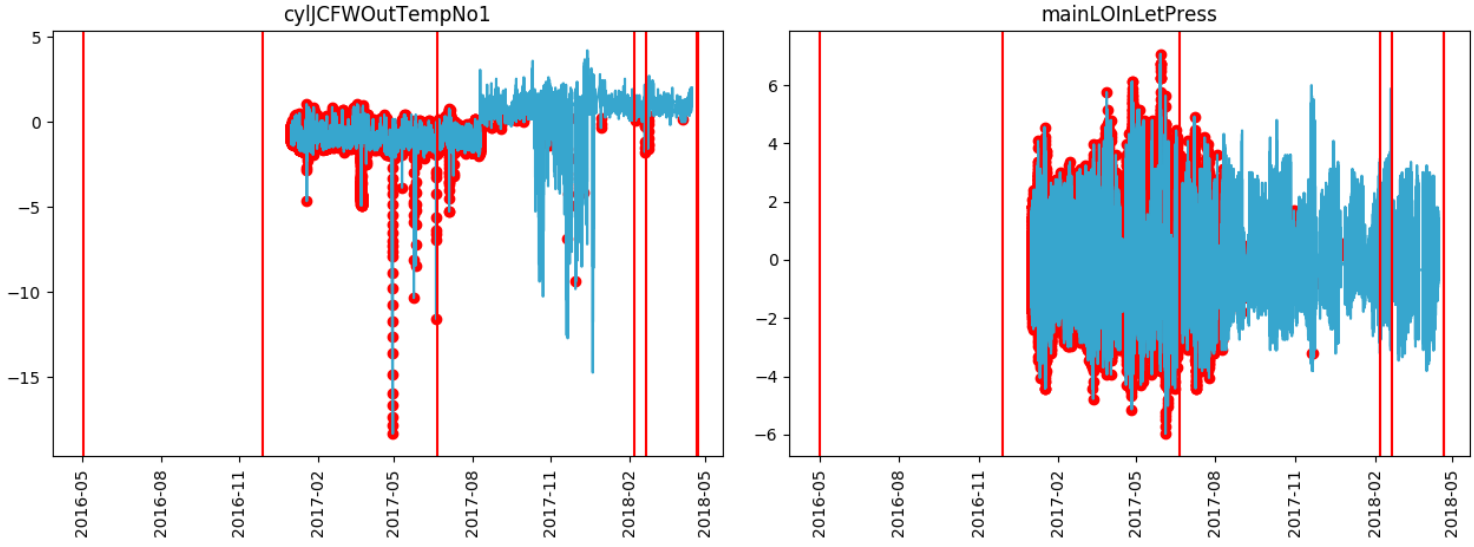
6.2 Ανίχνευση ανωμαλιών μέσω συσταδοποίησης και χρήσης ανταγωνιστικού αυτοκωδικοποιητή

Στο συγκεκριμένο πείραμα, ερευνούμε την ικανότητα του ανταγωνιστικού αυτοκωδικοποιητή να ανιχνεύει ανώμαλα δεδομένα μέσω ενός κατηγορικού διανύσματος στην έξοδο του κωδικοποιητή του συστήματος. Συγκεκριμένα, υλοποιήσαμε την αρχιτεκτονική του σχήματος 3.13 και κάναμε τρία πειράματα με διαφορετικές κατηγορικές κατανομές ενώ η επιβαλλόμενη κατανομή που αφορά το διάνυσμα προβολής του κωδικοποιητή ήταν σε όλα τα πειράματα μια τυπική γκαουσιανή κατανομή διάστασης δέκα. Οι εποχές που εκπαιδεύσαμε το δίκτυο ήταν 10, το μέγεθος του συνόλου εισόδου 32 και ως βελτιστοποιητές για όλα τα επιμέρους στοιχεία του δικτύου $Nadam$ με τιμές μάθησης 0.0025. Οι τρεις διαφορετικές κατηγορικές κατανομές που εφαρμόσαμε στα τρία διαφορετικά πειράματα ήταν δυαδικές διακριτές κατανομές με πιθανότητες $p_1(X_0) = p_1(X_1) = 0.5$, $p_2(X_0) = 0.8$, $p_2(X_1) = 0.2$ και $p_3(X_0) = 0,95$, $p_3(X_1) = 0.05$.

6.2.1 Πείραμα συσταδοποίησης $p_1(X_0) = p_1(X_1) = 0.5$

Το πρώτο πείραμα έγινε με σκοπό να ερευνήσουμε την ικανότητα του ανταγωνιστικού αυτοκωδικοποιητή στην συσταδοποίηση ολόκληρης της χρονοσειράς αφού τα ποσοστά της κατηγορικής κατανομής δεν εκφράζουν την πεποίθησή μας για ανώμαλα δεδομένα.

Τα αποτελέσματα της συσταδοποίησης φαίνονται παρακάτω στα οποία οι κλάσεις έχουν ποσοστά 0.55 και 0.45 αντίστοιχα ενώ με κόκκινες κάθετες γραμμές φαίνονται οι καταγραμμένες βλάβες.

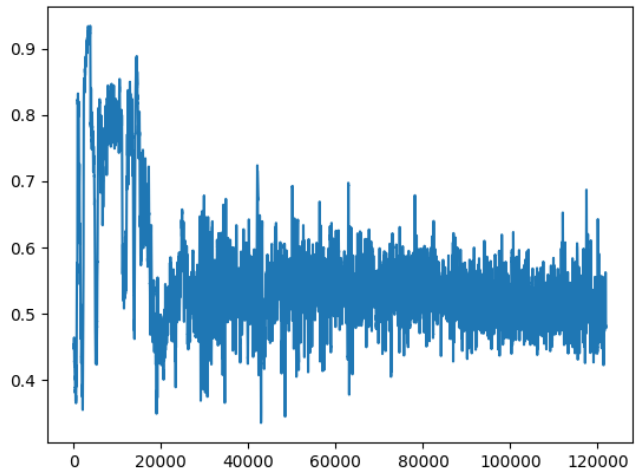


Σχήμα 6.1: Διάγραμμα ανίχνευσης ανωμαλιών αυτοκωδικοποιητή με εφαρμογή συσταδοποίησης 1

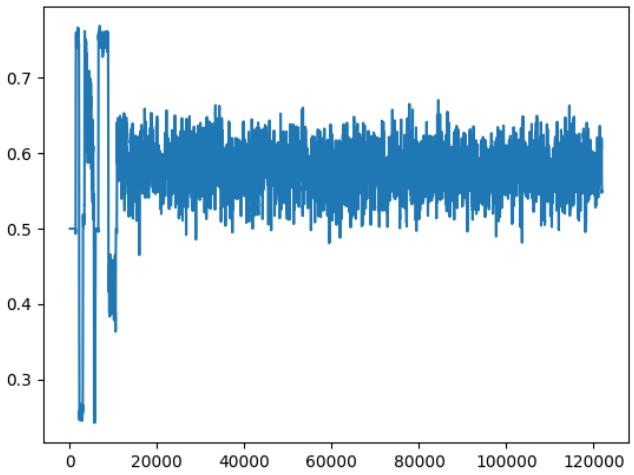
Όπως φαίνεται παραπάνω τα αποτελέσματα της συσταδοποίησης είναι η δημιουργία δύο

χρονολογικώς καλώς ορισμένων συστάδων στις οποίες υπάρχει εμφανής αλλαγή των κατανομών των δυο χαρακτηριστικών. Τα περισσότερα από τα υπόλοιπα χαρακτηριστικά του συνόλου εκπαίδευσης υφίστανται παρόμοια μεταβολή στην κατανομή τους. Σύμφωνα με τα παραπάνω μπορούμε να συμπεράνουμε ότι συγκεκριμένο δίκτυο είναι ικανό να δημιουργεί συστάδες σε χρονοσειρές. Στην συνέχεια θα ερευνήσουμε την συμπεριφορά του δικτύου σε διαφορετικές πρότερες πεποιθήσεις για συστάδες στο σύνολο δεδομένων.

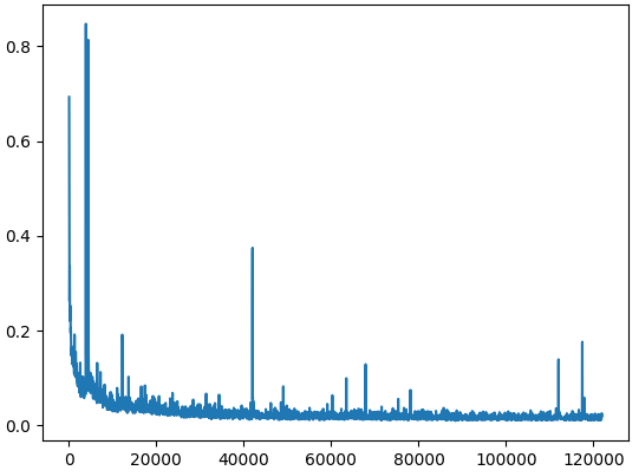
Όσον αφορά την διαδικασία της εκπαίδευσης παρακάτω φαίνονται τα διαγράμματα της εκπαίδευσης του δικτύου.



(α′) Δίκτυο διαχωριστή συνεχούς κατανομής



(β′) Δίκτυο διαχωριστή διακριτής κατανομής



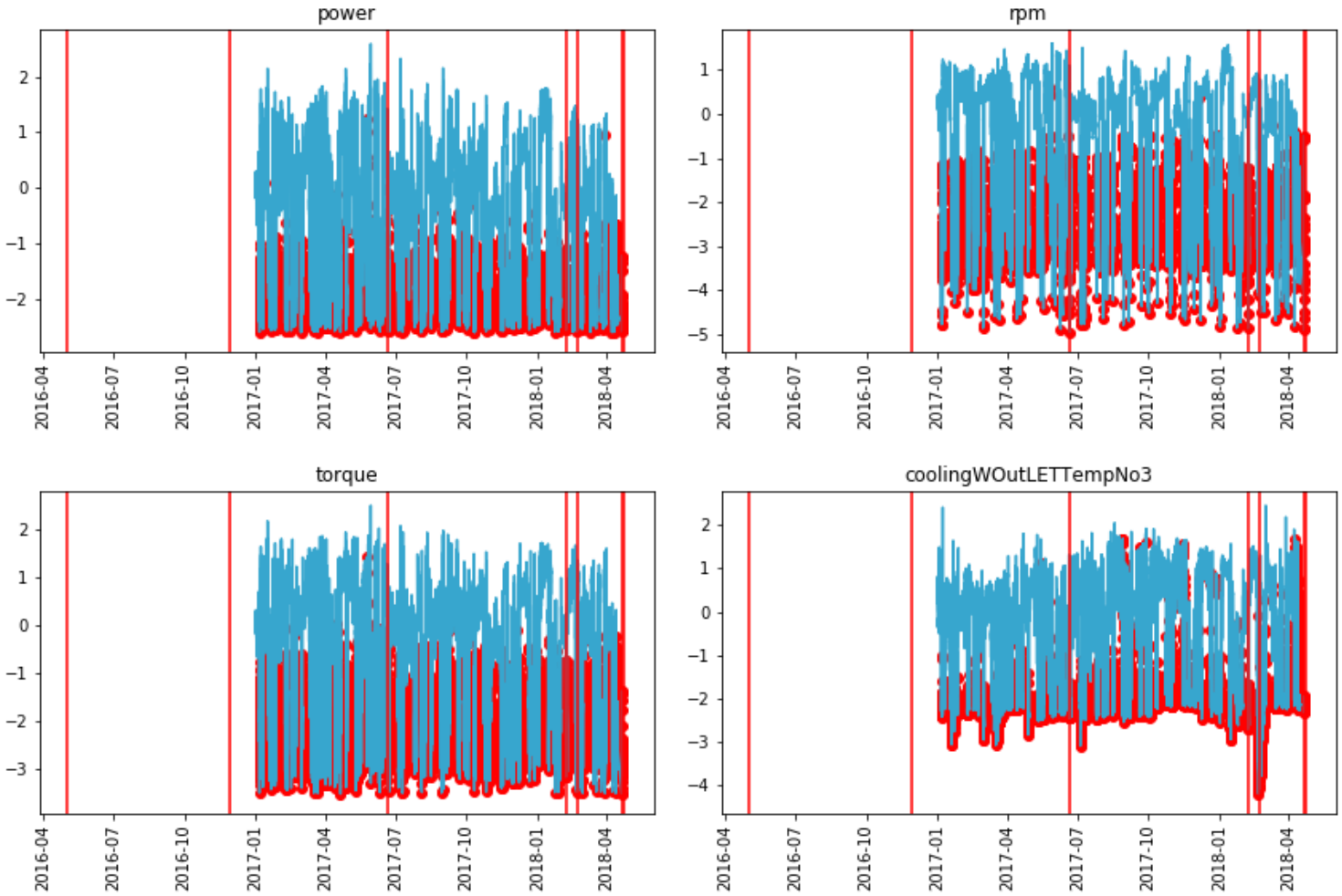
(γ′) Δίκτυο Αυτοκωδικοποιητή

Σχήμα 6.2: Διάγραμμα εκπαίδευσης ανταγωνιστικού αυτοκωδικοποιητή

6.2.2 Πείραμα συσταδοποίησης $p_2(X_0) = 0.8, p_2(X_1) = 0.2$

Σαν ένα δεύτερο πείραμα επιλέξαμε νέες πιθανότητες για την διακριτή κατανομή $p_2(X_0) = 0.8, p_2(X_1) = 0.2$.

Τα αποτελέσματα της συσταδοποίησης φαίνονται παρακάτω στα οποία οι κλάσεις έχουν ποσοστά 0.77 και 0.23 αντίστοιχα ενώ με κόκκινες κάθετες γραμμές φαίνονται οι καταγραμμένες βλάβες.



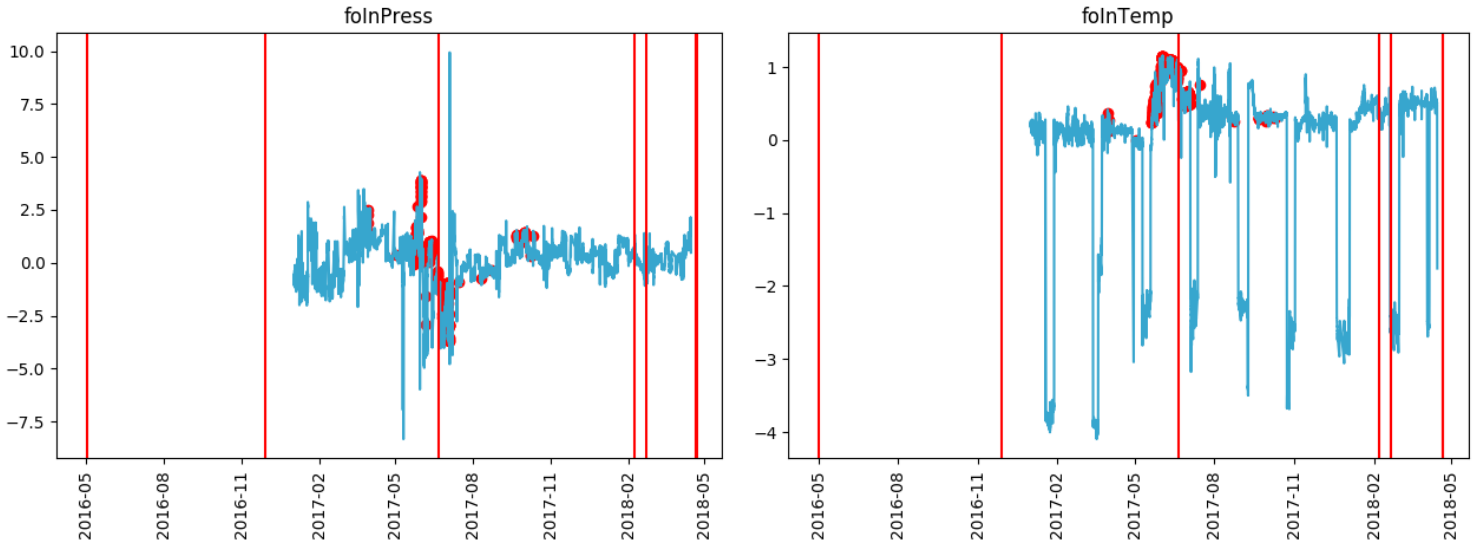
Σχήμα 6.3: Διάγραμμα ανίχνευσης ανωμαλιών αυτοκωδικοποιητή με εφαρμογή συσταδοποίησης 2

Στα παραπάνω διαγράμματα φαίνεται ότι οι συστάδες οργανώθηκαν με βάση την ενεργειακή κατάσταση του σκάφους. Συγκεκριμένα η μεγαλύτερη συστάδα ορίζεται σαν την συστάδα υψηλής ενεργειακής κατανάλωσης ενώ η μικρότερη συστάδα υποδεικνύει καταστάσεις όπου το σύστημα του πλοίου πλέει με χαμηλότερες στροφές ανά λεπτό.

6.2.3 Πείραμα συσταδοποίησης $p_3(X_0) = 0,95, p_3(X_1) = 0.05$

Το τελευταίο πείραμα συσταδοποίησης εστιάζει στην ανίχνευση ανωμαλιών. Τα ποσοστά που αφορούν πρότερη πεποίθηση μας σχετικά με τα ανώμαλα δεδομένα είναι της τάξεως του 0.05 ενώ τα κανονικώς παραγόμενα δεδομένα αφορούν το 0.95 του συνόλου των δεδομένων.

Τα αποτελέσματα της συσταδοποίησης φαίνονται παρακάτω στα οποία οι κλάσεις έχουν ποσοστά 0.98 και 0.02 αντίστοιχα ενώ με κόκκινες κάθετες γραμμές φαίνονται οι καταγραμμένες βλάβες.



Σχήμα 6.4: Διάγραμμα ανίχνευσης ανωμαλιών αυτοκωδικοποιητή με εφαρμογή συσταδοποίησης 3

Παραπάνω φαίνεται ότι η συσταδοποίηση είχε σαν αποτέλεσμα την ανίχνευση μιας βλάβης. Τα χαρακτηριστικά που φαίνονται παραπάνω είναι η θερμοκρασία και η πίεση του καυσίμου τα οποία φαίνονται να παρουσιάζουν ασυνήθιστη συμπεριφορά λίγο πριν μια επερχόμενη βλάβη. Η παραπάνω διαδικασία δεν μπορεί βέβαια να λειτουργήσει σαν ανιχνευτής ανωμαλιών σε πραγματικό χρόνο αφού η συσταδοποίηση απαιτεί έναν αρκετά μεγάλο αριθμό από συνεχόμενες ανωμαλίες ώστε να τις ανιχνεύσει ενώ παράλληλα υπάρχει μεγάλος βαθμός στοχαστικότητας στην δημιουργία των συστάδων. Πάρα ταύτα μπορεί να βοηθήσει στην κατανόηση των προβλημάτων που οδήγησαν στην βλάβη και η συγκεκριμένη πληροφορία να χρησιμοποιηθεί σε νέο σύστημα για την ανίχνευση βλαβών σε μοντέλα ήμι-επιβλεπόμενης μάθησης.

6.2.4 Συμπεράσματα και περιορισμοί της συσταδοποίησης

Σαν συμπέρασμα από την διαδικασία της συσταδοποίησης έχουμε ότι το δίκτυο του ανταγωνιστικού αυτοκωδικοποιητή είναι ικανό να δημιουργεί σύνθετες συστάδες αναλόγως με την πεποίθηση μας για τις εκάστοτε ομάδες. Όσο περισσότερο δυσανάλογες είναι οι πιθανότητες των ομάδων τόσο περισσότερο στοχαστική γίνεται και η διαδικασία της εκπαίδευσης με αποτέλεσμα οι ίδιες υπέρ παράμετροι του μοντέλου να δίνουν διαφορετικά αποτελέσματα. Ο

λόγος είναι ότι υπάρχουν πολλοί διαφορετικοί τρόποι το δίκτυο να δημιουργεί τις συστάδες. Ακόμα σε ότι αφορά το τελευταίο πείραμα πολλές φορές το δίκτυο δημιουργούσε μια μοναδική συστάδα στα δεδομένα. Το τελευταίο αποτελεί έναν περιορισμό του δικτύου και ενδιαφέρουσα είναι η έρευνα που θα μπορούσε να γίνει για την βελτίωση της ικανότητας του αυτοκωδικοποιητή να δημιουργεί αρκετές μικρές συστάδες στα δεδομένα.

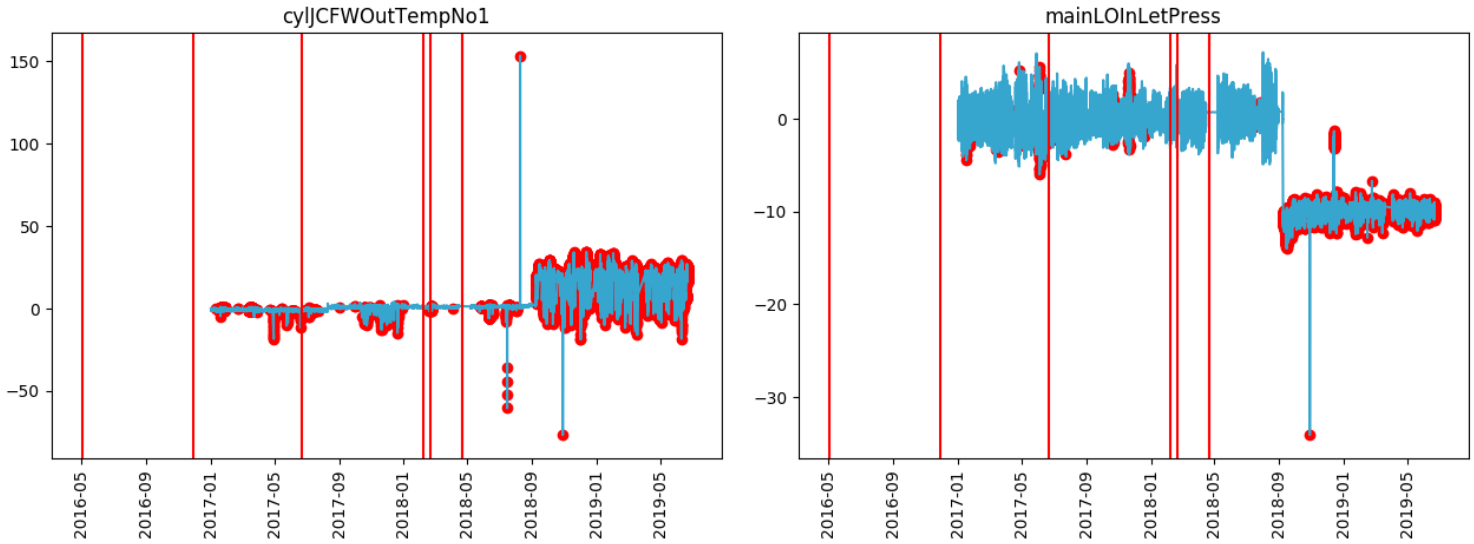
6.3 Ανίχνευση ανωμαλιών με χρήση ανταγωνιστικού αυτοκωδικοποιητή

Σαν δεύτερο πείραμα μελετούμε την ικανότητα του δικτύου του αυτοκωδικοποιητή να ανιχνεύει ανωμαλίες στα δεδομένα μέσω του ενδιάμεσου χώρου προβολής που πλέον ακολουθεί την επιβαλλόμενη κατανομή και συγκεκριμένα μια τυπική γκαουσιανή κατανομή δέκα διαστάσεων. Το δίκτυο είναι το ίδιο με του προηγουμένου πειράματος. 3.13.

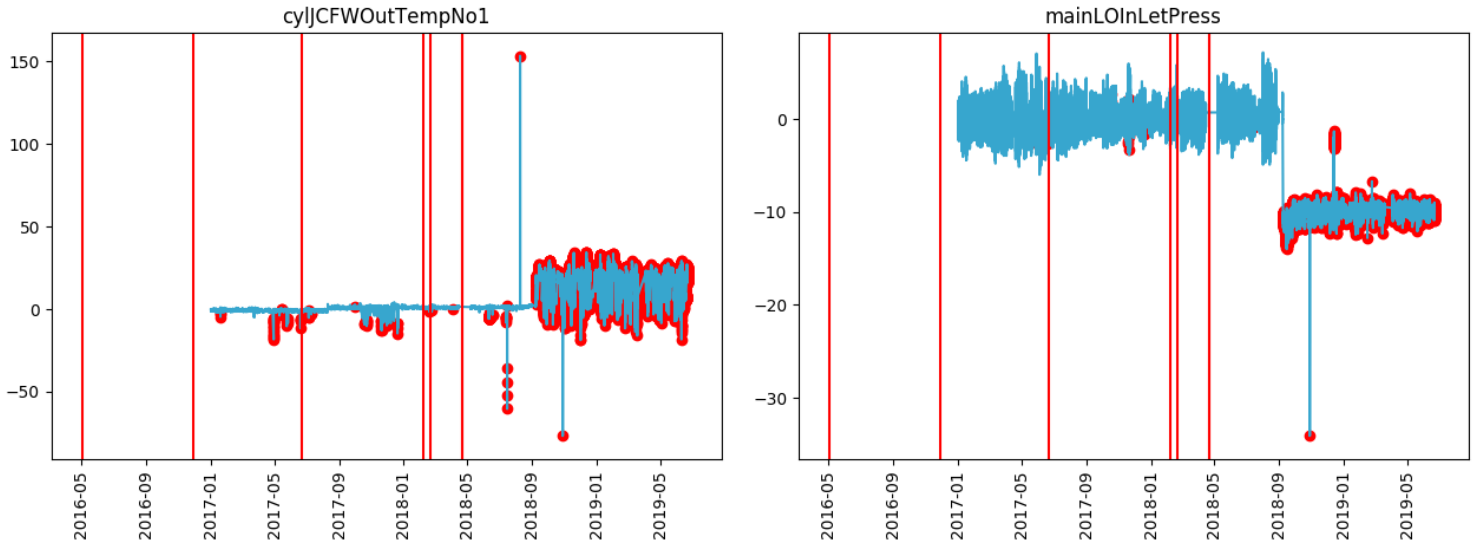
6.3.1 Ανίχνευση ανωμαλιών μέσω του χώρου προβολής αυτοκωδικοποιητή

Η ανίχνευση ανωμαλιών μέσω του χώρου προβολής του αυτοκωδικοποιητή είναι ικανή αφού η διαδικασία της εκπαίδευσης αναγκάζει τον αυτοκωδικοποιητή παρόμοια δεδομένα να προβάλλονται κοντά στον χώρο ώστε να μπορέσει εν τέλει να ακολουθήσει την επιθυμητή κατανομή. Ο τρόπος με τον όποιο γίνεται η ανίχνευση περιγράφεται στο τρίτο κεφάλαιο. Περισσότερο αναλυτικά θέσαμε δυο διαφορετικά κατώφλια αφού είχε πλέον σχηματιστεί ο χώρος προβολής ανάλογως με την πιθανοφάνεια των αποτελεσμάτων $P(\hat{Z})$. Τα δυο κατώφλια αποτελούν οι τιμές $P(x_i)$ που δίνουν τα δυο διαφορετικά διανύσματα $x_i = i * x_{μον}$ όπου το $x_{μον}$ αποτελεί το μοναδιαίο διάνυσμα δέκα διαστάσεων και $i \in \{2, 3\}$. Μια ακόμα προσπάθεια έγινε με χρήση τις $mahalanobis$ απόστασης των δεδομένων.

Τα αποτελέσματα για τα δυο διαφορετικά κατώφλια φαίνονται παρακάτω και τα ποσοστά τους σε ανώμαλα δεδομένα είναι 20 % και 15 %.



Σχήμα 6.5: Διάγραμμα ανίχνευσης ανωμαλιών αυτοκωδικοποιητή με εκμετάλλευση χώρου προβολής 1

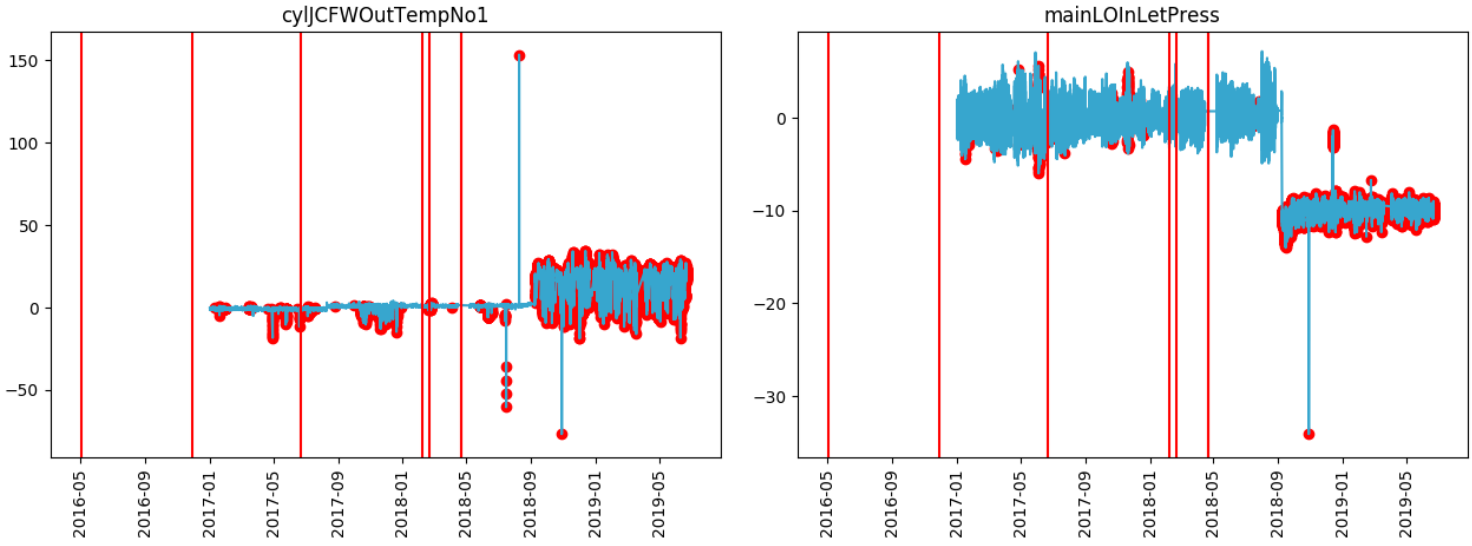


Σχήμα 6.6: Διάγραμμα ανίχνευσης ανωμαλιών ανταγωνιστικού αυτοκωδικοποιητή με εκμετάλλευση χώρου προβολής 2

Τα παραπάνω ποσοστά είναι αρκετά κοντά, διότι το τελευταίο χρονολογικά κομμάτι της χρονοσειράς αποτελεί ένα μεγάλο ποσοστό του συνόλου των δεδομένων και το όποιο δεν υπήρχε στο σύνολο εκπαίδευσης. Το γεγονός ότι το συγκεκριμένο σύνολο δεν αποτελούσε μέρος του συνόλου εκπαίδευσης είναι ιδιαίτερα ενθαρρυντικό καθώς η προβολή του είναι εκτός των ορίων των κατωφλιών ενώ η σειρά είναι αρκετά διαφορετική από τα υπόλοιπα δεδομένα. Έτσι, μπορούμε να υποθέτουμε ότι σημεία εκτός κατανομής θα προβάλλονται διαφορετικά από τα δεδομένα εκπαίδευσης πράγμα το όποιο δεν συνέβαινε με τον απλό αυτοκωδικοποιητή του

προηγουμένου κεφαλαίου.

Τα αποτελέσματα με χρήση της απόστασης *mahalanobis* φαίνονται παρακάτω:

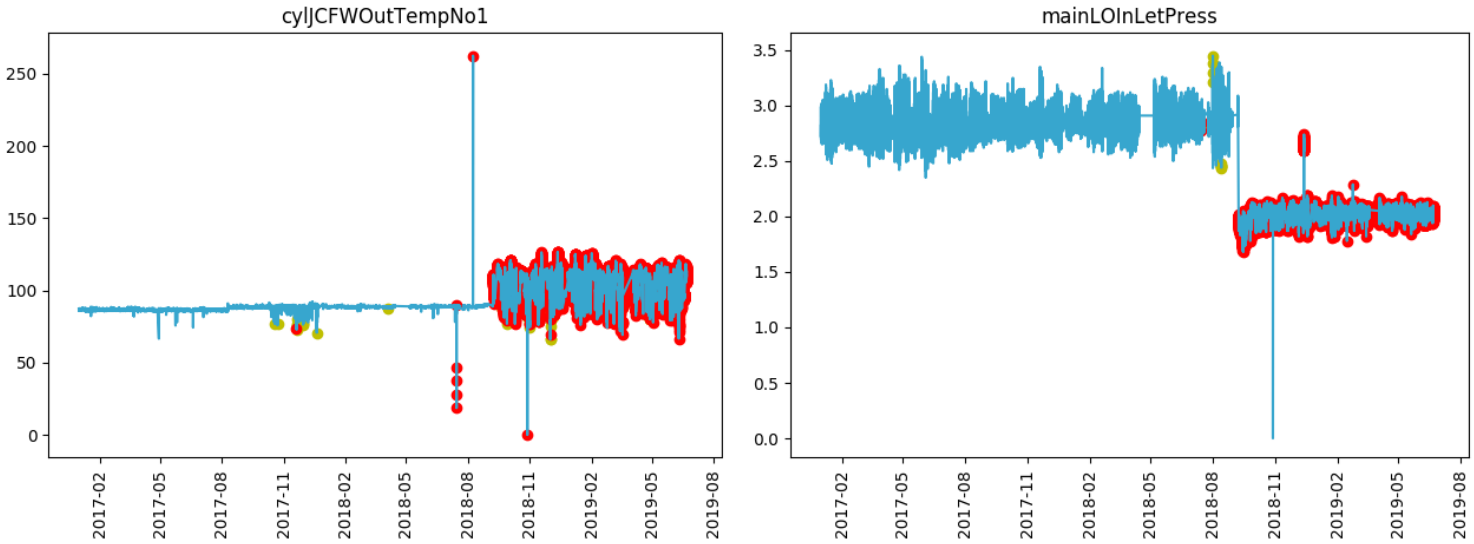


Σχήμα 6.7: Διάγραμμα ανίχνευσης ανωμαλιών ανταγωνιστικού αυτοκωδικοποιητή με εκμετάλλευση χώρου προβολής και την *mahalanobis* απόσταση

Τα παραπάνω αποτελέσματα μοιάζουν αρκετά με τα προηγούμενα καθώς αλλάξαμε μόνο τον τρόπο που ορίζονται τα κατώφλια απόφασης και η κατανομή έμεινε η ίδια. Το παραπάνω είναι ιδιαίτερα χρήσιμο καθώς πλέον έχουμε μια καλώς ορισμένη κατανομή όπου τα κατώφλια απόφασης δεν επηρεάζονται από τα ανώμαλα δεδομένα.

6.3.2 Ανίχνευση ανωμαλιών μέσω σφάλματος ανακατασκευής αυτοκωδικοποιητή

Με σκοπό την σύγκριση των αποτελεσμάτων της ανίχνευσης ανωμαλιών του αυτοκωδικοποιητή του προηγουμένου κεφαλαίου και του παρόντος παραθέτουμε τα αποτελέσματα του δεύτερου με χρήση των εμπειρικώς ορισμένων κατωφλιών όπως παρουσιάστηκε στην παράγραφο 5.6.2.



Σχήμα 6.8: Διάγραμμα ανίχνευσης ανωμαλιών ανταγωνιστικού αυτοκωδικοποιητή με χρήση σφάλματος ανακατασκευής

Τα αποτελέσματα είναι αρκετά όμοια με αυτά του 5.6.2 πράγμα ιδιαίτερα ικανοποιητικό και για τα δύο συστήματα. Το ενδιαφέρον είναι ότι στο δεύτερο δίκτυο δεν εφαρμόσαμε τεχνικές ομαλοποίησης που εφαρμόσαμε στο πρώτο δίκτυο. Ο λόγος που έχουμε παρόμοια αποτελέσματα είναι ότι τον παράγοντα της ομαλοποίησης, που αποτρέπει το δίκτυο στο να μαθαίνει τις πιθανές ανωμαλίες των δεδομένων, εισάγουν πλέον τα ανταγωνιστικά δίκτυα.

Κεφάλαιο 7

Επίλογος

Σαν τελευταίο κεφάλαιο της διπλωματικής είναι ιδιαίτερα σημαντικό να παρουσιάσουμε τα συμπεράσματα των αποτελεσμάτων, καθώς επίσης να αναφέρουμε πιθανές μελλοντικές επεκτάσεις οι οποίες θα μπορούσαν να βελτιώσουν την ποιότητα των αποτελεσμάτων του συστήματος ανίχνευσης ανωμαλιών.

7.1 Σύνοψη και συμπεράσματα

Στην παρούσα διπλωματική εργασία έγινε έρευνα σχετικά με την ανίχνευση ανωμαλιών σε δεδομένα που αφορούν μηχανολογικά χαρακτηριστά πλοίων εμπορικού σκοπού. Υλοποιήσαμε έναν αυτοκωδικοποιητή και ένα ανατροφοδοτούμενο νευρωνικό δίκτυο και τα εμπλουτίσαμε με υποσυστήματα ώστε να ξεπεράσουμε ορισμένες δυσκολίες που παρουσιάστηκαν στην πορεία. Συγκεκριμένα, αρχίσαμε με την υλοποίηση του αυτοκωδικοποιητή και ορίσαμε στατικώς κατώφλια απόφασης για την ανίχνευση ανωμαλιών. Τα συγκεκριμένα κατώφλια ήταν δύσκολο να ορισθούν με συνέπεια αρκετά από τα δεδομένα εισόδου να χαρακτηρίζονται σαν ανώμαλα. Παρατηρήσαμε ότι ο αυτοκωδικοποιητής επισήμανε καλά στιγμιαίες ανωμαλίες (point anomalies) οι οποίες συμβαίνουν αρκετά συχνά στον χρόνο. Για τον λόγο αυτό, εφαρμόσαμε στα αποτελέσματα έναν αλγόριθμο ομαδοποίησης ώστε να επισημαίνουμε περιοχές με συνεχόμενες ανωμαλίες. Στην συνέχεια, ορίσαμε εμπειρικά τα κατώφλια λαμβάνοντας υπόψιν την φυσική στην οποία υπόκειται κάθε χαρακτηριστικό ξεχωριστά με τα οποία λάβαμε πολύ ικανοποιητικά αποτελέσματα μειώνοντας τον αριθμό των ανωμαλιών. Το επόμενο βήμα ήταν να εκπαιδεύσουμε τον αυτοκωδικοποιητή στο πέρασμα του χρόνου με σκοπό να εντοπίσουμε περιοχές όπου αλλάζει η κατανομή των δεδομένων, χαρακτηριστικό ιδιαίτερα σημαντικό για ένα σύστημα εντοπισμού ανωμαλιών. Για την επίτευξη των παραπάνω χωρίσαμε το σύνολο δεδομένων σε περιοχές εκπαιδεύοντας περιοδικά το μοντέλο λαμβάνοντας υπόψιν τα αποτελέσματα της ανίχνευσης ανωμαλιών. Με στόχο την περαιτέρω επαλήθευση των αποτελεσμάτων υλοποιήσαμε ένα ανατροφοδοτούμενο νευρωνικό δίκτυο όπου με χρήση ενός παράθυρου σταθερού μεγέθους προέβλεπε την επομένη χρονική στιγμή. Εφαρμόζοντας ξανά κατώφλια απόφασης εντοπίσαμε τις ανωμαλίες στα δεδομένα που αυτή την φορά ήταν λιγότερες, αφού το συγκεκριμένο μοντέλο ήταν ικανό να μοντελοποιεί καλύτερα τις ανωμαλίες δεδομένων των συνθηκών

(contextual anomalies) αλλά και των συνεχόμενων ανωμαλιών με αποτέλεσμα λιγότερη επεξεργασία των αποτελεσμάτων του δικτύου. Έτσι, το δεύτερο δίκτυο αναδείχθηκε περισσότερο ικανό αφού με λιγότερη ρύθμιση στις υπέρ-παραμέτρους κατάφερε να δώσει περισσότερο σταθερά αποτελέσματα στην διαδικασία της ανίχνευσης. Υλοποιήσαμε έναν ανταγωνιστικό αυτοκωδικοποιητή του οποίου το πλεονέκτημα είναι ότι ο χώρος προβολής του αναγκάζεται να ακολουθεί μια επιθυμητή κατανομή μέσω των ανταγωνιστικών δικτύων. Έτσι, πέρα από το σφάλμα ανακατασκευής είχαμε στην διάθεση μας τον χώρο προβολής του αυτοκωδικοποιητή για ανίχνευση ανωμαλιών. Τα αποτελέσματα ήταν ιδιαίτερα ικανοποιητικά. Τέλος, ερευνήσαμε την ικανότητα του παραπάνω αυτοκωδικοποιητή στο να δημιουργεί συστάδες από δεδομένα μέσω μιας επιβαλλόμενης διακριτής κατανομής στην έξοδο της κωδικοποίησης. Εφαρμόσαμε διαφορετικές πιθανότητες στις συστάδες που αναγκάσαμε τον αυτοκωδικοποιητή να δημιουργεί και παρατηρήσαμε αύξηση της στοχαστικότητας των αποτελεσμάτων όσο περισσότερο δυσανάλογες ήταν οι πιθανότητες των κλάσεων.

7.2 Μελλοντικές επεκτάσεις

Τα δυο κυριότερα προβλήματα που αντιμετωπίσαμε είναι ότι στο συγκεκριμένο σύνολο δεδομένων δεν υπάρχουν ετικέτες εκπαίδευσης και ότι περιέχει μεγάλο ποσοστό θορύβου καθώς οι μετρήσεις που έχουμε στην διάθεσή μας προέρχονται από σύστημα αισθητήρων σε ένα αρκετά απρόβλεπτο περιβάλλον. Τα παραπάνω προσδίδουν μεγάλη αβεβαιότητα σχετικά με τα αποτελέσματα των μοντέλων. Μια ιδέα για μελλοντικές επεκτάσεις της εργασίας είναι υλοποίηση ενός συστήματος ενεργής μάθησης (Active Learning) [20]. Όπως αναφέρεται και στο παραπάνω άρθρο πολλές φορές το πρόβλημα της μη επιβλεπόμενης ανίχνευσης ανωμαλιών είναι αδύνατο να λυθεί χωρίς τη χρήση κάποιας πρότερης γνώσης. Οι συγγραφείς του άρθρου προτείνουν ένα σύστημα το οποίο θα λειτουργεί συνεργατικά με τους χρήστες του οι οποίοι θα είναι ικανά καταρτισμένοι ώστε να διορθώνουν τα λάθη του μοντέλου. Με αυτόν τον τρόπο αναδεικνύεται μια ενεργή διαδικασία μάθησης. Ένα πρόβλημα που ανακύπτει στην συγκεκριμένη προσέγγιση είναι ότι η διαδικασία είναι σχετικά αργή και εξαρτάται σε μεγάλο βαθμό από τον χρήστη που διορθώνει πρακτικά το μοντέλο. Όσον αφορά τα εναλλακτικά μοντέλα μηχανικής μάθησης υπάρχουν πολλές διαφορετικές επιλογές τις οποίες μπορούμε να εφαρμόσουμε. Μια από τις πιο ενδιαφέρουσες επιλογές είναι η χρήση ενός γενετικού ανταγωνιστικού δικτύου (GAN) όπως περιγράφεται στο [14]. Οι επιλογές στην ανίχνευση ανωμαλιών σε ένα τέτοιο δίκτυο συνοψίζονται στην χρήση του διαχωριστή (Descriminator) είτε στην χρήση του γεννήτορα (Generator) ή ενός συνδυασμού των δυο. Τέλος, ιδιαίτερα ενδιαφέρουσα είναι η επέκταση του αυτοκωδικοποιητή σε ένα σύστημα ήμι-επιβλεπόμενης μάθησης ώστε να μπορέσουμε να εξαλείψουμε την στοχαστικότητα των αποτελεσμάτων στην περίπτωση των δυσανάλογων κλάσεων στην διαδικασία την συσταδοποίησης.

Βιβλιογραφία

[1] Mohammad Iquebal Akhter and Dr. Mohammad Gulam Ahamad. "Detecting Telecommunication Fraud using Neural Networks through Data Mining". In: 2012.

[2] Babak Alipanahi et al. "Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning". In: *Nature biotechnology* 33 (July 2015). DOI: 10.1038/nbt.3300.

[3] Dzmitry Bahdanau, Kyunghyun Cho, and Yoshua Bengio. *Neural Machine Translation by Jointly Learning to Align and Translate*. cite arxiv:1409.0473Comment: Accepted at ICLR 2015 as oral presentation. 2014. URL: http://arxiv.org/abs/1409.0473.

[4] Laura Beggel, Michael Pfeiffer, and Bernd Bischl. "Robust Anomaly Detection in Images using Adversarial Autoencoders". In: *CoRR* abs/1901.06355 (2019). arXiv: 1901.06355. URL: http://arxiv.org/abs/1901.06355.

[5] Markus M. Breunig et al. "LOF: Identifying Density-based Local Outliers". In: *SIGMOD Rec.* 29.2 (May 2000), pp. 93–104. ISSN: 0163-5808. DOI: 10.1145/335191.335388. URL: http://doi.acm.org/10.1145/335191.335388.

[6] Raghavendra Chalapathy and Sanjay Chawla. "Deep Learning for Anomaly Detection: A Survey". In: *CoRR* abs/1901.03407 (2019). arXiv: 1901.03407. URL: http://arxiv.org/abs/1901.03407.

[7] Varun Chandola, Arindam Banerjee, and Vipin Kumar. "Anomaly Detection: A Survey". In: *ACM Comput. Surv.* 41.3 (July 2009), 15:1–15:58. ISSN: 0360-0300. DOI: 10.1145/1541880.1541882. URL: http://doi.acm.org/10.1145/1541880.1541882.

[8] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep Learning*. http://www.deeplearningbook.org. MIT Press, 2016.

[9] Ian Goodfellow et al. "Generative Adversarial Nets". In: *Advances in Neural Information Processing Systems 27*. Ed. by Z. Ghahramani et al. Curran Associates, Inc., 2014, pp. 2672–2680. URL: http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf.

[10] G. Hinton et al. "Deep Neural Networks for Acoustic Modeling in Speech Recognition: The Shared Views of Four Research Groups". In: *IEEE Signal Processing Magazine* 29.6 (Nov. 2012), pp. 82–97. ISSN: 1053-5888. DOI: 10.1109/MSP.2012.2205597.

[11] D. K. Iakovidis et al. "Detecting and Locating Gastrointestinal Anomalies Using Deep Learning and Iterative Cluster Unification". In: *IEEE Transactions on Medical Imaging* 37.10 (Oct. 2018), pp. 2196–2210. ISSN: 0278-0062. DOI: 10.1109/TMI.2018.2837002.

[12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: (2012). Ed. by F. Pereira et al., pp. 1097–1105. URL: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

[13] Valentin Leveau and Alexis Joly. *Adversarial autoencoders for novelty detection*. Research Report. Inria - Sophia Antipolis, Feb. 2017. URL: https://hal.inria.fr/hal-01636617.

[14] Dan Li et al. *MAD-GAN: Multivariate Anomaly Detection for Time Series Data with Generative Adversarial Networks*. 2019. eprint: arXiv:1901.04997.

[15] Fei Tony Liu, Kai Ming Ting, and Zhi-Hua Zhou. "Isolation Forest". In: Jan. 2009, pp. 413–422. DOI: 10.1109/ICDM.2008.17.

[16] Alireza Makhzani et al. *Adversarial Autoencoders*. 2015. eprint: arXiv:1511.05644.

[17] Pankaj Malhotra et al. "Long Short Term Memory Networks for Anomaly Detection in Time Series". In: Apr. 2015.

[18] Luis Martí et al. "Anomaly Detection Based on Sensor Data in Petroleum Industry Applications". In: *Sensors* 15 (Feb. 2015), pp. 2774–2797. DOI: 10.3390/s150202774.

[19] M. Mutlu Yapici, A. Tekerek, and N. Topaloglu. "Convolutional Neural Network Based Offline Signature Verification Application". In: (Dec. 2018), pp. 30–34. DOI: 10.1109/IBIGDELFT.2018.8625290.

[20] Tiago Pimentel et al. *A Generalized Active Learning Approach for Unsupervised Anomaly Detection*. 2018. eprint: arXiv:1805.09411.

[21] Tiago Pimentel et al. "A Generalized Active Learning Approach for Unsupervised Anomaly Detection". In: (2018). eprint: arXiv:1805.09411.

[22] *Prognostics and Health Management of Engineering Systems*. Springer, 2017.

[23] L. Ren et al. "Remaining Useful Life Prediction for Lithium-Ion Battery: A Deep Learning Approach". In: *IEEE Access* 6 (2018), pp. 50587–50598. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2018.2858856.

[24] SamanehSorournejad et al. "A Survey of Credit Card Fraud Detection Techniques: Data and Technique Oriented Perspective". In: (Nov. 2016).

[25] Nitish Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *Journal of Machine Learning Research* 15 (2014), pp. 1929–1958. URL: http://jmlr.org/papers/v15/srivastava14a.html.

[26] J. Wei et al. "Unsupervised Anomaly Detection for Traffic Surveillance Based on Background Modeling". In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. June 2018, pp. 129–1297. DOI: 10.1109/CVPRW.2018.00025.

[27] D. Wulsin et al. "Semi-Supervised Anomaly Detection for EEG Waveforms Using Deep Belief Nets". In: (Dec. 2010), pp. 436–441. DOI: 10.1109/ICMLA.2010.71.

[28] Yanfang Ye et al. "A Survey on Malware Detection Using Data Mining Techniques". In: *ACM Comput. Surv.* 50.3 (June 2017), 41:1–41:40. ISSN: 0360-0300. DOI: 10.1145/3073559. URL: http://doi.acm.org/10.1145/3073559.

[29] Rui Zhao et al. "Deep Learning and Its Applications to Machine Health Monitoring: A Survey". In: *CoRR* abs/1612.07640 (2016). arXiv: 1612.07640. URL: http://arxiv.org/abs/1612.07640.