



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ, ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ ΕΡΓΑΣΤΗΡΙΟ ΕΥΦΥΩΝ ΣΥΣΤΗΜΑΤΩΝ

Πρόβλεψη της τιμής του κρυπτονομίσματος με αλγόριθμους βαθιάς μηχανικής μάθησης

Μελέτη και υλοποίηση

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΠΟΝΤΙΚΗ Κ. ΜΑΡΙΑΣ



Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π

Συνεπιβλέπουσα: Παρασκευή Τζούβελη
Εργαστηριακό και Διδακτικό Προσωπικό Ε.Μ.Π.

Αθήνα, Μάρτιος 2022



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ, ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ ΕΡΓΑΣΤΗΡΙΟ ΕΥΦΥΩΝ
ΣΥΣΤΗΜΑΤΩΝ

Πρόβλεψη της τιμής του κρυπτονομίσματος με αλγόριθμους βαθιάς μηχανικής μάθησης

Μελέτη και υλοποίηση

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΠΟΝΤΙΚΗ Κ. ΜΑΡΙΑΣ

Επιβλέπων: Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π

Συνεπιβλέπουσα: Παρασκευή Τζούβελη
Εργαστηριακό και Διδακτικό Προσωπικό Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 21 Μαρτίου 2022.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....
Στέφανος Κόλλιας
Καθηγητής Ε.Μ.Π

.....
Ανδρέας Γ. Σταφυλοπάτης
Καθηγητής Ε.Μ.Π

.....
Γεώργιος Στάμου
Καθηγητής Ε.Μ.Π

Αθήνα, Μάρτιος 2022



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ, ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ ΕΡΓΑΣΤΗΡΙΟ ΕΥΦΥΩΝ

ΣΥΣΤΗΜΑΤΩΝ

Copyright © – All rights reserved. Με την επιφύλαξη παντός δικαιώματος.

Μαρία Ποντίκη, 2022.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα.

Το περιεχόμενο αυτής της εργασίας δεν απηχεί απαραίτητα τις απόψεις του Τμήματος, του Επιβλέποντα, ή της επιτροπής που την ενέκρινε.

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ ΕΥΘΥΝΗΣ

Με πλήρη επίγνωση των συνεπειών του νόμου περί πνευματικών δικαιωμάτων, δηλώνω ενυπογράφως ότι είμαι αποκλειστικός συγγραφέας της παρούσας Πτυχιακής Εργασίας, για την ολοκλήρωση της οποίας κάθε βοήθεια είναι πλήρως αναγνωρισμένη και αναφέρεται λεπτομερώς στην εργασία αυτή. Έχω αναφέρει πλήρως και με σαφείς αναφορές, όλες τις πηγές χρήσης δεδομένων, απόψεων, θέσεων και προτάσεων, ιδεών και λεκτικών αναφορών, είτε κατά κυριολεξία είτε βάσει επιστημονικής παράφρασης. Αναλαμβάνω την προσωπική και ατομική ευθύνη ότι σε περίπτωση αποτυχίας στην υλοποίηση των ανωτέρω δηλωθέντων στοιχείων, είμαι υπόλογος έναντι λογοκλοπής, γεγονός που σημαίνει αποτυχία στην Πτυχιακή μου Εργασία και κατά συνέπεια αποτυχία απόκτησης του Τίτλου Σπουδών, πέραν των λοιπών συνεπειών του νόμου περί πνευματικών δικαιωμάτων. Δηλώνω, συνεπώς, ότι αυτή η Πτυχιακή Εργασία προετοιμάστηκε και ολοκληρώθηκε από εμένα προσωπικά και αποκλειστικά και ότι, αναλαμβάνω πλήρως όλες τις συνέπειες του νόμου στην περίπτωση κατά την οποία αποδειχθεί, διαχρονικά, ότι η εργασία αυτή ή τμήμα της δεν μου ανήκει διότι είναι προϊόν λογοκλοπής άλλης πνευματικής ιδιοκτησίας.

(Υπογραφή)

.....
Μαρία Ποντίκη

Μάρτιος 2022

Περίληψη

Το Bitcoin είναι ένα αποκεντρωμένο κρυπτονόμισμα, το οποίο είναι ένας τύπος ψηφιακού περιουσιακού στοιχείου που παρέχει τη βάση για peer-to-peer χρηματοοικονομικές συναλλαγές που βασίζονται στην τεχνολογία blockchain. Ένα από τα κύρια προβλήματα με τα αποκεντρωμένα κρυπτονομίσματα είναι η αστάθεια της τιμής τους, η οποία υποδηλώνει την ανάγκη μελέτης του υποκείμενου μοντέλου. Επιπλέον, οι τιμές Bitcoin παρουσιάζουν μη στάσιμη συμπεριφορά, όπου η στατιστική κατανομή των δεδομένων αλλάζει με την πάροδο του χρόνου.

Σκοπός της παρούσας διπλωματικής εργασίας είναι η χρήση μοντέλων παλινδρόμησης που βασίζονται στη μηχανική μάθηση για την πρόβλεψη των κινήσεων των τιμών και των τιμών Bitcoin βραχυπρόθεσμα και μεσοπρόθεσμα. Στην παρούσα εργασία η πρόβλεψη της τιμής του κρυπτονομίσματος πραγματοποιείται με μοντέλα που βασίζονται στη μηχανική μάθηση για μία, επτά, τριάντα και ενενήντα ημέρες. Στόχος της διπλωματικής εργασίας είναι η μελέτη διαφορετικών μοντέλων βαθιάς μηχανικής μάθησης για την επίλυση ενός προβλήματος παλινδρόμησης. Μελετούνται αναδρομικά και συνελκτικά μοντέλα αφού πρώτα έχει γίνει ανάλυση και επεξεργασία ακολουθιακών δεδομένων.

Λέξεις Κλειδιά

Bitcoin, Blockchain, Κρυπτονόμισμα, Πρόβλεψη Χρονοσειρών, Μηχανική Μάθηση, Νευρωνικά Δίκτυα, Επαναληπτικά Νευρωνικά Δίκτυα

Abstract

Bitcoin is a decentralized cryptocurrency, which is a type of digital asset that provides the basis for peer-to-peer financial transactions based on blockchain technology. One of the main problems with decentralized cryptocurrencies is their price volatility, which indicates the need to study the underlying model. In addition, Bitcoin values show non-stationary behavior, where the statistical distribution of data changes over time.

The purpose of this dissertation is to use regression models based on machine learning to predict price movements and prices Bitcoin in the short and medium term. In the present thesis the cryptocurrency price prediction includes models based on machine learning for one, seven, thirty and ninety of days. The aim of the dissertation is to study different models of deep learning to solve a regression problem. RNN and CNN models are studied after first analyzing and processing data sequences.

Keywords

Bitcoin, Blockchain, Cryptocurrency, Time Series Forecasting, Machine Learning, Neural Networks, Recurrent Neural Networks

στους γονείς μου

Ευχαριστίες

Η ολοκλήρωση της παρούσας Διπλωματικής Εργασίας σηματοδοτεί το τέλος των προπτυχιακών μου σπουδών. Θα ήθελα καταρχήν να ευχαριστήσω ιδιαίτερα τον καθηγητή μου κ. Στέφανο Κόλλια, που μου έδωσε την ευκαιρία να εκπονήσω την διπλωματική μου στο Εργαστήριο Ευφών Συστημάτων αλλά και τους κ. Ανδρέα - Γεώργιο Σταφυλοπάτη και κ.Γεώργιο Στάμου , που μου έκαναν την τιμή να βρίσκονται στην τριμελή εξεταστική επιτροπή. Θα ήθελα επίσης, να ευχαριστήσω την συνεπιβλέπουσα κ. Παρασκευή Τζούβελη για την καθοδήγησή της και την εξαιρετική συνεργασία που είχαμε όλο το αυτό το διάστημα. Οι ιδέες της και η αμέριστη συμπαράσταση της στάθηκαν καθοριστικές. Ταυτόχρονα, θα ήθελα να ευχαριστήσω τους φίλους μου, τον αδερφό μου Βαγγέλη και τον φίλο μου Δημήτρη για την αμέριστη βοήθεια και τη συνεχή εμπύχωση όλα αυτά τα χρόνια. Χωρίς εκείνους το μακρινό ταξίδι των σπουδών μου θα φάνταζε αδύνατο. Πάνω από όλα όμως οφείλω να ευχαριστήσω την μητέρα μου, Γαρυφαλλιά, για τις αμέτρητες θυσίες που έκανε όλα αυτά τα χρόνια, για να φτάσω σε αυτό το σημείο σήμερα, μέσα από την αγάπη της, το ενδιαφέρον της, τον χρόνο της και την υποστήριξή της. Γι' αυτό, θα ήθελα να αφιερώσω σε αυτήν, αλλά και στον εκλιπόντα πατέρα μου Κωνσταντίνο την παρούσα εργασία.

Αθήνα, Μάρτιος 2022

Μαρία Ποντίκη

Περιεχόμενα

Περίληψη	1
Abstract	3
Ευχαριστίες	7
1 Εισαγωγή	15
1.1 Πληροφορίες για το bitcoin και το blockchain	15
1.2 Αντικείμενο της διπλωματικής	17
1.3 Οργάνωση του τόμου	18
I Θεωρητικό Μέρος	19
2 Θεωρητικό υπόβαθρο	21
2.1 Bitcoin	21
2.1.1 Ιστορική Αναδρομή	21
2.1.2 Εξόρυξη του Bitcoin	24
2.1.3 Τεχνική Ανάλυση	25
2.1.4 Ανάλυση χρονοσειρών σε οικονομικό πλαίσιο	26
2.2 Τεχνολογία Blockchain	26
2.2.1 Η εμφάνιση της αποκεντρωμένης οικονομίας	27
2.2.2 Οι υποσχέσεις της αποκεντρωμένης οικονομίας	27
2.2.3 Κύρια επιχειρηματικά μοντέλα στην αποκεντρωμένη οικονομία	28
3 Νευρωνικά Δίκτυα	33
3.1 Εισαγωγή στη Μηχανική Μάθηση	33
3.1.1 Random Forest Regressor	33
3.1.2 Είδη μηχανικής μάθησης	34
3.1.3 Είδη προβλημάτων μάθησης	35
3.2 Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks-ANN)	36
3.3 Επαναληπτικά Νευρωνικά Δίκτυα	38
3.3.1 Απλά Επαναληπτικά Νευρωνικά Δίκτυα	38
3.3.2 Back Propagation Through Time-(BPTT)	41
3.3.3 Φραγμένα Επαναληπτικά Δίκτυα -GRU	44
3.3.4 Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης - LSTM	47
3.3.5 Αμφίδρομα Επαναληπτικά Νευρωνικά Δίκτυα (Bidirectional RNNs)	50

3.4	Υβριδικά Νευρωνικά Δίκτυα (Hybrid Neural Networks)	52
3.4.1	Εισαγωγή στα Συνελκτικά Νευρωνικά Δίκτυα(Convolutional Neural Networks)	52
3.4.2	1D Convolutional Neural Networks	57
3.4.3	Ζεύξη Επαναληπτικών και Συνελκτικών Νευρωνικών Δικτύων (CNN-RNNs)	61
3.5	Χρονικά Συνελκτικά Δίκτυα (Temporal Convolutional Network)	62
II	Πρακτικό Μέρος	71
4	Ανάλυση τεχνικών και αρχιτεκτονικών	73
4.1	Περιγραφή Συνόλου Δεδομένων Προβλήματος	73
4.1.1	Γνωριμία με το dataset	73
4.1.2	Επιλογή Χαρακτηριστικών (Feature Selection)	74
4.1.3	Περαιτέρω επεξεργασία στα καινούργια datasets	76
4.2	Αρχιτεκτονικές Μοντέλων Χρονοσειρών	76
4.2.1	Επαναληπτικά Μοντέλα	76
4.2.2	Συνελκτικά Μοντέλα	78
4.2.3	Υβριδικά-Συνελκτικά Μοντέλα	79
5	Πειραματικά αποτελέσματα	81
5.1	Μετρικές Αξιολόγησης	81
5.1.1	Μετρικές Τετραγωνικού Σφάλματος	81
5.1.2	Μετρικές Απόλυτου Σφάλματος	82
5.1.3	Μετρικές Ακρίβειας	82
5.2	Αποτελέσματα και αξιολόγηση Μοντέλων	82
III	Επίλογος	93
6	Επίλογος	95
6.1	Συμπεράσματα	95
6.2	Μελλοντικές Επεκτάσεις	96
	Βιβλιογραφία	102

Κατάλογος Σχημάτων

2.1	Bitcoin (BTC) price per day from October 2013 to February 7,2022 (in U.S. dollars)	23
2.2	Μεθοδολογία Συναλλαγών για συναλλαγές bitcoin[1]	25
2.3	Η άνοδος των χρηματοοικονομικών υπηρεσιών που βασίζονται σε blockchain	29
2.4	The rise of ICOs and IEOs	31
3.1	Biological Neural Network	37
3.2	Artificial Neural Network	38
3.3	Δομή Perceptron	38
3.4	Συναρτήσεις Ενεργοποίησης	39
3.5	Απλό Αναδρομικό Νευρωνικό Δίκτυο σε τρία χρονικά βήματα[2]	41
3.6	Υπολογιστικό γράφημα που αναδεικνύει τις εξαρτήσεις για ένα μοντέλο RNN σε τρία χρονικά βήματα.Τα μη χρωματιστά πλαίσια αναπαριστούν τις μεταβλητές ενώ τα χρωματισμένα πλαίσια και οι κύκλοι αντιπροσωπεύουν τελεστές πράξεων	42
3.7	Gated Recurrent Unit	45
3.8	Update gate	46
3.9	Reset gate	46
3.10	Current memory content	47
3.11	Hidden State	48
3.12	Αναπαράσταση LSTM με τις αντίστοιχες πύλες	49
3.13	Σχηματική Αναπαράσταση forget gate	50
3.14	Σχηματική Αναπαράσταση input/update gate	51
3.15	Σχηματική Αναπαράσταση Cell State	52
3.16	Σχηματική Αναπαράσταση output gate	53
3.17	Σχηματική Αναπαράσταση Αμφίδρομου RNN	53
3.18	Σχηματική Αναπαράσταση RGB εικόνας και οι διαστάσεις της	54
3.19	Σχηματική Αναπαράσταση gray εικόνας και οι διαστάσεις της	55
3.20	Νευρωνικό δίκτυο με πολλά συνελκτικά στρώματα	56
3.21	Τεχνική stride με βήμα 2	57
3.22	Τεχνική zero-padding	57
3.23	Συνάρτηση Ενεργοποίησης ReLU	58
3.24	Σχηματική Αναπαράσταση Max Pooling Layer	58
3.25	Σχηματική Αναπαράσταση Fully Connected Layer	59
3.26	Σχηματική Αναπαράσταση Flattening	59

3.27	Υπολογισμός συνέλιξης για 1D-CNN με input channels = output channels = 1	60
3.28	Υπολογισμός συνέλιξης για δύο διαδοχικά στοιχεία εξόδου	60
3.29	Υπολογισμός συνέλιξης για 1D-CNN με input channels = 1	61
3.30	Η αρχιτεκτονική του δικτύου CNN-RNN	62
3.31	Zero padding with kernel size=3 and input length=4	63
3.32	Dilation σε στιβαγμένα επίπεδα με kernel size=3	64
3.33	2-σταλμένο στρώμα με input length=4 kernel size=3	65
3.34	Σχηματική αναπαράσταση δικτύου με διασταλμένα συνελκτικά στρώματα με πλήρη κάλυψη	65
3.35	Σχηματική αναπαράσταση δικτύου με διασταλμένα συνελκτικά στρώματα με οπές	66
3.36	Σχηματική αναπαράσταση αιτιατού δικτύου με kernel size=3 dilation=2 μετά τη βελτίωση	68
3.37	Σχηματική αναπαράσταση residual block έπειτα από Activation, Normalization, Regularization	69
3.38	Σχηματική αναπαράσταση ολοκληρωμένου δικτύου TCN	69
4.1	Ακατέργαστα χαρακτηριστικά από τα οποία δημιουργούνται οι τεχνικοί δείκτες	74
5.1	Τιμή του BTC από το 2013 έως το 2020	83
5.2	Σύγκριση πραγματικών τιμών και τιμών πρόβλεψης για το Interval1	87
5.3	Σύγκριση πραγματικών τιμών και τιμών πρόβλεψης για το Interval2	88
5.4	Σύγκριση πραγματικών τιμών και τιμών πρόβλεψης για το Interval3	88
5.5	Σύγκριση πραγματικών τιμών και τιμών πρόβλεψης για το Interval3 με χρονικό ορίζοντα 7 ημερών	89
5.6	Σύγκριση πραγματικών τιμών και τιμών πρόβλεψης για το Interval3 με χρονικό ορίζοντα 30 ημερών	89
5.7	Σύγκριση πραγματικών τιμών και τιμών πρόβλεψης για το Interval3 με χρονικό ορίζοντα 90 ημερών	90
5.8	Μετρική MAE για EOD predictions στα διαφορετικά Intervals	90
5.9	Μετρική MAPE για EOD(End Of Day) predictions στα διαφορετικά Intervals	91
5.10	Μετρική MAE των μοντέλων που προβλέπουν την τιμή BTC της n-οστής ημέρας στο Interval 3	91
5.11	Μετρική MAPE των μοντέλων που προβλέπουν την τιμή BTC της n-οστής ημέρας στο Interval 3	92

Κατάλογος Πινάκων

4.1	Υπερπαράμετροι μοντέλου RNN	77
4.2	Υπερπαράμετροι αμφίδρομου μοντέλου GRU	77
4.3	Υπερπαράμετροι αμφίδρομου μοντέλου LSTM	78
4.4	Υπερπαράμετροι απλών μοντέλων LSTM και GRU	78
4.5	Υπερπαράμετροι μοντέλων deep GRU και deep LSTM	78
4.6	Υπερπαράμετροι μοντέλου TCN	79
4.7	Υπερπαράμετροι CNN-RNN	80
4.8	Υπερπαράμετροι CNN-LSTM	80
4.9	Υπερπαράμετροι CNN-GRU	80
5.1	Μετρικές απόδοσης για το Interval 1	84
5.2	Μετρικές απόδοσης για το Interval 2	85
5.3	Μετρικές απόδοσης για το Interval 3	85
5.4	Μετρικές απόδοσης για το Interval 3 με 7-ήμερη πρόβλεψη	85
5.5	Μετρικές απόδοσης για το Interval 3 30-ημέρων πρόβλεψη	86
5.6	Μετρικές απόδοσης για το Interval 3 90-ημερών πρόβλεψη	86

Εισαγωγή

1.1 Πληροφορίες για το bitcoin και το blockchain

Το κρυπτονόμισμα είναι ένα ψηφιακό νόμισμα που χρησιμοποιεί κρυπτογραφία για να εξασφαλίσει τις σχετικές διαδικασίες σε συναλλαγές και παραγωγή μονάδων. Το Bitcoin, συγκεκριμένα, ήταν το πρώτο αποκεντρωμένο στον κόσμο κρυπτονόμισμα. Το bitcoin εφευρέθηκε το 2008 από μία άγνωστη ομάδα χρησιμοποιώντας το όνομα Satoshi Nakamoto και ξεκίνησε το 2009 όταν ο πηγαίος κώδικάς του δόθηκε στο διαδίκτυο σαν ελεύθερο λογισμικό. Τα bitcoins παράγονται ως αποτέλεσμα μιας υπολογιστικής διεργασίας που μπορεί να πραγματοποιηθεί από οποιοδήποτε αρκούντως ισχυρό ηλεκτρονικό υπολογιστή. Η διεργασία αποτελείται από μια σειρά πολύπλοκων μαθηματικών υπολογισμών, και είναι γνωστή με τον όρο εξόρυξη (mining). Τα bitcoins μπορούν, μέσω ειδικών συναλλαγματικών πρακτορειών να εξαργυρωθούν με άλλα νομίσματα, καθώς και να χρησιμοποιηθούν για να αγοράσουν προϊόντα και υπηρεσίες. Έρευνα που έγινε από το πανεπιστήμιο Κέμπριτζ εκτιμά ότι το 2017, υπήρχαν από 2,9 έως 5,8 εκατομμύρια μοναδικοί χρήστες που έχουν στην κατοχή τους ένα "πορτοφόλι κρυπτονομισμάτων", δηλαδή ένα ψηφιακό λογαριασμό που περιέχει όλες τις πληροφορίες που πιστοποιούν την κατοχή των bitcoins από κάθε χρήστη. Η έρευνα κατέδειξε ότι το bitcoin ήταν η δημοφιλέστερη επιλογή κρυπτονομίσματος από την πλειοψηφία των χρηστών. [3] [4]

Στα κεντρικά νομίσματα, η κυβέρνηση ή άλλες εταιρικές οντότητες έχουν τον έλεγχο της προσφοράς νομίσματος και την εκτύπωση νέων χρημάτων. Αντίθετα, το Bitcoin είναι ένα αποκεντρωμένο νόμισμα, που σημαίνει ότι καμία οντότητα δεν είναι υπεύθυνη για τη δημιουργία νέων μονάδων ή bitcoin. Άρα οι χρήστες του δικτύου εν δυνάμει μπορούν να επιβεβαιώσουν μια συναλλαγή π.χ. μεταφοράς bitcoin από τον έναν χρήστη στον άλλον χωρίς την ανάγκη κάποιου είδους οργανισμού, όπως μια τράπεζα.

Τα τελευταία χρόνια, το Bitcoin έχει γίνει αντικείμενο ενδιαφέροντος από τα μέσα ενημέρωσης και τους επενδυτές. Αυτή η αύξηση της προσοχής μπορεί να αποδοθεί στα καινοτόμα χαρακτηριστικά, την απλότητα, τη διαφάνεια και την αυξανόμενη δημοτικότητα του. Παράλληλα, θέτει μεγάλες προκλήσεις και ευκαιρίες για τους υπεύθυνους χάραξης πολιτικής, τους οικονομολόγους, τους επιχειρηματίες και τους καταναλωτές. Στατιστικές έρευνες δείχνουν ότι το Bitcoin είναι μια κερδοσκοπική επένδυση και όχι ένα εναλλακτικό νόμισμα ή μέσο ανταλλαγής. [5]

Το Bitcoin κατάφερε να διατηρήσει τη θέση του ως σημείο αναφοράς στην αγορά κρυ-

πυρονομισμάτων παρά τον μεγάλο ανταγωνισμό από τον αυξανόμενο αριθμό νέων κρυπτονομισμάτων. Αυτό συνέβη όχι μόνο λόγω της πρώτης παρουσίας αυτού του κρυπτονομίσματος αλλά και λόγω της υψηλής εκτίμησης σε αξία, δηλαδή συνδέεται πλέον με το ψηφιακό νόμισμα. Ωστόσο, το Bitcoin είναι επίσης γνωστό ότι πάσχει από υψηλά επίπεδα ασταθούς τιμής. Αυτές οι διακυμάνσεις των τιμών έχουν προβληματίσει και δυσκολέψει ανθρώπους του κλάδου, όπως επενδυτές, οικονομολόγους κλπ .

Ενώ το Bitcoin χρησιμοποιείται από απλούς ανθρώπους και λόγω της έλλειψης συνάφειας του με άλλα περιουσιακά στοιχεία, έχει γίνει μια ελκυστική επιλογή για τους επενδυτές. Επομένως, η δυνατότητα πρόβλεψης τιμών θα ήταν μεγάλη βοήθεια. Λαμβάνοντας υπόψη τη σημασία του θέματος, σκοπός της παρούσας εργασίας είναι η πρόβλεψη τιμής bitcoin με αλγορίθμους Deep learning.

Τεχνολογία Blockchain

Το Blockchain είναι η τεχνολογία που επιτρέπει την ύπαρξη και τη χρήση κρυπτονομισμάτων. Η τεχνολογία Blockchain εφευρέθηκε για να υποστηρίξει το Bitcoin, το πιο γνωστό κρυπτονόμισμα. Πρόκειται για έναν ειδικό τύπο βάσης δεδομένων. Οι συναλλαγές δεν διέπονται από ένα μεμονωμένο μέρος, αλλά ολόκληρο το ιστορικό συναλλαγών καταγράφεται σε ένα αποκεντρωμένο, κατανεμημένο σύστημα. Για να διασφαλιστεί ότι το ιστορικό συναλλαγών του δικτύου δεν χειραγωγείται από κανέναν, η κοινότητα πίσω από το δίκτυο πρέπει να συμφωνήσει σε μια κοινή «πραγματικότητα».

Το Blockchain είναι ένας τύπος διαμοιραζόμενης βάσης δεδομένων που διαφέρει από μια τυπική βάση δεδομένων στον τρόπο με τον οποίο αποθηκεύει πληροφορίες. Οι αλυσίδες μπλοκ αποθηκεύουν δεδομένα σε μπλοκ που στη συνέχεια συνδέονται μεταξύ τους μέσω κρυπτογραφίας. Αυτά τα δεδομένα έρχονται σε μπλοκ, οπότε δημιουργούνται μπλοκ ψηφιακών δεδομένων. Αυτά τα μπλοκ συνδέονται μεταξύ τους και αυτό κάνει τα δεδομένα τους αμετάβλητα. Όταν ένα μπλοκ δεδομένων συνδέεται με τα άλλα μπλοκ, τα δεδομένα του δεν μπορούν ποτέ να αλλάξουν ξανά. Θα είναι δημόσια διαθέσιμο σε όποιον θέλει να το δει ξανά, ακριβώς με τον τρόπο που προστέθηκε κάποτε στο blockchain. Αυτό είναι αρκετά επαναστατικό, γιατί μας επιτρέπει να κρατάμε αρχεία σχεδόν για οτιδήποτε μπορούμε να σκεφτούμε (για να αναφέρουμε μερικά: δικαιώματα ιδιοκτησίας, ταυτότητες, χρηματικά υπόλοιπα, ιατρικά αρχεία), χωρίς να κινδυνεύουμε από κάποιον να παραποιήσει αυτά τα αρχεία. Εάν αγοράσω ένα σπίτι αυτήν τη στιγμή και προσθέσω μια φωτογραφία των δικαιωμάτων ιδιοκτησίας σε ένα blockchain, θα μπορώ πάντα και για πάντα να αποδεικνύω ότι είχα αυτά τα δικαιώματα σε εκείνο το σημείο. Κανείς δεν μπορεί να αλλάξει αυτή την πληροφορία όταν είναι σε blockchain. Έτσι, το Blockchain είναι ένας τρόπος για να αποθηκεύουμε δεδομένα και να τα κάνουμε αμετάβλητα. Αυτό είναι αρκετά πρωτοπόρο, πως δουλεύει όμως; [6]

Η τεχνολογία Blockchain επιτυγχάνει αποκεντρωμένη ασφάλεια και εμπιστοσύνη με διάφορους τρόπους. Αρχικά, τα νέα μπλοκ αποθηκεύονται πάντα γραμμικά και χρονολογικά. Δηλαδή, προστίθενται πάντα στο «τέλος» του blockchain. Αφού προστεθεί ένα μπλοκ στο τέλος της αλυσίδας μπλοκ, είναι εξαιρετικά δύσκολο να αλλάξουν τα περιεχόμενα του μπλοκ εκτός εάν η πλειοψηφία του δικτύου έχει καταλήξει σε συναίνεση για να το κάνει. Αυτό συμβαίνει

επειδή κάθε μπλοκ περιέχει το δικό του κατακερματισμό, μαζί με τον κατακερματισμό του μπλοκ πριν από αυτό, καθώς και τη χρονική σήμανση που αναφέρθηκε προηγουμένως. Οι κωδικοί κατακερματισμού δημιουργούνται από μια μαθηματική συνάρτηση που μετατρέπει τις ψηφιακές πληροφορίες σε μια σειρά αριθμών και γραμμμάτων. Εάν αυτές οι πληροφορίες υποβληθούν σε επεξεργασία με οποιονδήποτε τρόπο, τότε αλλάζει και ο κωδικός κατακερματισμού.

Για παράδειγμα έστω ότι ένας χάκερ, ο οποίος εκτελεί επίσης έναν κόμβο σε ένα δίκτυο blockchain, θέλει να αλλάξει ένα blockchain και να κλέψει κρυπτονομίσματα από όλους τους άλλους. Εάν επρόκειτο να τροποποιήσει το δικό του μεμονωμένο αντίγραφο, δεν θα ευθυγραμμίζεται πλέον με το αντίγραφο όλων των άλλων. Όταν όλοι οι άλλοι διασταυρώνουν τα αντίγραφά τους ο ένας εναντίον του άλλου, θα έβλεπαν αυτό το ένα αντίγραφο να ξεχωρίζει και η εκδοχή αυτού του χάκερ της αλυσίδας θα απορριφθεί ως παράνομη.

Για να πετύχει ένα τέτοιο hack θα απαιτούσε ο χάκερ να ελέγχει και να τροποποιεί ταυτόχρονα το 51 τοις εκατό ή περισσότερα από τα αντίγραφα του blockchain, έτσι ώστε το νέο τους αντίγραφο να γίνει το πλειοψηφικό αντίγραφο και, επομένως, η συμφωνημένη αλυσίδα. Μια τέτοια επίθεση θα απαιτούσε επίσης ένα τεράστιο ποσό χρημάτων και πόρων, καθώς θα χρειαζόταν να επαναληφθούν όλα τα μπλοκ επειδή θα είχαν πλέον διαφορετικές χρονικές σημάνσεις και κωδικούς κατακερματισμού.[7]

Λόγω του μεγέθους πολλών δικτύων κρυπτονομισμάτων και του πόσο γρήγορα αναπτύσσονται, το κόστος για να επιτευχθεί μια τέτοια κακόβουλη ενέργεια θα ήταν πιθανώς ανυπέβλητη. Αυτό θα ήταν όχι μόνο εξαιρετικά ακριβό αλλά και πιθανότατα άκαρπο. Κάνοντας κάτι τέτοιο δεν θα περνούσε απαρατήρητο, καθώς τα μέλη του δικτύου θα έβλεπαν δραστικές αλλαγές στο blockchain. Στη συνέχεια, τα μέλη του δικτύου θα μεταθούν σκληρά σε μια νέα έκδοση της αλυσίδας που δεν έχει επηρεαστεί. Αυτό θα προκαλούσε την κατακόρυφη πτώση της αξίας της έκδοσης του διακριτικού στην επίθεση, καθιστώντας την επίθεση τελικά άσκοπη, καθώς ο κακόβουλος χρήστης έχει τον έλεγχο ενός άχρηστου περιουσιακού στοιχείου.

1.2 Αντικείμενο της διπλωματικής

Στην παρούσα διπλωματική εργασία εφαρμόζονται αλγόριθμοι βαθιάς μηχανικής μάθησης με σκοπό την πρόβλεψη της τιμής του κρυπτονομίσματος Bitcoin. Λόγω της αστάθειας της τιμής του αλλά και ταυτόχρονα λόγω της διασημότητας του και της ολοένα αυξανόμενης χρήσης και επενδυτικής δραστηριότητας του, καθίσταται επιτακτική η πρόβλεψη τιμής κλεισίματος της επόμενης μέρας αλλά και με χρονικό ορίζοντα κάποιων ημερών ή και μηνών. Όσο μεγαλύτερος ο χρονικός ορίζοντας της πρόβλεψης τόσο λιγότερο προσεγγιστική είναι η πρόβλεψη στην πραγματική τιμή.

Έπειτα από ενδελεχή ανάλυση των δεδομένων και προεπεξεργασία, διαχειριστήκαμε τα δεδομένα ως μια χρονοσειρά με χαρακτηριστικά (features) τους τεχνικούς δείκτες. Αυτοί οι τεχνικοί δείκτες που έχουν προκύψει από τεχνική ανάλυση χρησιμοποιούν την έννοια των προτύπων τιμών από το παρελθόν και προβλέπουν τις μελλοντικές κινήσεις των τιμών.

Η ανάλυση και μερική οπτικοποίηση χρονοσειρών είναι επίσης ένα θέμα που μελετήθηκε στην παρούσα διπλωματική. Η ανάλυση που έγινε βασίστηκε στα χαρακτηριστικά του

συνόλου δεδομένων μας και στόχευε στη βαθύτερη κατανόηση του συνόλου καθώς και στην ευκολότερη μοντελοποίηση του.

1.3 Οργάνωση του τόμου

Αφού έγινε μια σύντομη εισαγωγή με σκοπό να εξηγηθούν οι βασικοί όροι και έννοιες καθώς και το κίνητρο που οδήγησε σε αυτή τη διπλωματική, ακολουθεί μια σύντομη περιγραφή της εργασίας και των Κεφαλαίων που έπονται. Αρχικά, στο Κεφάλαιο 2 αναλύονται όροι και έννοιες σχετικές με τα κρυπτονομίσματα και την τεχνολογία πίσω από αυτά. Στο Κεφάλαιο 3, περιγράφονται τα νευρωνικά δίκτυα και συγκεκριμένα τα αναδρομικά νευρωνικά δίκτυα, τα οποία και επιλέχθηκαν. Επίσης σε αυτή την ενότητα παρουσιάζονται και αιτιολογούνται τα μοντέλα μηχανικής μάθησης που θα χρησιμοποιηθούν. Στη συνέχεια, στο Κεφάλαιο 4 παρουσιάζεται αναλυτικά η διαδικασία συλλογής των δεδομένων. Ακολουθεί το Κεφάλαιο 5, όπου παρουσιάζονται και εξηγούνται τα πειράματα και τα αποτελέσματα της εργασίας. Σε αυτό το κεφάλαιο γίνεται παρουσίαση και σύγκριση των αποτελεσμάτων για όλα τα μοντέλα που εξετάστηκαν. Τέλος, η εργασία ολοκληρώνεται στο Κεφάλαιο 6, όπου πραγματοποιείται η αποτίμησή της ενώ παρουσιάζονται και ιδέες για μελλοντικές επεκτάσεις και έρευνα.

Μέρος I

Θεωρητικό Μέρος

Κεφάλαιο 2

Θεωρητικό υπόβαθρο

Στο Κεφάλαιο πραγματοποιείται ανάλυση του θεωρητικού υπόβαθρου της διπλωματικής εργασίας και των τεχνικών που θα χρησιμοποιηθούν. Στην πρώτη ενότητα θα γίνει αναφορά στο Bitcoin, τα χαρακτηριστικά του και την τεχνολογία που χρησιμοποιείται πίσω απ' αυτό. Στη δεύτερη ενότητα θα αναλυθεί θεωρητικά η λειτουργία και η εφαρμογή των νευρωνικών δικτύων ξεκινώντας με την περιγραφή του απλού νευρωνικού δικτύου, και στην πορεία θα μελετηθούν τα αναδρομικά νευρωνικά δίκτυα με επεκτάσεις σε CNN και TCN δίκτυα.

2.1 Bitcoin

2.1.1 Ιστορική Αναδρομή

- **Η ιστορία του Bitcoin**

Το Bitcoin δημιουργήθηκε από την αναταραχή της Μεγάλης Ύφεσης του 2008 καθώς η δυσπιστία προς τις τράπεζες και το ρόλο τους στο χρηματοπιστωτικό σύστημα αυξήθηκε. Ένα άτομο ή μια ομάδα ανθρώπων με το όνομα Satoshi Nakamoto εξέδωσε μια λευκή βίβλο για να αντιμετωπίσει τον κεντρικό έλεγχο των χρημάτων και την εμπιστοσύνη που απαιτείται για τη διαχείριση των μετρητών των πολιτών.

Στο παραδοσιακό χρηματοπιστωτικό σύστημα, οι συναλλαγές μπορούν να αντιστραφούν ή να υποστούν παρέμβαση από τρίτους και το κόστος συναλλαγής μπορεί να αθροιστεί. Το Bitcoin παρουσιάστηκε ως ένας τρόπος συναλλαγών χωρίς την ανάμειξη τρίτων. Αντίθετα, το σύστημα Bitcoin χρησιμοποιεί κρυπτογραφική απόδειξη για να διατηρήσει την ακεραιότητα του δικτύου αντί να βασίζεται σε τράπεζες και άλλα ιδρύματα.

Στις 3 Ιανουαρίου 2009, το blockchain ξεκίνησε τη λειτουργία του όταν εξορύχθηκε το πρώτο μπλοκ, που ονομάζεται genesis block. Η πρώτη δοκιμαστική συναλλαγή πραγματοποιήθηκε περίπου μία εβδομάδα αργότερα. Τους πρώτους μήνες της ύπαρξής του, ήταν δυνατή η απόκτησή του μόνο από εξορύκτες (miners) που επικυρώνουν το blockchain Bitcoin.

Θα χρειαζόταν περισσότερο από ένα χρόνο για να πραγματοποιηθεί η πρώτη οικονομική συναλλαγή, όταν ένας άνδρας από τη Φλόριντα διαπραγματεύτηκε να παραδώσει

δύο πίτσες, αξίας 25 δολαρίων, έναντι 10.000 BTC στις 22 Μαΐου 2010. Αυτή η συναλλαγή ουσιαστικά καθιέρωσε την αρχική πραγματική -Παγκόσμια τιμή ή αξία του Bitcoin στα 4 Bitcoins ανά λεπτό(cent), λέει ο Grabowski.[8] Τις πρώτες μέρες, οι πρώτες συναλλαγές με Bitcoin διαπραγματεύονταν σε διαδικτυακά φόρουμ με ανθρώπους που αντάλλασαν για αγαθά και υπηρεσίες με αντάλλαγμα το Bitcoin καθώς η αξία του Bitcoin ήταν αρχικά αυθαίρετη.

Στη συνέχεια, το 2011, οι εξορύκτες- ιδιοκτήτες εξελιγμένου υπολογιστικού εξοπλισμού που ανταγωνίζονται για να είναι οι πρώτοι που θα επαληθεύσουν τις συναλλαγές Bitcoin και θα κερδίσουν ανταμοιβές που πληρώνονται σε Bitcoin[9]- άρχισαν να δημιουργούν άλλα δίκτυα, όπως το Ethereum και το Litecoin[10], και άρχισαν να βελτιώνουν τον κώδικα πίσω από το blockchain του Bitcoin, προσαρμόζοντάς το για διαφορετικές χρήσεις. Αυτή η ευρύτερη βάση εφαρμογών έφερε περισσότερα άτομα, γεγονός που συνέβαλε εν μέρει στην αύξηση της αξίας του Bitcoin. Μόλις το Bitcoin έγινε διαθέσιμο στα χρηματιστήρια το 2010, έγινε ευκολότερη η αγορά, η πώληση, το εμπόριο και η αποθήκευση. Χάρη σε αυτές τις ανταλλαγές, το Bitcoin έχει μια αξιόπιστη τιμή έναντι του δολαρίου των ΗΠΑ.

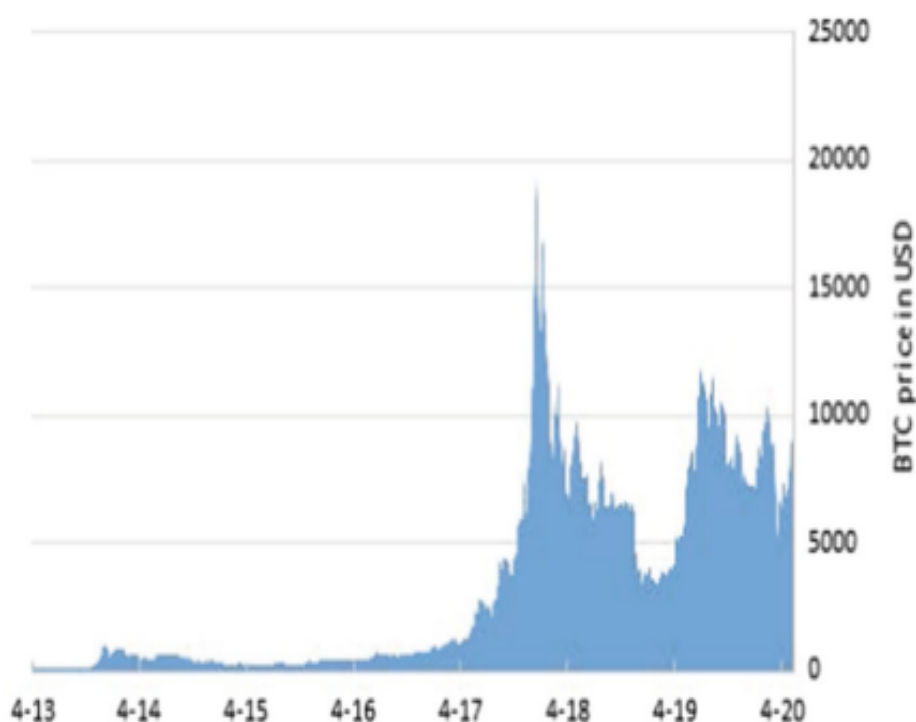
Το Φεβρουάριο του 2011, η τιμή του Bitcoin πέρασε το όριο του 1 δολαρίου. Τα πρώτα χρόνια του καθώς αναπτυσσόταν, η τιμή του ήταν κάτω από 2 δολάρια. Τον Ιούνιο του 2011, χτύπησε την πρώτη του φούσκα, εκτοξεύοντας στα 31 δολάρια περίπου πριν βυθιστεί ξανά σε μονοψήφιο εύρος. Σχεδόν δύο χρόνια αργότερα, τον Απρίλιο του 2013, το Bitcoin έφτασε τα 200 δολάρια. Μέχρι τα τέλη Νοεμβρίου του ίδιου έτους, άξιζε πάνω από 1.000 δολάρια. Στη συνέχεια αυξήθηκε στα 10.000 δολάρια τον Νοέμβριο του 2017 και μέχρι τη μέγιστη τιμή του κοντά στα 68.990 δολάρια τον Νοέμβριο του 2021. Αυτό δεν σημαίνει ότι ήταν μια ομαλή διαδρομή. Λόγω αυτών των εξελίξεων, η αγορά του Bitcoin έχει γίνει αρκετά ώριμη. Υπάρχουν έξυπνες και αποτελεσματικές ανταλλαγές και οι βασικοί θεσμικοί παίκτες υιοθετούν τα απαραίτητα μέτρα για τη δημιουργία μιας βιώσιμης και ανταγωνιστικής αγοράς για την εμπορία και την επένδυση του Bitcoin και άλλων κρυπτονομισμάτων.

• Το bitcoin σήμερα

Σήμερα, ένα Bitcoin αξίζει περίπου 40.000 δολάρια. Απέχει πολύ από την υψηλότερη τιμή όλων των εποχών, αλλά και πολύ από την χαμηλότερη τιμή του μετά την κορύφωση του 2018, μόλις λίγο πάνω από 3.000 δολάρια. Το Bitcoin είναι ένας καλός δείκτης της αγοράς κρυπτονομισμάτων, επειδή είναι το κυρίαρχο κρυπτονομίσμα ανά κεφαλαιοποίηση και η υπόλοιπη αγορά τείνει να ακολουθεί τις τάσεις του.

Η τιμή του Bitcoin είχε μια ραγδαία άνοδο το 2021 και τον Νοέμβριο έφτασε σε μια νέα υψηλή τιμή όλων των εποχών όταν ξεπέρασε τα 68.000 δολάρια. Αυτό το τελευταίο υψηλό ρεκόρ ακολουθεί τις υψηλές τιμές άνω των 60.000 δολαρίων τον Απρίλιο και τον Οκτώβριο, και έπειτα από μια καλοκαιρινή πτώση σε λιγότερο από 30.000 δολάρια τον Ιούλιο.

Πολλοί ειδικοί λένε ότι είναι μόνο θέμα χρόνου το Bitcoin να φτάσει τα 100.000 δολάρια. Το παρελθόν του Bitcoin μπορεί να παρέχει κάποιες ενδείξεις για το τι να



Σχήμα 2.1: *Bitcoin (BTC) price per day from October 2013 to February 7, 2022 (in U.S. dollars)*

[11]

περιμένουμε στο μέλλον.

Η ταυτότητα του δημιουργού του Bitcoin παραμένει μυστήριο. Μέχρι σήμερα, κανείς δεν ξέρει ποιος είναι ή ποιος ήταν ο Satoshi Nakamoto. Είναι ένα θέμα όχι μόνο για συζήτηση, αλλά υπάρχουν και εικασίες και –ίσως αναπόφευκτα– θεωρίες συνωμοσίας. Αυτές οι θεωρίες αφθονούν, συμπεριλαμβανομένου του Bitcoin να είναι στην πραγματικότητα ένα μυστικό έργο μιας μεγάλης οντότητας ή μιας υπηρεσίας πληροφοριών. Άλλοι πιστεύουν ότι είναι ένα «έργο παγίδας», το οποίο, όταν γίνει αρκετά μεγάλο, μια κακόβουλη οντότητα που περιμένει υπομονετικά για κάποιο χρονικό διάστημα θα καταλάβει ξαφνικά τον έλεγχο.

• Το Bitcoin αύριο

Η επόμενη δεκαετία θα μπορούσε να αποδειχθεί σημαντική για την εξέλιξη του Bitcoin. Πέρα από τις επαναστάσεις στο χρηματοοικονομικό οικοσύστημα, υπάρχουν δύο τομείς στο πεδίο του Bitcoin στους οποίους οι επενδυτές πρέπει να δώσουν ιδιαίτερη προσοχή.

Επί του παρόντος, το κρυπτονόμισμα βρίσκεται ανάμεσα στο να είναι μια μονάδα αποθηκευτικής αξίας και ένα μέσο για καθημερινές συναλλαγές. Οι θεσμικοί επενδυτές είναι πρόθυμοι να συμμετάσχουν στη δράση και να επωφεληθούν από την αστάθεια των τιμών, ακόμη και όταν κυβερνήσεις σε όλο τον κόσμο, όπως η Ιαπωνία, το έχουν δηλώσει ως έγκυρη μορφή πληρωμής για αγαθά.

Αναμφισβήτητα οι μεγαλύτερες αποτυχίες του Bitcoin και άλλων κρυπτονομισμάτων τα προηγούμενα χρόνια έγκεινται στην ασφάλεια.

Η ενσωμάτωση του Bitcoin ως μηχανισμού πληρωμών (ή εν προκειμένω, η αυξανόμενη ελκυστικότητά του ως κατηγορία περιουσιακών στοιχείων) δεν θα συμβεί χωρίς τεχνολογικές βελτιώσεις στο σύστημα του. Για να θεωρείται βιώσιμο επενδυτικό περιουσιακό στοιχείο ή τρόπος πληρωμής, το blockchain του Bitcoin θα πρέπει να μπορεί να χειρίζεται εκατομμύρια συναλλαγές σε σύντομο χρονικό διάστημα.

Η επόμενη δεκαετία θα φέρει μια έκρηξη πληρωμών χαμηλού κόστους και υψηλής ταχύτητας που θα μεταμορφώσουν την ανταλλακτική αξία με τρόπο όμοιο όπως το Διαδίκτυο μεταμόρφωσε την ανταλλαγή πληροφοριών.

2.1.2 Εξόρυξη του Bitcoin

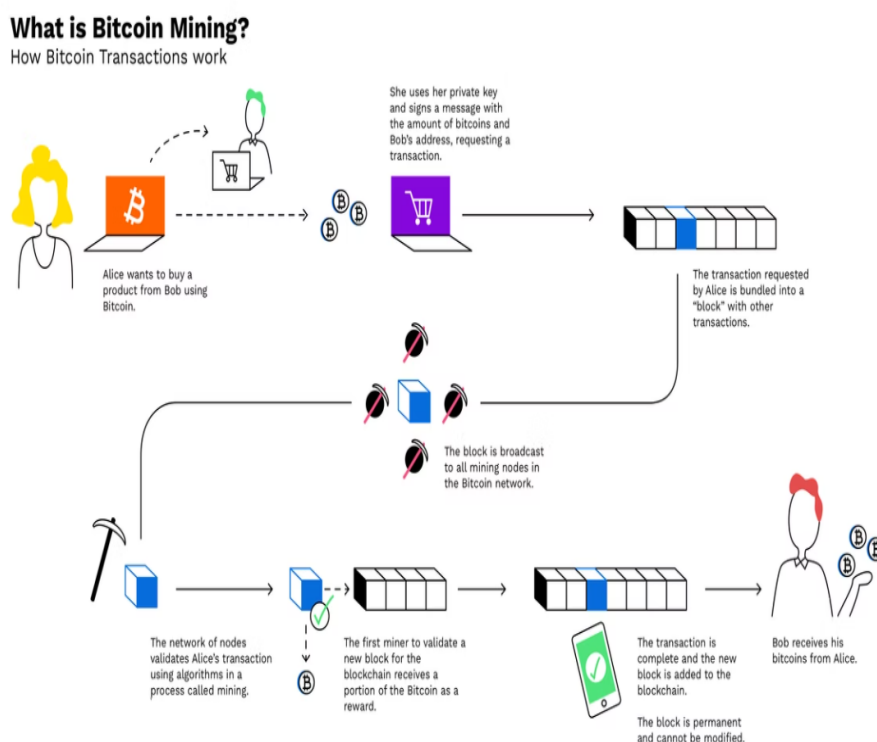
Τα κρυπτονομίσματα δημιουργούνται με εξόρυξη μέσω του blockchain χρησιμοποιώντας την τεχνολογία κρυπτογραφίας. Αυτή είναι η μέθοδος έκδοσης νέου κρυπτονομίσματος. Το σύστημα blockchain αποτελείται από χρήστες, προγραμματιστές, εξορύκτες, συντηρητές κόμβων και τις αλληλεπιδράσεις που διασφαλίζουν τη λειτουργικότητα των κατανεμημένων λογισμικών. Μια τέτοια διαδικασία εξόρυξης απαιτεί από τους εξορύκτες να έχουν κεφαλαιουχικά έξοδα για την αγορά του λογισμικού και του υλικού. Η εξόρυξη άλλων νομισμάτων που χρησιμοποιεί πολλούς διαφορετικούς αλγόριθμους απαιτεί τη χρήση καρτών γραφικών υψηλής τεχνολογίας και υψηλής ταχύτητας. Για έναν νέο εξορύκτη, χρειάζεται να καταχωρήσει ένα πορτοφόλι-δηλαδή και μια κρυπτογραφημένη τραπεζική συναλλαγή στο Διαδίκτυο που μπορεί να αποθηκεύσει και να αποδεχτεί το κρυπτονόμισμα. Σε αυτήν την τοποθεσία, θα λαμβάνονται τα Bitcoin και θα υπάρχει επιλογή ελέγχου αυτών. Τα πορτοφόλια Bitcoin παρέχουν τη δυνατότητα να ελέγχονται τα δεδομένα Blockchain. Κάθε διεύθυνση Bitcoin αντιπροσωπεύεται από έναν συνδυασμό ψηφιακής υπογραφής και αντίστοιχου ιδιωτικού κλειδιού. Όταν ένας εξορύκτης είναι σε θέση να λύσει τους πολύπλοκους μαθηματικούς υπολογισμούς που είναι ενσωματωμένο το μπλοκ στο σύστημα blockchain, τα ψηφιακά νομίσματα θα ανταμειφθούν και θα μεταφερθούν στο πορτοφόλι που έχει προκαθοριστεί νωρίτερα.[12]

Σύμφωνα με πολλά από τα πρωτόκολλα κρυπτονομισμάτων, ο τρόπος με τον οποίο λειτουργεί η εξόρυξη είναι με την επικύρωση της συναλλαγής μέσω σύνδεσης με το μπλοκ που είχε γίνει αποδεκτό νωρίτερα. Η τεχνολογία blockchain καταγράφει κάθε συναλλαγή στη μονάδα της. Ένα μοναδικό αναγνωριστικό εκχωρείται σε κάθε μπλοκ και στο μπλοκ που προηγείται. Αυτό ονομάζεται πρωτόκολλο απόδειξης εργασίας (Proof-of-Work είναι ένα πρωτόκολλο επαλήθευσης μιας συναλλαγής και ενημέρωσης όλων των υπόλοιπων). Οι χρήστες ή οι εξορύκτες πρέπει να εργαστούν για την επικύρωση ή την απόδειξη ότι είναι οι πραγματικές οντότητες.

Οι πιο επιτυχημένοι εξορύκτες ανταμείβονται με νέα Bitcoin εάν καταφέρουν να προσθέσουν επιτυχώς ένα νέο μπλοκ στην αλυσίδα των μπλοκ. Στη σημερινή εποχή, το έπαθλο δεν αποδίδεται σε ένα μεμονωμένο άτομο, διότι κανείς στον κόσμο δεν έχει αρκετή υπολογιστική ισχύ στη διάθεσή του, ώστε να λύσει τις πολύπλοκες μαθηματικές πράξεις που χρειάζονται για ένα επιτυχές μπλοκ.

Συνεπώς, οι εξορύκτες συνενώνονται στις επονομαζόμενες «δεξαμενές εξόρυξης» και συνεργάζονται ενώνοντας τις δυνάμεις τους. Στη συνέχεια, το έπαθλο μοιράζεται αναλογικά με την εργασία που προηγήθηκε. Αυτοί με τη μεγαλύτερη υπολογιστική ισχύ λαμβάνουν μεγαλύτερο έπαθλο.

Οι εξορύκτες αρχικά ανταμείβονταν με 50 Bitcoin, αλλά το 2012 το έπαθλο υποδιπλασιάστηκε στα 25 Bitcoin. Ο πιο πρόσφατος υποδιπλασιασμός συνέβη το 2016 και μέχρι τη στιγμή που μιλάμε, το τρέχον έπαθλο είναι 12,5 Bitcoin. Ο επόμενος υποδιπλασιασμός θα λάβει χώρα όταν ο αριθμός των μπλοκ αγγίξει τις 630.000. Το έπαθλο υποδιπλασιάζεται περίπου κάθε τέσσερα χρόνια. [13]



Σχήμα 2.2: Μεθοδολογία Συναλλαγών για συναλλαγές bitcoin[1]

2.1.3 Τεχνική Ανάλυση

Η τεχνική ανάλυση που εφαρμόζεται στα κρυπτονομίσματα περιλαμβάνει τη χρήση μαθηματικών δεικτών που βασίζονται σε δεδομένα προηγούμενων τιμών για να προσπαθήσουμε να προβλέψουμε τις μελλοντικές τάσεις. Η βασική ιδέα είναι ότι οι αγορές συμπεριφέρονται σύμφωνα με ορισμένα μοτίβα και ότι όταν καθιερωθούν, οι τάσεις που κινούνται προς μια συγκεκριμένη κατεύθυνση συχνά συνεχίζουν στην ίδια πορεία για κάποιο χρονικό διάστημα.

Η τεχνική ανάλυση θεωρείται ως ένα μέσο που ενημερώνει τους εμπόρους σχετικά με τη μη θεμελιώδη τιμή και τους καθοριστικούς παράγοντες αυτής, που είναι πιο σημαντικό στις βραχυπρόθεσμες και άκαμπτες αγορές συναλλαγματικών ισοτιμιών, όπου ο βαθμός μεταβλητότητας δεν μπορεί να εξηγηθεί μόνο με θεμελιώδη στοιχεία.

Σε γενικές γραμμές, οι επενδυτές θέλουν να αγοράζουν όταν οι αγορές είναι χαμηλές, ώστε να μπορούν να πουλήσουν υψηλότερα κάποια στιγμή στο μέλλον, και έτσι να έχουν

κέρδος. Η διεξαγωγή τεχνικής ανάλυσης πριν από την είσοδο σε μια θέση είναι ένας τρόπος για να εντοπιστούν τα επίπεδα τιμών που μπορεί να θεωρηθούν χαμηλά.

Δεν υπάρχει ενιαία, ολοκληρωμένη μέθοδος για τεχνική ανάλυση κρυπτονομίσματος. Κάθε έμπορος προτιμάει να χρησιμοποιεί διαφορετικούς δείκτες και πιθανότατα θα τους ερμηνεύσει ελαφρώς διαφορετικά. Θα πρέπει επίσης να σημειωθεί ότι καμία τεχνική ανάλυση δεν είναι σχεδόν 100 τοις εκατό προγνωστική.

Υπάρχουν εκατοντάδες πιθανοί προγνωστικοί παράγοντες που μπορούν να προκύψουν από την τεχνική ανάλυση. Αυτοί παρέχουν δύο χρήσιμα πλεονεκτήματα. Πρώτον, απαιτούν μόνο ιστορικά δεδομένα της τιμής και τον όγκο των συναλλαγών και επομένως είναι εύκολο να εφαρμοστούν. Δεύτερον, ο τεράστιος αριθμός των προγνωστικών δεικτών είναι κατάλληλος για μηχανική μάθηση, καθώς είναι ένα εργαλείο που ενδείκνυται για την αντιμετώπιση πολλαπλών χαρακτηριστικών. [14]

2.1.4 Ανάλυση χρονοσειρών σε οικονομικό πλαίσιο

Η αξία του Bitcoin σε δολάριο ΗΠΑ μεταβάλλεται καθημερινά, γεγονός που καθιστά δύσκολη την πρόβλεψη μελλοντικών τιμών Bitcoin. Οι τιμές του Bitcoin είναι ιδιαίτερα δύσκολο να προβλεφθούν, καθώς έχει ακόμη πιο ασταθή χαρακτήρα από τα περισσότερα χρηματιστήρια. [15] Η γνώση της μελλοντικής αξίας του Bitcoin, φυσικά, αποφέρει κέρδη για τους επενδυτές διότι γνωρίζουν πότε να αγοράσουν ή να πουλήσουν Bitcoin και να αποκτήσουν το μέγιστο κέρδος. Για παράδειγμα, όταν αναμένεται να αυξηθεί η τιμή του Bitcoin μπορούν να επενδύσουν και να αγοράσουν ενώ όταν πρόκειται να πέσει η τιμή μπορούν να πουλήσουν και να είναι κερδισμένοι. Η πρόβλεψη μπορεί να είναι δυνατή εάν υπάρχει σχέση μεταξύ ιστορικών δεδομένων και της μελλοντικής αξίας Bitcoin. Στη συνέχεια, η μηχανική μάθηση μπορεί να χρησιμοποιηθεί για την αναγνώριση προτύπων μεταξύ ιστορικών δεδομένων και μελλοντικών τιμών Bitcoin.

Η πρόβλεψη μπορεί να επιτευχθεί χρησιμοποιώντας ταξινόμηση ή παλινδρόμηση. Η κύρια διαφορά μεταξύ ταξινόμησης και παλινδρόμησης είναι ότι η ταξινόμηση εξαρτάται από μεταβλητές που δεν είναι ταξινομημένες ενώ η παλινδρόμηση εξαρτάται από διατεταγμένες μεταβλητές συνεχούς αξίας. Η παλινδρόμηση αποσκοπεί σε προβλέψεις ακριβείας μελλοντικών τιμών του Bitcoin ενώ η ταξινόμηση επιτρέπει προβλέψη που έγκειται σε περιορισμένο αριθμό κλάσεων. Αυτή η εργασία στοχεύει στην πρόβλεψη των τιμών Bitcoin χρησιμοποιώντας παλινδρόμηση. [16]

2.2 Τεχνολογία Blockchain

Ως νέος τομέας της χρηματοοικονομικής τεχνολογίας, η αποκεντρωμένη οικονομία μπορεί να αναδιαμορφώσει τη δομή της σύγχρονης οικονομίας και τη δημιουργία ενός νέου τοπίου για την επιχειρηματικότητα και την καινοτομία, εκπληρώνοντας τις υποσχέσεις και τις προκλήσεις των αποκεντρωμένων επιχειρηματικών μοντέλων.

2.2.1 Η εμφάνιση της αποκεντρωμένης οικονομίας

Για αιώνες, τα χρηματοπιστωτικά ιδρύματα διαδραμάτισαν σημαντικό ρόλο στη διαμεσολάβηση και τη διάρθρωση των οικονομικών συναλλαγών που διαφορετικά θα ήταν δύσκολο να γίνουν. Με κατανεμημένη εμπιστοσύνη και αποκεντρωμένες πλατφόρμες που ενεργοποιούνται από την τεχνολογία blockchain, οι επιχειρηματίες και οι καινοτόμοι έχουν αναγνωρίσει τις δυνατότητες δημιουργίας ενός ανοιχτού χρηματοπιστωτικού συστήματος που έχει περιορισμένη ή καθόλου συμμετοχή από χρηματοπιστωτικά ιδρύματα. Με αυτή την κίνηση, σκοπεύουν να μειώσουν το κόστος συναλλαγής, να διευρύνουν την οικονομική ένταξη, να ενισχύσουν την ανοιχτή πρόσβαση, να ενθαρρύνουν την καινοτομία και να δημιουργήσουν νέες επιχειρηματικές ευκαιρίες. [17]

2.2.2 Οι υποσχέσεις της αποκεντρωμένης οικονομίας

- Αποκέντρωση

Τα χρηματοπιστωτικά ιδρύματα μπορούν να αναπτυχθούν για να κυριαρχήσουν στις οικονομικές δραστηριότητες ως βασικοί μεσάζοντες που διευκολύνουν τις χρηματοπιστωτικές συναλλαγές. Όταν ένα κεντρικό χρηματοπιστωτικό ίδρυμα αποκτά κυριαρχία, μπορεί να συσσωρεύσει δυσανάλογη ισχύ και κέρδη στην αγορά. Σε ένα αποκεντρωμένο χρηματοπιστωτικό σύστημα, αντίθετα, οι χρηματοοικονομικές συναλλαγές διευκολύνονται όχι από τα κεντρικά ιδρύματα αλλά από αποκεντρωμένα δίκτυα peer-to-peer. Μειώνοντας τη συμμετοχή κεντρικών ιδρυμάτων, τα αποκεντρωμένα δίκτυα μπορούν να μειώσουν το κόστος συναλλαγών χωρίς να επιβαρύνουν με μονοπωλιακά κόστη. Όταν ένα αποκεντρωμένο δίκτυο peer-to-peer αποκτά κυριαρχία, καμία μεμονωμένη οντότητα δεν μπορεί να συγκεντρώσει επαρκή μονοπωλιακή ισχύ για να μονοπωλήσει το δίκτυο, επιτρέποντας σε όλους να επωφεληθούν από τα αποτελέσματα του δικτύου για να διευρυνθούν δυνατότητες συναλλαγής.

- Καινοτομία

Μια αποκεντρωμένη πλατφόρμα δεν έχει μέρος ελέγχου και ως εκ τούτου, επιτρέπει την ανοιχτή πρόσβαση και την καινοτομία -χωρίς άδεια— δηλαδή, οι προγραμματιστές μπορούν ελεύθερα να δημιουργήσουν και να πειραματιστούν με νέες εφαρμογές χωρίς να ζητήσουν άδεια. Με αυτήν τη διευκόλυνση, οι αποκεντρωμένες πλατφόρμες ενδυναμώνουν τους προγραμματιστές εξασφαλίζοντας πρόσβαση, επιτρέποντας τους να εξελιχθούν. Αντίθετα, οι κεντρικές χρηματοοικονομικές υπηρεσίες και πλατφόρμες συχνά προστατεύουν τα πνευματικά δικαιώματα αποκλείοντας άλλους από τη χρήση πολύτιμων τεχνολογιών. Αυτή η ελευθερία των αποκεντρωμένων συστημάτων μπορεί να επιταχύνει τον ρυθμό της χρηματοοικονομικής καινοτομίας καθώς και να αυξήσει τον βαθμό ανταγωνισμού της αγοράς, οδηγώντας δυνητικά σε νεότερες, καλύτερες και φθηνότερες χρηματοοικονομικές υπηρεσίες.

- Διαλειτουργικότητα

Η αποκεντρωμένη χρηματοδότηση μπορεί να ενισχύσει τη διαλειτουργικότητα. Η παραδοσιακή χρηματοδότηση τείνει να αυξάνει τα εμπόδια στις συναλλαγές. Για παράδειγμα, διαφορετικά χρηματοπιστωτικά ιδρύματα πρέπει να διατηρούν τα δικά τους

λογιστικά βιβλία, επομένως μια χρηματοοικονομική υπηρεσία μπορεί να μην είναι διαλειτουργική με μια άλλη. Ως αποτέλεσμα, η μεταφορά κεφαλαίου και αξίας μπορεί να γίνει δαπανηρή και δυσκίνητη. Αντίθετα, η αποκεντρωμένη χρηματοδότηση βασίζεται στο δημόσιο blockchain και σε ανοιχτά πρότυπα, αυξάνοντας τη διαλειτουργικότητα σε διάφορες υπηρεσίες. Με υψηλή διαλειτουργικότητα, το οικονομικό κεφάλαιο και η αξία μπορεί να ρέει απρόσκοπτα σε διαφορετικές υπηρεσίες και σύνορα, δημιουργώντας ενδεχομένως ένα διαδίκτυο αξίας.

- Χωρίς σύνορα

Η κεντρική οικονομία δε λειτουργεί εύκολα εκτός συνόρων, καθώς συνδέεται με συγκεκριμένες γεωγραφικές τοποθεσίες και με συγκεκριμένα νομίσματα. Σαν αποτέλεσμα, η διασυνοριακή μετακίνηση κεφαλαίου και αξίας συχνά αντιμετωπίζει τριβές και καθυστερήσεις. Αντίθετα, η αποκεντρωμένη χρηματοδότηση είναι εγγενώς χωρίς σύνορα και έτσι επιτρέπει τη χρηματοδότηση χωρίς σύνορα, καθώς δεν συνδέεται με γεωγραφικές τοποθεσίες ή νομίσματα. Βασιζόμενοι σε κρυπτονομίσματα χωρίς σύνορα, η αποκεντρωμένη οικονομία μπορεί να χρησιμοποιηθεί από οποιονδήποτε σε όλο τον κόσμο. Με την αποκεντρωμένη χρηματοδότηση, επομένως, η μεταφορά αξίας σε ολόκληρο τον κόσμο μπορεί να γίνει όπως εύκολη όσο η αποστολή ενός email, καταργώντας τα εμπόδια στις παγκόσμιες μεταφορές αξίας.

- Διαφάνεια

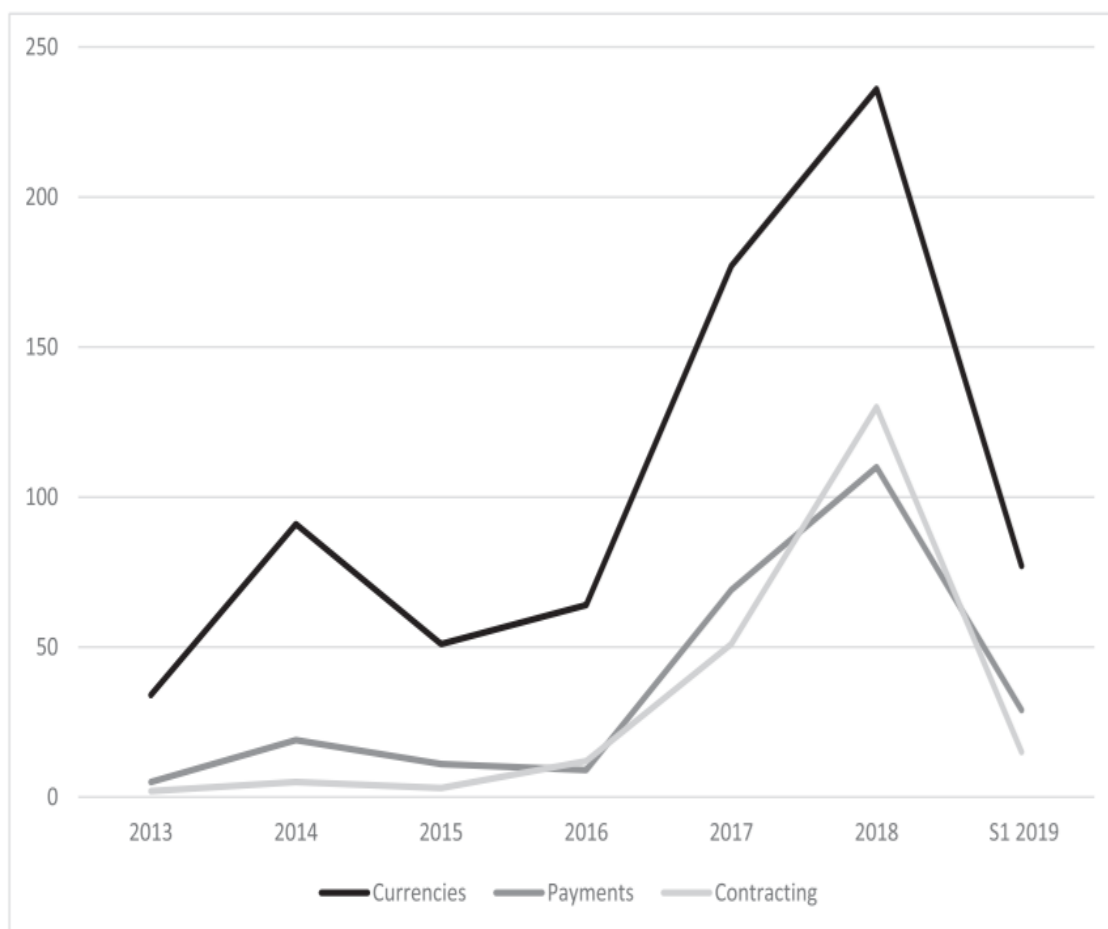
Η αποκεντρωμένη οικονομία μπορεί επίσης να ενισχύσει τη διαφάνεια στο χρηματοπιστωτικό σύστημα. Η κεντρική οικονομία δεν μπορεί να έχει πλήρη διαφάνεια, καθώς τα κεντρικά χρηματοπιστωτικά ιδρύματα πρέπει να διασφαλίσουν τα κεντρικά τους βιβλία περιορίζοντας την πρόσβαση. Αντίθετα, η αποκεντρωμένη χρηματοδότηση εξασφαλίζει τα δημόσια βιβλία της μέσω κατανεμημένης συναίνεσης και ριζικής διαφάνειας. Καταγράφει συναλλαγές σε δημόσια λογιστικά βιβλία που μπορούν εύκολα να προβληθούν και να επαληθευτούν. Με τα δημόσια λογιστικά βιβλία, η αποκεντρωμένη χρηματοδότηση δημιουργεί κατανεμημένη εμπιστοσύνη, έτσι τα μέρη που συναλλάσσονται μπορούν να συναλλάσσονται με άλλα χωρίς προϋπάρχουσες σχέσεις ή αξιόπιστο ενδιάμεσο, διευρύνοντας την κλίμακα και το εύρος των πιθανών συναλλαγών.

2.2.3 Κύρια επιχειρηματικά μοντέλα στην αποκεντρωμένη οικονομία

Η τεχνολογία Blockchain μπορεί να οδηγήσει στην εμφάνιση νέων επιχειρηματικών μοντέλων που προηγουμένως δεν ήταν βιώσιμα. Στις χρηματοοικονομικές βιομηχανίες, η τεχνολογία blockchain μπορεί να μειώσει τη συμμετοχή κεντρικών ιδρυμάτων, να ενθαρρύνει τον πειραματισμό και να διευρύνει την πρόσβαση σε χρηματοπιστωτικές υπηρεσίες. Το Σχήμα 2.3 απεικονίζει την άνοδο ορισμένων επιχειρηματικών μοντέλων. Επί του παρόντος, τα αποκεντρωμένα νομίσματα είναι το πιο κοινό μοντέλο, ακολουθούμενο από συμβάσεις και πληρωμές.

- Αποκεντρωμένες συμβάσεις

Τα συμβόλαια είναι απαραίτητα για τις αγορές, τις εταιρείες και τα άτομα, καθώς διευκολύνουν τις συνεργασίες και τις συναλλαγές. Ωστόσο, τα συμβόλαια μπορεί να είναι



Σχήμα 2.3: Η άνοδος των χρηματοοικονομικών υπηρεσιών που βασίζονται σε blockchain [18]

περίπλοκα και δαπανηρά, λόγω του κόστους της διαπραγμάτευσης, της σύνταξης, της επιβολής και της επαναδιαπραγμάτευσης συμφωνιών. Παραδοσιακά, τα μέρη που πραγματοποιούνται οι συναλλαγές συχνά βασίζονται σε χρηματοοικονομικούς ενδιάμεσους για να δημιουργήσουν εμπιστοσύνη και να μειώσουν το κόστος συναλλαγής. Τα τελευταία χρόνια, η τεχνολογία blockchain έχει αρχίσει να διευκολύνει τις χρηματοοικονομικές συμβάσεις αντικαθιστώντας τους ενδιάμεσους χρηματοπιστωτικούς οργανισμούς με έξυπνα συμβόλαια, οδηγώντας στην άνοδο των χρηματοοικονομικών συμβάσεων peer-to-peer. Τα έξυπνα συμβόλαια είναι προγράμματα που εκτελούνται αυτόματα όταν πληρούνται προκαθορισμένες προϋποθέσεις (δηλ. κανόνες) στα πρωτόκολλα. Τα έξυπνα συμβόλαια υπόσχονται να μειώσουν την πολυπλοκότητα και το κόστος των συμβάσεων, χάρη στη διαφάνεια, την αμεταβλητότητα, την αυτοματοποίηση και τον προγραμματισμό τους.

- Αποκεντρωμένα νομίσματα

Τα εθνικά νομίσματα υπάρχουν εδώ και αιώνες. Μέχρι πριν από μερικές δεκαετίες, τα νομίσματα υποστηρίζονταν από πολύτιμα μέταλλα όπως ο χρυσός και βασίστηκαν σε κεντρικούς θεσμούς για τη διαχείριση της υποστήριξης των νομισμάτων με χρυσό. Τα τελευταία χρόνια, ωστόσο, τα νομίσματα δεν υποστηρίζονται πλέον από χρυσό. Σήμερα,

η αξία ενός νομίσματος είναι καθαρά συνδεδεμένη με την εμπιστοσύνη που έχουν οι άνθρωποι στη συγκεκριμένη χώρα, οικονομία, κυβέρνηση και κεντρική τράπεζα.

Το Bitcoin είναι το πρώτο αποκεντρωμένο κρυπτονόμισμα που δεν εκδίδεται από καμία χώρα αλλά μέσω μιας αποκεντρωμένης τεχνολογίας. Σε αντίθεση με τα παραδοσιακά νομίσματα που εκδίδονται από τις κεντρικές τράπεζες, το πρόγραμμα προμήθειας του Bitcoin είναι σταθερό και δεν μπορεί να αλλάξει. Δεδομένης της αποκέντρωσής του, το Bitcoin έχει γίνει το κύριο κατάστημα αξίας στη βιομηχανία blockchain και αναφέρεται συχνά ως ψηφιακός χρυσός. Όπως ο χρυσός, το Bitcoin είναι εγγενώς χωρίς σύνορα και μπορεί να αποθηκευτεί και να μεταφερθεί χωρίς εμπλοκή οποιασδήποτε κεντρικής οντότητας. Εκτός από το Bitcoin, άλλα μεγάλα αποκεντρωμένα κρυπτονομίσματα είναι τα Ether, Litecoin, Monero κ.ο.κ.

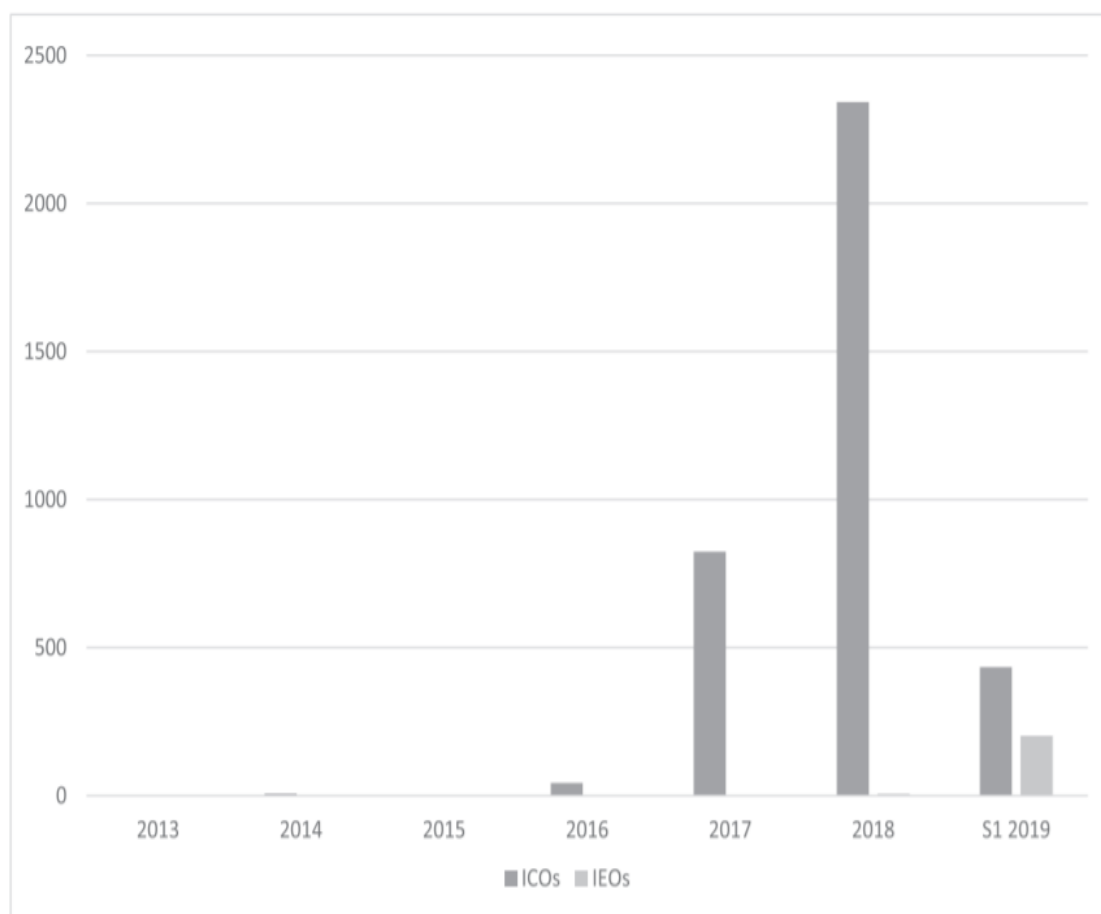
- Αποκεντρωμένες υπηρεσίες πληρωμών

Τα κεντρικά δίκτυα πληρωμών—όπως Visa, PayPal και το SWIFT—διευκολύνουν το διαδικτυακό και το εκτός σύνδεσης εμπόριο, αλλά συνήθως χρεώνουν σχετικά υψηλές χρεώσεις για τις υπηρεσίες τους, ιδίως για διασυνοριακές πληρωμές. Το κόστος των υπηρεσιών πληρωμών παραμένει υψηλό, ακόμη και αφού το Διαδίκτυο έχει μειώσει σημαντικά το κόστος μιας μεγάλης ποικιλίας υπηρεσιών. Αποκεντρωμένα δίκτυα πληρωμών—όπως το Libra (Libra Association, 2019) και το Bitcoin Lightning Network (Poon and Dryja, 2016)—υπόσχονται να πραγματοποιήσουν πληρωμές χαμηλού κόστους, άμεσες και παγκόσμιες, επιλύοντας προβλήματα που σχετίζονται με τις παραδοσιακές υπηρεσίες πληρωμών. Με ελάχιστη χρέωση, για παράδειγμα, Bitcoin Lightning Network προσφέρει άμεσες, ασφαλείς και μη αναστρέψιμες υπηρεσίες πληρωμής. Λόγω των χαμηλών χρεώσεων των συναλλαγών, οι έμποροι μπορούν να μειώσουν σημαντικά το κόστος τους και να βελτιώσουν την κερδοφορία τους. Δεδομένου ότι οι αποκεντρωμένες υπηρεσίες πληρωμών βασίζονται στην τεχνολογία blockchain και τα κρυπτονομίσματα που είναι εγγενώς παγκόσμια, οι διασυνοριακές πληρωμές μπορούν πλέον να γίνουν γρήγορες και φθηνές. Η Ripple είναι αυτή τη στιγμή πρωτοπόρος σε αυτόν τον τομέα και έχει συνεργαστεί με πολλά χρηματοπιστωτικά ιδρύματα για τη βελτίωση της αποτελεσματικότητας της διασυνοριακής μεταφοράς χρήματος. [19]

- Αποκεντρωμένη συγκέντρωση χρημάτων

Τα τελευταία χρόνια, τα Initial Coin Offering (ICO) έχουν αναδειχθεί ως ένας καινοτόμος μηχανισμός χρηματοδότησης για εγχειρήματα πρώιμου σταδίου, δίνοντας τη δυνατότητα σε επιχειρηματίες να συγκεντρώσουν δισεκατομμύρια δολάρια από παγκόσμιους επενδυτές. Το ICO είναι ένας δυνητικά ισχυρός τρόπος για ένα έργο να συγκεντρώσει κεφάλαια. Ως νέα μορφή συγκέντρωσης χρημάτων, επιτρέπει για ένα έργο την άντληση κεφαλαίων από επενδυτές σε όλο τον κόσμο - χάρη στη διαφάνεια της τεχνολογίας blockchain, των έξυπνων συμβολαίων και του ανοιχτού πηγαίου κώδικα. Βασιζόμενοι στην κατανεμημένη εμπιστοσύνη που δημιουργείται από το blockchain, η αποκεντρωμένη συγκέντρωση κεφαλαίων μπορεί να μειώσει την τριβή στη συγκέντρωση κεφαλαίων, να διευκολύνει πρόσβαση στο κεφάλαιο και, ως εκ τούτου, να προωθήσει την επιχειρηματικότητα και την καινοτομία. Πολλά κακόβουλα ICO και οι αβέβαιες

προοπτικές ακόμη και για εκείνες τις startups που δεν αποτελούν απάτη αποτελούν ένα σοβαρό πρόβλημα. Μια μέθοδος επίλυσης αυτών των προβλημάτων ήταν η συμμετοχή ενός τρίτου μέρους που θα μπορούσε να εξαλείψει τα κακόβουλα ICO και να βοηθήσει υποσχόμενες startups να προσελκύσουν επενδύσεις. Τα ανταλλακτήρια κρυπτογράφησης ανέλαβαν αυτόν τον ρόλο τρίτου μέρους. Έτσι εμφανίστηκαν για πρώτη φορά οι Αρχικές Προσφορές Ανταλλαγής (IEO). Σε αντίθεση με τα ICO, τα IEO βασίζονται σε ανταλλακτήρια κρυπτονομισμάτων για να διασφαλιστεί η αξιοπιστία των πιθανών έργων και να συνδεθούν έργα υψηλής ποιότητας με πιθανούς επενδυτές. Το Σχήμα 2.4 δείχνει την άνοδο των ICO και IEOs τις πρόσφατες περιόδους.



Σχήμα 2.4: *The rise of ICOs and IEOs*
[18]

Κεφάλαιο 3

Νευρωνικά Δίκτυα

Στο κεφάλαιο αυτό αναλύεται η έννοια της μηχανικής μάθησης, και γίνεται απαρίθμηση των διαφορετικών ειδών μηχανικής μάθησης ενώ αναλύονται και τα είδη των προβλημάτων ανάλογα με το ζητούμενο που καλούμαστε σε κάθε πρόβλημα να προβλέψουμε. Στη συνέχεια, γίνεται μια αναλυτική περιγραφή των αρχιτεκτονικών στις οποίες είναι βασισμένη η εργασία.

3.1 Εισαγωγή στη Μηχανική Μάθηση

Η μηχανική μάθηση (ML) είναι η επιστημονική μελέτη αλγορίθμων και στατιστικών μοντέλων που χρησιμοποιούν τα συστήματα υπολογιστών για να κάνουν ακριβείς προβλέψεις μαθαίνοντας από δεδομένα. Κάθε φορά που η μηχανή αναζήτησης Ιστού όπως το Google χρησιμοποιείται για την αναζήτηση στο διαδίκτυο, ένας από τους λόγους που λειτουργεί τόσο καλά είναι επειδή ένας αλγόριθμος εκμάθησης έχει μάθει πώς να ταξινομεί ιστοσελίδες. Οι μέθοδοι της μηχανικής μάθησης έχουν ρει εφαρμογές σε διάφορες περιοχές όπως η μηχανική όραση, η επεξεργασία υσικής γλώσσας, οι μηχανές αναζήτησης παγκόσμιου ιστού, η ιατρική διάγνωση, η οικονομική ανάλυση, η ταξινόμηση αλληλουχιών DNA, η αναγνώριση ωνής και η αναγνώριση οπτικών χαρακτήρων, η ανάπτυξη ευφυών συστημάτων συμμετοχής σε παιχνίδια, η ανάπτυξη λογισμικού, η αναγνώριση αστρονομικών αντικειμένων, η ομποτική, κτλ. Μία ενδιαφέρουσα περιοχή εφαρμογής μεθόδων μηχανικής μάθησης, στην οποία κατατάσσεται και η παρούσα εργασία, αποτελεί η Οικονομική Ανάλυση.

3.1.1 Random Forest Regressor

Τα τυχαία δάση είναι ένας από τους πιο δημοφιλείς αλγόριθμους μηχανικής μάθησης. Είναι τόσο επιτυχημένα επειδή παρέχουν γενικά καλή προγνωστική απόδοση, χαμηλή υπερπροσαρμογή και εύκολη ερμηνεία. Αυτή η ερμηνευσιμότητα δίνεται από το γεγονός ότι είναι εύκολο να εξαχθεί η σημασία κάθε μεταβλητής (feature) στην απόφαση δέντρου. Με άλλα λόγια, είναι εύκολο να υπολογίσουμε πόσο συμβάλλει κάθε μεταβλητή στην απόφαση.[20] Το Random Forest είναι ένα πολύ ισχυρό μοντέλο τόσο για παλινδρόμηση όσο και για ταξινόμηση. Μπορεί επίσης να δώσει τη δική του ερμηνεία της σημασίας των χαρακτηριστικών, η οποία μπορεί να σχεδιαστεί και να χρησιμοποιηθεί για την επιλογή του πιο ενημερωτικού συνόλου χαρακτηριστικών σύμφωνα, για παράδειγμα, με μια διαδικασία Εξάλειψης

Αναδρομικών Χαρακτηριστικών (Recursive Feature Elimination). Η αναδρομική εξάλειψη χαρακτηριστικών (RFE) είναι μια μέθοδος επιλογής χαρακτηριστικών που ταιριάζει σε ένα μοντέλο και καταργεί το πιο αδύναμο χαρακτηριστικό (ή χαρακτηριστικά) μέχρι να επιτευχθεί ο καθορισμένος αριθμός χαρακτηριστικών. Τα χαρακτηριστικά ταξινομούνται με βάση τα χαρακτηριστικά (feature importances) του μοντέλου εξαλείφοντας έτσι αναδρομικά ένα μικρό αριθμό χαρακτηριστικών ανά βρόχο.[21] Αν χρησιμοποιηθεί σωστά, η σημασία των χαρακτηριστικών (Feature Selection) μπορεί να μας δώσει αποτελεσματική βελτιστοποίηση (επιλογή χαρακτηριστικών).[22]

3.1.2 Είδη μηχανικής μάθησης

Στη μηχανική μάθηση υπήρχαν παραδοσιακά οι όροι μάθηση με επίβλεψη (supervised learning) και μάθηση χωρίς επίβλεψη (unsupervised learning). Η μάθηση με επίβλεψη ταυτίζεται με την πρώτη κατηγορία προβλημάτων, δηλαδή της ταξινόμησης και παλινδρόμησης. Το όνομα προέρχεται από το γεγονός ότι σε αυτά τα προβλήματα υπάρχει κάποιος «επιβλέπων» ο οποίος μας παρέχει την τιμή εξόδου της συνάρτησης για τα δεδομένα που εξετάζουμε. Η μάθηση χωρίς επίβλεψη ταυτιζόταν παραδοσιακά με το πρόβλημα της ομαδοποίησης. Ο λόγος είναι ότι στην ομαδοποίηση δεν υπάρχει κάποιος «επιβλέπων» αφού δε γνωρίζουμε πόσες, ποιες και αν υπάρχουν ομάδες.

- **Επιβλεπόμενη Μάθηση (Supervised Learning)**

Η πλειοψηφία των προβλημάτων μηχανικής εκμάθησης επιλύονται με επιβλεπόμενη μάθηση. Η επιβλεπόμενη μάθηση είναι όταν έχουμε μεταβλητές εισόδου (x) και μια μεταβλητή εξόδου (Y) και κάνουμε χρήση ενός αλγορίθμου για να μάθουμε τη συνάρτηση αντιστοίχισης από την είσοδο στην έξοδο, $Y = f(X)$. Ο στόχος είναι να προσεγγίσουμε τη συνάρτηση αντιστοίχισης τόσο καλά ώστε για νέα δεδομένα εισόδου (x) να προβλεφθούν σωστά οι μεταβλητές εξόδου (Y) για αυτά τα δεδομένα.

Ονομάζεται επιβλεπόμενη μάθηση επειδή η διαδικασία ενός αλγορίθμου που μαθαίνει από το σύνολο δεδομένων εκπαίδευσης μπορεί να θεωρηθεί ως ένας "δάσκαλος" που επιβλέπει τη διαδικασία μάθησης. Γνωρίζουμε τις σωστές απαντήσεις, ο αλγόριθμος κάνει επαναληπτικά προβλέψεις για τα δεδομένα εκπαίδευσης και διορθώνεται από τον δάσκαλο. Η εκμάθηση σταματά όταν ο αλγόριθμος επιτύχει ένα αποδεκτό επίπεδο απόδοσης.

- **Μη Επιβλεπόμενη Μάθηση (Unsupervised Learning)**

Η μη επιβλεπόμενη μάθηση είναι όταν υπάρχουν μόνο δεδομένα εισόδου (X) και καμία αντίστοιχη μεταβλητή εξόδου. Ο στόχος της μάθησης χωρίς επίβλεψη είναι να μοντελοποιήσει την υποκείμενη δομή ή κατανομή στα δεδομένα προκειμένου να μάθουμε περισσότερα για τα δεδομένα. Ονομάζεται μη επιβλεπόμενη μάθηση, επειδή σε αντίθεση με την επιβλεπόμενη μάθηση παραπάνω δεν υπάρχουν σωστές προβλέψεις και δεν υπάρχει "δάσκαλος". Οι αλγόριθμοι αφήνονται στις δικές τους επινοήσεις για να ανακαλύψουν και να παρουσιάσουν την ενδιαφέρουσα δομή των δεδομένων. Τα προβλήματα μάθησης χωρίς επίβλεψη μπορούν να ομαδοποιηθούν περαιτέρω σε προβλήματα ομαδοποίησης και συσχέτισης.

- Ημιεπιβλεπόμενη Μάθηση (Semisupervised Learning)

Τα προβλήματα στα οποία ο όγκος δεδομένων εισόδου (X) είναι μεγάλος και μόνο μερικά από τα δεδομένα φέρουν την ένδειξη (Y) ονομάζονται προβλήματα μάθησης ημιεπιβλεπόμενα. Αυτά τα προβλήματα βρίσκονται μεταξύ τόσο της εποπτευόμενης όσο και της μη εποπτευόμενης μάθησης. Ένα καλό παράδειγμα είναι ένα αρχείο φωτογραφιών όπου μόνο ορισμένες από τις εικόνες φέρουν ετικέτα, (π.χ. σκύλος, γάτα, άτομο) και οι περισσότερες είναι χωρίς ετικέτα. Πολλά προβλήματα μηχανικής μάθησης του πραγματικού κόσμου εμπίπτουν σε αυτόν τον τομέα. Αυτό οφείλεται στο γεγονός ότι μπορεί να είναι δαπανηρή ή χρονοβόρα η επισήμανση δεδομένων. Ενώ τα δεδομένα χωρίς ετικέτα είναι φθηνά και εύκολα συλλέγονται και αποθηκεύονται. Μπορεί να χρησιμοποιηθούν τεχνικές εκμάθησης χωρίς επίβλεψη για να ανακαλύψουμε και να μάθουμε τη δομή στις μεταβλητές εισόδου. Μπορούμε επίσης να χρησιμοποιήσουμε τεχνικές εποπτευόμενης εκμάθησης για να κάνουμε τις βέλτιστες προβλέψεις για τα δεδομένα χωρίς ετικέτα, να τροφοδοτήσουμε αυτά τα δεδομένα στον αλγόριθμο εποπτευόμενης μάθησης ως δεδομένα εκπαίδευσης και να χρησιμοποιήσουμε το μοντέλο για να κάνουμε προβλέψεις για καινούργια δεδομένα.

- Ενισχυτική Μάθηση (Reinforcement learning)

Η ενισχυτική μάθηση είναι ένας τομέας της Μηχανικής Μάθησης. Πρόκειται για τη λήψη κατάλληλων ενεργειών για τη μεγιστοποίηση της ανταμοιβής σε μια συγκεκριμένη κατάσταση. Χρησιμοποιείται από διάφορα λογισμικά και μηχανήματα για να βρει την καλύτερη δυνατή συμπεριφορά ή διαδρομή που πρέπει να ακολουθήσει σε μια συγκεκριμένη κατάσταση. Η ενισχυτική μάθηση διαφέρει από την εποπτευόμενη μάθηση με τρόπο που στην εποπτευόμενη μάθηση τα δεδομένα εκπαίδευσης έχουν μαζί τους το κλειδί απάντησης, έτσι το μοντέλο εκπαιδεύεται με τη σωστή απάντηση, ενώ στην ενισχυτική μάθηση, δεν υπάρχει απάντηση, αλλά ο ενισχυτικός παράγοντας αποφασίζει αν θα εκτελέσει τη δεδομένη εργασία. Ελλείψει συνόλου δεδομένων εκπαίδευσης, είναι βέβαιο ότι θα διδαχθεί από την εμπειρία του.

3.1.3 Είδη προβλημάτων μάθησης

Μερικά από τα πιο συνήθη προβλήματα μηχανικής μάθησης που έχουν μελετηθεί και συνεχίζουν και μελετώνται είναι τα παρακάτω:

- Ταξινόμηση (Classification)

Η ταξινόμηση είναι μια διαδικασία εύρεσης μιας συνάρτησης που βοηθά στη διαίρεση του συνόλου δεδομένων σε κλάσεις με βάση διαφορετικές παραμέτρους. Στην ταξινόμηση, ένα πρόγραμμα υπολογιστή εκπαιδεύεται στο σύνολο δεδομένων (training set) και βάσει αυτής της εκπαίδευσης, κατηγοριοποιεί τα δεδομένα σε διαφορετικές κατηγορίες. Η αποστολή του αλγορίθμου ταξινόμησης είναι να βρει τη συνάρτηση αντιστοίχισης για να αντιστοιχίσει την είσοδο (x) στη διακριτή έξοδο (y).

- Παλινδρόμηση (Regression)

Η παλινδρόμηση είναι μια διαδικασία εύρεσης των συσχετίσεων μεταξύ εξαρτημένων

και ανεξάρτητων μεταβλητών. Βοηθά στην πρόβλεψη των συνεχών μεταβλητών όπως η πρόβλεψη των τάσεων της αγοράς, η πρόβλεψη των τιμών των κατοικιών κ.λπ. Η αποστολή του αλγόριθμου παλινδρόμησης είναι να βρει τη συνάρτηση αντιστοίχισης για να αντιστοιχίσει τη μεταβλητή εισόδου(x) στη συνεχή μεταβλητή εξόδου (y).

- Συσταδοποίηση (Clustering)

Η συσταδοποίηση είναι το έργο της διαίρεσης του πληθυσμού ή των σημείων δεδομένων σε έναν αριθμό ομάδων, έτσι ώστε τα σημεία δεδομένων στις ίδιες ομάδες να είναι ομοιογενή με άλλα σημεία δεδομένων στην ίδια ομάδα και ανόμοια με τα σημεία δεδομένων σε άλλες ομάδες. Είναι βασικά μια συλλογή αντικειμένων με βάση την ομοιότητα και την ανομοιότητα μεταξύ τους.

- Μείωση της διαστατικότητας (Dimensionality reduction)

Όπως υποδηλώνει το όνομα, χρησιμοποιούμε μείωση διαστάσεων για να αφαιρέσουμε τις λιγότερο σημαντικές πληροφορίες (μερικές φορές περιττές στήλες) από ένα σύνολο δεδομένων. Στην πράξη, συχνά υπάρχουν σύνολα δεδομένων με εκατοντάδες ή και χιλιάδες στήλες (ονομάζονται επίσης χαρακτηριστικά), επομένως η μείωση του συνολικού αριθμού είναι ζωτικής σημασίας. Η πιο δημοφιλής μέθοδος μείωσης διαστάσεων είναι η Principal Component Analysis (PCA), η οποία μειώνει τη διάσταση του χώρου χαρακτηριστικών με την εύρεση νέων διανυσμάτων που μεγιστοποιούν τη γραμμική διακύμανση των δεδομένων. Το PCA μπορεί να μειώσει τη διάσταση των δεδομένων δραματικά και χωρίς να χάσει πάρα πολλές πληροφορίες όταν οι γραμμικές συσχετίσεις των δεδομένων είναι ισχυρές.[23]

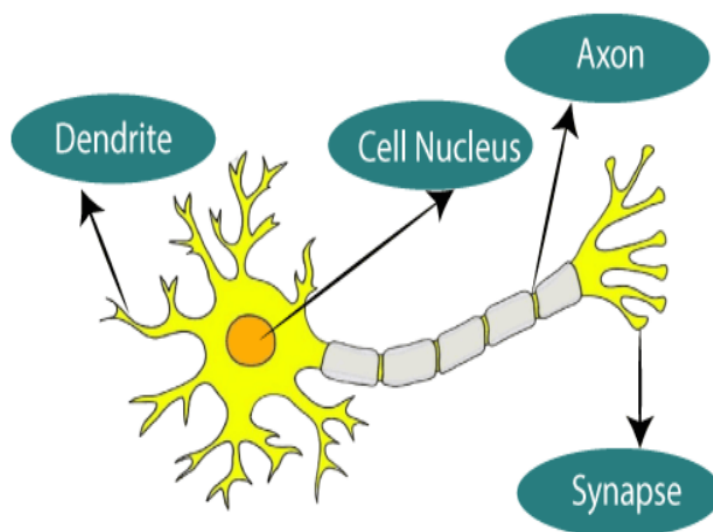
3.2 Τεχνητά Νευρωνικά Δίκτυα (Artificial Neural Networks-ANN)

Ο όρος «Τεχνητά Νευρωνικά Δίκτυα» προέρχεται από τα βιολογικά νευρωνικά δίκτυα που αναπτύσσουν τη δομή του ανθρώπινου εγκεφάλου. Παρόμοια με τον ανθρώπινο εγκέφαλο που έχει νευρώνες διασυνδεδεμένους μεταξύ τους, τα τεχνητά νευρωνικά δίκτυα έχουν επίσης νευρώνες που είναι διασυνδεδεμένοι μεταξύ τους σε διάφορα στρώματα των δικτύων. Αυτοί οι νευρώνες είναι γνωστοί ως κόμβοι. Το Σχήμα 3.1 που δίνεται απεικονίζει το τυπικό διάγραμμα του Βιολογικού Νευρωνικού Δικτύου.[24]

Το τυπικό τεχνητό νευρωνικό δίκτυο μοιάζει με την παρακάτω εικόνα του Σχήματος 3.2.

Οι δενδρίτες από το βιολογικό νευρωνικό δίκτυο αντιπροσωπεύουν τις εισόδους στα τεχνητά νευρωνικά δίκτυα, ο πυρήνας του κυττάρου αντιπροσωπεύει τους κόμβους, η σύναψη αντιπροσωπεύει τα βάρη και ο άξονας αντιπροσωπεύει την έξοδο.[27] [28] [29] [30]

Η βασική μονάδα υπολογισμού σε ένα νευρωνικό δίκτυο είναι ο νευρώνας, συχνά ονομαζόμενος και ως κόμβος. Λαμβάνει εισόδους από άλλους κόμβους ή από μια εξωτερική πηγή και υπολογίζει μια έξοδο. Κάθε είσοδος πολλαπλασιάζεται με το αντίστοιχο βάρος (Weight) και υπολογίζεται το ολικό άθροισμα των γινομένων. Ο κόμβος ή νευρώνας εφαρμόζει μια συνάρτηση ενεργοποίησης (Activation function) σε αυτό το άθροισμα και έτσι προκύπτει η έξοδος του.[31] Παρακάτω στο Σχήμα 3.3 φαίνεται η αναπαράσταση ενός νευρώνα, καθώς και η εξίσωση της εξόδου.



Σχήμα 3.1: *Biological Neural Network*
[25]

Το παρακάτω δίκτυο παίρνει ως εισόδους τα x_1, x_2, \dots, x_n που έχουν για βάρη τα w_1, w_2, \dots, w_n αντίστοιχα. Επιπλέον για κάθε είσοδο υπάρχει ακόμα μια είσοδος με τιμή 1 με βάρος b η οποία ονομάζεται πόλωση (bias). Η συνάρτηση f είναι μη γραμμική και ονομάζεται συνάρτηση ενεργοποίησης (activation function). Ο σκοπός της συνάρτησης ενεργοποίησης είναι να εισάγει μη γραμμικότητα στην έξοδο ενός νευρώνα. Αυτό είναι σημαντικό καθώς σχεδόν όλα τα πραγματικά δεδομένα είναι μη γραμμικά.

$$k = f\left(\sum_{x=1}^i (w_i * x_i + b_i)\right)$$

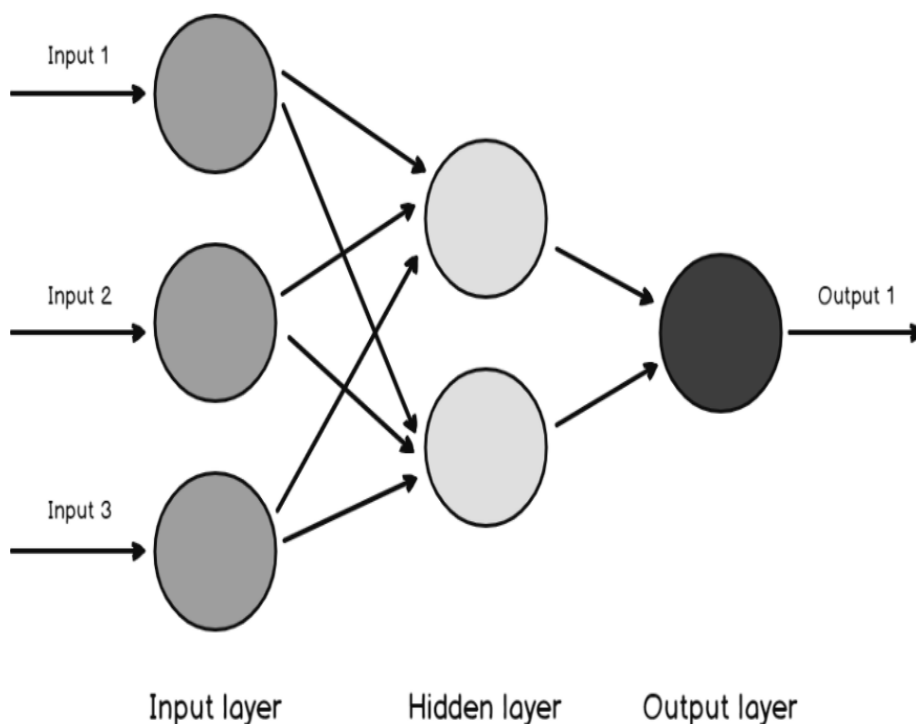
to

(3.2)

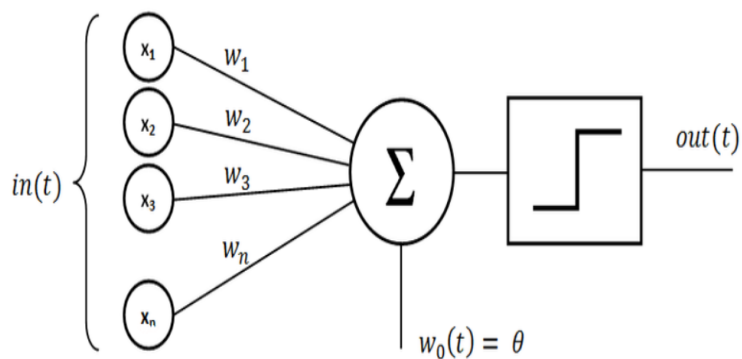
Συνάρτηση Ενεργοποίησης

Οι συναρτήσεις ενεργοποίησης καθορίζουν αν ένας νευρώνας πρέπει να ενεργοποιηθεί ή όχι και είναι τμηματικά παραγωγίσιμες συναρτήσεις προκειμένου να μετατρέπουν τα σήματα εισόδου σε διακριτές τιμές εξόδου. Παραδείγματα τέτοιων συναρτήσεων είναι η σιγμοειδής συνάρτηση, η υπερβολική εφαπτομένη και η ReLU (Rectified Linear Unit) όπως φαίνονται και παρακάτω στο Σχήμα 3.4

- Σιγμοειδής (sigmoid) συνάρτηση είναι εκείνη που περιορίζει την είσοδό της στο διάστημα $[0, 1]$ και έτσι αποτελεί χρήσιμη επιλογή σε περιπτώσεις που επιθυμούμε για παράδειγμα το δίκτυο να υπολογίζει πιθανότητες
- Υπερβολική εφαπτομένη (tanh) είναι η συνάρτηση, η οποία επίσης περιορίζει τις εισόδους της στο διάστημα $[-1, 1]$.
- διορθωτική γραμμική μονάδα (rectified linear unit - ReLU) που ορίζεται ως η μέγιστη



Σχήμα 3.2: *Artificial Neural Network* [26]



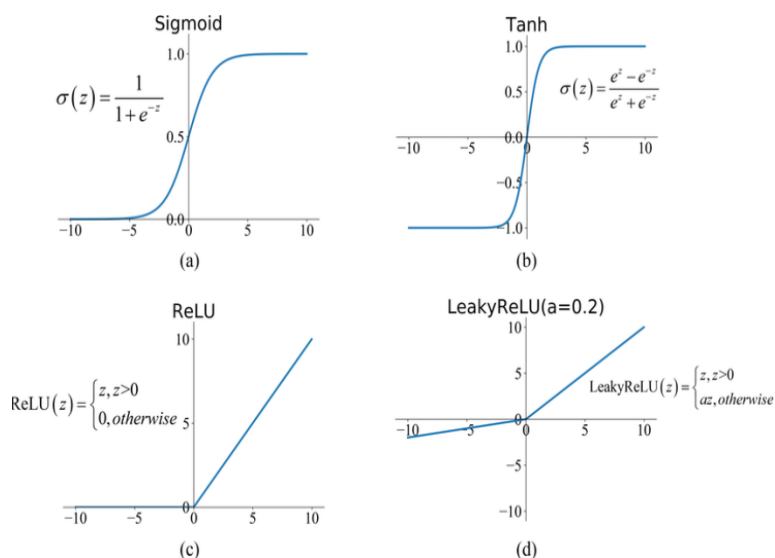
Σχήμα 3.3: *Δομή Perceptron* [32]

τιμή μεταξύ της εισόδου x και του 0.[33]

3.3 Επαναληπτικά Νευρωνικά Δίκτυα

3.3.1 Απλά Επαναληπτικά Νευρωνικά Δίκτυα

Τα δίκτυα αυτά αποτελούν και τη βάση της εργασίας γύρω από τα οποία πραγματοποιήθηκαν τα διάφορα πειράματα. Ένα επαναλαμβανόμενο νευρωνικό δίκτυο (RNN) είναι ένας ειδικός τύπος τεχνητού νευρωνικού δικτύου προσαρμοσμένο να λειτουργεί για δεδομένα χρονοσειρών ή δεδομένα που περιλαμβάνουν ακολουθίες. Τα RNN έχουν την έννοια



Σχήμα 3.4: Συναρτήσεις Ενεργοποίησης [34]

της «μνήμης» που τους βοηθά να αποθηκεύουν τις καταστάσεις ή τις πληροφορίες προηγούμενων εισόδων για να δημιουργήσουν την επόμενη έξοδο της ακολουθίας.

Για την σωστή κατανόηση των RNN, απαραίτητη είναι η γνώση των «κανονικών» νευρωνικών δικτύων προώθησης τροφοδοσίας και των διαδοχικών δεδομένων.[29] Τα διαδοχικά δεδομένα είναι βασικά απλώς ταξινομημένα δεδομένα στα οποία τα σχετικά γεγονότα διαδέχονται το ένα το άλλο. Παραδείγματα είναι τα οικονομικά δεδομένα ή η αλληλουχία DNA. Ο πιο δημοφιλής τύπος διαδοχικών δεδομένων είναι ίσως τα δεδομένα χρονοσειρών, τα οποία είναι απλώς μια σειρά σημείων δεδομένων που παρατίθενται με χρονική σειρά.[35]

Για να αναλύσουμε την αρχιτεκτονική των RNNs θα χρησιμοποιήσουμε ένα παράδειγμα γλωσσικού μοντέλου. Για ένα τέτοιο μοντέλο αντιστοιχίζουμε τα κειμενικά δεδομένα σε tokens, όπου αυτά τα tokens μπορούν να θεωρηθούν ως μια ακολουθία διακριτών παρατηρήσεων. Σε ένα τέτοιο μοντέλο ο στόχος είναι ο υπολογισμός της από κοινού πιθανότητας $P(x_1, x_2, \dots, x_T)$, όπου x_1, \dots, x_T οι ακολουθίες κειμένου (συνήθως ονομάζονται tokens ή characters) και T το συνολικό μήκος κειμένου. Η πιθανότητα εμφάνισης μίας ακολουθίας κειμένου x_t τη δεδομένη χρονική στιγμή $t \in (1, T)$ είναι $P(x_t|x_1, \dots, x_{t-1})$.

Ως εκ τούτου, θα ήταν πρότιμότερο να χρησιμοποιήσουμε ένα ελαφρώς λανθάνον μοντέλο που θα κάνει την εξής προσέγγιση:

$$P(x_t|x_1, \dots, x_{t-1}) \approx P(x_t|h_{t-1}). \quad (3.3)$$

Το h_{t-1} είναι η κρυφή κατάσταση και αποθηκεύει τις πληροφορίες της ακολουθίας έως το χρονικό βήμα $t-1$. Στη γενική περίπτωση, το Hidden State σε κάθε χρονικό βήμα t μπορεί να υπολογιστεί με βάση τόσο την τρέχουσα είσοδο x_t όσο και από την προηγούμενη κρυφή κατάσταση, h_{t-1} ως:

$$\mathbf{h}_t = f(x_t, h_{t-1}) \quad (3.4)$$

,όπου f μία καταλλήλως ορισμένη συνάρτηση. Η κρυφή κατάσταση h_t είναι αυτή που εμπεριέχει όλη την πληροφορία που έχουμε δει έως τη χρονική στιγμή t .

Ας υποθέσουμε, ότι έχουμε μια δέσμη της εισόδου, $X_t \in R^{n+d}$ τη χρονική στιγμή t . Με άλλα λόγια, για μια δέσμη εισόδου που εμπεριέχει n σε πλήθος παραδείγματα ανά είσοδο, κάθε γράμμη του X_t αντιστοιχεί σε ένα τέτοιο παράδειγμα για το συγκεκριμένο χρονικό βήμα t . Ακόμη, με $H_t \in R^{n+d}$ σημειώνουμε την κρυφή κατάσταση κατά το χρονικό βήμα t . Σε αντίθεση με το MLP, εδώ έχουμε αποθηκεύσει τη κρυφή κατάσταση H_{t-1} από το προηγούμενο χρονικό βήμα και για αυτό το λόγο εισάγουμε μία νέα παράμετρο βάρους $W_{hh} \in R^{h \times h}$ για να περιγράψουμε το πως θα χρησιμοποιήσουμε αυτή τη κρυφή κατάσταση του προηγούμενου χρονικού βήματος στο τρέχον βήμα. Συγκεκριμένα, ο υπολογισμός της κρυφής κατάστασης του τρέχοντος χρονικού βήματος καθορίζεται τόσο από την είσοδο του τρέχοντος χρονικού βήματος όσο και από την κρυφή κατάσταση του προηγούμενου χρονικού βήματος ως εξής:

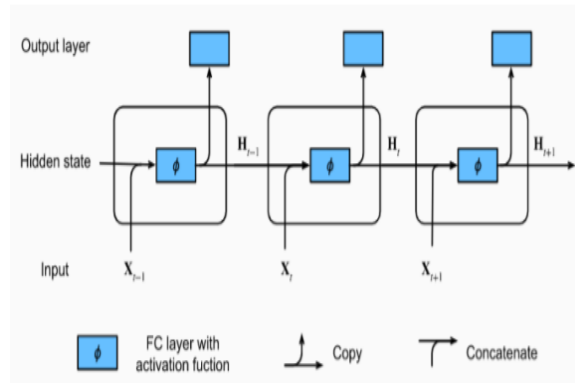
$$H_t = (X_t * W_{xh} + H_{t-1} * W_{hh} + b_h) \quad (3.5)$$

Συγκριτικά με την εξίσωση ενός Hidden Layer παρατηρούμε ότι υπάρχει ένας επιπλέον όρος, ο $H_{t-1} * W_{hh}$. Ο όρος αυτός συνδέει τις κρυφές καταστάσεις H_t και H_{t-1} μεταξύ γειτονικών χρονικών βημάτων. Άρα παρατηρούμε ότι η ιστορικότητα της πληροφορίας των ακολουθιών εισόδου διατηρείται σε ικανοποιητικό επίπεδο έως το τρέχον βήμα. Αυτός είναι και ο λόγος που η μεταβλητή H ονομάζεται κρυφή κατάσταση και τα νευρωνικά δίκτυα και συνεπώς με τέτοιες κρυφές καταστάσεις ονομάζονται Αναδρομικά Νευρωνικά Δίκτυα. Τα δίκτυα που η κρυφή τους κατάσταση ορίζεται από την σχέση 3.7 είναι τα Απλά Νευρωνικά Δίκτυα. Για κάθε χρονικό βήμα t , η έξοδος του RNN υπολογίζεται με παρόμοιο τρόπο με του MLP από την σχέση :

$$O_t = H_t * W_{hq} + b_q \quad (3.6)$$

Στη Σχέση (3.5) οι παράμετροι συμπεριλαμβάνουν τα βάρη $W_{xh} \in R^{d \times h}$ και $W_{hh} \in R^{h \times h}$ και τον όρο bias $b_h \in R^{1 \times h}$ για το επίπεδο κρυφής κατάστασης ενώ στη Σχέση (3.6) οι παράμετροι περιλαμβάνουν τα βάρη $W_{hq} \in R^{h \times q}$, και τον όρο bias $b_q \in R^{1 \times q}$ για το εξωτερικό επίπεδο. Ένα πλεονέκτημα των RNN είναι ότι χρησιμοποιούν πάντα αυτές τις παραμέτρους πράγμα που σημαίνει ότι το κόστος παραμετροποίησης ενός RNN δεν αυξάνεται με την αύξηση του αριθμού των χρονικών βημάτων.

Στην εικόνα 3.5 απεικονίζεται ένα RNN σε τρία συνεχόμενα χρονικά βήματα. Σε οποιαδήποτε χρονική στιγμή t η έξοδος του Hidden State μπορεί να υπολογιστεί ως εξής: Συνετώνοντας (concatenation) την είσοδο X_t της τρέχουσας χρονικής στιγμής t και την κρυφή κατάσταση H_{t-1} του προηγούμενου χρονικού βήματος $t-1$ και στη συνέχεια τροφοδοτώντας το αποτέλεσμα της συνένωσης σε ένα πλήρως συνδεδεμένο επίπεδο (fully-connected layer) με συνάρτηση ενεργοποίησης ϕ . Η έξοδος ενός τέτοιου πλήρους συνδεδεμένου επιπέδου είναι η κρυφή κατάσταση H_t του τρέχοντος χρονικού βήματος t . Η κρυφή κατάσταση του τρέχοντος βήματος t , H_t θα συμμετάσχει στον υπολογισμό της κρυφής κατάστασης H_{t+1} του επόμενου χρονικού βήματος $t+1$. Επιπλέον, θα τροφοδοτηθεί στο πλήρως συνδεδεμένο στρώμα εξόδου για να υπολογιστεί η έξοδος O_t για τη χρονική στιγμή t . [36] [37]



Σχήμα 3.5: Απλό Αναδρομικό Νευρωνικό Δίκτυο σε τρία χρονικά βήματα[2]

3.3.2 Back Propagation Through Time-(BPTT)

Ο τρόπος εκπαίδευσης των επαναληπτικών νευρωνικών δικτύων γίνεται μέσω του αλγορίθμου Back Propagation through time. Είναι μία ειδική κατηγορία του αλγορίθμου Back Propagation που αποτελεί τον αλγόριθμο εκπαίδευσης των τεχνητών νευρωνικών δικτύων όπως το MLP. Στην παράγραφο αυτή παρουσιάζεται ο αλγόριθμος εκπαίδευσης για τα επαναληπτικά νευρωνικά δίκτυα [38].

Θεωρούμε ένα RNN χωρίς παραμέτρους μεροληψίας(bias), του οποίου η συνάρτηση ενεργοποίησης στο κρυφό επίπεδο είναι η ταυτοτική ($f(x) = x$). Για το χρονικό βήμα t , ας είναι η είσοδος μεμονωμένου παραδείγματος και η ετικέτα $X_t \in R_d$ και y_t , αντίστοιχα. Η κρυφή κατάσταση $h_t \in R_h$ και η έξοδος $o_t \in R_q$ υπολογίζονται ως:

$$\mathbf{h}_t = X_t * W_{hx} + h_{t-1} * W_{hh} \quad (3.7)$$

και

$$O_t = h_t * W_{qh} \quad (3.8)$$

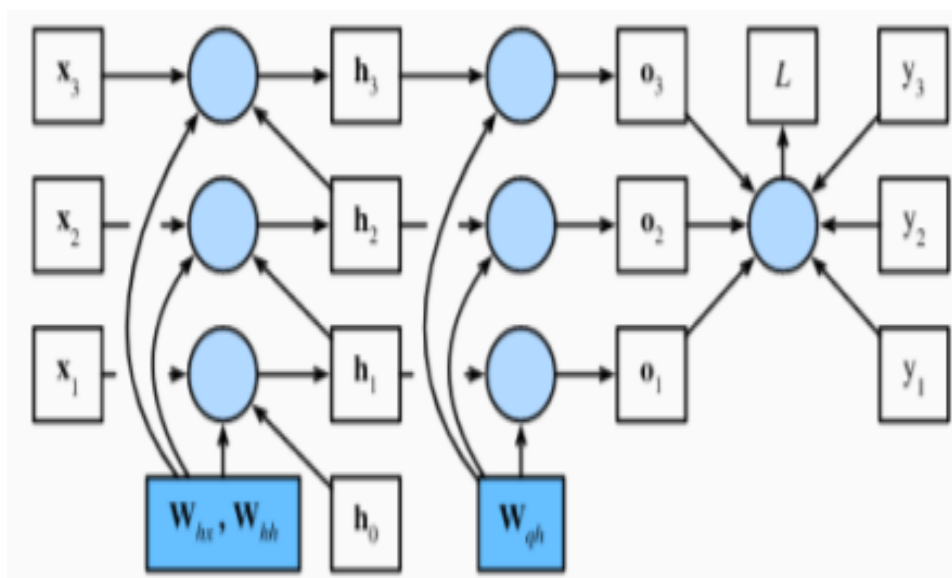
,όπου $W_{hx} \in R^{h \times d}$, $W_{hh} \in R^{h \times h}$ και $W_{qh} \in R^{q \times h}$.

Συμβολίζουμε με $l(o_t, y_t)$ την απώλεια (loss) στο χρονικό βήμα t , δηλαδή τη διαφορά της εξόδου o_t και της ετικέτας y_t . Συνεπώς, η αντικειμενική συνάρτηση απωλειών (loss function) για συνολικά T χρονικά βήματα από την αρχή της ακολουθίας εισόδου είναι η ακόλουθη:

$$L = \frac{1}{T} \sum_{t=1}^T l(o_t, y_t) \quad (3.9)$$

Προκειμένου να απεικονιστεί η εξάρτηση μεταξύ των μεταβλητών μοντέλου και των παραμέτρων κατά τον υπολογισμό του RNN, μπορούμε να σχεδιάσουμε ένα υπολογιστικό γράφημα για το μοντέλο, όπως φαίνεται στο σχήμα ;; διακρίνεται ξεκάθαρα ότι ο υπολογισμός της κρυφής κατάστασης στο τρίτο χρονικό βήμα, h_3 , εξαρτάται από τις παραμέτρους των πρώτων W_{hx} , W_{hh} , την κρυφή κατάσταση του προηγούμενου χρονικού βήματος h_2 και την τρέχουσα είσοδο x_3 . Σημειώνουμε ότι οι παράμετροι του μοντέλου είναι W_{hx} , W_{hh} και W_{qh} . Γενικά, η εκπαίδευση αυτού του μοντέλου απαιτεί τον υπολογισμό της κλίσης της συνάρτησης απωλειών ως προς αυτές τις παραμέτρους, δηλαδή $\frac{\partial L}{\partial W_{hx}}$, $\frac{\partial L}{\partial W_{hh}}$ και $\frac{\partial L}{\partial W_{qh}}$. Οι εξαρτήσεις μεταξύ των

μεγεθών μπορούν να ρεθούν διασχίζοντας το γράφημα του σχήματος 3.6 προς την αντίθετη κατεύθυνση των βελών, αποθηκεύοντας τις εκάστοτε παραγώγους.



Σχήμα 3.6: Υπολογιστικό γράφημα που αναδεικνύει τις εξαρτήσεις για ένα μοντέλο RNN σε τρία χρονικά βήματα. Τα μη χρωματιστά πλαίσια αναπαριστούν τις μεταβλητές ενώ τα χρωματισμένα πλαίσια και οι κύκλοι αντιπροσωπεύουν τελεστές πράξεων

[39]

Η εκπαίδευση ενός τέτοιου μοντέλου προϋποθέτει τον υπολογισμό της κλίσης της συνάρτησης απωλειών σε σχέση με αυτές τις παραμέτρους, δηλαδή: $\frac{\partial L}{\partial W_{hx}}, \frac{\partial L}{\partial W_{hh}}$ και $\frac{\partial L}{\partial W_{gh}}$. Για να αναδειχθούν οι εξαρτήσεις μεταξύ των μεγεθών, μπορούμε να προσπελάσουμε το γράφημα της εικόνας 3.6 προς την αντίθετη κατεύθυνση των βελών, υπολογίζοντας και αποθηκεύοντας τις αντίστοιχες κλίσεις (παραγώγους). Για μεγαλύτερη ευκολία στην έκφραση του πολλαπλασιασμού πινάκων και διανυσμάτων διαφορετικών διαστάσεων στον κανόνα της αλυσίδας, εισάγουμε τον τελεστή prod. Υποθέτοντας, ότι έχουμε 2 συναρτήσεις $Y = f(X)$ και $Z = g(Y)$, όπου οι είσοδοι και οι έξοδοι X, Y, Z είναι τανυστές αυθαίρετων διαστάσεων. Κάνοντας χρήση του κανόνα της αλυσίδας, μπορούμε να υπολογίσουμε το παράγωγο του Z ως προς το X μέσω της σχέσης:

$$\frac{\partial Z}{\partial X} = \text{prod}\left(\frac{\partial Z}{\partial Y}, \frac{\partial Y}{\partial X}\right) \quad (3.10)$$

Η χρήση του τελεστή prod εμπεριέχει όλες τις απαραίτητες ενέργειες που θα πρέπει να γίνουν στους τανυστές X, Y, Z , όπως αντιστροφή και εναλλαγή θέσης των εισόδων. Με τη βοήθεια λοιπόν του τελεστή prod μπορούμε να συνεχίσουμε την ανάλυση μας. Η διαφορίση της συνάρτησης απωλειών ως προς την έξοδο του μοντέλου o_t για κάποια χρονική στιγμή t δίνεται από τη σχέση:

$$\frac{\partial L}{\partial o_t} = \frac{\partial l(o_t, y_t)}{\partial Y} \in R^q \quad (3.11)$$

Από τον κανόνα της αλυσίδας και με γνώση της σχέσης 3.11 θα υπολογίσουμε την κλίση της συνάρτησης απωλειών $\partial L / \partial W_{qh} \in R^{q \times h}$:

$$\frac{\partial L}{\partial W_{qh}} = \sum_{t=1}^T \text{prod}\left(\frac{\partial L}{\partial o_t}, \frac{\partial o_t}{\partial W_{qh}}\right) = \sum_{t=1}^T \frac{\partial L}{\partial o_t} h_t^T \quad (3.12)$$

Έπειτα παρατηρούμε την εικόνα 3.5, όπου στο τελευταίο χρονικό βήμα T η συνάρτηση απωλειών L εξαρτάται από την κρυφή κατάσταση h_T μόνο μέσω της o_T . Άρα μέσω του κανόνα της αλυσίδας θα υπολογίσουμε ξανά την κλίση $\partial L / \partial h_T \in R^h$:

$$\frac{\partial L}{\partial h_T} = \text{prod}\left(\frac{\partial L}{\partial o_T}, \frac{\partial o_T}{\partial h_T}\right) = W_{qh}^T \frac{\partial L}{\partial o_T} h_t^T \quad (3.13)$$

Σύμφωνα με τον κανόνα της αλυσίδας, η κλίση $\partial L / \partial h_t \in R^h$ για οποιοδήποτε χρονικό βήμα $t < T$ μπορεί να υπολογιστεί επαναληπτικά ως:

$$\frac{\partial L}{\partial h_t} = \text{prod}\left(\frac{\partial L}{\partial h_{t+1}}, \frac{\partial h_{t+1}}{\partial h_t}\right) + \text{prod}\left(\frac{\partial L}{\partial o_t}, \frac{\partial o_t}{\partial h_t}\right) = W_{hh}^T \frac{\partial L}{\partial h_{t+1}} + W_{qh}^T \frac{\partial L}{\partial o_t} \quad (3.14)$$

Αναλύοντας την προηγούμενη αναδρομική σχέση 3.14 για οποιοδήποτε χρονικό βήμα $t < T$ μπορούμε να πάρουμε:

$$\frac{\partial L}{\partial h_t} = \sum_{i=1}^T W_{hh}^{T-i} W_{qh}^T \frac{\partial L}{\partial o_{t+T-i}} \quad (3.15)$$

Στο σχήμα 3.15 αίνεται η εξάρτηση της συνάρτησης απωλειών από τις παραμέτρους W_{hx} και W_{hh} στο κρυφό επίπεδο, μέσω των κρυφών καταστάσεων h_1, \dots, h_T . Για τον υπολογισμό της κλίσης της L ως προς αυτές τις παραμέτρους, $\partial L / \partial W_{hx} \in R^{h \times d}$ και $\partial L / \partial W_{hh} \in R^{h \times h}$ και εφαρμόζουμε για μια ακόμη φορά τον κανόνα της αλυσίδας όπου προκύπτει:

$$\frac{\partial L}{\partial W_{hx}} = \sum_{t=1}^T \text{prod}\left(\frac{\partial L}{\partial h_t}, \frac{\partial h_t}{\partial W_{hx}}\right) = \sum_{t=1}^T \frac{\partial L}{\partial h_t} x_t^T \quad (3.16)$$

$$\frac{\partial L}{\partial W_{hh}} = \sum_{t=1}^T \text{prod}\left(\frac{\partial L}{\partial h_t}, \frac{\partial h_t}{\partial W_{hh}}\right) = \sum_{t=1}^T \frac{\partial L}{\partial h_t} h_{t-1}^T \quad (3.17)$$

,όπου η ποσότητα $\partial L / \partial h_t$ που υπολογίζεται επαναληπτικά από τις σχέσεις 3.14 και 3.15 είναι ο βασικός παράγοντας που επηρεάζει την αριθμητική σταθερότητα.

Εν κατακλείδι κι ενώ ο αλγόριθμος Back Propagation Through Time είναι αρκετά χρονοβόρος και περίπλοκος έχει ένα πλεονέκτημα. Αποθηκεύει τις ενδιάμεσες μεταβλητές του αλγορίθμου με τη σειρά υπολογισμού τους ώστε να αποφεύγονται οι αχρείαστοι επιπρόσθετοι υπολογισμοί, όπως για παράδειγμα αποθηκεύει την Δ / η_t που χρησιμοποιείται στους υπολογισμούς των 3.16, 3.17, επιταχύνοντας σημαντικά τον χρόνο εκτέλεσης του αλγορίθμου.

Μειονεκτήματα BPTT στα Αναδρομικά Επαναληπτικά Δίκτυα

Όταν χρησιμοποιούμε το BPTT (backpropagation through time) στο RNN, γενικά αντιμετωπίζουμε προβλήματα όπως η έκρηξη της κλίσης και η εξαφάνιση της κλίσης. Το σοβαρότερο πρόβλημα που ενδέχεται να προκαλέσει ο αλγόριθμος Back Propagation through time στα

RNNs είναι ότι για ιδιαίτερα υψηλές δυνάμεις πινάκων μπορούμε να οδηγηθούμε σε αποκλίνουσες ή εξαφανιζόμενες ιδιοτιμές. Αυτό σημαίνει, ότι οι κλίσεις (παράγωγοι) κατά την εκτέλεση του αλγορίθμου εκρήγνυνται σε πολύ μεγάλες τιμές ή εξαφανίζονται σε πολύ μικρές. Για το λόγο αυτό υπήρξε ανάγκη για εξεύρεση περιπλοκότερων ακολουθιακών μοντέλων που μπορούν να αντιμετωπίσουν αυτό το πρόβλημα. Δυστυχώς, αν εφαρμόσουμε τα παραπάνω βήματα, δεν θα είμαστε τόσο ευχαριστημένοι με τα αποτελέσματα. Αυτό συμβαίνει επειδή το απλούστερο μοντέλο RNN έχει ένα σημαντικό μειονέκτημα, που ονομάζεται πρόβλημα εξαφάνισης κλίσης, το οποίο το εμποδίζει να είναι ακριβές. Με λίγα λόγια, το πρόβλημα προέρχεται από το γεγονός ότι σε κάθε βήμα κατά τη διάρκεια της εκπαίδευσης χρησιμοποιούμε τα ίδια βάρη για να υπολογίσουμε το o_t . Αυτός ο πολλαπλασιασμός γίνεται επίσης κατά την αντίστροφη διάδοση. Όσο περισσότερο κινούμαστε προς τα πίσω, τόσο μεγαλύτερο γίνεται το σήμα σφάλματός μας. Αυτό σημαίνει ότι το δίκτυο αντιμετωπίζει δυσκολίες στην απομνημόνευση τιμών από μακριά στη σειρά και κάνει προβλέψεις με βάση μόνο τις πιο πρόσφατες. Για το λόγο αυτό υπήρξε ανάγκη για εξεύρεση περιπλοκότερων ακολουθιακών μοντέλων που μπορούν να αντιμετωπίσουν αυτό το πρόβλημα. Γι' αυτό θα αναλύσουμε και θα χρησιμοποιήσουμε πιο ισχυρά μοντέλα όπως το LSTM και το GRU.

3.3.3 Φραγμένα Επαναληπτικά Δίκτυα -GRU

Για να λύθει το πρόβλημα της εξαφάνισης κλίσης ενός τυπικού RNN, η GRU χρησιμοποιεί, τη λεγόμενη, πύλη ενημέρωσης(update gate) και πύλη επαναφοράς(reset gate). Βασικά, αυτά είναι δύο διανύσματα που αποφασίζουν ποιες πληροφορίες πρέπει να περάσουν στην έξοδο. Το ιδιαίτερο με αυτά είναι ότι μπορούν να εκπαιδευτούν στο να διατηρούν πληροφορίες από παλιά, χωρίς να τις "ξεπλένουν" στο χρόνο ή να αφαιρούν πληροφορίες που είναι άσχετες με την πρόβλεψη.

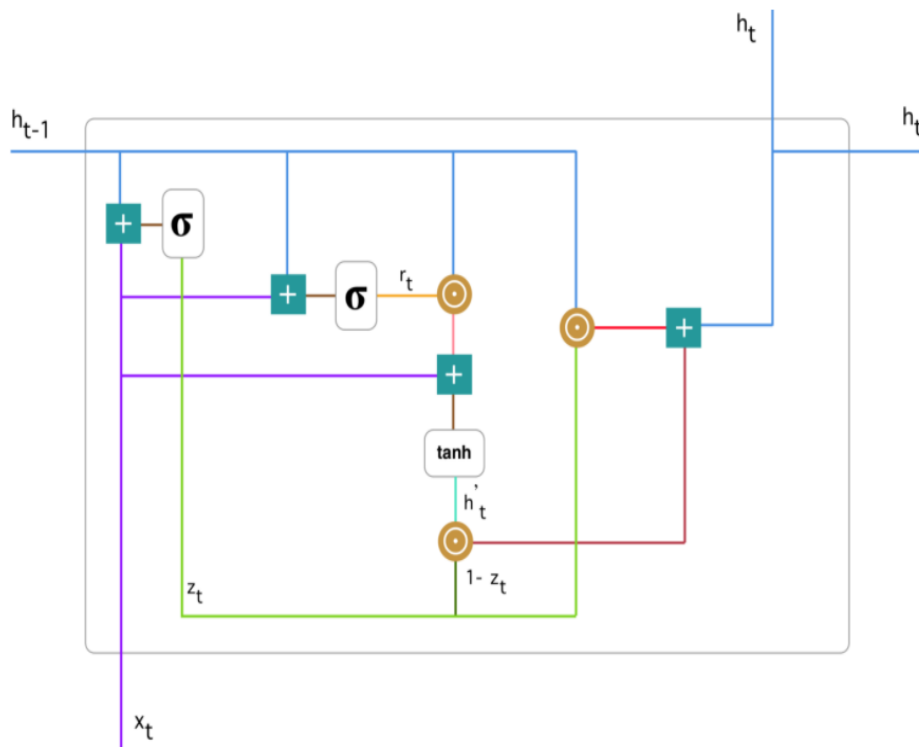
Θα αναλύσουμε τη λειτουργία του φραγμένου δικτύου την χρονική στιγμή t με βάση το παρακάτω Σχήμα 3.7:

1.Update gate

Θα αρχίσουμε με τον υπολογισμό της πύλης ενημέρωσης z_t για το χρονικό βήμα t χρησιμοποιώντας τον τύπο:

$$z_t = \sigma(W^{(z)}x_t + U^{(z)}h_{t-1}) \quad (3.18)$$

Όταν το x_t είναι συνδεδεμένο στη μονάδα δικτύου, πολλαπλασιάζεται με το βάρος του $W^{(z)}$. Το ίδιο ισχύει και για το h_{t-1} που περιέχει τις πληροφορίες για τις προηγούμενες μονάδες $t-1$ και πολλαπλασιάζεται με το βάρος του $U^{(z)}$. Και τα δύο αποτελέσματα προστίθενται μαζί και εφαρμόζεται μια συνάρτηση ενεργοποίησης (σιγμοειδής) για να συμπίψει το αποτέλεσμα μεταξύ 0 και 1. Η πύλη ενημέρωσης βοηθά το μοντέλο να προσδιορίσει πόσες από τις πληροφορίες του παρελθόντος (από προηγούμενα χρονικά βήματα) πρέπει να διαβιβαστούν στο μέλλον. Αυτό είναι πολύ σημαντικό γιατί το μοντέλο μπορεί να αποφασίσει να αντιγράψει όλες τις πληροφορίες από το παρελθόν και να εξαλείψει τον κίνδυνο εξαφάνισης κλίσης.



Σχήμα 3.7: *Gated Recurrent Unit*
[40]

2.Reset gate

Ουσιαστικά, αυτή η πύλη χρησιμοποιείται από το μοντέλο για να αποφασίσει πόσες από τις προηγούμενες πληροφορίες θα "ξεχάσουμε". Ο τύπος της πύλης επαναφοράς είναι ο ακόλουθος:

$$r_t = \sigma(W^{(r)}x_t + U^{(r)}h_{t-1}) \quad (3.19)$$

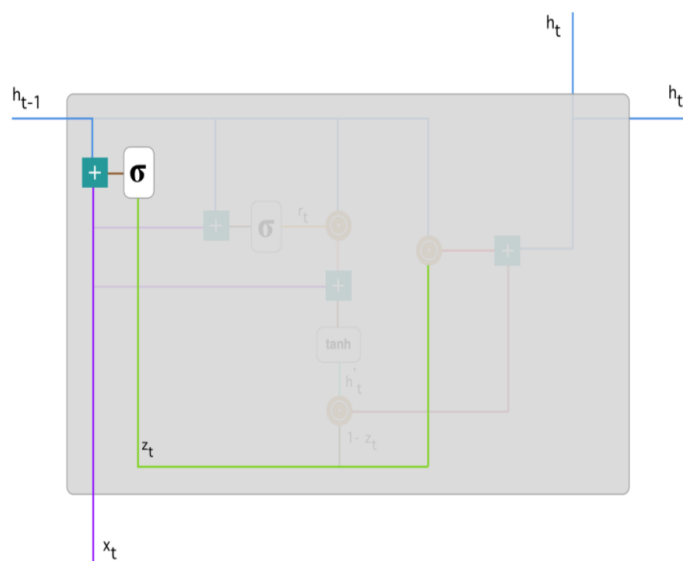
Όπως παρατηρούμε οι υπολογισμοί των πυλών επαναφοράς και ενημέρωσης μοιάζουν αρκετά. Αυτό που αλλάζει είναι τα βάρη ενημέρωσης και η μετέπειτα λειτουργία της πύλης. Όπως και πριν, πολλαπλασιάζουμε τις προηγούμενες πληροφορίες h_{t-1} και την είσοδο x_t της τρέχουσας χρονικής στιγμής με τα αντίστοιχα βάρη τους, αθροίζουμε τα αποτελέσματα και εφαρμόζουμε τη σιγμοειδή συνάρτηση.

3.Current memory content

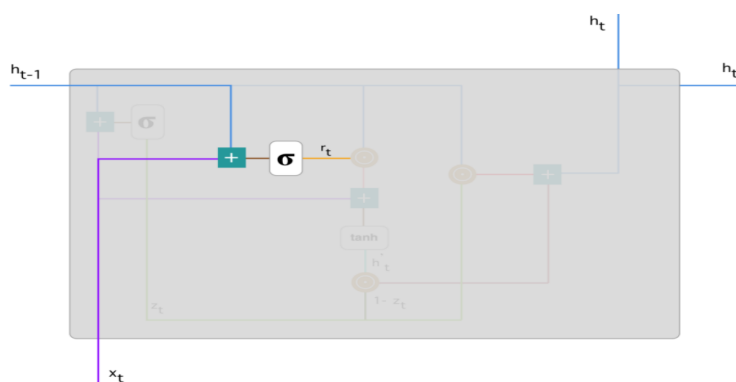
Ας δούμε πώς ακριβώς οι δύο προηγούμενες πύλες που αναλύσαμε θα επηρεάσουν την τελική έξοδο. Αρχικά, ξεκινάμε με τη χρήση της πύλης επαναφοράς. Παρουσιάζουμε ένα νέο περιεχόμενο μνήμης που θα χρησιμοποιεί την πύλη επαναφοράς για να αποθηκεύει τις σχετικές πληροφορίες από το παρελθόν. Υπολογίζεται ως εξής: σκακ

$$h'_t = \tanh(W^{(r)}x_t + r_t \odot U^{(r)}h_{t-1}) \quad (3.20)$$

Σε αυτό το βήμα πολλαπλασιάζεται η είσοδος x_t με το βάρος W -μωβ γραμμή- και h_{t-1} με



Σχήμα 3.8: *Update gate*
[40]



Σχήμα 3.9: *Reset gate*
[40]

το βάρος Υ-γαλάζια γραμμή. Έπειτα υπολογίζεται το γινόμενο Hadamard (στοιχειακά) μεταξύ της πύλης επαναφοράς r_t -πορτοκαλί γραμμή- και του γινομένου $U h_{t-1}$ -ροζ γραμμή. Αυτό θα καθορίσει τι πρέπει να αφαιρεθεί από τα προηγούμενα χρονικά βήματα. Αθροίζονται λοιπόν-καφέ γραμμή- όπως φαίνεται και στη Σχέση 3.20 τα γινόμενα που αναφέραμε παραπάνω και εφαρμόζεται η μη γραμμική συνάρτηση ενεργοποίησης \tanh και προκύπτει το h'_t -πράσινη γραμμή- σύμφωνα και με το σχήμα 3.10 .

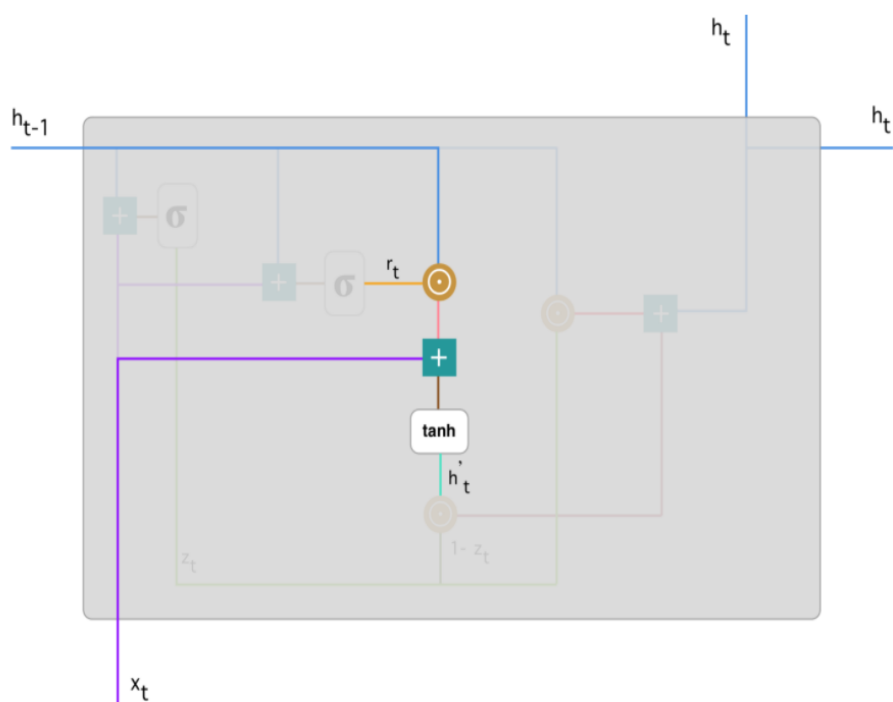
4. Final memory at current time step

Ως τελευταίο βήμα, το δίκτυο πρέπει να υπολογίσει το h_t — διάνυσμα που διατηρεί πληροφορίες για την τρέχουσα μονάδα και τις μεταβιβάζει στο δίκτυο. Για να γίνει αυτό χρειάζεται η πύλη ενημέρωσης(update gate). Καθορίζει τι θα συλλέξει από το τρέχον περιεχόμενο μνήμης — h'_t και τι από τα προηγούμενα βήματα — h_{t-1} . Αυτό γίνεται ως εξής:

σκακ

$$h_t = (z_{tt-1} + (1 - z_t) \odot h'_t) \quad (3.21)$$

Όταν η πύλη ενημέρωσης Z_t είναι κοντά στο 1, διατηρούμε απλώς την παλιά κατάσταση.



Σχήμα 3.10: *Current memory content*
[40]

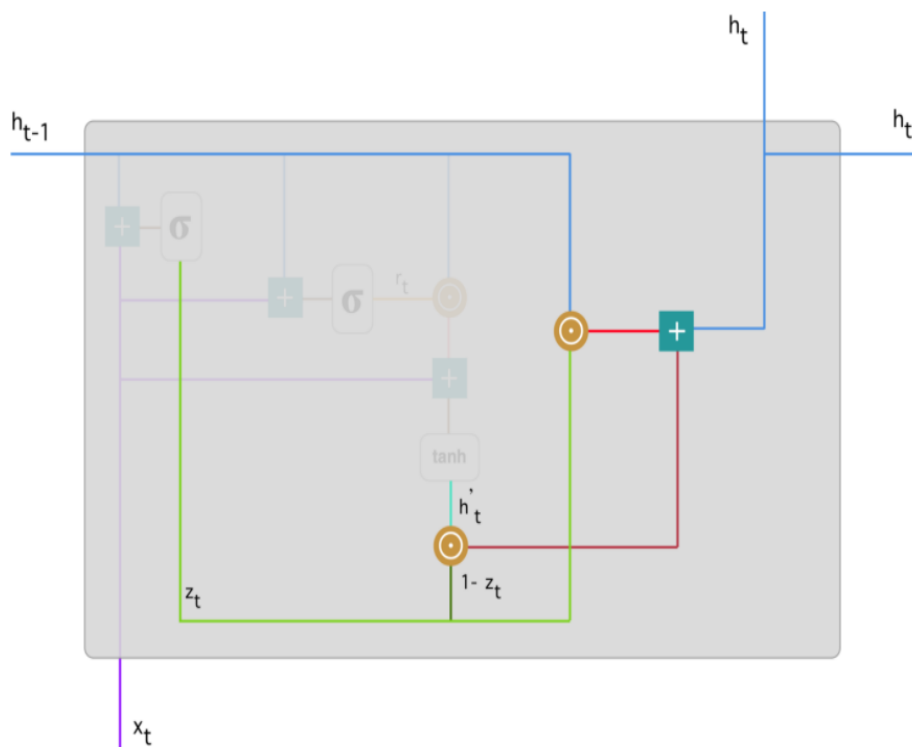
Σε αυτήν την περίπτωση, οι πληροφορίες από το X_t ουσιαστικά αγνοούνται, παρακάμπτοντας έτσι το βήμα t στην αλυσίδα εξάρτησης της πληροφορίας. Αντιθέτως, όποτε η Z_t είναι κοντά στο 0, το νέα κρυφή κατάσταση h_t πλησιάζει πιο πολύ στην κατάσταση h'_t .

Τέλος, τα δίκτυα GRU μπορούν να αποθηκεύουν και να φιλτράρουν τις πληροφορίες χρησιμοποιώντας τις πύλες ενημέρωσης και επαναφοράς τους. Αυτό εξαλείφει το πρόβλημα της εξαφάνισης κλίσης, καθώς το μοντέλο δεν κρατάει μόνο τη νέα είσοδο κάθε φορά, αλλά διατηρεί τις σχετικές πληροφορίες και τις μεταβιβάζει στα επόμενα χρονικά βήματα του δικτύου. Εάν εκπαιδευτούν προσεκτικά, μπορούν να έχουν εξαιρετικά καλή απόδοση ακόμη και σε πολύπλοκα προβλήματα. [41]

3.3.4 Δίκτυα Μακράς Βραχυπρόθεσμης Μνήμης - LSTM

Τα LSTM έχουν τη φύση να "θυμούνται" πληροφορίες για μεγάλες χρονικές περιόδους. Η πρόκληση για την αντιμετώπιση της μακροπρόθεσμης διατήρησης πληροφοριών και της βραχυπρόθεσμης παράκαμψης εισροών σε μοντέλα λανθάνουσας μεταβλητής υπάρχει εδώ και πολύ καιρό. Μία από τις πρώτες προσεγγίσεις για να αντιμετωπιστεί αυτό ήταν η μακροπρόθεσμη βραχυπρόθεσμη μνήμη (LSTM) [Hochreiter Schmidhuber, 1997]. Μοιράζεται πολλές από τις ιδιότητες της GRU. Είναι ενδιαφέρον ότι τα LSTM έχουν ελαφρώς πιο περίπλοκο σχεδιασμό από τα GRU, αλλά προηγούνται των GRU σχεδόν δύο δεκαετίες.[42]

Το LSTM εισάγει ένα νέο είδος κελιού/κατάστασης που ονομάζεται κελί μνήμης (memory cell), που σχεδιάστηκε για την καταγραφή πρόσθετων πληροφοριών το οποίο μοιάζει αρκετά με την γνωστή κρυφή κατάσταση. Για να ελέγξουμε το κελί μνήμης χρειαζόμαστε κάποιες



Σχήμα 3.11: *Hidden State*
[40]

πύλες. Απαιτείται μία πύλη, για να διαβάσει τις καταχωρήσεις από το κελί μνήμης, την οποία θα αναφέρουμε και ως Πύλη Εξόδου (Output Gate). Ακόμη, μια δεύτερη πύλη χρειάζεται για να αποφασίσουμε πότε να διαβάσουμε δεδομένα από το κελί μνήμης, το όνομα της οποίας είναι Πύλη Εισόδου (Input Gate). Τέλος, χρειαζόμαστε ένα μηχανισμό για επαναφορά του περιεχομένου του κελιού μνήμης. Τη δουλειά αυτή αναλαμβάνει η λεγόμενη Πύλη Λήθης (Forget Gate).[44]

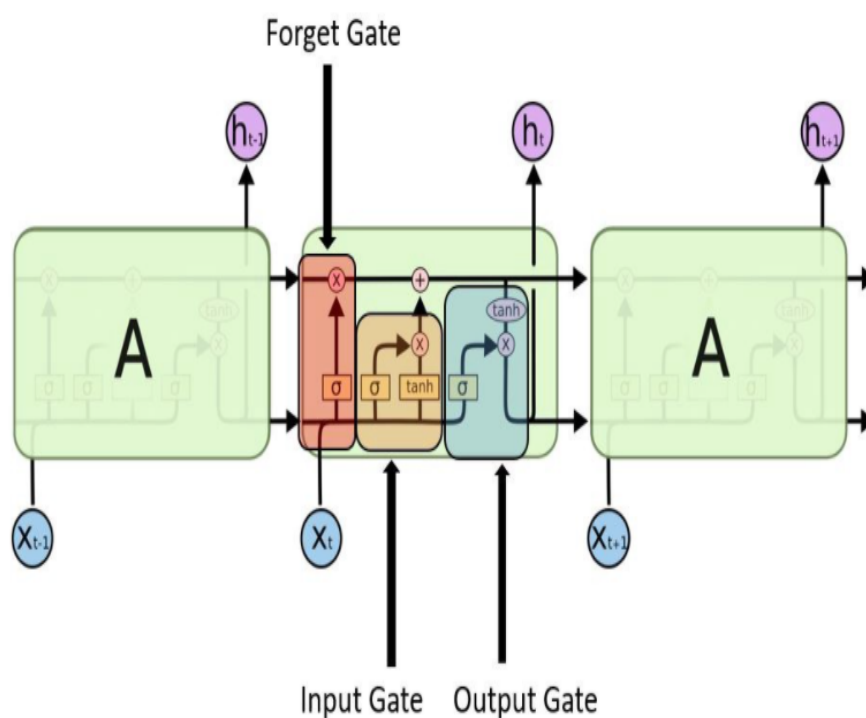
1. Forget Gate

Η συγκεκριμένη πύλη αποφασίζει πόσο από το παρελθόν πρέπει να “θυμόμαστε”. Αυτή η πύλη αποφασίζει ποιες πληροφορίες θα παραληφθούν από το κελί στη συγκεκριμένη χρονική σήμανση. Αποφασίζεται από τη σιγμοειδή συνάρτηση. κοιτάζει την προηγούμενη κατάσταση h_{t-1} και την είσοδο περιεχομένου X_t και βγάζει έναν αριθμό μεταξύ 0 (παραλείψτε αυτό) και 1 (κρατήστε αυτό) για κάθε αριθμό στην κατάσταση κελιού C_{t-1} .

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3.22)$$

2. Input Gate

Για να ενημερώσουμε την κατάσταση του κελιού (cell state), έχουμε την πύλη εισόδου. Αρχικά, περνάμε την προηγούμενη κρυφή κατάσταση (hidden state, h_{t-1}) και την τρέχουσα είσοδο x_t σε μια σιγμοειδή συνάρτηση (Σχέση 3.23). Εκεί αποφασίζεται ποιες τιμές θα ενημερωθούν μετατρέποντας τις τιμές μεταξύ 0 και 1. Τιμές 0 σημαίνει μη σημαντικές και απορρίπτονται ενώ 1 σημαίνει σημαντικές και τις κρατάμε. Επίσης περνάει η κρυφή κατάσταση



Σχήμα 3.12: Αναπαράσταση LSTM με τις αντίστοιχες πύλες [43]

της προηγούμενης χρονικής στιγμής και η είσοδος στη συνάρτηση \tanh για να συμπιέσουν οι τιμές μεταξύ -1 και 1 και να βοηθήσουν στη ρύθμιση του δικτύου (Σχέση(3.24) όπως φαίνεται και στο Σχήμα 3.14.

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (3.23)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (3.24)$$

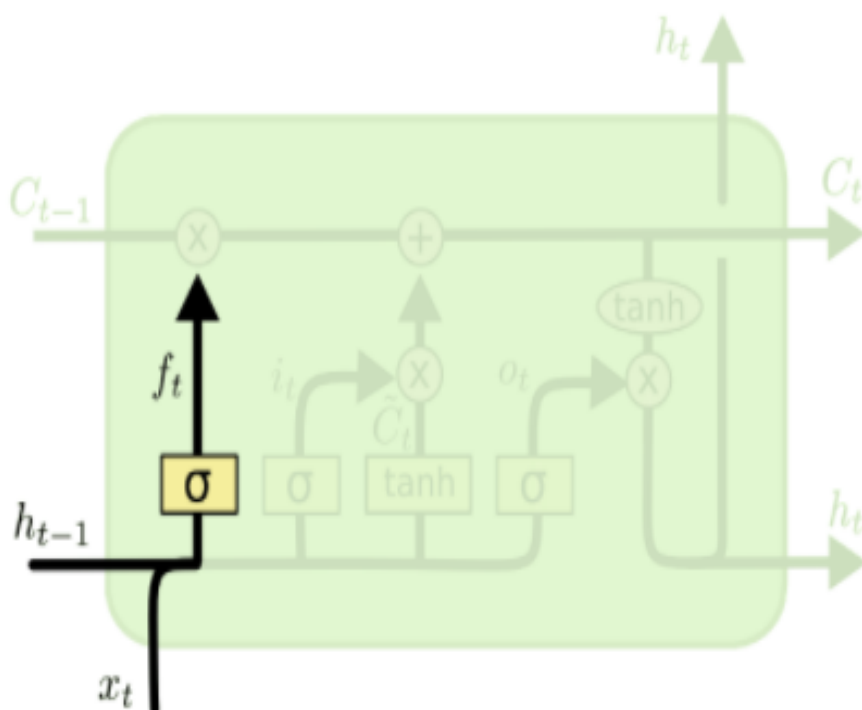
3.Cell State

Τώρα θα πρέπει να έχουμε αρκετές πληροφορίες για να υπολογίσουμε την κατάσταση του κελιού(cell state). Πρώτον, η κατάσταση του κελιού πολλαπλασιάζεται σημειακά με το διάνυσμα λήθης. Αυτό έχει τη δυνατότητα απόρριψης τιμών στην κατάσταση του κελιού εάν πολλαπλασιαστεί με τιμές κοντά στο 0. Στη συνέχεια παίρνουμε την έξοδο από την πύλη εισόδου και κάνουμε μια σημειακή προσθήκη που ενημερώνει την κατάσταση του κελιού σε νέες τιμές που το νευρωνικό δίκτυο θεωρεί σχετικές. Αυτό μας δίνει τη νέα κατάσταση κελιού 3.25.

$$C_t = F_t \odot C_{t-1} + I_t \odot \tilde{C}_t \quad (3.25)$$

4.Output Gate

Στην τελευταία αυτή πύλη η είσοδος και η κρυφή κατάσταση της αμέσως προηγούμενης τιμής περνάνε από μία σιγμοειδή συνάρτηση και έπειτα η κατάσταση του κελιού περνάει από



Σχήμα 3.13: Σχηματική Αναπαράσταση forget gate [43]

την tanh που δίνει βαρύτητα στις τιμές που περνούν καθορίζοντας το επίπεδο σπουδαιότητάς τους που κυμαίνεται από -1 έως 1 και πολλαπλασιάζεται με την έξοδο του Sigmoid. Τέλος, το καινούριο κελί μνήμης C_t και η κρυφή κατάσταση h_t προωθούνται στο επόμενο χρονικό βήμα[45].

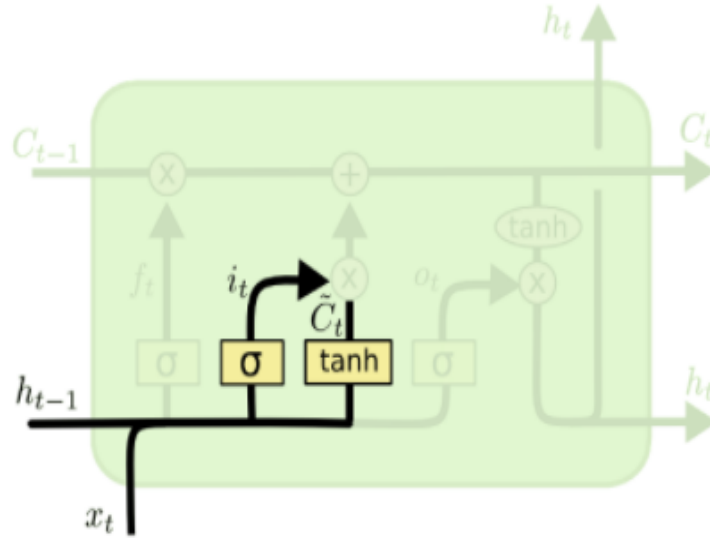
$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (3.26)$$

$$h_t = o_t * \tanh(C_t) \quad (3.27)$$

Συνεπώς, η πύλη εισόδου I_t είναι αυτή που καθορίζει το πόσο λαμβάνουμε υπόψη μας τα νέα δεδομένα μέσω του υποψήφιου κελιού, ενώ η πύλη λήθης F_t διευθετεί την ποσότητα που θα διατηρήσουμε από το παλιό κελί μνήμης (της προηγούμενης χρονικής στιγμής) C_{t-1} . Με απλό τρόπο, τα δίκτυα LSTM έχουν κάποια εσωτερικά κελιά κατάστασης που λειτουργούν ως κελιά μακροπρόθεσμης ή βραχυπρόθεσμης μνήμης. Η έξοδος του δικτύου LSTM διαμορφώνεται από την κατάσταση αυτών των κυψελών. Αυτή είναι μια πολύ σημαντική ιδιότητα όταν χρειαζόμαστε η πρόβλεψη του νευρωνικού δικτύου να εξαρτάται από το ιστορικό πλαίσιο των εισόδων και όχι μόνο από την τελευταία είσοδο. [46]

3.3.5 Αμφίδρομα Επαναληπτικά Νευρωνικά Δίκτυα (Bidirectional RNNs)

Η μετατροπή ενός μονόδρομου RNN σε αμφίδρομο γίνεται με την προσθήκη ενός ακόμα κρυφού επιπέδου το οποίο διαχειρίζεται τα δεδομένα είσοδου με αντίθετη σειρά από ότι το



Σχήμα 3.14: Σχηματική Αναπαράσταση input/update gate [43]

πρώτο κρυφό επίπεδο. Αντί δηλαδή να τρέχουμε ένα RNN μόνο στη λειτουργία προώθησης ξεκινώντας από το πρώτο διακριτικό, ξεκινάμε ένα άλλο από το τελευταίο διακριτικό που τρέχει από πίσω προς τα εμπρός. Το κύριο πλεονέκτημα των αμφίδρομων RNNs είναι ότι γνωρίζουν την πληροφορία τόσο από το παρελθόν προς το μέλλον όσο και από το μέλλον προς το παρελθόν με αποτέλεσμα να μπορούν να κάνουν πιο ευέλικτους και ακριβείς υπολογισμούς. Το Σχήμα 3.18 απεικονίζει την αρχιτεκτονική ενός αμφίδρομου RNN με ένα μόνο κρυφό στρώμα.

Για οποιοδήποτε χρονικό βήμα t , δίνεται μια είσοδος $X_t \in R^{n \times d}$. Στην αμφίδρομη αρχιτεκτονική, υποθέτουμε ότι οι κρυφές καταστάσεις προς τα εμπρός και προς τα πίσω για αυτό το χρονικό βήμα είναι $\vec{H}_t \in R^{n \times h}$ και $\overleftarrow{H}_t \in R^{n \times h}$.

Οι ενημερώσεις κρυφής κατάστασης προς τα εμπρός και προς τα πίσω είναι οι εξής:

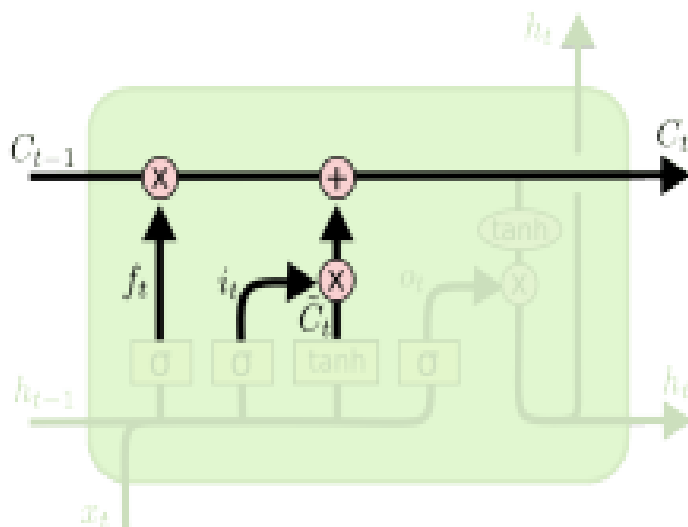
$$\vec{H}_t = f(X_t * W_{xh}^{(f)} + \vec{H}_{t-1} * W_{hh}^{(f)} + b_h^{(f)}) \quad (3.28)$$

$$\overleftarrow{H}_t = f(X_t * W_{xh}^{(b)} + \overleftarrow{H}_{t+1} * W_{hh}^{(b)} + b_h^{(b)}) \quad (3.29)$$

,όπου τα βάρη $W_{xh}^{(f)} \in R^{d \times h}$, $W_{hh}^{(f)} \in R^{h \times h}$, $W_{xh}^{(b)} \in R^{d \times h}$, $W_{hh}^{(b)} \in R^{h \times h}$ και προκαταλήψεις $b^{(f)}_h \in R^{(1h)}$ και $b^{(b)}_h \in R^{1 \times h}$ είναι όλες οι παράμετροι του μοντέλου.

Στη συνέχεια, συνενώνουμε τις εμπρός και πίσω κρυφές καταστάσεις \vec{H}_t και \overleftarrow{H}_t για να λάβουμε την κρυφή κατάσταση $H_t \in R^{n \times 2h}$ που θα τροφοδοτηθεί στο επίπεδο εξόδου. Στα βαθιά αμφίδρομα RNN με πολλαπλά κρυφά επίπεδα, τέτοιες πληροφορίες μεταβιβάζονται ως είσοδος στο επόμενο αμφίδρομο επίπεδο. Τέλος, το επίπεδο εξόδου υπολογίζει την έξοδο $O_t \in R^{n \times q}$:

$$O_t = H_t * W_{hq} + b_q \quad (3.30)$$



Σχήμα 3.15: Σχηματική Αναπαράσταση Cell State [43]

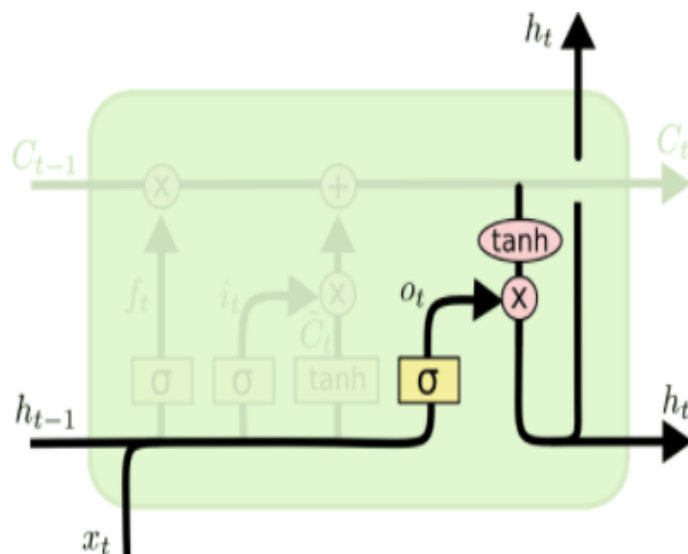
3.4 Υβριδικά Νευρωνικά Δίκτυα (Hybrid Neural Networks)

3.4.1 Εισαγωγή στα Συνελκτικά Νευρωνικά Δίκτυα (Convolutional Neural Networks)

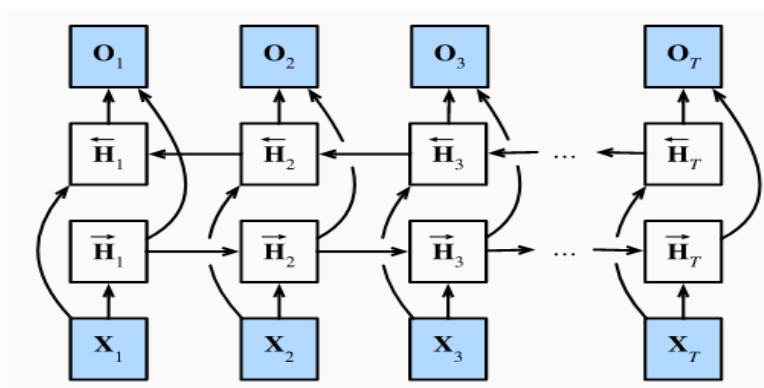
Αυτή η υποενότητα εισάγει τα συνελκτικά νευρωνικά δίκτυα (CNN), μια ισχυρή οικογένεια νευρωνικών δικτύων που έχουν σχεδιαστεί ακριβώς για αυτόν τον σκοπό. Οι αρχιτεκτονικές που βασίζονται στο CNN είναι πλέον πανταχού παρούσες στον τομέα της όρασης υπολογιστών και έχουν γίνει τόσο κυρίαρχες που σχεδόν κανείς σήμερα δεν θα ανέπτυξε μια εμπορική εφαρμογή ή θα συμμετείχε σε διαγωνισμό που σχετίζεται με την αναγνώριση εικόνας, την ανίχνευση αντικειμένων ή τη σημασιολογική τμηματοποίηση, χωρίς να βασιστεί σε αυτή την προσέγγιση.[48]

Τα σύγχρονα CNN οφείλουν τον σχεδιασμό τους σε εμπνεύσεις από τη βιολογία, τη θεωρία των ομάδων και μια υγιή δόση πειραματικών μαεστριών. Εκτός από την αποτελεσματικότητά τους ως δείγματα για την επίτευξη ακριβών μοντέλων, τα CNN τείνουν να είναι υπολογιστικά αποδοτικά, τόσο επειδή απαιτούν λιγότερες παραμέτρους από τις πλήρως συνδεδεμένες αρχιτεκτονικές όσο και επειδή οι συνελίξεις είναι εύκολο να παραλληλιστούν μεταξύ των πυρήνων GPU. Κατά συνέπεια, οι ειδικοί συχνά εφαρμόζουν CNN όποτε είναι δυνατόν, και όλο και περισσότερο αναδεικνύονται ως αξιόπιστοι ανταγωνιστές ακόμη και σε εργασίες με μονοδιάστατη δομή ακολουθίας, όπως ανάλυση ήχου, κειμένου και χρονοσειρών, όπου χρησιμοποιούνται συνελκτικά επαναλαμβανόμενα νευρωνικά δίκτυα. Ορισμένες έξυπνες προσαρμογές των CNN τα έχουν επίσης φέρει σε εφαρμογή σε δεδομένα δομημένων γραφημάτων και σε συστήματα συστάσεων.[49] [50]

Αρχικά, θα αναλύσουμε τις βασικές λειτουργίες που αποτελούν τη ραχοκοκαλιά όλων των συνελκτικών δικτύων. Αυτά περιλαμβάνουν τα ίδια τα συνελκτικά στρώματα, λεπτομέρειες όπως το padding και το stride, τα στρώματα συγκέντρωσης που χρησιμοποιούνται



Σχήμα 3.16: Σχηματική Αναπαράσταση output gate [43]



Σχήμα 3.17: Σχηματική Αναπαράσταση Αμφίδρομου RNN [47]

για τη συγκέντρωση πληροφοριών σε παρακείμενες χωρικές περιοχές, τη χρήση πολλών καναλιών σε κάθε στρώμα και μια προσεκτική συζήτηση για τη δομή των σύγχρονων αρχιτεκτονικών.[51] Θα αναφερθούμε και στα 1-dimensional CNN που αποτελούν και κομμάτι υλοποίησης στην εργασία μας.[52]

Στα μαθηματικά, η συνέλιξη μεταξύ δύο συναρτήσεων, ας υποθέσουμε $f, g: \mathbb{R}^d \rightarrow \mathbb{R}$ ορίζεται ως:

$$(f * g)(x) = \int f(z)g(x - z) dz \quad (3.31)$$

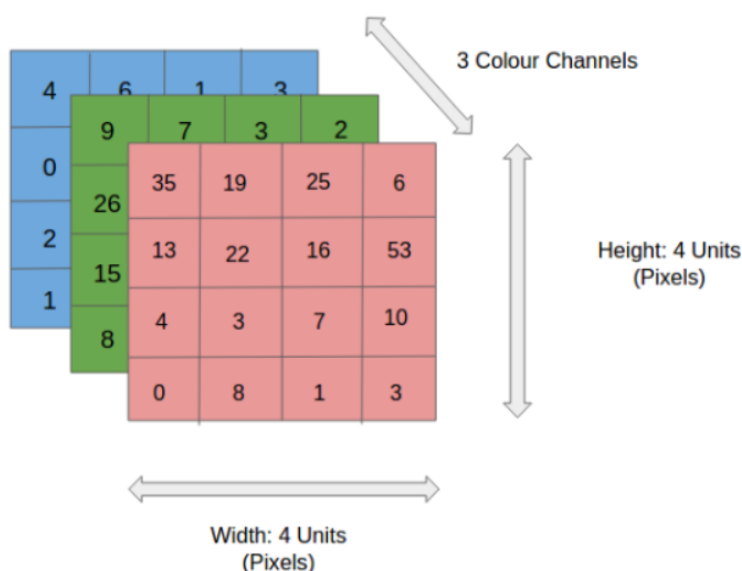
Δηλαδή, μετράμε την επικάλυψη μεταξύ f και g όταν μια συνάρτηση «αναστρέφεται» και μετατοπίζεται κατά x . Όποτε έχουμε διακριτές συναρτήσεις, το ολοκλήρωμα μετατρέπεται σε άθροισμα:

$$(f * g)(i) = \sum_a f(a)g(i + a) \quad (3.32)$$

Για διδιάστατους ταυστές, έχουμε ένα αντίστοιχο άθροισμα με δείκτες (a,b) για f και (i-a,j-b) για g, αντίστοιχα:

$$(f * g)(i,j) = \sum_a \sum_b f(a,b)g(i-a,j-b) \quad (3.33)$$

Πριν πάμε στη λειτουργία του CNN, ας καλύψουμε τα βασικά, όπως τι είναι μια εικόνα και πώς αναπαρίσταται. Μια εικόνα RGB 3.18 δεν είναι παρά ένας πίνακας τιμών εικονοστοιχείων με τρία επίπεδα, ενώ μια εικόνα σε κλίμακα του γκρι 3.19 είναι η ίδια αλλά έχει ένα μόνο επίπεδο. Για απλότητα, ας επιμείνουμε στις εικόνες σε κλίμακα του γκρι καθώς προσπαθούμε να κατανοήσουμε πώς λειτουργούν τα CNN. [53]



Σχήμα 3.18: Σχηματική Αναπαράσταση RGB εικόνας και οι διαστάσεις της [53]

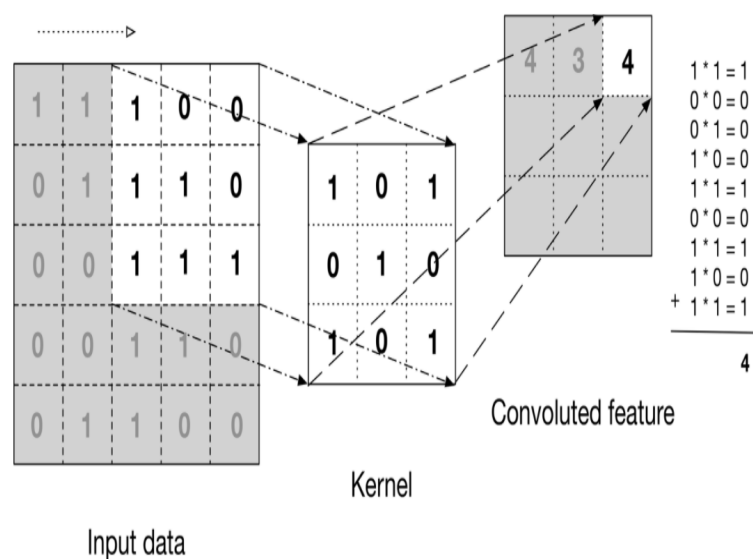
Η παρακάτω εικόνα 3.19 δείχνει τι είναι μια συνέλιξη. Παίρνουμε ένα φίλτρο(kernel) (μήτρας 3×3) και το εφαρμόζουμε στην εικόνα εισόδου για να λάβουμε το συνεκτικό χαρακτηριστικό. Αυτό το σύνθετο χαρακτηριστικό μεταβιβάζεται στο επόμενο επίπεδο.

Τεχνικά, τα μοντέλα CNN προκειμένου να κάνουν σωστή εκμάθηση σε επίπεδο εκπαίδευσης και τεστ, κάθε εικόνα εισόδου θα την περάσει από μια σειρά επιπέδων συνέλιξης με φίλτρα (Kernels), Pooling, πλήρως συνδεδεμένα επίπεδα (FC) και θα εφαρμόσει τη συνάρτηση Softmax για να ταξινομήσει ένα αντικείμενο με πιθανολογικές τιμές μεταξύ 0 και 1. Το παρακάτω σχήμα ;; είναι μια πλήρης ροή του CNN για την επεξεργασία μιας εικόνας εισόδου και ταξινομεί τα αντικείμενα με βάση τις τιμές. Όλα αυτά τα στάδια θα αναλυθούν εκτενώς ώστε να γίνει αντιληπτό τι λειτουργία διενεργείται σε κάθε στάδιο.

[55] [56]

Convolution Layer

Η συνέλιξη είναι το πρώτο επίπεδο που εξάγει χαρακτηριστικά από μια εικόνα εισόδου. Η



Σχήμα 3.19: Σχηματική Αναπαράσταση gray εικόνας και οι διαστάσεις της [54]

λειτουργία ενός συνελκτικού επιπέδου, όπως ανερώνει το όνομα του, στηρίζεται στην έννοια της συνέλιξης. Ως πυρήνας μπορεί να οριστεί ένα παράθυρο καθορισμένης διάστασης $n \times n$ ίδιου άθους με την εικόνα (για παράδειγμα 3 σε περίπτωση RGB εικόνας). Πολλαπλασιάζονται λοιπόν τα εικονοστοιχεία της εισόδου (pixels) με τα αντίστοιχα στοιχεία του kernel και προκύπτει ο πίνακας εξόδου. Αυτό το στάδιο μπορούμε να το δούμε και στην εικόνα ;:

Stride

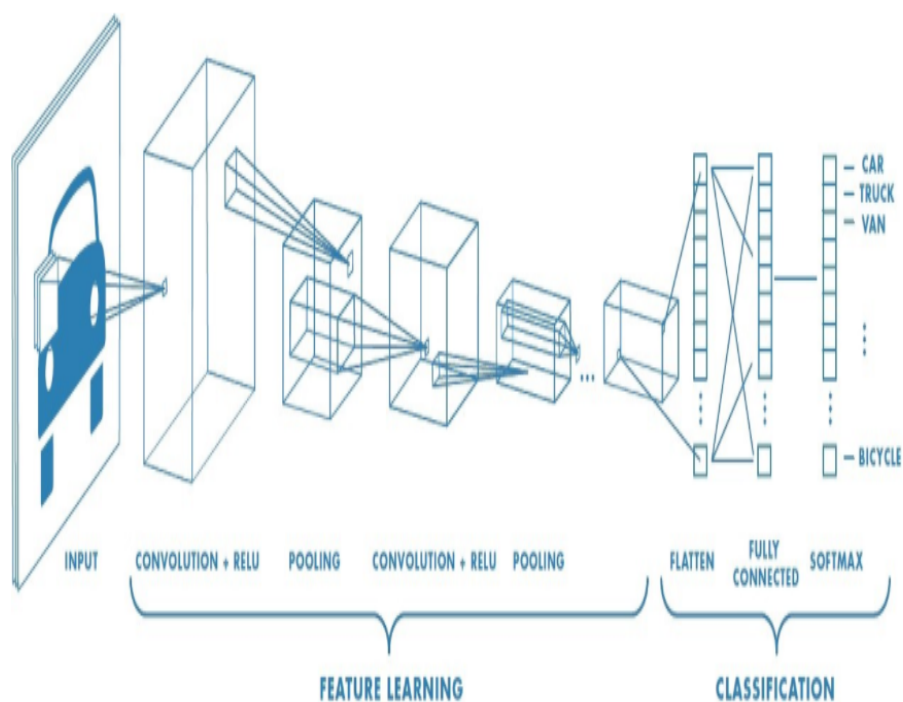
Διασκελισμός(stride) είναι ο αριθμός των εικονοστοιχείων που μετατοπίζονται στον πίνακα εισόδου. Όταν ο διασκελισμός είναι 1 τότε μετακινούμε τα φίλτρα σε 1 pixel κάθε φορά-μετακινούμενοι είτε οριζοντια είτε κάθετα. Όταν ο διασκελισμός είναι 2 τότε μετακινούμε τα φίλτρα σε 2 pixel κάθε φορά και ούτω καθεξής. Το παρακάτω σχήμα 3.21 δείχνει ότι η συνέλιξη θα λειτουργούσε με ένα βήμα 2.

Padding

Μερικές φορές το φίλτρο δεν ταιριάζει απόλυτα στην εικόνα εισόδου, έτσι έχουμε δύο επιλογές: Συμπληρώνουμε την εικόνα με μηδενικά (zero-padding) έτσι ώστε να ταιριάζει με τις διαστάσεις του kernel-filter και να πραγματοποιούνται ομαλά οι μαθηματικές πράξεις πινάκων. Ή πετάμε ένα μέρος της εικόνας όπου το φίλτρο δεν χωρούσε για να γίνουν οι πράξεις. Αυτό ονομάζεται έγκυρη συμπλήρωση που διατηρεί έγκυρο μόνο μέρος της εικόνας. Ένα παράδειγμα του zero-padding που είναι και πιο συχνά εφαρμόσιμο φαίνεται στο Σχήμα 3.22.

Rectified Linear Unit(ReLU)

Η συνάρτηση ενεργοποίησης ReLU που φαίνεται και στο Σχήμα 3.23 εξηγηρετεί μη γραμμική λειτουργία. Η έξοδος είναι $f(x) = \max(0, x)$. Ο σκοπός του ReLU είναι να εισάγει τη μη γραμμικότητα στο ConvNet μας. Αυτό συμβαίνει διότι τα δεδομένα του πραγματικού κόσμου που χρησιμοποιούνται στα CNN είναι μη γραμμικά. Υπάρχουν άλλες μη γραμμικές συναρτήσεις όπως η tanh ή η sigmoid που μπορούν επίσης να χρησιμοποιηθούν αντί για την



Σχήμα 3.20: Νευρωνικό δίκτυο με πολλαπλά συνεπικτικά στρώματα [57]

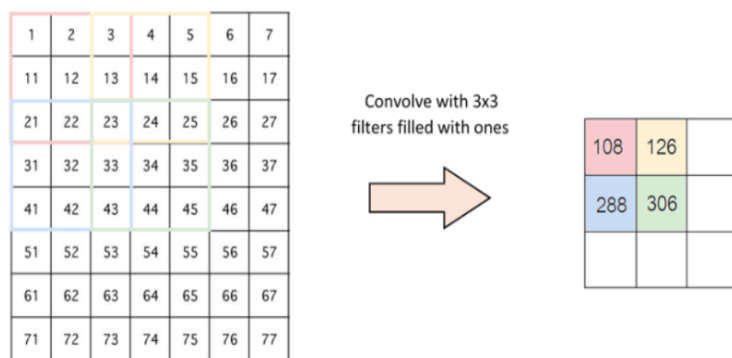
ReLU. Οι περισσότεροι από τους επιστήμονες δεδομένων χρησιμοποιούν την ReLU, καθώς από άποψη απόδοσης είναι καλύτερη από τα άλλα δύο.

Pooling Layers

Η υποδειγματοληψία μειώνει τη διάσταση κάθε χάρτη, αλλά διατηρεί σημαντικές πληροφορίες. Τα επίπεδα υποδειγματοληψίας σκοπεύουν στη μείωση των διαστάσεων των χαρτών ενεργοποίησης που προκύπτουν από τα συνεπικτικά επίπεδα. Πρακτικά, χωρίζουν τον χάρτη σε μη επικαλυπτόμενα μεταξύ τους τμήματα και για κάθε ένα από αυτά επιλέγουν μία αντιπροσωπευτική τιμή. Η χωρική συγκέντρωση μπορεί να είναι διαφόρων τύπων: Max Pooling, Mean Pooling, Sum Pooling. Το Max Pooling λαμβάνει το μεγαλύτερο στοιχείο από τον υποχάρτη χαρακτηριστικών. Το Mean Pooling λαμβάνει τον μέσο όρο των στοιχείων κάθε μη επικαλυπτόμενου υποχάρτη ενώ το Sum Pooling λαμβάνει το άθροισμα όλων των στοιχείων στον χάρτη χαρακτηριστικών. Ένα χαρακτηριστικό παράδειγμα Max Pooling μπορούμε να το δούμε και στην εικόνα 3.24.

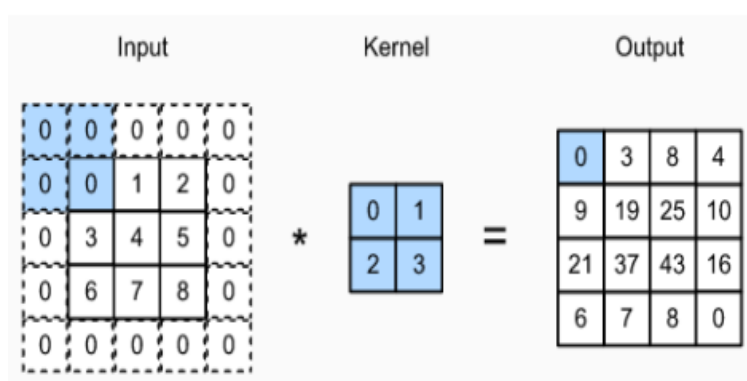
Fully Connected Layer

Το στρώμα που ονομάζουμε στρώμα Fully Connected, ουσιαστικά ισοπεδώσαμε τη μήτρα μας σε διάνυσμα και την τροφοδοτήσαμε σε ένα πλήρως συνδεδεμένο στρώμα σαν ένα νευρωνικό δίκτυο 3.25. Τα πλήρως συνδεδεμένα επίπεδα αποτελούν τα τελευταία επίπεδα του δικτύου. Η είσοδος στο πλήρως συνδεδεμένο στρώμα είναι η έξοδος από το τελικό Pooling ή Convolutional Layer, το οποίο ισοπεδώνεται και στη συνέχεια τροφοδοτείται στο πλήρως συνδεδεμένο στρώμα. Η έξοδος από το τελικό (και οποιοδήποτε) επίπεδο συγκέντρωσης και συνέλιξης είναι ένας τρισδιάστατος πίνακας, για να ισοπεδωθεί, πρέπει να ξετυλιχθούν όλες



[57]

Σχήμα 3.21: Τεχνική stride με βήμα 2



Σχήμα 3.22: Τεχνική zero-padding

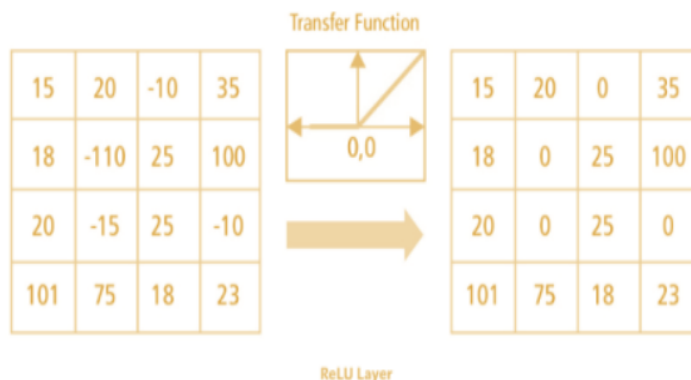
[58]

οι τιμές τους σε ένα διάνυσμα.

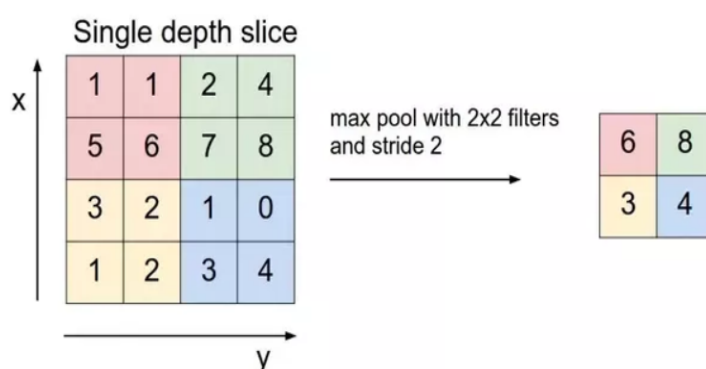
Αυτό το Flattened vector 3.26 συνδέεται στη συνέχεια με μερικά πλήρως συνδεδεμένα επίπεδα που είναι ίδια με τα τεχνητά νευρωνικά δίκτυα και εκτελούν τις ίδιες μαθηματικές πράξεις. Αφού περάσει μέσα από τα πλήρως συνδεδεμένα επίπεδα, το τελικό στρώμα χρησιμοποιεί τη συνάρτηση ενεργοποίησης softmax (αντί για ReLU) που χρησιμοποιείται για να ταξινομηθεί η έξοδος σε συγκεκριμένη κλάση (ταξινόμηση).[59]

3.4.2 1D Convolutional Neural Networks

Μέχρι πρόσφατα, το θέμα της μοντελοποίησης ακολουθιών στο πλαίσιο της βαθιάς μάθησης είχε συνδεθεί σε μεγάλο βαθμό με επαναλαμβανόμενες αρχιτεκτονικές νευρωνικών δικτύων όπως το LSTM και το GRU. Οι Bai et al. προτείνουν ότι αυτός ο τρόπος σκέψης είναι απαρχαιωμένος και ότι τα συνελκτικά δίκτυα θα πρέπει να λαμβάνονται υπόψη ως ένας από τους κύριους υποψήφιους αλγόριθμους κατά τη μοντελοποίηση διαδοχικών δεδομένων. Κατάφεραν να δείξουν ότι τα συνελκτικά δίκτυα μπορούν να επιτύχουν καλύτερη απόδοση από τα RNN σε πολλές εργασίες, αποφεύγοντας τα κοινά μειονεκτήματα των επαναλαμβανόμενων μοντέλων, όπως το πρόβλημα της κλίσης έκρηξης/εξαφάνισης ή η έλλειψη διατήρησης μνήμης. Επιπλέον, η χρήση ενός συνελκτικού δικτύου αντί για ένα επαναλαμβανόμενο μπορεί να οδηγήσει σε βελτιώσεις στην απόδοση καθώς επιτρέπει παράλληλο υπολογισμό



Σχήμα 3.23: Συνάρτηση Ενεργοποίησης ReLU [57]

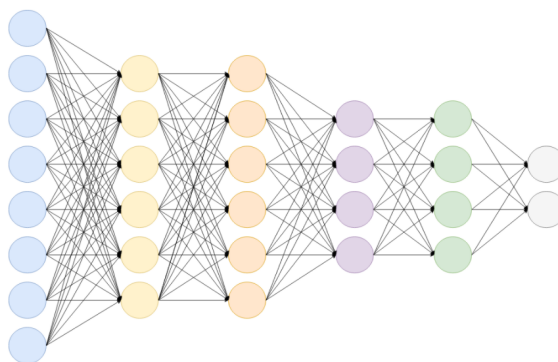


Σχήμα 3.24: Σχηματική Αναπαράσταση Max Pooling Layer [57]

των εξόδων.

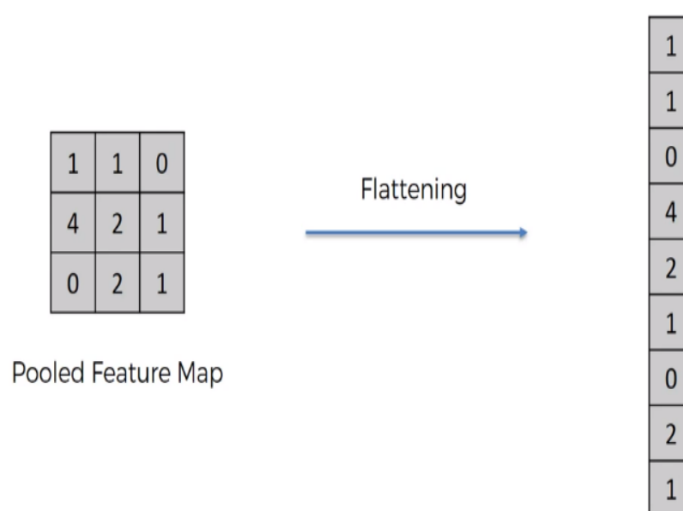
Ένα 1D συνελκτικό δίκτυο παίρνει ως είσοδο έναν τρισδιάστατο τανυστή και επίσης εξάγει έναν τρισδιάστατο τανυστή. Ο τανυστής εισόδου έχει το σχήμα (batch size, input length, input size) και ο τανυστής εξόδου έχει το σχήμα (batch size, output length, output size). Δεδομένου ότι κάθε στρώμα έχει το ίδιο μήκος εισόδου και εξόδου, μόνο η τρίτη διάσταση των τανυστών εισόδου και εξόδου διαφέρει. Στη μονομεταβλητή περίπτωση, το input size και το output size θα είναι και τα δύο ίσα με ένα. Στη γενικότερη περίπτωση πολλαπλών μεταβλητών, το μέγεθος εισόδου και το μέγεθος εξόδου ενδέχεται να διαφέρουν, καθώς μπορεί να μην θέλουμε να προβλέψουμε κάθε στοιχείο της ακολουθίας εισόδου. [57]

Μπορούμε να δούμε ότι για να υπολογίσουμε ένα στοιχείο της εξόδου, εξετάζουμε μια σειρά από διαδοχικά στοιχεία μήκους kernel size της εισόδου. Στο παρακάτω παράδειγμα επιλέξαμε kernel size= 3. Για να λάβουμε την έξοδο, παίρνουμε το γινόμενο κουκίδων της υποακολουθίας της εισόδου και ένα διάνυσμα πυρήνα με μαθησιακά βάρη του ίδιου μήκους. Για να ληφθεί το επόμενο στοιχείο της εξόδου, εφαρμόζεται η ίδια διαδικασία, αλλά το παράθυρο μεγέθους kernel size της ακολουθίας εισόδου μετατοπίζεται προς τα δεξιά κατά ένα στοιχείο (για τους σκοπούς αυτού του μοντέλου πρόβλεψης, ο βηματισμός ορίζεται πάντα σε 1). Το ίδιο σύνολο βαρών πυρήνα θα χρησιμοποιηθεί για τον υπολογισμό κάθε εξόδου ενός συνελκτικού επιπέδου. Το παρακάτω Σχήμα 3.28 δείχνει δύο διαδοχικά στοιχεία εξόδου και τις αντίστοιχες υποακολουθίες εισόδου απ' όπου προήλθαν. Για να γίνει πιο απλή η



[57]

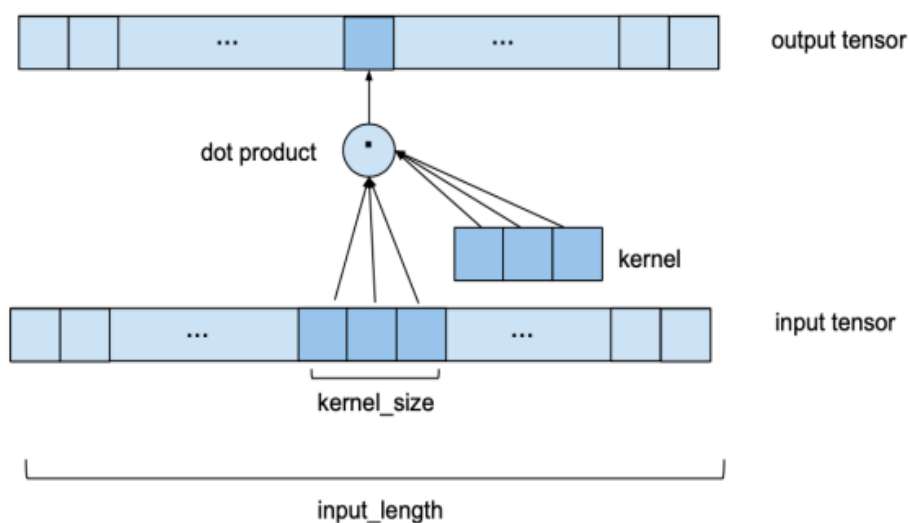
Σχήμα 3.25: Σχηματική Αναπαράσταση Fully Connected Layer

Σχήμα 3.26: Σχηματική Αναπαράσταση Flattening
[60]

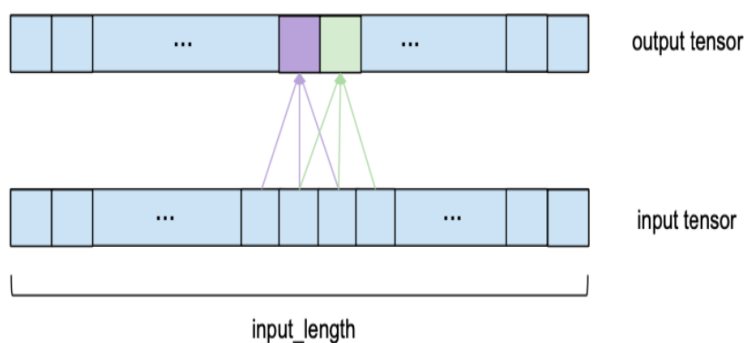
απεικόνιση, το γινόμενο κουκίδων με το διάνυσμα του πυρήνα δεν εμφανίζεται πλέον, αλλά λαμβάνει χώρα για κάθε στοιχείο εξόδου με τα ίδια βάρη πυρήνα.

Για να επιβεβαιωθεί ότι η ακολουθία εξόδου έχει το ίδιο μήκος με την ακολουθία εισόδου, εφαρμόζεται κάποια μηδενική επένδυση. Αυτό σημαίνει ότι προστίθενται επιπλέον εγγραφές μηδενικής τιμής είτε στην αρχή είτε στο τέλος του τανυστή εισόδου για να διασφαλιστεί ότι η έξοδος έχει το επιθυμητό μήκος. Το πώς ακριβώς γίνεται αυτό θα εξηγηθεί σε επόμενες ενότητες.

Στην περίπτωση που έχουμε πολλά κανάλια εισόδου, δηλαδή το nr input channels είναι μεγαλύτερο από 1. Σε αυτήν την περίπτωση, η διαδικασία που περιγράφεται παραπάνω επαναλαμβάνεται για κάθε κανάλι εισόδου, αλλά κάθε φορά με διαφορετικό πυρήνα. Αυτό οδηγεί σε ενδιάμεσα διανύσματα εξόδου nr input channels και έναν αριθμό βαρών πυρήνα μεγέθους $kernel * nr$ input channels. Στη συνέχεια, όλα τα ενδιάμεσα διανύσματα εξόδου αθροίζονται για να ληφθεί το τελικό διάνυσμα εξόδου. Κατά κάποιο τρόπο αυτό ισοδυναμεί με μια 2D συνέλιξη με έναν τανυστή εισόδου σχήματος (input size, nr input channels) και έναν πυρήνα σχήματος (kernel size, nr input channels) όπως φαίνεται στο παρακάτω Σχήμα 3.29.



Σχήμα 3.27: Υπολογισμός συνέλιξης για 1D-CNN με $input\ channels = output\ channels = 1$ [61]



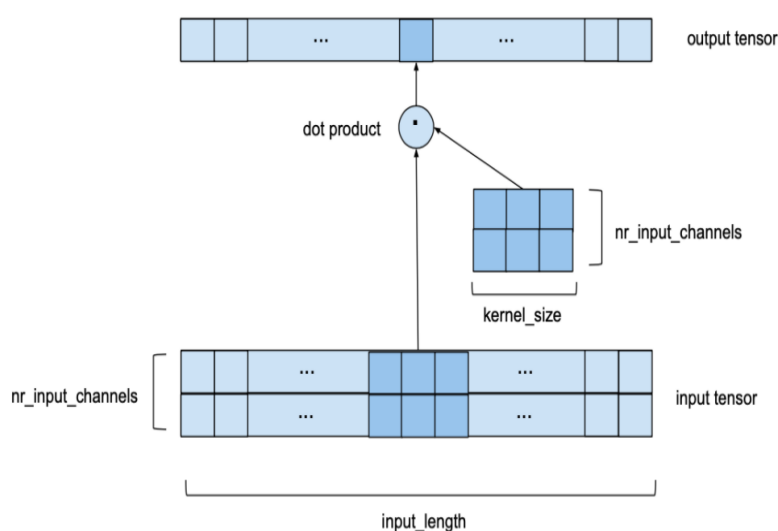
Σχήμα 3.28: Υπολογισμός συνέλιξης για δύο διαδοχικά στοιχεία εξόδου [61]

Εξακολουθεί να είναι ωστόσο 1D με την έννοια ότι το παράθυρο κινείται μόνο κατά μήκος ενός μόνο άξονα, αλλά έχουμε μια 2D συνέλιξη σε κάθε βήμα με την έννοια ότι χρησιμοποιούμε έναν δισδιάστατο πίνακα πυρήνα.

Για αυτό το παράδειγμα τα $nr\ input\ channels$ είναι ίσα με 2. Τώρα, αντί για ένα διάνυσμα πυρήνα να ολισθαίνει πάνω από μια μονοδιάστατη ακολουθία εισόδου, έχουμε ένα $nr\ input\ channels$ από τον πίνακα $kernel_size$ που ολισθαίνει κατά μήκος μιας ευρείας σειράς μήκους ($input\ length * nr\ input\ channels$).

Εάν και τα δύο κανάλια $input\ channels$ και $nr\ output\ channels$ είναι μεγαλύτερα από 1, η παραπάνω διαδικασία απλά επαναλαμβάνεται για κάθε κανάλι εξόδου με διαφορετικό πίνακα πυρήνα. Στη συνέχεια, τα διανύσματα εξόδου στοιβάζονται το ένα πάνω στο άλλο με αποτέλεσμα έναν τανυστή εξόδου σχήματος ($input\ length, nr\ output\ channels$). Ο αριθμός των βαρών του πυρήνα σε αυτήν την περίπτωση είναι ίσος με το ($kernel\ size * nr\ input\ channels * nr\ output\ channels$).

Οι δύο μεταβλητές $nr\ input\ channels$ και $nr\ output\ channels$ εξαρτώνται από τη θέση



Σχήμα 3.29: Υπολογισμός συνέλιξης για 1D-CNN με $input_channels = 1$ [61]

του επιπέδου εντός του δικτύου. Το πρώτο επίπεδο θα έχει $nr_input_channels = input_size$ και το τελευταίο επίπεδο θα έχει $nr_output_channels = output_size$. Όλα τα άλλα επίπεδα θα χρησιμοποιούν τον ενδιάμεσο αριθμό καναλιού που δίνεται από το `num_filters`.

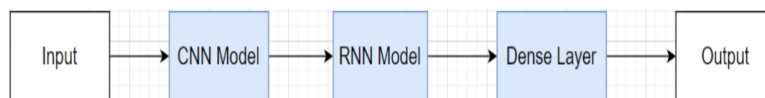
3.4.3 Ζεύξη Επαναληπτικών και Συνελκτικών Νευρωνικών Δικτύων (CNN-RNNs)

Σε αυτή την παράγραφο θα συνδυάσουμε τα συνελκτικά δίκτυα με τα επαναληπτικά. Ο συνδυασμός των CNN με τα RNN ενδείκνυται για την αντιμετώπιση προβλημάτων πρόβλεψης ακολουθίας. Συνδυάζοντας αυτές τις δύο αρχιτεκτονικές έχουμε τη δυνατότητα τόσο να εξάγουμε τα σημαντικότερα από τα χαρακτηριστικά των δεδομένων, όσο και να αξιοποιήσουμε αυτά σε μία χρονική σειρά διατηρώντας την όποια πληροφορία χρειαζόμαστε από το παρελθόν. [62]

Σε αυτή την εργασία παρουσιάζουμε τρεις διαφορετικούς αλγόριθμους με τους οποίους μπορούμε να ενώσουμε αυτές τις 2 κατηγορίες νευρωνικών δικτύων. Η αρχιτεκτονική των μοντέλων είναι η απλή τοποθέτηση για αρχή του συνελκτικού δικτύου (για το φιλτράρισμα των σημαντικότερων χαρακτηριστικών των δεδομένων εισόδου) και στη συνέχεια το επαναληπτικό δίκτυο που θα διαχειριστεί αυτά τα χαρακτηριστικά ως χρονική ακολουθία. Η αρχιτεκτονική αυτή είναι ιδιαίτερα απλή και κατανοητή, καθώς απλώς τοποθετεί σε σειρά τα δύο είδη δικτύων με το μοναδικό σημείο το οποίο χρήζει προσοχής να είναι αυτό όπου το κομμάτι της εξόδου του 1ου δικτύου αποτελεί είσοδο για το 2ο. Τα CNN-RNNs αναπτύχθηκαν για προβλήματα πρόβλεψης ακολουθιακών δεδομένων και την εφαρμογή δημιουργίας περιγραφικών κειμένων από ακολουθίες εικόνων. Το δικό μας πρόβλημα έγκειται στην κατηγορία προβλημάτων NLP όπου σκοπός είναι η ανγνώριση ομιλίας ή πρόβλεψη χρονοσειρών. Πιο συγκεκριμένα το CNN χρησιμοποιείται ως εξολκέας χαρακτηριστικών και το RNN ως ο κύριος επεξεργαστής των δεδομένων.

Μια τέτοια δομή ενός CNN-RNN παρουσιάζεται στην εικόνα 3.30, στην οποία μπορούμε

να δούμε ότι το δίκτυο απαρτίζεται από 3 διακριτά μέρη, το CNN μοντέλο, το RNN μοντέλο, και το πλήρως διασυνδεδεμένο (Dense) στρώμα εξόδου. Στην παρούσα εργασία ως επαναληπτικό δίκτυο χρησιμοποιήθηκε αρχικά το RNN, έπειτα το LSTM και τέλος το GRU.



Σχήμα 3.30: Η αρχιτεκτονική του δικτύου CNN-RNN

3.5 Χρονικά Συνελικτικά Δίκτυα (Temporal Convolutional Network)

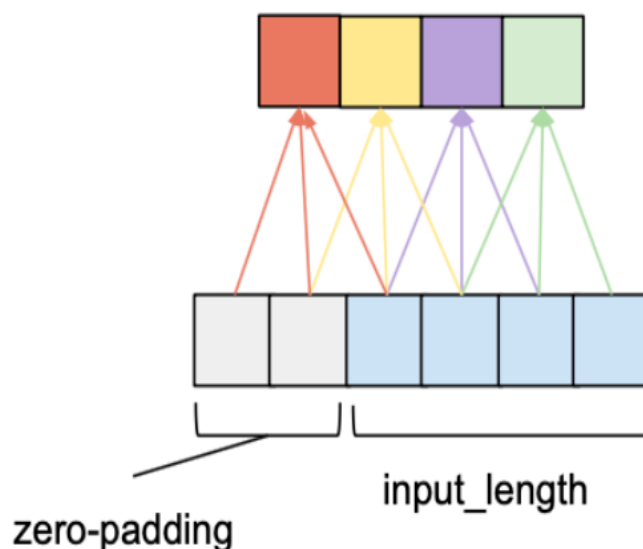
Η χρήση ενός συνελικτικού δικτύου αντί ενός αναδρομικού μπορεί να οδηγήσει σε ελιτώσεις απόδοσης καθώς επιτρέπει τον παράλληλο υπολογισμό των εξόδων. Η αρχιτεκτονική που θα αναλυθεί ονομάζεται Temporal Convolutional Network (TCN) και α εξηγηθεί στην παρούσα ενότητα. Ένα TCN, συντομογραφία του Temporal Convolutional Network δηλαδή Χρονικό Συνελικτικό δίκτυο, αποτελείται από διασταλμένα (dilated), ά (causal) 1D συνελικτικά στρώματα με τα ίδια μήκη εισόδου και εξόδου. Στις παρακάτω υποενότητες αυτοί οι όροι αναπτύσσονται διεξοδικά.

Αιτιώδης Συνέλιξη (Causal Convolution)

Για να είναι αιτιατό ένα συνελικτικό επίπεδο, πρέπει για κάθε i του συνόλου $(0, \dots, \text{μήκος εισόδου} - 1)$ το ιστό στοιχείο της ακολουθίας εξόδου να εξαρτάται μόνο από τα στοιχεία της ακολουθίας εισόδου με δείκτες $(0, \dots, i)$. Με άλλα λόγια, ένα στοιχείο στην ακολουθία εξόδου μπορεί να εξαρτάται μόνο από στοιχεία που βρίσκονται πριν από αυτό στην ακολουθία εισόδου. Όπως αναφέρθηκε προηγουμένως, για να διασφαλίσουμε ότι ένας τανυστής εξόδου έχει το ίδιο μήκος με τον τανυστή εισόδου, πρέπει να εφαρμόσουμε μηδενική επένδυση (zero-padding). Εάν εφαρμόσουμε μόνο μηδενική επένδυση στην αριστερή πλευρά του τανυστή εισόδου, τότε θα εξασφαλιστεί η αιτιακή συνέλιξη. Δεδομένου ότι δεν υπάρχει επένδυση στη δεξιά πλευρά της ακολουθίας εισόδου, το τελευταίο στοιχείο από το οποίο εξαρτάται είναι το τελευταίο στοιχείο της εισόδου. Τώρα εξετάστε το δεύτερο έως το τελευταίο στοιχείο εξόδου της ακολουθίας εξόδου. Το παράθυρο του πυρήνα του μετατοπίζεται προς τα αριστερά κατά ένα σε σύγκριση με το τελευταίο στοιχείο εξόδου, που σημαίνει ότι η δεξιά εξάρτησή του στην ακολουθία εισόδου είναι το δεύτερο προς το τελευταίο στοιχείο της ακολουθίας εισόδου. Ακολουθεί επαγωγή ότι για κάθε στοιχείο στην ακολουθία εξόδου, η τελευταία του εξάρτηση στην ακολουθία εισόδου έχει τον ίδιο δείκτη με το ίδιο. Το παρακάτω σχήμα δείχνει ένα παράδειγμα με $\text{input length}=4$ και $\text{kernel size}=3$.

Μπορούμε να δούμε στο Σχήμα 3.31 ότι με ένα αριστερό μηδενικό συμπλήρωμα 2 καταχωρήσεων μπορούμε να επιτύχουμε το ίδιο μήκος εξόδου, ενώ τηρούμε τον κανόνα της αιτιότητας. Στην πραγματικότητα, χωρίς διαστολή, ο αριθμός των εγγραφών μηδενικής συμπλήρωσης που απαιτούνται για τη διατήρηση του μήκους εισόδου είναι πάντα ίσος με το $(\text{size kernel}-1)$. Ένα σημαντικό μειονέκτημα αυτού του αστικού σχεδιασμού είναι ότι για να επιτύχουμε να επεξεργαστούμε χρονοσειρές μεγάλου μήκους, χρειαζόμαστε ένα εξαιρετικά

αθύ δίκτυο ή πολύ μεγάλα μεγέθη πυρήνα, κανένα από τα οποία δεν ήταν ιδιαίτερα εφικτό κατά την πρώτη εισαγωγή των μεθόδων. Έτσι, στην επόμενη υποενότητα, περιγράψουμε πώς μπορούν να ενσωματωθούν τεχνικές από σύγχρονες συνελκτικές αρχιτεκτονικές σε ένα TCN για να επιτρέψουν τόσο την ύπαρξη βαθιών δικτύων όσο και την επεξεργασία χρονοσειρών μεγάλου μήκους.



Σχήμα 3.31: *Zero padding with kernel size=3 and input length=4* [61]

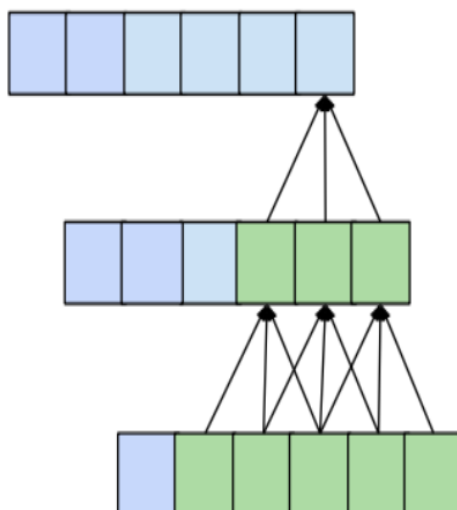
Διαστολή(Dilation)

Ένα επιθυμητό ποιοτικό γνώρισμα ενός μοντέλου πρόβλεψης είναι η τιμή μιας συγκεκριμένης καταχώρισης στην έξοδο να εξαρτάται από όλες τις προηγούμενες εγγραφές στην είσοδο, δηλαδή όλες τις εγγραφές που έχουν δείκτη μικρότερο ή ίσο με τον εαυτό της. Αυτό επιτυγχάνεται όταν το πεδίο των δεικτών, που σημαίνει το σύνολο των καταχωρήσεων της αρχικής εισόδου που επηρεάζουν μια συγκεκριμένη καταχώριση της εξόδου, έχει μέγεθος *input length*. Αυτό το ονομάζουμε επίσης «πλήρη κάλυψη ιστορικού». Όπως έχουμε δει προηγουμένως, ένα συμβατικό συνελκτικό επίπεδο κάνει μια καταχώριση στην έξοδο να εξαρτάται από τις εγγραφές *kernel size* της εισόδου που έχουν δείκτη μικρότερο ή ίσο με τον εαυτό της. Για παράδειγμα, αν έχουμε *kernel size=3*, το 5ο στοιχείο στην έξοδο θα εξαρτάται από τα στοιχεία 3, 4 και 5 της εισόδου. Αυτή η προσέγγιση επεκτείνεται όταν στοιδιάζουμε πολλά στρώματα το ένα πάνω στο άλλο. Στο παρακάτω σχήμα μπορούμε να δούμε ότι στοιδιάζοντας δύο επίπεδα με *kernel size=3* έχουμε ένα δεκτικό μέγεθος πεδίου 5.

Γενικότερα, ένα 1D συνελκτικό δίκτυο με *n* επίπεδα και ένα *kernel size=k* έχει ένα δεκτικό πεδίο μεγέθους *r*.

$$r = 1 + n * (k - 1) \quad (3.34)$$

Για να γνωρίζουμε πόσα επίπεδα χρειάζονται για την πλήρη κάλυψη, μπορούμε να ορίσουμε το μέγεθος του πεδίου υποδοχής(*r*) σε *input length l* και να λύσουμε τον αριθμό των επιπέδων *n* (πρέπει να στρογγυλοποιήσουμε προς τα πάνω σε περίπτωση μη ακέραιων



Σχήμα 3.32: *Dilation σε σιβαγμένα επίπεδα με kernel size=3*
[61]

τιμών):

$$n = (l - 1)/(k - 1) \quad (3.35)$$

Αυτό σημαίνει ότι, δεδομένου ενός σταθερού kernel size, ο αριθμός των επιπέδων που απαιτούνται για την πλήρη κάλυψη ιστορικού είναι γραμμικός στο μήκος του τανυστή εισόδου, γεγονός που θα έχει ως αποτέλεσμα δίκτυα που γίνονται πολύ βαθιά πολύ γρήγορα, οδηγώντας σε μοντέλα με πολύ μεγάλο αριθμό παραμέτρων που χρειάζονται περισσότερο χρόνο για να προπονηθούν. Επιπλέον, ένας μεγάλος αριθμός στρωμάτων έχει αποδειχθεί ότι οδηγεί σε προβλήματα υποβάθμισης που σχετίζονται με την κλίση της συνάρτησης απώλειας. Ένας τρόπος για να αυξηθεί το μέγεθος του δεκτικού πεδίου, διατηρώντας παράλληλα τον αριθμό των στρωμάτων σχετικά μικρό, είναι η εισαγωγή διαστολής στο συνελκτικό δίκτυο.

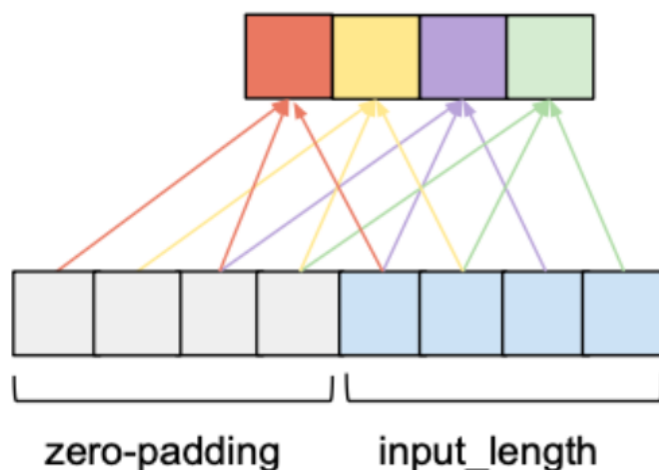
Η διαστολή στο πλαίσιο ενός συνελκτικού στρώματος αναφέρεται στην απόσταση μεταξύ των στοιχείων της ακολουθίας εισόδου που χρησιμοποιούνται για τον υπολογισμό μιας καταχώρησης της ακολουθίας εξόδου. Έτσι, ένα συμβατικό συνελκτικό στρώμα θα μπορούσε να θεωρηθεί ως ένα 1-διασταλμένο στρώμα, καθώς τα στοιχεία εισόδου για 1 τιμή εξόδου είναι γειτονικά. Το παρακάτω σχήμα δείχνει ένα παράδειγμα 2-διασταλμένου στρώματος με input length= 4 και kernel size=3.

Γενικότερα, ένα d-διασταλμένο στρώμα με μέγεθος πυρήνα k έχει ένα δεκτικό πεδίο που εκτείνεται σε μήκος

$$r = 1 + d * (k - 1) \quad (3.36)$$

Εάν το d είναι σταθερό, αυτό θα απαιτεί ακόμα έναν αριθμό γραμμικό στο μήκος του τανυστή εισόδου για να επιτευχθεί πλήρης κάλυψη δεκτικού πεδίου (μόλις μειώσαμε τη σταθερά).

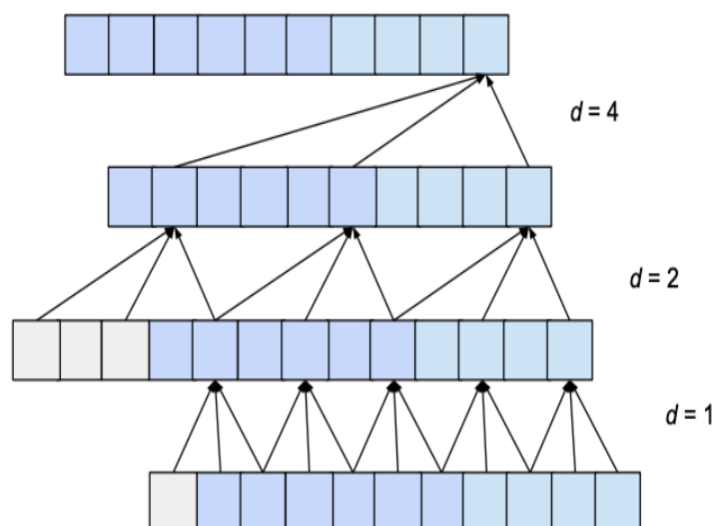
Αυτό το πρόβλημα μπορεί να αντιμετωπιστεί αυξάνοντας την τιμή του d εκθετικά καθώς προχωράμε προς τα πάνω μέσα από τα επίπεδα. Για αυτό, επιλέγουμε έναν σταθερό ακέραιο αριθμό βάσης διαστολής b που θα μας επιτρέψει να υπολογίσουμε τη διαστολή d ενός



Σχήμα 3.33: 2-σταθμένο στρώμα με $input\ length=4$ $kernel\ size=3$
[61]

συγκεκριμένου στρώματος ως συνάρτηση του αριθμού των στρωμάτων κάτω από αυτό, i , ως $d = b^i$.

Το παρακάτω σχήμα 3.34 δείχνει ένα δίκτυο με $input\ length= 10$, $kernel\ size= 3$ και $dilation\ base= 2$ που έχει ως αποτέλεσμα 3 διευρυμένα συνελκτικά στρώματα για πλήρη κάλυψη.



Σχήμα 3.34: Σχηματική αναπαράσταση δικτύου με διασταθμένα συνελκτικά στρώματα με πλήρη κάλυψη

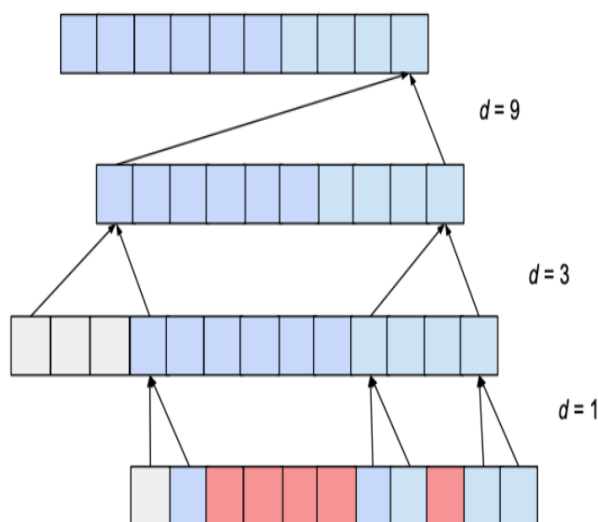
[61]

Εδώ δείχνουμε μόνο την επιρροή των εισόδων που επηρεάζουν την τελευταία τιμή της εξόδου. Ομοίως, εμφανίζονται μόνο οι εγγραφές μηδενικής συμπλήρωσης που είναι απαραίτητες για την τελευταία τιμή εξόδου. Σαφώς, η τελευταία τιμή εξόδου εξαρτάται από

το σύνολο της κάλυψης εισόδου. Στην πραγματικότητα, δεδομένου των υπερπαραμέτρων, θα μπορούσε να χρησιμοποιηθεί ένα input length έως και 15 διατηρώντας παράλληλα την πλήρη κάλυψη δεκτικού πεδίου. Σε γενικές γραμμές, κάθε πρόσθετο επίπεδο προσθέτει μια τιμή $d \cdot (k-1)$ στο τρέχον εύρος του πεδίου υποδοχής, όπου το d υπολογίζεται ως $d=b^i$, αντιπροσωπεύοντας τον αριθμό των επιπέδων κάτω από το νέο μας επίπεδο. Συνεπώς, το πλάτος του δεκτικού πεδίου w ενός TCN με εκθετική διαστολή της βάσης b , του μεγέθους του πυρήνα k και του αριθμού των στρωμάτων n δίνεται από τον τύπο:

$$w = 1 + \sum_{i=0}^{n-1} (k-1) * b^i = 1 + (k-1) * \frac{b^n - 1}{b - 1} \quad (3.37)$$

Ωστόσο, ανάλογα με τις τιμές b και k , αυτό το δεκτικό πεδίο μπορεί να έχει «οπές». Αξίζει να παρατηρήσουμε το ακόλουθο δίκτυο 3.35 με dilation base= 3 και kernel size= 2: Το πεδίο λήψης καλύπτει όντως ένα εύρος που είναι μεγαλύτερο από το μέγεθος εισόδου (δηλαδή 15). Ωστόσο, το δεκτικό πεδίο έχει τρύπες σε αυτό. Δηλαδή υπάρχουν καταχωρήσεις στην ακολουθία εισόδου από τις οποίες δεν εξαρτάται η τιμή εξόδου (που φαίνεται με κόκκινο). Για να λύσουμε αυτό το πρόβλημα, πρέπει είτε να αυξήσουμε το μέγεθος του πυρήνα στο 3 ή να μειώσουμε τη βάση διαστολής στο 2. Γενικά, για ένα δεκτικό πεδίο χωρίς τρύπες, το μέγεθος του πυρήνα k πρέπει να είναι τουλάχιστον τόσο μεγάλο όσο η βάση διαστολής b .



Σχήμα 3.35: Σχηματική αναπαράσταση δικτύου με διασταθμένα συνελκτικά στρώματα με οπές

[61]

Λαμβάνοντας υπόψη αυτές τις παρατηρήσεις, μπορούμε να υπολογίσουμε πόσα επίπεδα χρειάζεται το δίκτυό μας για πλήρη κάλυψη ιστορικού. Δεδομένου του μεγέθους του πυρήνα k , της βάσης διαστολής b , όπου $k \geq b$ και του μήκους εισόδου l , η ακόλουθη ανισότητα πρέπει να ισχύει για την πλήρη κάλυψη του ιστορικού:

$$1 + (k-1) * \frac{b^n - 1}{b - 1} \geq l \quad (3.38)$$

Μπορούμε να λύσουμε ως προς n και να πάρουμε τον ελάχιστο αριθμό απαιτούμενων επι-

πέδων ως:

$$\log_b \left[\frac{(l-1)(b-1)}{k-1} + 1 \right] \quad (3.39)$$

Μπορούμε να δούμε ότι ο αριθμός των επιπέδων είναι πλέον λογαριθμικός παρά γραμμικός στο μήκος της εισόδου. Αυτό συνιστά σημαντική βελτίωση, η οποία μπορεί να επιτευχθεί χωρίς να θυσιαστεί η δεκτική κάλυψη πεδίου.

Τώρα το μόνο πράγμα που μένει να καθοριστεί είναι ο αριθμός των εγγραφών μηδενικής συμπλήρωσης που απαιτούνται σε κάθε επίπεδο. Δεδομένης μιας βάσης διαστολής b , ενός μεγέθους πυρήνα k και ενός αριθμού στρωμάτων i κάτω από το τρέχον στρώμα μας, ο αριθμός των εγγραφών μηδενικής συμπλήρωσης p που απαιτούνται για το τρέχον στρώμα υπολογίζεται ως εξής:

$$p = b^i * (k - 1) \quad (3.40)$$

Υπολειμματικά Μπλοκ (Residual Blocks)

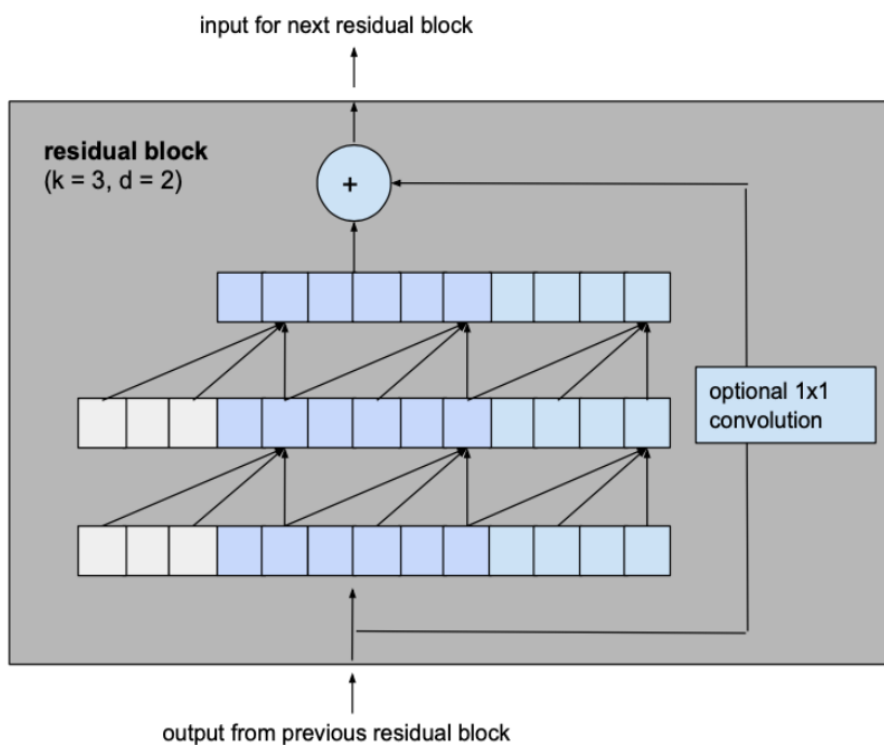
Η μεγαλύτερη τροποποίηση που κάνουμε στο βασικό μοντέλο που εισήχθη προηγουμένως είναι να αλλάξουμε το θεμελιώδες δομικό στοιχείο του μοντέλου από ένα απλό 1D αιτώδες συνελκτικό στρώμα σε ένα υπολειπόμενο μπλοκ που αποτελείται από 2 στρώματα με τον ίδιο παράγοντα διαστολής και μια υπολειπόμενη σύνδεση.

Ας εξετάσουμε ένα στρώμα με συντελεστή διαστολής $d=2$ και kernel size $k=3$ από το βασικό μοντέλο για να δούμε πώς αυτό μεταφράζεται σε ένα υπολειπόμενο μπλοκ του βελτιωμένου μοντέλου.

Η έξοδος των δύο συνελκτικών στρωμάτων θα προστεθεί στην είσοδο του υπολειπόμενου μπλοκ για να παραχθεί η είσοδος για το επόμενο μπλοκ. Για όλα τα εσωτερικά μπλοκ του δικτύου, δηλαδή όλα εκτός από το πρώτο και το τελευταίο, το πλάτος του καναλιού εισόδου και εξόδου είναι το ίδιο, επονομαζόμενα και ως αριθμός φίλτρων. Δεδομένου ότι το πρώτο συνελκτικό στρώμα του πρώτου υπολειπόμενου μπλοκ και το δεύτερο συνελκτικό στρώμα του τελευταίου υπολειπόμενου μπλοκ μπορεί να έχουν διαφορετικά πλάτη καναλιών εισόδου και εξόδου, το πλάτος του υπολειπόμενου τανυστή μπορεί να χρειαστεί να ρυθμιστεί, κάτι που γίνεται χρησιμοποιώντας μια συνέλιξη 1×1 . Αυτή η αλλαγή επηρεάζει τον υπολογισμό για τον ελάχιστο αριθμό απαιτούμενων στρωμάτων για πλήρη κάλυψη. Άρα πρέπει να σκεφτούμε πόσα υπολειπόμενα μπλοκ είναι απαραίτητα για να επιτευχθεί μια πλήρης δεκτική κάλυψη πεδίου. Η προσθήκη ενός υπολειπόμενου μπλοκ σε ένα TCN προσθέτει διπλάσιο πλάτος δεκτικού πεδίου από ό,τι όταν προστίθεται ένα βασικό αιτώδες στρώμα, καθώς περιλαμβάνει 2 τέτοια επίπεδα. Έτσι το συνολικό μέγεθος του δεκτικού πεδίου r ενός TCN με βάση διαστολής b , μέγεθος πυρήνα k με $k \geq b$ και αριθμό υπολειπόμενων μπλοκ n μπορεί να υπολογιστεί ως:

$$r = 1 + \sum_{i=0}^{n-1} 2 * (k - 1) * b^i = 1 + 2 * (k - 1) * \frac{b^n - 1}{b - 1} \quad (3.41)$$

που οδηγεί σε έναν ελάχιστο αριθμό υπολειπόμενων μπλοκ n για πλήρη κάλυψη ιστορικού



Σχήμα 3.36: Σχηματική αναπαράσταση αιτιατού δικτύου με $kernel\ size=3$ $dilation=2$ μετά τη βεβίωση

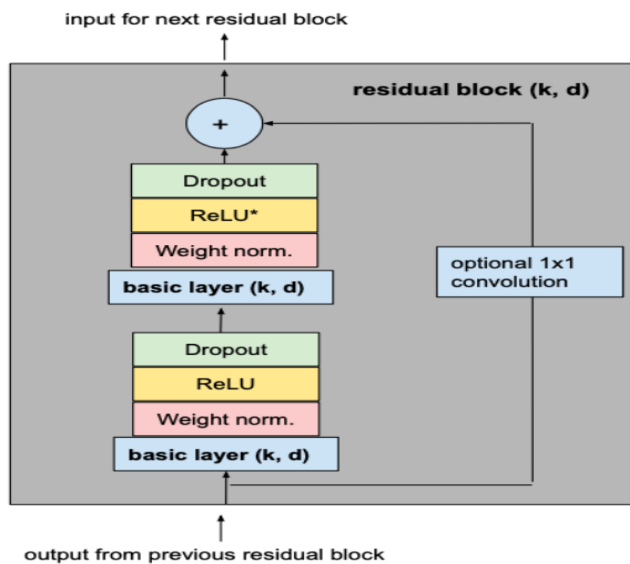
[61]

input length l :

$$n = \log_b \left[\frac{(l-1)(b-1)}{2 * (k-1)} + 1 \right] \quad (3.42)$$

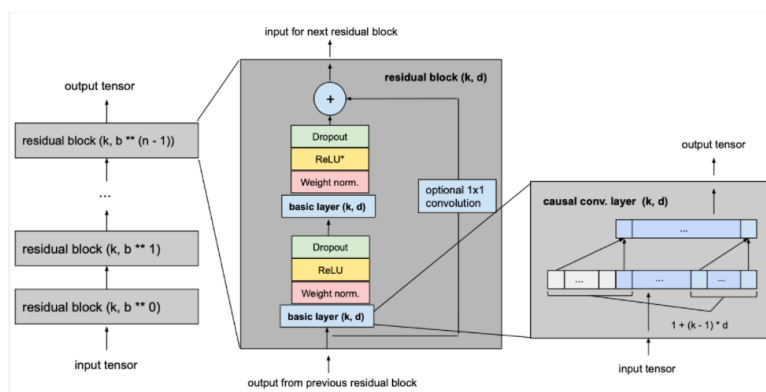
Για να κάνουμε το TCN μας κάτι περισσότερο από ένα υπερβολικά πολύπλοκο μοντέλο γραμμικής παλινδρόμησης, πρέπει να προστεθούν συναρτήσεις ενεργοποίησης πάνω από τα συνελκτικά επίπεδα για να εισαγάγουν μη γραμμικότητες. Η συνάρτηση ενεργοποίησης ReLU προστίθενται στα υπολειπόμενα μπλοκ μετά από και τα δύο συνελκτικά στρώματα. Για να ομαλοποιηθεί η είσοδος των κρυφών επιπέδων (που εξουδετερώνει το πρόβλημα της έκρηξης της κλίσης μεταξύ άλλων), εφαρμόζεται κανονικοποίηση βάρους σε κάθε συνελκτικό στρώμα. Προκειμένου να αποφευχθεί η υπερπροσαρμογή, η τακτοποίηση εισάγεται μέσω εγκατάλειψης μετά από κάθε συνελκτική στρώση σε κάθε υπολειπόμενο μπλοκ. Το παρακάτω σχήμα δείχνει το τελικό υπολειπόμενο μπλοκ. Ο αστερίσκος στη δεύτερη μονάδα ReLU υποδεικνύει ότι υπάρχει σε κάθε επίπεδο εκτός από το τελευταίο, καθώς θέλουμε η τελική μας έξοδος να μπορεί να λάβει και αρνητικές τιμές.

Η παρακάτω εικόνα 3.37 δείχνει το τελικό μας μοντέλο TCN με l ίσο με input length, k ίσο με kernel size, b ίσο με dilation base, $k \geq b$ και με ελάχιστο αριθμό υπολειπόμενων μπλοκ για πλήρη κάλυψη ιστορικού n , όπου το n μπορεί να υπολογιστεί από τις άλλες τιμές όπως εξηγήθηκε παραπάνω.[61]



Σχήμα 3.37: Σχηματική αναπαράσταση residual block έπειτα από Activation, Normalization, Regularization

[61]



Σχήμα 3.38: Σχηματική αναπαράσταση ολοκληρωμένου δικτύου TCN

[61]

Μέρος 

Πρακτικό Μέρος

Κεφάλαιο 4

Ανάλυση τεχνικών και αρχιτεκτονικών

Στο κεφάλαιο αυτό παρουσιάζεται η μελέτη που έγινε για την υλοποίηση του συστήματος. Αρχικά περιγράφονται τα δεδομένα που χρησιμοποιήθηκαν η προεπεξεργασία τους και στη συνέχεια η αρχιτεκτονική κάθε μοντέλου μάθησης. Για την επεξεργασία των δεδομένων, χρησιμοποιήθηκε η γλώσσα προγραμματισμού Python ενώ έγινε χρήση του TensorFlow και συγκεκριμένα του Keras, μιας βιβλιοθήκης του TensorFlow, που αποτελεί Διεπαφή Προγραμματισμού Εφαρμογών - ΔΠΕ (API) υψηλού επιπέδου για Μηχανική Μάθηση.[63]

4.1 Περιγραφή Συνόλου Δεδομένων Προβλήματος

4.1.1 Γνωριμία με το dataset

Τα χαρακτηριστικά BTC και τα δεδομένα τιμών είναι διαθέσιμα online στο <https://bitinfocharts.com> και τα αντλήσαμε από ένα csv αρχείο[64]. Συλλέχθηκαν περισσότερα από 700 χαρακτηριστικά(features) βασισμένα σε τεχνικούς δείκτες. Από αυτό το μεγάλο σύνολο δεδομένων χωρίσαμε σε τρία επιμέρους υποσύνολα μικρότερης χρονικής διάρκειας από το αρχικό dataset. Το πρώτο υποσύνολο(interval₁) περιέχει δεδομένα από 01/04/2013 έως 01/04/2016, το δεύτερο υποσύνολο(interval₂) από 01/04/2013 έως 01/04/2017 ενώ το τελευταίο υποσύνολο(interval₃) από 01/04/2013 έως 31/12/2019. Όλα τα υποσύνολα θα χρησιμοποιηθούν για end-of-day πρόβλεψη ενώ το τρίτο υποσύνολο που αποτελείται και από τα περισσότερα δεδομένα θα χρησιμοποιηθεί και για πρόβλεψη 7,30 και 90 ημερών. Για την πρόβλεψη τιμής της ημέρας, ο διαχωρισμός train-test είναι ο ίδιος εκτός από το ότι η στήλη τιμής μετατοπίζεται προς τα πάνω (ή ισοδύναμα, προς τα πίσω στον τομέα χρόνου) με βάση τον αριθμό απαιτούμενων ημερών. Για παράδειγμα, για τις προβλέψεις της 7ης ημέρας, η στήλη τιμής Price μετατοπίζεται προς τα πάνω κατά 7 ημέρες. Αυτό δίνει τη δυνατότητα στα μοντέλα παλινδρόμησης να μάθουν τη σχέση μεταξύ των χαρακτηριστικών και μελλοντικών τιμών. Τα χαρακτηριστικά είναι επεξηγηματικά χαρακτηριστικά που βασίζονται σε διαφορετικά στατιστικά δεδομένα σχετικά με τη λειτουργία του ίδιου του blockchain, καθώς και τους τεχνικούς δείκτες της αγοράς.

Τεχνικοί Δείκτες

Η πρόβλεψη για τις τιμές των κρυπτονομισμάτων έχει υποκείμενες αλληλεξαρτήσεις που είναι δύσκολο να κατανοηθούν και να μοντελοποιηθούν. Για παράδειγμα, υπάρχουν στατιστικοί

παράγοντες όπως η διακύμανση και η τυπική απόκλιση που αλλάζει με την πάροδο του χρόνου. Αυτές οι αλληλεξαρτήσεις εμφανίζονται ως τεχνικοί δείκτες. Οι τεχνικοί δείκτες που συναντάμε στο dataset είναι: Απλος Κινητός μέσος όρος (SMA), Εκθετικός Κινητός Μέσος (EMA), Δείκτης Σχετικής Ισχύος (RSI), Σταθμισμένος Κινητός Μέσος όρος (WMA), Τυπική Απόκλιση (STD), Διακύμανση (VAR), Τριπλή εκθετική κίνηση (TRIX) και Ρυθμός Αλλαγής(ROC). Αυτοί οι τεχνικοί δείκτες υπολογίζονται σε συνδυασμό με χρονικές περιόδους όπως 3, 7, 14, 30 και 90 μέρες. Οι τιμές κλεισίματος στο τέλος της ημέρας θεωρούνται ακατέργαστες τιμές. Τα ακατέργαστα χαρακτηριστικά, πάνω στα οποία αυτοί οι τεχνικοί δείκτες βασίζονται, δίνονται στο Σχήμα 4.1. Οι τεχνικοί δείκτες δείχνουν ιδιότητες που δεν βρίσκονται εύκολα στα ακατέργαστα χαρακτηριστικά: πράγματα όπως οι διακυμάνσεις και τα χαρακτηριστικά της σειράς τυπικών αποκλίσεων. Για παράδειγμα, δείχνουν πώς η τιμή BTC σχετίζεται με την τυπική απόκλιση των συναλλαγών ή του κατακερματισμού σε περιόδους 30 ημερών αμφότερα και όχι μόνο στις ακατέργαστες συναλλαγές και τους κατακερματισμούς.

Features	Description	Features	Description
Transactions	The number of sent and received Bitcoin payments	Median transaction fee	The median of transaction fees in Bitcoin
Block size	Transactional information cryptographically linked in the blockchain. The maximum block size is currently set at 1 megabyte	Average transaction fee	Each transaction can have an associated transaction fee determined by the sender. The transaction fee is received by the miners who verify the transaction. Transactions with higher fees incentivize the Bitcoin miners to process them sooner than transactions with lower fees
Sent from addresses	These are distinct Bitcoin addresses from which payments are made everyday	Block time	The time required to mine one block. Usually, it is around 10 min but can fluctuate depending on the hashrate of the network
Difficulty	The daily average mining difficulty. The difficulty is computed by the network after a specified number of blocks have been created so that the time it requires to mine a block remains around 10 min	Hashrate	The daily total computational capacity of the Bitcoin network. Hashrate indicates the speed of a computer in completing an operation
Average transaction value	The average value of the transactions in Bitcoin	Median transaction value	The median value of the transactions in Bitcoin
Mining profitability	The profitability in USD/day for 1 terahash per second (THash/s)	Active addresses	The number of unique addresses participating in a transaction by either sending or receiving Bitcoins
Sent BTC	The total Bitcoins sent daily	Top 100 to total	The ratio of Bitcoins stored in the top 100 accounts to all the other accounts of Bitcoin
Fee-to-reward ratio	The ratio of the fee sent in a transaction to the reward for verifying that transaction by the other users		

Σχήμα 4.1: Ακατέργαστα χαρακτηριστικά από τα οποία δημιουργούνται οι τεχνικοί δείκτες [65]

Κάθε υποσύνολο με τον αντίστοιχο χρονικό ορίζοντα πρόβλεψης χρησιμοποιείται για την επιλογή των τεχνικών δεικτών που χρησιμεύουν ως εισροές (προγνωστικοί παράγοντες) στην ανάλυση του δέντρου αποφάσεων.

4.1.2 Επιλογή Χαρακτηριστικών (Feature Selection)

Προεπεξεργασία Δεδομένων

Με στόχο τη βελτίωση της διαχείρισης του dataset, στις περιπτώσεις που λείπουν δεδομένα συμπληρώθηκαν με τη μέθοδο της γραμμικής παρεμβολής(Interpolation method) όπου είναι

δυνατόν ενώ στις υπόλοιπες περιπτώσεις χρησιμοποιήθηκαν οι πιο συχνά παρατηρούμενες τιμές. Η επιλογή χαρακτηριστικών, η οποία είναι ένα κρίσιμο μέρος της προεπεξεργασίας δεδομένων, είναι απαραίτητη για τη βελτίωση της απόδοσης του μοντέλου. Τα χαρακτηριστικά εξήχθησαν και κλαδεύτηκαν επαναληπτικά χρησιμοποιώντας ένα πλήθος διαφορετικών προσεγγίσεων. Πρώτον, η σημασία των χαρακτηριστικών προσδιορίστηκε χρησιμοποιώντας μια μέθοδο συνόλου που βασίζεται σε δάσος τυχαίας απόφασης (Random Forest Regressor). Στη συνέχεια το πλέον μειωμένο σύνολο χαρακτηριστικών πρέπει να ελεγχθεί για Πολυ-συγγραμμικότητα και αλληλοσυσχέτισεις. Η πολυσυγγραμμικότητα εμφανίζεται όταν δύο ή περισσότερες ανεξάρτητες μεταβλητές συσχετίζονται σε μεγάλο βαθμό μεταξύ τους σε ένα μοντέλο παλινδρόμησης. Αυτό σημαίνει ότι μια ανεξάρτητη μεταβλητή μπορεί να προβλεφθεί από μια άλλη ανεξάρτητη μεταβλητή σε ένα μοντέλο παλινδρόμησης. Για παράδειγμα, ύψος-βάρος, οικιακό εισόδημα-κατανάλωση νερού, χιλιόμετρα-τιμή αυτοκινήτου, χρόνος μελέτης-ελεύθερος χρόνος κ.λπ. Η πολυσυγγραμμικότητα μπορεί να ανιχνευθεί με διάφορες μεθόδους. Στην παρούσα εργασία, θα επικεντρωθούμε στον πιο συνηθισμένο - VIF (Variable Inflation Factors). Το VIF καθορίζει την ισχύ της συσχέτισης μεταξύ των ανεξάρτητων μεταβλητών. Προβλέπεται λαμβάνοντας μια μεταβλητή και παλινδρομώντας την έναντι κάθε άλλης μεταβλητής. Η βαθμολογία VIF μιας ανεξάρτητης μεταβλητής αντιπροσωπεύει πόσο καλά εξηγείται η μεταβλητή από άλλες ανεξάρτητες μεταβλητές. Η τιμή R^2 προσδιορίζεται για να μάθουμε πόσο καλά περιγράφεται μια ανεξάρτητη μεταβλητή από τις άλλες ανεξάρτητες μεταβλητές. Μια υψηλή τιμή R^2 σημαίνει ότι η μεταβλητή συσχετίζεται σε μεγάλο βαθμό με τις άλλες μεταβλητές. Αυτό αποτυπώνεται από το VIF που υποδηλώνεται παρακάτω:

$$VIF = \frac{1}{1 - R^2} \quad (4.1)$$

Έτσι, όσο πιο κοντά είναι η τιμή R^2 στο 1, τόσο μεγαλύτερη είναι η τιμή του VIF και τόσο μεγαλύτερη είναι η πολυσυγγραμμικότητα της συγκεκριμένης ανεξάρτητης μεταβλητής. Οι τιμές της μετρικής VIF αρχίζουν από 1 και δεν έχουν άνω όριο. Όταν το VIF = 1 δεν υπάρχει πολυσυγγραμμικότητα της ανεξάρτητης μεταβλητής με τις υπόλοιπες ενώ για τιμές VIF μεγαλύτερες από 5 ή 10 σημαίνουν υψηλή πολυσυγγραμμικότητα μεταξύ της ανεξάρτητης μεταβλητής και των υπολοίπων. Η απόρριψη μεταβλητών θα πρέπει να είναι μια επαναληπτική διαδικασία που ξεκινά με τη μεταβλητή που έχει τη μεγαλύτερη τιμή VIF και είναι μεγαλύτερη ενός κατωφλίου. Επειδή η τάση της αποτυπώνεται σε μεγάλο βαθμό από άλλες μεταβλητές αν κάνουμε drop αυτό το χαρακτηριστικό, θα παρατηρήσουμε ότι οι τιμές VIF και για άλλες μεταβλητές θα μειωθεί επίσης, αν και σε διαφορετικό βαθμό. [66] Τέλος, από τα χαρακτηριστικά που έχουν απομείνει ως τα πιο σημαντικά εφαρμόζουμε ένα score και για μεταβλητές που παρουσιάζουν μεγάλη εξάρτηση με τις υπόλοιπες τις εξαλείφουμε αφού μπορούμε από τις υψηλά σχετιζόμενες να πάρουμε την πληροφορία που χρειαζόμαστε. Αυτή τη διαδικασία την εφαρμόζουμε για κάθε interval με τον αντίστοιχο χρονικό ορίζοντα του. Έτσι, δημιουργούμε 6 διαφορετικά datasets με διαφορετικά features το καθένα.

4.1.3 Περαιτέρω επεξεργασία στα καινούργια datasets

Για όλα τα μοντέλα παλινδρόμησης, το σύνολο δεδομένων χωρίστηκε σε δύο sets: $train_{set}$ και $test_{set}$. Το 20 τοις εκατό των δεδομένων διατηρήθηκαν για τεστ και το 80 τοις εκατό των δεδομένων χρησιμοποιήθηκαν για εκπαίδευση.

Για την εκπαίδευση των μοντέλων, τα χαρακτηριστικά κλιμακώθηκαν χρησιμοποιώντας την ισχυρή κλιμάκωση (robust scaler). Όταν τα δεδομένα αποτελούνται συχνά από πολλές διαφορετικές μεταβλητές εισόδου ή χαρακτηριστικά (στήλες) και το καθένα μπορεί να έχει διαφορετικό εύρος τιμών ή μονάδων μέτρησης είναι αναγκαία η χρήση Robust Scaler. Εάν υπάρχουν μεταβλητές εισόδου που έχουν πολύ μεγάλες τιμές σε σχέση με τις άλλες μεταβλητές εισόδου, αυτές οι μεγάλες τιμές μπορεί να κυριαρχούν ή να παραμορφώνουν ορισμένους αλγόριθμους μηχανικής εκμάθησης. Το αποτέλεσμα είναι ότι οι αλγόριθμοι δίνουν το μεγαλύτερο μέρος της προσοχής τους στις μεγάλες τιμές και αγνοούν τις μεταβλητές με μικρότερες τιμές. Ως εκ τούτου, είναι φυσιολογικό να κλιμακώνονται οι μεταβλητές εισόδου σε ένα κοινό εύρος ως τεχνική προετοιμασίας δεδομένων πριν από την τοποθέτηση ενός μοντέλου. [67] Έπειτα εφαρμόζεται και ο MinMax Scaler ο οποίος μετασχηματίζει τις τιμές στο εύρος (0,1).

4.2 Αρχιτεκτονικές Μοντέλων Χρονοσειρών

4.2.1 Επαναληπτικά Μοντέλα

Όλα τα επαναληπτικά μοντέλα αναμένουν ότι το σχήμα εισόδου θα είναι τρισδιάστατο με [δείγματα, χρονικά βήματα, χαρακτηριστικά] ή [samples, time steps, features], επομένως, πρέπει να αναδιαμορφώσουμε το μεμονωμένο δείγμα εισόδου πριν κάνουμε την πρόβλεψη.

RNN

Ο πρώτος αλγόριθμος που εφαρμόστηκε μέσω των νευρωνικών δικτύων είναι ο RNN. Για την εκπαίδευσή του χρησιμοποιήθηκαν τα δεδομένα εκπαίδευσης (training data) και τα δεδομένα ελέγχου (test data) όπως έχουν χωριστεί από το σύνολο δεδομένων μας. Το νευρωνικό δίκτυο RNN, έχει μια πληθώρα παραμέτρων που πρέπει να προσδιορίσουμε πριν γίνει η εκπαίδευση του. Κάποιες από αυτές είναι, ο αριθμός επιπέδων του RNN, ο αριθμός των μονάδων που θα έχει το κάθε επίπεδο, η συνάρτηση ενεργοποίησης του. Ακόμα, για την εκπαίδευση του μοντέλου θα πρέπει να οριστεί ο αλγόριθμος βελτιστοποίησης optimization αλλά και το batch size, δηλαδή μετά από πόσα πακέτα δεδομένων μορφής (batch size, timesteps, features) θα ανανεώνονται οι κλίσεις (gradients). Ως προς τη δομή του νευρωνικού δικτύου υπάρχουν επιπλέον υπερπαραμέτροι που καθορίζουν αν το επαναληπτικό μοντέλο θα επιστρέφει το σύνολο των κρυφών καταστάσεων, Η, για κάθε χρονικό βήμα t (return sequences=True) ή αν θα επιστρέφει μόνο το διάλυμα κρυφής κατάστασης του τελευταίου βήματος χρόνου (return sequences=False). Στον πίνακα 4.1 που ακολουθεί παρατηρούμε στην πρώτη στήλη τα διάφορα είδη υπερπαραμέτρων που ελέχθησαν, στη δεύτερη τις διάφορες εναλλακτικές και στην τρίτη τη βέλτιστη τιμή που πήραμε μετά από δοκιμές για τη βελτιστοποίηση των παραμέτρων.

Αναρομικά Δίκτυα LSTM και GRU

Υπερπαράμετροι	Συνδυασμός	Βέλτιστο
Neurons	[16,32,64,128,256,512]	256
Dropout	[None,0.1,0.2]	0.1
Dense Layers	[1,2,3]	1
Batch size	[8,16,32,64]	16
Optimizer	[Adam,Rmsprop,SGD]	Adam
Activation function	[tanh,ReLU]	ReLU
Loss function	[mse,mae,logcosh]	logcosh
epochs	[2000,500,1000,2000,5000]	5000

Πίνακας 4.1: Υπερπαράμετροι μοντέλου RNN

Στη συνέχεια αναπτύχθηκαν τα μοντέλα LSTM και GRU των οποίων τα οι στρώσεις δικτύου περιορίζονται σε μονάδα. Πρόκειται για την απλούστερη μορφή αυτών των δικτύων. Οι υπερπαράμετροι που ελέχθησαν για αυτά τα δύο μοντέλα είναι ίδιες και μπορούμε να τις διακρίνουμε στον Πίνακα 5.4.

Αμφίδρομα και Βαθιά Επαναληπτικά Μοντέλα

Αυτά τα μοντέλα αποτελούνται από παραπάνω από μία στρώση δικτύου, πρόκειται δηλαδή για στοιβαγμένα δίκτυα το ένα έπειτα από το άλλο. Αυτό σημαίνει ότι η πολυπλοκότητα αυξάνεται και οι υπολογισμοί γίνονται πιο χρονοβόροι, οδηγώντας πολλές φορές σε πιο αξιόπιστα μοντέλα συγκριτικά με τα απλά επαναληπτικά μοντέλα. Όσον αφορά τα βαθιά επαναληπτικά δίκτυα, τα οποία μπορεί να αποτελούνται είτε από απλές στρώσεις δικτύων είτε από αμφίδρομες, οι επιλογές των υπερπαραμέτρων που έχουμε είναι η ίδια με την προσθήκη του αριθμού των επιπέδων-στρώσεων από επαναληπτικά δίκτυα που χρησιμοποιήθηκαν. Θα πρέπει επίσης να αναφερθεί ότι στην περίπτωση αυτών των μοντέλων στις ενδιάμεσες στρώσεις η μεταβλητή `return sequences` τίθεται `True` προκειμένου να είναι εφικτή η είσοδος ενός 3D τανυστή στην επόμενη στρώση ενώ στην τελευταία στρώση αυτή η παράμετρος συνηθίζεται να τίθεται ίση με `False` αφού μας ενδιαφέρει η τελευταία κρυφή κατάσταση του δικτύου.

Στους παρακάτω πίνακες 5.2 και 5.3 αναπαρίστανται οι υπερπαράμετροι με τις οποίες δουλέψαμε για τα Αμφίδρομα Επαναληπτικά Δίκτυα και οι βέλτιστες επιλογές των δικτύων αυτών έκαστος.

Υπερπαράμετροι	Συνδυασμός	Βέλτιστο
Neurons	[16,32,64,128,256,512]	256
Dropout	[None,0.1,0.2]	0.1
Dense Layers	[1,2,3]	1
Batch size	[8,16,32,64]	32
Optimizer	[Adam,Rmsprop,SGD]	Adam
Activation function	[tanh,ReLU]	ReLU
Loss function	[mse,mae,logcosh]	logcosh
epochs	[2000,500,1000,2000,5000]	5000

Πίνακας 4.2: Υπερπαράμετροι αμφίδρομου μοντέλου GRU

Υπερπαράμετροι	Συνδυασμός	Βέλτιστο
Neurons	[16,32,64,128,256,512]	128
Dropout	[None,0.1,0.2]	0.1
Dense Layers	[1,2,3]	1
Batch size	[8,16,32,64]	32
Optimizer	[Adam,Rmsprop,SGD]	Adam
Activation function	[tanh,ReLU]	ReLU
Loss function	[mse,mae,logcosh]	logcosh
epochs	[2000,500,1000,2000,5000]	5000

Πίνακας 4.3: Υπερπαράμετροι αμφίδρομου μοντέλου LSTM

Υπερπαράμετροι	Συνδυασμός	Βέλτιστο
Neurons	[16,32,64,128,256,512]	256
Dropout	[None,0.1,0.2]	0.2
Dense Layers	[1,2,3]	1
Batch size	[8,16,32,64]	32
Optimizer	[Adam,Rmsprop,SGD]	Adam
Activation function	[tanh,ReLU]	ReLU
Loss function	[mse,mae,logcosh]	mae
epochs	[2000,500,1000,2000,5000]	5000

Πίνακας 4.4: Υπερπαράμετροι απλών μοντέλων LSTM και GRU

Υπερπαράμετροι	Συνδυασμός	Βέλτιστο
Recurrent layers	[1,2,3]	2
Neurons 1ου σταδίου	[16,32,64,128,256]	256
Neurons 2ου σταδίου	[16,32,64,128,256,512]	256
Dropout	[None,0.1,0.2]	0.1
Dense Layers	[1,2,3]	1
Batch size	[8,16,32,64]	16 ή 32
Optimizer	[Adam,Rmsprop,SGD]	Adam
Activation function	[tanh,ReLU]	ReLU
Loss function	[mse,mae,logcosh]	logcosh ή mae
epochs	[2000,500,1000,2000,5000]	5000 ή 4000

Πίνακας 4.5: Υπερπαράμετροι μοντέλων deep GRU και deep LSTM

4.2.2 Συνελικτικά Μοντέλα

Η επόμενη μέθοδος που εφαρμόστηκε ήταν μέσω των νευρωνικών δικτύων CNN και TCN. Για την εκπαίδευσή τους χρησιμοποιήθηκαν τα δεδομένα εκπαίδευσης (training data) και τα δεδομένα ελέγχου (test data) όπως ακριβώς και προηγουμένως. Προκειμένου να πετύχουμε τα βέλτιστα αποτελέσματα για την εκπαίδευση της αρχιτεκτονικής μας χρειάστηκε να ελέγξουμε αρκετές υπερπαραμέτρους. Οι βασικές υπερπαραμέτροι που σχετίζονται με τη διαδικασία εκπαίδευσης είναι οι ίδιες και με τα προηγούμενα πειράματα. Για το TCN επίπεδο μελετήσαμε την παράμετρο use skip connectios, η οποία καθορίζει αν θα έχουμε σύνδεση μεταξύ της εισόδου ενός υπολειμματικού μπλοκ με την έξοδο του και το μέγεθος

πυρήνα που χρησιμοποιούν τα επίπεδα των μονοδιάστατων CNNs. Ακόμη, ελέγχθηκε ο αριθμός των φίλτρων (nb size) που χρησιμοποιήθηκε από τα επίπεδα των CNNs (η αντίστοιχη παράμετρος στα LSTMs είναι το unit). Άλλη μια παράμετρος είναι η παράμετρος του γεμίσματος (padding) που καθορίζει αν α είναι το γέμισμα αιτιώδης όπως στη θεωρητική υλοποίηση ή όχι. Τέλος, πειραματιστήκαμε και με τον ρυθμό εγκατάλειψης (dropout rate) στο επίπεδο εγκατάλειψης καθώς και τις συναρτήσεις ενεργοποίησης τόσο στο επίπεδο του TCN όσο και στα πυκνά επίπεδα (Dense layers). Στο πίνακα παρουσιάζονται αναλυτικά οι τιμές που δοκιμάστηκαν αλλά και οι βέλτιστες που προέκυψαν έπειτα από δοκιμή.

Υπερπαράμετροι	Συνδυασμός	Βέλτιστο
TCN layers	[1,2,3]	1
kernel size	[2,3,4,5]	3
padding	[causal,same, valid]	causal
Skip Connectors	[True,False]	True
True Dilations	[(1,2,4,8), (1,2,4,8,16), (1,2,4,8,16,32)]	[1,2,4,8,16,32]
Dense Layers	[1,2,3]	1
Batch size	[8,16,32,64]	16 ή 32
Optimizer	[Adam,Rmsprop,SGD]	Adam
Activation function	[tanh,ReLU]	ReLU
Loss function	[mse,mae,logcosh]	mse
epochs	[2000,500,1000,2000,5000]	5000

Πίνακας 4.6: Υπερπαράμετροι μοντέλου TCN

4.2.3 Υβριδικά-Συνελικτικά Μοντέλα

Τα μοντέλα αυτά συνίστανται από συνδυασμό επαναληπτικών και συνελικτικών δικτύων. Η διαφορά με τα προηγούμενα μοντέλα είναι ότι εκτός από τις παραμέτρους του επαναληπτικού μοντέλου θα πρέπει να διαλέξουμε και τις παραμέτρους του συνελικτικού. Μερικές από τις υπερπαραμέτρους του συνελικτικού δικτύου είναι ο αριθμός των φίλτρων, το μέγεθος πυρήνα(kernel_size), η χρήση padding και η συνάρτηση ενεργοποίησης του συνελικτικού στρώματος. Οι παράμετροι για Optimizer και Loss Function παρέμειναν ίδιες όπως πριν. Όπως θα παρατηρήσουμε και για τα τρία διαφορετικά μοντέλα οι βέλτιστες υπερπαραμέτροι δε διαφέρουν ιδιαίτερα από μοντέλο σε μοντέλο.

Υπερπαράμετροι	Συνδυασμός	Βέλτιστο
CNN Filters	[16,32,64,128,256]	128
Kernel Size	[2,3,4]	2
CNN activation	[relu,softmax]	relu
Padding	[same,valid,causal]	causal/causal
RNN Neurons	[16,32,64,128,256,512]	256
RNN Activation	[tanh, ReLU]	relu
Dense Layers	[1,2,3]	2
Dropout	[None,0.1,0.2]	0.1/0.1
Batch size	[8,16,32,64]	32
Optimizer	[Adam,Rmsprop,SGD]	Adam
Activation function	[tanh,ReLU]	ReLU
Loss function	[mse,mae,logcosh]	logcosh
epochs	[2000,500,1000,2000,5000]	2000

Πίνακας 4.7: Υπερπαράμετροι CNN-RNN

Υπερπαράμετροι	Συνδυασμός	Βέλτιστο
CNN dFilters	[16,32,64,128,256]	128
Kernel Size	[2,3,4]	2
CNN activation	[relu,softmax]	relu
Padding	[same,valid,causal]	causal/causal
LSTM Neurons	[16,32,64,128,256,512]	256
LSTM Activation	[tanh, ReLU]	relu
Dense Layers	[1,2,3]	2
Dropout	[None,0.1,0.2]	0.1/0.1
Batch size	[8,16,32,64]	64
Optimizer	[Adam,Rmsprop,SGD]	Adam
Activation function	[tanh,ReLU]	ReLU
Loss function	[mse,mae,logcosh]	logcosh
epochs	[100,200,300,400,500,1000]	2000

Πίνακας 4.8: Υπερπαράμετροι CNN-LSTM

Υπερπαράμετροι	Συνδυασμός	Βέλτιστο
CNN dFilters	[16,32,64,128,256]	128
Kernel Size	[2,3,4]	2
CNN activation	[relu,softmax]	relu
Padding	[same,valid,causal]	causal/causal
GRU Neurons	[16,32,64,128,256,512]	128
RNN Activation	[tanh, ReLU]	relu
Dense Layers	[1,2,3]	2
Dropout	[None,0.1,0.2]	0.1/0.1
Batch size	[8,16,32,64]	32
Optimizer	[Adam,Rmsprop,SGD]	Adam
Activation function	[tanh,ReLU]	ReLU
Loss function	[mse,mae,logcosh]	logcosh
epochs	[100,200,300,400,500,1000]	2000

Πίνακας 4.9: Υπερπαράμετροι CNN-GRU

Πειραματικά αποτελέσματα

Για την αξιολόγηση της απόδοσης των μοντέλων παλινδρόμησης, χρησιμοποιούνται οι ακόλουθες μετρήσεις: μέσο απόλυτο σφάλμα (MAE), ρίζα μέσου τετραγωνικού σφάλματος (RMSE) και μέσος όρος απόλυτου ποσοστού σφάλματος (MAPE). Ένα μοντέλο με χαμηλές τιμές στις μετρικές απόδοσης MAE, MAPE και RMSE είναι επιθυμητό. Στο πλαίσιο της πρόβλεψης της τιμής του BTC, για παράδειγμα, ένα MAE=5 σημαίνει ότι η προβλεπόμενη τιμή είναι ± 5 USD από την πραγματική τιμή. Το RMSE υποδεικνύει την εξάπλωση των σφαλμάτων πρόβλεψης. Ένα μοντέλο που προβλέπει περιστασιακά ακανόνιστες τιμές θα έχουν υψηλότερη τιμή RMSE, αν και μπορεί να έχει ακόμα χαμηλότερο MAE. Επομένως, τα μοντέλα θα πρέπει να αξιολογηθούν με σεβασμό και στις τρεις μετρήσεις.

5.1 Μετρικές Αξιολόγησης

5.1.1 Μετρικές Τετραγωνικού Σφάλματος

Mean Squared Error (MSE)

Στις στατιστικές αναλύσεις, το μέσο τετραγωνικό σφάλμα (MSE) μετρά τον μέσο όρο των τετραγώνων των σφαλμάτων—δηλαδή τη μέση τετραγωνική διαφορά μεταξύ των εκτιμώμενων τιμών και της πραγματικής αξίας. Καθώς προκύπτει από το τετράγωνο της Ευκλείδειας απόστασης, είναι πάντα μια θετική τιμή που μειώνεται καθώς το σφάλμα πλησιάζει το μηδέν. Η εξίσωση που περιγράφει αυτή τη Σχέση 5.1 δίνεται παρακάτω:

$$MSE = \frac{1}{n} \cdot \sum_{i=1}^n (Y_i - \tilde{Y}_i)^2 \quad (5.1)$$

Root Mean Squared Error (RMSE)

Το (RMSE) είναι η τετραγωνική ρίζα του μέσου όρου του τετραγώνου του συνόλου του σφάλματος. Η χρήση του RMSE είναι πολύ συνηθισμένη και θεωρείται μια εξαιρετική μετρική σφάλματος για αριθμητικές προβλέψεις. Στη σχέση 5.2 όπου Y_i είναι οι i -οστη παρατήρηση, η αντίστοιχη πρόβλεψη και n ο αριθμός των παρατηρήσεων που είναι διαθέσιμες για ανάλυση. [68]

$$RMSE = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (Y_i - \tilde{Y}_i)^2} \quad (5.2)$$

5.1.2 Μετρικές Απόλυτου Σφάλματος

Mean Absolute Error (MAE)

Το μέσο απόλυτο σφάλμα είναι μια μετρική αξιολόγησης μοντέλου που χρησιμοποιείται για μοντέλα παλινδρόμησης. Το MAE μετρά το μέσο μέγεθος των σφαλμάτων σε ένα σύνολο προβλέψεων, χωρίς να λαμβάνει υπόψη την κατεύθυνσή τους. Μετρά την ακρίβεια για συνεχείς μεταβλητές. Με λίγα λόγια το MAE είναι ο μέσος όρος του δείγματος επαλήθευσης των απόλυτων τιμών των διαφορών μεταξύ της πρόβλεψης και της αντίστοιχης παρατήρησης. Το MAE είναι μια γραμμική βαθμολογία που σημαίνει ότι όλες οι επιμέρους διαφορές σταθμίζονται εξίσου στο μέσο όρο. [69]. Η μετρική υπολογίζεται ως:

$$MAE = \frac{1}{n} \cdot \sum_{i=1}^n |(Y_i - \tilde{Y}_i)|^2 \quad (5.3)$$

Mean Absolute Percentage Error (MAPE)

Παρόλα αυτά θα πρέπει να ποσοτικοποιήσουμε την παραπάνω μετρική για να γίνεται πιο εύκολα κατανοητή. Έτσι, ορίζουμε το μέσο απόλυτο ποσοστιαίο σφάλμα (Mean Absolute Percentage Error - (MAPE)) ως εξής:

$$MAPE = \left(\frac{1}{n} \cdot \sum_{i=1}^n \left| \frac{(Y_i - \tilde{Y}_i)^2}{\tilde{Y}_i} \right| \right) \cdot 100\% \quad (5.4)$$

5.1.3 Μετρικές Ακρίβειας

Συντελεστής Προσδιορισμού

Ο συντελεστής προσδιορισμού, που συμβολίζεται με R^2 είναι η αναλογία της διακύμανσης στην εξαρτημένη μεταβλητή που προβλέπεται από τις ανεξάρτητες μεταβλητές.

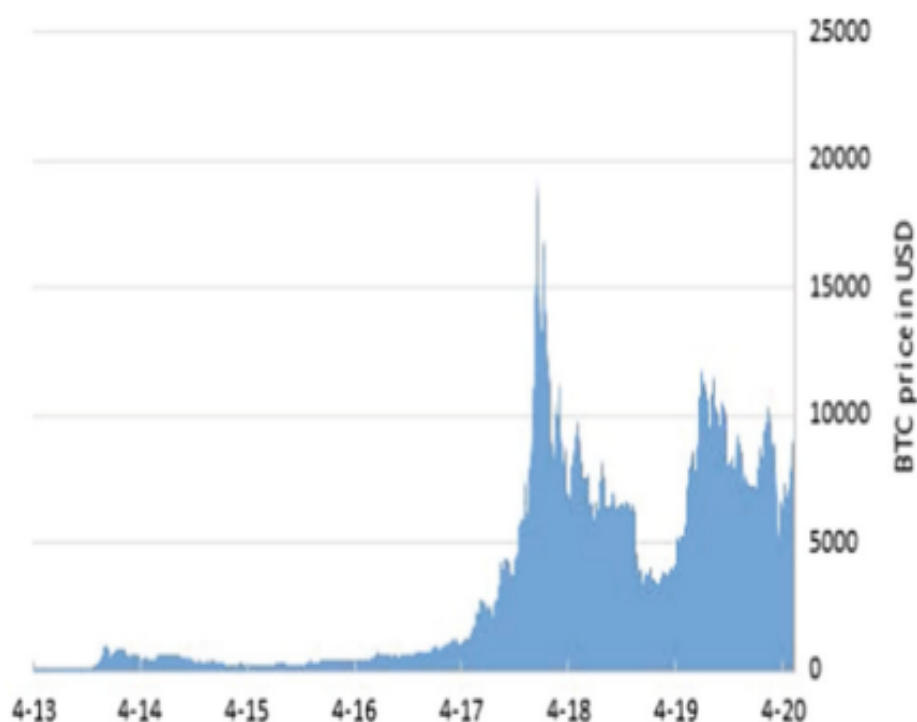
Είναι μια στατιστική που χρησιμοποιείται στο πλαίσιο στατιστικών μοντέλων των οποίων ο κύριος σκοπός είναι είτε η πρόβλεψη μελλοντικών αποτελεσμάτων είτε ο έλεγχος υποθέσεων, με βάση άλλες σχετικές πληροφορίες. Παρέχει ένα μέτρο του πόσο καλά παρατηρούμενα αποτελέσματα αναπαράγονται από το μοντέλο, με βάση το ποσοστό της συνολικής διακύμανσης των αποτελεσμάτων που εξηγείται από το μοντέλο.

Αυτός ο συντελεστής μας δείχνει την αλληλοσυσχέτιση των προβλέψεων του μοντέλου με την ανεξάρτητη μεταβλητή (παρατηρήσεις) σε κλίμακα του 0 - 100%. Ο μαθηματικός τύπος μπορεί να δοθεί από την παρακάτω σχέση:

$$R^2 = \frac{\text{model' s variance}}{\text{Total variance}} \quad (5.5)$$

5.2 Αποτελέσματα και αξιολόγηση Μοντέλων

Στους Πίνακες 5.1, 5.2, 5.3, 5.4, 5.5 φαίνονται αναλυτικά τα σκορ όλων των μοντέλων ως προς τις μετρικές MAE, RMSE και MAPE. Σε κάθε μέθοδο τα καλύτερα σκορ τονίζονται με έντονη γραφή. Στις εικόνες 5.1, 5.2, 5.3 και 5.4 μπορούμε να δούμε σε διαγράμματα τις καμπύλες προβλεπόμενων-πραγματικών τιμών των μοντέλων που έδωσαν τα καλύτερα αποτελέσματα και παρουσιάζονται στους προαναφεθέντες Πίνακες.



Σχήμα 5.1: Τιμή του BTC από το 2013 έως το 2020 [65]

Στο διάστημα I (interval 1), από τον Απρίλιο του 2013 έως τον Απρίλιο του 2016, οι τιμές BTC δεν παρουσίασαν μεγάλη αστάθεια όπως φαίνεται στο Σχήμα 5.1. Σε αυτό το διάστημα οι μετρικές απόδοσης όλων των μοντέλων παρουσίασαν πολύ κοντινές τιμές. Με μικρή διαφορά το μοντέλο που εμφάνισε το μικρότερο δείκτη MAE=7.091 και το μικρότερο MAPE=2.21% είναι το **Bidirectional GRU** ενώ σημείωσε και μία από τις μικρότερες τιμές RMSE=12.73. Το μοντέλο TCN σημείωσε τη μικρότερη τιμή RMSE. Ο υψηλότερος MAPE αναφέρεται από το μοντέλο Simple GRU με 12.7% ενώ το μεγαλύτερο MAE=8.608 σημείωσε το μοντέλο CNN-LSTM. Τέλος, το μοντέλο CNN-RNN παρουσίασε τη μεγαλύτερη τιμή RMSE=14.98.

Στο διάστημα II (Interval 2), από τον Απρίλιο του 2013 έως τον Απρίλιο του 2017, παρατηρούμε ότι οι τιμές του BTC αρχίζουν να αυξάνονται αισθητά προς το τέλος της περιόδου. Παρ' όλα αυτά σε γενικές γραμμές είναι σχετικά σταθερό όπως το Interval I. Οι τιμές των μετρικών απόδοσης δεν έχουν αισθητή διαφορά από εκείνες του interval1. Παρ' όλα αυτά τη μικρότερη τιμή του δείκτη MAE=7.09 την εμφάνισε το μοντέλο Simple LSTM, την ελάχιστη τιμή RMSE=16.46 το μοντέλο Simple GRU ενώ το μοντέλο LSTM παρουσίασε το ελάχιστο MAPE=2.190%. Το CNN-LSTM παρουσίασε τις χειρότερες τιμές δεικτών.

Στο διάστημα III (Interval 3), από τον Απρίλιο του 2013 έως τον Δεκέμβριο του 2019 και για πρόβλεψη επόμενης ημέρας, παρατηρούμε ότι οι τιμές του BTC παρουσιάζουν μεγάλη μεταβλητότητα και αστάθεια γεγονός που καθίσταται πρόκληση για το πόσο καλά θα ανταποκριθούν τα μοντέλα. Οι τιμές των μετρικών απόδοσης εκτοξεύονται σε σχέση με τις τιμές που είχαμε στα Interval 1 και interval 2. Για παράδειγμα τη μικρότερη τιμή του δείκτη MAE=61.52 την εμφάνισε το μοντέλο TCN, την ελάχιστη τιμή RMSE=136.20 το μοντέλο TCN

ενώ το μοντέλο BiLSTM παρουσίασε το ελάχιστο MAPE=3.02%. Το Simple-LSTM παρουσίασε τις χειρότερες τιμές δεικτών. Το TCN είναι το πιο αποδοτικό μοντέλο για πρόβλεψη με χρονικό ορίζοντα της επόμενης μέρας.

Ενώ λοιπόν στα Interval 1 και Interval 2 οι μετρικές απόδοσης είναι εξαιρετικά χαμηλές, γεγονός που δείχνει ότι οι προβλέψεις είναι πολύ προσεγγιστικές στις πραγματικές, απεναντίας, στο Interval 3 για ορίζοντα πρόβλεψης 7 ημερών οι μετρικές απόδοσης έχουν μια αυξητική τάση. Αυτό είναι αναμενόμενο καθώς οι τιμές του BTC παρουσιάζουν ραγδαίες αυξομειώσεις, κάτι που δυσχεραίνει την απόδοση των μοντέλων που εκπαιδεύουμε. Παρόλα αυτά και για 7-μέρες χρονικό ορίζοντα πρόβλεψης τα μοντέλα παλινδρόμησης προσεγγίζουν αρκετά "καλά" τις πραγματικές τιμές. Το μοντέλο GRU παρουσιάζει το ελάχιστο MAE=88.53, το μοντέλο TCN έχει τον μικρότερο δείκτη RMSE=187.09 και το μοντέλο BiGRU έχει τον χαμηλότερο δείκτη MAPE=4.33%. Αντίθετα τα μοντέλα Simple RNN και CNN-RNN έχουν τους υψηλότερους δείκτες και συνεπώς τη χειρότερη απόδοση.

Για τον χρονικό ορίζοντα πρόβλεψης 30 ημερών παρατηρούμε μια περαιτέρω αύξηση των μετρικών απόδοσης αλλά όχι τόσο μεγάλη όση αναμέναμε μετά την 7-ήμερη πρόβλεψη και τις αντίστοιχες μετρικές. Το μοντέλο Bi-LSTM έχει τη μικρότερη τιμή MAE=88.96 και MAPE=4.91%. Το μοντέλο TCN φαίνεται να έχει τη μικρότερη RMSE τιμή ενώ τα μοντέλα Simple RNN και Simple GRU έχουν τη χειρότερη απόδοση.

Για τον χρονικό ορίζοντα 90 ημερών το μοντέλο δε δείχνει ιδιαίτερη επιδείνωση σε σχέση με την πρόβλεψη 30 ημερών, συμπεριφέρεται με παρόμοιο τρόπο ενώ σε κάποιες περιπτώσεις οι μετρικές δείχνουν και κάποια βελτίωση από την προηγούμενη πρόβλεψη. Το μοντέλο που αποδίδει καλύτερα είναι το GRU με χαμηλότερο δείκτη MAE=98.84, το BiGRU με δείκτη RMSE=217.37 και το CNN-LSTM με μικρότερη τιμή MAPE=4.42%. Αντίθετα τα μοντέλα Simple RNN και Simple LSTM έχουν τους υψηλότερους δείκτες και συνεπώς τη χειρότερη απόδοση.

Μοντέλο	MAE	RMSE	MAPE
Simple RNN	8.158108	14.372104	2.617568
Simple LSTM	8.108963	13.033773	2.77248
Simple GRU	8.253673	13.437015	2.933764
LSTM	7.5066	13.20578	2.384818
GRU	7.352739	13.099481	2.298477
BiLSTM	7.337561	13.023606	2.290239
BiGRU	7.091009	12.736201	2.211676
TCN	7.693732	12.666187	2.511594
CNN-RNN	8.567291	14.984704	2.772404
CNN-LSTM	8.608418	14.70643	2.716578
CNN-GRU	8.204023	14.411878	2.628904

Πίνακας 5.1: Μετρικές απόδοσης για το Interval 1

Όπως παρατηρούμε στα παρακάτω γραφήματα 5.2,5.3,5.4,5.5,5.6 και 5.7 όπου αναπαρίστανται οι τιμές πρόβλεψης και οι πραγματικές τιμές των BTC, σε όλα τα διαστήματα αλλά και σε όλους τους χρονικούς ορίζοντες πρόβλεψης οι προβλέψεις είναι αρκετά ικανοποιητικές αφού οι γραφικές παραστάσεις σε πολλά σημεία συμπίπτουν. Ένα πρόβλημα

Μοντέλο	MAE	RMSE	MAPE
Simple RNN	8.973143	16.863983	2.26944
Simple LSTM	8.759564	16.532671	2.234573
Simple GRU	9.194032	16.468202	2.330984
LSTM	8.788158	17.61168	2.190988
GRU	8.882583	17.251038	2.193604
BiLSTM	8.89036	17.067012	2.283041
BiGRU	8.790593	16.754536	2.249394
TCN	9.077539	16.66291	2.355532
CNN-RNN	9.415475	17.503809	2.497659
CNN-LSTM	10.065247	19.784362	2.594564
CNN-GRU	9.051724	17.58063	2.248444

Πίνακας 5.2: Μετρικές απόδοσης για το Interval 2

Μοντέλο	MAE	RMSE	MAPE
Simple RNN	70.083855	168.720888	3.08591
Simple LSTM	83.608207	187.983993	4.125778
Simple GRU	80.666549	185.452033	4.094719
LSTM	65.123712	165.062747	2.964271
GRU	67.403052	160.806129	2.94709
BiLSTM	65.22873	159.210777	3.027591
BiGRU	65.879502	160.945003	3.053086
TCN	61.52406	136.200515	3.512636
CNN-RNN	75.237509	182.838267	3.672533
CNN-LSTM	74.421479	175.652198	4.298847
CNN-GRU	79.01724	183.132027	4.423437

Πίνακας 5.3: Μετρικές απόδοσης για το Interval 3

Μοντέλο	MAE	RMSE	MAPE
Simple RNN	175.315739	381.159195	6.946289
Simple LSTM	113.613855	234.896344	6.444
Simple GRU	154.594405	323.553746	7.23869
LSTM	104.429432	234.139283	5.079625
GRU	88.536938	204.344249	4.512378
BiLSTM	98.185932	223.736799	5.084289
BiGRU	90.493427	213.163971	4.338806
TCN	99.692691	187.095107	5.800371
CNN-RNN	137.253363	304.434478	6.436186
CNN-LSTM	103.619961	216.16476	6.350573
CNN-GRU	129.014865	280.544072	6.381098

Πίνακας 5.4: Μετρικές απόδοσης για το Interval 3 με 7-ήμερη πρόβλεψη

που παρατηρείται σε όλα τα γραφήματα είναι η αδυναμία της ακριβούς πρόβλεψης ακραίων τιμών (outliers). Οι προβλέψεις προσπαθούν να προσεγγίσουν την ακραία τιμή έως κάποιο

Μοντέλο	MAE	RMSE	MAPE
Simple RNN	201.604757	428.591431	10.830014
Simple LSTM	122.509877	223.455714	9.907053
Simple GRU	149.050163	267.572158	15.133951
LSTM	108.15502	223.720425	5.594543
GRU	95.571665	202.028221	5.166491
BiLSTM	88.967939	199.782562	4.910627
BiGRU	108.328395	222.534152	6.48084
TCN	97.196161	88.447879	6.391606
CNN-RNN	140.982097	283.135831	7.645444
CNN-LSTM	99.014274	204.621183	5.528128
CNN-GRU	145.432984	324.597881	6.77233

Πίνακας 5.5: Μετρικές απόδοσης για το Interval 3 30-ημέρων πρόβλεψη

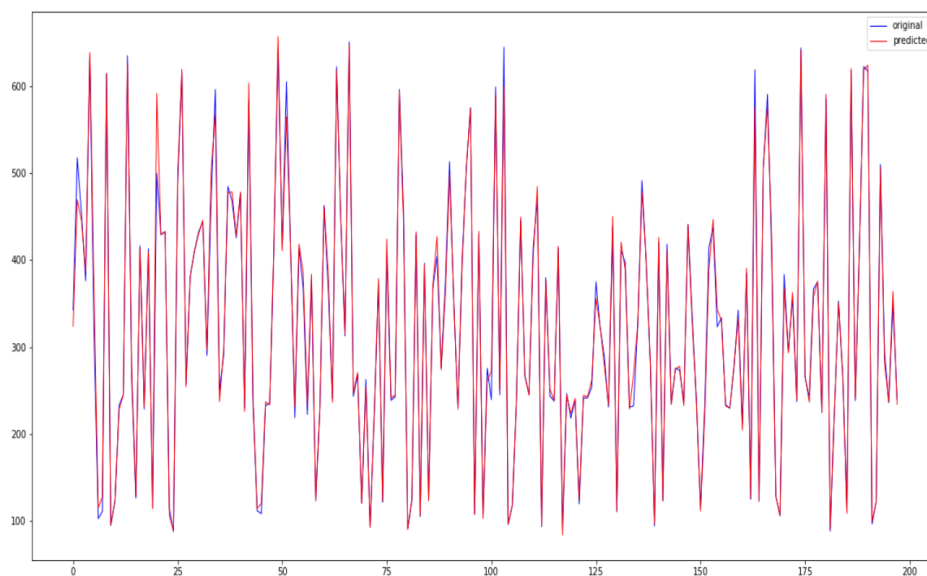
Μοντέλο	MAE	RMSE	MAPE
Simple RNN	185.043058	364.182612	9.720179
Simple LSTM	144.855931	265.899717	11.230097
Simple GRU	152.024862	285.421809	11.202169
LSTM	99.375496	220.252472	5.472118
GRU	98.843967	228.025651	4.748213
BiLSTM	134.447002	252.262074	10.952218
BiGRU	100.113445	217.372419	5.122996
TCN	121.629886	243.33387	7.105039
CNN-RNN	135.720899	294.194379	7.500999
CNN-LSTM	103.740988	251.53621	4.427266
CNN-GRU	126.751344	276.462627	6.568654

Πίνακας 5.6: Μετρικές απόδοσης για το Interval 3 90-ημερών πρόβλεψη

σημείο αλλά αυτό δεν επιτυγχάνεται πάντα.

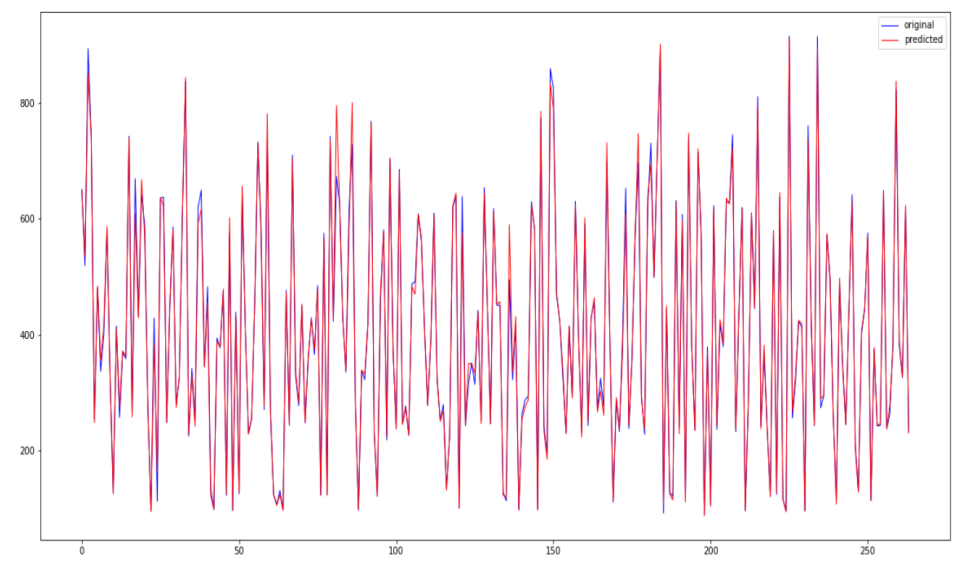
Από τα bar plots στα Σχήματα 5.8 και 5.9 παρατηρούμε τους δείκτες MAE και MAPE αντίστοιχα για τα 3 intervals με EOD χρονικό ορίζοντα πρόβλεψης. Για το δείκτη MAE παρατηρούμε ότι είναι πολύ μικρός για τα πρώτα 2 intervals ενώ για το τρίτο εκτοξεύονται οι τιμές του. Ο δείκτης MAPE ενώ για το Interval 3 έχει την υψηλότερη τιμή για το Interval 2 φαίνεται να έχει χαμηλότερες τιμές απ' ό,τι για το Interval 1. Αυτό μας δείχνει ότι η μέση διαφορά μεταξύ της προβλεπόμενης τιμής και της πραγματικής ελαχιστοποιείται στο 2ο Interval.

Στα bar plots που αναπαρίστανται στα Σχήματα 5.10 και 5.11 παρατηρούμε ότι για το Interval 3 οι προβλέψεις που έχουν να κάνουν με χρονικό ορίζοντα EOD έχουν τις χαμηλότερες τιμές μετρικών με αποτέλεσμα να είναι το πιο αποδοτικό μοντέλο για το Διάστημα 3. Ακολουθεί το Διάστημα 3 με χρονικό ορίζοντα 7 ημερών, με εμφανώς αυξημένες τις τιμές των μετρικών και χειρότερηση της απόδοσης των μοντέλων. Όσον αφορά για τις προβλέψεις σε βάθος χρόνου 30 ημερών και 90 ημερών οι μετρικές απόδοσης παρουσιάζουν τις μέγιστες τιμές πράγμα που σημαίνει ότι δεν έχουμε τόση μεγάλη ακρίβεια στις προβλέψεις. Αυτό που αξίζει να σημειωθεί είναι ότι κάποια μοντέλα που προβλέπουν για χρονικό ορίζοντα 90-

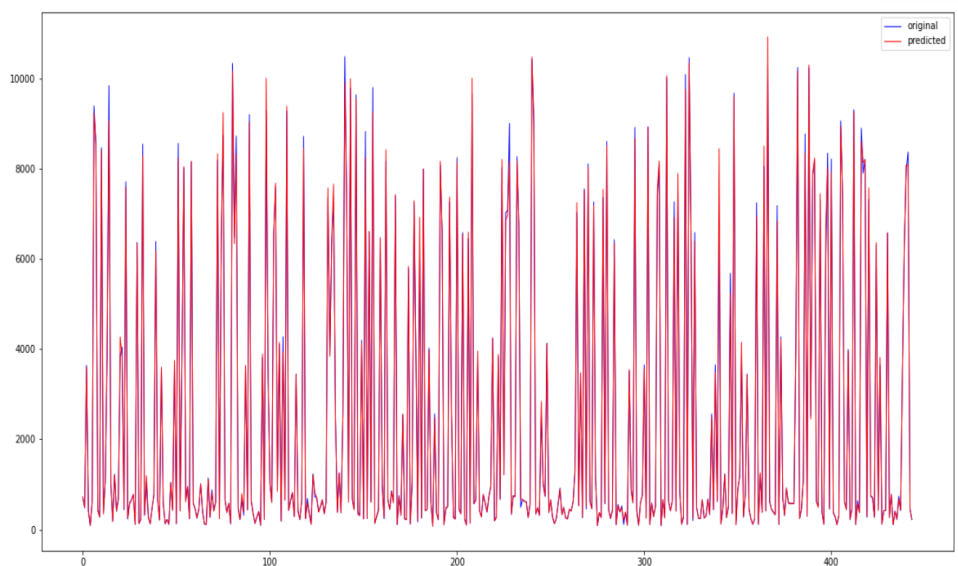


Σχήμα 5.2: Σύγκριση πραγματικών τιμών και τιμών πρόβλεψης για το Interval1

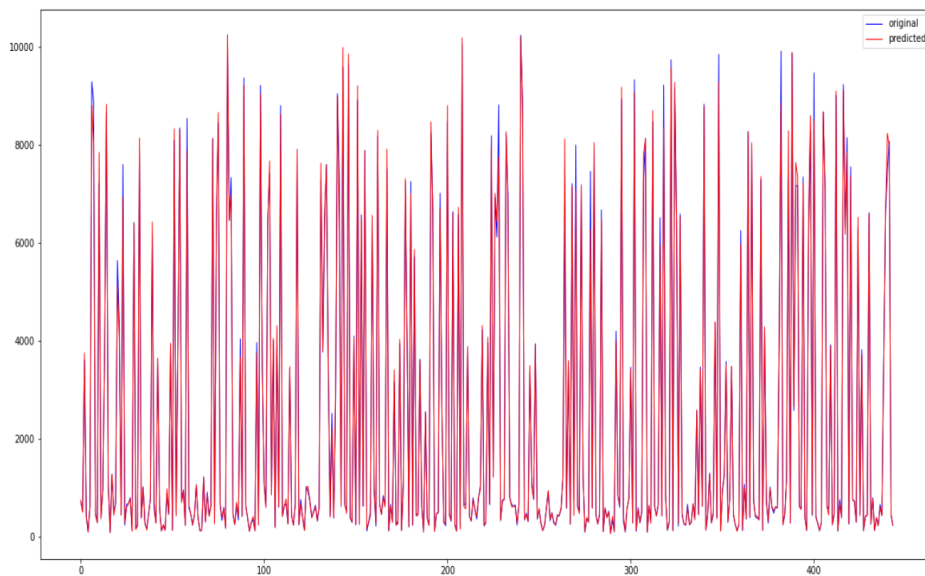
ημερών παρουσιάζουν μικρότερες τιμές στις μετρικές τους απ' ότι για πρόβλεψη 30-ημερών. Αυτό υποδεικνύει ότι όταν ο χρονικός ορίζοντας ξεπεράσει κάποιο όριο ημερών πρόβλεψης από κει και στο εξής δε δυσχεραίνουν οι προβλέψεις των μοντέλων.



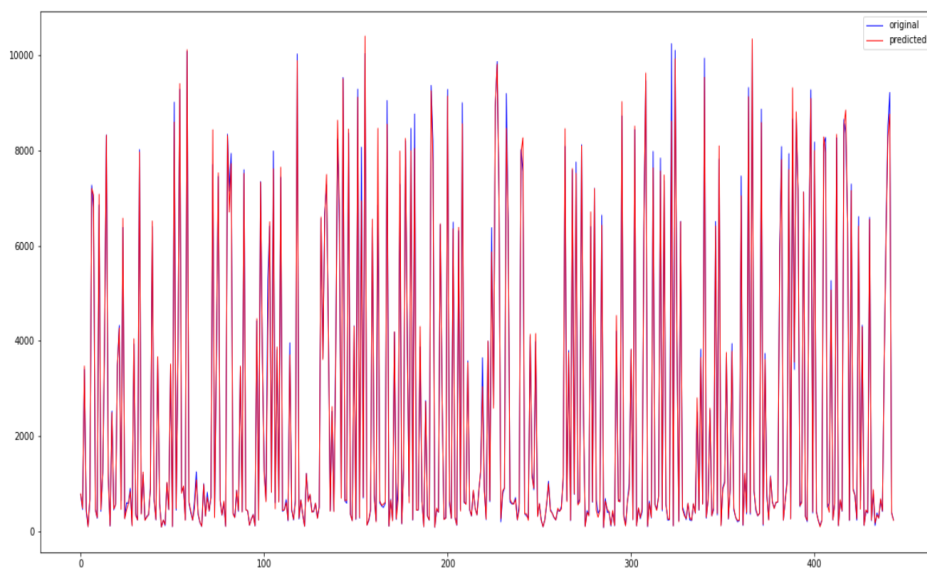
Σχήμα 5.3: Σύγκριση πραγματικών τιμών και τιμών πρόβλεψης για το Interval2



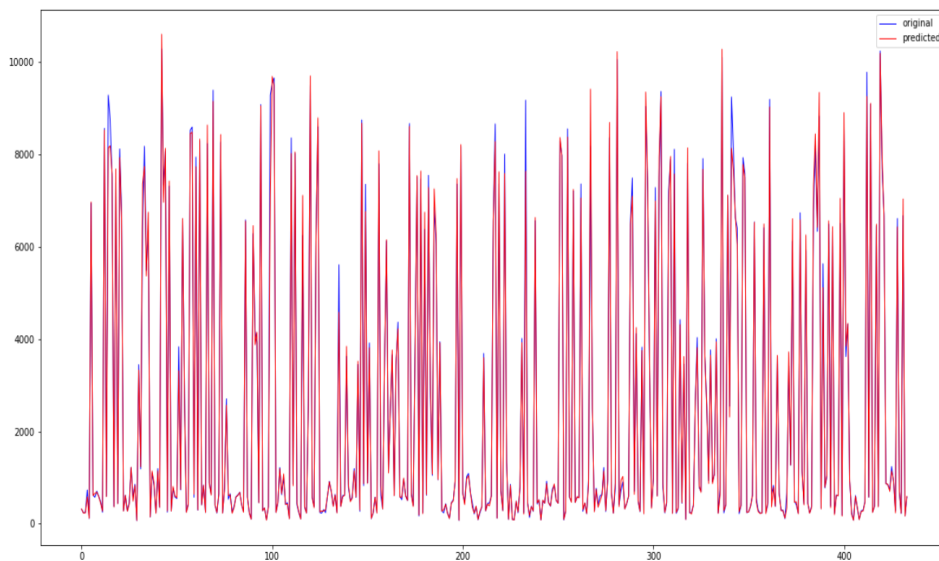
Σχήμα 5.4: Σύγκριση πραγματικών τιμών και τιμών πρόβλεψης για το Interval3



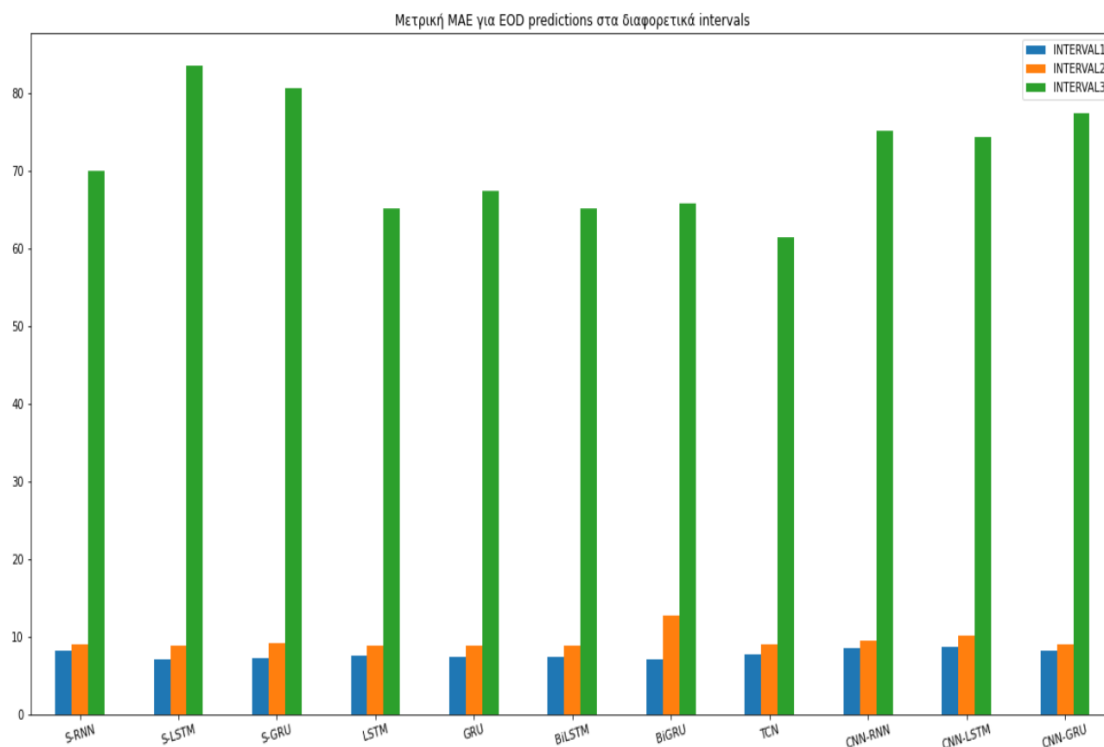
Σχήμα 5.5: Σύγκριση πραγματικών τιμών και τιμών πρόβλεψης για το Interval3 με χρονικό ορίζοντα 7 ημερών



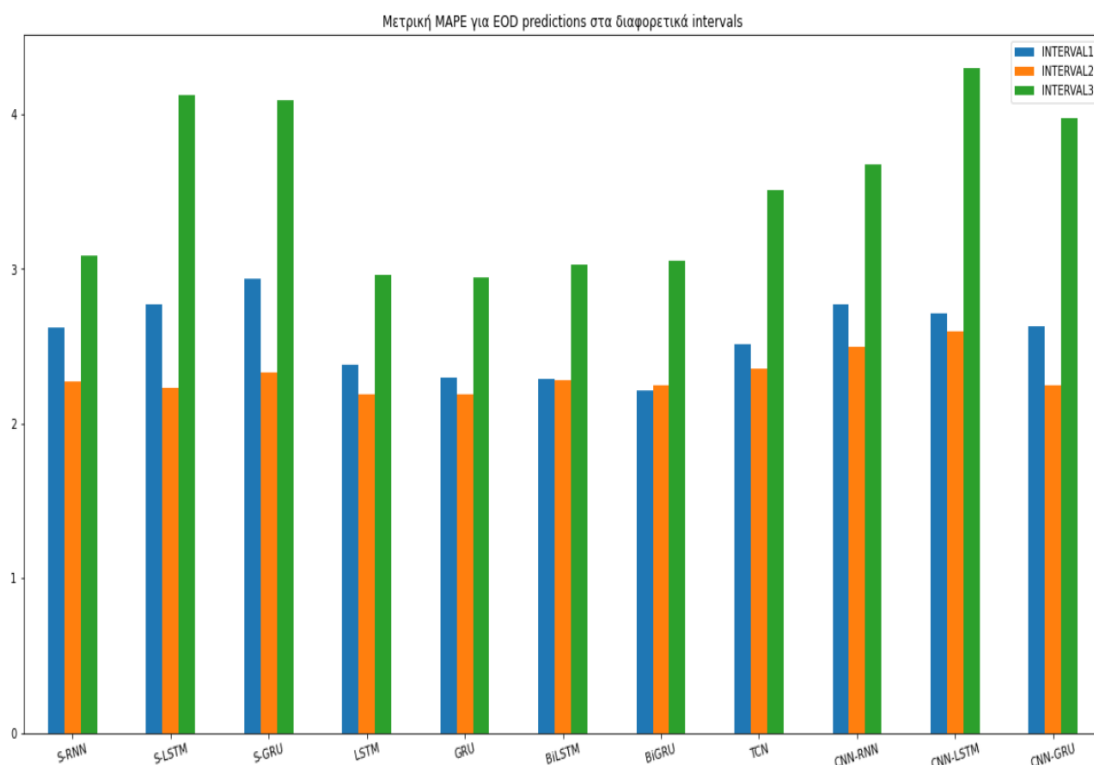
Σχήμα 5.6: Σύγκριση πραγματικών τιμών και τιμών πρόβλεψης για το Interval3 με χρονικό ορίζοντα 30 ημερών



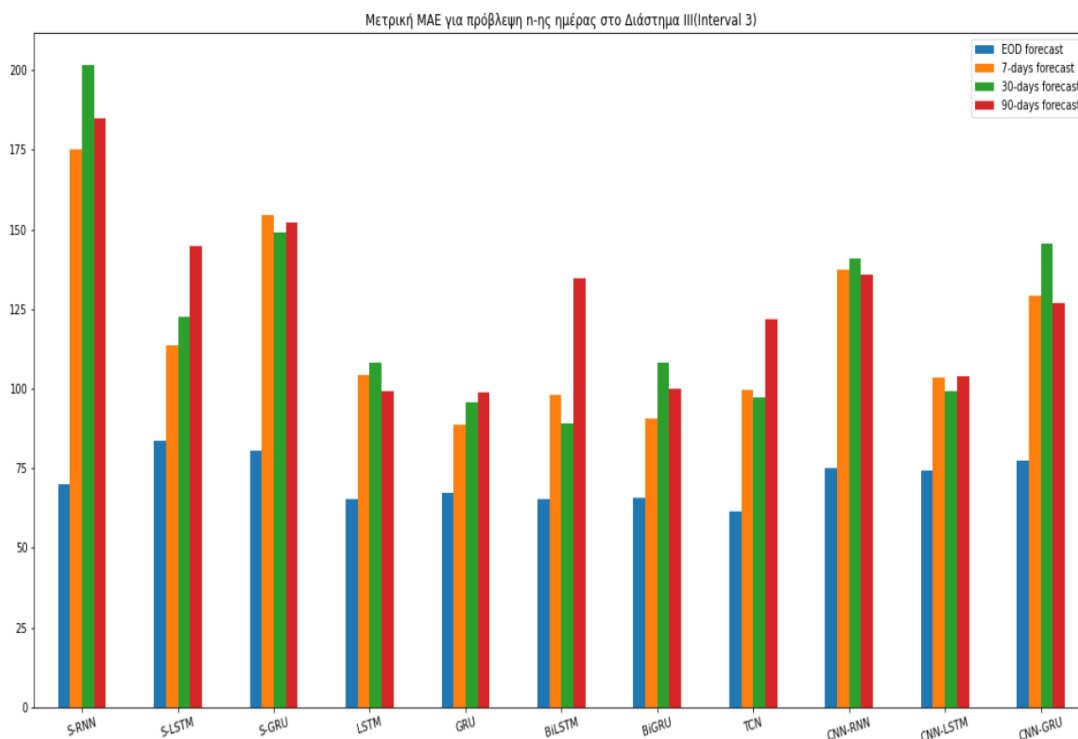
Σχήμα 5.7: Σύγκριση πραγματικών τιμών και τιμών πρόβλεψης για το Interval3 με χρονικό ορίζοντα 90 ημερών



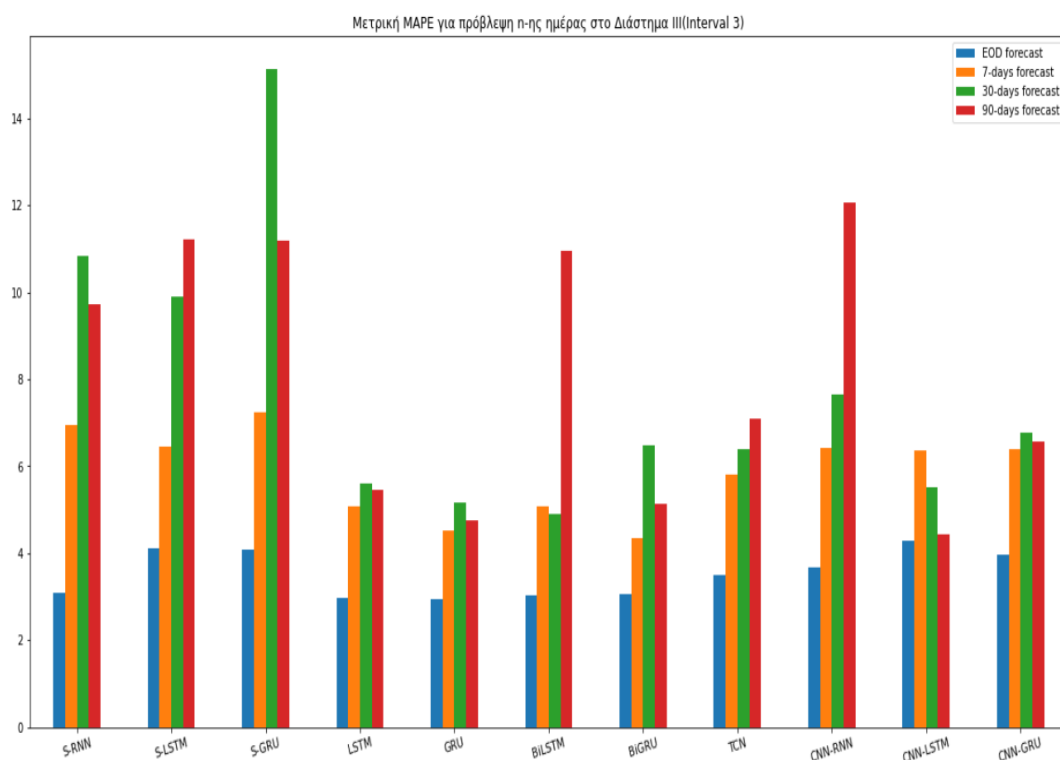
Σχήμα 5.8: Μετρική MAE για EOD predictions στα διαφορετικά Intervals



Σχήμα 5.9: Μετρική MAPE για EOD(End Of Day) predictions στα διαφορετικά Intervals



Σχήμα 5.10: Μετρική MAE των μοντέλων που προβλέπουν την τιμή BTC της n-οστής ημέρας στο Interval 3



Σχήμα 5.11: Μειωτική MAPE των μοντέλων που προβλέπουν την τιμή BTC της n-οστής ημέρας στο Interval 3

Μέρος **III**

Επίλογος

Κεφάλαιο 6

Επίλογος

Στο κεφάλαιο αυτό θα παρουσιαστούν τα συμπεράσματα που προέκυψαν έπειτα από την εκτέλεση του πειράματος μας καθώς και μελλοντικές επεκτάσεις που θα μπορούσαν να βελτιώσουν τα αποτελέσματα αυτής της μελέτης.

6.1 Συμπεράσματα

Στόχος της παρούσας διπλωματικής εργασίας είναι η πρόβλεψη της τιμής του κρυπτονομίσματος. Τα τελευταία χρόνια το μεγάλο ενδιαφέρον για απόκτηση και πώληση κρυπτονομισμάτων είναι αδιαμφισβήτητο γεγονός. Γι' αυτό το λόγο οι τιμές του BTC παρουσιάζουν μεγάλη αστάθεια και μεταβλητότητα γεγονός που αποτελεί πρόκληση για την πρόβλεψη της τιμής της επόμενης ημέρας.

Για την επίλυση αυτού του προβλήματος κάνουμε χρήση ενός dataset btc-data.csv [65]. Πρόκειται για ένα dataset με περίπου 700 χαρακτηριστικά (features) και ημερήσια δεδομένα σχεδόν 10 ετών. Από αυτό το μεγάλο σύνολο χαρακτηριστικών, ένα μικρότερο υποσύνολο χαρακτηριστικών επιλέχθηκε μέσω της διαδικασίας Επιλογής Χαρακτηριστικών (Feature Selection). Τα χαρακτηριστικά μετά το Feature Selection αφορούν ακατέργαστα χαρακτηριστικά συνδυασμένα με τεχνικούς δείκτες και διάφορες χρονικές περιόδους υπολογισμού αυτών των δεικτών. Αυτή η επιλογή γίνεται διότι οι τεχνικοί δείκτες έχουν ιδιότητες που δεν βρίσκονται εύκολα στο ακατέργαστα χαρακτηριστικά και δίνουν χρήσιμη πληροφορία στην πρόβλεψη της τιμής του btc. Αυτή τη διαδικασία του Feature Selection την επαναλαμβάνουμε για τα τρία διαφορετικά διαστήματα που επιλέξαμε να μελετήσουμε με χρονικό ορίζοντα πρόβλεψης 1 ημέρας αλλά και για το τρίτο Διάστημα για το οποίο θα προβλέψουμε την τιμή του btc για χρονικό ορίζοντα 7, 30 και 90 ημερών. Προφανώς, για κάθε Interval αλλά και για κάθε χρονικό ορίζοντα πρόβλεψης τα features διαφέρουν και αυτό είναι αναμενόμενο καθώς διαφορετικές μεταβλητές επηρεάζουν την τιμή του btc για πρόβλεψη της επόμενης ημέρας και διαφορετικές μεταβλητές θα επηρεάσουν την τιμή για πρόβλεψη των 90 ημερών. Η διαδικασία που ακολουθείται για το Feature Selection είναι μέσω του αλγορίθμου Recursive Feature Elimination, ή RFE για συντομία καθώς πρόκειται για ένα δημοφιλή αλγόριθμο επιλογής χαρακτηριστικών. Το RFE είναι δημοφιλές επειδή είναι εύκολο στη διαμόρφωση και τη χρήση του και επειδή είναι αποτελεσματικό στην επιλογή εκείνων των χαρακτηριστικών (στήλων) σε ένα σύνολο δεδομένων εκπαίδευσης που είναι περισσότερο ή πιο σχετικές με την πρόβλεψη της μεταβλητής στόχου. Στη συνέχεια απαιτείται έλεγχος του

μειωμένου dataset για πολυσυγγραμικότητα και αλληλοσυσχετίσεις. Αυτό το επιτυγχάνουμε μέσω του δείκτη VIF που καθορίζει τη συσχέτιση μεταξύ των ανεξάρτητων μεταβλητών. Για τιμές $VIF > 5$ η μεταβλητή αυτή μπορεί να αφαιρεθεί από το dataset. Τέλος αφού χωρίσαμε το dataset σε train και test σε 80% και 20% αντίστοιχα εφαρμόσαμε κανονικοποίηση στα δεδομένα του εκάστοτε υποσυνόλου δεδομένων για να διώξουμε τα ακραία σημεία (outliers) από τα χαρακτηριστικά και να μην εστιάσουν οι αλγόριθμοι στις ακραίες τιμές. Έπειτα από την προεπεξεργασία των δεδομένων, ακολούθησε το πειραματικό κομμάτι της διπλωματικής. Αρχικά περιγράφηκαν οι βασικές αρχιτεκτονικές των συνελκτικών δικτύων, η αρχιτεκτονική των αναδρομικών δικτύων και η υβριδική αρχιτεκτονική. Κατόπιν, δοκιμάστηκαν τα νευρωνικά δίκτυα RNN, GRU, LSTM μέσω της πρώτης αρχιτεκτονικής και στη συνέχεια τα χρονικά συνελκτικά νευρωνικά δίκτυα (TCN). Μέσω της δεύτερης αρχιτεκτονικής έγινε συνδυασμός όλων των παραπάνω νευρωνικών με μονοδιάστατα CNNs. Όλα τους είχαν ικανοποιητική απόδοση με το πιο αποδοτικό να είναι το νευρωνικό δίκτυο TCN, στη συνέχεια το GRU, το BiGRU, LSTM και BiLSTM. Τα υβριδικά μοντέλα δεν έδειξαν κάποια ιδιαίτερη βελτίωση στην πρόβλεψη τιμών ενώ σε αρκετές περιπτώσεις είχαν τα υψηλότερα σκορ, γεγονός που τα καθιστά λιγότερο αποδοτικά. Ομοίως και το RNN νευρωνικό δίκτυο φάνηκε να υπολείπεται των υπολοίπων, κάτι το οποίο ήταν αναμενόμενο και οφείλεται στην αδυναμία των RNNs να αποθηκεύουν στη μνήμη τους πληροφορίες από μεγάλες σε μήκος χρονοσειρές. Στα Intervals 1-2 η πρόβλεψη ήταν αρκετά κοντινή στις πραγματικές τιμές κάτι το οποίο δεν ισχύει και για το Interval 3 όπου οι μεταβολές στις τιμές του Bitcoin δυσκόλεψαν τα μοντέλα να κάνουν πιο accurate προβλέψεις.

6.2 Μελλοντικές Επεκτάσεις

Η μελέτη που έγινε στα πλαίσια της παρούσας διπλωματικής εργασίας θα μπορούσε σίγουρα να επεκταθεί και ίσως να επιφέρει και βελτίωση των μοντέλων. Συγκεκριμένα: 1. Εφαρμογή της μεθόδου Cross-validation ίσως επιφέρει βελτίωση στα μοντέλα βαθιάς μάθησης. 2. Καλύτερη προεπεξεργασία των δεδομένων, όπως για παράδειγμα με χρήση PCA. 3. Εξερεύνηση άλλων νευρωνικών δικτύων που επεξεργάζονται χρονοσειρές όπως Κωδικοποιητές - Αποκωδικοποιητές ή Transfer Learning.

Σε μελλοντική εργασία, θα μπορούσαν να διερευνηθούν άλλοι παράγοντες που ενδέχεται να επηρεάσουν τις τιμές στην αγορά κρυπτονομισμάτων, όπως για παράδειγμα η επίδραση που έχουν τα μέσα κοινωνικής δικτύωσης και τα tweets που αφορούν την τιμή και τον όγκο συναλλαγών των κρυπτονομισμάτων. Συνεπώς η ανάλυση tweets χρησιμοποιώντας τεχνικές επεξεργασίας φυσικής γλώσσας και ανάλυση συναισθήματος μπορεί να αποτελέσουν καινοτόμα επιλογή στην πρόβλεψη τιμής κρυπτονομίσματος.

Βιβλιογραφία

- [1] What is “Bitcoin mining” and how does mining work? <https://www.bitpanda.com/academy/en/lessons/what-is-bitcoin-mining-and-how-does-mining-work/>. Ημερομηνία Πρόσβασης:12-03-2022.
- [2] Aston Zhang, Zachary C Lipton, Mu Li και Alexander J Smola. *Dive into deep learning*. *arXiv preprint arXiv:2106.11342*, 2021.
- [3] Arvind Narayanan, Joseph Bonneau, Edward Felten, Andrew Miller και Steven Goldfeder. *Bitcoin and cryptocurrency technologies: a comprehensive introduction*. Princeton University Press, 2016.
- [4] *Bitcoin*. <https://el.wikipedia.org/wiki/Bitcoin>. Ημερομηνία Πρόσβασης:12-03-2022.
- [5] Andrew Urquhart. *What causes the attention of Bitcoin?* *Economics Letters*, 166:40–44, 2018.
- [6] Michael Crosby, Pradan Pattanayak, Sanjeev Verma, Vignesh Kalyanaraman και others. *Blockchain technology: Beyond bitcoin*. *Applied Innovation*, 2(6-10):71, 2016.
- [7] Congcong Ye, Guoqiang Li, Hongming Cai, Yonggen Gu και Akira Fukuda. *Analysis of Security in Blockchain: Case Study in 51 2018 5th International Conference on Dependable Systems and Their Applications (DSA)*, σελίδες 15-24, 2018.
- [8] Paulina Likos και Coryanne Hicks. *The History of Bitcoin, the First Cryptocurrency*. <https://money.usnews.com/investing/articles/the-history-of-bitcoin>. Ημερομηνία Πρόσβασης:12-03-2022.
- [9] ERIC ROSENBERG. *How Does Bitcoin Mining Work?* <https://www.thebalance.com/how-does-bitcoin-mining-work-5088328>. Ημερομηνία Πρόσβασης:12-03-2022.
- [10] *Litecoin vs Ethereum*. <https://www.thebalance.com/how-does-bitcoin-mining-work-5088328>. Ημερομηνία Πρόσβασης:12-03-2022.
- [11] *Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach*. <https://link.springer.com/content/pdf/10.1007/s00521-020-05129-6.pdf>. Ημερομηνία Πρόσβασης:12-03-2022.
- [12] Muhammad Ashraf FAUZI, Norazha PAIMAN και Zarina OTHMAN. *Bitcoin and cryptocurrency: Challenges, opportunities and future works*. *The Journal of Asian Finance, Economics, and Business*, 7(8):695–704, 2020.

- [13] Karl J O'Dwyer και David Malone. *Bitcoin mining and its energy footprint*. 2014.
- [14] Stanvan der Avoird. *Prediction and technical analysis of the Bitcoin crypto currency using machine learning*.
- [15] NATHAN REIFF. *Why Bitcoin Price Predictions Are Unreliable*. <https://www.investopedia.com/tech/why-bitcoin-price-predictions-are-unreliable/>. Ημερομηνία Πρόσβασης:12-03-2022.
- [16] George H Chen, Stanislav Nikolov και Devavrat Shah. *A latent source model for nonparametric time series classification*. *Advances in neural information processing systems*, 26, 2013.
- [17] Yan Chen και Cristiano Bellavitis. *Blockchain disruption and decentralized finance: The rise of decentralized business models*. *Journal of Business Venturing Insights*, 13:e00151, 2020.
- [18] *Decentralized Finance: Blockchain Technology and the Quest for an Open Financial System*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3418557. Ημερομηνία Πρόσβασης:17-01-2022.
- [19] Yan Chen και Cristiano Bellavitis. *Decentralized finance: Blockchain technology and the quest for an open financial system*. *Stevens Institute of Technology School of Business Research Paper*, 2019.
- [20] Akash Dubey. *Feature selection Using Random forest-The Wisdom of Crowds*. <https://towardsdatascience.com/feature-selection-using-random-forest-26d7b747597f>. Ημερομηνία Πρόσβασης:28-01-2022.
- [21] *Recursive Feature Elimination*. https://www.scikit-yb.org/en/latest/api/model_selection/rfecv.html. Ημερομηνία Πρόσβασης:30-01-2022.
- [22] GIANLUCA MALATO. *Feature selection with Random Forest*. <https://www.yourdatateacher.com/2021/10/11/feature-selection-with-random-forest/>. Ημερομηνία Πρόσβασης:28-01-2022.
- [23] Taiwo Oladipupo Ayodele. *Types of machine learning algorithms*. *New advances in machine learning*, 3:19-48, 2010.
- [24] Paraskevi Tzouveli, Andreas Schmidt, Michael Schneider, Antonis Symvonis και Stefanos Kollias. *Adaptive Reading Assistance for the Inclusion of Students with Dyslexia: The AGENT-DYSL Approach*. *Proceedings of the 2008 Eighth IEEE International Conference on Advanced Learning Technologies*, σελίδες 167-171, 2008.
- [25] *Build your first artificial neural networks using Pytorch*. <https://www.analyticsvidhya.com/blog/2021/08/build-your-first-artificial-neural-networks-using-pytorch/>. Ημερομηνία Πρόσβασης:12-03-2022.

- [26] Implement Artificial Neural Networks (ANNs) in SQL Server. <https://www.sqlshack.com/implement-artificial-neural-networks-anns-in-sql-server/>. Ημερομηνία Πρόσβασης:15-02-2022.
- [27] Bruno Spilak. *Deep neural networks for cryptocurrencies price prediction*. Μεταπτυχιακή διπλωματική εργασία, Humboldt-Universität zu Berlin, 2018.
- [28] Manolis Wallace, Ilias Maglogiannis, Kostas Karpouzis, George Kormentzas και Stefanos Kollias. *Intelligent one-stop-shop travel recommendations using an adaptive neural network and clustering of history*. *Information Technology & Tourism*, 6(3):181-193, 2003.
- [29] S Kollias και Dimitris Anastassiou. *Adaptive training of multilayer neural networks using a least squares estimation technique*. 1988.
- [30] Manolis Wallace, Nicolas Tsapatsoulis και Stefanos Kollias. *Intelligent initialization of resource allocating RBF networks*. *Neural Networks*, 18(2):117-122, 2005.
- [31] Phivos Mylonas, Evaggelos Spyrou, Yannis Avrithis και Stefanos Kollias. *Using visual context and region semantics for high-level concept detection*. *IEEE Transactions on Multimedia*, 11(2):229-243, 2009.
- [32] Perceptron. https://commons.wikimedia.org/wiki/File:Perceptron_moj.png. Ημερομηνία Πρόσβασης:17-01-2022.
- [33] Bekir Karlik και A Vehbi Olgac. *Performance analysis of various activation functions in generalized MLP architectures of neural networks*. *International Journal of Artificial Intelligence and Expert Systems*, 1(4):111-122, 2011.
- [34] *Reconstruction of porous media from extremely limited information using conditional generative adversarial networks*. https://www.researchgate.net/publication/335845675_Reconstruction_of_porous_media_from_extremely_limited_information_using_conditional_generative_adversarial_networks. Ημερομηνία Πρόσβασης:17-01-2022.
- [35] G Caridakis, A Raouzaïou, K Karpouzis και S Kollias. *Synthesizing Gesture Expressivity Based on Real Sequences*. *Workshop Programme*, τόμος 10, σελίδα 19.
- [36] Francesco Caliva, Fabio Sousa De Ribeiro, Antonios Mylonakis, Christophe Demazière, Paolo Vinai, Georgios Leontidis και Stefanos Kollias. *A deep learning approach to anomaly detection in nuclear reactors*. *2018 International joint conference on neural networks (IJCNN)*, σελίδες 1-8. IEEE, 2018.
- [37] Dimitrios Kollias, Anastasios Arsenos, Levon Soukissian και Stefanos Kollias. *Micov19d: Covid-19 detection through 3-d chest ct image analysis*. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, σελίδες 537-544, 2021.
- [38] Paul J Werbos. *Backpropagation through time: what it does and how to do it*. *Proceedings of the IEEE*, 78(10):1550-1560, 1990.

- [39] 8.7. *Backpropagation Through Time – Dive into Deep Learning 0.16.6 documentation*. https://d2l.ai/chapter_recurrent_neural_networks/language_models_and_dataset.html. Ημερομηνία Πρόσβασης:12-03-2022.
- [40] *Understanding GRU Networks*. <https://towardsdatascience.com/understanding-gru-networks-2ef37df6c9be>. Ημερομηνία Πρόσβασης:20-11-2021.
- [41] *Understanding GRU Networks*. <https://towardsdatascience.com/understanding-gru-networks-2ef37df6c9be>. Ημερομηνία Πρόσβασης:20-11-2021.
- [42] 9.2. *Long Short-Term Memory (LSTM)-Dive into deep learning*.
- [43] *Understanding LSTM Networks*. <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>. Ημερομηνία Πρόσβασης:20-11-2021.
- [44] Bashar Alhnaity, Stefanos Kollias, Georgios Leontidis, Shouyong Jiang, Bert Schamp και Simon Pearson. *An autoencoder wavelet based deep neural network with attention mechanism for multi-step prediction of plant growth*. *Information Sciences*, 560:35-50, 2021.
- [45] *Recurrent Neural Networks and LSTM explained*. <https://purnasaigudikandula.medium.com/recurrent-neural-networks-and-lstm-explained-7f51c7f6bbb9>. Ημερομηνία Πρόσβασης:22-11-2021.
- [46] *The magic of LSTM neural networks*. <https://medium.com/datathings/the-magic-of-lstm-neural-networks-6775e8b540cd>. Ημερομηνία Πρόσβασης:22-11-2021.
- [47] 9.4. *Bidirectional Recurrent Neural Networks*. https://d2l.ai/chapter_recurrent_neural_networks/language_models_and_dataset.html. Ημερομηνία Πρόσβασης:12-03-2022.
- [48] Dimitrios Kollias, Miao Yu, Athanasios Tagaris, Georgios Leontidis, Andreas Stafylopatis και Stefanos Kollias. *Adaptation and contextualization of deep neural network models*. *2017 IEEE symposium series on computational intelligence (SSCI)*, σελίδες 1-8. IEEE.
- [49] *DIVE INTO DEEP LEARNING*. https://d2l.ai/chapter_convolutional_neural_networks/index.html. Ημερομηνία Πρόσβασης:17-01-2022.
- [50] Ilianna Kollia, Andreas Georgios Stafylopatis και Stefanos Kollias. *Predicting Parkinson’s disease using latent information extracted from deep neural networks*. *2019 International Joint Conference on Neural Networks (IJCNN)*, σελίδες 1-8. IEEE, 2019.
- [51] D Kollias, N Bouas, Y Vlaxos, V Brillakis, M Seferis, I Kollia, L Sukissian, J Wingate και S Kollias. *Deep Transparent Prediction through Latent Representation Analysis*. *arXiv preprint arXiv:2009.07044*, 2020.
- [52] Dimitrios Kollias, Athanasios Tagaris, Andreas Stafylopatis, Stefanos Kollias και Georgios Tagaris. *Deep neural architectures for prediction in healthcare*. *Complex & Intelligent Systems*, 4(2):119-131, 2018.

- [53] *Introduction to Convolutional Neural Networks (CNN)*. <https://www.analyticsvidhya.com/blog/2021/05/convolutional-neural-networks-cnn/>. Ημερομηνία Πρόσβασης:15-02-2022.
- [54] *6. Convolutional Neural Networks – Dive into Deep Learning 0.17.0 documentation*. https://d2l.ai/chapter_convolutional-neural-networks/index.html. Ημερομηνία Πρόσβασης:17-01-2022.
- [55] *Understanding of Convolutional Neural Network (CNN) – Deep Learning*. <https://medium.com/@RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148>. Ημερομηνία Πρόσβασης:18-02-2022.
- [56] Fabio De Sousa Ribeiro, Francesco Calivá, Mark Swainson, Kjartan Gudmundsson, Georgios Leontidis και Stefanos Kollias. *Deep bayesian self-training*. *Neural Computing and Applications*, 32(9):4275-4291, 2020.
- [57] *Introduction to 1D Convolutional Neural Networks in Keras for Time Sequences*. <https://blog.goodaudience.com/introduction-to-1d-convolutional-neural-networks-in-keras-for-time-sequences-3a7ff801a2cf>. Ημερομηνία Πρόσβασης:10-01-2022.
- [58] *6.3. Padding and Stride*. https://d2l.ai/chapter_convolutional-neural-networks/padding-and-strides.html. Ημερομηνία Πρόσβασης:17-01-2022.
- [59] *Convolutional Neural Network- An Introduction to Convolutional Neural Networks*. <https://towardsdatascience.com/convolutional-neural-network-17fb77e76c05>. Ημερομηνία Πρόσβασης:17-02-2022.
- [60] *CNN:Step 3-Flattening*. https://medium.com/@PK_KwanG/cnn-step-2-flattening-50ee0af42e3e. Ημερομηνία Πρόσβασης:12-03-2022.
- [61] Francesco Lässig. *Temporal Convolutional Networks and Forecasting*. <https://unit8.com/resources/temporal-convolutional-networks-and-forecasting/>. Ημερομηνία Πρόσβασης:18-02-2022.
- [62] Dimitris Kollias, Y Vlaxos, M Seferis, Ilianna Kollia, Levon Sukissian, James Wingate και S Kollias. *Transparent adaptation in deep medical image diagnosis*. *International Workshop on the Foundations of Trustworthy AI Integrating Learning, Optimization and Reasoning*, σελίδες 251-267. Springer, 2020.
- [63] *An end-to-end open source machine learning platform*. Ημερομηνία Πρόσβασης:15-11-2021.
- [64] *Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach*. <https://link.springer.com/content/pdf/10.1007/s00521-020-05129-6.pdf>. Ημερομηνία Πρόσβασης:18-11-2021.

- [65] *Time-series forecasting of Bitcoin prices using high-dimensional features: a machine learning approach.* <https://link.springer.com/content/pdf/10.1007/s00521-020-05129-6.pdf>. Ημερομηνία Πρόσβασης:02-03-2022.
- [66] aniruddha. *What is Multicollinearity? Here's Everything You Need to Know.* <https://www.analyticsvidhya.com/blog/2020/03/what-is-multicollinearity/>. Ημερομηνία Πρόσβασης:19-01-2022.
- [67] Jeff Hale. *Scale, Standardize, or Normalize with Scikit-Learn.* <https://towardsdatascience.com/scale-standardize-or-normalize-with-scikit-learn-6ccc7d176a02>. Ημερομηνία Πρόσβασης:31-01-2022.
- [68] *How to Calculate Root Mean Square Error (RMSE) in Excel.* <https://gisgeography.com/root-mean-square-error-rmse-gis/>. Ημερομηνία Πρόσβασης:28-02-2022.
- [69] *Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE).* http://www.eumetrain.org/data/4/451/english/msg/ver_cont_var/uos3/uos3_ko1.htm. Ημερομηνία Πρόσβασης:28-02-2022.