



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ – ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ**

**Πρόγραμμα Μεταπτυχιακών Σπουδών**

**« ΠΡΟΗΓΜΕΝΑ ΣΥΣΤΗΜΑΤΑ ΠΛΗΡΟΦΟΡΙΚΗΣ - ΑΝΑΠΤΥΞΗ  
ΛΟΓΙΣΜΙΚΟΥ ΚΑΙ ΤΕΧΝΗΤΗΣ ΝΟΗΜΟΣΥΝΗΣ »**

**Μεταπτυχιακή Διατριβή**

Τίτλος Διατριβής	«Χρήση Ανατροφοδοτούμενου Νευρωνικού Δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων.» “Using recurrent neural networks and time series for data prediction.”
Όνοματεπώνυμο Φοιτητή	Ψαρρού Αικατερίνη
Πατρώνυμο	Νικόλαος
Αριθμός Μητρώου	ΜΠΣΠ21064
Επιβλέπων	Διονύσιος Σωτηρόπουλος, Επίκουρος καθηγητής

23 Οκτωβρίου 2023

---

### **Τριμελής Εξεταστική Επιτροπή**

Διονύσιος Σωτηρόπουλος  
Επίκουρος καθηγητής

Γεώργιος Τσιχριντζής  
Καθηγητής

Ευάγγελος Σακκόπουλος  
Αναπληρωτής καθηγητής

## Περίληψη

Δεν υφίσταται καμία αμφιβολία πως το να γνωρίζει κανείς το μέλλον μπορεί να αποτελέσει το μεγαλύτερο πλεονέκτημα αυτής της ζωής. Ο άνθρωπος που διαθέτει την ικανότητα να προβλέψει το τι έχει να του προσφέρει το αύριο καθίσταται ικανός να προσεγγίσει αποτελεσματικότερα τους στόχους του. Δεν υφίσταται επίσης καμία αμφιβολία για την συνδρομή των νευρωνικών δικτύων στην σημερινή εποχή αλλά και το πλήθος ερευνητικών εφαρμογών που έχει προσδώσει. Τα νευρωνικά δίκτυα αποτελούν ισχυρά εργαλεία για την πρόβλεψη μέσω της ανάλυσης χρονοσειρών και αυτό αποτελεί και τον πυρήνα της παρούσας διπλωματικής εργασίας. Παρόλο που γνωρίζουμε ότι είναι αδύνατο να γνωρίζουμε το μέλλον, πλήρως, μπορούμε να προσπαθήσουμε να προβλέψουμε το τι έρχεται.

Τόσο τα νευρωνικά δίκτυα όσο και η ανάλυση κοινωνικών δικτύων είναι δύο διακριτά πεδία, παρόλα αυτά είναι δυνατό να συσχετιστούν με διάφορους τρόπους, όπως θα αποδειχθεί με την εν λόγω προσέγγιση, θέτοντας ως σημείο εστίασης την ανάλυση όπως και την αποσαφήνιση της δομής των κοινωνικών δικτύων. Το twitter ως ένας από τους βασικότερους πυλώνες του των κοινωνικών δικτύων θα αποτελέσει την πηγή δεδομένων της υλοποίησης του συγκεκριμένου ερευνητικού σχεδιασμού προκειμένου να επιτευχθούν οι στόχοι της παρούσας εργασίας.

Στην παρούσα διπλωματική εργασία πρόκειται να πραγματοποιηθεί οπτικοποίηση του ημερησίου όγκου αναρτήσεων περιορίζοντας το σύνολο χρηστών σε εκείνους που έχουν κάνει παραπάνω από έναν αριθμό tweets, πρόβλεψη του αριθμού των tweets που δημοσιεύονται κάθε λεπτό βασισμένο στα δεδομένα των προηγούμενων 24 ωρών και τέλος πρόβλεψη του αριθμού των tweets ανά λεπτό βασισμένο σε ιστορικά δεδομένα, χρησιμοποιώντας το μοντέλο βαθιάς μάθησης που ονομάζεται LSTM (Long Short-Term Memory). Έπειτα θα γίνει γνωστοποίηση των μεθόδων πρόβλεψης χρονοσειρών με χρήση ανατροφοδοτούμενου νευρωνικού δικτύου, η διαδικασία δημιουργίας του dataset καθώς και τα αποτελέσματα των προβλέψεων των δύο μελετών περίπτωσης.

Λέξεις κλειδιά: Νευρωνικό δίκτυο, Χρονοσειρές, Δεδομένα twitter, Πρόβλεψη, LSTM.

## “Using Recurrent Neural Networks and Time Series for Data Prediction.”

### **Abstract**

There is no doubt that knowing the future can be the greatest advantage in life. A person who has the ability to predict what the future holds can approach their goals more effectively. There is also no doubt about the contribution of neural networks in today's era and the numerous research applications they have provided. Neural networks are powerful tools for prediction through time series analysis, and this forms the core of this thesis.

Although we know that it is impossible to fully know the future, we can try to predict what is coming. Both neural networks and social network analysis are distinct fields; however, they can be correlated in various ways, as will be demonstrated in this approach. The focus will be on analyzing and clarifying the structure of social networks. Twitter, as one of the major pillars of social networks, will be the data source for implementing this research design to achieve the goals of this work.

In this thesis, the following tasks will be carried out: visualization of the daily volume of posts, limiting the set of users to those who have posted more than a certain number of tweets, predicting the number of tweets published every minute based on the data from the previous 24 hours, and finally, predicting the number of tweets per minute based on historical data using the Long Short-Term Memory (LSTM) deep learning model. The methods for time series prediction using a feedback neural network will be disclosed, along with the process of creating the dataset and the results of the two case studies.

Keywords: Neural network, Time series, Twitter data, Prediction, LSTM.

## ΠΕΡΙΕΧΟΜΕΝΑ

---

<b>1. Εισαγωγή στην Ανάλυση Κοινωνικών Δικτύων</b> .....	9
<b>1.1 Τι είναι ένα κοινωνικό δίκτυο;</b> .....	9
<b>1.1.1 Οι κορυφαίες 10 κοινωνικές δικτυώσεις παγκοσμίως:</b> .....	11
<b>1.2 Ανάλυση κοινωνικών δικτύων</b> .....	14
<b>1.3 Μηχανική μάθηση</b> .....	15
<b>1.3.1 Κατηγορίες μηχανικής μάθησης</b> .....	18
<b>1.4 Νευρωνικό Δίκτυο</b> .....	22
<b>1.4.1 Τύποι νευρωνικών δικτύων:</b> .....	25
<b>1.4.2 Βαθιά Νευρωνικά Δίκτυα (Deep learning)</b> .....	32
<b>1.5 Σύνδεση Νευρωνικών και Κοινωνικών δικτύων</b> .....	34
<b>2. Πρόβλεψη Χρονοσειρών με χρήση Ανατροφοδοτούμενων Νευρωνικών Δικτύων</b> .....	39
<b>2.1 Τι είναι οι Χρονοσειρές;</b> .....	39
<b>2.2 Ανάλυση χρονοσειρών</b> .....	41
<b>2.3 Μοτίβα – Χαρακτηριστικά των χρονοσειρών</b> .....	43
<b>2.4 Ανατροφοδοτούμενα Νευρωνικά Δίκτυα (Recurrent neural network (RNN)</b> .....	51
<b>2.4.1 Προετοιμασία των δεδομένων και κατασκευή του μοντέλου RNN</b> .....	53
<b>2.4.2 Πρόβλεψη χρονοσειρών με χρήση ανατροφοδοτούμενων νευρωνικών δικτύων.</b> .....	55
<b>3. Μελέτες περίπτωσης: Εφαρμογή Ανατροφοδοτούμενου Νευρωνικού Δικτύου σε πραγματικά δεδομένα</b> .....	57
<b>3.4 Χρονοσειρές όγκου δεδομένων</b> .....	67
<b>3.4.1 Η πρόβλεψη μελλοντικών τιμών με βάση τα υπάρχοντα δεδομένα</b> .....	67
<b>3.4.2 Κατηγορίες μοντέλων πρόβλεψης</b> .....	68
<b>4. Μεθοδολογία ανάπτυξης μοντέλου βαθιάς μάθησης LSTM σε χρονοσειριακά δεδομένα.</b> .....	70
<b>4.1. «Περίπτωση Μάτι και Περίπτωση Τέμπερ»</b> .....	70

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

<b>4.2 Συλλογή και επεξεργασία δεδομένων</b> .....	71
<b>4.2.1 Βιβλιοθήκες &amp; Περιβάλλον Ανάπτυξης</b> .....	73
<b>4.3 Επιλογή μοντέλου LSTM</b> .....	74
<b>4.3.1 Διαδικασία</b> .....	74
<b>4.3.2 Επισκόπηση Στόχων Υλοποίησης</b> .....	74
<b>4.3.3 Βιβλιοθήκες και ρόλοι τους:</b> .....	75
<b>4.4 Δημιουργία Datasets για το LSTM:</b> .....	75
<b>4.5 Αποτελέσματα οπτικοποίησης ημερησίου volume αναρτήσεων</b> .....	77
<b>4.5.2 Αποτελέσματα Τέμνη LTSM :</b> .....	88
<b>5. Συμπεράσματα – μελλοντική εργασία</b> .....	95
<b>6. Βιβλιογραφική αναφορά</b> .....	97

## Κατάλογος Εικόνων

<b>Εικόνα 1: Εφαρμογές κοινωνικής δικτύωσης</b>	<b>11</b>
<b>Εικόνα 2: Διάγραμμα επιβλεπόμενης μάθησης</b>	<b>19</b>
<b>Εικόνα 3: Διάγραμμα μη επιβλεπόμενης μάθησης</b>	<b>22</b>
<b>Εικόνα 4: Κανονικό νευρωνικό δίκτυο</b>	<b>25</b>
<b>Εικόνα 5: Διάγραμμα συνελκτικού νευρωνικού δικτύου</b>	<b>27</b>
<b>Εικόνα 6: Διάγραμμα ανατροφοδοτούμενου Νευρωνικού δικτύου</b>	<b>28</b>
<b>Εικόνα 7: Αναπαράσταση νευρωνικού δικτύου LSTM</b>	<b>29</b>
<b>Εικόνα 8: Νευρωνικό δίκτυο stacked autoencoder</b>	<b>30</b>
<b>Εικόνα 9: Νευρωνικό δίκτυο variational autoencoder</b>	<b>32</b>
<b>Εικόνα 10: Μορφή ενσωματωμένου layer</b>	<b>36</b>
<b>Εικόνα 11: εικονογράφηση Graph Neural Network</b>	<b>37</b>
<b>Εικόνα 12: διάγραμμα χρονοσειράς</b>	<b>40</b>
<b>Εικόνα 13: Μηνιαίες πωλήσεις αντιδιαβητικών φαρμάκων στην Αυστραλία</b>	<b>44</b>
<b>Εικόνα 14: Εποχικό μοτίβο στις πωλήσεις φαρμάκων</b>	<b>45</b>
<b>Εικόνα 15: Τέσσερα παραδείγματα χρονοσειρών που οπτικοποιούν διαφορετικά μοτίβα των τάσεων</b>	<b>45</b>
<b>Εικόνα 16: Διαγράμματα περιπτώσεων με χαρακτηριστικά χρονοσειρών</b>	<b>47</b>
<b>Εικόνα 17: Διάγραμμα χρονοσειράς με πρόβλεψη</b>	<b>49</b>
<b>Εικόνα 18: Διαδρομή πληροφορίας εντός RNN νευρωνικού δικτύου</b>	<b>52</b>
<b>Εικόνα 19: Πυρκαγιά στο Μάτι, 2008</b>	<b>59</b>
<b>Εικόνα 20: Σιδηροδρομικό δυστύχημα Τέμπη, 2023</b>	<b>60</b>
<b>Εικόνα 21: Κώδικας με εγκατάσταση Tweepy</b>	<b>66</b>
<b>Εικόνα 22: κώδικας Python με χρήση μοναδικών κλειδιών</b>	<b>66</b>
<b>Εικόνα 23: Αρχείο mati.csv</b>	<b>72</b>
<b>Εικόνα 24: Αρχείο tempi.csv</b>	<b>73</b>
<b>Εικόνα 25: Προσδιορισμός unique authors author_volume_df</b>	<b>78</b>
<b>Εικόνα 26: Δείγμα με author_volume_df για τους 25.000 unique authors</b>	<b>79</b>
<b>Εικόνα 27: Διάγραμμα Daily evolution of tweets</b>	<b>80</b>
<b>Εικόνα 28: Διάγραμμα Author ID να έχει τουλάχιστον 1 tweet στο dataset</b>	<b>81</b>
<b>Εικόνα 29: Διάγραμμα Author ID να έχει τουλάχιστον 400 tweets στο dataset</b>	<b>82</b>
<b>Εικόνα 30: Προσδιορισμός unique authors author_volume_df</b>	<b>83</b>

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

<b>Εικόνα 31: Unique authors to Author volume 40.000 – 60.000 .....</b>	<b>84</b>
<b>Εικόνα 32: Unique authors to Author volume 50-100 .....</b>	<b>85</b>
<b>Εικόνα 33: authors με tweets ανά χρήστη <math>\geq 1</math> .....</b>	<b>87</b>
<b>Εικόνα 34: authors με tweets ανά χρήστη <math>\geq 750</math> .....</b>	<b>87</b>
<b>Εικόνα 35: Επιδόσεις μοντέλου εκπαίδευσης Τέμπη .....</b>	<b>90</b>
<b>Εικόνα 36: Εκπαιδευμένο μοντέλο περίπτωσης Τέμπη.....</b>	<b>91</b>
<b>Εικόνα 37: Επιδόσεις μοντέλου εκπαίδευσης Τέμπη .....</b>	<b>93</b>
<b>Εικόνα 38: Εκπαιδευμένο μοντέλο περίπτωσης Μάτι .....</b>	<b>94</b>



## **1. Εισαγωγή στην Ανάλυση Κοινωνικών Δικτύων**

---

### **1.1 Τι είναι ένα κοινωνικό δίκτυο;**

Ένα κοινωνικό δίκτυο είναι μια ομάδα κοινωνικά-σχετιζόμενων κόμβων οι οποίοι είναι συνδεδεμένοι με μία ή περισσότερες συσχετίσεις. Οι κόμβοι, ή τα μέλη του δικτύου, είναι τα στοιχεία που είναι συνδεδεμένα διαμέσου των κοινών συσχετισμών τους και των οποίων οι διατάξεις που σχηματίζουν μας δημιουργούν και το ενδιαφέρον να τα μελετήσουμε. Ο όρος κοινωνικό δίκτυο συνεπάγεται με την ύπαρξη συνδέσεων τόσο στον πραγματικό όσο και στον ψηφιακό κόσμο. Σήμερα, αυτός ο όρος χρησιμοποιείται κυρίως για να αναφερθεί στις διαδικτυακές κοινωνικές επικοινωνίες. Το διαδίκτυο έχει καταστήσει δυνατή την εύρεση και τη σύνδεση ανθρώπων με άλλους που ίσως δεν θα είχαν την δυνατότητα να γνωριστούν διαφορετικά.

Τα κοινωνικά δίκτυα συνιστούν οι ιστότοποι και οι εφαρμογές που επιτρέπουν σε χρήστες και οργανισμούς να συνδέονται, να επικοινωνούν, να μοιράζονται πληροφορίες και να δημιουργούν σχέσεις. Οι άνθρωποι μπορούν να συνδεθούν με άλλους στην ίδια περιοχή, την οικογένεια, τους φίλους και όσους έχουν τα ίδια ενδιαφέροντα. Τα κοινωνικά δίκτυα είναι ένα από τα πιο σημαντικά μέσα επικοινωνίας του διαδικτύου σήμερα.

Δημοφιλείς κοινωνικές δικτυώσεις, όπως το Facebook, το Twitter, το Instagram και το TikTok, επιτρέπουν στα άτομα να διατηρούν κοινωνικές σχέσεις, να παραμένουν ενημερωμένα και να έχουν πρόσβαση σε πληροφορίες, καθώς και να μοιράζονται μια πληθώρα πληροφοριών.

Το διαδικτυακό κοινωνικό δίκτυο εξαρτάται από την τεχνολογία και τη σύνδεση στο διαδίκτυο. Οι χρήστες μπορούν να έχουν πρόσβαση σε ιστοτόπους κοινωνικών δικτύων χρησιμοποιώντας τους υπολογιστές

τους, τα tablet ή τα smartphones τους. Οι περισσότεροι ιστότοποι κοινωνικών δικτύων λειτουργούν με μια υποκείμενη βάση δεδομένων που μπορεί να αναζητηθεί και να αποθηκεύει με προηγμένες γλώσσες προγραμματισμού, όπως η Python. Αυτό συμβαίνει σε μια προσπάθεια οργάνωσης αποθήκευσης και ανάκτησης δεδομένων με εύκολο και κατανοητό τρόπο. Για παράδειγμα, το Tumblr χρησιμοποιεί τέτοιου είδους προϊόντα και υπηρεσίες στην καθημερινή του λειτουργία, όπως το Google Analytics, και το WordPress. [1]

Με ένα τόσο ευρύ φάσμα ιστοτόπων, εφαρμογών και υπηρεσιών που υπάρχουν στο διαδίκτυο, δεν υπάρχει ένας και μοναδικός ακριβής ορισμός για ένα κοινωνικό δίκτυο. Ωστόσο, τα κοινωνικά δίκτυα έχουν μερικά κοινά χαρακτηριστικά που τα κάνουν να ξεχωρίζουν.

- a. Ένα κοινωνικό δίκτυο επικεντρώνεται στο περιεχόμενο που δημιουργούν οι ίδιοι οι χρήστες. Οι χρήστες κυρίως προβάλλουν και αλληλοεπιδρούν με περιεχόμενο που δημιουργήθηκε από άλλους χρήστες. Κινητοποιούνται να δημοσιεύουν κείμενο ή φωτογραφίες και να τα μοιράζονται με άλλους.
- b. Τα κοινωνικά δίκτυα επιτρέπουν στον χρήστη ή στον οργανισμό να δημιουργήσει ένα προφίλ. Το προφίλ περιέχει πληροφορίες για το άτομο και μια κεντρική σελίδα με το περιεχόμενο που δημοσιεύουν.
- c. Ένα κοινωνικό δίκτυο διαθέτει την δυνατότητα να δομήσει μόνιμες συνδέσεις με άλλους χρήστες. Αυτές οι συνδέσεις ονομάζονται συνήθως "φίλοι" ή "ακόλουθοι" των άλλων χρηστών. Χαρακτηριστικό τους είναι πως επιτρέπουν στους χρήστες να βρίσκουν άλλους χρήστες και να δημιουργούν δίκτυα σχέσεων. Συχνά, ένας αλγόριθμος συνδράμει στο να προτείνει άλλους χρήστες και οργανισμούς με τους οποίους ενδέχεται να θέλουν να συνδεθούν.



Εικόνα 1: Εφαρμογές κοινωνικής δικτύωσης

### 1.1.1 Οι κορυφαίες 10 κοινωνικές δικτυώσεις παγκοσμίως:

- 1) Facebook: 2.96 δισεκατομμύρια μηνιαίοι ενεργοί χρήστες  
Το Facebook είναι το μεγαλύτερο κοινωνικό δίκτυο, με σχεδόν 3 δισεκατομμύρια άτομα να το χρησιμοποιούν μηνιαία. Αυτό σημαίνει ότι περίπου το 37% του παγκόσμιου πληθυσμού είναι χρήστες του Facebook. Περισσότερες από 200 εκατομμύρια επιχειρήσεις (κυρίως μικρές επιχειρήσεις) χρησιμοποιούν εργαλεία του Facebook, και περισσότερο από 7 εκατομμύρια διαφημιστές προωθούν ενεργά την επιχείρησή τους στο Facebook, κάτι που το καθιστά μια ασφαλή επιλογή εάν θέλετε να έχετε παρουσία στα κοινωνικά μέσα.

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

- 2) YouTube: 2.2 δισεκατομμύρια μηνιαίοι ενεργοί χρήστες:  
Το YouTube είναι μια πλατφόρμα κοινοποίησης βίντεο όπου οι χρήστες παρακολουθούν δισεκατομμύρια ώρες βίντεο καθημερινά. Εκτός από το ότι είναι το δεύτερο μεγαλύτερο κοινωνικό δίκτυο, το YouTube συχνά αναφέρεται ως ο δεύτερος μεγαλύτερος μηχανισμός αναζήτησης μετά τη Google, τη μητρική εταιρεία του.
  
- 3) WhatsApp: 2 δισεκατομμύρια μηνιαίοι ενεργοί χρήστες:  
Το WhatsApp είναι μια εφαρμογή ανταλλαγής μηνυμάτων που χρησιμοποιείται από ανθρώπους σε περισσότερες από 180 χώρες. Αρχικά, οι χρήστες χρησιμοποιούσαν το WhatsApp για να στέλνουν κείμενα μηνύματα στην οικογένεια και τους φίλους τους. Σταδιακά, οι άνθρωποι άρχισαν να επικοινωνούν με επιχειρήσεις μέσω του WhatsApp. Η επαγγελματική πλατφόρμα του WhatsApp επιτρέπει στις επιχειρήσεις να παρέχουν υποστήριξη πελατών και να μοιράζονται ενημερώσεις με τους πελάτες τους σχετικά με τις αγορές τους. Για τις μικρές επιχειρήσεις υπάρχει το WhatsApp Business app, ενώ οι μεγάλες επιχειρήσεις μπορούν να χρησιμοποιήσουν το WhatsApp Business API.
  
- 4) Instagram: πάνω 2.35 δισεκατομμύρια μηνιαίοι ενεργοί χρήστες:  
Ως ένα οπτικό κοινωνικό δίκτυο, το Instagram επιτρέπει στους χρήστες να επεξεργαστούν και να ανεβάσουν φωτογραφίες αλλά και βίντεο μικρής διάρκειας μέσω μιας κινητής εφαρμογής. Έτσι μπορεί να μοιραστεί κανείς μια ευρεία γκάμα περιεχομένου, όπως φωτογραφίες, βίντεο, Stories, reels, live videos και IGTV για βίντεο με μεγαλύτερη διάρκεια. Οι χρήστες μπορούν να προσθέσουν μια λεζάντα σε κάθε δημοσίευσή τους και να χρησιμοποιήσουν hashtags και γεωγραφικά geotags για να καταχωρίσουν αυτές τις δημοσιεύσεις παρέχοντας την

δυνατότητα να είναι «αναζητήσιμες» και από άλλους χρήστες εντός της εφαρμογής.

- 5) WeChat: 1.26 δισεκατομμύρια μηνιαίοι ενεργοί χρήστες:  
Το WeChat κυκλοφόρησε το 2011 από την Tencent, μία από τις μεγαλύτερες τεχνολογικές εταιρείες της Κίνας. Όπως το WhatsApp και το Messenger, το WeChat αρχικά ήταν μια εφαρμογή ανταλλαγής μηνυμάτων, πλέον έχει εξελιχθεί σε μια πολυλειτουργική πλατφόρμα. Εκτός από την ανταλλαγή μηνυμάτων και κλήσεων, οι χρήστες μπορούν να κάνουν online αγορές, να πληρώνουν λογαριασμούς, να αγοράζουν τρόφιμα, να μεταφέρουν χρήματα, να κάνουν κρατήσεις, να κλείνουν ταξί και άλλα πολλά.
- 6) TikTok: 1 δισεκατομμύρια μηνιαίοι ενεργοί χρήστες:  
Το TikTok είναι μια εφαρμογή κοινωνικής δικτύωσης για τη δημιουργία και κοινοποίηση βίντεο μικρού μήκους. Παρά το γεγονός ότι κυκλοφόρησε μόλις το 2017, είναι μία από τις ταχύτερα αναπτυσσόμενες εφαρμογές στον κόσμο και πρόσφατα ξεπέρασε τη Google ως το πιο επισκεφθέν διαδικτυακό site. Το TikTok επιτρέπει στους χρήστες να δημιουργούν και να κοινοποιούν βίντεο διάρκειας από 15 έως 60 δευτερολέπτων, διαθέτοντας έναν τεράστιο κατάλογο εφέ ήχου, αποσπασμάτων μουσικής και φίλτρων για να βελτιώνουν τα βίντεο και να τα καθιστούν πιο ελκυστικά.
- 7) Telegram: 550 εκατομμύρια μηνιαίοι ενεργοί χρήστες:  
Το Telegram είναι μια δωρεάν εφαρμογή ανταλλαγής μηνυμάτων που λειτουργεί σε πολλές συσκευές χωρίς περιορισμούς στο μέγεθος των πολυμέσων. Το ξεχωριστό χαρακτηριστικό του Telegram είναι η κρυπτογράφηση end-to-end για όλες τις

δραστηριότητες, συμπεριλαμβανομένων των συνομιλιών, των ομάδων και των πολυμέσων που κοινοποιούνται ανάμεσα στους συμμετέχοντες. Η εστίαση της εφαρμογής στην ασφάλεια έχει προσελκύσει περισσότερους πελάτες με τα χρόνια, ειδικά όταν το WhatsApp ανακοίνωσε αλλαγές στην πολιτική απορρήτου του που θα του επέτρεπαν να μοιράζεται πληροφορίες με τη μητρική εταιρεία Meta. [2]

## **1.2 Ανάλυση κοινωνικών δικτύων**

Η ανάλυση κοινωνικών δικτύων αποτελεί την έρευνα και περιήγηση στις κοινωνικές δομές με την χρήση της θεωρίας των γραφημάτων και της έννοιας του δικτύου. Χαρακτηρίζει έτσι κόμβους, συνδέσεις μεταξύ αυτών αλλά και ακμές προκειμένου να δημιουργήσει μια συνδεσμολογία που καθιστά την διαδικασία αυτή και ως «η μελέτη της δομής», στηριζόμενη τόσο στις θεμελιώδεις βάσεις της κοινωνιολογίας όσο και στις μαθηματικές δομές της θεωρίας των γραφημάτων.

Ο όρος δομή αναφέρεται στις κανονικότητες που δημιουργεί το μοτίβο των σχέσεων μεταξύ των μεμονωμένων αυτών οντοτήτων αλλά και του απόηχου των ομαδοποιήσεων ή ταξινομήσεων τους. Όταν διενεργείται μια τέτοιου είδους ανάλυση θεωρούμε πως η δομή του δικτύου αλλά και οι ιδιότητες του έχουν αξιολογικό ενδιαφέρον προς μελέτη.

Κατά την διάρκεια λοιπόν αυτής της εργασίας θα έχουμε την ευκαιρία να εστιάσουμε περισσότερο στην δομή του δικτύου και λιγότερο στα μεμονωμένα χαρακτηριστικά ή συμπεριφορές των μερών του δικτύου.

Ένα σημείο αναφοράς της διπλωματικής αυτής αποτέλεσε η διαφοροποίηση της από άλλες μελέτες που κατά κανόνα χρησιμοποιούν τα “attribute data” τα δεδομένα δηλαδή τα οποία περιγράφουν τα στοιχειώδη γνωρίσματα μιας οντότητας όπως π.χ. συμπεριφορά. Η ανάλυση κοινωνικών δικτύων που θα επιδιώξουμε σε

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

αυτή την προσέγγιση δεν απαιτεί μόνο τέτοιου είδους δεδομένα αλλά χρησιμοποιεί στην δομή της και τα “relational data”, τα οποία είναι δεδομένα όπου η οντότητα απεικονίζεται, μέσω συνδέσεων, ως η ομαδοποίηση των μεμονωμένων οντοτήτων που διαθέτουν κοινές ιδιότητες σε ένα δίκτυο. [3]

Η ανάλυση κοινωνικών δικτύων αποτυπώνει σε δύο επίπεδα, το micro και το macro, την θεμελιώδη δομή ενός δικτύου. Στο micro επίπεδο το μοτίβο των συσχετίσεων της δομής του δικτύου αποτυπώνεται στην μικρή κλίμακα και διαθέτει την ιδιότητα να είναι εύκολα αντιληπτό και αναγνωρίσιμο. Στο macro επίπεδο, το μοτίβο των συσχετίσεων της δομής του δικτύου εκτυλίσσεται στη μεγάλη κλίμακα χωρίς όμως να μπορεί να γίνει τόσο εύκολα αναγνωρίσιμο όσο το προηγούμενο. Παρά την διαφοροποίησή τους όμως έχουν ως κοινό άξονα, το θεμελιώδες χαρακτηριστικό τους, την οπτικοποίηση των μεταξύ τους αλληλεπιδράσεων.[4]

### **1.3 Μηχανική μάθηση**

Η μηχανική μάθηση είναι ένα πεδίο της τεχνητής νοημοσύνης που επιτρέπει στα συστήματα να μαθαίνουν και να βελτιώνονται από την εμπειρία χωρίς να προγραμματίζονται ρητά. Έχει γίνει ένα ολοένα και πιο δημοφιλές θέμα τα τελευταία χρόνια λόγω των πολλών πρακτικών εφαρμογών της σε πολλούς τομείς. Αποτελεί έτσι θα λέγαμε το μέσο για να έρθει κανείς σε επαφή με την τεχνητή νοημοσύνη. Καθορίστηκε τη δεκαετία του 1950 από τον πρωτοπόρο της τεχνητής νοημοσύνης Arthur Samuel ως "το πεδίο μελέτης που δίνει στους υπολογιστές τη δυνατότητα να μαθαίνουν χωρίς να προγραμματίζονται ρητά." [5]

Η μηχανική μάθηση αποτελεί τον πυρήνα της τεχνητής νοημοσύνης. Οι εφαρμογές μηχανικής μάθησης «μαθαίνουν» από την εμπειρία (ή για

να είμαστε ακριβείς, από δεδομένα) όπως το κάνουν οι άνθρωποι χωρίς κάποια άμεση προγραμματισμένη εντολή. Όταν έρχονται σε επαφή με νέα δεδομένα, αυτές οι εφαρμογές μαθαίνουν, αναπτύσσονται, αλλάζουν και εξελίσσονται από μόνες τους. Έτσι η μηχανική μάθηση «εκμεταλλεύεται» αλγορίθμους που μαθαίνουν από τα δεδομένα εντός ενός επαναληπτικού διαδικαστικού προτύπου.

Σε υψηλό επίπεδο, η μηχανική μάθηση είναι η δυνατότητα προσαρμογής σε νέα δεδομένα ανεξάρτητα και μέσω επαναλήψεων. Οι εφαρμογές μαθαίνουν από προηγούμενους υπολογισμούς και συναλλαγές και χρησιμοποιούν τη "συναναγνώριση μοτίβων" για να παράγουν αξιόπιστα και ενημερωμένα αποτελέσματα.

Καλής ποιότητας δεδομένα τροφοδοτούν τις μηχανές, και διάφοροι αλγόριθμοι χρησιμοποιούνται για την κατασκευή μοντέλων μηχανικής μάθησης προκειμένου να εκπαιδευτούν οι μηχανές με τα δεδομένα αυτά. Η επιλογή του αλγορίθμου εξαρτάται από τον τύπο των δεδομένων που υπάρχουν και τον τύπο της δραστηριότητας που χρειάζεται να αυτοματοποιηθεί.

Στον παραδοσιακό προγραμματισμό, θα τροφοδοτούσαμε με τα δεδομένα εισόδου ένα πρόγραμμα γραμμένο και δοκιμασμένο σε έναν υπολογιστή προκειμένου να παράγει έξοδο. Όταν πρόκειται για μηχανική μάθηση, τα δεδομένα εισόδου, μαζί με την έξοδο, τροφοδοτούνται στη μηχανή κατά τη διάρκεια της φάσης μάθησης, και αυτή εξετάζει ένα πρόγραμμα για τον εαυτό της.[6]

Ο στόχος της τεχνητής νοημοσύνης είναι να δημιουργήσει υπολογιστικά μοντέλα που εκδηλώνουν "ευφυείς συμπεριφορές" όπως οι άνθρωποι, σύμφωνα με τον Boris Katz, κύριο ερευνητή και μέλος της ομάδας InfoLab στο CSAIL. Αυτό σημαίνει πως οι μηχανές πλέον αποκτούν την δυνατότητα να αναγνωρίζουν μια οπτική σκηνή, να κατανοούν ένα κείμενο γραμμένο στη φυσική γλώσσα ή να εκτελούν μια ενέργεια στο φυσικό κόσμο.



Η Μηχανική Μάθηση είναι, αναμφισβήτητα, ένα από τα πιο συναρπαστικά δομικά «κομμάτια» της Τεχνητής Νοημοσύνης καθώς ολοκληρώνει το έργο της μάθησης. Απώτερος σκοπός αυτής της εργασίας, μεταξύ άλλων, είναι να κατανοήσουμε τι κάνει τη Μηχανική Μάθηση να λειτουργεί και ως εκ τούτου πώς μπορεί να χρησιμοποιηθεί στο μέλλον.

Ως διαδικασία Μηχανικής Μάθησης ξεκινά με την εισαγωγή εκπαιδευτικών δεδομένων στον επιλεγμένο αλγόριθμο. Τα εκπαιδευτικά δεδομένα είναι γνωστά ή άγνωστα δεδομένα για την ανάπτυξη του τελικού αλγορίθμου Μηχανικής Μάθησης. Νέα εισαγόμενα δεδομένα τροφοδοτούνται στον αλγόριθμο μηχανικής μάθησης για να ελεγχθεί αν ο αλγόριθμος λειτουργεί σωστά.[7]

Η πρόβλεψη και τα αποτελέσματα στη συνέχεια ελέγχονται μεταξύ τους. Εάν η πρόβλεψη και τα αποτελέσματα δεν ταιριάζουν, ο αλγόριθμος εκπαιδεύεται ξανά πολλές φορές μέχρι να λάβουμε το επιθυμητό αποτέλεσμα. Αυτό επιτρέπει στον αλγόριθμο μηχανικής μάθησης να μαθαίνει συνεχώς μόνος του και να παράγει τη βέλτιστη απάντηση, αυξάνοντας σταδιακά την ακρίβειά του με τον χρόνο.

Υπάρχουν επτά βήματα της μηχανικής μάθησης τα οποία παρατίθενται παρακάτω:

1. Συλλογή Δεδομένων
2. Προετοιμασία των δεδομένων
3. Επιλογή ενός μοντέλου
4. Εκπαίδευση
5. Αξιολόγηση
6. Ρύθμιση υπερπαραμέτρων
7. Πρόβλεψη

Αυτά τα βήματα αντιπροσωπεύουν τη διαδικασία επίλυσης προβλημάτων μέσω της μηχανικής μάθησης, όπου η συλλογή και η προετοιμασία των δεδομένων είναι σημαντικοί πυλώνες για την επίτευξη επιτυχημένων αποτελεσμάτων. Αυτό καθιστά και την επιλογή ενός κατάλληλου μοντέλου κρίσιμη συμπεριλαμβανομένης και της αξιολόγησης των αποτελεσμάτων αλλά και της πρόβλεψης που αποτελεί και το τελικό στάδιο της διαδικασίας.[8]

### 1.3.1 Κατηγορίες μηχανικής μάθησης

- Επιβλεπόμενη μάθηση (Supervised Learning) :

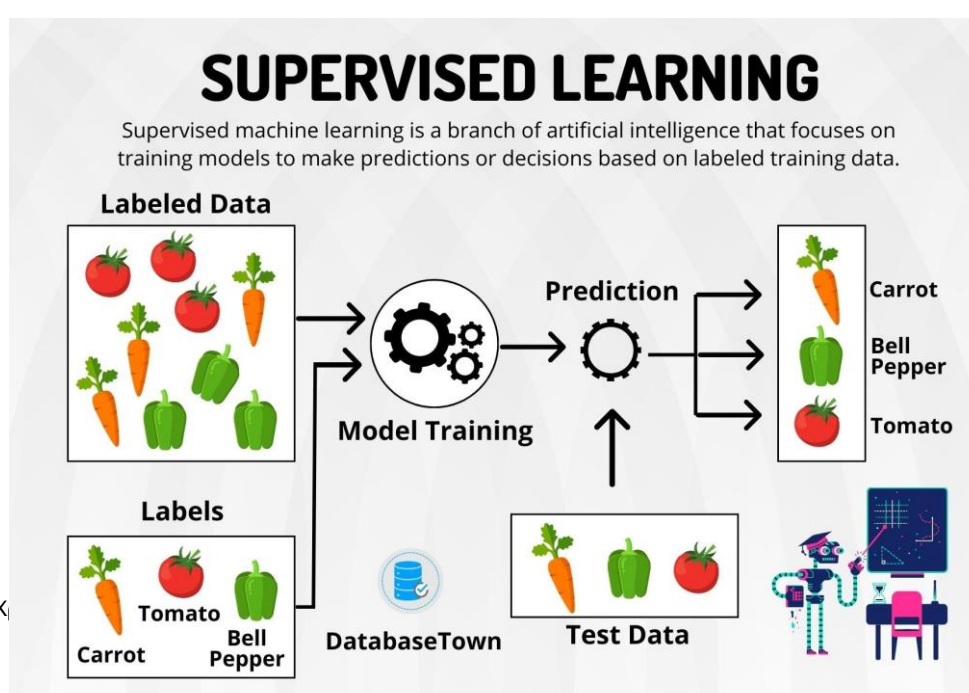
Η επιβλεπόμενη μάθηση αποτελεί μια μέθοδο μηχανικής μάθησης στην οποία τα μοντέλα εκπαιδεύονται χρησιμοποιώντας δεδομένα με ετικέτες. Στην επίβλεψη, τα μοντέλα πρέπει να βρουν τη συνάρτηση αντιστοίχισης που συνδέει τη μεταβλητή εισόδου (X) με τη μεταβλητή εξόδου (Y). [9]

$$Y = f(X)$$

Η επιβλεπόμενη μάθηση χρειάζεται την επίβλεψη για την εκπαίδευση του μοντέλου, παρόμοια με τον τρόπο που ένας μαθητής μαθαίνει πράγματα μέσω της καθοδήγησης ενός δασκάλου. Ακολούθως εξηγούνται οι βασικές αρχές λειτουργίας της επίβλεψης:

- I. Δεδομένα με Ετικέτες: Παίρνουμε ένα σύνολο δεδομένων που περιλαμβάνει δεδομένα εισόδου (χαρακτηριστικά) και τις αντίστοιχες σωστές εξόδους (ετικέτες ή στόχους). Αυτά τα δεδομένα χρησιμοποιούνται για την εκπαίδευση του μοντέλου.

- II. Εκπαίδευση: Το μοντέλο μηχανικής μάθησης παρουσιάζεται με τα δεδομένα που έχουν αποκτήσει ετικέτες, και μαθαίνει να αντιστοιχίζει τα δεδομένα εισόδου στη σωστή έξοδο, εντοπίζοντας πρότυπα και σχέσεις μέσα στα δεδομένα.
- III. Δοκιμές και Αξιολόγηση: Αφού το μοντέλο εκπαιδευτεί, ελέγχεται σε νέα, μη αναγνωρισμένα δεδομένα για να αξιολογηθεί η απόδοσή του. Οι προβλέψεις του μοντέλου συγκρίνονται με τα πραγματικά σωστά αποτελέσματα για να μετρηθεί η ακρίβεια και η απόδοσή του.
- IV. Προσαρμογή: Με βάση τα αποτελέσματα της αξιολόγησης, το μοντέλο μπορεί να προσαρμοστεί ή να βελτιωθεί για να βελτιώσει την απόδοσή του. Αυτή η διαδικασία μπορεί να περιλαμβάνει την αλλαγή των υπερπαραμέτρων, την προσθήκη περισσότερων δεδομένων ή τη δοκιμή διαφορετικών αλγορίθμων.
- V. Πρόβλεψη: Αφού πετύχει την εκπαίδευση και την αξιολόγηση, το μοντέλο μπορεί να χρησιμοποιηθεί για να κάνει προβλέψεις ή κατηγοριοποιήσεις σε νέα δεδομένα. [10]



Εικόνα 2: Διάγραμμα επιβλεπόμενης μάθησης

- Μη επιβλεπόμενη μάθηση (Unsupervised Learning):

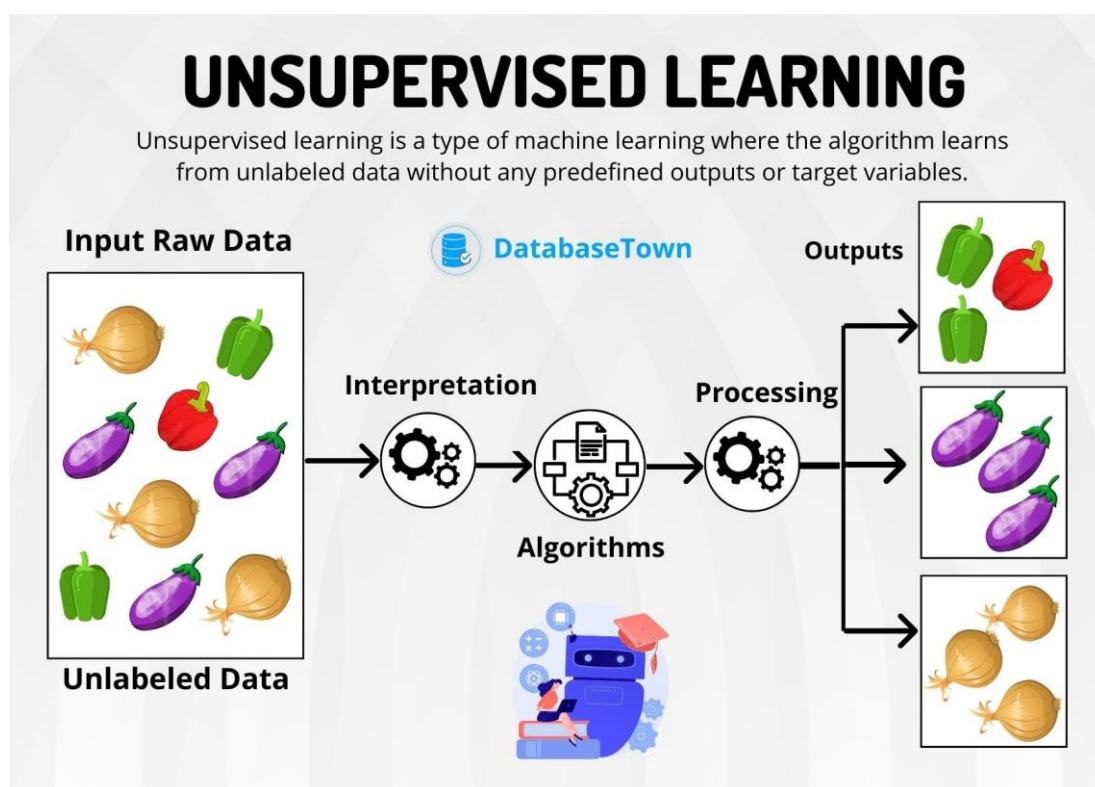
Η μη επιβλεπόμενη μηχανική μάθηση είναι μια άλλη μέθοδος της μηχανικής μάθησης, στην οποία εξάγονται μοτίβα από τα χωρίς ετικέτες δεδομένα εισόδου. Ο στόχος της μη επιβλεπόμενης μάθησης είναι να εντοπίσει τη δομή και τα πρότυπα από τα δεδομένα εισόδου. Έτσι αυτού του είδους η μάθηση δεν χρειάζεται καμία επίβλεψη αντ' αυτού, βρίσκει πρότυπα από τα δεδομένα μόνη της.

Η μη υποστηριζόμενη μάθηση λειτουργεί εξετάζοντας τα δεδομένα για να βρει πρότυπα, δομές ή κρυμμένες σχέσεις χωρίς την ανάγκη για επισημασμένες εξόδους ή επίβλεψη. Εδώ είναι μια γενική επισκόπηση του πώς λειτουργούν οι αλγόριθμοι μη επιτηρούμενης μάθησης:

- I. Συλλογή Δεδομένων: Χρησιμοποιεί ένα σύνολο δεδομένων που περιέχει είσοδο, όπου κάθε δείγμα περιλαμβάνει πολλές χαρακτηριστικές ή ιδιότητες. Αυτά τα δεδομένα μπορούν να συλλεγούν από διάφορες πηγές και ενδέχεται να είναι επαρκώς δομημένα.
  - II. Προ επεξεργασία Δεδομένων: Πριν την εφαρμογή αλγορίθμων μη επιτηρούμενης μάθησης, είναι απαραίτητο να γίνει προ επεξεργασία δεδομένων. Αυτό περιλαμβάνει εργασίες όπως η αντιμετώπιση ελλιπών τιμών, η κανονικοποίηση ή κλιμάκωση των χαρακτηριστικών και η αφαίρεση άσχετων ή θορυβωδών δεδομένων.
- Επιλογή Αλγορίθμου μη επιτηρούμενης Μάθησης: Εφόσον επιλεγθεί ο κατάλληλος αλγόριθμος βάσει του συγκεκριμένου προβλήματος που επιδιώκεται να επιλυθεί. Οι συνηθέστεροι αλγόριθμοι που χρησιμοποιούνται περιλαμβάνουν :

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

- Συσταδοποίηση (clustering): Ομαδοποίηση των δειγμάτων δεδομένων σε συστάδες ή τμήματα βάσει ομοιοτήτων. Η συσταδοποίηση γίνεται μέσω της μεθόδου K-Means.
- Μείωση Διαστάσεων: Μείωση του αριθμού των χαρακτηριστικών ενώ διατηρούνται οι σημαντικές πληροφορίες. Σε αυτού του είδους την μάθηση μη επιβλεπόμενης μάθησης χρησιμοποιείται η ανάλυση κύριων συνιστωσών (Principal Component Analysis - PCA) ως μια τεχνική επαναπροσδιορισμού των συντεταγμένων συνόλου δεδομένων στο ιδανικότερο για την ανάλυση αυτών. Επίσης χρησιμοποιείται και η τεχνική t-distributed Stochastic Neighbor Embedding (t-SNE) για την οπτικοποίηση και εξερεύνηση δεδομένων ιδιαίτερα διάστασης σε ένα χαμηλότερης διάστασης χώρο, καθιστώντας πιο εύκολο τον κατανοητό της βαθύτερης δομής και των σχέσεων μέσα στα δεδομένα [11]
- Εξόρυξη Κανόνων Συσχέτισης: Εντοπισμός ενδιαφερόντων σχέσεων ή συσχετίσεων μεταξύ στοιχείων δεδομένων. Οι μέθοδοι Apriori και FP-Growth χρησιμοποιούνται για την ανάλυση των κανόνων συσχέτισης στον τομέα της ανάλυσης του καλαθιού αγορών.
- Εκπαίδευση του Μοντέλου: Στην μη επιτηρούμενη μάθηση, δεν υπάρχει η έννοια των δεδομένων με ετικέτες ή στόχων εξόδου. Αντί για αυτό, ο αλγόριθμος μη επιτηρούμενης μάθησης μαθαίνει πρότυπα αναλύοντας τα ίδια τα δεδομένα.
- Ανακάλυψη Προτύπων: Ο αλγόριθμος μη επιτηρούμενης μάθησης εξετάζει τα δεδομένα για τον εντοπισμό προτύπων ή δομών. Για παράδειγμα, στη συσταδοποίηση, μπορεί να καθορίσει πώς πρέπει να ομαδοποιηθούν τα δείγματα δεδομένων σε συστάδες, ενώ στη μείωση της διάστασης, εντοπίζει τα πιο σημαντικά χαρακτηριστικά.



Εικόνα 3: Διάγραμμα μη επιβλεπόμενης μάθησης

## 1.4 Νευρωνικό Δίκτυο

Τα νευρωνικά δίκτυα είναι ένα ευρέως χρησιμοποιούμενο, συγκεκριμένο είδος αλγορίθμων μηχανικής μάθησης. Τα τεχνητά νευρωνικά δίκτυα είναι μοντελοποιημένα πάνω στον ανθρώπινο εγκέφαλο, όπου χιλιάδες ή εκατομμύρια κόμβοι επικοινωνούν μεταξύ τους και οργανώνονται σε στρώματα.

Σε ένα τεχνητό νευρωνικό δίκτυο, κύτταρα ή κόμβοι είναι συνδεδεμένοι, με κάθε κύτταρο να επεξεργάζεται εισόδους και να παράγει έξοδο που στέλνεται σε άλλα νευρώνες. Τα δεδομένα με ετικέτα κινούνται μέσα από τους κόμβους ή τα κύτταρα, με κάθε κύτταρο να εκτελεί διαφορετική λειτουργία. Σε ένα νευρωνικό δίκτυο που έχει εκπαιδευτεί να αναγνωρίζει εάν μια φωτογραφία περιλαμβάνει γάτα ή όχι, οι διάφοροι κόμβοι θα αξιολογήσουν την

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

πληροφορία και θα φτάσουν σε μια έξοδο που υποδηλώνει εάν μια φωτογραφία περιέχει γάτα.

Τα νευρωνικά δίκτυα αποτελούνται από στρώματα συνδεδεμένων νευρώνων και μαθαίνουν προσαρμόζοντας τα βάρη των συνδέσεων μεταξύ των νευρώνων. Είναι επί της ουσίας ένας τύπος υπολογιστικού συστήματος που σχεδιάστηκε για να μιμείται τη λειτουργία του ανθρώπινου εγκεφάλου. Είναι παρόμοια με άλλους αλγόριθμους μηχανικής μάθησης, αλλά αποτελούνται από έναν μεγάλο αριθμό συνδεδεμένων κόμβων επεξεργασίας, ή νευρώνων, που μπορούν να μάθουν να αναγνωρίζουν πρότυπα εισόδου δεδομένων. Τα νευρωνικά δίκτυα είναι ικανά για βαθιά μάθηση, που είναι ένας τύπος μάθησης που τους επιτρέπει να εξάγουν πολύπλοκα πρότυπα από δεδομένα. Τα βαθιά νευρωνικά δίκτυα αποτελούνται από πολλά επίπεδα νευρώνων και μπορούν να μάθουν να αναγνωρίζουν πρότυπα που είναι πολύ δύσκολα να αναγνωρίσουν οι άνθρωποι.[12]

Αποτελούν έτσι ένα ισχυρό εργαλείο που μπορεί να χρησιμοποιηθεί για την επίλυση διάφορων προβλημάτων. Μία από τις πιο γνωστές εφαρμογές των νευρωνικών δικτύων είναι η αναγνώριση εικόνων. Τα νευρωνικά δίκτυα μπορούν να εκπαιδευτούν να αναγνωρίζουν αντικείμενα σε ψηφιακές εικόνες, έχοντας μια ευρεία γκάμα δυνητικών εφαρμογών. Η παρακάτω λίστα αναφέρει διάφορα είδη προβλημάτων που μπορούν να αντιμετωπιστούν χρησιμοποιώντας νευρωνικά δίκτυα:

- Αναγνώριση φωνής
- Αναγνώριση εικόνας
- Αναγνώριση ομιλίας
- Μετάφραση γλώσσας
- Αναγνώριση προτύπων
- Ανίχνευση ανωμαλιών

Για να χρησιμοποιηθούν τα νευρωνικά δίκτυα αποτελεσματικά, υπάρχουν μερικές βασικές απαιτήσεις που πρέπει να τηρηθούν:

- 1) Το σύνολο δεδομένων πρέπει να είναι αρκετά μεγάλο ώστε να παρέχει στο νευρωνικό δίκτυο αρκετές πληροφορίες για να μάθει και να γενικεύει σε άγνωστα δεδομένα με ακρίβεια. Όσο πιο μεγάλη η ύπαρξη ποικιλομορφίας τόσο πιο αντιπροσωπευτικά είναι τα δεδομένα και τόσο καλύτερα θα είναι τα δυνητικά αποτελέσματα.
- 2) Για προβλήματα εποπτευόμενης μάθησης, τα δεδομένα πρέπει να είναι καλά οργανωμένα, ώστε το νευρωνικό δίκτυο να μπορεί να αναγνωρίζει πρότυπα με ακρίβεια.
- 3) Τα δεδομένα πρέπει να είναι υψηλής ποιότητας, χωρίς ακραία σημεία ή σφάλματα.
- 4) Πρέπει να σχεδιαστεί κατάλληλη αρχιτεκτονική νευρωνικού δικτύου όσον αφορά τους τύπους νευρωνικού δικτύου, τον αριθμό των στρωμάτων, τον αριθμό των κόμβων σε αυτά τα στρώματα, κ.λπ.
- 5) Κατάλληλη περίοδος εκπαίδευσης: Ο αλγόριθμος εκπαίδευσης πρέπει να είναι κατάλληλος, συμπεριλαμβανομένης της στοχαστικής καθοδήγησης κίνησης ή της ανάκλησης, καθώς και της κατάλληλης επιλογής υπερπαραμέτρων, όπως η ταχύτητα μάθησης, το μέγεθος του πακέτου, κ.λπ.
- 6) Το νευρωνικό δίκτυο πρέπει να έχει στην διάθεσή του αρκετό χρόνο για εκπαίδευση στα δεδομένα πριν από την υλοποίησή του. Αυτό εξαρτάται επίσης από τους πόρους υπολογισμού (μνήμη, επεξεργαστική ισχύ, κ.λπ.) που χρησιμοποιούνται για την εκπαίδευση.

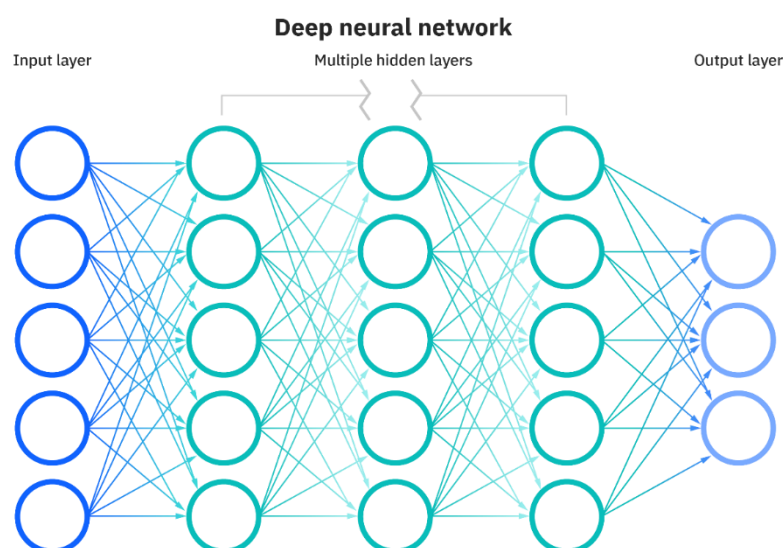


7) Τα νευρωνικά δίκτυα μπορεί να είναι αργά στην εκπαίδευση και να απαιτούν μεγάλο όγκο δεδομένων εκπαίδευσης. Μπορεί επίσης να είναι δύσκολο να τα ερμηνεύσει κανείς και μερικές φορές, να προκύπτουν λάθη που είναι δύσκολο να τα κατανοήσει κανείς. Παρά τα περιορισμένα αυτά χαρακτηριστικά, που αναφέρθηκαν, τα νευρωνικά δίκτυα αποτελούν μια ευρέως χρησιμοποιούμενη τεχνική για την μηχανική μάθηση και την τεχνητή νοημοσύνη.[13]

#### 1.4.1 Τύποι νευρωνικών δικτύων:

- Κανονικό τεχνητό νευρωνικό δίκτυο (ANN):

Τα κανονικά τεχνητά νευρωνικά δίκτυα είναι νευρωνικά δίκτυα με πολλά κρυφά στρώματα. Αποτελούνται από αρκετούς διασυνδεδεμένους κόμβους επεξεργασίας, ή νευρώνες, που επικοινωνούν μεταξύ τους μέσω συνάψεων. Σε ένα κανονικό νευρωνικό δίκτυο, κάθε νευρώνας (κόμβος) λαμβάνει είσοδο από αρκετούς άλλους νευρώνες και παράγει έξοδο που περνά σε άλλους νευρώνες στο δίκτυο.



Εικόνα 4: Κανονικό νευρωνικό δίκτυο

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

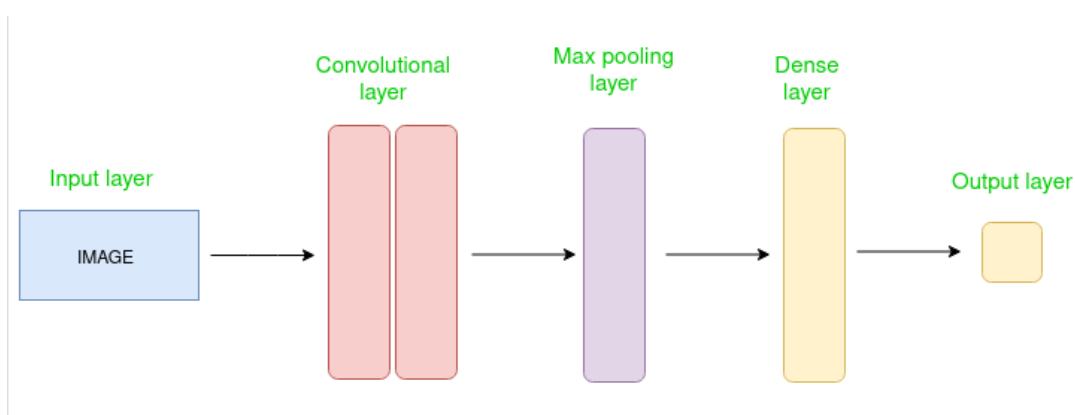
Η δύναμη των συνδέσεων μεταξύ νευρώνων (κόμβων), που είναι γνωστή ως συναπτικά βάρη, καθορίζει πόση επιρροή έχει ένας νευρώνας (κόμβος) σε έναν άλλο. Τα κανονικά νευρωνικά δίκτυα είναι σχεδιασμένα να μαθαίνουν προσαρμόζοντας τα βάρη ανάλογα με τα εισερχόμενα δεδομένα. Τα κανονικά νευρωνικά δίκτυα εκπαιδεύονται χρησιμοποιώντας μια τεχνική που ονομάζεται αναδρομική διάδοση (backpropagation). Η αναδρομική διάδοση περιλαμβάνει τη ρύθμιση των βαρών των συνδέσεων μεταξύ των κόμβων με βάση το πόσο καλά αποδίδει το δίκτυο σε ένα σύνολο εκπαίδευσης δεδομένων. Είναι ικανά να εκτελούν πολύπλοκες εργασίες, όπως η αναγνώριση προτύπων και οι προβλέψεις. [14]

- Συνελικτικό νευρωνικό δίκτυο (CNN):

Τα CNN είναι ένας τύπος νευρωνικού δικτύου που είναι κατάλληλος για εργασίες ταξινόμησης εικόνων. Τα CNN νευρωνικά δίκτυα αποτελούνται από μια σειρά συνελικτικών στρωμάτων που εξάγουν χαρακτηριστικά από τις εικόνες, και από στρώματα συσσώρευσης που μειώνουν τις διαστάσεις των χαρακτηριστικών χαρτών. Έτσι ένα συνελικτικό στρώμα εφαρμόζει ένα σύνολο φίλτρων σε μια εισερχόμενη εικόνα, με αποτέλεσμα να δημιουργηθεί ένα σύνολο χαρακτηριστικών χαρτών που αντιπροσωπεύουν διάφορα πρότυπα στην εικόνα, και ένα στρώμα συσσώρευσης χρησιμοποιείται για τη μείωση της διαστατικότητας της εξόδου από το συνελικτικό στρώμα με αποδεδειγμένη δειγματοληψία των χαρακτηριστικών χαρτών. Τα CNN λειτουργούν με τον τρόπο του συνελικτικού φιλτραρίσματος μιας εικόνας με ένα σύνολο φίλτρων που σχεδιάζονται για τον εντοπισμό ενός ή περισσότερων συγκεκριμένων χαρακτηριστικών στην εικόνα.

Η έξοδος του CNN στη συνέχεια τροφοδοτείται σε ένα πλήρως συνδεδεμένο στρώμα, το οποίο παράγει την τελική ταξινόμηση. Τα CNN μπορούν να μάθουν να ανιχνεύουν πολύπλοκα πρότυπα στις εικόνες και έχουν χρησιμοποιηθεί για διάφορες εργασίες, όπως η ανίχνευση και η ταξινόμηση αντικειμένων. Χρησιμοποιούνται επίσης για την αναγνώριση προσώπων και την αναγνώριση ανθρώπων σε φωτογραφίες και βίντεο. Επιπλέον έχουν χρησιμοποιηθεί για τη δημιουργία αυτόνομων αυτοκινήτων.

Τέλος χρησιμοποιούνται επίσης σε εφαρμογές ιατρικής εικονικής απεικόνισης. Για παράδειγμα, τα CNN μπορούν να χρησιμοποιηθούν για τον εντοπισμό συμπτωμάτων νόσου σε ακτινογραφίες και CT. Τέλος, τα CNN χρησιμοποιούνται σε εφαρμογές επεξεργασίας φυσικής γλώσσας.



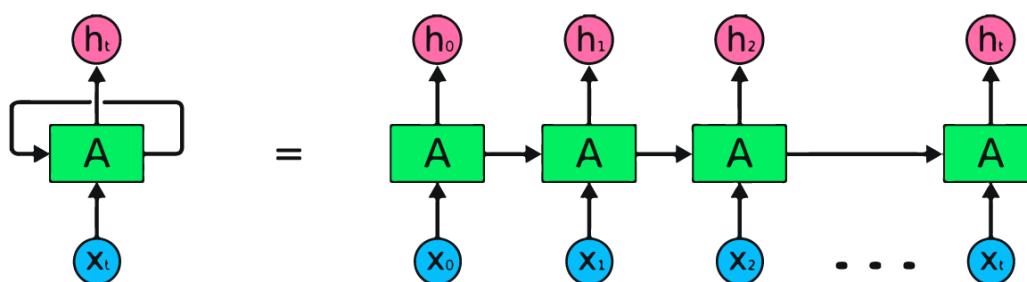
Εικόνα 5: Διάγραμμα συνελκτικού νευρωνικού δικτύου.

- Ανατροφοδοτούμενο Νευρωνικό δίκτυο (RNN):

Τα RNN είναι βαθιά νευρωνικά δίκτυα που έχουν τη δυνατότητα να αποθηκεύουν πληροφορίες από προηγούμενους υπολογισμούς και να τις προωθούν προς τα εμπρός για να εργαστούν σε αυτά τα δεδομένα με συνεχόμενο τρόπο.

**\*ΑΝΑΛΥΣΗ ΑΚΡΙΒΗΣ ΣΤΟ ΚΕΦΑΛ 2.4 \***

Η παρακάτω εικόνα είναι μια απλή αναπαράσταση των αναδρομικών νευρωνικών δικτύων. Εάν προβλέπουμε τις τιμές των μετοχών χρησιμοποιώντας απλά δεδομένα [45,56,45,49,50,...], κάθε είσοδος από το  $X_0$  έως το  $X_t$  θα περιέχει μια παλαιότερη τιμή. Για παράδειγμα, το  $X_0$  θα έχει την τιμή 45, το  $X_1$  θα έχει την τιμή 56, και αυτές οι τιμές χρησιμοποιούνται για να προβλέψουν τον επόμενο αριθμό σε μια ακολουθία.



Εικόνα 6: Διάγραμμα ανατροφοδοτούμενου Νευρωνικού δικτύου

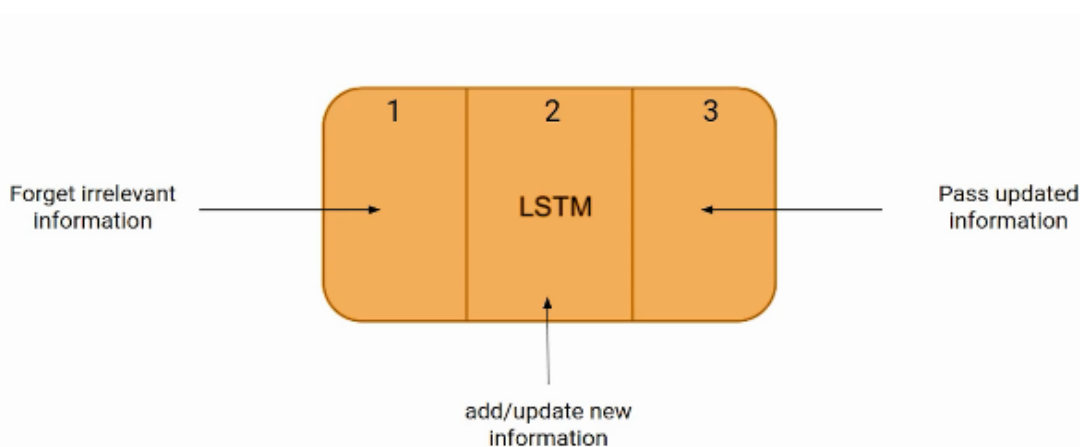
Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

- Long Short-Term Memory (LSTM):

Τα LSTM είναι βαθιά νευρωνικά δίκτυα που διαθέτουν έναν μηχανισμό για να αποθηκεύουν πληροφορίες για μεγάλα χρονικά διαστήματα. Αυτό τους επιτρέπει να μαθαίνουν από εμπειρίες που εκτείνονται σε πολλά χρονικά βήματα. Τα LSTM είναι ένας τύπος τεχνητού νευρωνικού δικτύου που χρησιμοποιείται για την επεξεργασία ακολουθιακών δεδομένων.

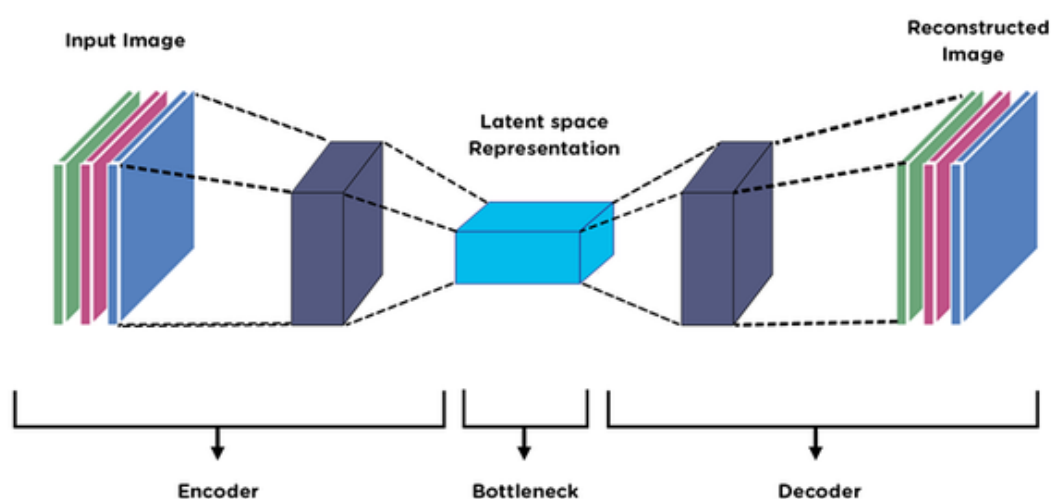
Αντίθετα από τα παραδοσιακά νευρωνικά δίκτυα, τα LSTM μπορούν να θυμούνται μακροπρόθεσμες εξαρτήσεις, κάτι που τα καθιστά κατάλληλα για εργασίες όπως η μηχανική μετάφραση και η αναγνώριση φωνής. Τα LSTM χρησιμοποιούνται επίσης σε πολλές άλλες εφαρμογές, όπως η πρόβλεψη των τιμών των μετοχών και η δημιουργία μουσικής.[15]

Ένα από τα βασικά πλεονεκτήματα των LSTM είναι ότι μπορούν να εκπαιδευτούν σε σχετικά μικρά σύνολα δεδομένων. Αυτό συμβαίνει διότι τα LSTM είναι σε θέση να μαθαίνουν από τα πλαίσια, γεγονός που τους επιτρέπει να γενικεύουν καλύτερα από άλλους τύπους νευρωνικών δικτύων. Ως αποτέλεσμα, τα LSTM έχουν γίνει ένας από τους πιο δημοφιλείς τύπους νευρωνικών δικτύων για εφαρμογές επεξεργασίας φυσικής γλώσσας (NLP).



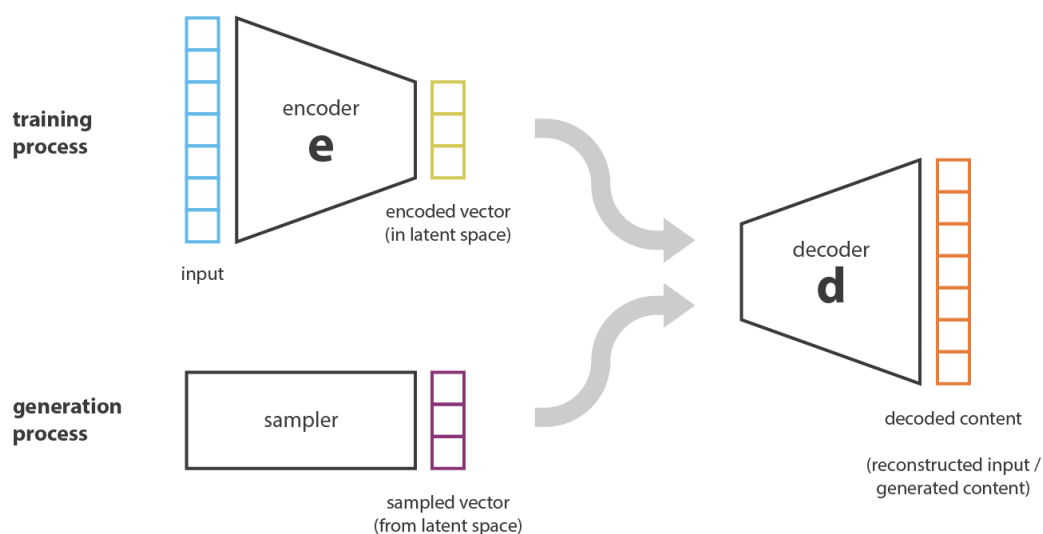
Εικόνα 7: Αναπαράσταση νευρωνικού δικτύου LSTM

- **Stacked Autoencoders** : Αποτελεί ένα τύπο νευρωνικού δικτύου που χρησιμοποιείται για τη μάθηση αποτελεσματικών αναπαραστάσεων των δεδομένων. Είναι ένα σταθμισμένο αυτοματοποιημένο δίκτυο κωδικοποίησης όπου τα εισερχόμενα δεδομένα κωδικοποιούνται αρχικά από ένα νευρωνικό δίκτυο, και στη συνέχεια το αποτέλεσμα αυτού του δικτύου χρησιμοποιείται ως είσοδος σε ένα άλλο δίκτυο κωδικοποιητή. Με το να στοιβάζονται πολλοί κωδικοποιητές ο ένας πάνω στον άλλον αυτό καθιστά δυνατή την εκμάθηση όλο και πιο πολύπλοκων αναπαραστάσεων των δεδομένων. Αυτό μπορεί να είναι χρήσιμο για εργασίες όπως η αναγνώριση εικόνας, όπου απαιτείται μια βαθιά κατανόηση των δεδομένων για την πραγματοποίηση ακριβών προβλέψεων. Επιπλέον αυτή η κατηγορία νευρωνικού δικτύου μπορεί να χρησιμοποιηθεί για την αρχικοποίηση των βαρών ενός βαθιού νευρωνικού δικτύου, πράγμα που μπορεί να βελτιώσει την απόδοση του δικτύου.



Εικόνα 8:Νευρωνικό δίκτυο stacked autoencoder

- Variational autoencoders: Αποτελεί ένα τύπο νευρωνικού δικτύου που χρησιμοποιείται ευρέως, η βασική ιδέα πίσω από αυτούς τους αυτόνομους κωδικοποιητές είναι να μάθουν μια κρυφή αναπαράσταση των δεδομένων που έχει μικρότερη διάσταση από την είσοδο. Αυτό επιτυγχάνεται εκπαιδύοντας το δίκτυο να ελαχιστοποιεί την απόκλιση KL μεταξύ της κρυφής αναπαράστασης και των εισερχομένων δεδομένων. Αυτή η κατηγορία νευρωνικού δικτύου έχουν αποδειχθεί ότι είναι πολύ αποτελεσματική στην εκμάθηση πολύπλοκων κατανομών και έχουν χρησιμοποιηθεί για εφαρμογές όπως η δημιουργία εικόνων και η δημιουργία κειμένου. Η διαφορά μεταξύ των variational encoders και των stacked encoders είναι ότι οι τελευταίοι μαθαίνουν μια συμπιεσμένη έκδοση των εισερχόμενων δεδομένων, ενώ οι variational μια πιθανοτική κατανομή.[16]



Εικόνα 9:Νευρωνικό δίκτυο variational autoencoder

### 1.4.2 Βαθιά Νευρωνικά Δίκτυα (Deep learning)

Τα βαθιά νευρωνικά δίκτυα είναι νευρωνικά δίκτυα με πολλά επίπεδα. Το επίπεδο δίκτυο μπορεί να επεξεργαστεί μεγάλες ποσότητες δεδομένων και να καθορίσει το "βάρος" κάθε σύνδεσης στο δίκτυο π.χ. σε ένα σύστημα αναγνώρισης εικόνων, ορισμένα επίπεδα του νευρωνικού δικτύου μπορεί να ανιχνεύσουν τα μεμονωμένα χαρακτηριστικά ενός προσώπου, όπως μάτια, μύτη ή στόμα, ενώ ένα άλλο επίπεδο θα μπορούσε να πει εάν αυτά τα χαρακτηριστικά εμφανίζονται με έναν τρόπο που υποδηλώνει ένα πρόσωπο.[17]

Όπως και τα νευρωνικά δίκτυα, η βαθιά μάθηση μοντελοποιείται με βάση τον τρόπο λειτουργίας του ανθρώπινου εγκεφάλου και τροφοδοτεί πολλές εφαρμογές μηχανικής μάθησης, όπως αυτόνομα οχήματα, chatbots και ιατρικές διαγνώσεις. Η βαθιά μάθηση απαιτεί μεγάλη υπολογιστική ισχύ, πράγμα που δημιουργεί ανησυχίες για την οικονομική και περιβαλλοντική της βιωσιμότητα.

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων



- a) Πρόβλεψη τιμής κατοικίας: Ένα τυπικό τεχνητό νευρωνικό δίκτυο (ANN) μπορεί να χρησιμοποιηθεί για την πρόβλεψη των τιμών ακινήτων στην αγορά ακινήτων. Η προσέγγιση της βαθιάς μάθησης μπορεί επίσης να χρησιμοποιηθεί για την πρόβλεψη των τιμών κατοικιών σε μια συγκεκριμένη περιοχή, πόλη ή χώρα με υψηλή ακρίβεια και χαμηλό κίνδυνο. Τα εισερχόμενα δεδομένα μπορεί να είναι διάφορα χαρακτηριστικά των κατοικιών και η πρόβλεψη εξόδου θα είναι μια εκτίμηση τιμολόγησης. Αυτό είναι ένα πρόβλημα επιβλεπόμενης μάθησης.
- b) Εάν ο χρήστης θα κάνει κλικ σε μια διαφήμιση: Ένα τυπικό ANN μπορεί να χρησιμοποιηθεί για να προβλέψει εάν ένας χρήστης θα κάνει κλικ σε μια διαφήμιση ή όχι. Τα δεδομένα εισόδου είναι η διαφήμιση και η πληροφορία του χρήστη και η έξοδος μπορεί να είναι μια ετικέτα όπως κλικ (1) ή μην κλικ (0). Αυτό είναι ένα πρόβλημα επιβλεπόμενης μάθησης.
- c) Πρόβλεψη καιρού: Ένα αναδρομικό νευρωνικό δίκτυο (RNN) ή LSTM μπορεί να χρησιμοποιηθεί για την πρόβλεψη του καιρού, καθώς τα δεδομένα είναι χρονικά ή ακολουθίας και βασίζονται σε σειρές χρόνου. Σε αυτή την περίπτωση μπορεί να χρησιμοποιηθεί ένα προσαρμοσμένο ή υβριδικό μοντέλο που κατασκευάζεται χρησιμοποιώντας αλγόριθμους δικτύων που βασίζονται στον χρόνο (RNN / LSTM) και CNN. [18]

## 1.5 Σύνδεση Νευρωνικών και Κοινωνικών δικτύων

Τα κοινωνικά δίκτυα, όπως το Facebook, το Twitter και το Instagram, έχουν διευκολύνει σημαντικά την επικοινωνία μεταξύ χρηστών στον παγκόσμιο ιστό. Η ανάλυση των κοινωνικών δικτύων βοηθά στη σύνοψη των ενδιαφερόντων και των απόψεων των χρηστών (κόμβοι), στην ανακάλυψη προτύπων από τις αλληλεπιδράσεις (σύνδεσμοι) μεταξύ των χρηστών και στον εντοπισμό των γεγονότων που λαμβάνουν χώρα σε διαδικτυακές πλατφόρμες. Οι πληροφορίες που αποκτώνται από την ανάλυση των κοινωνικών δικτύων μπορεί να είναι ιδιαίτερα αξιόλογες για πολλές εφαρμογές. Μερικά τυπικά παραδείγματα περιλαμβάνουν την εξυπηρέτηση στοχευμένων διαφημίσεων στο διαδίκτυο, την προσωπική σύσταση, την κοινωνική υγεία, την ανάλυση της κοινωνικής επίδρασης και την ανάλυση των ακαδημαϊκών δικτύων.

Ένα κεντρικό πρόβλημα στην ανάλυση κοινωνικών δικτύων είναι πώς να εξάγουμε χρήσιμα χαρακτηριστικά από μη-Ευκλείδεια δομημένα δίκτυα, ώστε να επιτραπεί η χρήση μοντέλων μηχανικής μάθησης για συγκεκριμένη ανάλυση. Για παράδειγμα, στην περίπτωση της πρότασης νέων φίλων σε έναν χρήστη σε ένα κοινωνικό δίκτυο, η κύρια πρόκληση μπορεί να είναι πώς να ενσωματώσουμε τους χρήστες του δικτύου σε ένα χαμηλής διάστασης χώρο, ώστε η κοντινότητα μεταξύ των χρηστών να μπορεί να μετρηθεί εύκολα με μετρικές απόστασης. Για την επεξεργασία πληροφοριών δομής σε δίκτυα, οι περισσότερες προηγούμενες προσπάθειες βασίζονταν κυρίως σε χειροποίητα χαρακτηριστικά, όπως πυρήνες (kernel functions), στατιστικά δικτύων (π.χ. συντελεστές συστάσεων) ή άλλα προσεκτικά σχεδιασμένα χαρακτηριστικά. Ωστόσο, η διαδικασία αυτή της μηχανικής χαρακτηριστικών μπορεί να είναι πολύ χρονοβόρα και ακριβή, καθιστώντας την ανεπαρκή για πολλές εφαρμογές στην πραγματική ζωή. [19]

Ο στόχος είναι δομήσουμε μια συνάρτηση μετασχηματισμού που αντιστοιχεί κόμβους, υποδομές ή ακόμη και ολόκληρα δίκτυα σε διανύσματα χαρακτηριστικών χαμηλής διάστασης, όπου οι χωρικές σχέσεις μεταξύ των διανυσμάτων αντανakλούν τις δομές ή τα περιεχόμενα στο αρχικό δίκτυο. Δεδομένων αυτών των διανυσματικών χαρακτηριστικών, μοντέλα μηχανικής μάθησης όπως μοντέλα ταξινόμησης, μοντέλα συστάσεων και μοντέλα εντοπισμού εκκεντρικών αξιολογήσεων μπορούν να χρησιμοποιηθούν απευθείας για στόχευση στις εφαρμογές.

Μαζί με τη σημαντική βελτίωση της απόδοσης που επιτεύχθηκε με τη βαθιά μάθηση στην αναγνώριση εικόνων, την εξόρυξη κειμένου και την επεξεργασία φυσικής γλώσσας, η ανάπτυξη μεθόδων αναπαράστασης δικτύου χρησιμοποιώντας μοντέλα νευρωνικών δικτύων έχει λάβει αυξημένη προσοχή τα τελευταία χρόνια.

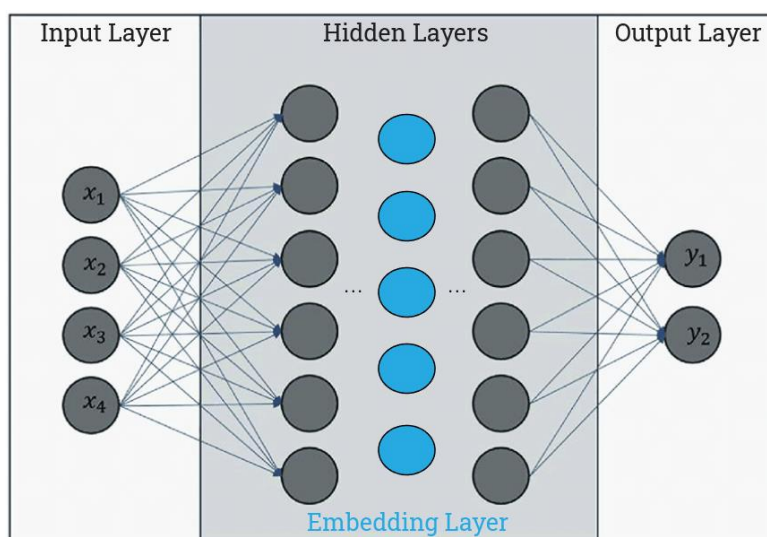
Τα νευρωνικά δίκτυα και η ανάλυση κοινωνικών δικτύων είναι δύο διακριτά πεδία, αλλά μπορούν να σχετίζονται με διάφορους τρόπους, ειδικά όταν πρόκειται για την ανάλυση και την κατανόηση κοινωνικών δικτύων.

Εδώ παρατίθενται μερικοί τρόποι με τους οποίους τα νευρωνικά δίκτυα μπορούν να «συναντηθούν» και να «συμβαδίσουν» με την ανάλυση κοινωνικών δικτύων:

1. Node Embeddings: Τα νευρωνικά δίκτυα μπορούν να χρησιμοποιηθούν για τη δημιουργία ενσωματώσεων (embeddings) για τους κόμβους (χρήστες) σε ένα κοινωνικό δίκτυο. Αυτές οι ενσωματώσεις αναπαριστούν τους κόμβους ως πυκνούς σε διανύσματα εντός ενός χώρου χαμηλής διάστασης, αποτυπώνοντας τη δομική και περιβαλλοντική πληροφορία του δικτύου. Αυτές οι ενσωματώσεις μπορούν να χρησιμοποιηθούν για διάφορες εργασίες στην ανάλυση κοινωνικών δικτύων, όπως

η κατηγοριοποίηση κόμβων, η πρόβλεψη συνδέσμων και η ανίχνευση κοινοτήτων.[20]

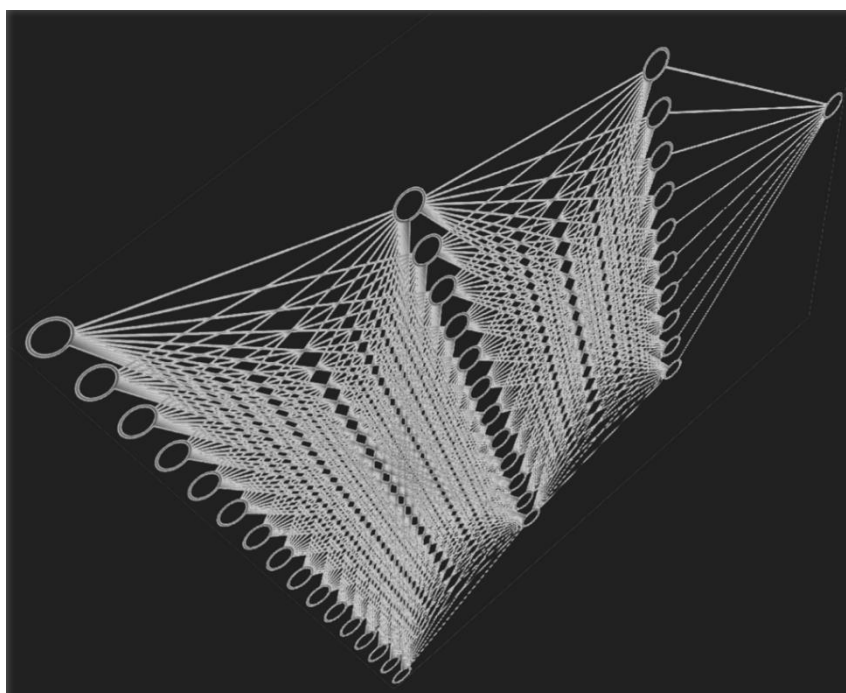
Το ενσωματωμένο layer είναι ένας τύπος κρυφού layer σε ένα νευρωνικό δίκτυο που απεικονίζει πληροφορίες εισόδου από έναν υψηλής διάστασης χώρο σε έναν χαμηλότερης διάστασης χώρο, επιτρέποντας έτσι στο δίκτυο να μάθει περισσότερα για τη σχέση μεταξύ των εισόδων και να επεξεργάζεται τα δεδομένα πιο αποτελεσματικά.



Εικόνα 10:Μορφή ενσωματωμένου layer

2. Graph Neural Networks (GNNs) : Τα GNNs είναι ένα είδος νευρωνικού δικτύου που σχεδιάστηκε για να λειτουργεί απευθείας με δεδομένα που έχουν δομή γράφου, όπως τα κοινωνικά δίκτυα. Τα GNNs μπορούν να μάθουν να συγκεντρώνουν πληροφορίες από τους γείτονες ενός κόμβου και να ενημερώνουν τη δική του αναπαράσταση ανάλογα. Αυτό τους επιτρέπει να ανιχνεύουν πολύπλοκες σχέσεις και πρότυπα μέσα σε κοινωνικά δίκτυα, κάνοντάς τα χρήσιμα για εργασίες όπως η κατηγοριοποίηση κόμβων και η σύσταση.

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων



Εικόνα 11: εικονογράφιση Graph Neural Network

3. Αναγνώριση κοινοτήτων: Τα νευρωνικά δίκτυα μπορούν να βοηθήσουν στην αναγνώριση κοινοτήτων εντός ενός κοινωνικού δικτύου. Με το να εκπαιδεύονται μοντέλα για τον ομαδοποιητικό χαρακτηρισμό κόμβων με παρόμοια μοτίβα συνδεσιμότητας, μπορούν να ανακαλυφθούν κοινότητες ή ομάδες χρηστών με παρόμοια ενδιαφέροντα ή συμπεριφορές. Οι μέθοδοι ομαδοποίησης είναι βασισμένες σε νευρωνικά δίκτυα μπορούν να είναι πιο ευέλικτες και προσαρμοστικές από τους παραδοσιακούς αλγόριθμους ανίχνευσης κοινοτήτων.

4. Ανάλυση συναισθημάτων: Τα νευρωνικά δίκτυα, ιδιαίτερα τα αναδρομικά νευρωνικά δίκτυα (RNNs) και τα συνελκτικά νευρωνικά δίκτυα (CNNs), χρησιμοποιούνται συχνά για την ανάλυση συναισθημάτων σε κείμενα που εντοπίζονται στα μέσα κοινωνικής δικτύωσης. Η ανάλυση των συναισθημάτων μπορεί να βοηθήσει στην κατανόηση του συναισθηματικού τόνου των συνομιλιών στα κοινωνικά

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

δίκτυα, πράγμα που είναι πολύτιμο για την παρακολούθηση της εικόνας μιας εταιρείας, την ανάλυση των αντιδράσεων των πελατών και άλλα.

5. Ανίχνευση ανωμαλιών: Τα νευρωνικά δίκτυα μπορούν να χρησιμοποιηθούν για την ανίχνευση ανωμαλιών σε κοινωνικά δίκτυα. Με το να εκπαιδεύονται μοντέλα στην κανονική συμπεριφορά του δικτύου, οποιεσδήποτε αποκλίσεις από αυτό το μάθημα μπορούν να εντοπιστούν ως ανωμαλίες. Αυτό μπορεί να βοηθήσει στον εντοπισμό ασυνήθιστων ή ενδεχομένως κακόβουλων δραστηριοτήτων σε κοινωνικά δίκτυα.

6. Προτεινόμενα συστήματα: Τα νευρωνικά δίκτυα χρησιμοποιούνται συχνά για την κατασκευή συστημάτων συστάσεων για κοινωνικά δίκτυα. Αυτά τα συστήματα χρησιμοποιούν τις προτιμήσεις και τη συμπεριφορά των χρηστών για να προτείνουν σχετικό περιεχόμενο, φίλους ή προϊόντα. Τα μοντέλα βαθιάς μάθησης μπορούν να αναγνωρίσουν περίπλοκες προτιμήσεις χρηστών και να βελτιώσουν την ακρίβεια των προτάσεων.

7. Link Prediction: Τα νευρωνικά δίκτυα μπορούν να προβλέπουν την πιθανότητα μελλοντικών συνδέσεων (συνδέσμων) μεταξύ χρηστών σε ένα κοινωνικό δίκτυο. Αυτό είναι ιδιαίτερα χρήσιμο για την πρόταση δυνητικών φίλων ή συνδέσεων σε χρήστες. [21]

Συνοψίζοντας, τα νευρωνικά δίκτυα μπορούν να έχουν σημαντικό ρόλο στη βελτίωση της ανάλυσης των κοινωνικών δικτύων με τον εντοπισμό νοήματος στις αναπαραστάσεις, την ανίχνευση προτύπων και ανωμαλιών και τη βελτίωση της συνολικής κατανόησης των πολύπλοκων κοινωνικών αλληλεπιδράσεων και συμπεριφορών μέσα στο δίκτυο. Αυτές οι τεχνικές είναι όλο και πιο αξιόλογες στην εποχή των μεγάλων δεδομένων και των κοινωνικών μέσων ενημέρωσης.

## **2. Πρόβλεψη Χρονοσειρών με χρήση Ανατροφοδοτούμενων Νευρωνικών Δικτύων**

---

### **2.1 Τι είναι οι Χρονοσειρές;**

Οι χρονοσειρές ορίζονται ως μια ακολουθία ποσοτικών παρατηρήσεων οι οποίες έχουν ταξινομηθεί σε μια συγκεκριμένη χρονολογική σειρά, και αφορούν συγκεκριμένα χρονικά διαστήματα. Επί της ουσίας τα δεδομένα των χρονοσειρών είναι δεδομένα καταχωρημένα πάνω σε τακτικά διαστήματα ή χρονικές περιόδους. Κάθε μη μεταβλητή τιμή η οποία είναι συνάρτηση του χρόνου μπορεί να αποτελέσει μέρος μιας χρονοσειράς. Συχνά παρουσιάζουν τάσεις, εποχικά πρότυπα και κυκλικά πρότυπα, και αποτελούν σημαντικό εργαλείο για την ανάλυση και την πρόβλεψη των μελλοντικών εξελίξεων.

Η παραπάνω ιδέα μας ακούγεται αρκετά γνώριμη καθώς αποτελεί μέρος της καθημερινότητάς μας χωρίς να το γνωρίζουμε. Οποτεδήποτε έχει κανείς μας χρησιμοποιήσει το smartwatch του προκειμένου να δει πόσα βήματα έχει κάνει εντός της εβδομάδας ή να σημειώσει τα καθημερινά έξοδα του για παράδειγμα ένα μήνα, έχει επισήμως καταγράψει δεδομένα χρονοσειρών.[22]

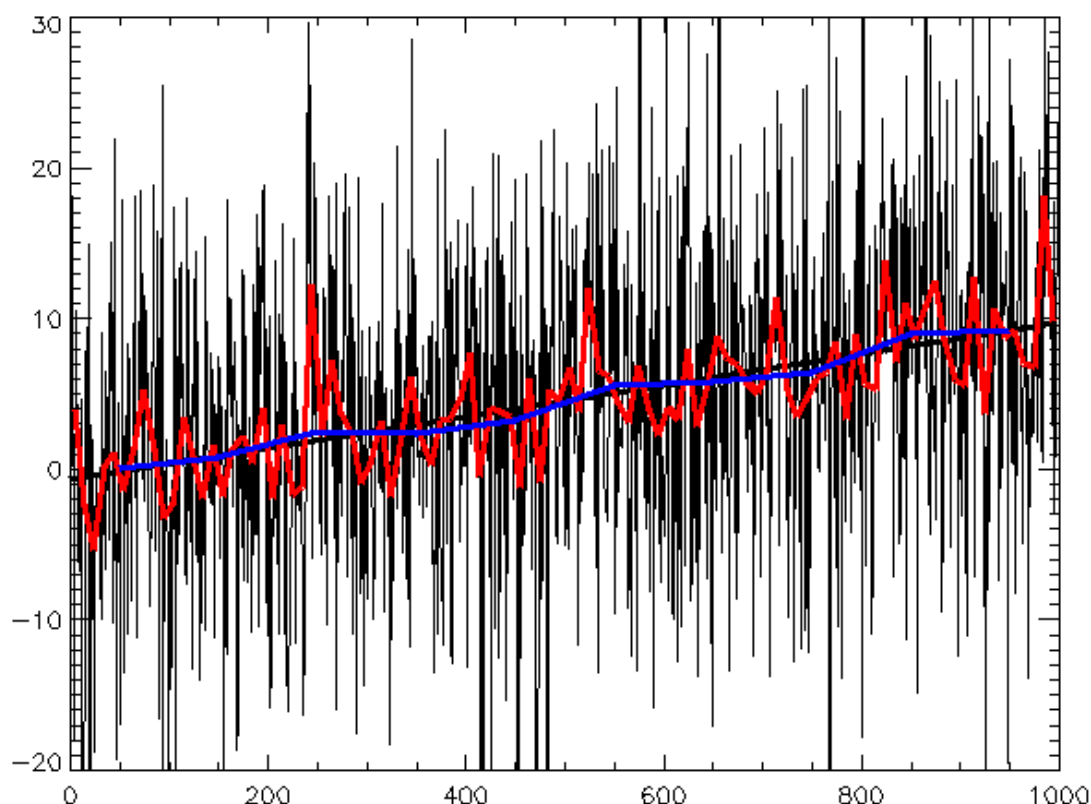
Τα δεδομένα των χρονοσειρών διαφέρουν όσον αφορά :

- a) Την χρονική τους εξάρτηση. Οι ακολουθίες δεν είναι ανεξάρτητες μεταξύ τους με την τρέχουσα παρατήρηση να εξαρτάται από τις προηγούμενες. Π.χ. Η θερμοκρασία σήμερα δεν μπορεί να προγνωστεί ανεξάρτητα αλλά εξαρτάται από τις χθεσινές καιρικές συνθήκες.
- b) Τις τάσεις, οι οποίες μπορούν να αυξήσουν ή να μειώσουν τα δεδομένα στο μήκος της χρονικής αυτής συνιστώσας κυρίως μέσω των εποχικών τάσεων. Π.χ. Η πώληση των αδιάβροχων

μπουφάν αυξάνεται ή μειώνεται χρόνο με το χρόνο εξαρτώμενη από το πόσο θα βρέξει την συγκεκριμένη χρονιά.

Η ανάλυση των χρονοσειρών πραγματοποιείται σε τρία στάδια, τα οποία είναι τα εξής:

- a. Η περιγραφή, η οποία επιτυγχάνεται με τη χρήση διαφόρων γραφημάτων για τα δεδομένα.
- b. Η επεξήγηση, γίνεται με τη βοήθεια ορισμένων μοντέλων για την καλύτερη κατανόηση των μηχανισμών που δημιουργήσαν τις χρονοσειρές.
- c. Η πρόβλεψη, γίνεται χρησιμοποιώντας ένα μοντέλο έτσι ώστε να προβλεφθούν οι μελλοντικές τιμές της χρονοσειράς.



Εικόνα 12: διάγραμμα χρονοσειράς



## 2.2 Ανάλυση χρονοσειρών

Η ανάλυση χρονοσειρών γίνεται μέσω της επεξεργασίας των δεδομένων αλλά και μέσω της χρήσης μαθηματικών εργαλείων και αναζητά εντός των δεδομένων αυτών, όντας συστατικά των χρονοσειρών, να μάθουν όχι μόνο τι συνέβη αλλά πότε και γιατί συνέβη όπως και το τι είναι πιθανό να συμβεί στο μέλλον. Η ανάλυση των δεδομένων με χρονοσφραγίδα συμβάλλει στην παραγωγή σημαντικών στατιστικών και συμβάλλει στη λήψη αποφάσεων. Για παράδειγμα, χρησιμοποιείται για την κατανόηση και την ερμηνεία των τάσεων και των προτύπων. Επιπλέον, χρησιμοποιείται στην πρόβλεψη, την κατηγοριοποίηση, την τμηματοποίηση, την περιγραφική και την παρέμβαση σε αναλύσεις κ.λπ.

Το στοιχείο του χρόνου σε δεδομένα χρονοσειρών σημαίνει πως τα δεδομένα είναι ταξινομημένα χρονικά. Σε αυτή την κατηγορία δεδομένων, κάθε είσοδος προηγείται και ακολουθεί από μια άλλη διαθέτοντας μάλιστα ηλεκτρονική χρονοσήμανση, αποδεικνύοντας έτσι ότι βρίσκονταν στο χρονικό σημείο εκείνο. Η ανάλυση χρονοσειρών αποτελείται από μεθόδους που προσπαθούν να κατανοήσουν τέτοιες χρονοσειρές για να κάνουν προβλέψεις. [23]

Οι χρονοσειρές μπορούν να αναλυθούν σε τέσσερα στοιχεία, καθένα από τα οποία εκφράζει ένα συγκεκριμένο κομμάτι της κίνησης των τιμών της χρονοσειράς. Η εφαρμογή της είναι συνήθως ορατή σε μη-σταθερά δεδομένα για να παρατηρήσουμε πώς αλλάζουν κάποια πράγματα με την πάροδο του χρόνου. Παραδείγματα περιλαμβάνουν την κρίσιμη εφαρμογή της στο χρηματιστήριο, τα ιστορικά δεδομένα, τις πωλήσεις, το απόθεμα και την ανάλυση του καιρού. Επομένως, η εμπειρία σε αυτόν τον τομέα είναι σημαντική για επαγγελματίες από τον χώρο του λιανικού εμπορίου, της χρηματοοικονομίας, των πωλήσεων, της πρόβλεψης καιρού κ.λπ.

Η ανάλυση χρονοσειρών χρησιμοποιείται για μη-σταθερά δεδομένα - πράγματα που διακυμαίνονται διαρκώς με τον χρόνο ή επηρεάζονται από αυτόν. Βιομηχανίες όπως οι οικονομικές, η χρηματοοικονομία και το λιανεμπόριο χρησιμοποιούν συχνά την ανάλυση χρονοσειρών, καθώς το νόμισμα και οι πωλήσεις αλλάζουν συνεχώς. Η ανάλυση του χρηματιστηρίου αποτελεί εξαιρετικό παράδειγμα της ανάλυσης χρονοσειρών σε δράση, ειδικά με τους αυτοματοποιημένους αλγορίθμους συναλλαγών. Επίσης, η ανάλυση χρονοσειρών είναι ιδανική για την πρόβλεψη αλλαγών στον καιρό, βοηθώντας τους μετεωρολόγους να προβλέπουν όλα, από τον καιρικό προγραμματισμό της αυριανής αναφοράς μέχρι τα μελλοντικά έτη της κλιματικής αλλαγής. [24]

Παραδείγματα ανάλυσης χρονοσειρών σε δράση περιλαμβάνουν:

- Δεδομένα καιρού
- Μετρήσεις βροχής
- Αναγνώσεις θερμοκρασίας
- Παρακολούθηση παλμών (EKG)
- Παρακολούθηση εγκεφάλου (EEG)
- Τριμηνιαίες πωλήσεις
- Τιμές μετοχών
- Αυτοματοποιημένες συναλλαγές με μετοχές
- Προβλέψεις βιομηχανίας
- Επιτόκια

Εφόσον η ανάλυση χρονοσειρών περιλαμβάνει πολλές κατηγορίες ή παραλλαγές δεδομένων, οι αναλυτές πρέπει μερικές φορές να δημιουργήσουν πολύπλοκα μοντέλα. Ωστόσο, οι αναλυτές δεν μπορούν να λάβουν υπόψη όλες τις διακυμάνσεις, και δεν μπορούν να γενικεύσουν ένα συγκεκριμένο μοντέλο σε κάθε δείγμα. Μοντέλα που είναι πολύπλοκα ή προσπαθούν να κάνουν πολλά πράγματα μπορεί να οδηγήσουν σε έλλειψη προσαρμογής. Η έλλειψη προσαρμογής οδηγεί σε ανεπαρκή διάκριση μεταξύ τυχαίου σφάλματος και πραγματικών σχέσεων αφήνοντας την ανάλυση παραμορφωμένη και τις προβλέψεις εσφαλμένες.

Ο στόχος την ανάλυσης χρονοσειρών:

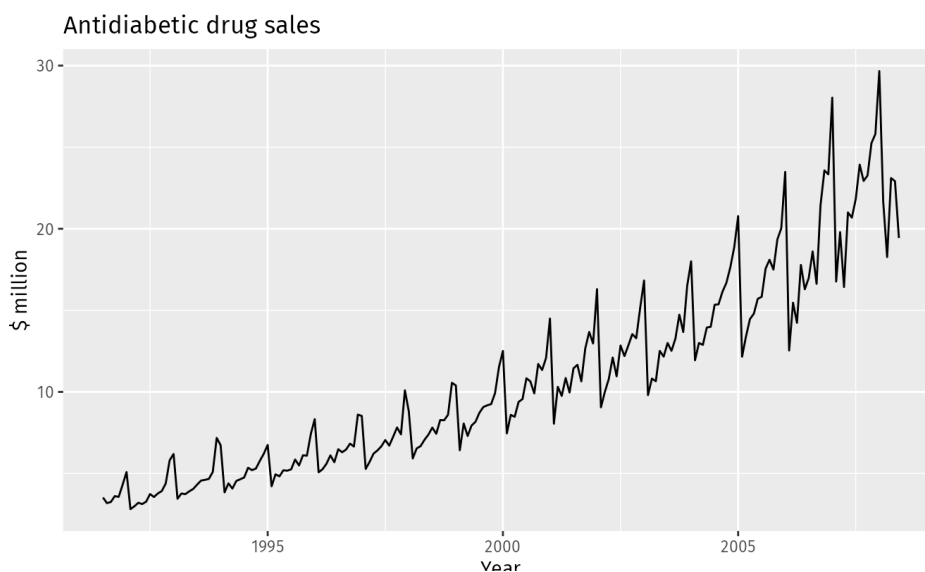
- Να αναγνωρίσει τις υποβόσκουσες δυνάμεις που οδηγούν στην δημιουργία μιας συγκεκριμένης τάσης εντός του μοτίβου των χρονοσειρών.
- Να προβλέψει τις μελλοντικές αξίες των χρονοσειρών μεταβλητών.
- Να εντοπίσει στα δεδομένα που αποτελούν τις χρονοσειρές τα συστηματικά μοτίβα που υπάρχουν και τον τυχαίο θόρυβο που ίσως αυτά διαθέτουν. Αν αφαιρέσουμε τυχόν θόρυβο τότε το συστηματικό μοτίβο θα γίνει πλήρως εμφανές, κάτι που εκτελείται με την αποσύνθεση των χρονοσειρών.

### **2.3 Μοτίβα – Χαρακτηριστικά των χρονοσειρών**

Αν επανέλθουμε πάλι στο παράδειγμα της εφαρμογής στο κινητό που καταγράφει δεδομένα και ελέγχει τις ζωτικές μας πληροφορίες, το πρώτο πράγμα που θα παρατηρήσουμε ανοίγοντας την εφαρμογή είναι τα στατιστικά μας στο διάγραμμα που θα έχει δημιουργηθεί και αν αυτά έχουν ανοδική ή καθοδική τάση. Αυτό μας υποδεικνύει τότε τα στατιστικά είναι σε γενικές γραμμές βελτιωμένα ή όχι. Αυτό αποτελεί και το πρώτο από τα τρία μοτίβα των χρονοσειρών:

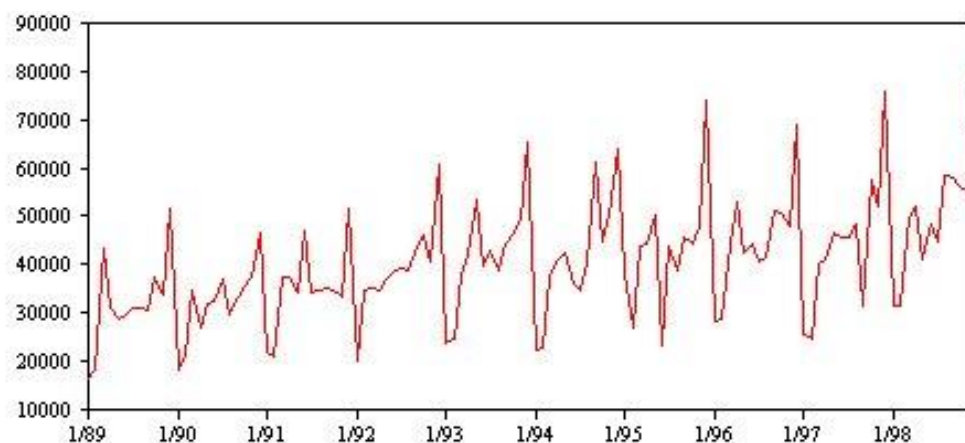
Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

- I. Τάση: Ένα τέτοιο μοτίβο εμφανίζεται όταν υπάρχει μια μακροπρόθεσμη αύξηση ή μείωση στα δεδομένα, η οποία δεν χρειάζεται να είναι γραμμική. Ορισμένες φορές αναφερόμαστε στην τάση ως την «μεταβλητή κατεύθυνση» ιδιαιτέρως όταν η πορεία που χαράζεται πηγαιίνει από μια ανοδική τάση σε μία καθοδική.



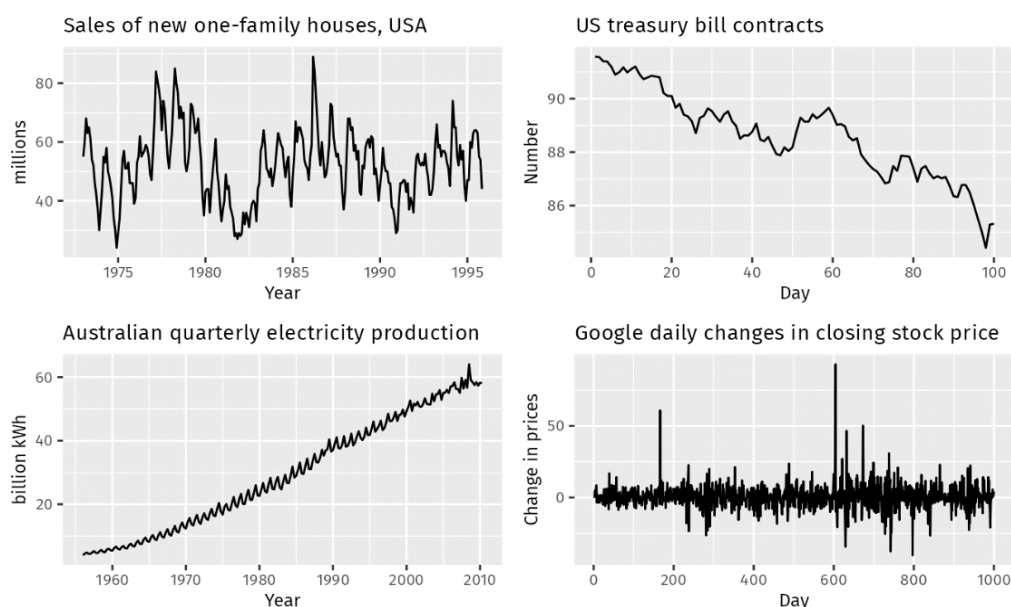
Εικόνα 13: Μηνιαίες πωλήσεις αντιδιαβητικών φαρμάκων στην Αυστραλία

- II. Εποχικότητα: Ένα εποχικό μοτίβο προκύπτει όταν μια χρονοσειρά επηρεάζεται από εποχικούς παράγοντες όπως η εποχή ή μέρα της εβδομάδας. Η εποχικότητα είναι σταθερή και γνωστής συχνότητας. Οι μηνιαίες πωλήσεις των αντιδιαβητικών φαρμάκων πάνω σε μια εποχικότητα η οποία προκαλείται από την αλλαγή του κόστους των φαρμάκων στο τέλος του χρόνου.



Εικόνα 14:: Εποχικό μοτίβο στις πωλήσεις φαρμάκων

III. Κυκλικότητα : Προκύπτει όταν ο όγκος των δεδομένων αυξάνεται και μειώνεται και δεν είναι μιας σταθερής συχνότητας. Αυτές οι αυξομειώσεις παρατηρούνται κυρίως σε οικονομικές συνθήκες και συνήθως σχετίζονται με την επιχειρηματική κυκλικότητα. Η διάρκεια αυτών των αυξομειώσεων είναι το λιγότερο δύο χρόνια. Αυτές



Εικόνα 15: Τέσσερα παραδείγματα χρονοσειρών που οπτικοποιούν διαφορετικά μοτίβα των τάσεων

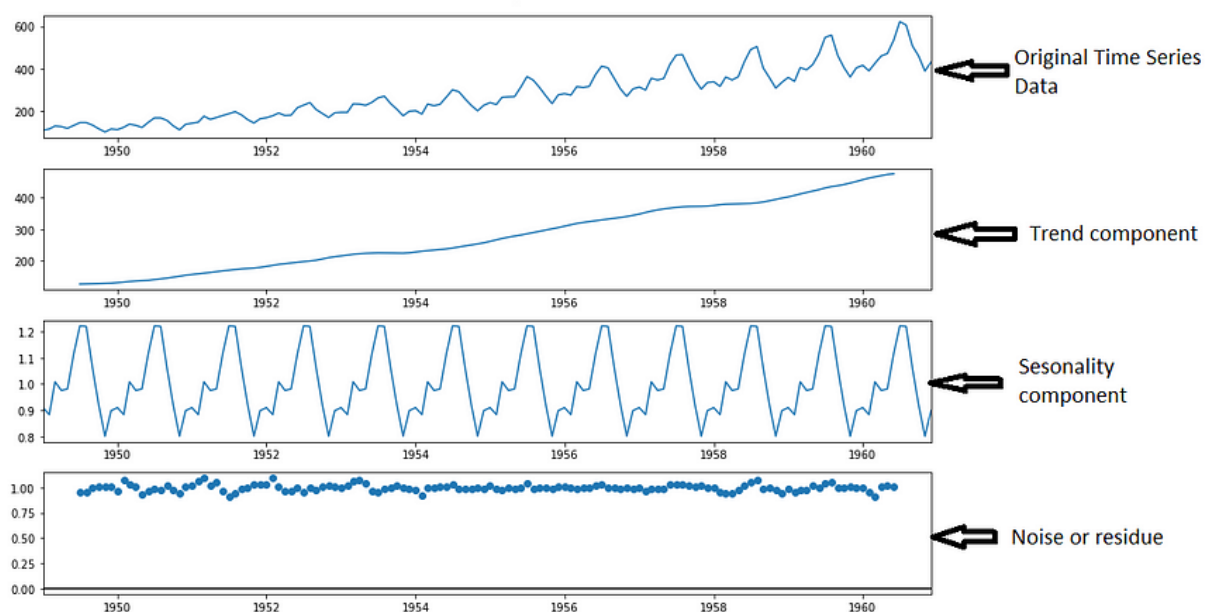
Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

Πολλοί άνθρωποι συγχέουν την κυκλικότητα με την εποχικότητα ενώ στην ουσία είναι αρκετά διαφορετικές. Αν οι αυξομειώσεις δεν ανήκουν σε σταθερή συχνότητα τότε έχουμε το μοτίβο της κυκλικότητας. Αν η συχνότητα είναι αμετάβλητη και συσχετισμένη με κάποια ημερολογιακή μορφή τότε το μοτίβο είναι εποχιακό.

Σε γενικές γραμμές το μέσο μήκος μιας κυκλικότητας δεν είναι μεγαλύτερο από το μήκος ενός εποχικού μοτίβου καθώς το μέγεθος των επαναλήψεων τείνει να είναι πιο ασταθές από τις επαναλήψεις των εποχιακών μοτίβων. Σε πολλές μάλιστα περιπτώσεις χρονοσειρών περιλαμβάνονται και οι τρεις τάσεις. Όταν επιλέγει κανείς μια μέθοδο πρόβλεψης θα πρέπει σε πρώτη φάση να αναγνωρίσει τα μοτίβα που «κρύβουν» μέσα τους τα δεδομένα και έπειτα να επιλέξει την κατάλληλη μέθοδο που μπορεί να αποτυπώσει τα μοτίβα καταλλήλως.

Οι μηνιαίες πωλήσεις κατοικιών (πάνω αριστερά) δείχνουν έντονη εποχικότητα κατά τη διάρκεια κάθε έτους, καθώς και κάποια έντονη κυκλική συμπεριφορά με περίοδο περίπου 6-10 ετών. Δεν φαίνεται να υπάρχει κάποια εμφανής τάση στα δεδομένα κατά τη διάρκεια αυτής της περιόδου. Οι συμβάσεις του αμερικανικού υπουργείου Οικονομικών (πάνω δεξιά) δείχνουν αποτελέσματα από την αγορά του Σικάγο για 100 συνεχόμενες εμπορικές ημέρες το 1981. Εδώ δεν υπάρχει εποχικότητα, αλλά υπάρχει μια προφανής φθίνουσα τάση. Ίσως, αν είχαμε μια πολύ μεγαλύτερη σειρά δεδομένων, θα διακρίναμε ότι αυτή η φθίνουσα τάση αποτελεί πραγματικά μέρος μιας μακράς κυκλικής κίνησης, αλλά όταν την βλέπουμε για μόνο 100 ημέρες, φαίνεται να είναι μια τάση. Η τριμηνιαία παραγωγή ηλεκτρικής ενέργειας στην Αυστραλία (κάτω αριστερά) δείχνει μια έντονη αυξητική τάση, με έντονη εποχικότητα. Δεν υπάρχουν ενδείξεις κάποιας κυκλικής συμπεριφοράς εδώ. Η ημερήσια μεταβολή στην τιμή κλεισίματος της Google (κάτω δεξιά) δεν έχει τάση, εποχικότητα ή κυκλική συμπεριφορά. Υπάρχουν τυχαίες διακυμάνσεις που δεν φαίνεται να είναι πολύ προβλέψιμες, και δεν υπάρχουν έντονα πρότυπα που θα βοηθούσαν στην ανάπτυξη ενός μοντέλου πρόβλεψης. [25]

Εικόνα 16: Διαγράμματα περιπτώσεων με χαρακτηριστικά χρονοσειρών



## 2.3 Πρόβλεψη & Χρονοσειρές

Η πρόβλεψη των χρονοσειρών αποτελεί μία από τις πιο εφαρμοζόμενες τεχνικές στην επιστήμη των δεδομένων σε επιχειρήσεις, τη χρηματοοικονομία, τη διαχείριση της αλυσίδας εφοδιασμού, την παραγωγή και τον σχεδιασμό αποθεμάτων. Πολλά προβλήματα πρόβλεψης περιλαμβάνουν ένα χρονικό στοιχείο και απαιτούν την εκτίμηση δεδομένων χρονοσειράς, ή απλά πρόβλεψη χρονοσειράς.

Η πρόβλεψη των χρονοσειρών σημαίνει να προβλέπετε ή να εκτιμάτε τη μελλοντική τιμή κατά τη διάρκεια ενός χρονικού διαστήματος. Περιλαμβάνει τη δημιουργία μοντέλων με βάση προηγούμενα δεδομένα και την εφαρμογή τους για την παρακολούθηση παρατηρήσεων και την καθοδήγηση μελλοντικών στρατηγικών αποφάσεων. Η μελλοντική πρόβλεψη βασίζεται στο γεγονός ότι η μελλοντική εξέλιξη βασίζεται σε ήδη υπάρχοντα συμβάντα. Οι

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

χρονοσειρές προσθέτουν μια εξάρτηση από τη χρονική σειρά μεταξύ των παρατηρήσεων. Αυτή η εξάρτηση αποτελεί τόσο περιορισμό όσο και δομή που παρέχει πηγή πρόσθετων πληροφοριών. Πριν συζητήσουμε τις μεθόδους πρόβλεψης χρονοσειράς, ας καθορίσουμε πιο αναλυτικά την πρόβλεψη των χρονοσειρών. [26]

Στόχος της πρόβλεψης χρονοσειρών είναι η πρόβλεψη μελλοντικών τιμών ή γεγονότων με βάση τα ιστορικά δεδομένα ταξινομημένα χρονικά. Η πρόβλεψη χρονοσειρών αποτελεί επίσης σημαντικό πεδίο της μηχανικής μάθησης και μπορεί να αναπαρασταθεί ως πρόβλημα επιβλεπόμενης μάθησης.

Οι χρονοσειρές αναπαριστούν δεδομένα που οργανώνονται χρονικά, δηλαδή σε συγκεκριμένες χρονικές στιγμές. Συχνά παρουσιάζουν τάσεις, εποχικά πρότυπα και κυκλικά πρότυπα, και αποτελούν σημαντικό εργαλείο για την ανάλυση και την πρόβλεψη των μελλοντικών εξελίξεων.

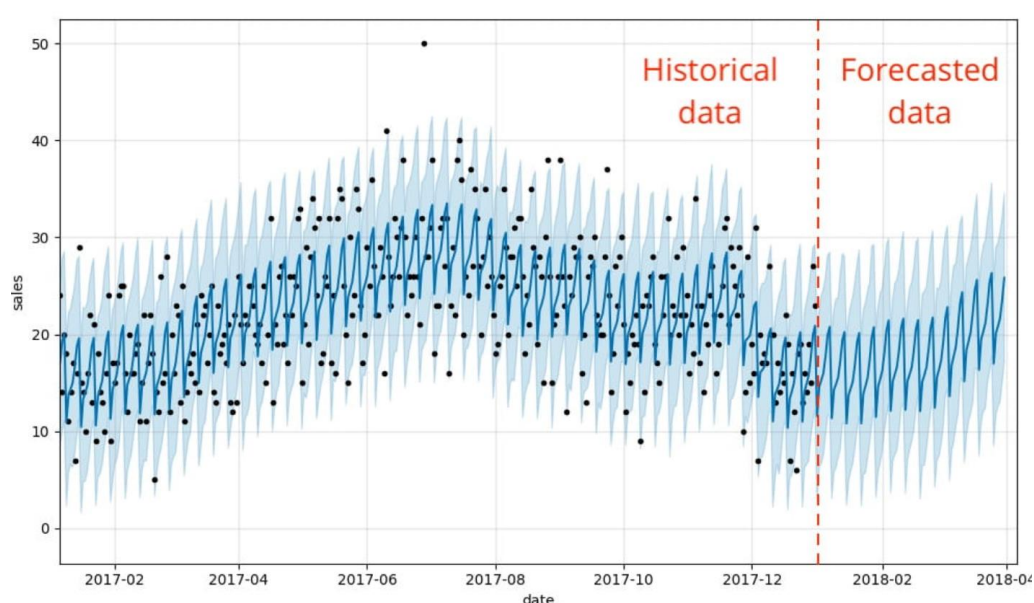
Η πρόβλεψη των χρονοσειρών ξεκινά με μια ιστορική χρονοσειρά. Οι αναλυτές εξετάζουν τα ιστορικά δεδομένα και ελέγχουν για μοτίβα χρονικής ανάλυσης, όπως τάσεις, εποχικά μοτίβα, κυκλικά μοτίβα και τακτικότητα. Πολλοί τομείς εντός των οργανισμών, μάρκετινγκ, της χρηματοοικονομίας και των πωλήσεων, χρησιμοποιούν κάποια μορφή πρόβλεψης χρονοσειράς για να αξιολογήσουν πιθανές τεχνικές δαπάνες και τη ζήτηση των καταναλωτών. Τα μοντέλα για τα δεδομένα χρονοσειράς μπορούν να έχουν πολλές μορφές και να αντιπροσωπεύουν διάφορες στοχαστικές διαδικασίες.

Μεθοδολογίες μηχανικής μάθησης αποτελούν η Παλινδρόμηση, τα νευρωνικά δίκτυα, οι Μηχανές Υποστήριξης Διανυσμάτων, οι Διαστολές Δέντρων και μπορούν να εφαρμοστούν για την πραγματοποίηση προβλέψεων σε χρονοσειρές. Αυτή η διαδικασία βασίζεται στην κατασκευή μοντέλων χρησιμοποιώντας ιστορικά δεδομένα και στη χρήση αυτών των μοντέλων για την πρόβλεψη μελλοντικών παρατηρήσεων. [27]



Τα μοντέλα χρονοσειρών χρησιμοποιούνται για την πρόβλεψη γεγονότων βασισμένα σε επαληθευμένα ιστορικά δεδομένα. Κοινί τύποι περιλαμβάνουν τα ARIMA, τα μοντέλα βάσης λείανσης και το κινητό μέσο. Δεν διαθέτουν όλα τα μοντέλα την δυνατότητα να παράγουν τα ίδια αποτελέσματα για τα ίδια δεδομένα, επομένως είναι κρίσιμο να καθορίσουμε ποιο λειτουργεί καλύτερα βάσει της συγκεκριμένης χρονοσειράς.

Εικόνα 17: Διάγραμμα χρονοσειράς με πρόβλεψη



### Στόχοι Πρόβλεψης

Όταν επιδιώκεται η προσπάθεια προβλέψεων, είναι σημαντικό να κατανοήσουμε τον στόχο μας. Για να οριοθετήσουμε τις λεπτομέρειες του προβλήματος που θέλουμε να προβλέψουμε συνίστανται να λαμβάνουμε υπόψιν:

- Τον όγκο των διαθέσιμων δεδομένων, όσο περισσότερα δεδομένα τόσο προσφέρουν μεγαλύτερες δυνατότητες για εξερευνητική ανάλυση δεδομένων, δοκιμή και προσαρμογή μοντέλων, καθώς και πιστότητα μοντέλου.

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

- Τον απαιτούμενο χρονικό ορίζοντα προβλέψεων, συνήθως οι σύντομοι χρονικοί ορίζοντες είναι πιο εύκολοι να προβλεφθούν με μεγαλύτερη εμπιστοσύνη από τους μακροπρόθεσμους.
- Συχνότητα ενημέρωσης προβλέψεων, οι προβλέψεις μπορεί να χρειάζεται να ενημερώνονται συχνά με την πάροδο του χρόνου ή να χρειάζεται να γίνονται μία φορά και να παραμένουν στατικές. Η ενημέρωση των προβλέψεων καθώς νέες πληροφορίες γίνονται διαθέσιμες συχνά οδηγεί σε πιο ακριβείς προβλέψεις.
- Χρονική συχνότητα πρόβλεψης, οι προβλέψεις μπορούν να γίνουν σε χαμηλότερες ή υψηλότερες συχνότητες, επιτρέποντας την εκμετάλλευση της δειγματοληψίας και υπερδειγματοληψίας δεδομένων κάτι που μπορεί να προσφέρει οφέλη κατά την διάρκεια της μοντελοποίησης. [28]

## **2.4 Ανατροφοδοτούμενα Νευρωνικά Δίκτυα (Recurrent neural network (RNN))**

Τα RNN είναι βαθιά νευρωνικά δίκτυα που έχουν τη δυνατότητα να αποθηκεύουν πληροφορίες από προηγούμενους υπολογισμούς και τις προωθούν προς τα εμπρός ώστε να εργαστούν πάνω σε αυτά τα δεδομένα με ακολουθιακό τρόπο. Αυτός ο τύπος νευρωνικού δικτύου χρησιμοποιεί το αποτέλεσμα του προηγούμενου χρονικού βήματος ως είσοδο για το τρέχον χρονικό βήμα μαζί με τη νέα είσοδο.

Αυτό δημιουργεί μια σχέση μεταξύ των στοιχείων στην ακολουθία, επιτρέποντας στα RNN να μάθουν τις εξαρτήσεις μεταξύ τους. Έτσι τα RNN καθίστανται κατάλληλα για εργασίες όπως η αναγνώριση ομιλίας όπου η κατανόηση του πλαισίου είναι κρίσιμη. Τα RNN μπορούν επίσης να χρησιμοποιηθούν για την πρόβλεψη χρονοσειρών, όπως οι κινήσεις στην αγορά μετοχών ή τα μοτίβα καιρού. Τα τελευταία χρόνια, τα RNN έχουν επιτύχει σημαντικές επιτυχίες σε πολλές προκλητικές εργασίες.

Στο δίκτυο αυτό παράγεται ένα αποτέλεσμα, η συνάρτηση απώλειας ( $C$ ) που υποδεικνύει πόσο καλά προέβλεψε το αποτέλεσμα το RNN δίκτυο. Η ιδέα είναι πως το δίκτυο αυτό εκτελεί αντίστροφη διάδοση δεδομένων για να ελαχιστοποιήσει την απώλεια. Μέσω αυτής της μεθόδου υπολογίζεται η κλίση της συνάρτησης απώλειας ως προς κάθε βάρη στο δίκτυο. Αυτό δημιουργεί και ένα πρόβλημα καθώς κατά τη διάρκεια της εκπαίδευσης, η κλίση η οποία δείχνει πόσο πρέπει να προσαρμοστούν τα βάρη του δικτύου μπορεί να γίνει ελάχιστα μικρή. Έτσι γεννάται ένα καινούργιο πρόβλημα, το της εξαφάνισης της κλίσης (Vanishing Gradient Problem) το οποίο καθιστά την εκπαίδευση των RNN δύσκολη. [29]

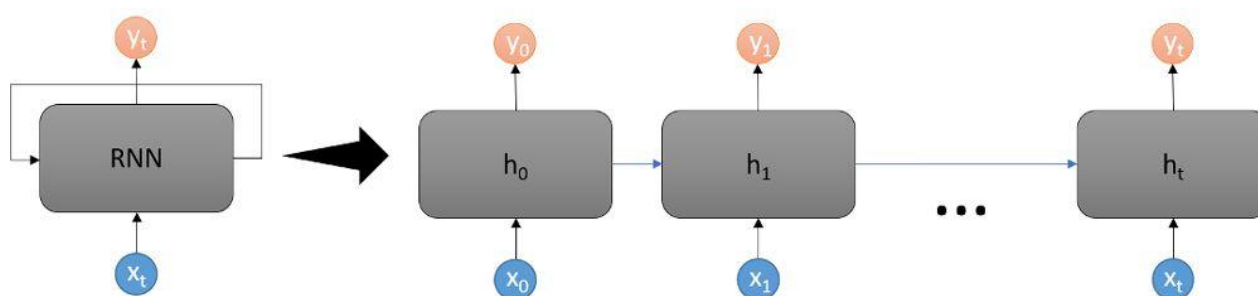
Κύριος δείκτης του προβλήματος αυτών των δικτύων είναι η σιγμοειδής συνάρτηση, η οποία είναι μία από τις πιο δημοφιλείς συναρτήσεις ενεργοποίησης που χρησιμοποιούνται για την ανάπτυξη βαθιών νευρωνικών δικτύων. Η χρήση της σιγμοειδούς συνάρτησης περιόρισε την εκπαίδευση βαθιών νευρωνικών δικτύων επειδή προκάλεσε το πρόβλημα της εξαφάνισης της κλίσης, που αναφέραμε. Αυτό οδηγεί το

νευρωνικό δίκτυο να μαθαίνει με πιο αργό ρυθμό ή σε ορισμένες περιπτώσεις, να μην μάθει καθόλου.

Σε επίπεδο δικτύων η χρήση της σιγμοειδούς συνάρτησης μπορεί να χρησιμοποιηθεί επειδή η μικρή τιμή της κλίσης δεν αποτελεί πρόβλημα. Όταν όμως πρόκειται για βαθιά δίκτυα, το πρόβλημα της εξαφάνισης της κλίσης μπορεί να έχει σημαντική επίδραση στην απόδοση. Τα βάρη του δικτύου παραμένουν αμετάβλητα καθώς η κλίση εξαφανίζεται. Κατά τη διάρκεια της αντίστροφης διάδοσης, ένα νευρωνικό δίκτυο μαθαίνει ενημερώνοντας τα βάρη και τις παραμέτρους του για να μειώσει τη συνάρτηση απώλειας. Σε ένα δίκτυο με εξαφανισμένη κλίση, τα βάρη δεν μπορούν να ενημερωθούν, και έτσι το δίκτυο δεν μπορεί να μάθει. Ως αποτέλεσμα, η απόδοση του δικτύου να μειωθεί.

Ένας τρόπος για να αντιμετωπιστεί το πρόβλημα της εξαφάνισης της κλίσης είναι να αντικατασταθεί η συνάρτηση ενεργοποίησης του δικτύου. Αντί για τη σιγμοειδή συνάρτηση, μπορεί να χρησιμοποιηθεί μια συνάρτηση ενεργοποίησης όπως η ReLU (Rectified Linear Unit).

Οι Rectified Linear Units (ReLU) είναι συναρτήσεις ενεργοποίησης που δημιουργούν ένα θετικό γραμμικό αποτέλεσμα όταν εφαρμόζονται σε θετικές τιμές εισόδου. Αν η είσοδος είναι αρνητική, η συνάρτηση θα επιστρέψει μηδέν.



Εικόνα 18: Διαδρομή πληροφορίας εντός RNN νευρωνικού δικτύου

Η παραπάνω εικόνα δείχνει πώς η πληροφορία ενός μοναδικού κελιού ενός RNN κινείται μέσα από το δίκτυο για μια ακολουθία δεδομένων.

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

Τα εισερχόμενα δεδομένα επηρεάζονται από την κρυφή κατάσταση του κελιού για να παράγουν την έξοδο, και η κρυφή κατάσταση περνά στο επόμενο βήμα του χρόνου. [30]

#### **2.4.1 Προετοιμασία των δεδομένων και κατασκευή του μοντέλου RNN**

Η προετοιμασία των δεδομένων είναι κρίσιμη για ακριβείς προβλέψεις χρονοσειρών με RNNs. Η διαχείριση των ελλειπτικών τιμών και των ακραίων τιμών, η κλιμάκωση των δεδομένων και η δημιουργία κατάλληλων ζευγαριών εισόδου-εξόδου είναι απαραίτητα. Η αφαίρεση της εποχικότητας και της τάσης βοηθάει στον εντοπισμό μοτίβων, ενώ η επιλογή του κατάλληλου μήκους ακολουθίας ισορροπεί τις μακροπρόθεσμες και βραχυπρόθεσμες εξαρτήσεις.

Η κατασκευή του μοντέλου RNN περιλαμβάνει μια σειρά σημαντικών βημάτων που συνεισφέρουν συλλογικά στην απόδοση και την ακρίβειά του μοντέλου.

Προτού εφαρμόσουμε οποιονδήποτε μοντέλο πρόβλεψης, είναι αναγκαία η προεπεξεργασία των δεδομένων της χρονοσειράς. Ορισμένες κοινές τεχνικές προεπεξεργασίας περιλαμβάνουν:

- **Χειρισμός Λειψών Δεδομένων:** Ο χειρισμός των λειψών τιμών είναι κρίσιμος καθώς κενά στα δεδομένα μπορεί να επηρεάσουν την απόδοση του μοντέλου. Μπορούν να χρησιμοποιηθούν τεχνικές όπως η ενδιάμεση τιμολόγηση ή η γέμιση μπροστά/πίσω.
- **Κανονικοποίηση Δεδομένων:** Η κανονικοποίηση των δεδομένων εξασφαλίζει ότι όλα τα χαρακτηριστικά βρίσκονται στην ίδια κλίμακα, εμποδίζοντας οποιοδήποτε μεμονωμένο χαρακτηριστικό να κυριαρχεί στη διαδικασία μάθησης του μοντέλου.

- Απομάκρυνση της Τάσης: Η αφαίρεση του τάσης από τα δεδομένα μπορεί να βοηθήσει στην καλύτερη κατανόηση των υποκείμενων προτύπων και στην πρόβλεψη ακριβών.
- Εποχική Προσαρμογή: Για δεδομένα με εποχικότητα, μπορούν να εφαρμοστούν μέθοδοι εποχικής προσαρμογής όπως εποχιακές διαφοροποιήσεις ή εποχιακή αποσύνθεση.
- Εξομάλυνση: Τεχνικές εξομάλυνσης όπως οι κινούμενοι μέσοι όροι μπορούν να χρησιμοποιηθούν για τη μείωση του θορύβου και την ανάδειξη των υποκείμενων προτύπων.
- Διαίρεση σε Σύνολα Εκπαίδευσης-Δοκιμής: Είναι κρίσιμο να διαχωρίσετε τα δεδομένα σε σύνολα εκπαίδευσης και δοκιμής ενώ εξασφαλίζετε τη διατήρηση της χρονολογικής σειράς. Αυτό επιτρέπει στο μοντέλο να μάθει από τα προηγούμενα δεδομένα εισόδου του συνόλου εκπαίδευσης και να αξιολογήσει την απόδοσή του σε μη έχοντα δεδομένα στο μέλλον.

Η κατασκευή και εκπαίδευση ενός αποτελεσματικού μοντέλου Επαναλαμβανόμενων Νευρωνικών Δικτύων (RNNs) για προβλέψεις χρονοσειρών απαιτεί μια προσέγγιση που ισορροπεί την αρχιτεκτονική του μοντέλου και τεχνικές εκπαίδευσης. Αυτή η ενότητα εξερευνά όλα τα απαραίτητα βήματα για την κατασκευή και εκπαίδευση ενός μοντέλου RNN. Η διαδικασία περιλαμβάνει την προετοιμασία των δεδομένων, τον καθορισμό της αρχιτεκτονικής του μοντέλου, την κατασκευή του μοντέλου, τη βελτιστοποίηση των υπερπαραμέτρων και στη συνέχεια την αξιολόγηση της απόδοσης του μοντέλου.[31]

#### **2.4.2 Πρόβλεψη χρονοσειρών με χρήση ανατροφοδοτούμενων νευρωνικών δικτύων.**

Η πρόβλεψη της χρονοσειράς αποτελεί ένα κλάδο της ανάλυσης δεδομένων και της προβλεπτικής μοντελοποίησης που στοχεύει στο να επιτυγχάνονται προβλέψεις σχετικά με μελλοντικές τιμές βασισμένες σε ιστορικά σημεία δεδομένων σε χρονολογική σειρά. Σε μια χρονοσειρά, τα δεδομένα συλλέγονται και καταγράφονται σε τακτά χρονικά διαστήματα (ανά ώρα, ημερήσια, μηνιαία ή ετήσια).

Τα Αναδρομικά Νευρωνικά Δίκτυα (RNNs) είναι μοντέλα βαθιάς μάθησης που μπορούν να χρησιμοποιηθούν για την ανάλυση της χρονοσειράς, με αναδρομικές συνδέσεις που τους επιτρέπουν να διατηρούν πληροφορίες από προηγούμενα χρονικά βήματα.

Δημοφιλείς παραλλαγές περιλαμβάνουν το Long Short-Term Memory (LSTM) και το Gated Recurrent Unit (GRU), τα οποία μπορούν να μάθουν μακροπρόθεσμες εξαρτήσεις. Εμείς στο επόμενο κεφάλαιο που αφορά την υλοποίηση της εφαρμογής μας θα επικεντρωθούμε στα δίκτυα LSTM.

Ένα αναδρομικό νευρωνικό δίκτυο (RNNs) ανήκει σε μια κατηγορία τεχνητών νευρωνικών δικτύων όπου οι συνδέσεις μεταξύ των κόμβων σχηματίζουν έναν απευθείας γράφο κατά μήκος μιας ακολουθίας. Αυτό του επιτρέπει να εκδηλώνει χρονική δυναμική συμπεριφορά για μια χρονική ακολουθία. Σε αντίθεση με τα νευρωνικά δίκτυα, τα RNN μπορούν να χρησιμοποιήσουν την εσωτερική τους κατάσταση (μνήμη) για την επεξεργασία ακολουθιών εισόδων. Αυτό τα καθιστά εφαρμόσιμα για την πρόβλεψη τιμών αλλά και για την αναγνώριση γραφής και φωνής.[32]

Η πρόβλεψη χρονοσειρών χρησιμοποιώντας Αναδρομικά Νευρωνικά Δίκτυα (RNN) είναι μια ισχυρή τεχνική για την πρόβλεψη βασισμένη σε ακολουθίες δεδομένων.

Ας δούμε παρακάτω πώς λειτουργεί:

- **Προετοιμασία των Δεδομένων:**  
Ξεκινάμε με τα ιστορικά δεδομένα χρονοσειράς, τα οποία περιλαμβάνουν μια ακολουθία σημείων δεδομένων ταξινομημένων χρονολογικά. Αυτά τα δεδομένα πρέπει να υποστούν προεπεξεργασία, η οποία μπορεί να περιλαμβάνει τη χειρισμό των τιμών που απουσιάζουν, την κλιμάκωση και τη διαίρεση τους σε σύνολα εκπαίδευσης και δοκιμής.
- **Αρχιτεκτονική του Μοντέλου:**  
Στη συνέχεια, ορίζουμε την αρχιτεκτονική ενός μοντέλου RNN. Τα RNN είναι σχεδιασμένα να λειτουργούν με ακολουθίες δεδομένων, όπου η σειρά των σημείων δεδομένων έχει σημασία. Η βασική ιδέα πίσω από τα RNN είναι ότι διατηρούν μια κρυφή κατάσταση που διατηρεί τις πληροφορίες από προηγούμενες χρονικές στιγμές. Αυτή η κρυφή κατάσταση ενημερώνεται σε κάθε χρονικό βήμα καθώς τα νέα δεδομένα εισάγονται στο δίκτυο.[33]
- **Εκπαίδευση:**  
Κατά τη διάρκεια της φάσης εκπαίδευσης, το μοντέλο RNN μαθαίνει να ανιχνεύει μοτίβα και εξαρτήσεις στα δεδομένα της χρονοσειράς. Αυτό το κάνει προσαρμόζοντας τα εσωτερικά του βάρη και τις αποκλίσεις για να ελαχιστοποιήσει το σφάλμα πρόβλεψης στα δεδομένα εκπαίδευσης. Η διαδικασία εκπαίδευσης συνήθως περιλαμβάνει τον υπολογισμό της οπισθοδρόμησης μέσω του χρόνου (BPTT), που είναι μια παραλλαγή της οπισθοδρόμησης σχεδιασμένη για σειριακά δεδομένα.
- **Πρόβλεψη:** Αφού το RNN εκπαιδευτεί, μπορεί να χρησιμοποιηθεί για πρόβλεψη νέων δεδομένων χρησιμοποιώντας το.



Τα Αναδρομικά Νευρωνικά Δίκτυα (RNNs) φέρνουν ένα μοναδικό πλεονέκτημα στην πρόβλεψη των χρονοσειρών, επιτρέποντάς μας να αποτυπώνουμε περίπλοκες χρονικές εξαρτήσεις. Η εξερεύνηση των μοντέλων Long Short-Term Memory (LSTM) και Gated Recurrent Unit (GRU) είναι οι ευκαιρίες να έρθουμε σε επαφή με την πραγματική δυνατότητα της προβλεπτικής ανάλυσης.

### **3. Μελέτες περίπτωσης: Εφαρμογή Ανατροφοδοτούμενου Νευρωνικού Δικτύου σε πραγματικά δεδομένα**

Στα πλαίσια της παρούσας διπλωματικής εργασίας ως πηγή άντλησης δεδομένων, της υλοποίησης που πραγματοποιήθηκε, επιλέχθηκε το twitter καθώς αποτελεί ένα μέσο πληροφόρησης που μας προσφέρει ένα αρκετά «προσγειωμένο» περιεχόμενο κοντά στην πραγματικότητα εφόσον προκύπτει άμεσα από τους ίδιους τους χρήστες. Έτσι στο επόμενο κεφάλαιο θα δούμε αναλυτικότερα τι υλοποιήθηκε, τα αποτελέσματα του και το πως συνδυάστηκαν πρακτικώς όλες οι έννοιες που συζητήθηκαν στα προηγούμενα κεφάλαια. Στην παρούσα φάση θα δούμε τους βασικούς θεματικούς πυλώνες που επέλεξα να περιηγηθώ μέσω του twitter και πως αντλήθηκαν τα δεδομένα αυτά προκειμένου να δομηθεί ένα Ανατροφοδοτούμενο νευρωνικό δίκτυο με χρήση πραγματικών συμβάντων-γεγονότων. Ως ερευνητική μέθοδος χρησιμοποιήθηκε η μελέτη περίπτωσης δύο γεγονότων, της τεράστιας πυρκαγιάς στο Μάτι αλλά και του σιδηροδρομικού δυστυχήματος στα Τέμπη.

Η χρήση πραγματικών δεδομένων για τον όγκο δεδομένων των χρονοσειρών προτιμάται επειδή αντικατοπτρίζει τις πραγματικές συνθήκες, τα πρότυπα και τις διακυμάνσεις στο φαινόμενο που προσπαθούμε να μοντελοποιήσουμε και να αναλύσουμε. Παρατίθενται ορισμένοι λόγοι για τους οποίους προτιμάται το πραγματικό δεδομένο για την ανάλυση των χρονοσειρών περιλαμβάνουν:

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

- **Ρεαλιστική Αναπαράσταση:** Τα πραγματικά δεδομένα παρέχουν μια πιο ρεαλιστική αναπαράσταση της υποκείμενης διαδικασίας ή συστήματος. Περιλαμβάνουν τους διάφορους παράγοντες, το θόρυβο και τις ανωμαλίες που υπάρχουν στα πραγματικά δεδομένα, κάνοντας την ανάλυση πιο ακριβή και εφαρμοστή στον πραγματικό κόσμο.
- **Πολυπλοκότητα:** Τα δεδομένα των χρονοσειρών σε πραγματικές εφαρμογές είναι συχνά πολύπλοκα και επηρεάζονται από πολλούς παράγοντες. Τα πραγματικά δεδομένα αντικατοπτρίζουν αυτήν την πολυπλοκότητα, επιτρέποντας μια πιο σφαιρική ανάλυση της υποκείμενης δυναμικής.
- **Αναγνώριση Προτύπων:** Η ανάλυση των χρονοσειρών συχνά περιλαμβάνει την ανίχνευση προτύπων, τάσεων και εποχικών διακυμάνσεων. Τα πραγματικά δεδομένα επιτρέπουν τον εντοπισμό των λεπτών προτύπων και των ανωμαλιών που ενδέχεται να μην υπάρχουν σε συνθετικά ή δημιουργημένα δεδομένα.
- **Επαλήθευση του Μοντέλου:** Κατά την ανάπτυξη μοντέλων πρόβλεψης χρονοσειρών ή προγνωστικών μοντέλων, η χρήση πραγματικών δεδομένων επιτρέπει καλύτερη επικύρωση του μοντέλου. Τα μοντέλα που εκπαιδεύονται με πραγματικά δεδομένα έχουν μεγαλύτερη πιθανότητα να αποδώσουν καλά όταν εφαρμόζονται σε νέα, μη έγκυρα δεδομένα. [34]

### 3.1 Μελέτη περίπτωσης «Κεφάλαιο Μάτι, 2008»

Οι πυρκαγιές στη Μάτι αποτελούν τη δεύτερη πιο θανατηφόρα πυρκαγιά του αιώνα μετά τις πυρκαγιές του 2009 στην Αυστραλία, κατά τις οποίες έχασαν τη ζωή τους 173 άνθρωποι. Είναι επίσης η έκτη πιο θανατηφόρα πυρκαγιά των τελευταίων 100 ετών και αποτελεί μια από τις μεγαλύτερες εθνικές τραγωδίες του ελληνικού κράτους.

Η πυρκαγιά, ξεκίνησε από την Πεντέλη, στις 23 Ιουλίου 2018, διαδόθηκε πολύ γρήγορα στις παραθαλάσσιες πόλεις του Μάτι και του Κόκκινου Λιμανακίου, ανατολικά της Αθήνας. Οι ανελέητες φλόγες προκάλεσαν τον θάνατο 102 ανθρώπων και τον τραυματισμό περίπου 250 άλλων, ενώ περίπου 2.500 σπίτια καταστράφηκαν ολοσχερώς ή υπέστησαν ζημιές. Οι φλόγες ήταν τόσο ισχυρές που παγίδευαν και έκαιγαν ανθρώπους μέσα στα σπίτια τους, τα αυτοκίνητά τους ή λίγα μέτρα μακριά από την παραλία. Χιλιάδες οχήματα και σπίτια καταστράφηκαν πριν ο πύρινος εφιάλτης τεθεί υπό έλεγχο ώρες αργότερα. Πολλά ζώα (άγρια και κατοικίδια) πέθαναν ή τραυματίστηκαν από τη φωτιά.[35]



Εικόνα 19: Πυρκαγιά στο Μάτι, 2008

### 3.2 Μελέτη περίπτωσης , «Κεφάλαιο Τέμπη, 2023»

Στις 28 Φεβρουαρίου 2023, σημειώθηκε μια καταστροφική σύγκρουση μεταξύ δύο τρένων στα Τέμπη στη βόρεια Ελλάδα. Η σύγκρουση, προκλήθηκε μεταξύ του επιβατικού InterCity 62 που μετέφερε περίπου 350 άτομα συγκρούστηκε με ένα εμπορικό τρένο προκαλώντας τον θάνατο τουλάχιστον 57 ατόμων. Το επιβατικό τρένο ήταν προγραμματισμένο να αναχωρήσει από την Αθήνα για τη Θεσσαλονίκη ενώ το εμπορικό από Θεσσαλονίκη προς Λάρισα. Ανακαλύφθηκε ότι στο επιβατικό τρένο είχε επιτραπεί να συνεχίσει στην λάθος ράγα και να προσπεράσει πινακίδες προειδοποίησης παρά την παρουσία του εμπορικού τρένου στην ίδια ράγα. Το περιστατικό συνέβη όταν το επιβατικό τρένο βγήκε από έναν σήραγγα στον δήμο Τεμπών. Τα πρώτα τέσσερα βαγόνια του επιβατικού τρένου εξετράπησαν, και τα δύο ρώτα πήραν φωτιά και καταστράφηκαν ολοσχερώς. Χαρακτηρίστηκε ως η μεγαλύτερη τραγωδία στην ιστορία των ελληνικών σιδηροδρόμων. [36]



Εικόνα 20: Σιδηροδρομικό δυστύχημα Τέμπη, 2023.

### 3.3 Δεδομένα Twitter & Ανάκτησή τους

Από τα πρώτα κίολας κεφάλαια αναφερθήκαμε στα κοινωνικά δίκτυα όντας δομική έννοια στην ερευνητική αυτή προσέγγιση. Παρακάτω θα προσπαθήσουμε μαζί να δούμε πως ένα από τα μεγαλύτερα κοινωνικά δίκτυα, το twitter, με μέρος του όγκου δεδομένων που διαθέτει μπορεί να μας βοηθήσει «να δούμε» πίσω από μια ανάρτηση. Ερχόμενοι έτσι πιο κοντά στο ζήτημα που επικεντρώνεται η συγκεκριμένη διπλωματική εργασία θα χρησιμοποιήσουμε την γλώσσα προγραμματισμού Python και τις απαραίτητες βιβλιοθήκες της προκειμένου να δούμε τα πως ξετυλίχθηκαν στο βάθος του χρόνου έως σήμερα τα γεγονότα στο Μάτι αλλά και στα Τέμπη.

Παρακάτω θα κάνουμε μια σύντομη επεξήγηση σε βασικούς όρους που θα συναντάμε από αυτό το κεφάλαιο και στο εξής προκειμένου να δομήσουμε ένα μοντέλο με βάση LTMS όπου θα προβλέπει το πλήθος των αναρτήσεων.

Επιλέχθηκε η χρήση της Python καθώς προσφέρει πολλά πλεονεκτήματα όσον αφορά την ανάλυση χρονοσειρών:

1. Φιλική προς τον χρήστη γλώσσα: Η Python είναι γνωστή για την απλότητα και τη φιλικότητά της προς τον χρήστη. Διαθέτει μια ευανάγνωστη σύνταξη που την καθιστά εύκολη στην μάθηση, ακόμη και για αρχάριους. Η καθαρή δομή του κώδικα στη Python προάγει αποτελεσματικές πρακτικές κωδικοποίησης, επιτρέποντας την επικέντρωση στην ανάλυση δεδομένων σειρών χρόνου.
2. Ανοικτού Κώδικα: Ένα από τα μεγάλα πλεονεκτήματα της Python είναι ότι πρόκειται για μια γλώσσα ανοικτού κώδικα. Αυτό σημαίνει ότι είναι δωρεάν διαθέσιμη για όλους και βελτιώνεται συνεχώς από μια ζωντανή κοινότητα προγραμματιστών. Η ανοικτή φύση της Python επιτρέπει στους επιστήμονες δεδομένων να έχουν πρόσβαση σε πλούσιους πόρους, εργαλεία και βιβλιοθήκες για την ανάλυση των δεδομένων σειρών χρόνου χωρίς πρόσθετο κόστος.

3. Εκτεταμένη υποστήριξη βιβλιοθηκών: Διαθέτει έναν εκτεταμένο κατάλογο εξειδικευμένων βιβλιοθηκών και εργαλείων που σχεδιάστηκαν ειδικά για την ανάλυση σειρών χρόνου. Αυτές οι βιβλιοθήκες, όπως η pandas, η NumPy, η statsmodels και η scikit-learn, παρέχουν διάφορες λειτουργίες και εργαλεία προσαρμοσμένα στις μοναδικές προκλήσεις της εργασίας με δεδομένα που εξαρτώνται από τον χρόνο. Απλοποιούν πολύπλοκες λειτουργίες, επιτρέποντάς να εστιάσουμε στην εξαγωγή σημαντικών ευρημάτων.

4. Διευκολύνει την επαναχρησιμοποίηση του κώδικα: Λόγω της μακράς διάρκειάς της και της ευρείας υιοθέτησής της, η Python διαθέτει έναν τεράστιο κώδικα που οι επιστήμονες δεδομένων και οι προγραμματιστές εφαρμογών μπορούν να αξιοποιήσουν για τις ανάγκες της ανάλυσης σειρών χρόνου τους.

Πολλές κοινές εργασίες, όπως το upload δεδομένων, η επιλογή, η μετατροπή και η οπτικοποίηση, έχουν ήδη υλοποιηθεί και κοινοποιηθεί από την κοινότητα της Python. Αυτό επιτρέπει να εξοικονομήσουμε χρόνο και προσπάθεια χρησιμοποιώντας υπάρχοντα κώδικα και λύσεις, επιταχύνοντας τη διαδικασία ανάλυσης. [37]

- Η βιβλιοθήκη Tweepy είναι μια βιβλιοθήκη για τη γλώσσα προγραμματισμού Python που παρέχει εύκολη πρόσβαση στο API του Twitter. Απλοποιεί τη διαδικασία αλληλεπίδρασης με τα δεδομένα του Twitter και επιτρέπει την εκτέλεση εργασιών όπως ανάκτηση tweets, ανάρτηση tweets, follow χρηστών και πολλά άλλα μέσω κώδικα Python. Το Tweepy καθιστά βολικό τον χειρισμό δεδομένων του Twitter στα έργα Python, συμπεριλαμβανομένης της εξαγωγής δεδομένων, της ανάλυσης και της αυτοματοποίησης των εργασιών που σχετίζονται με το Twitter.

- Η βιβλιοθήκη pandas: Είναι μια δημοφιλής βιβλιοθήκη της Python που χρησιμοποιείται για τον χειρισμό και την ανάλυση δεδομένων. Παρέχει δομές δεδομένων και λειτουργίες που είναι εύκολο να χρησιμοποιηθούν για την εργασία με δομημένα δεδομένα, όπως φύλλα εργασίας ή πίνακες SQL. Εδώ είναι μερικά βασικά χαρακτηριστικά και συστατικά της βιβλιοθήκης `pandas`:
  1. DataFrame: Η κύρια δομή δεδομένων στο `pandas` είναι το DataFrame. Είναι μια δισδιάστατη, μεταβλητή ως προς το μέγεθος, και ανομοιογενής δομή δεδομένων πίνακα με ετικεταρισμένους άξονες (γραμμές και στήλες). Τα DataFrames είναι παρόμοια με πίνακες SQL ή φύλλα εργασίας Excel και μπορούν να αποθηκεύουν δεδομένα διάφορων τύπων.
  2. Series: Ένα Series είναι ένας μονοδιάστατος αντικείμενο παρόμοιο με έναν πίνακα. Χρησιμοποιείται
- Αρχείο CSV (Comma-Separated Values) είναι ένα απλό και ευρέως χρησιμοποιούμενο αρχείο μορφής για την αποθήκευση δομημένων δεδομένων σε ένα απλό κείμενο μορφής. Σε ένα αρχείο CSV, τα δεδομένα οργανώνονται σε γραμμές και στήλες, με κάθε γραμμή να αντιπροσωπεύει ένα αρχείο, και κάθε στήλη να αντιπροσωπεύει ένα πεδίο ή χαρακτηριστικό αυτού του αρχείου.

Για να ανακτήσουμε τα δεδομένα από το Twitter και να τα επεξεργαστούμε ακολουθούμε τον εξής οδηγό:

- Δημιουργία λογαριασμού developer στο Twitter
- Συνδεόμαστε στην πλατφόρμα ανάπτυξης του Twitter στη διεύθυνση <https://developer.twitter.com/>
- Δημιουργούμε μια εφαρμογή ανάπτυξης Twitter και λαμβάνουμε τα απαραίτητα κλειδιά και τα διακριτικά του API

### Κλειδιά και Διακριτικά του API:

- Μόλις δημιουργήσουμε την εφαρμογή ανάπτυξης Twitter, διαθέτουμε τα ακόλουθα διαπιστευτήρια : Κλειδί Καταναλωτή (API Key), Μυστικό Καταναλωτή (API Secret Key), Κλειδί Πρόσβασης (Access Token), Μυστικό Κλειδί Πρόσβασης (Access Token Secret)

### Επιλογή Βιβλιοθήκης:

- Υπάρχει η δυνατότητα επιλογής διαφόρων γλωσσών προγραμματισμού, βιβλιοθηκών ή εργαλείων τρίτων για την πρόσβαση στα δεδομένα του Twitter. Συνηθισμένες επιλογές περιλαμβάνουν τη χρήση της Python με το Tweepy, ή τη χρήση εξειδικευμένων εργαλείων ανάλυσης κοινωνικών μέσων.

### Πιστοποίηση:

- Χρησιμοποιούμε τα κλειδιά και τα διακριτικά του API για να λάβουμε την πιστοποίηση στο API του Twitter μέσω της επιλεγμένης βιβλιοθήκης ή εργαλείου.
- Χρησιμοποιούμε τη βιβλιοθήκη ή το εργαλείο για να κάνουμε αιτήματα προς το API. Συνήθεις τύποι αιτήσεων περιλαμβάνουν: Ανάκτηση χρονολογιών χρηστών (tweets από έναν συγκεκριμένο χρήστη), Αναζήτηση tweets με βάση λέξεις-κλειδιά, hashtags ή αναφορές χρηστών.

### Επεξεργασία Δεδομένων:

- Αφού ανακτήσουμε τα δεδομένα, έχουμε την δυνατότητα να επιλέξουμε πως θα τα επεξεργαστούμε. Για παράδειγμα μπορούμε να κάνουμε sentiment analysis και να εξάγουμε συγκεκριμένες πληροφορίες από τα tweets.
- Μπορούμε να αποθηκεύσουμε τα δεδομένα σε διάφορες μορφές, συμπεριλαμβανομένων των μορφών CSV, JSON.[38]

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων



Για την εξαγωγή δεδομένων από το Twitter και την αποθήκευση τους ως αρχείο CSV, ακολουθούμε τα εξής βήματα:

Δημιουργία Λογαριασμού Ανάπτυξης Twitter:

- Μεταβαίνουμε στην πλατφόρμα ανάπτυξης του Twitter (<https://developer.twitter.com/en>)
- Κάνουμε σύνδεση με ήδη υπάρχον λογαριασμό Twitter ή δημιουργούμε έναν καινούργιο.
- Υποβάλλουμε αίτηση για έναν λογαριασμό developer και ακολουθούμε τα βήματα για τη δημιουργία ενός λογαριασμού developer και ενός project.

Δημιουργία Twitter App:

- Αφού συνδεθούμε, μεταβαίνουμε στο Developer Dashboard και δημιουργούμε μια νέα εφαρμογή συμπληρώνοντας τις απαιτούμενες πληροφορίες.

Λήψη Κλειδιών και Tokens της API:

- Αφού δημιουργήσουμε την εφαρμογή, πηγαίνουμε στην καρτέλα "Keys and Tokens".
- Δημιουργούμε τα μοναδικά κλειδιά API, τα μυστικά κλειδιά API, τα Access Token και τα μυστικά Access Token.

Εγκατάσταση μιας βιβλιοθήκης Python:

- Θα χρειαστούμε μια βιβλιοθήκη για την αλληλεπίδραση με την API του Twitter. Το Tweepy είναι μια δημοφιλής επιλογή όπως αναφέραμε παραπάνω. [39]

- Η εγκατάσταση της γίνεται χρησιμοποιώντας το pip:

```
Microsoft Windows [Version 10.0.22621.2283]
(c) Microsoft Corporation. Με επιφύλαξη κάθε νόμιμου δικαιώματος.

C:\Users\katep>pip install tweety
Collecting tweety
  Downloading tweety-0.1.6.tar.gz (4.8 kB)
  Preparing metadata (setup.py) ... done
Collecting beautifulsoup4
  Downloading beautifulsoup4-4.12.2-py3-none-any.whl (142 kB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 143.0/143.0 kB 941.4 kB/s eta 0:00:00
Requirement already satisfied: pandas in c:\users\katep\appdata\local\programs\python\python310\lib\site-packages (from tweety) (1.3.4)
Collecting selenium
  Downloading selenium-4.13.0-py3-none-any.whl (9.5 MB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 9.5/9.5 MB 3.1 MB/s eta 0:00:00
Collecting spacy
  Downloading spacy-3.6.1-cp310-cp310-win_amd64.whl (12.0 MB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 12.0/12.0 MB 4.6 MB/s eta 0:00:00
Collecting vaderSentiment
  Downloading vaderSentiment-3.3.2-py2.py3-none-any.whl (125 kB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 126.0/126.0 kB 737.0 kB/s eta 0:00:00
Collecting textblob
  Downloading textblob-0.17.1-py2.py3-none-any.whl (636 kB)
    ━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━━ 636.8/636.8 kB 3.3 MB/s eta 0:00:00
Collecting chromedriver-binary==74.0.3729.6.0
  Downloading chromedriver-binary-74.0.3729.6.0.tar.gz (3.3 kB)
  Preparing metadata (setup.py) ... done
Collecting sopsieve>1.2
  Downloading sopsieve-2.5-py3-none-any.whl (36 kB)
Requirement already satisfied: python-dateutil>=2.7.3 in c:\users\katep\appdata\local\programs\python\python310\lib\site
```

Εικόνα 21: Κώδικας με εγκατάσταση Tweepy

- Ενδεικτικός κώδικας Python με χρήση μοναδικών κλειδιών που μας προσφέρουν πρόσβαση στα δεδομένα:

```
import tweepy
import csv

consumer_key = 'YOUR_CONSUMER_KEY'
consumer_secret = 'YOUR_CONSUMER_SECRET'
access_token = 'YOUR_ACCESS_TOKEN'
access_token_secret = 'YOUR_ACCESS_TOKEN_SECRET'

auth = tweepy.OAuthHandler(consumer_key, consumer_secret)
auth.set_access_token(access_token, access_token_secret)
api = tweepy.API(auth)

# Define the search query
.. ..

Open a CSV file for writing
th open('twitter_data.csv', 'w', newline='') as file:
    writer = csv.writer(file)
    writer.writerow(['Tweet Text', 'Username', 'Date'])

# Iterate through the tweets and write to CSV
for tweet in tweepy.Cursor(api.search, q=query, lang="en").items(100):
    writer.writerow([tweet.text, tweet.user.screen_name, tweet.created_at])
```

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

Εικόνα 22: κώδικας Python με χρήση μοναδικών κλειδιών

### **3.4 Χρονοσειρές όγκου δεδομένων**

Ο όρος "όγκος δεδομένων χρονοσειρών" αναφέρεται στην ποσότητα των δεδομένων των χρονοσειρών που είναι διαθέσιμα ή χρησιμοποιούνται για ανάλυση. Αυτό μετρά το μέγεθος ή την κλίμακα του συνόλου δεδομένων των χρονοσειρών, συνήθως μετρώντας τον αριθμό των σημείων δεδομένων ή παρατηρήσεων που συλλέγονται κατά μήκος ενός συγκεκριμένου χρονικού διαστήματος. Ο όγκος των δεδομένων χρονοσειρών μπορεί να ποικίλλει ευρέως ανάλογα με την εφαρμογή και τη συχνότητα συλλογής δεδομένων, κυμαίνονται από μικρά σύνολα δεδομένων με λίγα σημεία δεδομένων έως μεγάλα σύνολα δεδομένων με εκατομμύρια ή ακόμη και δισεκατομμύρια σημεία δεδομένων.

#### **3.4.1 Η πρόβλεψη μελλοντικών τιμών με βάση τα υπάρχοντα δεδομένα**

Η Python προσφέρει μια ποικιλία βιβλιοθηκών και τεχνικών για τον προσδιορισμό των μελλοντικών τιμών σε σειρές χρόνου, και ένα δημοφιλές μέθοδος είναι το αυτοσυσχετιζόμενο ολοκληρωμένο κινητό μέσο (ARIMA) μοντέλο. Το ARIMA είναι μια ισχυρή και ευρέως χρησιμοποιούμενη προσέγγιση που συνδυάζει τα εξής τρία στοιχεία για να αντλήσει τα πρότυπα και τις τάσεις στα δεδομένα σειρών χρόνου:

1. Autoregressive (AR): Αυτό το στοιχείο αντιπροσωπεύει την αυτοσυσχέτιση των χρονοσειρών. Το ARIMA μοντέλο λαμβάνει υπόψη την αυτοσυσχέτιση για να προβλέψει μελλοντικές τιμές, δηλαδή η τρέχουσα τιμή εξαρτάται από προηγούμενες τιμές.

2. Integrated : Αυτό το στοιχείο αντιπροσωπεύει τον βαθμό ολοκλήρωσης της χρονοσειράς. Με απλά λόγια, αναδεικνύει πώς

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

διαφοροποιούνται τα δεδομένα για να χαρακτηριστούν στάσιμα (stationary) ή όχι. Η στάσιμη χρονοσειρά είναι πιο εύκολη για πρόβλεψη.

3. Moving Average (MA): Αυτό το στοιχείο αντιπροσωπεύει τον κινητό μέσο όρο των προηγούμενων τιμών στη χρονοσειρά. Βοηθά στην απομάκρυνση των τυχαίων διακυμάνσεων και επιτρέπει την ανίχνευση προτύπων.

Το ARIMA μοντέλο είναι ένα ισχυρό εργαλείο για την πρόβλεψη μελλοντικών τιμών σε σειρές χρόνου και μπορεί να εφαρμοστεί χρησιμοποιώντας την Python. [40]

### 3.4.2 Κατηγορίες μοντέλων πρόβλεψης

Στην ανάλυση σειρών χρόνου, υπάρχουν διάφορα μοντέλα πρόβλεψης για τον προσδιορισμό μελλοντικών τιμών βασισμένων σε ιστορικά δεδομένα. Κάθε μοντέλο έχει τα δικά του πλεονεκτήματα, περιορισμούς και καταλληλότητα για διάφορα είδη χρονοσειρών. Ας εξερευνήσουμε μερικά κοινά είδη μοντέλων πρόβλεψης:

1. Κινητός Μέσος (Moving Average - MA): Το μοντέλο του κινητού μέσου υπολογίζει τον μέσο όρο των προηγούμενων παρατηρήσεων για να προβλέψει τις μελλοντικές τιμές. Βοηθά στην εξάλειψη των σύντομων διακυμάνσεων και στον εντοπισμό των βασικών τάσεων στα δεδομένα.

2. Αυτοσυσχετιζόμενο (Autoregressive - AR): Το μοντέλο του αυτοσυσχετιζόμενου χρησιμοποιεί παρελθοντικές παρατηρήσεις και μια γραμμική εξίσωση παλινδρόμησης για να προβλέψει τις μελλοντικές τιμές. Υποθέτει ότι οι μελλοντικές τιμές εξαρτώνται από τις προηγούμενες τιμές με καθυστέρηση.

3. Αυτοσυσχετιζόμενο Κινητό Μέσος (Autoregressive Moving Average - ARMA): Το μοντέλο ARMA συνδυάζει τα μοντέλα AR και MA,

λαμβάνοντας υπόψη τόσο τις παρελθοντικές παρατηρήσεις όσο και το μέσο όρο των παρελθοντικών σφαλμάτων για τις προβλέψεις.

4. Αυτοσυσχετιζόμενο Ολοκληρωμένο Κινητό Μέσος (Autoregressive Integrated Moving Average - ARIMA): Το μοντέλο ARIMA επεκτείνει το μοντέλο ARMA με την ενσωμάτωση της διαφοροποίησης για να καταστήσει τη χρονοσειρά στάσιμη. Είναι κατάλληλο για μη στάσιμα δεδομένα με τάσεις και εποχικότητα.

5. Εκθετική Εξομάλυνση: Τα μοντέλα εκθετικής εξομάλυνσης εφαρμόζουν βάρη στις παρελθοντικές παρατηρήσεις, δίνοντας μεγαλύτερη σημασία στις πρόσφατες τιμές. Διάφορες παραλλαγές της εκθετικής εξομάλυνσης, όπως η Απλή Εκθετική Εξομάλυνση (SES), η Γραμμική Εκθετική Εξομάλυνση του Holt και η Εκθετική Εξομάλυνση του Holt-Winters, προσαρμόζονται σε διάφορα πρότυπα στα δεδομένα.

6. Εποχικό ARIMA : Το SARIMA είναι μια επέκταση του μοντέλου ARIMA που λαμβάνει υπόψη εποχικά πρότυπα στα δεδομένα. Περιλαμβάνει επιπλέον όρους για την ανίχνευση της εποχικότητας, κάνοντάς το κατάλληλο για χρονοσειρές με επαναλαμβανόμενα πρότυπα.

Κάθε μοντέλο έχει τη χρήση του ανάλογα με τα δεδομένα και τις ανάγκες πρόβλεψης. [41]

## **4. Μεθοδολογία ανάπτυξης μοντέλου βαθιάς μάθησης LTSM σε χρονοσειριακά δεδομένα.**

### **4.1. «Περίπτωση Μάτι και Περίπτωση Τέμπη»**

Στην αρχή του προηγούμενου κεφαλαίου αναφερθήκαμε συνοπτικά στις δυο μελέτες περίπτωσης που θα αποτελέσουν και την πηγή δεδομένων της παρακάτω ανάπτυξης του μοντέλου βαθιάς μάθησης LTSM. Τόσο η περίπτωση των Τεμπών όσο και στο Μάτι αποτέλεσαν δύο βαθύτατες τραγωδίες που έπληξαν την χώρα μας σε περιβαλλοντολογικό, κοινωνικό και ανθρωπιστικό επίπεδο.

Ο τρόπος με τον οποίο ένα γεγονός επηρεάζει τα κοινωνικά μέσα επικοινωνίας μπορεί να είναι σημαντικός για την κατανόηση των διαδικτυακών διαδικασιών και της αντίδρασης του κοινού. Μερικοί τρόποι με τους οποίους ένα γεγονός μπορεί να επηρεάσει τα κοινωνικά μέσα είναι:

- **Δημοσίευση και Κοινοποίηση:** Όταν συμβαίνει ένα σημαντικό γεγονός, οι άνθρωποι τείνουν να δημοσιεύουν σχετικές πληροφορίες στα κοινωνικά μέσα. Αυτό μπορεί να περιλαμβάνει κείμενα, φωτογραφίες, βίντεο ή ακόμα και ζωντανά μεταδόσεις. Οι χρήστες επίσης κοινοποιούν συχνά τις δημοσιεύσεις αυτές, επεκτείνοντας το εύρος της πληροφορίας.
- **Συζήτηση και Σχόλια:** Τα γεγονότα πυροδοτούν διαδικτυακές συζητήσεις και σχόλια. Οι άνθρωποι εκφράζουν τις απόψεις τους, συζητούν τα γεγονότα και ανταλλάσσουν απόψεις με άλλους χρήστες μέσω σχολίων και αναρτήσεων.
- **Επιδράσεις στα Θέματα Συνομιλίας (Trending Topics):** Τα κοινωνικά μέσα συχνά έχουν λειτουργία "Trending Topics" όπου τα πιο σημαντικά γεγονότα και λέξεις-κλειδιά της ημέρας παρουσιάζονται στους χρήστες. Ένα σημαντικό γεγονός μπορεί να επηρεάσει την κατάταξη αυτών των θεμάτων.

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

- **Επιρροή στη Διαδικτυακή Συμπεριφορά:**  
Τα γεγονότα μπορούν να επηρεάσουν τη συμπεριφορά των χρηστών στο διαδίκτυο. Για παράδειγμα, ένα γεγονός που σχετίζεται με την υγεία μπορεί να οδηγήσει σε αυξημένη αναζήτηση πληροφοριών σχετικά με την υγεία στο διαδίκτυο.
- **Πολιτικά Συμβάντα:** Εκδηλώσεις που αφορούν το διάστημα ή την πολιτική μπορούν να προκαλέσουν έντονες αντιδράσεις στα κοινωνικά μέσα, καθώς η επιρροή τους εκτείνεται πέρα από τα συγκεκριμένα γεγονότα.

Είναι σημαντικό να σημειωθεί ότι οι επιρροές αυτές μπορούν να είναι τόσο θετικές όσο και αρνητικές, και μπορούν να διαφέρουν ανάλογα με τη φύση και τον τύπο του γεγονότος. Επίσης, η διάδραση με τα κοινωνικά μέσα είναι σύνθετη και εξαρτάται από πολλούς παράγοντες, συμπεριλαμβανομένων των ατόμων, των κοινοτήτων.

Παρακάτω θα δούμε τις σχετικές υλοποιήσεις για τις δυο περιπτώσεις και πως το κοινό «αντέδρασε» και προσέλαβε όλο αυτόν τον όγκο πληροφόρησης που πυροδότησαν τα δύο γεγονότα. Τέλος θα παρατεθούν οι μεθοδολογίες, οι δυνητικοί στόχοι αλλά και τα αποτελέσματα που προέκυψαν κατά την ανάπτυξη του μοντέλου. [45]

#### **4.2 Συλλογή και επεξεργασία δεδομένων**

Η συλλογή των δεδομένων έγινε με βάση την μεθοδολογία που συζητήσαμε στο κεφάλαιο 3.3 και για τις δύο μελέτες περίπτωσης.

Έγινε χρήση της πλατφόρμας twitter developer όπου με την σχετική απόκτηση των API keys και των Tokens της εφαρμογής “mati case” & “tempri case”. Χρησιμοποιήθηκαν έτσι τα κλειδιά API, τα κλειδιά API Secret, το Access Token και το Access Token Secret για να πιστοποιηθεί η εφαρμογή κατά την υποβολή αιτημάτων στο Twitter API. Έτσι ξεκίνησα τα API requests στο Twitter για τη λήψη δεδομένων και συγκέντρωσα δυο αρχεία .csv για την κάθε περίπτωση ξεχωριστά, όπου

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

χρησιμοποιήθηκαν ως οι πυλώνες του περιεχομένου για τα αποτελέσματα που θα δούμε παρακάτω.

Ανάλογα με τον απώτερο σκοπό που επιθυμεί ο χρήστης του API υπάρχει η δυνατότητα να έχει κανείς πρόσβαση σε διάφορα σημεία εκκίνησης (endpoints) του Twitter API, όπως το Search API για την αναζήτηση των tweets, το Streaming API για δεδομένα σε πραγματικό χρόνο ή το User API για τη λήψη πληροφοριών που αφορούν συγκεκριμένους χρήστες. Στην προκειμένη περίπτωση έγινε χρήση του Search API.

### Αρχεία .CSV

Το αρχείο mati.csv αποτελείται από δεδομένα τα οποία αναφέρονται στο χρονικό εύρος από “2018-07-23” έως και “2023-05-31” δηλαδή από την μέρα που συνέβη το γεγονός έως και σήμερα.

Τα δεδομένα φορτώνονται από ένα αρχείο CSV με όνομα 'mati\_III.csv'. Το αρχείο αυτό περιέχει πληροφορίες σχετικά με tweets, όπως η ημερομηνία δημιουργίας, η γεωγραφική θέση, ο αριθμός των likes, retweets, αλλά και το αναγνωριστικό του χρήστη. Συνολικά για αυτή την περίπτωση χρησιμοποιήθηκαν 289.714 τιμές αναφοράς.

# A	T B	T C	# D	T E	# F	# G	# H	# I
1460475500	2018-07-23 18:38:32		1021464595108638720	eL	0	0	0	382
3288273220	2018-07-23 18:38:16		1021464529052618752	eL	0	0	0	382
601776175	2018-07-23 18:38:04		1021464480285746176	eL	0	0	0	382
846657624540090369	2018-07-23 18:37:49		1021464415869251584	eL	0	0	0	382
2429122875	2018-07-23 18:36:53		1021464181961371650	eL	0	0	0	382
873904374828994560	2018-07-23 18:36:51		1021464174654894080	eL	0	0	0	382
2639228129	2018-07-23 18:36:51		1021464174063439872	eL	0	0	0	6
387944906	2018-07-23 18:36:42		1021464134439833600	eL	0	0	0	382
38168419	2018-07-23 18:36:27		1021464071894429696	eL	0	0	0	109
972565354072494080	2018-07-23 18:36:12		1021464088820510724	eL	0	0	0	109
97532855595976704	2018-07-23 18:36:07		1021463989350535168	eL	0	0	0	382
2365542710	2018-07-23 18:35:53		1021463929766203395	eL	0	0	0	382
883452886130323459	2018-07-23 18:35:31		1021463837265088512	eL	0	0	0	382
967433594577793024	2018-07-23 18:35:25		1021463811923087361	eL	20	1	1	109
327966956	2018-07-23 18:35:09	2ee7eeaa84dbe65a	1021463747142078464	eL	3	0	0	0

Εικόνα 23: : Αρχείο mati.csv

X



Το αρχείο `tempri.csv` αποτελείται από δεδομένα τα οποία αναφέρονται στο χρονικό εύρος από “2023-02-28” έως και “2023-05-31” δηλαδή από την μέρα που συνέβη το γεγονός έως και σήμερα. Τα δεδομένα φορτώνονται από ένα αρχείο CSV με όνομα ‘`tempri.csv`’ και περιέχει πληροφορίες σχετικά με tweets, την ημερομηνία δημιουργίας, τη γεωγραφική θέση, των αριθμό των likes, retweets, αλλά και το αναγνωριστικό του χρήστη. Συνολικά για αυτή την περίπτωση χρησιμοποιήθηκαν 1.190.784 τιμές αναφοράς.

# A	T B	T C	# D	T E	# F	# G	# H	# I
1424448938281103362	2023-04-05 23:54:22		1643764045047033858	eL	0	0	0	1
485735885	2023-04-05 23:54:22		1643764043952320516	eL	0	0	0	1
1114692838128848896	2023-04-05 23:52:19		1643763529982562304	eL	0	0	0	49
1448791128844865540	2023-04-05 23:49:50		1643762902245965824	eL	1	0	1	1
2774804820	2023-04-05 23:49:24		1643762795106779137	eL	0	0	0	43
485735885	2023-04-05 23:47:01		1643762195287752705	eL	0	0	0	380
485735885	2023-04-05 23:46:03		1643761949363118080	eL	0	0	0	0
1391760354443661315	2023-04-05 23:45:21		1643761773521121280	eL	0	0	0	25
1352284537468948487	2023-04-05 23:40:42		1643760604262416384	eL	0	0	0	0
1235711744384917504	2023-04-05 23:38:36		1643760077491380225	eL	0	0	0	25
124918063	2023-04-05 23:33:33		1643758806323589120	qht	0	0	0	1
124918063	2023-04-05 23:33:12		1643758717731590146	qht	0	0	0	0
124918063	2023-04-05 23:32:56		1643758651797020672	qht	0	0	0	0
1495069539230437382	2023-04-05 23:31:36		1643758313212096512	eL	0	0	0	240
1287844230	2023-04-05 23:30:42		1643758089814908929	eL	0	0	0	24

Εικόνα 24: Αρχείο `tempri.csv`

#### 4.2.1 Βιβλιοθήκες & Περιβάλλον Ανάπτυξης

Ο κώδικας χρησιμοποιεί διάφορες βιβλιοθήκες όπως `pandas` και `numpy` για διαχείριση και επεξεργασία δεδομένων. Τα δεδομένα κλιμακώνονται χρησιμοποιώντας το `MinMaxScaler` από την `sklearn`. Το

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

μοντέλο πρόβλεψης είναι ένα νευρωνικό δίκτυο LSTM (Long Short-Term Memory) που δημιουργήθηκε με τη βοήθεια της βιβλιοθήκης Keras.

Για την αξιολόγηση του μοντέλου, χρησιμοποιούνται μετρικές όπως η RMSE (Root Mean Squared Error) και η MAE (Mean Absolute Error).

Τέλος, για την οπτικοποίηση των προβλέψεων χρησιμοποιείται η βιβλιοθήκη matplotlib.

### **4.3 Επιλογή μοντέλου LSTM**

#### **4.3.1 Διαδικασία**

Τα δεδομένα καθαρίζονται και οργανώνονται ανά λεπτό. Οι τιμές κλιμακώνονται στο εύρος [0,1]. Δημιουργούνται δεδομένα εκπαίδευσης και δοκιμής. Δημιουργείται το μοντέλο LSTM. Το μοντέλο εκπαιδεύεται με τα δεδομένα εκπαίδευσης. Γίνονται προβλέψεις για τα δεδομένα εκπαίδευσης και δοκιμής. Οι προβλέψεις συγκρίνονται με τα πραγματικά δεδομένα και υπολογίζονται τα σφάλματα RMSE και MAE. Τέλος, τα αποτελέσματα οπτικοποιούνται χρησιμοποιώντας ένα γράφημα.

#### **4.3.2 Επισκόπηση Στόχων Υλοποίησης**

Στόχοι του κώδικα αποτελούν:

1. Οπτικοποίηση του ημερησίου volume αναρτήσεων όταν σύνολο χρηστών περιοριστεί σε εκείνους που έχουν κάνει παραπάνω από έναν αριθμό tweets.
2. Πρόβλεψη του αριθμού των tweets που δημοσιεύονται κάθε λεπτό βασισμένο στα δεδομένα των προηγούμενων 24 ωρών και σε δεύτερη σκοπεύει να προβλέψει τον αριθμό των tweets ανά λεπτό βασισμένο σε ιστορικά δεδομένα, χρησιμοποιώντας το μοντέλο βαθιάς μάθησης που ονομάζεται LSTM (Long Short-Term Memory).

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

### 4.3.3 Βιβλιοθήκες και ρόλοι τους:

- pandas: Δημοφιλής βιβλιοθήκη για τη διαχείριση και ανάλυση δεδομένων.
- numpy: Χρησιμοποιείται για αριθμητικές λειτουργίες και λειτουργεί με πίνακες.
- sklearn.preprocessing: Ειδικότερα, η MinMaxScaler εισάγεται από αυτή τη μονάδα για νορμοποίηση του dataset ώστε όλες οι τιμές να βρίσκονται στο διάστημα [0,1].
- keras: Μια API νευρωνικών δικτύων υψηλού επιπέδου που επιτρέπει την εύκολη δημιουργία και εκπαίδευση μοντέλων.
- sklearn.metrics: Παρέχει εργαλεία για τον υπολογισμό κάποιων συνηθισμένων μετρήσεων για την αξιολόγηση των μοντέλων παλινδρόμησης.
- matplotlib: Μια βιβλιοθήκη για τη δημιουργία γραφημάτων στην Python.

### 4.4 Δημιουργία Datasets για το LSTM:

Dataset και Προ-επεξεργασία:

Το dataset temp1.csv φορτώνεται σε ένα DataFrame. Αυτό το dataset περιέχει πληροφορίες σχετικά με tweets, όπως ο συγγραφέας, η ημερομηνία δημιουργίας, η τοποθεσία, το περιεχόμενο, κ.ά. Τα tweets συγκεντρώνονται ανά λεπτό για να πάρουμε τον αριθμό των tweets για κάθε λεπτό. Τα δεδομένα στη συνέχεια νορμοποιούνται χρησιμοποιώντας το MinMaxScaler.

Καθορίζεται μια συνάρτηση create\_dataset, η οποία επεξεργάζεται τα χρονοσειριακά δεδομένα για να παραγάγει ακολουθίες εισόδου και αντίστοιχες εξόδους. Κάθε ακολουθία εισόδου περιέχει δεδομένα για την περασμένη εβδομάδα (1440 λεπτά \* 7 ημέρες) και το μοντέλο θα

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

μάθει να προβλέπει τον αριθμό των tweets για το επόμενο λεπτό βάσει αυτών των ιστορικών δεδομένων.

Διαίρεση Εκπαίδευσης-Δοκιμής:

Πραγματοποιείται δυναμική διαίρεση εκπαίδευσης-δοκιμής όπου ένα τμήμα των πιο πρόσφατων δεδομένων (15.000 σημεία δεδομένων) διατηρείται για δοκιμή, ενώ τα προηγούμενα δεδομένα χρησιμοποιούνται για εκπαίδευση.

LSTM Μοντέλο: Κατασκευάζεται ένα μοντέλο LSTM με τρεις στρώσεις:

- I. Πρώτη στρώση με 100 μονάδες και ενεργοποιημένη την επιλογή επιστροφής ακολουθίας για να περνάει τις εξόδους της στην επόμενη στρώση.
- II. Δεύτερη στρώση με 50 μονάδες, επιστρέφοντας επίσης ακολουθίες.
- III. Τρίτη στρώση με 50 μονάδες αλλά χωρίς επιστροφή ακολουθιών. Μια πυκνή στρώση με 1 νευρώνα, καθώς αυτό είναι ένα καθήκον παλινδρόμησης.

Το μοντέλο χρησιμοποιεί τη συνάρτηση απώλειας Mean Squared Error και τον βελτιστοποιητή Adam.

- Εκπαίδευση Μοντέλου: Το μοντέλο εκπαιδεύεται στα δεδομένα εκπαίδευσης ( $X_{train}$  και  $y_{train}$ ) για 500 εποχές με μέγεθος δέσμης 200.
- Προβλέψεις: Γίνονται προβλέψεις τόσο στα δεδομένα εκπαίδευσης όσο και στα δεδομένα δοκιμής.
- Μετρήσεις: Υπολογίζονται τα RMSE (Root Mean Squared Error) και MAE (Mean Absolute Error) για τα δεδομένα εκπαίδευσης και δοκιμής για την αξιολόγηση της απόδοσης του μοντέλου.
- Οπτικοποίηση: Χρησιμοποιώντας το matplotlib, «πλοτάρονται» οι πραγματικές εναντίον των προβλεπόμενων αριθμών tweets ανά λεπτό.

Ο κύριος στόχος του script είναι να δείξει πως εκπαιδεύεται ένα μοντέλο βαθιάς μάθησης (LSTM) σε χρονοσειριακά δεδομένα και πως αξιολογείται η απόδοσή του στην πρόβλεψη μελλοντικών σημείων δεδομένων.

Η ίδια διαδικασία ακολουθήθηκε και για το 'mati\_III.csv'.

#### **4.5 Αποτελέσματα οπτικοποίησης ημερησίου volume αναρτήσεων**

Στην πρώτη επιδίωξη αυτής της υλοποίησης περιορίστηκε το ημερήσιο volume των αναρτήσεων σε ένα υποσύνολο χρηστών το οποίο έχει παράξει παραπάνω από x αριθμό tweets. Στο αρχικό γράφημα, που θα δούμε παρακάτω (εικόνα 27), δημιουργήθηκε μέσω της plot και απλώς οπτικοποιεί τα δεδομένα εισόδου. Αυτό το γράφημα μας δίνει το ημερήσιο volume για χρήστες που έχουν κάνει tweets μεγαλύτερο ίσο το ένα και τους έχουμε εντός του dataset. Αυτό περιορίζει λίγο τα πράγματα όσο αφορά το πως θα εντοπίσουμε ποιοι είναι αυτοί οι χρήστες που φαίνεται να έχουν μια παραπάνω επιρροή στη διαδρομή της πληροφορίας; Ποιο θα είναι το ημερήσιο volume αναρτήσεων όταν σύνολο χρηστών περιοριστεί σε εκείνους που έχουν κάνει παραπάνω από έναν αριθμό tweets; Πως θα εξελιχθεί το daily volume αν βάλω μεγαλύτερο ίσο π.χ. του 800 και όχι του 1;

Αυτά είναι ερωτήματα που δημιουργήθηκαν και θα γίνει προσπάθεια να απαντηθούν προκειμένου να δούμε αν υπάρχουν εντός του dataset ή όχι χρήστες οι οποίοι φαίνεται να «επαναपुरοδοτούν» το φαινόμενο. Σε αυτή τη φάση λοιπόν έγινε ένα «γκρουπάρισμα» των tweets ανά author και δημιουργήθηκαν οι unique authors δηλαδή οι μοναδικοί αυτοί χρήστες που έγραψαν για τις περιπτώσεις των Τεμπών και της πυρκαγιάς στο Μάτι αντίστοιχα.

## «Μάτι»

Στην περίπτωση των πυρκαγιών στο Μάτι θα χρησιμοποιηθούν 289.714 τιμές αναφοράς, όπως προκύπτει από το `mati_III.csv`. Στο παρακάτω στιγμιότυπο (εικόνα 25) βλέπουμε πως οργανώνονται τα tweets ο συνολικός δηλαδή όγκος. Το `author volume` των πυρκαγιών δομείται με την ύπαρξη χρηστών που έχουν κάνει 1473 tweets αλλά και πολύ λιγότερα σύμφωνα με τα αποτελέσματα της `author volume df`. Τα 289.714 tweets βρέθηκε πως τα δημιούργησαν 30.000 unique authors σύμφωνα με το plot της εικόνας 25.



The screenshot shows a Jupyter Notebook cell with the following content:

```

289172      2804434455      2018-07-23      1021434672889331712      el
16:39:38+00:00

289173 rows x 12 columns

author_volume_df
[3] ✓ 0.0s
...

```

	author_id	author_volume
0	2349489651	1473
1	3173348733	1329
2	3300398932	1076
3	305738208	1045
4	1298248224	1031
...	...	...
30611	2926161501	1
30612	2926578735	1
30613	2926778105	1
30614	2927444029	1
30615	1641447243940265993	1

30616 rows x 2 columns

Cell 4

Εικόνα 25: Προσδιορισμός unique authors  
author\_volume\_df

Σε αντίθεση οι πρώτοι 50 φαίνεται να έχουν αρκετά tweets ενώ π.χ. οι 25.000 στους 30.000 έχουν πολύ μικρό αριθμό tweets, παρατηρείται δηλαδή μια αναντιστοιχία που θέλουμε να ερευνηθεί.

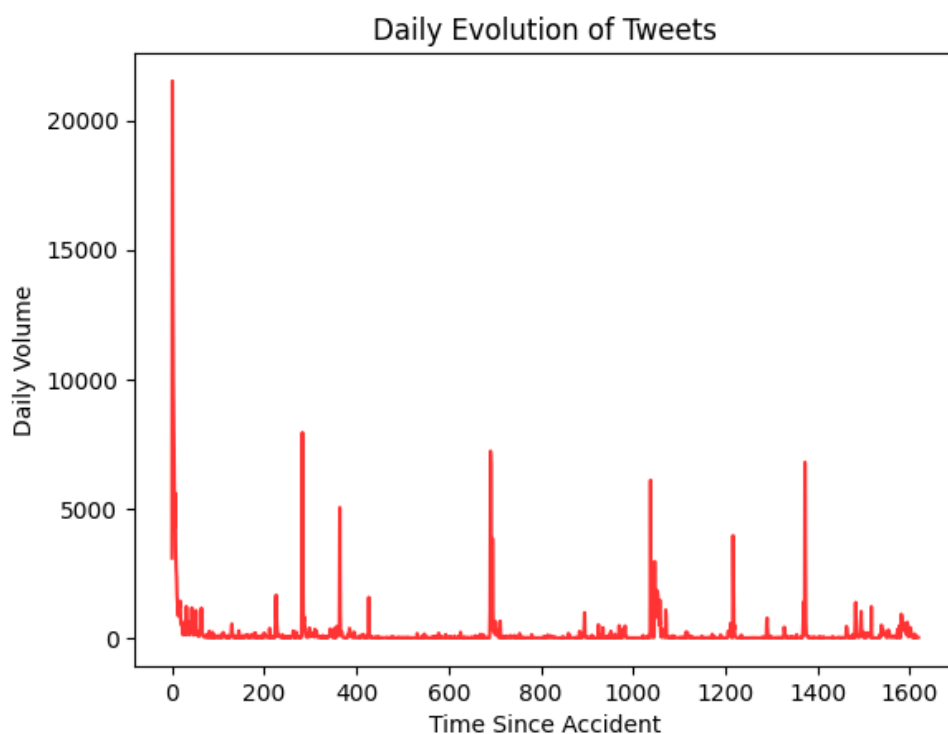
	author_id	author_volume
0	2349489651	1473
1	3173348733	1329
2	3300398932	1076
3	305738208	1045
4	1298248224	1031
...	...	...
24995	51715874	1
24996	51736822	1
24997	51839330	1
24998	48379790	1
24999	48329698	1

25000 rows × 2 columns

Εικόνα 26: Δείγμα με author\_volume\_df για τους 25.000 unique authors

Με βάση το παρακάτω γράφημα (Εικόνα 27) που δείχνει το daily volume των tweets, επιδιώκεται να δούμε πως αυτό θα αλλάξει μορφή αν περιοριστεί σε ένα υποσύνολο χρηστών που έχουν κάνει παραπάνω από ένα αριθμό tweets. Αν αυτές οι μικρές διαταραχές που παρατηρούμε στο παρακάτω γράφημα εξαφανίζονταν και εμφανίζονταν σε διακριτά σημεία πιθανόν να μας γνωστοποιήσαν αν αποτελούν αυτό το υποσύνολο που προσπαθεί ίσως να υποκινήσει το φαινόμενο και κατ' επέκταση να δημιουργήσει και τον όγκο των υπόλοιπων χρηστών. Επιδιώκεται δηλαδή να βρεθεί αν υπάρχει υποσύνολο χρηστών εντός του dataset το οποίο υποκινεί τους υπόλοιπους ή απλώς «ξαναζωντανεύει» το φαινόμενο και το παίρνουν και οι υπόλοιποι χρήστες. Έτσι οι κορυφές του γραφήματος στο Μάτι αποτυπώνονται σε μεγάλα χρονικά διαστήματα και υποδεικνύουν πότε αναμοχλεύεται το ζήτημα μέσα στο twitter.

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων



Εικόνα 27: Αρχικό Διάγραμμα Daily evolution of tweets

Στο πρώτο plot (εικόνα 27) οπτικοποιείται ο ημερήσιος όγκος αναφορών με το Keyword “mati”. Στον κάθετο άξονα βλέπουμε πως εξελίσσεται ο ημερήσιος όγκος αναφορών στο twitter σχετικά με την περίπτωση στο Μάτι. Παρατηρούμε μια εκτόξευση κυρίως από την στιγμή που συνέβη και για τις επόμενες 5 μέρες. Βλέπουμε πως με την πάροδο του χρόνου, η οποία παρουσιάζεται στον οριζόντιο άξονα στις 10 με 15 μέρες υπάρχει αρκετά αισθητή πτώση των αναφορών με παρουσιάσεις «αναζωπυρώσεων» της παρουσίας του γεγονότος στο twitter.

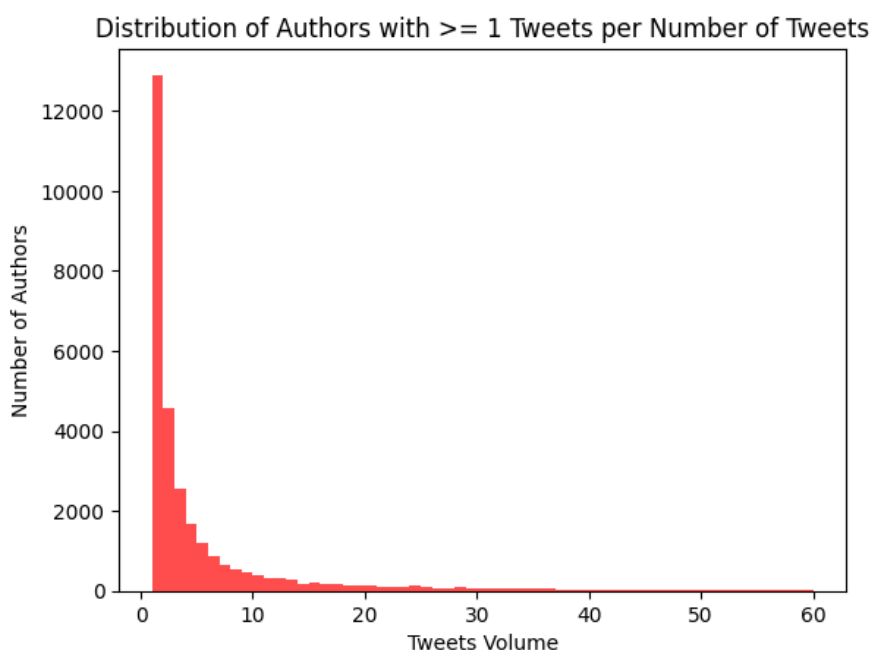
Στο δεύτερο plot της περίπτωσης Μάτι (Εικόνα 29) βλέπουμε την οπτικοποίηση των αναρτήσεων στο σύνολο των author ID'S που έχουν κάνει τουλάχιστον 1 tweet και υπάρχουν εντός του dataset. Αποτελεί την βάση για το τρίτο plot που ακολουθεί και εστιάζει στο πως θα

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων



συμπεριφερθεί το διάγραμμα εφόσον περιορίζεται το υποσύνολο χρηστών που έχουν κάνει τουλάχιστον  $x$  tweets. Υπάρχουν αρκετές διαταραχές που όπως και καθιστούν να γίνουν διακριτά σημεία που πιθανόν να αποτελούν αυτό το υποσύνολο που προσπαθεί ίσως να υποκινήσει το φαινόμενο.

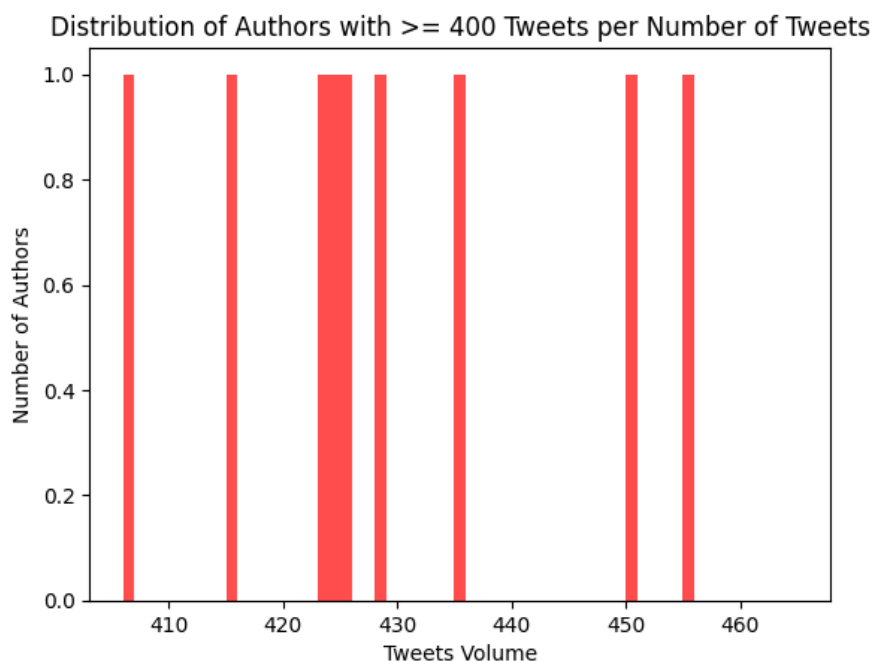
Εικόνα 28: Διάγραμμα Author ID να έχει τουλάχιστον 1 tweet στο dataset



Στο τρίτο αυτό Plot (Εικόνα 29) της περίπτωσης Μάτι βλέπουμε πως άλλαξε μορφή το γράφημα τώρα εφόσον περιορίστηκε υποσύνολο χρηστών που έχουν κάνει παραπάνω από 400 tweets. Οι μικρές διαταραχές που παρατηρούνταν στο προηγούμενο γράφημα φαίνεται να μην υπάρχουν πλέον και έχουν κάνει την εμφάνισή τους εμφανώς πιο διακριτά σημεία που πιθανόν να αποτελούν αυτό το υποσύνολο που προσπαθεί ίσως να υποκινήσει το φαινόμενο και κατ' επέκταση να δημιουργήσει και τον όγκο των υπόλοιπων χρηστών.

Τα αποτελέσματα είναι αξιοθαύμαστα καθώς το γράφημα άλλαξε εντελώς μορφή εφόσον περιορίστηκε υποσύνολο χρηστών που έχουν κάνει παραπάνω από 400 tweets. Οι μικρές διαταραχές που

παρατηρούνταν στο προηγούμενο γράφημα φαίνεται να μην υπάρχουν πλέον και έχουν κάνει την εμφάνισή τους εμφανώς διακριτά σημεία που πιθανόν να αποτελούν αυτό το υποσύνολο που προσπαθεί ίσως να υποκινήσει το φαινόμενο και κατ' επέκταση να δημιουργήσει και τον όγκο των υπόλοιπων χρηστών



Εικόνα 29: Διάγραμμα Author ID να έχει τουλάχιστον 400 tweets στο dataset

## «Τέμνη»

Στην περίπτωση των Τεμπών χρησιμοποιήθηκαν 1.190.784 τιμές αναφοράς, τα οποία πάλι «γκρουπάrouμε» ανά author και προκύπτουν οι unique authors που είναι 62.559. Δημιουργήθηκε δηλαδή ένα πλήθος από 1.190.784 tweets τα οποία όμως δημιουργήθηκαν από 62.559 authors και στο παρακάτω dataframe βλέπουμε πόσα tweets έχει κάνει ο καθένας από αυτούς.



The screenshot shows a Jupyter Notebook cell with the following content:

```
author_volume_df
[2] ✓ 0.0s
...
author_id  author_volume
0  1582379745085382657    3622
1    275106146            3605
2  1369581015178612737    3394
3  1492453528102739973    3065
4    179423854            2718
...
62555  910213192860815362         1
62556  910192689295056896         1
62557  909902048061161472         1
62558  909807474936606720         1
62559  1643722675720798216         1
62560 rows x 2 columns
```

Εικόνα 30: Προσδιορισμός unique authors author\_volume\_df

Όπως θα δούμε στα παρακάτω στιγμιότυπα (εικόνα 29) μέχρι και τις 20.000 authors έχει δημιουργηθεί 1 tweet από τον καθένα μετά τις 30.000 authors αρχίζουμε να βλέπουμε κάποια αύξηση στον αριθμό των tweets. Από τους 60.000 unique authors οι τελευταίοι 30.000 έχουν κάνει 1 και 2 tweets, στους 40.000 ανεβαίνει στο 5, στους 50.000 έχουν κάνει 14 tweets, στους 55.000 έχουμε 27 tweets, στους 60.000 είναι 89 tweets και αντίστοιχα οι πρώτοι 100 φτάνουν και στα 913 tweets έκαστος.

Εικόνα 31: Unique authors to Author volume 40.000 – 60.000

```
author_volume_df[: -40000]
```

✓ 0.0s

	author_id	author_volume
0	1582379745085382657	3622
1	275106146	3605
2	1369581015178612737	3394
3	1492453528102739973	3065
4	179423854	2718
...	...	...
22555	820430644606095360	5
22556	3008181904	5
22557	731675021681274880	5
22558	3021619477	5
22559	1435344111083692035	5

22560 rows × 2 columns

```
author_volume_df[: -60000]
```

✓ 0.0s

	author_id	author_volume
0	1582379745085382657	3622
1	275106146	3605
2	1369581015178612737	3394
3	1492453528102739973	3065
4	179423854	2718
...	...	...
2555	716321046366969858	89
2556	1513430995126697984	89
2557	356069137	89
2558	827939598965862402	89
2559	2359267094	89

2560 rows × 2 columns

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

Εικόνα 32: Unique authors to Author volume 50-100

The image contains two screenshots of a data viewer interface. The left screenshot shows a table with 100 rows and 2 columns, titled 'author\_volume\_df[:100]'. The right screenshot shows a table with 19 rows and 2 columns, titled 'author\_volume\_df[:50]'. Both tables have columns 'author\_id' and 'author\_volume'.

	author_id	author_volume
0	1582379745085382657	3622
1	275106146	3605
2	1369581015178612737	3394
3	1492453528102739973	3065
4	179423854	2718
...	...	...
95	1376867015059775488	918
96	1238907915932774401	917
97	1507045992557862918	915
98	601787871	913
99	1282744796935868422	913

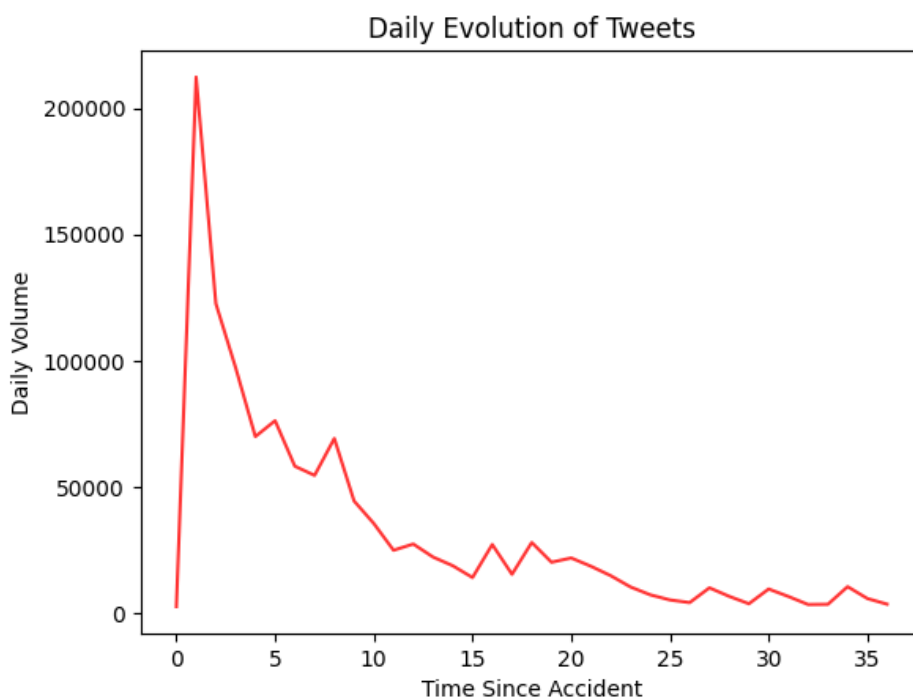
  

	author_id	author_volume
0	1582379745085382657	3622
1	275106146	3605
2	1369581015178612737	3394
3	1492453528102739973	3065
4	179423854	2718
5	1031398897	2679
6	1306580615890956289	2669
7	1358730673134182400	2493
8	1494721186974076941	2446
9	2998767646	2223
10	1872223814	2219
11	1233069623463030784	2095
12	1484509739484655616	2075
13	3347335019	2067
14	1588473165923196929	1963
15	883326229037953024	1961
16	577431005	1960
17	237719164	1944
18	1244628291404824578	1929

Στα παραπάνω στιγμιότυπα (Εικόνα 30) βλέπουμε πως υπάρχει μεγάλη μάζα ανθρώπων που έχουν κάνει 1 tweet και μερίδα ανθρώπων γύρω στους 100 που έχουν κάνει περίπου 1000 tweets, ενώ η μεγάλη εκτόξευση tweets γίνεται στους πρώτους 50 να κυμαίνονται από 1330 – 3622 tweets. Αυτοί οι 50 δεν ξέρουμε που ακριβώς βρίσκονται στο γράφημα daily volume of tweets προφανώς θα είναι στην αρχή του κατακόρυφου άξονα και πάρα πολύ δεξιά θα αναπαρίστανταν πάρα πολύ κοντά.

Ερχόμενοι πάλι σε επαφή με το αρχικό Plot της οπτικοποίησης δεδομένων του twitter αναφορικά με τα Τέμπη βλέπουμε παρακάτω το daily volume των tweets, επιδιώκεται να δούμε πως αυτό θα αλλάξει

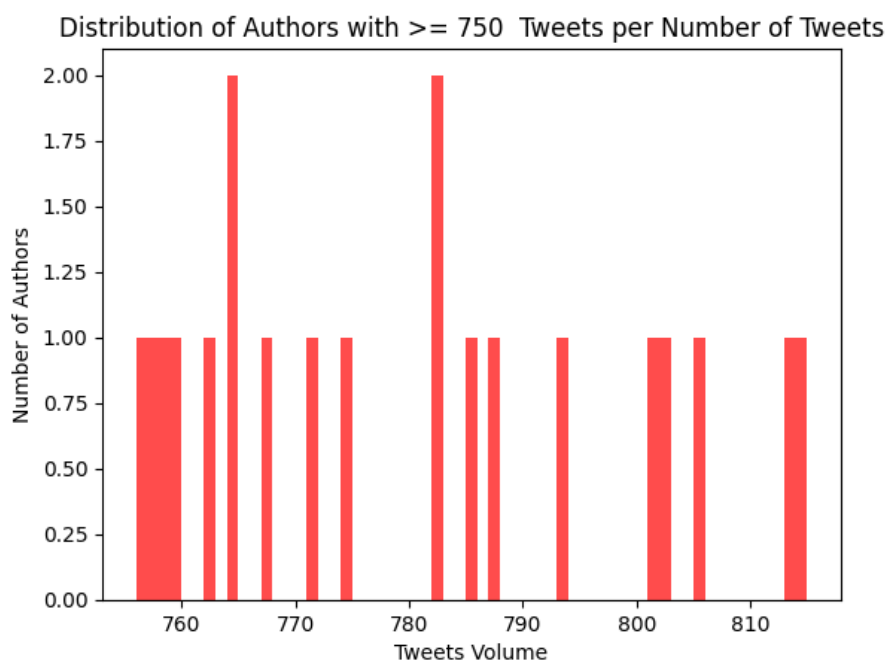
μορφή αν περιοριστεί σε ένα υποσύνολο χρηστών που έχουν κάνει παραπάνω από ένα αριθμό tweets.



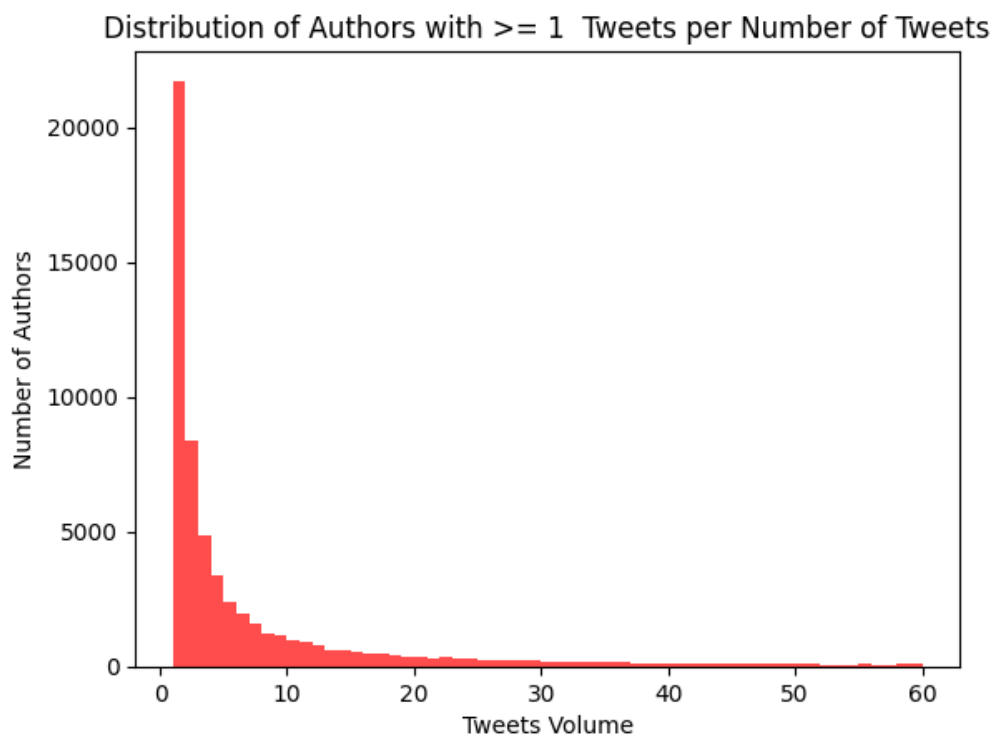
Εικόνα 33: Αρχικό Διάγραμμα Daily evolution of tweets

Όπως στην περίπτωση στο Μάτι έτσι και στα Τέμπη το αρχικό daily volume των tweets χρησιμοποιείται ως η βάση για να δούμε πως αυτό θα αλλάξει μορφή αν περιοριστεί σε ένα υποσύνολο χρηστών που έχουν κάνει παραπάνω από ένα αριθμό tweets. Σε αυτή την περίπτωση δεν έχουμε τόσες μικρές διαταραχές υπάρχει μια περισσότερη διακριτότητα ωστόσο δεν θεωρείται επαρκής προκειμένου να αποσαφηνιστεί αν οι authors με τα πιο πολλά tweets αποτελούν αυτό το υποσύνολο που προσπαθεί ίσως να υποκινήσει το φαινόμενο και κατ' επέκταση να δημιουργήσει και τον όγκο των υπόλοιπων χρηστών. Προκειμένου να βρεθεί αν υπάρχει σύνολο χρηστών που αναζωπυρώνει το φαινόμενο χρειάζεται να φιλτράρουμε λίγο το αρχικό αυτό διάγραμμα.

Εικόνα 34: authors με tweets ανά χρήστη  $\geq 750$



Εικόνα 35: authors με tweets ανά χρήστη  $\geq 1$



Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

Παίρνοντας το υποσύνολο των author ID'S το ημερήσιο volume των αναρτήσεων ορίζεται ως εξής όταν το σύνολο των χρηστών αν περιοριστεί σε εκείνους που έχουν κάνει παραπάνω από ένα αριθμό tweets, στην προκειμένη περίπτωση ορίστηκε το 750.

Τα αποτελέσματα είναι αξιοθαύμαστα καθώς το γράφημα άλλαξε εντελώς μορφή εφόσον περιορίστηκε υποσύνολο χρηστών που έχουν κάνει παραπάνω από 750 tweets. Οι μικρές διαταραχές που παρατηρούνταν στο προηγούμενο γράφημα φαίνεται να μην υπάρχουν πλέον και έχουν κάνει την εμφάνισή τους εμφανώς διακριτά σημεία που πιθανόν να αποτελούν αυτό το υποσύνολο που προσπαθεί ίσως να υποκινήσει το φαινόμενο και κατ' επέκταση να δημιουργήσει και τον όγκο των υπόλοιπων χρηστών

#### **4.5.2 Αποτελέσματα Τέμνη LSTM :**

Προτού αναλύσουμε τα αποτελέσματα του LSTM για τις δυο μελέτες περίπτωσης θα αποσαφηνίσουμε κάποιους όρους:

- Το RMSE (Root Mean Squared Error) είναι μια μετρική που χρησιμοποιείται για να αξιολογήσει την ακρίβεια ενός μοντέλου πρόβλεψης. Πρόκειται για την τετραγωνική ρίζα του μέσου τετραγωνικού σφάλματος και δίνει μια ενδεικτική αξία της απόκλισης μεταξύ των προβλεπόμενων και των πραγματικών τιμών. [46]
- Το MAE (Mean Absolute Error) είναι μια άλλη μετρική που αξιολογεί το σφάλμα μεταξύ των προβλεπόμενων και πραγματικών τιμών, υπολογίζοντας τον μέσο όρο των απόλυτων τιμών των διαφορών.



- Όταν αναφερόμαστε σε "εκπαίδευση" (train), αναφερόμαστε στα δεδομένα που χρησιμοποιήθηκαν για την εκπαίδευση του μοντέλου, δηλαδή για την προσαρμογή του στα δεδομένα.
- Το "δοκιμή" (test) αναφέρεται σε ένα σύνολο δεδομένων που δεν χρησιμοποιήθηκε κατά την εκπαίδευση και χρησιμεύει για την αξιολόγηση της απόδοσης του μοντέλου σε νέα, άγνωστα δεδομένα. Σε αυτήν την περίπτωση, το μοντέλο εμφανίζει χαμηλότερο σφάλμα στα δεδομένα εκπαίδευσης σε σύγκριση με τα δεδομένα δοκιμής, γεγονός που είναι συνηθισμένο στα μοντέλα πρόβλεψης

Έτσι φτάνοντας στο τελευταίο ζητούμενο προς διερεύνηση αξιολογούμε την απόδοση του μοντέλου πρόβλεψης LSTM:

Για τα δεδομένα εκπαίδευσης, το μοντέλο μας παρουσίασε εξαιρετικές επιδόσεις, με το RMSE να φτάνει το 0.73 και το MAE το 0.54. Αυτά τα αποτελέσματα υποδεικνύουν ότι το μοντέλο μας κατάφερε να εκπαιδευτεί πολύ καλά στα δεδομένα εκπαίδευσης και μπορεί να αναπαράγει τα pattern των χρονοσειρών με σχετικά χαμηλό σφάλμα.

Εικόνα 36: Επιδόσεις μοντέλου εκπαίδευσης Τέμπη

```

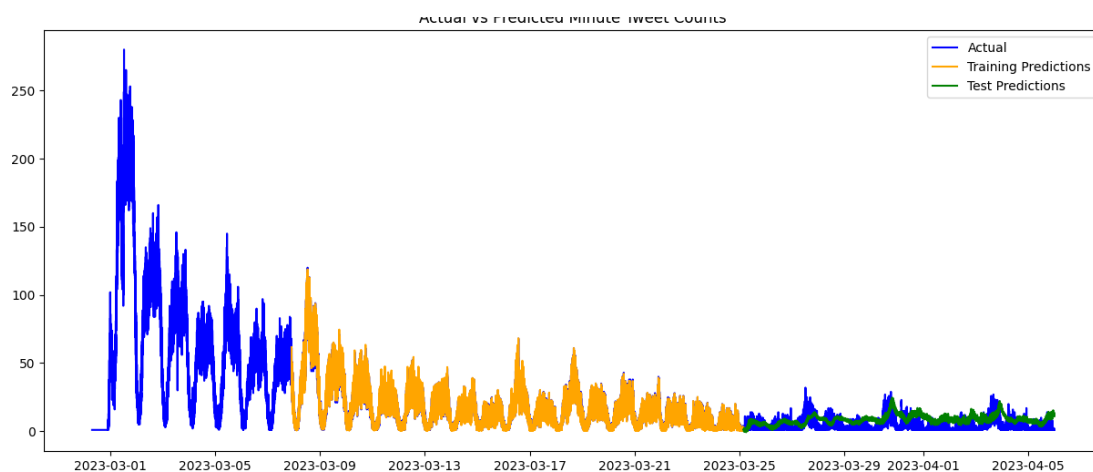
... Epoch 1/500
122/122 [=====] - 11s 67ms/step - loss: 7.7113e-04
Epoch 2/500
122/122 [=====] - 8s 64ms/step - loss: 3.7709e-04
Epoch 3/500
122/122 [=====] - 8s 65ms/step - loss: 3.6265e-04
Epoch 4/500
122/122 [=====] - 8s 64ms/step - loss: 3.4681e-04
Epoch 5/500
122/122 [=====] - 8s 67ms/step - loss: 3.5163e-04
Epoch 6/500
122/122 [=====] - 8s 64ms/step - loss: 3.3256e-04
Epoch 7/500
122/122 [=====] - 8s 65ms/step - loss: 3.1994e-04
Epoch 8/500
122/122 [=====] - 8s 69ms/step - loss: 3.0614e-04
Epoch 9/500
122/122 [=====] - 9s 72ms/step - loss: 3.1285e-04
Epoch 10/500
122/122 [=====] - 8s 66ms/step - loss: 3.0635e-04
Epoch 11/500
122/122 [=====] - 8s 64ms/step - loss: 3.0342e-04
Epoch 12/500
122/122 [=====] - 8s 65ms/step - loss: 3.0534e-04
Epoch 13/500
...
Train RMSE: 0.73
Test RMSE: 5.86
Train MAE: 0.54
Test MAE: 4.87

```

Ωστόσο, για τα δεδομένα δοκιμής, τα αποτελέσματα δεν είναι τόσο εντυπωσιακά. Το RMSE για τα δεδομένα δοκιμής ανέρχεται σε 5.86 και το MAE στο 4.87. Αυτό υποδεικνύει ότι, αν και το μοντέλο εκπαιδεύτηκε καλά στα δεδομένα εκπαίδευσης, δυσκολεύεται να γενικεύσει αποτελεσματικά σε νέα, αδιάθετα δεδομένα. Ένας πιθανός λόγος για αυτήν τη διαφορά στην απόδοση μπορεί να είναι ότι, όπως αναφέρετε, με το πέρασμα των ημερών μειώνονται τα tweets και αλλάζει το pattern. Αυτό υποδεικνύει ότι οι χρονοσειρές δεν είναι σταθερές, αλλά εξελίσσονται με τον καιρό.

Εάν το μοντέλο δεν έχει εκπαιδευτεί σε παρόμοιες αλλαγές στο παρελθόν, θα βρεί δύσκολο να προβλέψει με ακρίβεια αυτές τις νέες συμπεριφορές. Αυτό υπογραμμίζει την σημασία της διαρκούς αναθεώρησης και προσαρμογής των μοντέλων χρονοσειρών, ειδικά όταν παρατηρείται μια εξελικτική φύση στα δεδομένα.

Εικόνα 37: Εκπαιδευμένο μοντέλο περίπτωσης Τέμπη.



Με βάση το παραπάνω plot που αποτελεί και την πρόβλεψη που δημιουργείται για την περίπτωση των Τεμπών βλέπουμε την γενικότερη φιλοσοφία εφαρμογής του LSTM στα δεδομένα του twitter που χρησιμοποιήθηκαν.

Όσο αφορά το dataset δομήθηκε και υποδιαιρέθηκε σε train test όπου προκειμένου να εκπαιδύσουμε το μοντέλο βαθιάς μάθησης (LSTM) σε χρονοσειριακά δεδομένα επιλέχθηκαν κάποιες μέρες και βάλαμε μοντέλο να κοιτάξει μια εβδομάδα πίσω πόσα tweets είχαν γίνει και με βάση αριθμό αυτό προβλέπει επόμενα. Δηλαδή το feed των δεδομένων για την περίπτωση των Τεμπών ξεκινάει στις 28 Φεβρουαρίου η εκπαίδευση ξεκινάει μια εβδομάδα μετά στις 07 Μαρτίου 2023 και λαμβάνει υπόψιν την εβδομάδα που πέρασε, ουσιαστικά το διάστημα της μιας εβδομάδας αποτελεί το διάστημα που το δίκτυο «μαθαίνει» γι' αυτό και το train ξεκινάει από τότε και έπειτα. Τέλος χρησιμοποιήθηκαν 500 εποχές για τη εκπαίδευση του μοντέλου κάτι που σημαίνει ότι το μοντέλο εκπαιδεύτηκε τόσες φορές προκειμένου να επιτευχθεί αυτή η απόδοση.

Αποσαφήνιση χρωμάτων:

- Πορτοκαλί χρώμα μοντέλου = Training predictions είναι το σετ που έχουμε εκπαιδεύσει το μοντέλο.
- Μπλε χρώμα μοντέλου = Τα απτά δεδομένα του twitter.
- Πορτοκαλί πάνω σε μπλε στα δεξιά το τεστ ελέγχει πόσο καλά εκπαιδεύσαμε το σετ.
- Η μπλε σκιά πίσω από πορτοκαλί και πράσινο δείχνει πόσο πολύ κοντά είναι οι καμπύλες του μοντέλου κάτι που σηματοδοτεί πως έχουμε επιτύχει καλές προβλέψεις.

#### **4.6 Αποτελέσματα Μάτι LSTM:**

Τα αποτελέσματα του μοντέλου παρουσιάζονται ως εξής:

- Σφάλμα RMSE εκπαίδευσης: 0.42
- Σφάλμα RMSE δοκιμής: 0.64
- Σφάλμα MAE εκπαίδευσης: 0.28
- Σφάλμα MAE δοκιμής: 0.40

```

...
Epoch 1/500
122/122 [=====] - 11s 67ms/step - loss: 7.7113e-04
Epoch 2/500
122/122 [=====] - 8s 64ms/step - loss: 3.7709e-04
Epoch 3/500
122/122 [=====] - 8s 65ms/step - loss: 3.6265e-04
Epoch 4/500
122/122 [=====] - 8s 64ms/step - loss: 3.4681e-04
Epoch 5/500
122/122 [=====] - 8s 67ms/step - loss: 3.5163e-04
Epoch 6/500
122/122 [=====] - 8s 64ms/step - loss: 3.3256e-04
Epoch 7/500
122/122 [=====] - 8s 65ms/step - loss: 3.1994e-04
Epoch 8/500
122/122 [=====] - 8s 69ms/step - loss: 3.0614e-04
Epoch 9/500
122/122 [=====] - 9s 72ms/step - loss: 3.1285e-04
Epoch 10/500
122/122 [=====] - 8s 66ms/step - loss: 3.0635e-04
Epoch 11/500
122/122 [=====] - 8s 64ms/step - loss: 3.0342e-04
Epoch 12/500
122/122 [=====] - 8s 65ms/step - loss: 3.0534e-04
Epoch 13/500
...
Train RMSE: 0.73
Test RMSE: 5.86
Train MAE: 0.54
Test MAE: 4.87

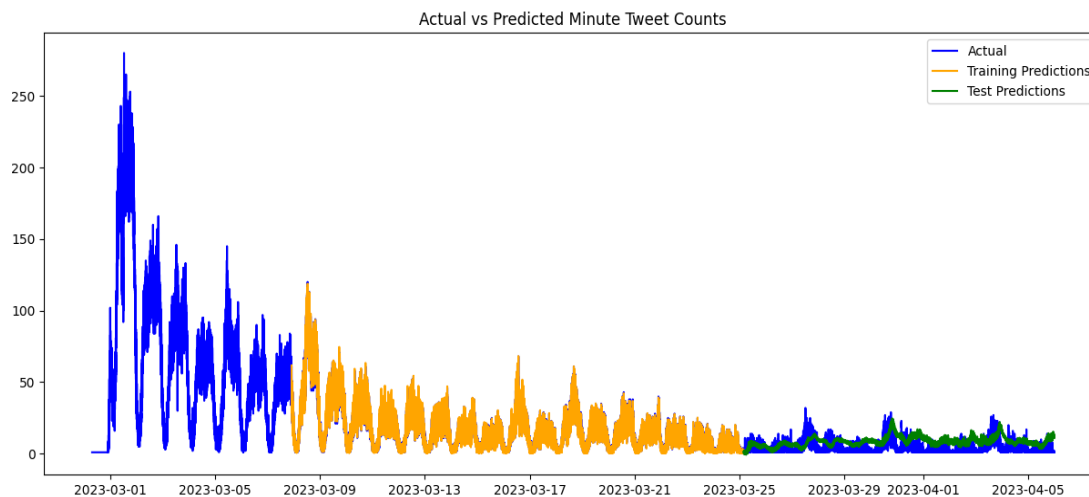
```

Εικόνα 38: Επιδόσεις μοντέλου εκπαίδευσης Τέμπη

Σε αυτήν την περίπτωση, το μοντέλο εμφανίζει χαμηλότερο σφάλμα στα δεδομένα εκπαίδευσης σε σύγκριση με τα δεδομένα δοκιμής, γεγονός που είναι συνηθισμένο στα μοντέλα πρόβλεψης. Συγκριτικά με την περίπτωση των Τεμπών τόσο τα δεδομένα δοκιμής όσο και τα αποτελέσματα είναι πολύ καλά. Το RMSE για τα δεδομένα δοκιμής ανέρχεται σε 0.73 και το MAE στο 4.87. Αυτό υποδεικνύει ότι, αν και το μοντέλο εκπαιδεύτηκε καλά στα δεδομένα εκπαίδευσης, δυσκολεύεται να γενικεύσει αποτελεσματικά σε νέα, αδιάθετα δεδομένα. Ένας πιθανός λόγος για αυτήν τη διαφορά στην απόδοση μπορεί να είναι ότι, όπως αναφέρετε, με το πέρασμα των ημερών μειώνονται τα tweets και αλλάζει το pattern. Αυτό υποδεικνύει ότι οι χρονοσειρές δεν είναι σταθερές, αλλά εξελίσσονται με τον καιρό. Εάν το μοντέλο δεν έχει εκπαιδευτεί σε παρόμοιες αλλαγές στο παρελθόν, θα βρεί δύσκολο να προβλέψει με ακρίβεια αυτές τις νέες συμπεριφορές. Αυτό υπογραμμίζει την σημασία της διαρκούς αναθεώρησης και προσαρμογής των μοντέλων χρονοσειρών, ειδικά όταν παρατηρείται μια εξελικτική φύση στα δεδομένα.

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

Εικόνα 39: Εκπαιδευμένο μοντέλο περίπτωσης Μάτι



#### Αποσαφήνιση χρωμάτων:

- Πορτοκαλί χρώμα μοντέλου = Training predictions είναι το σετ που έχουμε εκπαιδεύσει το μοντέλο.
- Μπλε χρώμα μοντέλου = Τα απτά δεδομένα του twitter.
- Πορτοκαλί πάνω σε μπλε στα δεξιά το τεστ ελέγχει πόσο καλά εκπαιδεύσαμε το σετ.

Η μπλε σκιά πίσω από πορτοκαλί και πράσινο δείχνει πόσο πολύ κοντά είναι οι καμπύλες του μοντέλου κάτι που σηματοδοτεί πως έχουμε επιτύχει καλές προβλέψεις

## 5. Συμπεράσματα – μελλοντική εργασία

Η παρούσα διπλωματική εργασία αποτέλεσε μια προσπάθεια πρόβλεψης δεδομένων των κοινωνικών δικτύων και συγκεκριμένα του twitter με χρήση νευρωνικού δικτύου LSTM αλλά των χρονοσειρών. Έτσι επιτευχθήκαν με αρκετά μεγάλο ποσοστό επιτυχίας η οπτικοποίηση του ημερησίου volume αναρτήσεων όταν σύνολο χρηστών περιοριστεί σε εκείνους που έχουν κάνει παραπάνω από έναν αριθμό tweets, η Πρόβλεψη του αριθμού των tweets που δημοσιεύονται κάθε λεπτό βασισμένο στα δεδομένα των προηγούμενων 24 ωρών αλλά και η πρόβλεψη του αριθμού των tweets ανά λεπτό βασισμένο σε ιστορικά δεδομένα, χρησιμοποιώντας το μοντέλο βαθιάς μάθησης που ονομάζεται LSTM (Long Short-Term Memory).

Σκοπός της προσέγγισης αυτής δεν ήταν η πρόγνωση των μελλοντικών tweets αλλά η αξιολόγηση της ακρίβειας των μοντέλων σε συγκεκριμένα χρονικά διαστήματα που αποτέλεσαν κομβικά σημεία για την χώρα μας. Οι μελέτες περίπτωσης αφορούσαν την θλιβερή πυρκαγιά στο μάτι το 2008 αλλά και το σιδηροδρομικό δυστύχημα των Τεμπών το 2023.

Οι δοκιμές αναφορικά με την εύρεση του όγκου αναρτήσεων των authors εντός του dataset και κατά πόσο κάποιος ή όχι έχουν μεγαλύτερη δυναμική επιρροή στην «επαναπυροδότηση» του φαινομένου θα λέγαμε πως ήταν αρκετά επιτυχείς. Όλες αυτές οι ανισοκατανομές που συναντήσαμε μας οδήγησαν στο να θεωρήσουμε πως υπάρχει αυτό το «σκοτεινό» σημείο που πρέπει να ερευνηθεί. Έτσι με την μεθοδολογία που ακολουθήθηκε δόθηκε μια καλύτερη εικόνα της κατεύθυνσης των διαταραχών του αρχικού γραφήματος που κατά προσέγγιση εμπεριέχουν αυτό το υποσύνολο που προσπαθεί ίσως να υποκινήσει το φαινόμενο και κατ' επέκταση να δημιουργήσει και τον όγκο των υπόλοιπων χρηστών..

Όσο αφορά τα αποτελέσματα εκπαίδευσης του LSTM ήταν εξαιρετικά ειδικά στην περίπτωση των Τεμπών με το RMSE να φτάνει το 0.73 και το MAE το 0.54. Η πρόβλεψη του νευρωνικού δικτύου είχε ως παράμετρο 500 εποχές κάτι που μας οδήγησε και σε ένα τόσο καλό αποτέλεσμα.

Αυτά τα αποτελέσματα υποδεικνύουν ότι το μοντέλο μας κατάφερε να εκπαιδευτεί πολύ καλά στα δεδομένα εκπαίδευσης και μπορεί να αναπαράγει τα pattern των χρονοσειρών με σχετικά χαμηλό σφάλμα.

Στην περίπτωση στο Μάτι το μοντέλο εμφάνισε χαμηλότερο σφάλμα στα δεδομένα εκπαίδευσης σε σύγκριση με τα δεδομένα δοκιμής, γεγονός που είναι συνηθισμένο στα μοντέλα πρόβλεψης. Ωστόσο, για τα δεδομένα δοκιμής, τα αποτελέσματα δεν είναι τόσο εντυπωσιακά όσο τις πρόβλεψης. Το RMSE για τα δεδομένα δοκιμής ανέρχεται σε 5.86 και το MAE στο 4.87. Αυτό υποδεικνύει ότι, αν και το μοντέλο εκπαιδεύτηκε καλά στα δεδομένα εκπαίδευσης, δυσκολεύεται να γενικεύσει αποτελεσματικά σε νέα, αδιάθετα δεδομένα. Συγκριτικά με την περίπτωση των Τεμπών τόσο τα δεδομένα δοκιμής όσο και τα αποτελέσματα είναι πολύ καλά.

Η έρευνα που πραγματοποιήθηκε σίγουρα διαθέτει περιθώρια βελτιστοποίησης. Η προσθήκη περισσότερων εποχών στην εκπαίδευση του μοντέλου είναι καθοριστικής σημασίας και θεωρώ πως με έναν εξοπλισμό ακόμα μεγαλύτερης υπολογιστικής ισχύος θα ήταν δυνατή η αύξησή τους. Τέλος κρίνεται σημαντική η μελλοντική σύγκριση διάφορων άλλων παραλλαγών LSTM, όπως οι Peehole Connections, Full Gradient.



## 6. Βιβλιογραφική αναφορά

1. [https://link.springer.com/referenceworkentry/10.1007/978-0-387-32833-1\\_401](https://link.springer.com/referenceworkentry/10.1007/978-0-387-32833-1_401)
2. <https://blog.quantinsti.com/stationarity/#:~:text=Stationary%20Time%20Series%20and%20Non%2DStationary%20Time%20Series,-A%20time%20series&text=A%20stationary%20time%20series%20has,to%20the%20long%2Drun%20mean.&text=A%20time%20series%20whose%20statistical%20properties%20change%20over%20time,a%20non%2Dstationary%20time%20series.>
3. <https://www.simplilearn.com/tutorials/statistics-tutorial/what-is-time-series-analysis>
4. <https://www.sciencedirect.com/topics/social-sciences/social-network-analysis>
5. <https://www.timescale.com/blog/what-is-time-series-analysis-with-examples-and-applications/>
6. <https://analyticsindiamag.com/a-beginners-guide-to-characteristics-of-a-time-series/>
7. <https://www.tableau.com/learn/articles/time-series-analysis>
8. <https://otexts.com/fpp2/tspatterns.html#fig:fourexamples>
9. [https://www.datacamp.com/tutorial/tutorial-for-recurrent-neural-network?utm\\_source=google&utm\\_medium=paid\\_search&utm\\_campaignid=19589720818&utm\\_adgroupid=143216588777&utm\\_device=c&utm\\_keyword=&utm\\_matchtype=&utm\\_network=g&utm\\_adpostion=&utm\\_creative=671350460558&utm\\_targetid=aud-299261629574:dsa-1947282172981&utm\\_loc\\_interest\\_ms=&utm\\_loc\\_physical\\_ms=20361&utm\\_content=dsa~page~community-tuto&utm\\_campaign=230119\\_1-sea~dsa~tofu-tutorials\\_2-b2c\\_3-eu\\_4-prc\\_5-na\\_6-na\\_7-le\\_8-pdsh-go\\_9-na\\_10-na\\_11-na-sep23&gclid=CjwKCAjwsKqoBhBPEiwALrrqiLy9ul9NJD5qfAcQPWXFllw6W-icdDDMhBEAgMazC3JrkVv2jhntoxoCilUQAvD\\_BwE](https://www.datacamp.com/tutorial/tutorial-for-recurrent-neural-network?utm_source=google&utm_medium=paid_search&utm_campaignid=19589720818&utm_adgroupid=143216588777&utm_device=c&utm_keyword=&utm_matchtype=&utm_network=g&utm_adpostion=&utm_creative=671350460558&utm_targetid=aud-299261629574:dsa-1947282172981&utm_loc_interest_ms=&utm_loc_physical_ms=20361&utm_content=dsa~page~community-tuto&utm_campaign=230119_1-sea~dsa~tofu-tutorials_2-b2c_3-eu_4-prc_5-na_6-na_7-le_8-pdsh-go_9-na_10-na_11-na-sep23&gclid=CjwKCAjwsKqoBhBPEiwALrrqiLy9ul9NJD5qfAcQPWXFllw6W-icdDDMhBEAgMazC3JrkVv2jhntoxoCilUQAvD_BwE)
10. <https://www.geeksforgeeks.org/machine-learning/>
11. <https://www.analyticsvidhya.com/blog/2021/03/inV.Stacked%20Autoencoder%20production-to-long-short-term-memory-lstm/>
12. <https://vitalflux.com/deep-neural-network-examples-from-real-life/>
13. <https://www.frontiersin.org/articles/10.3389/fdata.2019.00002/full>

14. <https://medium.com/analytics-vidhya/time-series-forecasting-a-complete-guide-d963142da33f>
15. <https://buffer.com/library/social-media-sites/#9-telegram-550-million-maus>
16. <https://www.kdnuggets.com/2022/02/vanishing-gradient-problem.html>
17. <https://www.timescale.com/blog/how-to-work-with-time-series-in-python/>
18. <https://www.mygreatlearning.com/blog/what-is-machine-learning/>
19. Λαμπριανάκη, Χριστίνα. "Μελέτη μεθόδων ανάλυσης κοινωνικών δικτύων στο πεδίο της τεχνολογικά υποστηριζόμενης μάθησης" University of Piraeus (Greece), 2022
20. Δημανόπουλος, Μιχαήλ. "Γραμμική παλινδρόμηση για την πρόβλεψη της ανόδου ή πτώσης του χρηματιστηρίου", University of Piraeus , 2022.
21. <https://towardsdatascience.com/time-series-forecasting-with-deep-learning-and-attention-mechanism-2d001fc871fc>
22. [https://www.tensorflow.org/tutorials/structured\\_data/time\\_series](https://www.tensorflow.org/tutorials/structured_data/time_series)
23. Deep learning CNN–LSTM framework for Arabic sentiment analysis using textual information shared in social networks: Abubakr H. Ombabi
24. <https://www.mathworks.com/help/deeplearning/gs/neural-network-time-series-prediction-and-modeling.html>
25. <https://saturncloud.io/blog/example-of-time-series-prediction-using-neural-networks-in-r/>
26. State-of-the-art in artificial neural network applications: Oludare Isaac Abiodun
27. Neural Networks in Civil Engineering: 1989–2000 Hojjat Adeli
28. How effective are neural networks at forecasting and prediction? A review and evaluation: Monica Adya
29. A Review of Recurrent Neural Networks: LSTM Cells and Network Architectures: Yong Yu.
30. <https://blog.dataiku.com/deep-learning-time-series-forecasting>
31. <https://towardsdatascience.com/a-quick-deep-learning-recipe-time-series-forecasting-with-keras-in-python-f759923ba64>
32. [https://www.sas.com/en\\_gb/insights/articles/analytics/a-guide-to-predictive-analytics-and-machine-learning.html](https://www.sas.com/en_gb/insights/articles/analytics/a-guide-to-predictive-analytics-and-machine-learning.html)
33. <https://towardsdatascience.com/classification-regression-and-prediction-whats-the-difference-5423d9efe4ec>
34. <https://towardsdatascience.com/mapping-with-matplotlib-pandas-geopandas-and-basemap-in-python-d11b57ab5dac>

Χρήση ανατροφοδοτούμενου Νευρωνικού δικτύου και χρονοσειρών για την πρόβλεψη δεδομένων

35. [https://www.datacamp.com/tutorial/tutorial-for-recurrent-neural-network?utm\\_source=google&utm\\_medium=paid\\_search&utm\\_campaignid=19589720818&utm\\_adgroupid=143216588777&utm\\_device=c&utm\\_keyword=&utm\\_matchtype=&utm\\_network=g&utm\\_adpostion=&utm\\_creative=671350460558&utm\\_targetid=aud-299261629574:dsa-1947282172981&utm\\_loc\\_interest\\_ms=&utm\\_loc\\_physical\\_ms=20361&utm\\_content=dsa~page~community-tuto&utm\\_campaign=230119\\_1-sea~dsa~tofu-tutorials\\_2-b2c\\_3-eu\\_4-prc\\_5-na\\_6-na\\_7-le\\_8-pdsh-go\\_9-na\\_10-na\\_11-na-sep23&gclid=CjwKCAjwsKqoBhBPEiwALrrqiLy9ul9NJD5qfAcQPWXFlw6W-jcdDDMhBEAgMazC3JrkVv2jhntoxoCilUQAvD\\_BwE](https://www.datacamp.com/tutorial/tutorial-for-recurrent-neural-network?utm_source=google&utm_medium=paid_search&utm_campaignid=19589720818&utm_adgroupid=143216588777&utm_device=c&utm_keyword=&utm_matchtype=&utm_network=g&utm_adpostion=&utm_creative=671350460558&utm_targetid=aud-299261629574:dsa-1947282172981&utm_loc_interest_ms=&utm_loc_physical_ms=20361&utm_content=dsa~page~community-tuto&utm_campaign=230119_1-sea~dsa~tofu-tutorials_2-b2c_3-eu_4-prc_5-na_6-na_7-le_8-pdsh-go_9-na_10-na_11-na-sep23&gclid=CjwKCAjwsKqoBhBPEiwALrrqiLy9ul9NJD5qfAcQPWXFlw6W-jcdDDMhBEAgMazC3JrkVv2jhntoxoCilUQAvD_BwE)
36. <https://medium.datadriveninvestor.com/step-by-step-time-series-analysis-d2f117554d7e?gi=41db5df45ea3>
37. <https://www.sciencedirect.com/science/article/abs/pii/S0169207020300996>
38. [https://www.geeksforgeeks.org/data-structures/?ref=shm\\_outind](https://www.geeksforgeeks.org/data-structures/?ref=shm_outind)
39. [https://www.datacamp.com/tutorial/tutorial-for-recurrent-neural-network?utm\\_source=google&utm\\_medium=paid\\_search&utm\\_campaignid=19589720818&utm\\_adgroupid=143216588777&utm\\_device=c&utm\\_keyword=&utm\\_matchtype=&utm\\_network=g&utm\\_adpostion=&utm\\_creative=671350460558&utm\\_targetid=aud-299261629574:dsa-1947282172981&utm\\_loc\\_interest\\_ms=&utm\\_loc\\_physical\\_ms=20361&utm\\_content=dsa~page~community-tuto&utm\\_campaign=230119\\_1-sea~dsa~tofu-tutorials\\_2-b2c\\_3-eu\\_4-prc\\_5-na\\_6-na\\_7-le\\_8-pdsh-go\\_9-na\\_10-na\\_11-na-sep23&gclid=CjwKCAjwsKqoBhBPEiwALrrqiLy9ul9NJD5qfAcQPWXFlw6W-jcdDDMhBEAgMazC3JrkVv2jhntoxoCilUQAvD\\_BwE](https://www.datacamp.com/tutorial/tutorial-for-recurrent-neural-network?utm_source=google&utm_medium=paid_search&utm_campaignid=19589720818&utm_adgroupid=143216588777&utm_device=c&utm_keyword=&utm_matchtype=&utm_network=g&utm_adpostion=&utm_creative=671350460558&utm_targetid=aud-299261629574:dsa-1947282172981&utm_loc_interest_ms=&utm_loc_physical_ms=20361&utm_content=dsa~page~community-tuto&utm_campaign=230119_1-sea~dsa~tofu-tutorials_2-b2c_3-eu_4-prc_5-na_6-na_7-le_8-pdsh-go_9-na_10-na_11-na-sep23&gclid=CjwKCAjwsKqoBhBPEiwALrrqiLy9ul9NJD5qfAcQPWXFlw6W-jcdDDMhBEAgMazC3JrkVv2jhntoxoCilUQAvD_BwE)
40. <https://towardsdatascience.com/predicting-stock-prices-using-a-keras-lstm-model-4225457f0233>
41. <https://www.neuraldesigner.com/blog/auto-associative-neural-networks>