ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Αυτόματη επιλογή μοντέλου πρόβλεψης για χρονοσειρές και εφαρμογή σε δεδομένα από καταστήματα τραπεζών

Αικατερίνη Λασκαράτου

Επιβλέπων καθηγητής: Βασιλική Καντερέ
Επίκουρη Καθηγήτρια Ε.Μ.Π.

Αθήνα, Φεβρουάριος 2023

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΤΕΧΝΟΛΟΓΙΑΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΥΠΟΛΟΓΙΣΤΩΝ

Αυτόματη επιλογή μοντέλου πρόβλεψης για χρονοσειρές και εφαρμογή σε δεδομένα από καταστήματα τραπεζών

Αικατερίνη Λασκαράτου

Επιβλέπων καθηγητής: Βασιλική Καντερέ
Επίκουρη Καθηγήτρια Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 22η Φεβρουαρίου 2023.

.....	.....	.....
Βασιλική Καντερέ	Γεώργιος Γκούμας	Δημήτριος Τσουμάκος
Επίκουρη Καθηγήτρια	Αναπληρωτής Καθηγητής	Αναπληρωτής Καθηγητής
ΣΗΜΜΥ Ε.Μ.Π	ΣΗΜΜΥ Ε.Μ.Π	ΣΗΜΜΥ Ε.Μ.Π

Αθήνα, 2023

.....
Αικατερίνη Λασκαράτου

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

Περίληψη

Τα τελευταία χρόνια παρατηρείται μία ραγδαία ανάπτυξη του Internet of Things, ενός δικτύου 'έξυπνων αντικειμένων' που έχουν την ιδιότητα να κατανοούν και να αντιδρούν στις μεταβολές του περιβάλλοντος. Είναι ιδιαίτερα σημαντικό τόσο για την οικονομία όσο και την κοινωνία, καθώς επιτρέπει την επικοινωνία μεταξύ των ανθρώπων, την επικοινωνία ανθρώπου με αντικείμενα και την επικοινωνία αντικειμένων με άλλα αντικείμενα. Ο κόσμος γίνεται όλο και πιο διασυνδεδεμένος και 'έξυπνος¨.

Τα 'έξυπνα' αντικείμενα προσλάμβανουν την πληροφορία από το περιβάλλον σε μορφή χρονοσειρών μέσα από τους αισθητήρες που διαθέτουν, και στη συνέχεια η πληροφορία αυτή (δεδομένα) υπόκειται σε ανάλυση. Η σωστή λειτουργία τους βασίζεται στην σωστή επεξεργασία και ανάλυση αυτών των χρονοσειρών.

Στην παρούσα εργασία μελετήθηκαν δεδομένα χρονοσειρών που συλλέχθηκαν από αισθητήρες εγκατεστημένους σε αντικείμενα που ήταν τοποθετημένα σε καταστήματα τραπεζών. Τα δεδομένα αφορούν τη θερμοκρασία ορισμένων δωματίων, την ενεργειακή κατανάλωση, την σύσταση διοξειδίου του άνθρακα και υγρασίας στην ατμόσφαιρα κα. Ακολούθησε επεξεργασία των δεδομένων και στη συνέχεια εφαρμόστηκαν ορισμένοι αλγόριθμοι πρόβλεψης ώστε να μελετηθεί η μελλοντική συμπεριφορά των χρονοσειρών. Στη συνέχεια, με βάση τα σφάλματα που προέκυψαν μετά την εφαρμογή των μοντέλων, βγήκαν ορισμένα συμπεράσματα σχετικά με το πόσο ακριβή στην πρόβλεψή τους είναι τα μοντέλα σε σχέση με τη μορφή των χρονοσειρών.

Κύριος στόχος της εργασίας είναι η μελέτη της συμπεριφοράς των μοντέλων πρόβλεψης ως προς τα διαφορετικά σύνολα δεδομένων και η αυτόματη επιλογή του κατάλληλου μοντέλου όταν είναι γνωστή η μορφή της χρονοσειράς. Σε αντίθετη περίπτωση, ο έλεγχος όλων των αλγορίθμων είναι ιδιαίτερα χρονοβόρος και μη πρακτικός.

Εκτός από τα μεμονωμένα μοντέλα, εφαρμόστηκαν τρεις μέθοδοι ensemble που συνδυάζουν τις ιδιότητες των αλγορίθμων και προέκυψαν συγκεκριμένα συμπεράσματα.

Λέξεις Κλειδιά

Πρόβλεψη Χρονοσειρών, Νευρωνικά Δίκτυα, Εκθετική εξομάλυνση, ARIMA, SARIMA, CNN, LSTM, GRU, Bidirectional LSTM, Random Forest, XGBoost, Prophet, τάση, εποχικότητα, Διαδίκτυο των Πραγμάτων

Abstract

In recent years there has been a rapid development of the Internet of Things, a network of "smart objects" that have the ability to understand and react to changes of the environment. It is particularly important for both the economy and society as it enables human-to-human communication, human-to-object communication and object-to-object communication. The world is becoming increasingly interconnected and "smart".

First, "smart" objects receive information from the environment in the form of time series through their sensors and then this information (data) is subjected to analysis. Their proper functioning is based on the correct processing and analysis of these time series. In this paper, time series data collected from sensors installed on objects placed in bank branches were studied. The data concern the temperature of some rooms, energy consumption, carbon dioxide and humidity composition in the atmosphere, etc. The data were processed and then some forecasting algorithms were applied to study the future behaviour of the time series. Then, based on the errors obtained after the application of the models, some conclusions were drawn on how accurate the models are in their prediction in relation to the shape of the time series.

The main objective of this paper is to study the behaviour of the forecasting models with respect to different data sets and to automatically select the appropriate model when the time series format is known. Otherwise, testing all algorithms is very time consuming and impractical.

Apart from the individual models, three ensemble methods combining the properties of the algorithms were applied and some conclusions were drawn.

Key Words

Time series, Forecasting, Neural Networks, Exponential Smoothing, ARIMA, SARIMA, CNN, LSTM, GRU, Bidirectional LSTM, Random Forest, XGBoost, Prophet, trend, seasonality, Internet of Things

Ευχαριστίες

Ευχαριστώ θερμά την καθηγήτρια κυρία Βασιλική Καντερέ, η οποία μου έδωσε τη δυνατότητα να ασχοληθώ με αυτό το ενδιαφέρον θέμα και να επεκτείνω τις γνώσεις μου. Επιπλέον, ιδιαίτερες ευχαριστίες οφείλω στον υποψήφιο διδάκτορα κύριο Πάρη Κερασιώτη για την διαθεσιμότητά του και την καθοδήγησή του καθ'όλη την διάρκεια του ακαδημαϊκού έτους, και το χρόνο που διέθετε για την επίλυση όποιων απορειών προέκυπταν. Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου για την στήριξη που μου παρέχει όλα αυτά τα χρόνια καθώς και τους φίλους που χωρίς αυτούς δεν θα μπορούσα να φανταστώ τα φοιτητικά μου χρόνια.

Περιεχόμενα

1 Internet of Things

Η έννοια του Internet of Things εμφανίστηκε για πρώτη φορά το 1999 από τον Βρετανό Kevin Ashton και περιέγραφε ένα σύστημα από αντικείμενα που μπορούν να συνδεθούν στο διαδίκτυο μέσω αισθητήρων. Σήμερα, το Διαδίκτυο των Πραγμάτων έχει εξελιχθεί σε ένα δημοφιλή όρο που περιγράφει ένα 'Ανοιχτό και ολοκληρωμένο δίκτυο έξυπνων αντικειμένων που έχουν την ικανότητα να οργανώνονται αυτόματα, να μοιράζονται πληροφορίες, δεδομένα και πόρους, να αντιδρούν και να ενεργούν απέναντι σε καταστάσεις και αλλαγές στο περιβάλλον¨. Το Διαδίκτυο των Πραγμάτων επιτρέπει την ανίχνευση και τον έλεγχο των αντικειμένων εξ αποστάσεως, στην υπάρχουσα υποδομή δικτύου, δημιουργώντας ευκαιρίες για πιο άμεση ενσωμάτωση του φυσικού κόσμου σε υπολογιστικά συστήματα με βελτιωμένη απόδοση και ακρίβεια. [1]
Οι λέξεις 'Διαδίκτυο' και 'Πράγματα' δηλώνουν ένα διασυνδεδεμένο παγκόσμιο δίκτυο που βασίζεται σε τεχνολογίες αισθητηριακών, επικοινωνιακών, δικτυακών εργασιών και επεξεργασίας πληροφοριών [1]. Είναι ένα δίκτυο που επιτρέπει την επικοινωνία ανθρώπου με άνθρωπο, ανθρώπου με αντικείμενα και αντικειμένων με αντικείμενα, παρέχοντας μία μοναδική ταυτότητα στα αντικείμενα.[5] Η τεχνική της εκχώρησης μοναδικής ταυτότητας σε ένα αντικείμενο ονομάζεται 'καθολικό μοναδικό αναγνωριστικό' (UUID) και είναι απαραίτητο για την επιτυχημένη ανάπτυξη υπηρεσιών σε ένα τεράστιο δίκτυο όπως το IoT. Τα αναγνωριστικά μπορεί να αναφέρονται σε ονόματα και διευθύνσεις.[2]
Τα αντικείμενα του IoT μπορούν να περιλαμβάνουν όχι μόνο ηλεκτρονικές συσκευές και τεχνολογικά προηγμένα προϊόντα, αλλά και αυτά που χρησιμοποιούνται καθημερινά, όπως είναι τρόφιμα, ρούχα, έπιπλα, ακόμα και κτήρια. Εκατομμύρια, ακόμη και δισεκατομμύρια πράγματα μπορούν να ενσωματωθούν απρόσκοπτα και αποτελεσματικά, με αποτέλεσμα το IoT να μπορεί να εφαρμοστεί ευρέως σε πολλούς τομείς.[2]
Στα παραπάνω αντικείμενα είναι ενσωματωμένοι αισθητήρες και ενεργοποιητές που συνδέονται μεταξύ τους μέσω ενσύρματων και ασύρματων δικτύων. Οι συσκευές μπορούν να αισθανθούν από το περιβάλλον και να επικοινωνήσουν με αυτό. Οι πληροφορίες που συλλέγονται προωθούνται στη συνέχεια για ανάλυση. Αυτό που είναι ιδιαίτερα επαναστατικό είναι ότι αυτά τα συστήματα πληροφοριών αρχίζουν τώρα να αναπτύσσονται και μερικά από αυτά να λειτουργούν, ακόμη και σε μεγάλο βαθμό, χωρίς την ανθρώπινη παρέμβαση.[5]
Οι αναδυόμενες ασύρματες αισθητηριακές τεχνολογίες έχουν επεκτείνει σημαντικά τις αισθητηριακές δυνατότητες των συσκευών και ως εκ τούτου η αρχική ιδέα του IoT επεκτείνεται και στο περιβάλλον της τεχνιτής νοημοσύνη και στον αυτόματο έλεγχο. Μέχρι σήμερα, ένας μεγάλος αριθμός από τεχνολογίες εμπλέκονται στο IoT, όπως ο ασύρματος αισθητήρας, δίκτυα, γραμμωτοί κώδικες, έξυπνη ανίχνευση, RFID, NFC, ασύρματες επικοινωνίες χαμηλής ενέργειας, cloud computing και πολλές άλλες.[1]
Η τεχνολογική πρόοδος φέρνει τεχνολογικές προόδους και στο Internet of Things, το οποίο θα αποτελέσει τη 'νέα γενιά' του Internet, στο οποίο θα είναι εφικτή η πρόσβαση στα φυσικά αντικείμενα μέσα από το διαδίκτυο [1]. Το IoT επεκτίνεται ταχύτατα με τις εκτιμήσεις να αναφέρουν ότι θα έχουν παγκόσμια οικονομική επίπτωση έως και 11,1 τρισεκατομμύρια δολάρια ετησίως μέχρι το 2025.[3]
Μια κρίσιμη απαίτηση του είναι η διαλειτουργικότητα μεταξύ ετερογενών συσκευών, ενώ επιπλέον η αρχιτεκτονική συστημάτων του πρέπει να συνδέει τον φυσικό με τον εικονικό κόσμο. Ακόμα, λόγω του

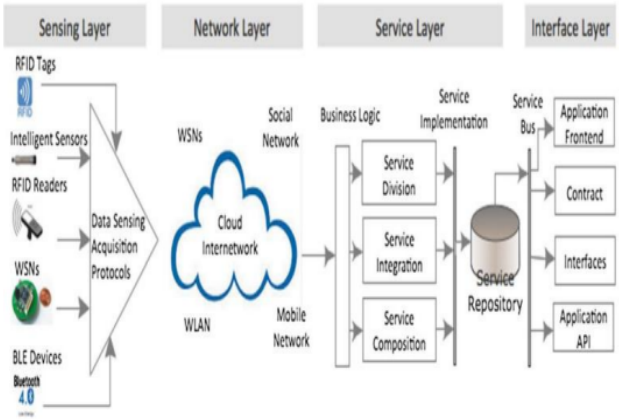
γεγονότος ότι τα αντικείμενα μπορεί να κινούνται γεωγραφικά και να χρειάζεται να αλληλεπιδρούν σε πραγματικό χρόνο, η αρχιτεκτονική του θα πρέπει να είναι προσαρμοστική και να καθιστά τις συσκευές ικανές να αλληλεπιδρούν με άλλα αντικείμενα δυναμικά.[1]
Το δίκτυο στο οποίο συνδέονται οι συσκευές αποτελείται από ορισμένα επίπεδα με διακεκριμένες λειτουργίες, επιτρέποντας τελικά την απαιτούμενη διαλειτουργικότητα μεταξύ των συσκευών:[2]

- Το επίπεδο ανίχνευσης στο οποίο είναι ενσωματωμένα όλα τα διαθέσιμα αντικείμενα για την αίσθηση της κατάστασής τους.[1]
  Λόγω του μεγάλου αριθμού των αισθητήρων σε πολύπλοκες εφαρμογές συστημάτων, οι έξυπνες συσκευές θα πρέπει να σχεδιάζονται ώστε να ελαχιστοποιούν τους απαιτούμενους πόρους και το κόστος. Επιπλέον, για τον σχεδιασμό του επιπέδου ανίχνευσης είναι σημαντικό να ληφθεί υπόψη το γεγονός ότι το IoT είναι ετερογενές, ότι τα δεδομένα που θα προκύψουν πρέπει να είναι προσβάσιμα και ανακτίσιμα και ότι τα αντικείμενα είναι οργανωμένα σε δίκτυα. Τέλος, οι αισθητήρες οφείλουν να διαθέτουν υψηλή ενεργειακή απόδοση αφού είναι συνεχώς ενεργοί, ώστε να αποκτούν δεδομένα σε πραγματικό χρόνο και να υποστηρίζουν τη συνύπαρξη διαφορετικών επικοινωνιών όπως είναι τα WLAN, ZigBee και Bluetooth.[1]

- Το επίπεδο δικτύου, για την υποστήριξη των ασύρματων ή ενσύρματων συνδέσεων μεταξύ των πραγμάτων.[1]
  Το επίπεδο δικτύου παρέχει στα αντικείμενα τη δυνατότητα να μοιράζονται τα δεδομένα με τα υπόλοιπα συνδεδεμένα αντικείμενα. Αυτό είναι απαραίτητο για την έξυπνη διαχείρηση και την επεξεργασία των δεδομένων στο IoT. Για να λειτουργήσει ομαλά και σωστά όμως, είναι σημαντικό να αντιμετωπιστούν ζητήματα όπως είναι η ενεργειακή απόδοση του δικτύου, οι τεχνολογίες εξόρυξης, επεξεργασίας και αναζήτησης δεδομένων και σημάτων, η ασφάλεια και το απόρρητο.[1]

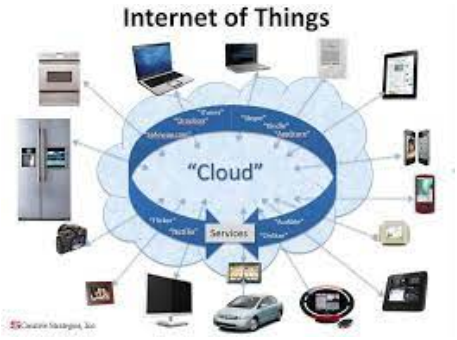
- Το επίπεδο υπηρεσιών για τη δημιουργία και τη διαχείρηση των υπηρεσιών που απαιτούνται.[1]
  Ένα πρακτικό επίπεδο υπηρεσιών αποτελείται από ένα ελάχιστο σύνολο κοινών απαιτήσεων των εφαρμογών, των διασυνδέσεων προγραμματισμού εφαρμογών (API) και των πρωτοκόλλων που υποστηρίζουν τις απαιτούμενες εφαρμογές και υπηρεσίες. Όλες οι δραστηριότητες που αφορούν τις υπηρεσίες, όπως είναι η ανταλλαγή πληροφοριών και η αποθήκευση, η διαχείριση δεδομένων, η βάση δεδομένων,οι μηχανές αναζήτησης και η επικοινωνία, πραγματοποιούνται στο επίπεδο υπηρεσιών. [1]

- Το επίπεδο διεπαφής το οποίο αποτελείται από τις απαραίτητες μεθόδους ή εφαρμογές για την αλληλεπίδραση των αντικειμένων με τους χρήστες.[1]
  Γενικά, εμπλέκεται μεγάλος αριθμός συσκευών που παρέχονται από διαφορετικούς προμηθευτές, και επομένως δεν ακολουθούν πάντα τα ίδια πρότυπα. Το ζήτημα της συμβατότητας μεταξύ των ετερογενών αντικειμένων αντιμετωπίζεται με έναν αποτελεσματικό μηχανισμό διεπαφής, ώστε να επιτρέπεται η ανταλλαγή, η επικοινωνία και η επεξεργασία συμβάντων(events).[1]



Σχήμα 1: Τα επίπεδα IOT[1]

Τα τελευταία χρόνια όλο και περισσότερο αυξάνονται οι 'έξυπνες' συσκευές, με αποτέλεσμα να απαιτείται τεχνολογία που να μπορεί να αναλύσει και να αποθηκεύσει τα δεδομένα με αποτελεσματικό τρόπο. Αυτή η τεχνολογία είναι το cloud, μία πλατφόρμα προσβάσιμη οποιαδήποτε στιγμή και από οποιοδήποτε μέρος, που επιτρέπει την κοινή χρήση πόρων μεταξύ των αντικειμένων. Το cloud computing αποτελεί ένα σημαντικό μέρος του IoT αφού παρέχει τη δυνατότητα αυξημένης ισχύος επεξεργασίας των δεδομένων που λαμβάνονται από τους αισθητήρες και σχετικά καλή αποθηκευτική χωρητικότητα. Επίσης, καθιστά εφικτή την επέκταση του Internet of Things και την ανάπτυξή του σε μεγάλη κλίμακα, διασυνδέοντας τις συσκευές με εκατομύρια αισθητήρες. [6]



Σχήμα 2: Cloud and Internet of Things[4]

Στην παρούσα εργασία οι μετρήσεις που χρησιμοποιήθηκαν προέκυψαν από ένα σύνολο συσκευών τοποθετημένων σε καταστήματα τραπεζών, αποτελούμενα από συστήματα αισθητήρων. Οι αισθητήρες αυτοί έχουν την ικανότητα παρακολούθησης φυσικών ή περιβαλλοντικών συνθηκών, όπως είναι η θερμοκρασία, ο ήχος, οι κραδασμοί, η πίεση, η κίνηση ή οι ρύποι. Τέτοια συστήματα αισθητήρων χρησιμοποιούνται συχνά και σε άλλους τομείς όπως είναι ο στρατός, η υγειονομική περίθαλψη και η ανίχνευση δασικών πυρκαγιών.[1]
Εν κατακλείδει, το IoT επιτρέπει τη συλλογή, την αποθήκευση και τη μετάδοση πληροφοριών σε αντικείμενα τα οποία είναι εξοπλισμένα με αισθητήρες. 'Έξυπνα αντικείμενα' έχουν χρησιμοποιηθεί ευρέως για την περιβαλλοντική παρακολούθηση, το λιανικό εμπόριο, την υγειονομική περίθαλψη, τη βιομηχανία τροφίμων και εστιατορίων, την ταξιδιωτική και τουριστική βιομηχανία και πολλούς άλλους τομείς. [7]
Είναι φανερό ότι το IoT είναι πολύ σημαντικό για την οικονομία και την κοινωνία. Θα συμβάλει σημαντικά στην αντιμετώπιση κοινωνικών ζητημάτων όπως είναι η παρακολούθηση της υγειονομικής περίθαλψης, η παρακολούθηση της καθημερινής ζωής και ο έλεγχος της κυκλοφοριακής συμφόρησης. Μέσω του Διαδικτύου των Πραγμάτων, ο κόσμος γίνεται όλο και πιο διασυνδεδεμένος και ευφυής. Δημιουργούνται τεράστιες ποσότητες δεδομένων, ενώ το κόστος αποθήκευσης μειώνεται.[3]
Για να πραγματοποιηθούν όμως όλα τα παραπάνω και να αξιοποιηθεί στο μέγιστο το Διαδίκτυο των Πραγμάτων, είναι απαραίτητη η σωστή λειτουργία των συσκευών, με αρχικό βήμα την σωστή ανάλυση και πρόβλεψη των δεδομένων εισόδου (χρονοσειρών) από τους αισθητήρες.

2 Εισαγωγή στις Χρονοσειρές

Οι χρονοσειρές αποτελούν μια συλλογή από παρατηρήσεις που έχουν συλλεχθεί με βάση τη χρονολογική τους σειρά.[8] Το σύνολο των στατιστικών μεθόδων για την ανάλυση των χρονοσειρών αναφέρεται ως ‘ time series analysis ¨. Η πρόβλεψη των χρονοσειρών είναι η διαδικασία κατά την οποία αναλύονται τα δεδομένα τους, με σκοπό την ανάπτυξη μοντέλου που περιγράφει την υποκειμενική σχέση μεταξύ τους και, τελικά, με βάση αυτό το μοντέλο, την προέκταση της χρονοσειράς στο μέλλον.[9]
Η σωστή ανάλυση και οι σωστές προβλέψεις είναι πολύ σημαντικές σε πολλούς τομείς της επιστήμης, της βιομηχανικής, της εμπορικής και της οικονομικής δραστηριότητας, καθώς, με βάση αυτές, οι υπεύθυνοι για τη λήψη αποφάσεων εκτιμούν μελλοντικές εμφανίσεις αστοχιών συστήματος για τον προγραμματισμό πόρων, τη διαχείριση αποθεμάτων και την ανάπτυξη ρεαλιστικών πολιτικών.[10] Το αποτέλεσμα μίας ενέργειας, δεν μπορεί να είναι γνωστό πριν από την πραγματοποίησή της, με ακρίβεια, αλλά η διαδικασία της πρόβλεψης προσθέτει μία σημαντική πληροφορία για αυτό, μειώνοντας το ρίσκο της απόφασης πραγματοποίησης της.
Τα δεδομένα μιας χρονοσειράς μεταβάλλονται σε χρονολογική σειρά κατά τη διάρκεια μίας περιόδου. Εάν οι εγγραφές ενσωματώνουν περισσότερα από ένα χαρακτηριστικά ή μεταβλητές, η σειρά ονομάζεται πολυμεταβλητή χρονοσειρά. Σε αντίθετη περίπτωση, ονομάζεται μονομεταβλητή. Επιπλέον, οι χρονοσειρές μπορούν να διαχωριστούν σε συνεχείς και διακριτές. Στις συνεχείς, η παρατήρηση των δεδομένων γίνεται σε συνεχή χρόνο κατά τη διάρκεια μιας περιόδου, ενώ στις διακριτές, η παρατήρηση των γεγονότων πραγματοποιείται σε συγκεκριμένους χρόνους ή σε χρόνους που απέχουν ίσες αποστάσεις μεταξύ τους. [11]
Για τη βαθύτερη κατανόηση των χρονοσειρών είναι σημαντικό να αναφερθούν ορισμένα από τα χαρακτηριστικά τους.

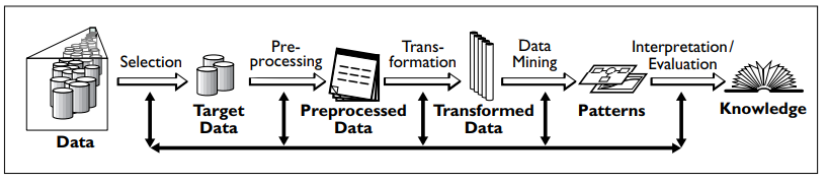
- **Trend-Τάση**
  Τάση είναι ένα μοτίβο που παρατηρείται σε μία χρονική περίοδο και εκφράζει τον μέσο ρυθμό μεταβολής της χρονοσειράς σε σχέση με το χρόνο. Δηλώνει την τάση των δεδομένων μακροπρόθεσμα για αύξηση ή μείωση. Ανοδική τάση υποδηλώνει αύξηση, ενώ πτωτική τάση μείωση. Δεν είναι απαραίτητο ότι θα ακολουθείται αυξητική ή μειωτική τάση σε όλη τη διάρκεια της παρατήρησης. Οι χρονοσειρές στις οποίες η συνάρτηση της τάσης παραμένει σταθερή με το χρόνο ονομάζονται στατικές.[11]
  Σε πολλές περιπτώσεις, το ενδιαφέρον επικεντρώνεται στην κατανόηση της εξελισσόμενης τάσης και την πρόβλεψή της, καθώς η πρόβλεψη σε συγκεκριμένα σημεία δεδομένων μπορεί να προσφέρει μικρό πλήθος πληροφοριών σχετικά με τη σημασιολογία και τη δυναμική της υποκείμενης διαδικασίας, οι παρατηρήσεις της οποίας σχηματίζουν τη χρονοσειρά. [12]

- **Seasonality-Εποχικότητα**
  Εποχικότητα είναι ένα άλλο χαρακτηριστικό των χρονοσειρών. Εκφράζει μία περιοδική διακύμανση όπου το ίδιο μοτίβο εμφανίζεται σε ένα χρονικό διάστημα.[11] Για παράδειγμα, εποχικότητα για μηνιαίες τιμές εμφανίζεται όταν παρατηρήσεις με διαφορά δώδεκα μηνών συσχετίζονται με κάποιον τρόπο. [13]

Η εποχικότητα μπορεί να είναι προσθετική (additive) ή πολλαπλασιαστική (multiplicative). Προσθετική χαρακτηρίζεται όταν η σειρά εμφανίζει σταθερές εποχιακές διακυμάνσεις, ανεξάρτητα από το συνολικό επίπεδό της. Στην πολλαπλασιαστική, από την άλλη, το μέγεθος των εποχιακών διακυμάνσεων ποικίλλει, ανάλογα με το συνολικό επίπεδο της σειράς. [10]

3 Ανακάλυψη γνώσης σε βάσεις δεδομένων

Η ανακάλυψη γνώσης σε βάσεις δεδομένων (Knowledge Discovery Database/KDD) ορίστηκε αρχικά, το 1991, ως η εξαγωγή προηγουμένως άγνωστων και δυνητικά χρήσιμων πληροφοριών από δεδομένα. Το 1996 ο ορισμός της εξελίχθηκε στη διαδικασία αναγνώρισης έγκυρων, πρωτότυπων, δυνητικά χρήσιμων και τελικά κατανοητών μοτίβων ή γνώσεων στα δεδομένα. Η διαδικασία αυτή περιλαμβάνει πολλές μεθόδους. [15]



Σχήμα 3: Knowledge Discovery Database[16]

3.1 Προεπεξεργασία δεδομένων

Οι χρονοσειρές είναι ένα σύνολο από δεδομένα τοποθετημένα σε χρονολογική σειρά. Αναφέρθηκε σε προηγούμενο κεφάλαιο η σημασία σωστής ανάλυσης και πρόβλεψής τους. Πριν όμως από τις διαδικασίες αυτές, είναι απαραίτητη η επεξεργασία τους.
Η προεπεξεργασία δεδομένων χρονοσειράς περιλαμβάνει τα ακόλουθα βήματα:[17]

- Απομάκρυνση του θορύβου από τα δεδομένα, με τεχνικές που χρησιμοποιούνται στην επεξεργασία σημάτων [17]
- Καθορισμός ενός συνόλου από χαρακτηριστικά που περιγράφουν τα δεδομένα [17]
- Χωρισμός της χρονικής κλίμακας σε διαστήματα, ώστε σε κάθε διάστημα να καθορίζεται και μία συνάρτηση για τα χαρακτηριστικά του συνόλου σε αυτό [17]
- Εξαγωγή συμβάντων(events) από ζεύγη παρακείμενων διαστημάτων και δημιουργία μιας βάσης δεδομένων από events. Είναι δυνατόν να επιβληθεί ένα συγκεκριμένο όριο ελάχιστου χρόνου μεταξύ διαδοχικών events. [17]

Πιο αναλυτικά η κύρια μέθοδος της προεπεξεργασίας είναι η εκκαθάριση δεδομένων.

3.1.1 Εκκαθάριση δεδομένων

Η εκκαθάριση δεδομένων (Data Cleansing) είναι μια διαδικασία που συνδέεται άμεσα με την απόκτηση και τον ορισμό δεδομένων ή εφαρμόζεται εκ των υστέρων με σκοπό τη βελτίωση της ποιότητας των δεδομένων σε ένα υπάρχον σύστημα. Περιλαμβάνει τρεις φάσεις:[15]

- Καθορισμός και προσδιορισμός τύπων σφαλμάτων

- Αναζήτηση και αναγνώριση περιπτώσεων σφαλμάτων
- Διόρθωση σφαλμάτων

Και οι τρεις αυτές φάσεις αποτελούν ένα σύνθετο πρόβλημα και για την αντιμετώπισή του μπορούν να εφαρμοστούν μια μεγάλη ποικιλία από εξειδικευμένες μεθόδους και τεχνολογίες. Η τρίτη φάση είναι πολύ δύσκολη να αυτοματοποιηθεί εκτός ενός αυστηρού και καλά καθορισμένου τομέα. [15]
Δυστυχώς, στην επιστήμη της πληροφορικής έχει πραγματοποιηθεί ελάχιστη βασική έρευνα που σχετίζεται άμεσα με τον εντοπισμό σφαλμάτων και τον καθαρισμό των δεδομένων. Δεν έχουν ακόμη δημοσιευθεί συγκρίσεις τεχνικών και μεθόδων καθαρισμού δεδομένων σε βάθος. [15]

3.2 Χειρισμός ελλειπόντων δεδομένων (Handling Missing Data)

Πολλά σύνολα δεδομένων της πραγματικής ζωής είναι ελλιπή, δηλαδή λείπουν ορισμένες τιμές των χαρακτηριστικών. Μερικές τιμές χαρακτηριστικών δεν καταγράφονται επειδή είναι άσχετες, ενώ άλλες φορές λείπουν τιμές χαρακτηριστικών επειδή ξεχάστηκαν ή τοποθετήθηκαν στο σύνολο, αλλά αργότερα διαγράφηκαν κατά λάθος. [15]
Γενικά, οι μέθοδοι για τον χειρισμό τιμών που λείπουν, ανήκουν είτε σε διαδοχικές μεθόδους, είτε σε παράλληλες. [15]
Οι διαδοχικές μέθοδοι περιλαμβάνουν τεχνικές που βασίζονται στη διαγραφή των εγγραφών των οποίων οι τιμές λείπουν, αντικατάσταση μιας τιμής από την πιο κοινή τιμή αυτής, αντιστοίχιση όλων των πιθανών τιμών σε εκείνη που λείπει, αντικατάστασή της με το μέσο όρο, εκχώρηση της αντίστοιχης τιμής που λαμβάνεται από την πλησιέστερη. [15]
Η δεύτερη ομάδα μεθόδων περιλαμβάνει εκείνες στις οποίες λαμβάνονται υπόψη οι ελλειπούσες τιμές των χαρακτηριστικών κατά την διαδικασία απόκτησης της γνώσης. [15]

3.3 Μείωση Διαστάσεων

Οι αλγόριθμοι εξόρυξης δεδομένων χρησιμοποιούνται για την αναζήτηση ουσιαστικών προτύπων σε σύνολα ακατέργαστων δεδομένων. Η διάσταση, ο αριθμός των χαρακτηριστικών των συνόλων δεδομένων, αποτελεί σοβαρό εμπόδιο για την αποτελεσματικότητα των περισσότερων αλγορίθμων Εξόρυξης Δεδομένων. Τεχνικές αρκετά αποτελεσματικές σε χαμηλές διαστάσεις δεν μπορούν να παρέχουν ουσιαστικά αποτελέσματα όταν ο αριθμός των εγγραφών υπερβαίνει ένα όριο χαρακτηριστικών. [15]
Οι αλγόριθμοι εξόρυξης δεδομένων είναι υπολογιστικά απαιτητικοί. Το κόστος τους είναι συνάρτηση της θεωρητικής πολυπλοκότητας του μοντέλου του αλγορίθμου και συσχετίζεται με τον χρόνο που απαιτείται για την εκτέλεση του και με το μέγεθος του συνόλου δεδομένων. [15]
Υπάρχουν τέσσερις κύριοι λόγοι για τους οποίους πολλές φορές η μείωση της διάστασης των δεδομένων καθίσταται απαραίτητη. Αυτοί είναι η μείωση του κόστους μάθησης, η αύξηση της απόδοσης μάθησης, η μείωση άσχετων διαστάσεων και η μείωση περιττών διαστάσεων. [15]
Η μείωση των περιττών και άσχετων διαστάσεων χωρίζεται σε δύο υποπροβλήματα: [15]

- Επιλογή χαρακτηριστικών
  Ο στόχος της επιλογής χαρακτηριστικών είναι ο εντοπισμός ορισμένων σημαντικών χαρακτηριστικών στο σύνολο δεδομένων και η απορρίψη των υπολοίπων. Η διαδικασία επιλογής χαρακτηριστικών μειώνει τη διάσταση του συνόλου των δεδομένων και επιτρέπει στους αλγόριθμους εκμάθησης να λειτουργούν ταχύτερα και πιο αποτελεσματικά. [15]

- Επιλογή εγγραφών
  Ορισμένες εγγραφές μπορεί να βοηθήσουν καλύτερα τη διαδικασία μάθησης από άλλες, οπότε γίνεται επιλογή των πιο σημαντικών. [15]

3.4 Διακριτοποίηση Δεδομένων

Η διακριτοποίηση είναι μια διαδικασία επεξεργασίας δεδομένων που μετατρέπει τα ποσοτικά δεδομένα σε ποιοτικά. Τις περισσότερες φορές οι αλγόριθμοι εξόρυξης χρησιμοποιούν ποσοτικά δεδομένα, όμως ορισμένες φορές είναι διαμορφωμένοι για τον χειρισμό ποιοτικών. [15]
Η αξιολόγηση των αλγορίθμων έχει δείξει ότι συχνά η διακριτοποίηση βοηθά στη βελτίωση της απόδοσης της μάθησης και στην κατανόηση των μαθησιακών αποτελεσμάτων. [15]
Η διακριτοποίηση, ουσιαστικά, ¨ενώνει' τις πραγματικές εφαρμογές εξόρυξης δεδομένων όπου κυριαρχούν τα ποσοτικά δεδομένα, με τους αλγόριθμους μάθησης, οι περισσότεροι από τους οποίους είναι πιο έμπειροι στη μάθηση από ποιοτικά δεδομένα. Ως εκ τούτου, η διακριτοποίηση παίζει σημαντικό ρόλο στην Εξόρυξη Δεδομένων και στην ανακάλυψη της γνώσης. Διαφορετικοί αλγόριθμοι μάθησης απαιτούν διαφορετικές στρατηγικές διακριτοποίησης. Δεν είναι ρεαλιστικό να επιδιώκεται μια καθολικά βέλτιστη προσέγγιση διακριτοποίησης. [15]

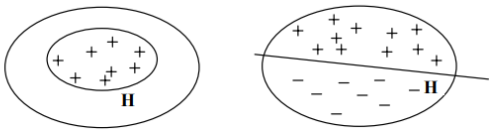
3.5 Εξόρυξη Δεδομένων

Η εξόρυξη δεδομένων (Data Mining) είναι η βασική εργασία στη διαδικασία του KDD. Περιλαμβάνει τις υπολογιστικές τεχνικές που έχουν ως σκοπό την εξαγωγή χρήσιμου μοτίβου ή γνώσης από τα δεδομένα, και συνήθως εκφράζεται με τη μορφή προγνωστικού ή περιγραφικού μοντέλου. [18]
Η έξοδος ενός αλγορίθμου εξόρυξης δεδομένων είναι τυπικά ένα μοτίβο ή ένα σύνολο μοτίβων που περιλαμβάνεται στα δεδομένα. Μόλις εξαχθεί, αυτή η γνώση μπορεί να αξιοποιηθεί και να χρησιμοποιηθεί για τη λήψη αποφάσεων.
Στην εξόρυξη δεδομένων χρησιμοποιούνται τα προεπεξεργασμένα δεδομένα για να δημιουργηθεί ένα μοντέλο που περιγράφει τη συμπεριφορά των χρονοσειρών. Στη συνέχεια εξάγονται κανόνες συσχέτισης από το μοντέλο και γίνεται επικύρωση της προγνωστικής ακρίβειας του μοντέλου. [17]
Οι εργασίες εξόρυξης μπορεί να είναι είτε προγνωστικές είτε περιγραφικές. Οι περισσότερες βασίζονται στη μηχανική μάθηση και τη στατιστική.

- Στην προγνωστική εξόρυξη δεδομένων πραγματοποιείται εστίαση σε μια συγκεκριμένη ιδιότητα ή χαρακτηριστικό, που ονομάζεται χαρακτηριστικό στόχος ή αντικείμενο των δεδομένων. Με βάση τις τιμές του χαρακτηριστικού στόχου, κατασκευάζεται ένα μοντέλο πρόβλεψης που είναι σε θέση να προβλέψει τις τιμές του σε νέες περιπτώσεις. Ο τύπος του μοντέλου εξαρτάται από τον τύπο του

χαρακτηριστικού στόχου. Εάν είναι ονομαστικό, θα παραχθεί μοντέλο ταξινόμησης, ενώ αν είναι αριθμητικό, θα παραχθεί ένα μοντέλο παλινδρόμησης. Παραδείγματα προογνωστικών τεχνικών είναι: κανόνες πρόβλεψης, δέντρα αποφάσεων, τεχνητά νευρωνικά δίκτυα, μοντέλα βασισμένα σε παραδείγματα και μαθηματικές εξισώσεις. [18]

- Στην περιγραφική εξόρυξη δεδομένων δεν υπάρχει συγκεκριμένο χαρακτηριστικό στόχου. Σκοπός είναι να γίνει κατασκευή ενός περιγραφικού μοντέλου που περιγράφει ενδιαφέρουσες κανονικότητες στα δεδομένα. Θα είναι ένα σύνολο με υπάρχουσες συσχετίσεις μεταξύ των χαρακτηριστικών. Δηλαδή, ένα σύνολο συστάδων όπου κάθε συστάδα περιγράφει ένα υποσύνολο παρόμοιων περιπτώσεων, ένα πιθανοτικό μοντέλο, που περιγράφει τις πιθανοτικές εξαρτήσεις μεταξύ των χαρακτηριστικών. Τεχνικές για την περιγραφική εξόρυξη δεδομένων περιλαμβάνουν κανόνες συσχέτισης, δέντρα ομαδοποίησης, επεκτατική ομαδοποίηση και πιθανολογικά μοντέλα. [18]



Σχήμα 4: (a)Descriptive (b)Predictive Data Mining [18]

3.6 Ερμηνεία των δεδομένων

Η ερμηνεία των δεδομένων Data Interpretation περιλαμβάνει τη διαδικασία ¨αποκρυπτογράφησης' των προτύπων που ανακαλύφθηκαν και ενδεχομένως την επιστροφή σε οποιοδήποτε από τα προηγούμενα βήματα, καθώς και πιθανή οπτικοποίηση των εξαγόμενων μοτίβων, αφαιρώντας περιττά ή άσχετα μοτίβα και μεταφράζοντας τα χρήσιμα με όρους κατανοητούς από τους χρήστες. [19]

4 Ορισμός του προβλήματος

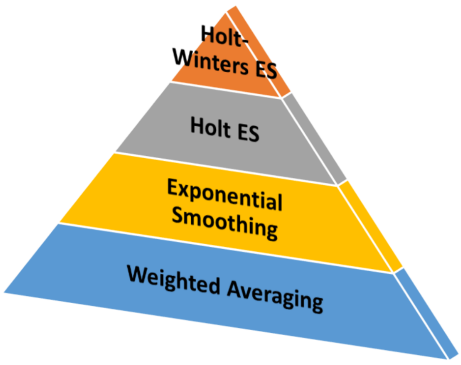
Στην παρούσα εργασία πραγματοποιήθηκε μελέτη δεδομένων που έχουν συλλεχθεί από ένα σύνολο συσκευών που ανήκουν στο Internet of Things και είναι τοποθετημένα σε τρία καταστήματα τραπεζών. Τα δεδομένα αυτά έχουν τη μορφή χρονοσειρών και αφορούν τη θερμοκρασία δωματίων, την μέτρηση του feedback, την ολική αρμονική παραμόρφωση τάσης στο φωτισμό, την κατανάλωση ενέργειας και ισχύος και την μέτρηση της υγρασίας.
Αφού πραγματοποιήθηκε περιγραφή της μορφής του εκάστοτε συνόλου δεδομένων, ακολούθησε η διαδικασία πρόβλεψης. Για την πρόβλεψη χρησιμοποιήθηκαν μερικοί στατιστικοί αλγόριθμοι, μοντέλα νευρωνικών δικτύων, ο αλγόριθμος του Facebook, Prophet, το XGBoost και ο Random Forest. Επιπλέον, μελετήθηκε μία παραλλαγή του Random Forest. Με βάση τα αποτελέσματα που προέκυψαν επιδιώχθηκε να εξαχθούν ορισμένα συμπεράσματα σχετικά με τη συμπεριφορά των παραπάνω αλγόριθμων ως προς τα χαρακτηριστικά της τάσης και της εποχικότητας των χρονοσειρών. Για τα συμπεράσματα δεν χρησιμοποιήθηκαν εξωγενείς παράγοντες, ούτε πολλυπλεξία χρονοσειρών.
Οι παρατηρήσεις διαχωρίστηκαν σε δύο υπό ομάδες. Η πρώτη αφορούσε τη συμπεριφορά των αλγορίθμων ξεχωριστά με βάση την εμφάνιση τάσης ή εποχικότητας σε ένα σύνολο δεδομένων. Η δεύτερη περιέγραφε ποιο μοντέλο πρόβλεψης είναι αποτελεσματικότερο ανάλογα με την ομάδα των δεδομένων, όπως ειναι για παράδειγμα τα δεδομένα που περιγράφουν τη θερμοκρασία δωματίων.
Στη συνέχεια, μελετήθηκαν τρείς τρόποι ensemble·το bagging, η μέθοδος υπολογισμού μέσου όρου και η στοίβαξη. Στο πρώτο χρησιμοποιήθηκε ο αλγόριθμος XGBoost, στο δεύτερο ο XGBoost και η παραλλαγή του Random Forest και στο τελευταίο, το LSTM, το Prophet και το Random Forest.
Έχουν πραγματοποιηθεί αρκετά πειράματα που συγκρίνουν μεθόδους πρόβλεψης και προσπαθούν να εξάγουν συμπεράσματα σχετικά με τη συμπεριφορά των αλγορίθμων στις διαφορετικές μορφές χρονοσειρών. Τα πειράματα αυτά έχουν εφαρμοστεί σε διάφορους τομείς. Στόχος αυτής της εργασίας είναι να πραγματοποιηθεί η συγκεκριμένη μελέτη σε δεδομένα από καταστήματα τραπεζών και με βάση τα συμπεράσματα που θα εξαχθούν να δοθεί η δυνατότητα σε μία μετέπειτα επέκταση της εργασίας, που θα έχει ως σκοπό την κατασκευή ενός αλγορίθμου ο οποίος με βάση ορισμένα στοιχεία των χρονοσειρών που θα δέχεται ως είσοδο, να ορίζει αυτόματα ποιά μέθοδος πρόβλεψης θα χρησιμοποιηθεί. Αυτόματοι αλγόριθμοι έχουν κατασκευαστεί κατά καιρούς, όχι όμως πάνω σε δεδομένα από φυσικά καταστήματα. Ένας αυτόματος αλγόριθμος πρόβλεψης καθίσταται απαραίτητος καθώς η διαδικασία δοκιμής διαφορετικών μεθόδων είναι ιδιαίτερα χρονοβόρα. Συνεπώς, επιλέγοντας αυτόματα ποιά μέθοδος είναι η καταλληλότερη, εξοικονομείται χρόνος και πόροι.

5 Αλγόριθμοι Πρόβλεψης

5.1 Στατιστικοί Αλγόριθμοι

5.1.1 Αλγόριθμοι ΕΞομάλυνσης

Η εξομάλυνση είναι μια στατιστική μέθοδος πρόβλεψης χρονοσειρών που χρησιμοποιείται για τη δημιουργία μιας προσεγγιστικής συνάρτησης με σκοπό την αφαίρεση των παρατυπιών από τα δεδομένα και την καταγραφή σημαντικών μοτίβων σε αυτά. Η τεχνική αυτή εμφανίστηκε για πρώτη φορά το 1950 με την απλή εκθετική εξομάλυνση. Στη συνέχεια το 1956 αναβαθμίστηκε από τον Charles Holt στη διπλή εκθετική εξομάλυνση, επιτρέποντας την ύπαρξη τάσης στη χρονοσειρά που αναλύεται. Το 1960, ο Holt συνεργαζόμενος με τον Peter Winters πρόσθεσαν και την δυνατότητα ύπαρξης εποχικότητας και έτσι προέκυψε τελικά η τριπλή εκθετική εξομάλυνση ή Holt-Winters εκθετική εξομάλυνση.[12]
Η τεχνική την εξομάλυνσης, συνολικά, είναι μια διαδικασία αναθεώρησης των εκτιμήσεων των συντελεστών ενός μοντέλου, στην προσπάθεια σωστής παρατήρησης των δεδομένων. [23] Πιο αναλυτικά, οι τεχνικές εξομάλυνσης χρησιμοποιούν τους σταθμισμένους μέσους όρους μιας προηγούμενης παρατήρησης ώστε να προβλέψουν μια νέα τιμή. Η κύρια ιδέα αυτής της τεχνικής είναι να επιβαρύνουν τις πρόσφατες τιμές σε μια χρονολογική σειρά. Εφαρμόζεται, δηλαδή, εκθετική αύξηση βαρών στις τιμές, με μεγαλύτερα βάρη στις πιο πρόσφατες. [24] Επιπλέον, για τον υπολογισμό των παραμέτρων, η τεχνική, λαμβάνει υπόψη τα λάθη, τις τάσεις και την εποχικότητα, τα οποία τα συγκεντρώνει είτε προσθετικά είτε πολλαπλασιαστικά. [11]



Σχήμα 5: Exponential Smoothing models[11]

(α′) *Απλή Εκθετική Εξομάλυνση*

Η απλή εκθετική εξομάλυνση είναι ένα από τα μοντέλα εκθετικής εξομάλυνσης. Χρησιμοποιείται για μονομεταβλητή παρατήρηση, που δεν έχει κάποια συγκεκριμένη τάση ή εποχικότητα. Εφαρμόζεται για μικρής εμβέλειας πρόβλεψη, συνήθως μόλις ένα μήνα. [10]

Έστω μία σειρά έχει επίπεδο L_t, δεν εμφανίζει τάση και εποχικότητα και έχει σταθερό θόρυβο. Η πρόβλεψη ισούται με το επίπεδο τη χρονική στιγμή t που θέλουμε να προβλέψουμε:[11]

$$F_k = L_t$$

Το επίπεδο L_t τη χρονική στιγμή t υπολογίζεται: [11]

$$L_t = aY_t + (1-a)L_{t-1}$$

Η παράμετρος 'alpha' είναι η παράμετρος εξομάλυνσης και κυμαίνεται μεταξύ των τιμών 0 και 1. [11] Η εξίσωση αυτή δηλώνει ότι το νέο επίπεδο προκύπτει από το επίπεδο την προηγούμενη στιγμή (L_{t-1}) ενημερωμένο με βάση την πληροφορία του πιο πρόσφατου σημείου (Y_t). [11]
Η μέθοδος αυτή καλείται *εκθετική* καθώς αν αναλύσουμε το L_{t-1} προκύπτει η εξίσωση κατά την οποία τα βάρη μειώνονται εκθετικά στο παρελθόν:[11]

$$L_t = aY_t + (1-a)L_{t-1} =$$

$$aY_t + (1-a)[(aY_{t-1} + (1-a)L_{t-1} = ... =$$

$$aY_t + a(1-a)Y_{t-1} + a(1-a)^2Y_{t-2} + ...$$

Όταν η παράμετρος α=1, τότε οι παλιές τιμές δεν επηρεάζουν την πρόβλεψη και αγνοούνται εντελώς, αφού οι συντελεστές τους ισούνται με 0. Όταν η παράμετρος α=0, η τρέχουσα παρατήρηση αγνοείται καθώς ο συντελεστής της γίνεται 0. Η νέα τιμή υπολογίζεται εξ ολοκλήρου από την προηγούμενη τιμή της τεχνικής εξομάλυνσης. Όλες οι τιμές στο τέλος θα έχουν ίδια τιμή και ίση με την αρχική τιμή L_0. [11] Γενικά, όσο πιο μικρή είναι η τιμή του α, τόσο πιο σημαντική είναι η επιλογή της αρχική τιμής L_0. [10]

(β′) *Διπλή Εκθετική Εξομάλυνση*

Η διπλή εκθετική εξομάλυνση χρησιμοποιείται στις χρονοσειρές που εμφανίζουν τάση αλλά όχι εποχικότητα. Η τεχνική αυτή προσθέτει χρονική τάση στην εξίσωση πρόβλεψης:[11]
Αν η τάση είναι προσθετική:[11]

$$F_{t+k} = L_t + kT_t$$

Αν η τάση είναι πολλαπλασιαστική:[11]

$$F_{t+k} = L_t * T_t^k$$

όπου:
L_t είναι το επίπεδο τη χρονική στιγμή t: [11]

$$L_t = aY_t + (1-a)(L_{t-1} + T_{t-1})$$

T_t είναι η τάση τη χρονική στιγμή t:[11]

$$T_t = b(L_t - L_{t-1}) + (1-b)T_{t-1}$$

L_{t-1} είναι το επίπεδο την προηγούμενη χρονική στιγμή, T_{t-1} είναι η τάση την προηγούμενη χρονική στιγμή, Y_t είναι το πιο πρόσφατο σημείο, k είναι ο αριθμός των μελλοντικών βημάτων που θέλουμε να προβλέψουμε και η παράμετρος β ελέγχει την ταχύτητα της προσαρμοζόμενης τάσης. Οι τιμές του β κυμαίνονται και αυτές από το 0 έως το 1. [11]

(γ′) *Τριπλή Εκθετική Εξομάλυνση*

Η τριπλή εκθετική εξομάλυνση είναι μια τεχνική που υπαγορεύει εκθετική εξομάλυνση τρεις φορές. Εφαρμόζεται σε δεδομένα που παρουσιάζουν τάση και εποχικότητα. Είναι γνωστή και ως μέθοδος Holt-Winters, από τα ονόματα των εφευρετών της. Έχει τη δυνατότητα να χειρίζεται αθροιστική ή πολλαπλασιαστική τάση και εποχικότητα. Υπάρχουν δύο κύρια μοντέλα με βάση το είδος της εποχικότητας (αθροιστική ή πολλαπλασιαστική). [10]
Έστω το μοντέλο έχει επίπεδο L_t, τάση και εποχικότητα με M εποχές.
Το αθροιστικό εποχικό μοντέλο υπολογίζεται: [11]

$$Forecast = estimate\ level + trend + seasonality\ at\ most\ recent\ time\ point$$
$$F_{(t+k)} =: L_t + kT_t + S_{(t+k-M)}$$

Το πολλαπλασιαστικό εποχικό μοντέλο υπολογίζεται: [11]

$$Forecast = estimate\ level\ x\ trend\ x\ seasonality\ at\ most\ recent\ time\ point$$
$$F_{(t+k)} = (L_t + kT_t)S_{(t+k-M)}$$

Στο μοντέλο αυτό υπάρχουν τρεις σταθερές: [11]

$$\text{Επίπεδο} : L_t = a(y_t/S_{t-M}) + (1-a)(L_{t-1} + T_{t-1})$$
$$\text{Τάση} : T_t = b(L_t - L_{t-1}) + (1-b)T_{t-1}$$

$$\text{Εποχικότητα} : S_t = g(Y_t/L_t) + (1+g)S_{t-M},$$

όπου S είναι η εποχική σταθερά.

Ως εκ τούτου η τεχνική της **εξομάλυνσης** χρησιμοποιείται για βραχυπρόθεσμα μοντέλα πρόβλεψης. Όταν η δεδομένη χρονοσειρά δεν εμφανίζει τάση ή εποχικότητα, τότε μπορεί να γίνει η χρήση της απλής εκθετικής εξομάλυνσης· αν υπάρχει τάση και όχι εποχικότητα, χρησιμοποιείται η διπλή εκθετική εξομάλυνση, ενώ όταν παρουσιάζει τάση και εποχικότητα, η χρήση της τριπλής εκθετικής εξομάλυνσης είναι απαραίτητη. Συνολικά η εξομάλυνση είναι πιο αποτελεσματική όταν τα δεδομένα χρονοσειρών κινούνται αργά με την πάροδο του χρόνου. Επιπλέον, υπάρχουν τόσοι υπερπαράμετροι, που καθιστούν την διαδικασία του συντονισμού δύσκολη και χρονοβόρα. [24]

5.1.2 Moving Average

Η τεχνική moving average, αποτελεί ίσως την πιο απλή μορφή πρόβλεψης και πολλές φορές χρησιμοποιείται ως μια τεχνική εξομάλυνσης ώστε να βρεθεί μία ευθεία γραμμή που να ενώνει τα δεδομένα της χρονοσειράς. Η τιμή της πρόβλεψης προκύπτει ως ο μέσος όρος των n προηγούμενων τιμών. Ο αριθμός n ορίζεται ως το window size.[11]
Η μέθοδος αυτή απαιτεί μεγάλο window size για να εξομαλύνει τον θόρυβο και να εντοπίσει την τάση. Όμως, μεγάλο μέγεθος παραθύρου μπορεί να καθυστερήσει την τάση, ελέγχοντας όλο και πιο παλαιές τιμές για να προβλέψει τις μελλοντικές. [24]
Η απλή μέθοδος του moving average αργεί να προσαρμοστεί στις νέες τάσεις και οι προβλεπόμενες τιμές υστερούν σε σχέση με την πραγματικότητα. Όπως και η τεχνική της εκθετικής εξομάλυνσης, δεν αντέχει σε μεγαλύτερο χρονικό πλαίσιο πρόβλεψης. [11]

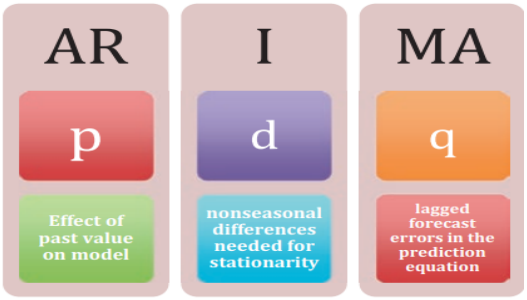
5.1.3 Autoregressive Models

Η παλινδρόμηση ή regression είναι μια εποπτευόμενη τεχνική εκμάθησης στη μηχανική μάθηση όπου προσπαθεί να εκτιμήσει τις μεταβλητές στόχους χρησιμοποιώντας έναν ή πολλαπλούς αναδρομείς. Πιο συγκεκριμένα, τα μοντέλα που χρησιμοποιούν την τεχνική αυτή προβλέπουν τις επερχόμενες τιμές με βάση τις προηγούμενες. [11]

(α′) ***ARIMA(Autoregressive Integrated Moving Average)***
Το 1970 ο μαθηματικός George Box και ο Gwilym Jenkins δημοσίευσαν το άρθρο 'Time Series: Forecasting and Control', στο οποίο, χρησιμοποιώντας ως βάση την ιδέα της μεθόδου 'Moving Average', περιέγραψαν το μοντέλο ARIMA (Autoregressive Integrated Moving Average) ή αλλιώς Box-Jenkins model. Το μοντέλο αυτό, όπως προκύπτει και από το όνομά του, είναι autoregressive, δηλαδή χρησιμοποιεί την εξαρτημένη σχέση ενός σημείου με κάποια παλαιότερα.[11] Επίσης, στο μοντέλο αυτό γίνεται χρήση της διαφοράς ενός σημείου και κάποιου προηγούμενου, μετατρέποντας τελικά μία σειρά από τιμές, σε μία σειρά από αλλαγές τιμών (integrated). Αυτό είναι απαραίτητο ώστε η τελική σειρά που θα αναλυθεί να είναι στατική κι ο βαθμός διαφοροποίησης εξαρτάται από το πόσες φορές χρειάζεται η σειρά να διαφορηθεί ώστε να μετατραπεί σε στατική. Ένα παράδειγμα του τελευταίου χαρακτηριστικού του μοντέλου αποτελεί ότι η αυριανή θερμοκρασία θα είναι παρόμοια με σήμερα, γιατί δεν έχει μεταβληθεί ιδιαίτερα μέσα στην εβδομάδα. [24]
Οι παράμετροι της μεθόδου είναι [11] :

- p: Ο αριθμός τον παρατηρήσεων που θα χρησιμοποιηθούν για την παλινδρόμηση. (trend autoregressive order) AR(q)
- d: Ο αριθμός των φορών που η παρατήρηση θα διαφορηθεί, δηλαδή, ο βαθμός της διαφοροποίησης. Αν η χρονοσειρά είναι στατική τότε $d = 0$. (trend differencing order) I(d)
- q: Ο αριθμός των παλαιών σφαλμάτων προβλέψεων. (trend moving average order) MA(q)

Έστω $X_{t-k}, k = 0, 1, 2, \ldots, p$ η στατική χρονοσειρά. Το $ARMA(p, q)$, το μοντέλο χωρίς τη διαφόρηση είναι ο γραμμικός συνδυασμός από τις παλαιές τιμές του X_t και τα σφάλματα e_t, όπως



Σχήμα 6: ARIMA[11]

περιγράφει η εξίσωση:

$$X_t = \Psi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \cdots + \phi_p X_{t-p} - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \ldots - \theta_q \epsilon_{t-q} + \epsilon_t$$

Όπου p και q οι τιμές των παραμέτρων του μοντέλου $ARIMA$, $\phi_1, \phi_2, \ldots, \phi_p, \theta_1, \theta_2, \ldots, \theta_q$ τα βάρη που θα υπολογιστούν, $\Psi 0$ η σταθερά της τάσης και $\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon t - q$ τα τυχαία σφάλματα. [25] Ένα σοβαρό μειονέκτημα του μοντέλου ARIMA είναι ότι δεν υποστηρίζει εποχικότητα. Για αυτό το λόγο δημιουργήθηκε το SARIMA που θα αναλυθεί στη συνέχεια.

(β') ***SARIMA(Seasonal ARIMA)***

Η συγκεκριμένη τεχνική του ARIMA, μπορεί να χειρίζεται εποχικότητα σε μία χρονοσειρά. [11] Το μοντέλο έχει μια γενική πολλαπλασιαστική μορφή, $SARIMA(p,d,q)X(P,D,Q)_m$. Το πρώτο μέρος (p,d,q) περιέχει τη σειρά των μη εποχικών παραμέτρων, ενώ η σειρά των εποχιακών παραμέτρων περιέχεται στο δεύτερο μέρος (P,D,Q).[28]
Οι παράμετροι του $SARIMA(p,d,q)X(P,D,Q)_m$ εκτός από τους p,d,q είναι: [11]

- P: seasonal autoregressive order
- D: seasonal differencing order
- Q: seasonal moving average order
- m: Ο αριθμός των βημάτων σε μία περίοδο

Ο γενικός τύπος είναι:

$$\phi_p(L)\Phi_P(L^m)\nabla^d\nabla^D X_t = \theta_q(L)\theta_Q(L^m)\epsilon_t + c$$

όπου φ και Φ είναι μη εποχική και εποχική autoregressive παράμετρος αντίστοιχα, ∇ είναι διαφορικός τελεστής και ϑ, Θ μη εποχική και εποχική παράμετρος του moving average αντίστοιχα.[26] Ο George Box και ο Gwilym Jenkins χρησιμοποίησαν ορισμένα απλοποιημένα βήματατα για να αποκτήσουν τις απαραίτητες πληροφορίες για την κατανόηση του μοντέλου ARIMA. Η μεθοδολογία Box-Jenkins (BJ) αποτελείται από τέσσερα επαναληπτικά βήματα: [28]

- Βήμα 1: Ταυτοποίηση
  Αυτό το βήμα επικεντρώνεται στην επιλογή του βαθμού διαφοροποίησης (d), της εποχιακής διαφοροποίησης (D), της αυτοπαλίνδρομης της μη εποχικής σειράς (p), της αυτοπαλίνδρομης της εποχικής σειράς (P), του αριθμού των παλαιών σφαλμάτων πρόβλεψης της μη εποχικής σειράς (q) και του αριθμού των παλαιών σφαλμάτων πρόβλεψης της εποχικής σειράς (Q).
  Ο αριθμός των παραμέτρων μπορεί να προσδιοριστεί παρατηρώντας τις αυτοσυσχετίσεις δειγμάτων (SAC) και τις μερικές αυτοσυσχετίσεις δειγμάτων ($SPAC$).
- Βήμα 2: Εκτίμηση
  Τα δεδομένα του συνόλου της ιστορίας χρησιμοποιούνται για την εκτίμηση των παραμέτρων του μοντέλου στο Βήμα 1.
- Βήμα 3: Διαγνωστικός Έλεγχος
  Η διαγνωστική δοκιμή χρησιμοποιείται για τον έλεγχο της επάρκειας του δοκιμαστικού μοντέλου.
- Βήμα 4: Πρόβλεψη
  Το τελικό μοντέλο στο Βήμα 3 χρησιμοποιείται για την πρόβλεψη των τιμών πρόβλεψης.

(γ′) *SARIMAX*

Το SARIMAX είναι μοντέλο του SARIMA, μόνο που επιτρέπει την επιρροή από εξωγενείς παράγοντες, όπως είναι η βροχή στην πρόβλεψη του καιρού.[24]
Η αλληλεξάρτηση μεταξύ διαφορετικών μεταβλητών δεν μπορεί να μελετηθεί χρησιμοποιώντας το μοντέλο ARIMA. Όμως, υπάρχει περίπτωση, τα προβλήματα της πραγματικής ζωής να περιέχουν κάποιου είδους σχέσεις μεταξύ διαφορετικών μεταβλητών. Οι συνεχώς μεταβαλλόμενοι κανόνες μιας πολυμεταβλητής χρονοσειράς δεν μπορούν να εκφραστούν χρησιμοποιώντας το μοντέλο ARIMA. Επομένως, παρατηρήθηκε η ανάγκη κατασκευής ενός μοντέλου που να μπορεί να λειτουργήσει σε μια πολυμεταβλητή χρονοσειρά.[29]
Γι αυτό, ένα σταθερό μοντέλο χρονοσειρών που εργάζεται σε πολλαπλές μεταβλητές, χρησιμοποιώντας το προϋπάρχον μοντέλο ARIMA, δημιουργήθηκε από τους Box και Jenkins.[29]
Η σταθερότητα μεταξύ χρονοσειρών εισόδου και εξωγενών χρονοσειρών είναι απαραίτητη. Επιπλέον, χρειάζεται να ικανοποιείται ο περιορισμός που απαγορεύει την ανάπτυξη στο πεδίο της ανάλυσης σε πολυμεταβλητές χρονοσειρές. Σε αυτό το πεδίο, οι Granger και Engle πρότειναν την έννοια της συνολοκλήρωσης. Αυτή η θεωρία αναφέρει ότι το υπόλοιπο των χρονοσειρών εισόδου και των χρονοσειρών εξόδου, μετά την παλινδρόμηση, απαιτείται να σταθεροποιηθεί, χωρίς να απαιτείται σταθεροποίηση της ίδιας της σειράς. Με τη βοήθεια της έννοιας της συνολοκλήρωσης, κατέστη δυνατή η ανάπτυξη της πολυμεταβλητής ανάλυσης χρονοσειρών. [29]

(δ′) *VARIMA(Vector Autoregressive Integrated Moving Average)*

Το Vector Autoregression (VAR) είναι ένα μοντέλο στοχαστικής διαδικασίας το οποίο χρησιμοποιείται για την εκμετάλλευση της γραμμικής σχέσης μεταξύ των πολλαπλών μεταβλητών δεδομένων χρονοσειράς. Με άλλα λόγια, είναι μια πολυμεταβλητή μέθοδος πρόβλεψης, που χρησιμοποιείται όταν δύο ή περισσότερες μεταβλητές χρονοσειρών έχουν ισχυρή εσωτερική σχέση

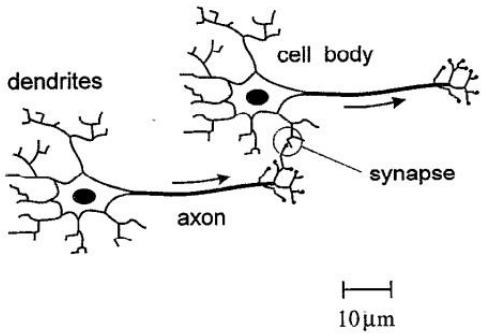
μεταξύ τους. Το VAR είναι μοντέλο αμφίδρομης κατεύθυνσης. Σε ένα μοντέλο μονής κατεύθυνσης, ένας προγνωστικός παράγοντας επηρεάζει τον στόχο, αλλά όχι το αντίστροφο. Σε ένα αμφίδρομο μοντέλο, οι μεταβλητές επηρεάζουν η μία την άλλη.[11]
Πιο αναλυτικά, το VARIMA είναι μια πολυμεταβλητή γενίκευση του μονομεταβλητού μοντέλου ARIMA. Τα χαρακτηριστικά του παρουσιάστηκαν για πρώτη φορά από τον Quenouille το 1957, ενώ το λογισμικό για την εφαρμογή τους έγινε διαθέσιμο τις δεκαετίες του 1980 και του 1990. Δεδομένου ότι τα μοντέλα VARIMA μπορούν να φιλοξενήσουν υποθέσεις σχετικά με την εξωγένεια και τις σύγχρονες σχέσεις, προσέφεραν νέες προκλήσεις στους μετεωρολόγους και τους υπεύθυνους χάραξης πολιτικής.[30]
Γενικά, τα μοντέλα VAR τείνουν να υποφέρουν από υπερβολική προσαρμογή με πάρα πολλές ελεύθερες ασήμαντες παραμέτρους. Ως αποτέλεσμα, αυτά τα μοντέλα μπορούν να παρέχουν κακές προβλέψεις εκτός δείγματος, παρόλο που η προσαρμογή εντός του δείγματος είναι καλή. [30]

(ε′) ***FARIMA (Fractional ARIMA) / ARFIMA (Fractionally Integrated ARMA***)
Το μοντέλο FARIMA περιλαμβάνει κλασματικό βαθμό διαφοράς, γεγονός που επιτρέπει μεγάλη μνήμη. Αυτό συμβαίνει, επειδή παρατηρήσεις που βρίσκονται σε απόσταση μεταξύ τους στο χρόνο μπορεί να μην έχουν αμελητέες εξαρτήσεις. [11]

Συνολικά, το μοντέλο ARIMA, εφαρμόζεται σε μακροπρόθεσμες σειρές οι οποίες εμφανίζουν ένα επαρκές μοτίβο και έχουν ορισμένη πληροφορία που θα βοηθήσει στην πρόβλεψη. Μεγάλο πλεονέκτημα αποτελεί η ευελιξία του, η σχετική ακρίβεια των προβλέψεων που προκύπτουν και η δυνατότητα επέκτασής του στην ανάλυση πολλαπλών χρονοσειρών. Κυρίως χρησιμοποιείται σε ad hoc αναλύσεις, δηλαδή σε αναλύσεις για τη βαθύτερη κατανόηση χρηματοοικονομικών δεδομένων. Παρόλα αυτά, η χρήση του μοντέλου απαιτεί υπολογιστικά δαπανηρή βελτιστοποίηση και συντονισμό. Ενώ ακόμα, τα αποτελέσματά του εξαρτώνται από την ικανότητα και την εμπειρία του εργατικού δυναμικού. Επιπλέον, τα παλινδρομικά μοντέλα που αναλύθηκαν είναι γραμμικά, ενώ αντίθετα είναι πιθανό, σε κάποια χρονοσειρά, η πρόβλεψη να είναι μία μη γραμμική συνάρτηση των προηγούμενων παρατηρήσεων. [11]

5.2 Νευρωνικά μοντέλα πρόβλεψης χρονοσειρών

Τα νευρωνικά μοντέλα πρόβλεψης χρονοσειρών εμφανίστηκαν λόγω της ανάγκης ανάπτυξης μοντέλων πρόβλεψης όταν τα δεδομένα εμφανίζουν μη γραμμική εξάρτηση και για αντοχή υψηλού θορύβου. [11]
Η έμπνευση για τα νευρωνικά δίκτυα προέρχεται από μελέτες των μηχανισμών για την επεξεργασία πληροφοριών σε βιολογικά νευρικά συστήματα, ειδικά στον ανθρώπινο εγκέφαλο. Ο ανθρώπινος εγκέφαλος αποτελείται από περίπου 10^{11} ηλεκτρικά ενεργά κύτταρα που ονομάζονται νευρώνες. Οι δενδρίτες παρέχουν ένα πλήθος από εισόδους και ο άξονας προσφέρει τις εξόδους. Η επικοινωνία μεταξύ διαφορετικών νευρώνων επιτυγχάνεται μέσω συνάψεων. Ένας νευρώνας συνδέεται με πολλούς χιλιάδες νευρώνες, με αποτέλεσμα την ύπαρξη τουλάχιστον 10^{14} συνάψεων στον εγκέφαλο. Παρόλο που κάθε νευρώνας είναι ένα σχετικά αργό σύστημα επεξεργασίας πληροφοριών, ο παραλληλισμός της επεξεργασίας πληροφοριών μέσα από τις πολλές συνάψεις, οδηγεί σε μία αποτελεσματική επεξεργαστική ισχύ και σε έναν υψηλό βαθμό ανοχής σφαλμάτων. Όταν ένα σήμα φτάσει σε μια σύναψη, ενεργοποιεί την απελευθέρωση χημικών νευροδιαβιβαστών που διασχίζουν τη σύναψη έως τον επόμενο νευρώνα. Κάθε σύναψη έχει ένα βάρος που καθορίζει την επίδραση της ώθησης στον μετασυναπτικό νευρώνα. [31]

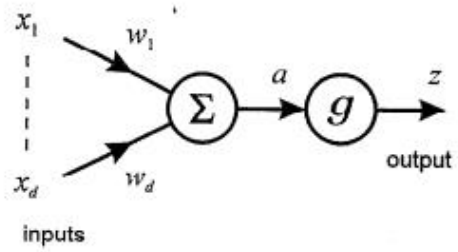


Σχήμα 7: Σχήμα νευρώνα [31]

Η ανάπτυξή των νευρωνικών δικτύων, προέκυψε από την προσπάθεια προσομοίωσης αυτών των βιολογικών νευρικών συστημάτων, μέσα από τον συνδυασμό απλών υπολογιστικών στοιχείων. Τα δίκτυα αυτά είναι μια κατηγορία διακριτών μοντέλων μη γραμμικής παλινδρόμησης. Αποτελούνται, συνήθως, από έναν μεγάλο αριθμό νευρώνων, δηλαδή γραμμικών ή μη γραμμικών στοιχείων, συνδεδεμένων με πολύπλοκους τρόπους και οργανωμένους σε στρώματα.[31]
Για πρώτη φορά, έγινε η περιγραφή ενός απλού μαθηματικού μοντέλου για έναν απλό νευρώνα σε μία δημοσίευση το 1943 από τον McCullock και τον Pitts. [31]
Ένας απλός perceptron ή νευρώνας, αρχικά, υπολογίζει τον γραμμικό συνδυασμό εισόδων. Στη συνέχεια μία πιθανώς μη γραμμική συνάρτηση ενεργοποίησης εφαρμόζεται στον γραμμικό συνδυασμό των εισόδων x_i πολλαπλασιασμένων με βάρος w_i και προκύπτει η έξοδος:[31]

$$a = \sum_{i=0}^{d} w_i x_i + w_0, , w_0 \text{ καλείται } bias$$

Στη συνέχεια, μία συνάρτηση ενεργοποίησης $g()$ αντιστοιχίζει μία οποιαδήποτε πραγματική είσοδο σε ένα συνήθως οριοθετημένο εύρος τιμών, από 0 έως 1 ή από -1 έως 1 και προκύπτει η έξοδος $z = g(a)$. Ορισμένες συναρτήσεις ενεργοποίησης αποτελούν:[11]



Σχήμα 8: Perceptron [11]

- Συνάρτηση δυαδικού βήματος:
$$f(x) = \begin{cases} 0, x < 0 \\ 1, x \geq 0 \end{cases}$$
- Γραμμική συνάρτηση ενεργοποίησης:
$$f(x) = x$$
- Μη γραμμικές συναρτήσεις ενεργοποίησης:
  - Sigmoid
  $$f(x) = \sigma(x) = \frac{1}{1 + e^{-x}}$$
  - tanh
  $$f(x) = \tanh = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$
  - Rectified Linear Unit (ReLU)
  $$f(x) = \begin{cases} 0, x \leq 0 \\ x, x > 0 \end{cases}$$
  - Softmax
  $$f(x) = \frac{x}{1 = e^{-x}}$$

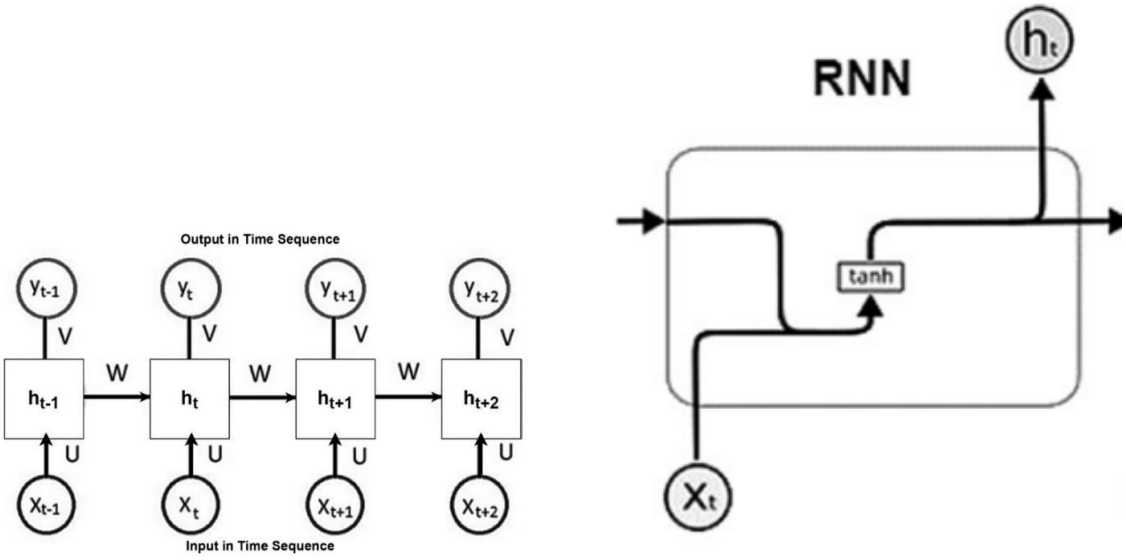
Ένας perceptron μπορεί να έχει μία ή περισσότερες εξόδους, οι οποίες έχουν ξεχωριστό bias και ξεχωριστό σετ από βάρη. Συνήθως χρησιμοποιείται η ίδια συνάρτηση ενεργοποίησης σε κάθε έξοδο. [31] Το πρόβλημα προσδιορισμού τιμών για τα βάρη στο νευρωνικό δίκτυο καλείται training.
Συμπερασματικά, τα τεχνητά νευρωνικά δίκτυα, όπως πολλές στατιστικές μέθοδοι, είναι ικανά να επεξεργάζονται τεράστιες ποσότητες δεδομένων και να κάνουν προβλέψεις αρκετά ακριβείς. «Μαθαίνουν» με τον ίδιο τρόπο που πολλοί στατιστικοί αλγόριθμοι κάνουν εκτιμήσεις, όμως πολύ πιο αργά από τους στατιστικούς. Τα νευρωνικά δίκτυα έχουν ένα σημαντικό μειονέκτημα, ότι παρέχουν μόνο χρονικούς υπολογισμούς σημείων, λαμβάνοντας μόνο ορισμένες προβλεπόμενες τιμές, χωρίς δείκτες εμπιστοσύνης των προβλέψεων. [11]

5.2.1 BackPropagation

Τα νευρωνικά δίκτυα εκπαιδευονται με τη διαδικασία του back propagation.Στα back propagation νευρωνικά δίκτυα, οι μαθηματικές σχέσεις μεταξύ των διαφόρων μεταβλητών δεν προσδιορίζονται. Αντίθετα, εκπαιδεύονται από ορισμένα παραδείγματα που τους δίνονται ως είσοδοι. [62]
Πιο αναλυτικά, κάθε κρυφός και εξερχόμενος νευρώνας επεξεργάζεται τις εισόδους του πολλαπλασιάζοντας κάθε είσοδο με το βάρος του. Στη συνέχεια αθροίζει το γινόμενο και «φιλτράρει» το άθροισμα μέσω της συνάρτησης μεταφοράς, μιας μη γραμμικής συνάρτησης, για να παραχθεί ένα αποτέλεσμα. Συνήθως, ως συνάρτηση μεταφοράς χρησιμοποιείται η σιγμοειδής καμπύλη.[62]
Το νευρωνικό δίκτυο "μαθαίνει" τροποποιώντας τα βάρη των νευρώνων με βάση τα σφάλματα μεταξύ των πραγματικών τιμών εξόδου και των τιμών εξόδου στόχου. Αυτό πραγματοποιείται μέσω της gradient descent στο άθροισμα των τετραγώνων των σφαλμάτων για όλο το σύνολο εκπαίδευσης. Οι αλλαγές στα βάρη είναι ανάλογες με το αρνητικό της παραγώγου του σφάλματος.
Ενα πέρασμα από το σύνολο εκπαίδευσης μαζί με την ενημέρωση των βαρών ονομάζεται κύκλος ή εποχή. Η εκπαίδευση επαναλαμβάνεται έως ότου ελαχιστοποιηθεί το σφάλμα όλων των εκπαιδεύσεων και εντός της ανοχής που καθορίζεται για το πρόβλημα. Στο τέλος της φάσης εκπαίδευσης, το νευρωνικό δίκτυο θα πρέπει να αναπαράγει σωστά τις στοχευόμενες τιμές εξόδου για τα δεδομένα εκπαίδευσης με την προϋπόθεση ότι τα σφάλματα είναι ελάχιστα, δηλαδή υπάρχει σύγκλιση. Τα σχετικά εκπαιδευμένα βάρη των νευρώνων αποθηκεύονται στη συνέχεια στη μνήμη του νευρικού δικτύου. Στην επόμενη φάση, το εκπαιδευμένο νευρωνικό δίκτυο τροφοδοτείται με ένα ξεχωριστό σύνολο δεδομένων. Σε αυτή τη φάση δοκιμής, οι προβλέψεις του νευρωνικού δικτύου (χρησιμοποιώντας τα εκπαιδευμένα βάρη) συγκρίνονται με τις τιμές εξόδου στόχου. Αυτό αξιολογεί την αξιοπιστία του νευρωνικού δικτύου για τη γενίκευση των σωστών αποκρίσεων για τα πρότυπα δοκιμών που μοιάζουν μόνο σε μεγάλο βαθμό με τα δεδομένα στο σύνολο εκπαίδευσης.[63]
Δεν υπάρχουν πρόσθετες ρυθμίσεις μάθησης ή βάρους κατά τη διάρκεια αυτής της φάσης. Μόλις διαπιστωθεί ότι οι φάσεις εκπαίδευσης και δοκιμής είναι επιτυχείς, το νευρωνικό δίκτυο μπορεί στη συνέχεια να χρησιμοποιηθεί σε πρακτικές εφαρμογές. Το νευρωνικό δίκτυο θα παράγει σχεδόν στιγμιαία αποτελέσματα της εξόδου για τις πρακτικές εισόδους που παρέχονται. Οι προβλέψεις θα πρέπει να είναι αξιόπιστες με την προϋπόθεση ότι οι τιμές εισόδου είναι εντός του εύρους που χρησιμοποιείται στο σετ εκπαίδευσης.[64]
Το σύνολο εκπαίδευσης πρέπει να περιλαμβάνει ένα αντιπροσωπευτικό δείγμα των δεδομένων το οποίο περιέχει τα διάφορα διακριτά χαρακτηριστικά του προβλήματος που είναι πιθανό να αντιμετωπίσει το νευρωνικό δίκτυο στην ολοκληρωμένη εφαρμογή. Ενα μεγάλο σύνολο εκπαίδευσης μειώνει τον κίνδυνο υποδειγματοληψίας της μη γραμμικής συνάρτησης αλλά αυξάνει τον χρόνο της διαδικασίας της εκπαίδευσης

5.2.2 Recurrent Neural Networks

Το RNN είναι ένας τύπος νευρωνικών δικτύων στο οποίο η κατάσταση κάθε νευρώνα αποθηκεύεται και προωθείται στο επόμενο επίπεδο. Στο συγκεκριμένο νευρωνικό δίκτυο, όλες οι είσοδοι και οι έξοδοι είναι ανεξάρτητες μεταξύ τους. Ονομάζεται recurrent καθώς εμφανίζει κύκλους ώστε να έχει πρόσβαση σε προηγούμενες πληροφορίες. Περιλαμβάνει ένα επίπεδο εισόδου, ένα κρυφό επίπεδο και ένα επίπεδο εξόδου.[11]
Έστω είσοδος η ακολουθία $X_{t-1}, X_t, X_{t+1}, X_{t+2}$. Περνάει από το κρυμμένο επίπεδο και παίρνει τις τιμές από το προηγούμενο βήμα και τα αντίστοιχα βάρη από το συγκεκριμένο επίπεδο. [11]



Σχήμα 9: RNN Network[11]

Τη χρονική στιγμή t δηλαδή[11]:

$$h_t = s(U * X_t + W * h_{t-1})$$

$$y_t = softmax(V * h_1)$$

όπου:

- σ = tanH συνάρτηση ενεργοποίησης
- U = διάνυσμα βαρών του κρυμμένου επιπέδου
- V = διάνυσμα βαρών του επιπέδου εξόδου
- W = το ίδιο διάνυσμα βαρών για διαφορετικές χρονικές στιγμές
- X = διάνυσμα εισόδου

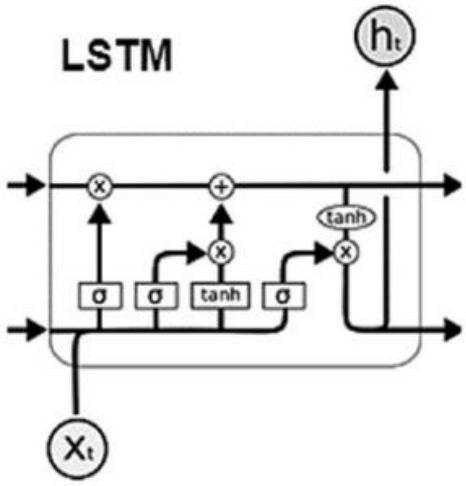
- U = διάνυσμα εξόδου

Εν κατακλείδει, το μοντέλο RNN χρειάζεται αρκετή ώρα να εκπαιδευτεί και να «μάθει» από τα δεδομένα. Ιδιαίτερα εάν οι ακολουθίες είναι μεγάλες, τα gradients που χρησιμοποιούνται για τον συντονισμό του βάρους και του bias, και υπολογίζονται κατά τη διάρκεια της εκπαίδευσης, είτε εξαφανίζονται, αφού πολλαπλασιάζονται συνεχώς με μικρές τιμές μικρότερες του 1, είτε γίνεται «έκρηξη», όταν πολλαπλασιάζονται με πολλές μεγάλες τιμές μεγαλύτερες του 1 (vanishing grading problem). Αυτό έχει ως αποτέλεσμα το μοντέλο να εκπαιδεύεται πολύ αργά. Πολλές φορές επίσης, οι τελικές παράμετροι και υπερπαράμετροι δεν ορίζονται σωστά, παρουσιάζοντας χαμηλή ακρίβεια. Εν γένει το RNN χρησιμοποιείται όταν έχει σημασία η σειρά των δεδομένων. [11]

5.2.3 Long Short-Term Memory (LSTM)

Τα early RNNs παρουσίαζαν short memory καθώς χρησιμοποιούσαν μόνο το προηγούμενο επίπεδο πληροφορίας. Για τις χρονοσειρές, λόγω του ότι συχνά τα μοτίβα επαναλαμβάνονται μετά από μεγάλες χρονικές αποστάσεις, χρειαζόταν ένα μοντέλο με long short term memory.Το μοντέλο LSTM, επιπλέον, επιλύει σε μεγάλο βαθμό το vanishing grading problem που αναφέρθηκε παραπάνω και εμφανίζεται στην περίπτωση των απλών RNNs.[11]
Το μοντέλο LSTM εμφανίστηκε για πρώτη φορά το 1997 από τον Sepp Hochreiter και τον Jurgen Schmidhuber.[11]
Περιλαμβάνει ένα κρυμμένο επίπεδο h και ένα έξτρα επίπεδο την «μνήμη» m, υπεύθυνα για τις ενημερώσεις και τις εξόδους. Έχει συνολικά τέσσερις πύλες[11]:

- Την Forget gate f με σιγμοειδή συνάρτηση ενεργοποίησης.
- Την Candidate gate C με συνάρτηση ενεργοποίησης την tanH.
- Την Input Gate I με σιγμοειδή συνάρτηση ενεργοποίησης.
- Την Output Gate O με σιγμοειδή συνάρτηση ενεργοποίησης.



Σχήμα 10: Σχήμα LSTM

Έστω μία είσοδος x_t στο σύστημα. Στην αρχή, συνδέεται με το προηγούμενο μοτίβο των δεδομένων της ακολουθίας της χρονοσειράς.[11]
Η εξίσωση της forget gate είναι:

$$f_t = s(W_f[h_{t-1}, x_t] + b_f)$$

Στη συνέχεια, αποφασίζεται αν η πληροφορία είναι σημαντική ή όχι. Η πύλη Input Gate αποφασίζει ποια τιμή θα αναβαθμιστεί. Στη συνέχεια η Candidate Gate παράγει ένα διάνυσμα με τις υποψήφιες τιμές. Οι εξισώσεις των δύο πυλών είναι[11]:

$$i_t = s(W_i[h_(t-1), x_t] + b_i)$$

$$\tilde{C_t} = tanh(W_C[h_(t-1), x_t] + b_C)$$

Με βάση τις τιμές αυτές, ανανεώνεται η τιμή του C_{t-1}[11]:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C_t}$$

Στο τέλος η πύλη Output αποφασίζει για την έξοδο με βάση την τιμή C_t. Οι εξισώσεις του επιπέδου εξόδου και του κρυμμένου επιπέδου αντίστοιχα είναι[11]:

$$o_t = \sigma(W_o[h_{t-1}, x_t] + b_0)$$

$$h_t = o_t * tanh(C_t)$$

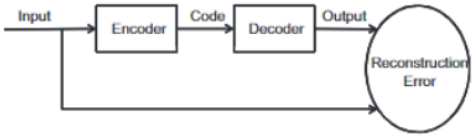
Περιλαμβάνει δύο τύπους:

- Peephole
- Peephole convolutional

Συνοψίζοντας, η ικανότητα του LSTM να ξεχνά, να θυμάται και να ενημερώνει τις πληροφορίες, προχωρά την πρόβλεψη χρονοσειρών σε ένα βήμα παρακάτω. [32] Έχει το πλεονέκτημα απέναντι σε άλλες μεθόδους να εντοπίζει μη γραμμική σχέση μεταξύ των δεδομένων, αλλά απαιτεί, όπως όλες οι deep learning μέθοδοι, υπολογιστικό χρόνο, επαρκή αριθμό δεδομένων και υπολογιστική ισχύ. Επιπλέον, λόγω του αριθμού των υπερπαραμέτρων, απαιτεί για τις αποφάσεις έναν έμπειρο forecaster. Για πιο πρακτικά ζητήματα, όπου το κόστος και ο χρόνος πρέπει να ληφθούν υπόψη, κυρίως προτιμάται το ARIMA. [24]
Προς το παρόν, τα RNNs και πιο συγκεκριμένα τα LSTM βρίσκονται στην αιχμή της τεχνολογίας. Το θετικό σε αυτά τα δίκτυα είναι ότι επιτρέπουν στο δίκτυο να έχουν πρόσβαση σε όλο το ιστορικό από τις προηγούμενες τιμές των χρονοσειρών.[11]

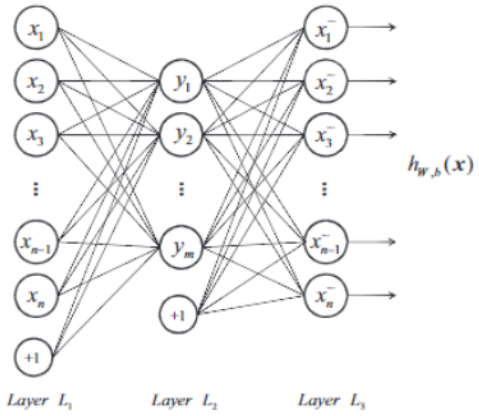
5.2.4 LSTM Auto-Encoders

Ο Auto-Encoder είναι μοντέλο νευρωνικού δικτύου με ίσο αριθμό νευρώνων στο στρώμα εισόδου και εξόδου και μικρότερο αριθμό στο κρυμμένο στρώμα. Έχει την ικανότητα να «μαθαίνει» τα κρυμμένα χαρακτηριστικά των δεδομένων εισόδου, το οποίο ονομάζεται encoding. Στη συνέχεια, μπορεί να «επαναδημιουργήσει» τα δεδομένα εισόδου από τα χαρακτηριστικά που «έμαθε» κατά τη διαδικασία του encoding. Η διαδικασία αυτή ονομάζεται decoding. [33]



Σχήμα 11: Σχήμα Auto-Encoder[33]

Δεδομένου ότι τα κρυφά στρώματα αποτελούνται από πολύ λιγότερους νευρώνες, για να ανακατασκευαστεί η είσοδος όσο το δυνατόν πιο «σωστά», τα βάρη στα κρυφά επίπεδα αποτυπώνουν μόνο τα πιο αντιπροσωπευτικά χαρακτηριστικά του πρωτοτύπου και αγνοούν τις λεπτομέρειες των δεδομένων εισόδου, όπως είναι ακραίες τιμές.



Σχήμα 12: Σχήμα Auto-Encoder[33]

Ο LSTM Auto-Encoder περιλαμβάνει ένα LSTM δίκτυο για encoder και ένα LSTM δίκτυο για decoder.[11]

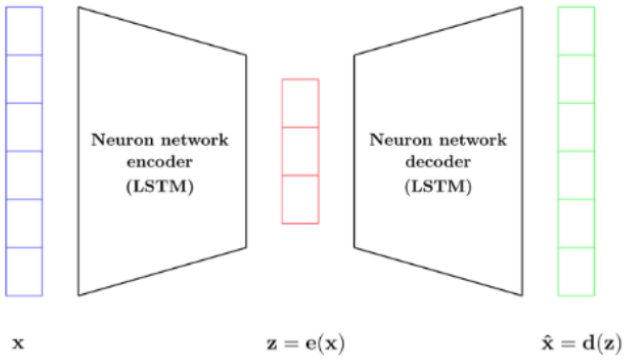


Fig. 3. An illustration of a LSTM Autoencoder network.

Σχήμα 13: Σχήμα LSTM Auto-Encoder

Έστω x η είσοδος του encoder, από όπου θα εξαχθούν τα νέα χαρακτηριστικά y. Η διαδικασία του encoding είναι μία γραμμική συνάρτηση της εισόδου x, ακολουθούμενη από μία μη γραμμική συνάρτηση ενεργοποίησης[11]:

$$y = f(Wx + b)$$

Στη συνέχεια με βάση τα νέα χαρακτηριστικά y ακολουθει η διαδικασία decoding, ώστε να ανακατασκευαστεί η είσοδος[11]:

$$x' = f(W'x + b')$$

Η έξοδος συγκρίνεται με την είσοδο και με βάση το σφάλμα ενημερώνονται τα βάρη του δικτύου. Ο autoencoder εκπαιδεύεται με σκοπό την ελαχιστοποίηση του reconstruction σφάλματος. [11]

$$L = \frac{1}{2}\sum_{x} ||x - \hat{x}||^2$$

Οι Auto-Encoders χρησιμοποιούνται με σκοπό την μείωση των διαστάσεων των δεδομένων, αλλά παράλληλα και τη διατήρηση των βασικών χαρακτηριστικών τους. Επιπλέον, ως μοντέλο βαθιάς εκμάθησης χωρίς επίβλεψη, οι auto-encoders μπορούν να χρησιμοποιηθούν για τη δημιουργία νέων δεδομένων που είναι διαφορετικά από τα δεδομένα που έχουν εκπαιδευτεί. Με αυτόν τον τρόπο, οι αυτόματοι κωδικοποιητές (variational auto-encoders) είναι παραγωγικά μοντέλα. [11]

5.2.4.1 Bidirectional LSTM

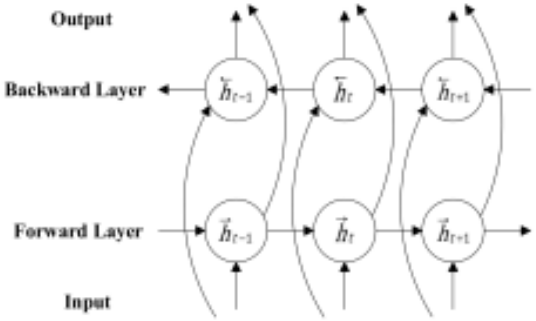
Στο Bidirectional LSTM για την εκπαίδευση του δικτύου, γίνεται χρήση τόσο των παρελθοντικών όσο και των μελλοντικών χαρακτηριστικών των δεδομένων για ένα συγκεκριμένο χρονικό πλαίσιο. [34]
Χρησιμοποιεί δύο επίπεδα τέτοια ώστε το ένα στρώμα να εκτελεί τις διαδικασίες(operations) που ακολουθούν την ίδια κατεύθυνση της ακολουθίας δεδομένων και το άλλο επίπεδο να εκτελεί τις πράξεις της αντίστροφης κατεύθυνσης από αυτή των δεδομένων. Με άλλα λόγια, τα δίκτυα BiLSTM έχουν δύο τρόπους για να περάσουν πληροφορίες: έναν από το παρελθόν στο μέλλον και έναν από το μέλλον στο παρελθόν.
Η συνολική δομή του δικτύου Bi-LSTM αποτελείται από δύο κύρια κρυφά στρώματα. Σε κάθε επίπεδο, στοιβάζονται πανομοιότυπα LSTM τα οποία θεωρούνται το ένα ως ένα πίσω κρυφό στρώμα και το άλλο ως ένα μπροστά κρυφό στρώμα.[36]
Έστω X_t η είσοδος τη χρονική στιγμή t με $\overrightarrow{H_t}$ και $\overleftarrow{H_t}$ τα κρυμένα επίπεδα της ίδιας κατεύθυνση της ακολουθίας δεδομένων και της αντίστροφης αντίστοιχα. Το συνολικό κρυμμένο επίπεδο H_t υπολογίζεται ολοκληρώνοντας τα $\overrightarrow{H_t}$ και $\overleftarrow{H_t}$. [36]

$$\overrightarrow{H_t} = tanh(X_t W_{xh}^{(f)} + \overrightarrow{H_{t-1}})W_{hh}^{(f)} + b_h(f)$$

$$\overleftarrow{H_t} = tanh(X_t W_{xh}^{(b)} + \overleftarrow{H_{t-1}})W_{hh}^{(b)} + b_h(b)$$

$$O_{fn} = H_t W_o + b_o$$

όπου $X_t W_{xh}^{(f)}, W_{hh}^{(f)}, b_h^{(f)}, X_t W_{xh}^{(b)}, W_{hh}^{(b)}, b_h^{(b)}$ οι παράμετροι του μοντέλου.



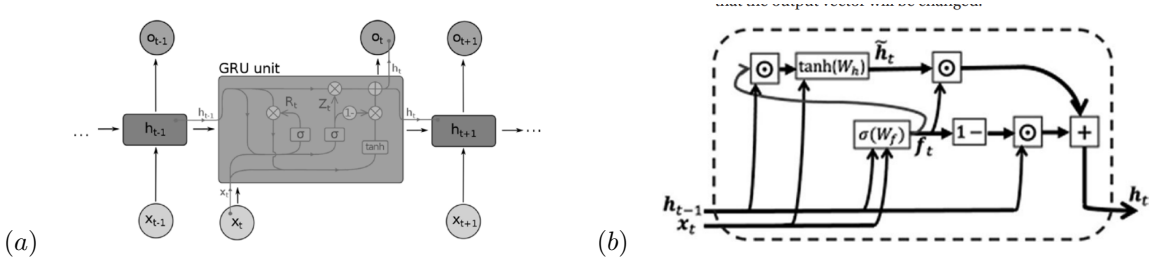
Σχήμα 14: Σχήμα BidirectionalLSTM

Συμπερασματικά, το BiLSTM είναι μια κατάλληλη λύση όταν προβλέπουμε δεδομένα πολλαπλών βημάτων, των οποίων οι έξοδοι χρησιμοποιούνται ως είσοδοι στα μελλοντικά βήματα. Διαθέτει ισχυρή μνήμη για να αποθηκεύει όλες τις χρήσιμες προηγούμενες και μελλοντικές λειτουργίες με υψηλή ακρίβεια. Επιπλέον, το δίκτυο αυτό είναι πιο ακριβές στην πρόβλεψη των φαινομένων με υψηλή στοχαστική και διακοπτόμενη συμπεριφορά επειδή δεν υπακούουν στην αναδρομική διαδικασία που ανατροφο-

δοτεί την προηγούμενη πληροφορία με επαναληπτικό τρόπο. [36] Το BiLSTM σε ορισμένες εφαρμογές, όπως η φωνητική ταξινόμηση, είναι περισσότερο αποτελεσματικό από το μονοκατευθυντικό LSTM.[35]

5.2.5 Gated Recurrent Units (GRU)

Το GRU εμφανίστηκε το 2014 από τον Kyunghyun Cho και είναι ένα είδος Recurrent Neural Network. Είναι παρόμοιο σε δομή με το LSTM, απλά έχει λιγότερες παραμέτρους. Συνολικά έχει δύο πύλες. Δεν περιλαμβάνει πύλη εξόδου, αλλά μία πύλη update και μία πύλη reset. Οι δύο πύλες είναι διανύσματα που αποφασίζουν ποια πληροφορία θα περάσει στην έξοδο. [37] Η πύλη update (z_t) ελέγχει την πληροφορία που προέρχεται από την προηγούμενη ενεργοποίηση και προσθέτει την νέα πληροφορία. Η reset gate εισέρχεται στην υποψήφια συνάρτηση ενεργοποίησης. [38]



(a) (b)

Σχήμα 15: GRU Networks
(a)Fully Gated Unit (b) Minimal Gated Unit

Έχει δύο παραλλαγές[11]:

- Fully Gated Unit
  Έστω είσοδος x_t, έξοδος h_t, update gate z_t , reset gate r_t, W, U και b παράμετροι:

$$h_0 = 0$$

$$z_t = \sigma_g(W_z x_t + U_z h_{t-1} + b_z)$$

$$r_t = \sigma_g(W_r x_t + U_r h_{t-1} + b_r)$$

$$h_t = z_t \odot h_{t-1} + (1 - z_t) \odot \phi_h(W_h x_t + U_h(r_t \odot h_{t-1}) + b_h)$$

  , όπου σ_g η $sigmoid$ συνάρτηση και ϕ_h η $tanH$ συνάρτηση ενεργοποίησης.

- Minimal Gated Unit
  Είναι παρόμοιο με το fully gated unit, με διαφορά όμως το reset gate το οποίο συγχωνεύεται σε ένα forget gate. Αλλάζοντας τις εξισώσεις ως εξής:

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

$$h_t = f_t \odot h_{t-1} + (1 - f_t) \odot \phi_h(W_h x_t + U_h(f_t \odot h_{t-1}) + b_h)$$

Το μοντέλο GRU εν γένει αποδίδει καλύτερα σε μικρότερα dataset. Είναι αρκετά παρόμοιο με το LSTM, οπότε είναι πιο δύσκολο να καθοριστεί ποιο μοντέλο είναι καταλληλότερο σε διαφορετικά προβλήματα. [11]

5.2.6 Convolution Neural Networks (CNN)

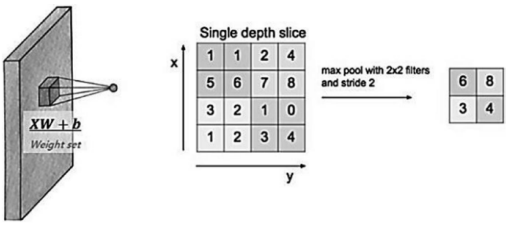
Το μοντέλο CNN δημοσιεύτηκε για πρώτη φορά το 1995 από τον LeCun και τον Bengio. Είναι ένας βιολογικά εμπνευσμένος τύπος deep neural network (DNN) που βασίζεται στην παρατήρηση ότι οι νευρώνες του οπτικού φλοιού μπορούν να αναγνωρίζουν με σωστό τρόπο ένα αντικείμενο ανεξάρτητα από την αλλαγή στη θέση ή τη διεύθυνση του. Γι αυτό, στο CNN, αντίθετα με τους πολυεπίπεδους perceptron, είναι δυνατή η αναγνώριση μοναδικών χαρακτηριστικών των αντικειμένων ανεξάρτητα από τη θέση και τον προσανατολισμό τους.[11]
Αποτελείται από μία ακολουθία συνελικτικών στρωμάτων, των οποίων η έξοδος συνδέεται μόνο με την είσοδο τοπικών περιοχών και τα οποία εξάγουν χαρακτηριστικά και συμπυκνώνουν πληροφορίες. Αυτά το στρώματα και οι συγκεντρώσεις επαναλαμβάνονται ανάμεσα στο επίπεδο εισόδου και εξόδου. Ο αριθμός των επαναλήψεων των επιπέδων μεταβάλλεται κάθε φορά ώστε να προκύψει το πιο σωστό αποτέλεσμα.[11]
Τα φίλτρα συνέλιξης αποτελούνται από μήτρες βαρών. Το φίλτρο, στο επίπεδο συνέλιξης, σαρώνει την εικόνα από πάνω αριστερά, προς τα κάτω δεξιά, διαβάζοντας τα χαρακτηριστικά της εικόνας και υπολογίζεται η συνέλιξη μεταξύ της εισόδου και του φίλτρου.[39] Αναζητά τα ίδια χαρακτηριστικά σε διαφορετικές θέσεις στην εικόνα εισόδου και τελικά δημιουργείται ένας χάρτης χαρακτηριστικών. Εάν το μέρος που πέρασε από το φίλτρο ταιριάζει με ένα χαρακτηριστικό, δίνεται υψηλή τιμή για να αυξηθεί η πιθανότητα ταξινόμησης της εικόνας σε μία κατηγορία.[11]
Για τις διαστάσεις των μήτρων στο επίπεδο συνέλιξης ισχύει:

- Input: (n x n x n_c)
- Filter: (f x f x n_c)
- Ouput: ($[n + 2p - \frac{f}{s} + 1]$ x $[n + 2p - \frac{f}{s} + 1]$ x n'_c)

Όπου, f το μέγεθος του φίλτρου, s το $stride$, p το pad, n το μέγεθος της εισόδου, και $ncÆ$ ο αριθμός των φίλτρων
Στη συνέχεια, πραγματοποιείται καταχώρηση των «χαρτών» των χαρακτηριστικών που έχουν δημιουργηθεί από τα διαφορετικά φίλτρα, σε ένα φύλλο, όπως φαίνεται στη εικόνα 16. Στη συνέχεια ένα CNN pooling layer συμπυκνώνει τους χάρτες σε μικρότερους, εξάγοντας έναν αριθμό από κάθε χάρτη. [11]
Υπάρχουν δύο είδη pooling:

- Max pooling: Επιλέγεται η μέγιστη τιμή από τα κελιά
- Average pooling: Επιλέγεται ο μέσος όρος των τιμών των κελιών.

Η διαδικασία επαναλαμβάνεται ορισμένες φορές, ώστε στο τέλος προκύπτει ένα νούμερο. Αυτή η τιμή αποτελεί τη στοχαστική εκτίμηση. [11]
Συνολικά, το CNN είναι αρκετά καλό στο να αναγνωρίζει απλά μοτίβα σε δεδομένα και στη συνέχεια να τα χρησιμοποιεί ώστε να δημιουργήσει πιο περίπλοκα μοτίβα σε υψηλότερα στρώματα. Το CNN μίας διάστασης είναι κυρίως χρήσιμο όταν χρειάζεται να εξαχθούν πληροφορίες για ορισμένα μέρη ενός



Σχήμα 16: Σχήμα CNN[11]

dataset, που βρίσκονται σε άσχετα μεταξύ τους σημεία σε αυτό. Είναι, επίσης, ιδανικό για την ανάλυση χρονοσειρών σε δεδομένα αισθητήρων. [11]
Λόγω των πολλών επιπέδων μπορούν να δουλέψουν καλά σε θορυβώδη συστήματα, απορρίπτοντας σε κάθε επίπεδο τον θόρυβο, εξάγοντας μόνο το ουσιώδη μοτίβα.
Το πλεονέκτημα του CNN έναντι του RNN είναι ότι λόγω της συνελικτικής δομής του δικτύου, ο αριθμός των εκπαιδευόμενων βαρών είναι μικρός, έχοντας πιο αποτελεσματική εκπαίδευση και πρόβλεψη.[11]

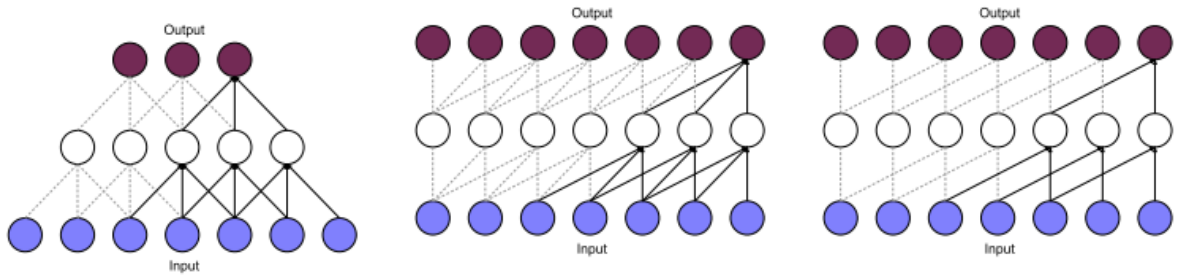
5.2.7 Temporal convolutional neural (TCN)

Το μοντέλο TCN αναπτύχθηκε το 2017 αρχικά για να εξετάζει μοτίβα μεγάλης εμβέλειας χρησιμοποιώντας μια ιεραρχία χρονικών συνελικτικών φίλτρων. [41] Είναι ένας τύπος συνελικτικού νευρωνικού δικτύου με σχεδιασμό που το καθιστά κατάλληλο για τον χειρισμό χρονοσειρών.[42]
Τα βασικά χαρακτηριστικά του είναι: [41]

- Περιλαμβάνει αιτιακές συνελίξεις. Η αιτιακή συνέλιξη διαφέρει από την τυπική συνέλιξη στο γεγονός ότι η συνελικτική λειτουργία που εκτελείται για να ληφθεί η έξοδος τη στιγμή t δεν λαμβάνει μελλοντικές τιμές ως εισόδους. Αυτό σημαίνει ότι με μέγεθος πυρήνα k, η έξοδος O_t προκύπτει από τις τιμές $Xt_{k-1}, Xt_{k-2}, ..., Xt_1, Xt$. [42]

- Έχει τη δυνατότητα να αντιστοιχήσει σε μία ακολουθία εξόδου μια ακολουθία από οποιοδήποτε μήκος, όπως το RNN.

Η προτεινόμενη αρχιτεκτονική είναι επηρεασμένη από γενικές συνελικτικές αρχιτεκτονικές για διαδοχικά δεδομένα που αναπτύχθηκαν πρόσφατα. Είναι απλή και χρησιμοποιεί αυτοπαλίνδρομική πρόβλεψη με πολύ μεγάλη μνήμη. Επιπλέον, επιτρέπει βαθιά δίκτυα όσο και πολύ μεγάλη ιστορία. [41]
Το παραπάνω επιτυγχάνεται μέσω μονοδιάστατων διευρυμένων συνελίξεων που επιτρέπουν την αναγνώριση πιο μακροπρόθεσμων μοτίβων.[41] Αυτή η συνέλιξη αυξάνει το πεδίο λήψης του δικτύου χωρίς τη χρήση λειτουργιών συγκέντρωσης(pooling operations), επομένως δεν υπάρχει σφάλμα ανάλυσης. Η διεύρυνση συνίσταται στην παράλειψη d τιμών μεταξύ των εισόδων της συνελικτικής λειτουργίας. [42]



Σχήμα 17: (a) Απλό Συνελικτικό Δίκτυο με 2 επίπεδα και μέγεθος πυρήνα 3 (b) Αιτιακό Συνελικτικό δίκτυο με 2 επίπεδα και μέγεθος πυρήνα 3 (c) Διευρυμένο Αιτιακό Συνελικτικό Δίκτυο με 2 επίπεδα, μέγεθος πυρήνα 3 και βαθμός διαστολής 2 [42]

Η διευρυμένη λειτουργία αιτιώδους συνέλιξης σε διαδοχικά στρώματα περιγράφεται:

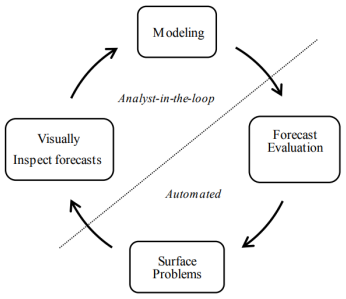
$$x_t^t \;=\; g(\sum_{k=0}^{K-1} w_l^k x_{(l-1)}^{(t-(kxd))} + b_l),$$

όπου x_l^t είναι η έξοδος του νευρώνα στη θέση t στο επίπεδο l, K είναι το μέγεθος του συνελικτικού πυρήνα, w_l^k είναι το βάρος στη θέση k, d είναι ο διασταλτικός παράγοντας της συνέλιξης, b_l το κατώφλι και g η συνάρτηση ενεργοποίησης.[42]

Μια άλλη τεχνική για την περαιτέρω αύξηση του δεκτικού πεδίου του δικτύου είναι η σύνδεση πολλών μπλοκ TCN. Μία τέτοια διαδικασία, όμως, οδηγεί σε βαθύτερες αρχιτεκτονικές με περισσότερες παραμέτρους που καθιστούν τη διαδικασία εκμάθησης πιο περίπλοκη.
Συμπερασματικά, τα παραπάνω χαρακτηριστικά του μοντέλου TCN, το καθιστούν μια κατάλληλη αρχιτεκτονική βαθιάς εκμάθησης για σύνθετα προβλήματα χρονοσειρών. Το κύριο πλεονέκτημα του είναι ότι, όπως και τα RNN, μπορούν να χειριστούν εισόδους μεταβλητού μήκους ολισθαίνοντας τον μονοδιάστατο αιτιατό συνελικτικό πυρήνα. Επίσης, τα μοντέλα TCN είναι πιο αποδοτικά στη μνήμη από τα επαναλαμβανόμενα δίκτυα. Αυτό συμβαίνει καθώς η κοινή αρχιτεκτονική συνέλιξης τα επιτρέπει να επεξεργάζονται μακριές ακολουθίες παράλληλα, σε αντίθεση με τα RNN στα οποία οι ακολουθίες εισόδου επεξεργάζονται διαδοχικά αυξάνοντας το χρόνο υπολογισμού.

5.3 Prophet

Το Prophet είναι ένα πλαίσιο ανοιχτού κώδικα από το facebook που χρησιμοποιείται για πλαισίωση και πρόβλεψη χρονοσειρών. Εστιάζει σε ένα προσθετικό μοντέλο το οποίο είναι μη γραμμικό. Είναι πανίσχυρο στο χειρισμό των δεδομένων που λείπουν και των αλλαγών εντός των τάσεων και γενικά χειρίζεται καλά τις ακραίες τιμές. Επιτρέπει, επίσης, να συσσσωρεύονται εξωγενείς μεταβλητές στο μοντέλο.[11]
Πιο συγκεκριμένα το μοντέλο Prophet αντιμετωπίζει δύο θέματα. Αρχικά ότι τα πιο αυτόματα μοντέλα παρουσίαζαν ακαμψία και ανικανότητα να δεχτούν πρόσθετες υποθέσεις, ενώ επίσης τα πιο ισχυρά εργαλεία απαιτούν έμπειρο αναλυτή με εξειδικευμένες γνώσεις.[11]
Είναι ικανό να χειρίζεται δεδομένα που λαμβάνονται ωριαία, μηνιαία και ετήσια με, ιδανικά, τουλάχιστον ένα πλήρες έτος ιστορικών δεδομένων, έντονη εποχικότητα, διακοπές ή εκδηλώσεις που δεν ακολουθούν ορισμένη εποχικότητα, όπως για παράδειγμα είναι η περίοδος των Χριστουγέννων, την έλλειψη δεδομένων και ακραίων τιμών, σημαντικές αλλαγές στις τάσεις, όπως η κυκλοφορία νέων λειτουργιών ή προϊόντων, και τάσεις που προσεγγίζουν ασυμπτωτικά ένα ανώτερο ή κατώτερο όριο.[11]



Σχήμα 18: Prophet Workflow[40]

Το Prophet βασίζεται στην τεχνική προσαρμογής καμπύλης του Μπεϋζιανού μοντέλου. Έχει εύκολα κατανοητές παραμέτρους και επίσης δεν απαιτεί πολλά δεδομένα χρονοσειρών για να κάνει πρόβλεψη. Η τεχνική είναι πιο κατάλληλη όταν τα δεδομένα χρονοσειρών έχουν ισχυρά εποχικά χαρακτηριστικά ως παράγοντες επιρροής. Διαθέτει επίσης εύκολα χρησιμοποιήσιμες και ερμηνεύσιμες βιβλιοθήκες.[40]
Όσο αφορά τη μορφή του, είναι ένα προσθετικό παλινδρομικό μοντέλο το οποίο περιλαμβάνει: [11]

- καμπύλη λογιστική ανάπτυξης
- ετήσια εποχική καμπύλη

- μηνιαία εποχική καμπύλη
- διακοπές ή άλλες εκδηλώσεις- γεγονότα
- προσθετικές από το χρήστη εποχικότητες, πχ ωριαία

Δηλαδή:

$$y(t) = g(t) + s(t) + h(t) + \epsilon_t$$

όπου:

- $g(t)$ = Τάση(γραμμική /λογιστική) Ο παράγοντας τάσης ($g(t)$) μπορεί να μοντελοποιηθεί με δύο τρόπους
  - Μοντέλο λογιστικής ανάπτυξης
    Αυτό το μοντέλο αντιπροσωπεύει την ανάπτυξη σε διάφορα στάδια. Στο πρώτο στάδιο παρατηρείται η ανάπτυξη περίπου εκθετικά και μετά από αυτό, μετά το στάδιο κορεσμού, μεγαλώνει γραμμικά. Το μοντέλο μπορεί να διατυπωθεί: [40]

    $$f(x) = \frac{L}{1 + e^{-k(x-x_0)}}$$

    όπου L αντιπροσωπεύει τη μέγιστη τιμή της καμπύλης, k είναι ο ρυθμός ανάπτυξης, x_0 είναι η τιμή του x στο σιγμοειδές σημείο
  - Piece-wise γραμμικό μοντέλο
    Είναι μια τροποποιημένη έκδοση του γραμμικού μοντέλου στο οποίο τα διαφορετικά εύρη του x έχουν γραμμικές σχέσεις με διαφορετικό τρόπο. Το μοντέλο μπορεί να διατυπωθεί: [40]

    $$y \ = \ \beta_0 \ + \ \beta_1 x \ + \beta_2 (x - c)^+ \ + \ \epsilon$$

- $s(t)$ = Περιοδικές μεταβολλές /εποχικότητα

  $$s(t) = \sum_{n=1}^{N} a_n cos(\frac{2\pi nt}{P}) + b_n sin(\frac{2\pi nt}{P}),$$

  όπου P είναι η περίοδος που έχει η χρονοσειρά. ($P = 365.35$ για δεδομένα ενός χρόνου, $P = 7$ για εβδομαδιαία δεδομένα, θεωρώντας ότι μετράμε την μεταβλητή σε ημέρες).[11]
  Για την εποχικότητα απαιτείται να υπολογιστούν οι 2N παράγοντες $b = [a1, b1, \ldots aN, bN]^T$ δημιουργώντας έναν πίνακα από διανύσματα για κάθε τιμή του t από τα δεδομένα.

- $h(t)$ = Επιρροές από διακοπές [11]
  Για κάθε περίοδο διακοπών i, τότε Di είναι το σύνολο των δεδομένων που ανήκουν σε αυτή την περίοδο και ισχύει:

  $$Z(t) = [l(teD_1), \ldots, l(teD_L)]$$

  $$h(t) = Z(t)k$$

  Για κ χρησιμοποιείται κ∼Νορμαλ(0,ν^2).

- ϵt = Αλλαγές που δεν περιλαμβάνονται στο μοντέλο

Η βασική διαδικασία του Prophet υλοποιείται χρησιμοποιώντας Stan, μία πιθανολογική γλώσσα προγραμματισμού, η οποία εκτελεί map optimization με σκοπό την εύρεση των παραμέτρων χρησιμοποιώντας τον αλγόριθμο Hamiltonian Monte σε λιγότερο από δευτερόλεπτο. Η γλώσσα αυτή, είναι συμβατή με την Python και την R και έτσι, μοιράζεται η ίδια διαδικασία προσαρμογής για την υλοποίηση και στις δύο γλώσσες. [11]

5.4 XGBOOST

Το XGBoost είναι μία συντομογραφία του πακέτου 'eXtreme Gradient Boosting'. Είναι μία αποτελεσματική και επεκτάσιμη εφαρμογή του πλαισίου ενίσχυσης της κλίσης (gradient boost) που δημιουργήθηκε από τον Chen το 2016. Η λέξη Extreme αναφέρεται στο ότι είναι ένας μεγάλος αλγόριθμος μηχανικής μάθησης με πολλά μέρη. [56]

Το μοντέλο αυτό σχεδιάστηκε ώστε να προσφέρει ταυτόχρονα ταχύτητα και υψηλή επίδοση χρησιμοποιώντας gradient-boosted decision trees. Προσφέρει τη δυνατότητα χρήσης μεγάλης ποικιλίας υπολογιστικών περιβαλλόντων όπως είναι η παραλληλοποίηση, οι υπολογισμοί εκτός πυρήνα και οι κατανεμημένοι υπολογισμοί με σκοπό τη διαχείριση μεγάλων δεδομένων. Υποστηρίζει ορισμένες αντικειμενικές συναρτήσεις, συμπεριλαμβανομένης της παλινδρόμησης, της ταξινόμησης και της κατάταξης. Επίσης, έχει τη δυνατότητα να εντοπίζει και να αντιμετωπίζει τις τιμές που λείπουν. [56]

Η ταχύτητα του αποτελεί βασικό πλεονέκτημά του, καθώς μπορεί να κάνει αυτόματα παράλληλους υπολογισμούς σε Windows και Linux με openmp, χρησιμοποιώντας όλους τους πυρήνες του μηχανήματος στο οποίο τρέχει και αξιοποιώντας όλη τη μνήμη και τους πόρους του hardware.Η ταχύτητα του είναι περισσότερο από δέκα φορές πιο μεγάλη από άλλα μοντέλα.[57] Σημαντική είναι και και η δυνατότητα του να δέχεται αραιή είσοδο.[59]

Επιπλέον, είναι εξαιρετικά επεκτάσιμο χρησιμοποιώντας κατανεμημένες ή παράλληλες υπολογιστικές και αλγοριθμικές βελτιστοποιήσεις όπως είναι ο αλγόριθμος εκμάθησης δέντρων, ένας αλγόριθμος για τον χειρισμό αραιών δεδομένων που θα παρουσιαστεί στη συνέχεια. [56]

Τέλος, το XGBoost είναι διαθέσιμο και μπορεί να συνδυαστεί σε πολλές πλατφόρμες cloud όπως είναι η Tianchi της Alibaba, η AWS, GCE και Azure. Επιπλέον, μπορεί να διασυνδέεται με συστήματα ροής δεδομένων cλουδ, όπως το Spark και το Flink. Η γλώσσες προγραμματισμού που μπορούν να το χρησιμοποιήσουν είναι αρκετές, όπως είναι η Python, Java, C++, R.[56]

Gradient Boosting

Ο αλγόριθμος ενίσχυσης (Boosting), είναι ένας αλγόριθμος μηχανικής μάθησης ο οποίος χρησιμοποιείται ώστε να μειωθεί το κατώφλι (bias) και η διακύμανση των δεδομένων. Μετατρέπει τους αδύναμους learners, οι οποίοι έχουν αδύναμη συσχέτιση με τους πραγματικούς ταξινομητές, σε ισχυρούς.Πιο συγκεκριμένα, κατά την εκπαίδευση στα δεδομένα που προστίθενται, αντιστοιχούν ορισμένα βάρη. Αν το δεδομένο ταξινομηθεί σωστά, μειώνεται το βάρος τους, σε αντίθετη περίπτωση το βάρος αυξάνεται. Στο Gradient Boosting απαιτούνται συνολικά m βήματα για να φτάσει η εκπαίδευση στο ολοκληρωμένο μοντέλο F. Στο βήμα $m+1$, το βασικό μοντέλο $h_{m+1}(x)$ θα εκπαιδευτεί και θα υπολογιστεί η τιμή $y - F_m$ για την πρόβλεψη σε αυτό το βήμα. [58]

$$F_{m+1} \;=\; F_m + h_{m+1}(x)$$

Συνεπώς ο υπολογισμός του στόχου βασίζεται στην έβρεση $h_{m+1}(x) \;=\; F_{m+1} \;-\; F_m$.

Τρόπος λειτουργίας του αλγορίθμου

Ο XGBOOST αποτελεί μία υλοποίηση του Gradient Boosting που συνδυάζει με γραμμικό τρόπο, πολλούς αδύναμους ταξινομητές σε έναν ισχυρό. Οι ταξινομητές μπορούν να είναι τόσο γραμμικοί όσο και $CART$. Στη συνάρτηση κόστους εκτελεί επέκταση $Taylor$ δεύτερης τάξης. [58] Η κύρια ιδέα του αλγορίθμου είναι να προσθέτει στο σύνολο συνεχώς αδύναμα δέντρα με διαφορετικά βάρη.Τα δέντρα θα

πρέπει να πλησιάσουν το υπολοιπώμενο από την προηγούμενη πρόβλεψη όσο το δυνατόν περισσότερο: [58]

$$\hat{y} \;=\; \sum_{k=1}^{K} f_k(x_i)\, f_k \in F$$

- $\hat{y}$ είναι η προβλεπόμενη τιμή
- F είναι το σύνολο με τα δέντρα
- f_k είναι ένα από τα δέντρα
- K είναι ο αριθμός των δέντρων στο σύνολο

Η τιμή του $\hat{y}_i$ πρέπει να είναι όσο το δυνατόν πιο κοντα στην πραγματική τιμή του y_i, χωρίς να χάνει την ικανότητα γενίκευσης.
Ο τύπος για την κανονικοποιημένη συνάρτηση είναι: [58]

$$Obj^{(t)} \;=\; \sum_{i=1}^{n} l(y_i, \hat{y_i}^{(t)}) + \sum_{i=1}^{t} \Omega(f_t) + constant(1)$$

- $l(y_i, \hat{y}_i)$ είναι η συνάρτηση σφάλματος. Μπορεί να είναι οποιαδήποτε συνάρτηση δευτέρου βαθμού διαφορίσιμη.
- $\Omega(f_t)$ περιγράφει την πολυπλοκότητα του μοντέλου. Όσο πιο μικρή είναι, τόσο καλύτερη είναι η ικανότητα γενίκευσης του μοντέλου. Η τιμή της εξαρτάται από τον αριθμό των κόμβων-φύλλων (T) και την τιμή που αντιπροσωπεύουν (w) όπως φαίνεται στην παρακάτω εξίσωση:

$$\Omega(f) \;= \gamma T + \frac{1}{2}\lambda ||w||^2$$

Ο τύπος (1), αντικαθιστώντας τη συνάρτηση σφάλματος με την αντίστοιχη σειρά $Taylor$ δευτέρου βαθμού, καθώς και την συνάρτηση πολυπλοκότητας μετατρέπεται:[58]

$$Obj^{(t)} \;\approx\; \sum_{i=1}^{n} l[g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)] + \Omega(f_t)$$

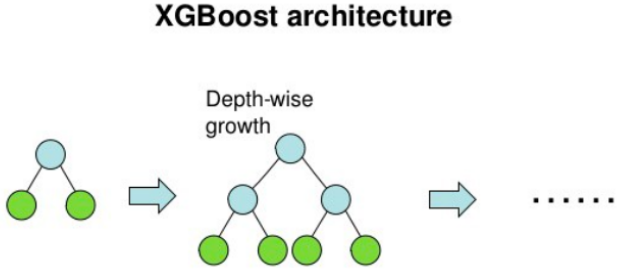
$$= \sum_{i=1}^{n} [g_i w_q(x_i) + \frac{1}{2} h_i w_q^2(x_i)] + \gamma T + \lambda \frac{1}{2} \sum_{j=1}^{T} w_j^2$$

$$= \sum_{j=1}^{T} [(\sum_{i \in I_j} g_i) w_j + \frac{1}{2} (\sum_{i \in I_j} h_i + \lambda) W_j^2] + \gamma T$$

- $g_i \;=\; \partial \hat{y}(t-1) l(y_i, \hat{y_i}^{(t-1)})$, πρώτου βαθμού παράγωγο
- $h_i \;=\; \partial^2 \hat{y}^{(t-1)} l(y_i, \hat{y_i}^{(t-1)})$, δευτέρου βαθμού παράγωγο
- I_j είναι το σύνολο των δεικτών από τα δείγματα σε κάθε κόμβο-φύλλο j

Για ένα δεδομένο $q(x_i)$, παίρνοντας το παράγωγο του w_j και εξισώνοντάς το με το 0, μπορεί να υπολογιστεί το καλύτερο βάρος w_j^* για ένα κόμβο-φύλλο j: [58]

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda}$$



Σχήμα 19: Αρχιτεκτονική XGBoost

5.5 Random Forest Regression

Ο αλγόριθμος Random Forest εμφανίστηκε πρώτη φορά από τον Breiman το 2001 και ήταν επέκταση μίας προηγούμενης εργασίας του σχετικά με το bagging. Είναι ένας αλγόριθμος εκμάθησης συνόλου που μπορεί να χειριστεί τόσο την ταξινόμηση υψηλής διάστασης όσο και την παλινδρόμηση. Η μέθοδος αυτή χρησιμοποιείται ευρέως σε πολλούς τομείς όπως η βιοστατιστική, τα χρηματοοικονομικά, στην παρακολούθηση του κλίματος και στην πρόγνωση του καιρού. Στον τομείς της οικονομίας δεν έχει εφαρμοστει σε σημαντικό βαθμό. [60]
Η RF είναι μια μέθοδος συνόλου που βασίζεται σε δέντρα όπου όλα εξαρτώνται από μία συλλογή τυχαίων μεταβλητών. Το συνολικό δάσος σχηματίζεται από πολλά δέντρα παλινδρόμησης που μαζί σχηματίζουν ένα σύνολο. Αφού εκπαιδευτούν τα μεμονωμένα δέντρα χρησιμοποιώντας δείγματα bootstrap, η τελική απόφαση λαμβάνεται από το σύνολο, με τον μέσο όρο της εξόδου, στην περίπτωση της παλινδρόμησης. Αυτή η διαδικασία, που ονομάζεται bagging, βελτιώνει τη σταθερότητα και την ακρίβεια του μοντέλου, μειώνει τη διακύμανση και βοηθά στην αποφυγή υπερπροσαρμογής (overfitting), μία 'συνήθεια' των δέντρων απόφασης.[60]
Το κατώφλι ($bias$) του συνόλου των δέντρων είναι το ίδιο με αυτό των μεμονωμένων δέντρων, αλλά η διακύμανση ελαττώνεται όταν μειώνεται η συσχέτιση μεταξύ των δέντρων. [60]
Το Random Forest παράγει μία περιορισμένη τιμή για το σφάλμα γενίκευσης. Το σφάλμα γενίκευσης εκτιμάται από το out-of-bag (OOB) σφάλμα. Όταν σχεδιάζεται το bootstrap δείγμα, μερικά δεδομένα παραλείπονται και δεν περιλαμβάνονται στο δείγμα. Αυτό ονομάζεται 'out-bag data¨. Η εκπαίδευση τερματίζεται όταν σταθεροποιηθεί το σφάλμα OOB. [60]

Τρόπος λειτουργίας του αλγορίθμου

Κατά τον αλγόριθμο αυτόν, επιλέγεται ένα σύνολο *bootstrap* L μεγέθους N από το σύνολο των δεδομένων. Στη συνέχεια για κάθε κόμβο του δέντρου επαναλαμβάνεται η εξής διαδικασία, μέχρι να επιτευχθεί το ελάχιστο μέγεθος κόμβων m:[60]

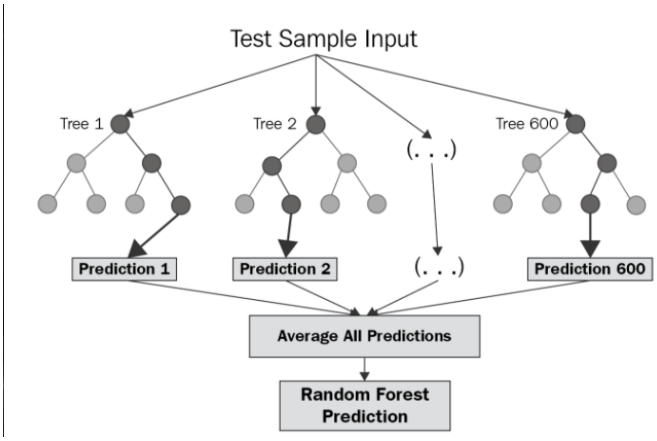
1. Επιλέγονται τυχαία F μεταβλητές από τις n συνολικές. Το F αντιπροσωπεύει το πλήθος των μεταβλητών εισόδου.

2. Διαλέγεται ο καλύτερος διαχωρισμός μεταξύ των F

3. Διαχωρίζεται ο κόμβος σε δύο παιδιά-κόμβους.

Η διαδικασία επαναλαμβάνεται για τα K δέντρα του συνόλου και στο τέλος προκύπτει το σύνολο των δέντρων T_k. Ο αριθμός K ορίζεται πειραματικά, καθώς η διαδικασία συνεχίζει μέχρι να σταθεροποιηθεί το σφάλμα OOB.[60]
Μετά από την παραπάνω εκπαίδευση, η πρόβλεψη για ένα σημείο x υπολογίζεται:

$$f(x) \;=\; \frac{1}{K}\sum_{k=1}^{K} T_k(x)$$



Σχήμα 20: Random Forest

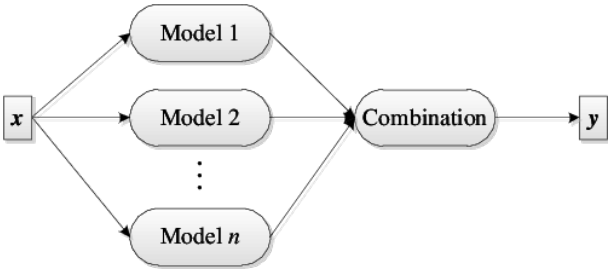
Η τυχαιότητας στο δάσος επιτυγχάνεται με δύο τρόπους: [61]

1. Το πρώτο μέρος της τυχαιότητας είναι ότι κάθε δέντρο τοποθετείται σε ένα ανεξάρτητο δείγμα *bootstrap* του αρχικού συνόλου δεδομένων.

2. Το δεύτερο μέρος προέρχεται από τη διάσπαση των κόμβων. Αντι σε κάθε διαχωρισμό να λαμβάνονται υπόψη όλες οι p μεταβλητές πρόβλεψης, χρησιμοποιούμε ένα τυχαίο υποσύνολο m προγνωστικών. Αυτό σημαίνει ότι τυχαία, διαφορετικές μεταβλητές πρόγνωσης βρίσκονται σε διαφορετικά δέντρα.

5.6 Μέθοδοι Ensemble

Η τεχνική Ensemble αντιμετωπίζει μια ομάδα στοιχείων ως σύνολο και όχι μεμονωμένα. Δημιουργεί πολλαπλά μοντέλα και τα συνδυάζει ώστε τελικά να προκύψουν τα καλύτερα αποτελέσματα. Οι μέθοδοι συνόλου βοηθούν στη βελτίωση της γενίκευσης του μοντέλου. Επιπλέον με την τεχνική αυτή αποφεύγονται τα ελαττώματα ενός μεμονομένου αλγορίθμουο ενώ επίσης μειώνει τον κίνδυνο επιλογής λάθος μοντέλου. Μάλιστα, μπορεί να αξιοποιήσει πλήρως τα πλεονεκτήματα ενός βασικού μοντέλου και να παρουσιάσει εξαιρετκή απόδοση πρόβλεψης. [43]



Σχήμα 21: Ensemble Diagram[27]

5.6.1 Bagging

Το Bagging ορίστηκε πρώτη φορά από τον Breiman το 1996, και προέρχεται από τις λέξεις bootstrap aggregating. Αποτελεί μία μέθοδο για τη βελτίωση των ασταθών σχημάτων εκτίμησης ή ταξινόμησης.[44]
Ο Breiman όρισε το bagging ως μία τεχνική μείωσης της διακύμανσης για διαδικασίες όπως είναι τα δέντρα απόφασης ή οι μέθοδοι που κάνουν επιλογή μεταβλητών και προσαρμογή σε γραμμικό μοντέλο. Έγινε ιδιαίτερα δημοφιλές λόγω της απλότητας της εφαρμογής και της δημοτικότητα της μεθοδολογίας bootstrap. [44]
Εκείνη την εποχή, όμως, παρουσιάστηκαν μόνο ευρετικά επιχειρήματα για τους λόγους που το bagging λειτουργούσε. Αργότερα, το 2002, αποδείχθηκε από τους Bühlmann και Yu ότι η τεχνική αυτή είναι μια λειτουργία εξομάλυνσης που αποδεικνύεται πλεονεκτική όταν στοχεύει στη βελτίωση της προγνωστικής απόδοσης των δέντρων παλινδρόμησης ή ταξινόμησης. Στην περίπτωση των δέντρων απόφασης, η θεωρία των δύο αυτών επιστημόνων, επιβεβαιώνει τη διαίσθηση του Breiman ότι το bagging είναι μια τεχνική μείωσης διακύμανσης, μειώνοντας επίσης το μέσο τετράγωνο σφάλμα (MSE).[44]
Η ιδέα του bagging είναι η επιθυμία να ταιριαχθούν πολλά ανεξάρτητα μοντέλα και να υπολογιστεί ο μέσος όρος των προβλέψεών τους προκειμένου να ληφθεί ένα μοντέλο με χαμηλότερη διακύμανση. Ωστόσο, δεν μπορούν, στην πράξη, να χωρέσουν πλήρως ανεξάρτητα μοντέλα γιατί θα απαιτούσε πάρα πολλά δεδομένα. Επομένως, δίνεται έμφαση στις καλές «κατά προσέγγιση ιδιότητες» της αντιπροσωπευτικότητας και της ανεξαρτησίας των δειγμάτων bootstrap ώστε να προσαρμοστούν μοντέλα που είναι σχεδόν ανεξάρτητα. [45]
Στην αρχή, δημιουργούνται πολλαπλά δείγματα bootstrap έτσι ώστε κάθε νέο δείγμα να λειτουργεί ως ένα άλλο σχεδόν ανεξάρτητο σύνολο δεδομένων που προέρχεται από το αρχικό σύνολο. Στη συνέχεια,

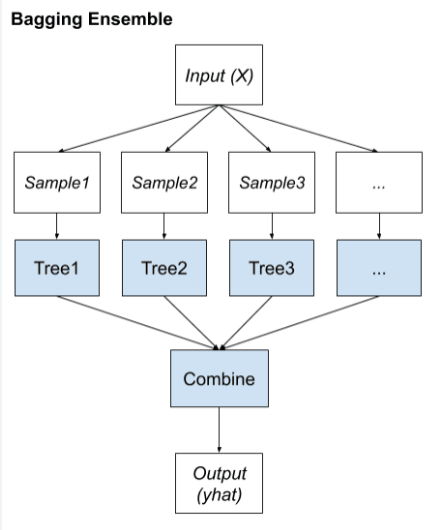
προσαρμόζεται ένας αδύναμος learner για καθένα από αυτά τα δείγματα και τελικά αυτά αθροίζονται υπολογίζοντας τον μέσο όρο των εξόδων τους και, έτσι, να δημιουργηθεί ένα μοντέλο συνόλου με μικρότερη απόκλιση.[45]
Πιο αναλυτικά, υποθέτοντας ότι υπάρχουν L bootstrap δείγματα μεγέθους B: [45]

$$\{z_1^1, z_2^1, ...z_B^1\}, \{z_1^2, z_2^2, ...z_B^2\}, ..., \{z_1^L, z_2^L, ...z_B^L\}$$

Υπάρχουν L σχεδόν ανεξάρτητοι αδύναμοι μαθητές (ένας σε κάθε σύνολο δεδομένων)

$$w_1(.), w_2(.), ..., w_L(.)$$

Στη συνέχεια συγκεντρώνονται οι έξοδοι σε ορισμένη διαδικασία υπολογισμού του μέσου όρου για να ληφθεί ένα μοντέλο συνόλου με μικρότερη απόκλιση. Υπάρχουν διάφοροι πιθανοί τρόποι για τη συγκέντρωση των πολλαπλών μοντέλων που έχουν τοποθετηθεί παράλληλα. [45]
Τέλος, θα λήταν σημαντικό να αναφερθεί ότι ένα από τα μεγάλα πλεονεκτήματα του bagging είναι ότι μπορεί να παραλληλιστεί. Τα διαφορετικά μοντέλα τοποθετούνται ανεξάρτητα το ένα από το άλλο και συνεπώς μπορούν να χρησιμοποιηθούν εντατικές τεχνικές παραλληλοποίησης εάν χρειάζεται. [45]



Σχήμα 22: Bagging

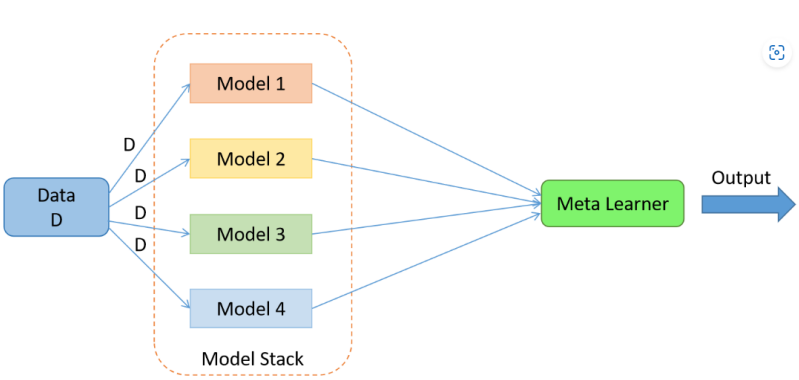
5.6.2 Μέθοδος υπολογισμού μέσου όρου

Ο μέσος όρος συνόλου (ή ο μέσος όρος) είναι η πιο απλή μέθοδος ensemble με σκοπό την εκμετάλλευση της πρόβλεψης διαφορετικών μοντέλων παλινδρόμησης. Αποτελεί μια κοινή και ευρέως χρησιμοποιούμενη μέθοδος συνόλου στα οποία οι προβλέψεις των μεμονομένων μοντέλων πρόβλεψης αντιμετωπίζονται εξίσου. Πιο αναλυτικά, κάθε μοντέλο πρόβλεψης εκπαιδεύεται ξεχωριστά και αφού γίνει η πρόβλεψη ανεξάρτητα από τα διαφορετικά μοντέλα, στο τέλος υπολογίζεται ο μέσος όρος της τελικής τιμής.

Ο μέσος όρος συνόλου βασίζεται στην ιδέα ότι τα συνιστώσα μοντέλα του δεν κάνουν συνήθως το ίδιο σφάλμα στα δεδομένα. Με αυτόν τον τρόπο, το μοντέλο συνόλου μειώνει τη διακύμανση στην πρόβλεψη, γεγονός που οδηγεί σε καλύτερες προβλέψεις. Τα πλεονεκτήματα αυτής της μεθόδου αποτελεί η απλότητα της εφαρμογής της καθώς και η εκμετάλλευση της ποικιλίας των σφαλμάτων των συνιστώσων μοντέλων της χωρίς να απαιτείται πρόσθετη εκπαίδευση. [46]

5.6.3 Στοίβαξη

Η στοίβαξη (stacking) είναι μια άλλη προσέγγιση ensemble που συνδυάζει διαφορετικά μοντέλα πρόβλεψης σε ένα μόνο μοντέλο. Αυτή η προσέγγιση εισάγει την έννοια της μετα-μάθησης. Αντιπροσωπεύει ένα ασυμπτωτικά βέλτιστο σύστημα μάθησης και στοχεύει στην ελαχιστοποίηση των σφαλμάτων. [48]
Η ιδέα είναι να δημιουργηθεί ένα μετα-σύνολο δεδομένων που περιέχει μια πλειάδα για κάθε πλειάδα στο αρχικό σύνολο δεδομένων. Ωστόσο, αντί να χρησιμοποιεί τα αρχικά χαρακτηριστικά εισαγωγής, χρησιμοποιεί την προβλεπόμενη ταξινόμηση των ταξινομητών ως γνωρίσματα εισόδου. [48]
Πιο αναλυτικά, κατά τη στοίβαξη τρέχουν μεμονωμένοι οι μαθητές (μαθητευόμενοι πρώτου επιπέδου). Το σύνολο εκπαίδευσης χρησιμοποιείται για την εκπαίδευση πρώτα από κάθε έναν από τους βασικούς ταξινομητές. Τα αποτελέσματα που προκύπτουν από την πρόβλεψη των μαθητών συνδιάζονται και δημιουργούν ένα νέο σύνολο το οποίο χρησιμοποιείται για να εκπαιδευτεί ο μεταμαθητής (μαθητή δεύτερου επιπέδου) ο οποίος συνδυάζει τις διαφορετικές προβλέψεις σε μια τελική.[47] Γενικά, το μοντέλο του επόμενου επιπέδου μαθαίνει από τα μοντέλα του προηγούμενου επιπέδου.



Σχήμα 23: Stacking

Τις περισσότερες φορές το αρχικό σύνολο δεδομένων χωρίζεται σε δύο υποσύνολα. Το πρώτο δεσμεύεται για να σχηματίσει το μετα-σύνολο δεδομένων και το δεύτερο υποσύνολο χρησιμοποιείται για τη δημιουργία των ταξινομητών βασικού επιπέδου. Συνεπώς, οι προβλέψεις των μετα-ταξινομητών αντικατοπτρίζουν την πραγματική απόδοση των αλγορίθμων μάθησης βασικού επιπέδου. [48]
Οι επιδόσεις στοίβαξης θα μπορούσαν να βελτιωθούν χρησιμοποιώντας πιθανότητες εξόδου για κάθε ετικέτα κλάσης από τους ταξινομητές βασικού επιπέδου. Σε τέτοιες περιπτώσεις, ο αριθμός των εισαγωγών πολλαπλασιάζεται ανάλογα τον αριθμό των κλάσεων. [48]

Συμπερασματικά, η κύρια βελτίωση στα προγνωστικά αποτελέσματα, όταν εφαρμόζεται το Stacking, είναι εμφανής όταν υπάρχει ποικιλομορφία μεταξύ των μοντέλων των διαφορετικών επιπέδων, δηλαδή μοντέλων με διαφορετικές στρατηγικές μάθησης. [48]

5.7 Σχετικές Εργασίες

Για την πρόβλεψη χρονοσειρών έχουν πραγματοποιηθεί πολλές μελέτες και έρευνες. Κυρίως έχει μελετηθεί η συμπεριφορά και η αποτελεσματικότητα διαφορετικών αλγορίθμων σε δεδομένα που αφορούν πολλούς τομείς της οικονομίας, της κοινωνίας και των επιστημών.

Ένα παράδειγμα αποτελεί η έρευνα των Sima Siami Namini και Akbar Siami Namin από το Τεχνολογικό Πανεπιστήμιο του Τέξας και της Neda Tavakoli από το Ινστιτούτο Τεχνολογίας της Georgia. Στόχος της μελέτης τους ήταν η κατανόηση για το αν και πώς οι αλγόριθμοι βασισμένοι σε βαθιά μάθηση, όπως είναι ο LSTM, παρουσιάζουν καλύτερα αποτελέσματα στην πρόβλεψη χρονοσειρών από τους παραδοσιακούς αλγόριθμους. Για αυτό, εφαρμόστηκε η μέθοδος ARIMA και η μέθοδος LSTM σε οικονομικά δεδομένα που είχαν συλλεχθεί από το Yahoo finance Website 1. Τα αποτελέσματα της έρευνας έδειξαν ότι οι αλγόριθμοι που βασίζονται σε βαθιά μάθηση όπως το LSTM ξεπερνά τους παραδοσιακούς αλγόριθμους όπως είναι το μοντέλο ARIMA. Πιο συγκεκριμένα, η μέση μείωση σε ποσοστά σφάλματος που παρατηρήθηκε από το LSTM ήταν μεταξύ 84 - 87% σε σύγκριση με το ARIMA, υποδεικνύοντας την μεγαλύτερη αποτελεσματικότητα του πρώτου προς το δεύτερο. [20]

Μία άλλη μελέτη του τομέα της μηχανικής πετρελέων του πανεπιστημιού του Wyoming και του Colorado School of Mines ασχολήθηκε με την πρόβλεψη της απόδοσης παραγωγής μη συμβατικών δεξαμενών λόγω της ετερογένειας των ιζημάτων, των περίπλοκων καναλιών ροής και της πολύπλοκης συμπεριφοράς της ρευστής φάσης. Οι παραδοσιακές μέθοδοι πρόβλεψης παραγωγής πετρελαίου, όπως για παράδειγμα η ανάλυση καμπύλης πτώσης και η προσωμοίωση reservoir για πρόβλεψη, είναι υποκειμενικές. Κατά την έρευνα πραγματοποιήθηκε σύγκριση των τριών αλγορίθμων: ARIMA, LSTM και Prophet. Στη συνέχεια, εφαρμόστηκαν για σύγκριση η ανάλυση της καμπύλης πτώσης και η προσωμοίωση reservoir για πρόβλεψη. [21]

Τα συμπεράσματα ήταν ότι τα μοντέλα μηχανικής μάθησης περιλαμβάνουν μια απλή ροή εργασίας, χωρίς προηγούμενη υπόθεση για τον τύπο της δεξαμενής, γρήγορη πρόβλεψη και αξιόπιστη απόδοση για μια τυπική κυμαινόμενη φθίνουσα καμπύλη. Επιπλέον, το μοντέλο Prophet κατέγραψε τις διακυμάνσεις της παραγωγής που προκαλούνται από τις χειμερινές επιπτώσεις, οι οποίες μπορούν να προσελκύσουν την προσοχή του χειριστή και να αποτρέψουν πιθανές αστοχίες. Κάτι τέτοιο δεν είχει διερευνηθεί και συζητηθεί ιδιαίτερα, από προηγούμενες μελέτες. Η εφαρμογή των μεθόδων έδειξε ότι οι ARIMA και LSTM αποδίδουν καλύτερα από το Prophet. Αυτό συμβαίνει καθώς δεν περιλαμβάνουν όλα τα δεδομένα παραγωγής πετρελαίου εποχιακές επιρροές. [21]

Ακόμα μία μελέτη σχετικά με σύγκριση χρονοσειρών πραγματοποιήθηκε από μία ομάδα ερευνητών σε δεδομένα εστιών της παθογόνου γρίπης των πτηνών (H5N1) στην Αίγυπτο. Εφαρμόστηκαν τα μοντέλα Random Forest και ARIMA. Παρατηρήθηκε ότι το μοντέλο Random Forest ξεπέρασε το ARIMA στην προγνωστική ικανότητα. Αυτό συνέβη πιθανόν για δύο λόγους σύμφωνα με τη μελέτη. Πρώτον, η σχέση μεταξύ των δεδομένων μπορεί να μην είναι γραμμική. Το μοντέλο ARIMA προϋποθέτει γραμμικές σχέσεις μεταξύ των προβλεπόμενων τιμών και αυτή η αδυναμία του μπορεί να έχει συμβάλει σε χαμηλότερη απόδοση. Δεύτερον, υπάρχει ένας μη φυσιολογικός θόρυβος στη χρονοσειρά κάτι που είναι μία από τις βασικές παραδοχές από το μοντέλο ARIMA. [22]

Υπάρχουν στην παγκόσμια βιβλιογραφία πολλές ακόμα μελέτες που συγκρίνουν διαφορετικούς αλγόριθ-

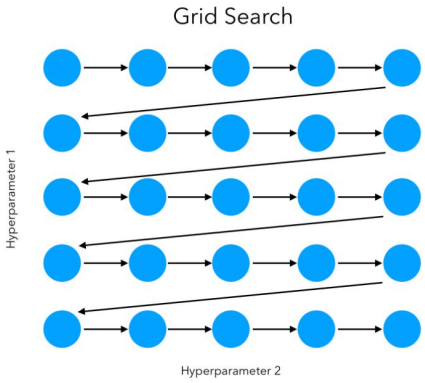
μους πρόβλεψης.

6 Μέθοδοι Βελτιστοποίησης

Βελτιστοποίηση είναι η διαδικασία της οποίας κύριος στόχος είναι η εύρεση της καλύτερης τιμής (μέγιστης ή ελάχιστης) μίας συνάρτησης $f(\theta_1, ..., \theta_m)$ με m παράμετρους $\theta_1, ..., \theta_m$. Οι παράμετροι αυτοί μπορεί να είναι περιορισμένοι (στην εκθετική κατανομή μπορούν να είναι μόνο θετικοί) είτε μη περιορισμένοι (γραμμική παλινδρόμηση).[49]
Ένας αποτελεσματικός βελτιστοποιητής είναι πολύ σημαντικός για την επιβεβαίωση ότι η καλύτερη λύση είναι προσβάσιμη. Η βάση ενός βελτιστοποιητή είναι ένας αλγόριθμος αναζήτησης ή βελτιστοποίησης που εφαρμόζεται σωστά ώστε να πραγματοποιηθεί η επιθυμητή αναζήτηση. Υπάρχουν στη βιβλιογραφία πολλοί αλγόριθμοι βελτιστοποίησης, όμως κανένας δεν είναι κατάλληλος για όλα τα προβλήματα.[50]

6.1 Αναζήτηση πλέγματος

Το grid search ή αναζήτηση πλέγματος κάνει μια πλήρη αναζήτηση σε ένα δεδομένο υποσύνολο του χώρου των υπερπαραμέτρων του αλγορίθμου εκπαίδευσης. [51]
Πιο συγκεκριμένα η μέθοδος αυτή ¨κατασκευάζει' ένα πλέγμα, αξιολογεί την αντικειμενική συνάρτηση σε όλα τα σημεία του πλέγματος και βρίσκει το σημείο που επιτυγχάνει την καλύτερη επιθυμητή τιμή της συνάρτησης.[51]
Για παράδειγμα αν μία μεταβλητή έχει ελάχιστη τιμή l_i και μέγιστη u_i τότε χωρίζει το διάστημα (l_i, p_i) σε $p_i - 1$ ίσα σημεία ώστε τα $x_i^{(1)}, x_i^{(2)}, ..., x_i^{p_i}$ να είναι σημεία στον x_i άξονα $(i = 1, 2, ..., n)$, όπου n το σύνολο των μεταβλητών.[51]
Επειδή ένας αλγόριθμος μηχανικής μάθησης μπορεί να περιλαμβάνει χώρους με πραγματικές ή απεριόριστες τιμές για ορισμένες παραμέτρους, είναι δυνατό να πρέπει να καθοριστεί ένα όριο για να εφαρμοστεί μια αναζήτηση πλέγματος. Η αναζήτηση πλέγματος περιλαμβάνει πολλές φορές χώρους υψηλών διαστάσεων, αλλά συχνά μπορεί εύκολα να παραλληλιστεί, καθώς οι τιμές των υπερπαραμέτρων με τα οποία λειτουργεί ο αλγόριθμος είναι συνήθως ανεξάρτητες μεταξύ τους. Επιπλέον, ο αλγόριθμος μπορεί να χρησιμοποιηθεί ώστε να βρεθεί ένα καλό αρχικό σημείο για μία ή περισσότερες μεθόδους.[52]

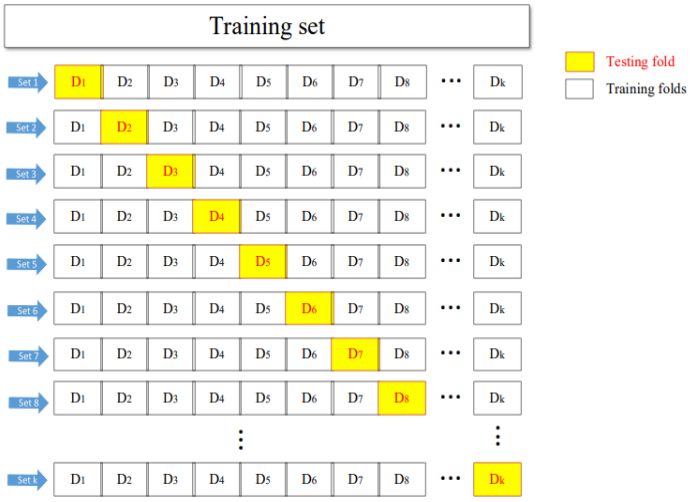


Σχήμα 24: Grid Search[53]

6.2 K-Fold

Η διασταυρούμενη επικύρωση k-fold ξεκινάει με την τυχαία διάσπαση των δεδομένων της σε K ομάδες. Στη συνέχεια, για κάθε ομάδα εκτελούνται οι ακόλουθες πράξεις[54]:

- Επιλογή μίας από τις ομάδες ώστε να γίνει το σύνολο που θα χρησιμοποιηθεί για τον έλεγχο της ακρίβειας του μοντέλου (test set
- Χρήση των υπόλοιπων K-1 ομάδων για την εκπαίδευση του μοντέλου (training sets)
- Πραγματοποίηση της εκπαίδευσης και αξιολόγηση του μοντέλου.



Σχήμα 25: K-Fold

Η μέθοδος K-Fold για την αξιολόγηση εκτιμά το συνολικό σφάλμα ως το μέσο όρο σφαλμάτων κάθε ομάδας. Συνεπώς, η μέθοδος εξαρτάται από δύο παράγοντες: το σύνολο που χρησιμοποιείται για την εκπαίδευση και τον διαχωρισμό σε ομάδες.[55]

7 Μετρικές παράμετροι σχετικά με την αξιολόγηση των δεδομένων

Για την αξιολόγηση των προβλεπόμενων τιμών θα χρησιμοποιηθούν ορισμένες μετρικές παράμετροι, οι οποίοι θα βοηθήσουν ώστε να γίνει κατανοητό το πόσο μακριά σε σχέση με τις πραγματικές τιμές είναι οι προβλέψεις.[65]
Θεωρείται ως:

- **MSE**: Δηλώνει πόσο κοντά είναι μία γραμμή παλινδρόμησης από ένα σύνολο σημείων, υπολογίζοντας την απόστασή τους. Στη συνέχεια αυτή η απόσταση υψώνεται στο τετράγωνο. Όσο μικρότερη είναι, δηλαδή όσο πιο κοντά στο μηδέν είναι το σφάλμα, τόσο καλύτερο θεωρείται το μοντέλο.

$$MSE(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$$

- **MAE**: Υπολογίζει το μέσο μέγεθος των σφαλμάτων σε ένα σύνολο προβλέψεων. Όσο μικρότερη είναι, δηλαδή όσο πιο κοντά στο μηδέν είναι το σφάλμα, τόσο καλύτερο θεωρείται το μοντέλο.

$$MAE(y, \hat{y}) = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i|^2$$

- **RMSE**: Υπολογίζει το μέσο μέγεθος των σφαλμάτων σε ένα σύνολο προβλέψεων. Όσο μικρότερη είναι, δηλαδή όσο πιο κοντά στο μηδέν είναι το σφάλμα, τόσο καλύτερο θεωρείται το μοντέλο.

$$RMSE(y, \hat{y}) = \sqrt{(1/n\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}$$

- **R-Squared**: Υπολογίζει το ποσοστό της διακύμανσης. Όσο μικρότερη είναι, δηλαδή όσο πιο κοντά στο μηδέν είναι το σφάλμα, τόσο καλύτερο θεωρείται το μοντέλο.

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^{n} y_i - \hat{y}_i{}^2}{\sum_{i=1}^{n} y_i - \overline{y}^2}$$

8 Επεξεργασία των δεδομένων και εφαρμογή των αλγόριθμων πρόβλεψης στην Python

Στο δεύτερο μέρος θα παρουσιαστεί ο τρόπος επεξεργασίας των δεδομένων πριν προωθηθούν στα μοντέλα πρόβλεψης και στη συνέχεια θα περιγραφούν οι συναρτήσεις που χρησιμοποιήθηκαν από τα μοντέλα ώστε να γίνει η τελική πρόβλεψη.

8.1 Επεξεργασία δεδομένων

Στην αρχή έγινε η φόρτωση των δεδομένων μέσω του *google drive*. Τα δεδομένα έχουν την μορφή πίνακα, ο οποίος περιλαμβάνει πολλά dataset ενωμένα.

	0	1	2	3	4
0	2	'2016-04-08T00:00:00.0000000'	22.608437	NULL	96
1	2	'2016-04-09T00:00:00.0000000'	23.285104	NULL	96
2	2	'2016-04-10T00:00:00.0000000'	22.444375	NULL	96
3	2	'2016-04-11T00:00:00.0000000'	20.529479	NULL	96
4	2	'2016-04-12T00:00:00.0000000'	19.844479	NULL	96
...	...	...	...	...	...
92696930	143534	'2022-06-12T00:00:00.0000000'	0.0	NULL	NULL
92696931	143534	'2022-06-13T00:00:00.0000000'	0.0	NULL	NULL
92696932	143534	'2022-06-14T00:00:00.0000000'	0.0	NULL	NULL
92696933	143534	'2022-06-15T00:00:00.0000000'	0.0	NULL	NULL
92696934	143534	'2022-06-16T00:00:00.0000000'	0.272727	NULL	NULL

92696935 rows × 5 columns

Σχήμα 26: Dataset

Η 1η στήλη περιέχει τον κωδικό που περιγράφει σε ποιό dataset ανήκει η κάθε γραμμή, η 2η περιλαμβάνει τις ημερομηνίες που καταγράφηκαν τα δεδομένα και η 3η τις τιμές των μετρήσεων για αυτές τις ημερομηνίες. Οι υπόλοιπες στήλες δεν περιέχουν κάποια σημαντική πληροφορία για την εργασία, οπότε και αφαιρέθηκαν

Αφού μετατράπηκε σε τύπο float η 1η στήλη, επιλέχθηκαν από τα δεδομένα οι γραμμές με τον αντίστοιχο κωδικό που μας ενδιέφερε. Επιπλέον μετατράπηκε η 2η στήλη σε τύπο datetime και ορίστηκε ως το index.

Είναι συχνό, λόγω σφάλματος του οργάνου μέτρησης ή για κάποιον άλλο λόγω να υπάρξουν στα δεδομένα ορισμένες χρονικές στιγμές που δεν υπάρχουν μετρήσεις. Για αυτόν τον λόγο ήταν σημαντικό στην αρχή να γίνει έλεγχος αν υπάρχουν *null* τιμές στα δεδομένα και στη συνέχεια αυτές να καλυφθούν. Προέκυψε ότι δεν υπάρχουν σε κανένα από τα σύνολα που χρησιμοποιήθηκαν οπότε δεν ήταν απαραίτητη κάποια αλλαγή. Επιπλέον ορίστηκε η συνάρτηση $timeseries\_evaluation\_metrics\_func$ η οποια χρησιμοποιήθηκε από όλους τους αλγορίθμους για να υπολογιστούν τα σφάλματα μεταξύ των

τιμών που προέκυψαν μετά την πρόβλεψη και των πραγματικών τιμών. Οι μετρικές παράμετροι που χρησιμοποιήθηκαν ήταν το $MSE, MAE, RMSE$ καιτο R^2.
Μετά την επεξεργασία αυτή, για το κάθε μοντέλο πρόβλεψης ο κώδικας διέφερε.

8.2 Στατιστικοί Αλγόριθμοι

Στον εκθετικό αλγόριθμο και στους αλγόριθμους $ARIMA$, στην αρχή, αφού ορίστηκαν τα μεγέθη του $test\_size$ και του $train\_size$, δημιουργήθηκε το $train$ και το $test\ set$. Το πρώτο σύνολο περιλαμβάνει τα δεδομένα που θα εκπαιδευτούν, ενώ το δεύτερο περιέχει τα δεδομένα που θα χρησιμοποιηθούν για τον έλεγχο της ακρίβειας πρόβλεψης του αλγορίθμου. Το μέγεθος του $test\ set$ ορίστηκε 30, δηλαδή θα προβλεφθούν οι 30 τελευταίες τιμές του αρχικού συνόλου δεδομένων.

8.2.1 SARIMA

Η πρόβλεψη με για τους αλγόριθμους ARIMA και SARIMA έγινε με τη βοήθεια της συνάρτησης $auto\_arima$ της βιβλιοθήκης $pmdarima$. Στον αλγόριθμο $SARIMA$, ως ορίσματα στις μεταβλητές p,q στη συνάρτηση ορίστηκαν οι τιμές από 1 έως 7 και στη μεταβλητή m οι τιμές 1,4,7,12 ανάλογα με το αν η εποχικότητα είναι ανά χρόνο, ανά τρίμηνο, ανά ημέρα ή ανά μήνα. Επιπλέον ορίστηκε η μεταβλητή $seasonal = True$.
Στη συνέχεια για κάθε m έγινε η πρόβλεψη με όλους τους συνδυασμούς των μεταβλητών p και q, εμφανίζοντας στο τέλος, για το κάθε m, το πιο ¨καλό' μοντέλο ως προς τα σφάλματα και η αντίστοιχη γραφική παράσταση που συγκρίνει την πρόβλεψη με την ¨αληθινή' τιμή.
Αντίστοιχη διαδικασία ακολουθήθηκε και για τον αλγόριθμο $ARIMA$, όμως η μεταβλητή $seasonal$ ορίστηκε $False$ και δεν υπήρχε η μεταβλητή m καθώς το μοντέλο δεν αναγνωρίζει την εποχικότητα.

8.2.2 Exponential Smoothing

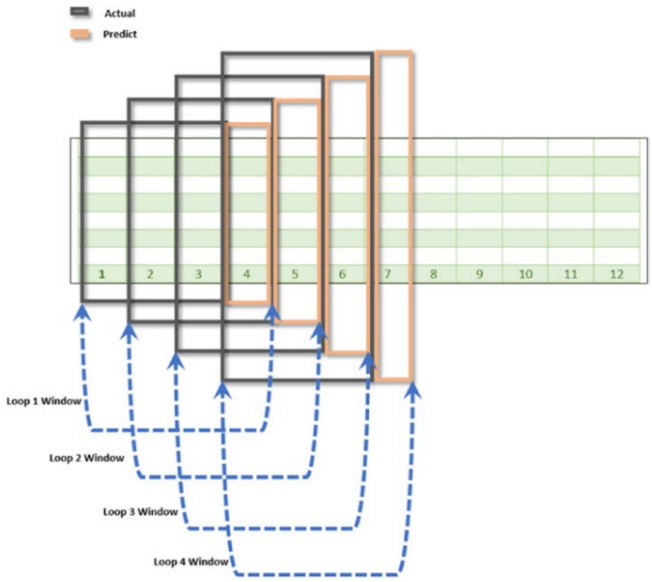
Για τον αλγόριθμο Exponential Smoothing κατασκευάστηκαν τρείς συναρτήσεις.
Η πρώτη με όνομα $best\_smoothing\_parametres()$ δοκιμάζει για τις μεταβλητές του αλγορίθμου όλες τις πιθανές τιμές ώστε στο τέλος, με βάση τα σφάλματα που προκείπτουν, να επιλεχθεί ο καλύτερος συνδυασμός.
Η δεύτερη συνάρτηση δέχεται ως όρισμα τις παραμέτρους που απαιτούνται για τη διαδικασία της πρόβλεψης και στη συνέχεια με βάση αυτές κάνει την απαραίτητη πρόβλεψη με χρήση της συνάρτησης $ExponentialSmoothing$ της αντίστοιχης βιβλιοθήκης.
Η τρίτη συναρτηση προβλέπει από μόνη της τις καλύτερους παραμέτρους και με βάση αυτές κάνει την αντίστοιχη πρόβλεψη.
Στη συνέχεια, με βάση τις παραπάνω συναρτήσεις αφού επιλέχθηκε από την πρώτη συνάρτηση ο κατάλληλος συνδυασμός των παραμέτρων, εφαρμόστηκε στα δεδομένα ο αλγόριθμος και στο τέλος εμφανίστηκαν τα σφάλματα και η γραφική παράσταση που συγκρίνει την πρόβλεψη με την ¨αληθινή' τιμή.

8.3 Νευρωνικά Δίκτυα

Στα Νευρωνικά δίκτυα, η προετοιμασία των δεδομένων της χρονοσειράς,πριν την διαδικασία της πρόβλεψης, ξεκίνησε με το $rescale$ των τιμών τους από 0 έως 1 για καλύτερα αποτελέσματα. Η μέθοδος κανονικοποίησης που χρησιμοποιήθηκε ήταν η $Min - Max\ Normalization$ κατα την οποία μετασχηματίζονται οι τιμές της χρονοσειρές γραμμικά σύμφωνα με τον τύπο:

$$X_{new} = \frac{X - min(X)}{max(X) - min(X)}$$

όπου:
X_{new} = Η νέα τιμή μετά την κανονικοποίηση
X = Η παλιά τιμή
$min(X)$ = Η ελάχιστη τιμή του συνόλου X
$max(X)$ = Η μέγιστη τιμή του συνόλου X
Στη συνέχεια τα δεδομένα χωρίστηκαν σε δύο ομάδες. Η συνάρτηση που χρησιμοποιήθηκε είχε ως ορίσματά της το σύνολο των δεδομένων και το $window\ size$. Με βάση το μέγεθος παραθύρου δημιουργούνται δύο μικρότερα σύνολα το X και το y. Στο πρώτο μπαίνουν με τη μορφή λίστας τόσα δεδομένα από το σύνολο όσο το μέγεθος παραθύρου και στη συνέχεια στο δεύτερο σύνολο μπαίνει η αμέσως επόμενη τιμή. Η διαδικασία αυτή επαναλαμβάνεται μέχρι να τοποθετηθούν όλες οι τιμές σε ένα από τα δύο σύνολα. Η παραπάνω συνάρτηση χρησιμοποιείται καθώς η πρόβλεψη γίνεται όπως φαίνεται στο σχήμα 12.



Σχήμα 27: Forecasting with Neural Networks[11]

Χρησιμοποιείται το σύνολο X για να προβλεφθούν οι τιμές του συνόλου y. Στην αρχή οι τιμές της λίστας στην πρώτη θέση του X θα προβλέψουν την τιμή του πρώτου στοιχείου του y. Και θα ακολουθήσουν όλες οι τιμές. Συνολικά θα έχουν προβλεφθεί $length$ (df) - $window\ size$ τιμές καθώς οι πρώτες $window\ size$ τιμές χρησιμοποιούνται μόνο για την πρώτη πρόβλεψη.
Συνεπώς, στην προετοιμασία των δεδομένων χρησιμοποιείται η συνάρτηση αυτή και δημιουργούνται τα σύνολα $X1$ και $y1$.
Στη συνέχεια ορίζονται τα μεγέθη που θα έχουν τα σύνολα $test$, $validation$ και $train$ και κατασκευάζονται.Τα δεδομένα από το $train\ set$ χρησιμοποιούνται για την εκπαίδευση του μοντέλου, το $validation\ set$ απαιτείται για την βελτιστοποίηση των παραμέτρων και τον ορισμό των βαρών και με τη βοήθεια του $test\ set$ θα ελεγχθεί η ακρίβεια του μοντέλου. Το $test\_size$ ορίζεται 30 καθώς πρέπει να προβλεφθούν οι 30 τελευταίες τιμές, ενώ το υπόλοιπο χωρίζεται στο $test\_size$ και στο val_size.
Το κάθε μοντέλο κατασκευάστηκε ξεχωριστά και έγινε $compile$ με βελτιστοποίηση που χρησιμοποιεί τον αλγόριθμο $Adam$, έναν στοχαστικό αλγόριθμο που χρησιμοποιει πρώτης τάξης $gradient$ και έχει μικρές απαιτήσεις μνήμης. Επίσης ως συνάρτηση απώλειας ορίστηκε το mse.
Ακολούθως, τοποθετήθηκαν οι τιμές των μεταβλητών που αντιστοιχούν στο $evaluation\ interval$ και στις εποχές και αφού εκπαιδεύτηκε το μοντέλο, έγιναν οι κατάλληλες προβλέψεις για το $train\ set$, το $validation\ set$ και το $test\ set$. Στο τέλος υπολογίστηκε το σφάλμα της πρόβλεψης και παρουσιάστηκε η γραφική παράσταση. Το κάθε μοντέλο, ανάλογα με τον αλγόριθμο, κατασκευάστηκε με διαφορετικά επίπεδα (layer).

8.4 Prophet

Για την πρόβλεψη με τον αλγόριθμο $Prophet$ χρησιμοποιήθηκε η συνάρτηση $Prophet$ της βιβλιοθληκης $prophet$. Στην επεξεργασία των δεδομένων πριν την διαδικασία της πρόβλεψης, έπρεπε να γίνουν αλλαγές στο όνομα του τίτλου της στήλης που είναι τα δεδομένα ώστε να γίνονται αποδεκτά απο την συνάρτηση. Η στήλη με τις ημερομηνίες ορίστηκε ως ds, ενώ η στήλη των τιμών ως y.
Αμέσως μετά ορίστηκαν τα σύνολα $train$ και $validate$. Στο πρώτο σύνολο, που με βάση αυτό έγινε η πρόβλεψη, προστέθηκαν όλα τα δεδομένα εκτός από τις τελευταίες 30 γραμμές, που χρησιμοποιήθηκαν για την αξιολόγηση του αλγορίθμου.
Το $train\ set$ εκπαιδεύτηκε με τη συνάρτηση $Prophet$. Με βάση την εκπαίδευση έγινε και η αντίστοιχη πρόβλεψη. Στο τέλος παρουσιάστηκαν τα σφάλματα και η γραφική παράσταση.
Τα ορίσματα της συνάρτησης $Prophet$ είναι:

- $growth$ που ρυθμίζει το $trend$ της σειράς και μπορεί να πάρει τις τιμές $flat$ αν δεν εμφανίζει κάποια τάση, $linear$ αν είναι γραμμική και $logistic$ αν ακολουθεί την συνάρτηση $\frac{L}{1+e^{-k(x-x_0)}}$, όπου L η μέγιστη τιμή, k ο ρυθμός αύξησης της καμπύλης και x_0 το μέσο στον άξονα των x.

- $seasonality\_mode$ που ορίζει αν η εποχικότητα είναι αθροιστική ή πολλαπλασιαστική

- $yearly\_seasonality$, $daily\_seasonality$ και $weekly\_seasonality$ για τις χρόνιες, ημερήσιες και εβδομαδιαίες εποχικότητες αντίστοιχα. Μπορούν να πάρουν τις τιμές $False$ αν δεν εμφανίζουν τα δεδομένα κάποια ή κάποιον ακέραιο αριθμό. Ο αριθμός αντιστοιχεί στο πλήθος των ημιτονοειδών

κυμματων που θα χρησιμοποιηθούν από τη συνάρτηση για την προσέγγιση των τιμών. Εξ ορισμού η τιμή που έχουν είναι 10, 3 και 4 αντίστοιχα.

- *seasonality_prior_scale* που περιγράφει την πιθανή κατανομή των τιμών πριν την πρόβλεψη. Εξ ορισμού του η τιμή είναι 10.
- *changepoint_prior_scale* που ορίζεται ως ο πιθανός αριθμός αλλαγών των τάσεων.

Σε ορισμένες περιπτώσεις έγινε η πρσθήκη άλλων παραγόντων εποχικότητας όπου ορίζοταν το όνομα της εποχικότητας και η περίοδος της.Επιπλέον προστέθηκε η εντολή κατά την οποία λαμβάνονται υπόψην οι διακοπές στην Ελλάδα εκείνο το μήνα. Η τελική πρόβλεψη των μελλοντικών τιμών έγινε με τη συνάρτηση $make\_future\_dataframe$. Το όρισμα *periods* δηλώνει πόσες περίδοι θα προβλεφθούν ενώ το $freq$ το είδος των περιόδων. Στη συγκεριμένη περίπτωση ορίστηκε ως D οπότε είναι ημερίσιο, και γι αυτό οι περίοδοι ήταν 30 (30 μέρες να προβλεφθούν).

8.5 XGBOOST

Για τον αλγόριθμο $XGBOOST$ χρησιμοποιήθηκε η συνάρτηση $xgboost$ της βιβλιοθηκης xgb.
Αρχικά κατασκευάστηκε η συνάρτηση κατά την οποία ¨σπάει' η χρονική στιγμή των δεδομένων σε ώρες, ημέρα της εβδομάδας, τρίμηνο, μήνα, χρονιά, ημέρα της χρονιάς και ημέρα του μήνα.
Στη συνέχεια, εφαρμόστηκε η παραπάνω συνάρτηση στο αρχικό *dataset* και σε εκείνο των *train* και των *test*.
Στο τέλος με τη βοήθεια της συνάρτησης $xgboost$ έγιναν οι κατάλληλες προβλέψεις και προβλήθηκαν τα σφάλματα και η αντίστοιχη γραφική παράσταση.

8.6 Random Forest Regression

Για την πρόβλεψη με τον αλγόριθμο Random Forest, στην αρχή δημιουργήθηκαν τα δύο σύνολα *train* και *test*.
Στη συνέχεια με τη χρήση της συνάρτηση $RandomForestRegressor()$ της βιβλιοθήκης $RandomForestRegressor$ εκπαιδεύτηκε το μοντέλο. Στη συνάρτηση ορίστηκε ο αριθμός των εκτιμητών μέσα από τη μεταβλητή $n\_estimators$ στους 1000.
Στο τέλος έγινε η κατάλληλη πρόβλεψη.

8.7 Παραλλαγή Random Forest Regression

Στην αρχή τα δεδομένα με τη συνάρτηση *series_to_supervised* διασπάστηκαν σε ύπο - ομάδες που αποτελούνταν από n τιμές των δεδομένων και από την επόμενη τιμή την οποία θα προβλέψουν κατά την εκπαίδευση του μοντέλου.
Στη συνέχεια, αφού χωρίστηκαν τα δεδομένα στα σύνολα train και test set, όπως και στην περίπτωση του απλού μοντέλου Random Forest, εκπαιδεύτηκε το μοντέλο με τη συνάρτηση $RandomForestRegressor()$ της βιβλιοθήκης $RandomForestRegressor$.

Στο τέλος, με τη βοήθεια της ίδιας συνάρτησης, έγιναν οι κατάλληλες προβλέψεις και προβλήθηκαν τα σφάλματα και η αντίστοιχη γραφική παράσταση.

9 Δεδομένα από καταστήματα Τραπεζών

Στο μέρος αυτό, θα γίνει παρουσιαση των δεδομένων από τα καταστήματα τραπεζών πάνω στα οποία θα γίνει η εφαρμογή των αλγορίθμων που προαναφέρθηκαν.

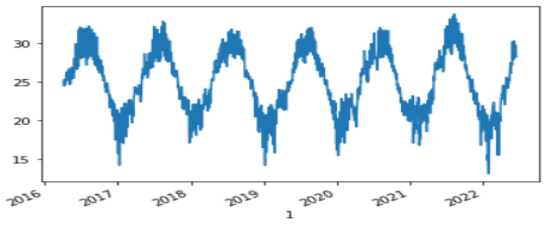
9.1 Κατάστημα 1ο

Το πρώτο κατάστημα που μελετήθηκε είναι εκείνο με κωδικό 10.

9.1.1 Controller Schneider > Groundfloor > Temperature2

Η μέτρηση αυτή αφορά την θερμοκρασία στο ισόγειο και μετρήθηκε με *Controller Schneider*. Στο παρακάτω διάγραμμα είναι φανερό, ότι η χρονοσειρά δεν παρουσιάζει ορισμένη τάση, όμως έχει έντονη εποχικότητα και κυρίως ανά χρόνο.

(a)



(b)

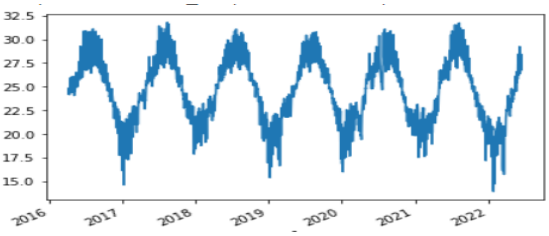
2016-04-07	3696.0	'2016-04-07T00:00:00.0000000'	25.257843
2016-04-08	3696.0	'2016-04-08T00:00:00.0000000'	24.457917
2016-04-09	3696.0	'2016-04-09T00:00:00.0000000'	25.009792
2016-04-10	3696.0	'2016-04-10T00:00:00.0000000'	25.094688
2016-04-11	3696.0	'2016-04-11T00:00:00.0000000'	24.675833
...	...	...	...
2022-06-12	3696.0	'2022-06-12T00:00:00.0000000'	29.136316
2022-06-13	3696.0	'2022-06-13T00:00:00.0000000'	29.482737
2022-06-14	3696.0	'2022-06-14T00:00:00.0000000'	29.714105
2022-06-15	3696.0	'2022-06-15T00:00:00.0000000'	28.201158
2022-06-16	3696.0	'2022-06-16T00:00:00.0000000'	28.205405

Σχήμα 28: Temperature 2 of GroundFloor
(a)Graph (b) Table

9.1.2 Controller Schneider > Groundfloor > Temperature1

Η μέτρηση αυτή αφορά την θερμοκρασία στο ισόγειο σε άλλο σημείο από προηγουμένως και μετρήθηκε με *Controller Schneider*. Όπως και στο προηγούμενο διάγραμμα που αφορούσε θερμοκρασία του ισόγειου, φαίνεται ότι η χρονοσειρά εμφανίζει μόνο εποχικότητα.

(a)



(b)

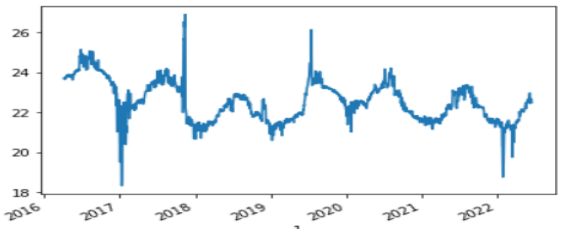
2016-04-07	3701.0	'2016-04-07T00:00:00.0000000'	24.712941
2016-04-08	3701.0	'2016-04-08T00:00:00.0000000'	24.195625
2016-04-09	3701.0	'2016-04-09T00:00:00.0000000'	24.529792
2016-04-10	3701.0	'2016-04-10T00:00:00.0000000'	24.872604
2016-04-11	3701.0	'2016-04-11T00:00:00.0000000'	24.356562
...	...	...	...
2022-06-12	3701.0	'2022-06-12T00:00:00.0000000'	27.87979
2022-06-13	3701.0	'2022-06-13T00:00:00.0000000'	28.378316
2022-06-14	3701.0	'2022-06-14T00:00:00.0000000'	28.429787
2022-06-15	3701.0	'2022-06-15T00:00:00.0000000'	26.757021
2022-06-16	3701.0	'2022-06-16T00:00:00.0000000'	27.001389

Σχήμα 29: Temperature 1 of GroundFloor
(a)Graph (b) Table

9.1.3 Controller Schneider > Data Room > DataRoom Temp

Η παρακάτω χρονοσειρά περιγράφει την θερμοκρασία στο *data room* και μετρήθηκε με *Controller Schneider*. Το διάγραμμα παρουσιάζει τη χρονοσειρά χωρίς κάποια εμφανή τάση αλλα με έντονη εποχικότητα. Η εποχικότητα είναι κυρίως ανά χρόνο, με μερικές διαφοροποιήσεις. Για παράδειγμα κάποιες χρονιές εμφανίζει έντονες πτώσεις, ενώ άλλες έντονες αυξήσεις.

(a)



(b)

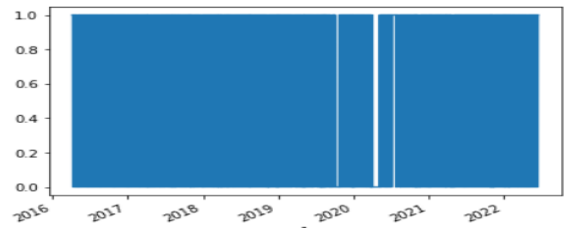
2016-04-07	3706.0	'2016-04-07T00:00:00.0000000'	23.686667
2016-04-08	3706.0	'2016-04-08T00:00:00.0000000'	23.661354
2016-04-09	3706.0	'2016-04-09T00:00:00.0000000'	23.742813
2016-04-10	3706.0	'2016-04-10T00:00:00.0000000'	23.749479
2016-04-11	3706.0	'2016-04-11T00:00:00.0000000'	23.682188
...	...	...	...
2022-06-12	3706.0	'2022-06-12T00:00:00.0000000'	22.607396
2022-06-13	3706.0	'2022-06-13T00:00:00.0000000'	22.616667
2022-06-14	3706.0	'2022-06-14T00:00:00.0000000'	22.668842
2022-06-15	3706.0	'2022-06-15T00:00:00.0000000'	22.519583
2022-06-16	3706.0	'2022-06-16T00:00:00.0000000'	22.47973

Σχήμα 30: Temperature of Data Room
(a)Graph (b) Table

9.1.4 Controller Schneider > Feedback > AC Feedback

Η μέτρηση αυτή υπολογίζει τη μέγιστη τιμή του *feedback* σε ένα *AC Feedback*. Η μέτρηση έγινε και πάλι με το όργανο *Controller Schneider*. Παρατηρείται ότι οι τιμές μεταβάλονται συνεχώς από 1 σε 0 με έντονη εποχικότητα.

(a)



(b)

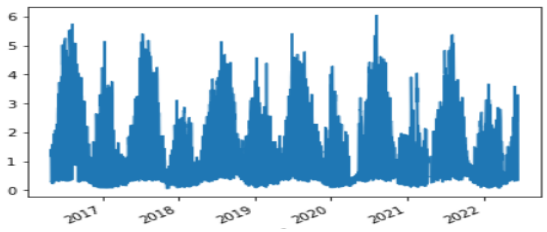
2016-04-07	3766.0	'2016-04-07T00:00:00.0000000'	1.0
2016-04-08	3766.0	'2016-04-08T00:00:00.0000000'	1.0
2016-04-09	3766.0	'2016-04-09T00:00:00.0000000'	1.0
2016-04-10	3766.0	'2016-04-10T00:00:00.0000000'	0.0
2016-04-11	3766.0	'2016-04-11T00:00:00.0000000'	0.0
...	...	...	...
2022-06-12	3766.0	'2022-06-12T00:00:00.0000000'	0.0
2022-06-13	3766.0	'2022-06-13T00:00:00.0000000'	0.0
2022-06-14	3766.0	'2022-06-14T00:00:00.0000000'	0.0
2022-06-15	3766.0	'2022-06-15T00:00:00.0000000'	1.0
2022-06-16	3766.0	'2022-06-16T00:00:00.0000000'	1.0

Σχήμα 31: Feedback of AC Feedback
(a)Graph (b) Table

9.1.5 Main Electricity Panel > HVAC > Active Power

Η μέτρηση αυτή αφορά την τιμή της *Active Energy* όσον αφοά τη θέρμανση, τον εξαερισμό και τον κλιματισμό από το όργανο *MainElectricityPanel*. Το διάγραμμα είναι αρκετά πυκνό, χωρίς κάποια τάση και με εποχικότητα. Οι τιμές μεταβάλλονται μεταξύ μίας μέγιστης και μίας ελάχιστης τιμής με ορισμένες διαφοροποιήσεις.

(a)



(b)

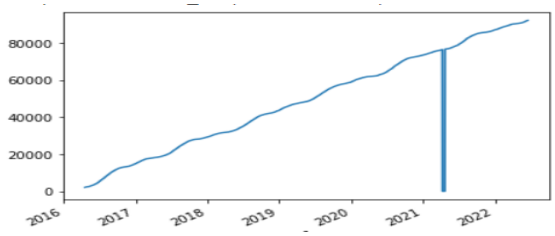
2016-04-26	3837.0	'2016-04-26T00:00:00.0000000'	1.178488
2016-04-27	3837.0	'2016-04-27T00:00:00.0000000'	1.410934
2016-04-28	3837.0	'2016-04-28T00:00:00.0000000'	1.373861
2016-04-29	3837.0	'2016-04-29T00:00:00.0000000'	1.267897
2016-04-30	3837.0	'2016-04-30T00:00:00.0000000'	0.283768
...	...	...	...
2022-06-12	3837.0	'2022-06-12T00:00:00.0000000'	0.353277
2022-06-13	3837.0	'2022-06-13T00:00:00.0000000'	0.328929
2022-06-14	3837.0	'2022-06-14T00:00:00.0000000'	0.354981
2022-06-15	3837.0	'2022-06-15T00:00:00.0000000'	3.300165
2022-06-16	3837.0	'2022-06-16T00:00:00.0000000'	2.419221

Σχήμα 32: Active Power of HVAC
(a)Graph (b) Table

9.1.6 Max calculation HVAC > Active Energy > Active Energy

Η μέτρηση αυτή παρουσιάζει την μέγιστη τιμή της *Active Energy* της θέρμανσης, του εξαερισμού και του κλιματισμού. Παρατηρείται αυξητική τάση με μικρή εποχικότητα. Το χρόνο 2021, μάλιστα, υπάρχει μία ξαφνική πτώση.

(a)



(b)

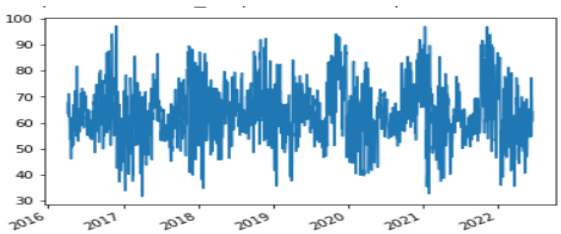
2016-04-19	3851.0	'2016-04-19T00:00:00.0000000'	2102.003906
2016-04-20	3851.0	'2016-04-20T00:00:00.0000000'	2145.173096
2016-04-21	3851.0	'2016-04-21T00:00:00.0000000'	2193.952881
2016-04-22	3851.0	'2016-04-22T00:00:00.0000000'	2222.170898
2016-04-23	3851.0	'2016-04-23T00:00:00.0000000'	2258.958008
...	...	...	...
2022-06-12	3851.0	'2022-06-12T00:00:00.0000000'	92039.40625
2022-06-13	3851.0	'2022-06-13T00:00:00.0000000'	92047.28125
2022-06-14	3851.0	'2022-06-14T00:00:00.0000000'	92055.8125
2022-06-15	3851.0	'2022-06-15T00:00:00.0000000'	92134.632812
2022-06-16	3851.0	'2022-06-16T00:00:00.0000000'	92156.929688

Σχήμα 33: Max Calculation of Active Energy of HVAC
(a)Graph (b) Table

9.1.7 Controller Schneider > Groundfloor > Humidity

Η χρονοσειρά απεικονίζει την μέτριση της υγρασίας στο ισόγειο που μετρήθηκε με *Controller Schneider*. Είναι πυκνό με έναν συνδιασμό εποχικοτήτων.

(a)



(b)

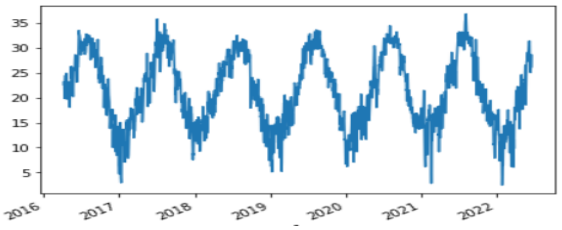
2016-04-07	3876.0	'2016-04-07T00:00:00.0000000'	67.677255
2016-04-08	3876.0	'2016-04-08T00:00:00.0000000'	63.282073
2016-04-09	3876.0	'2016-04-09T00:00:00.0000000'	64.559625
2016-04-10	3876.0	'2016-04-10T00:00:00.0000000'	71.198938
2016-04-11	3876.0	'2016-04-11T00:00:00.0000000'	67.067823
...	...	...	...
2022-06-12	3876.0	'2022-06-12T00:00:00.0000000'	77.313769
2022-06-13	3876.0	'2022-06-13T00:00:00.0000000'	70.700979
2022-06-14	3876.0	'2022-06-14T00:00:00.0000000'	69.746574
2022-06-15	3876.0	'2022-06-15T00:00:00.0000000'	60.276312
2022-06-16	3876.0	'2022-06-16T00:00:00.0000000'	64.322703

Σχήμα 34: Humidity of GroundFloor
(a)Graph (b) Table

9.1.8 Controller Schneider > Outdoor > Temperature

Το παρακάτω διάγραμμα παρουσιάζει την θερμοκρασία στον εξωτερικό χώρο του καταστήματος και υπολογίστηκε με *Controller Schneider*. Όπως και στα προηγούμενα διαγράμματα που αφορούσαν θερμοκρασίες, υπάρχει εποχικότητα, κυριώς ανά χρόνο, χωρίς κάποια συνολική τάση.

(a)



(b)

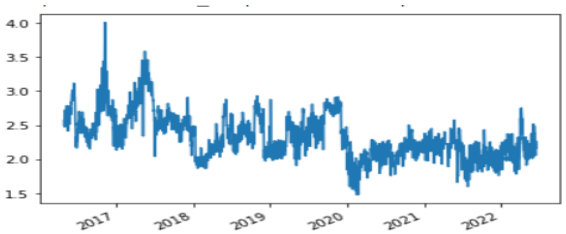
2016-04-08	2.0	'2016-04-08T00:00:00.0000000'	22.608437
2016-04-09	2.0	'2016-04-09T00:00:00.0000000'	23.285104
2016-04-10	2.0	'2016-04-10T00:00:00.0000000'	22.444375
2016-04-11	2.0	'2016-04-11T00:00:00.0000000'	20.529479
2016-04-12	2.0	'2016-04-12T00:00:00.0000000'	19.844479
...	...	...	...
2022-06-12	2.0	'2022-06-12T00:00:00.0000000'	25.03979
2022-06-13	2.0	'2022-06-13T00:00:00.0000000'	27.048632
2022-06-14	2.0	'2022-06-14T00:00:00.0000000'	27.745319
2022-06-15	2.0	'2022-06-15T00:00:00.0000000'	28.660521
2022-06-16	2.0	'2022-06-16T00:00:00.0000000'	26.16125

Σχήμα 35: Outdoor Temperature
(a)Graph (b) Table

9.1.9 Main Electricity Panel > Rest + Lighting > THD Voltage LN Avg

Η μέτρηση αυτή περιγράφει την ολική αρμονική παραμόρφωση τάσης στο φωτισμό. Τα δεδομένα της χρονοσειράς δεν παρουσιάζουν κάποια συγκεκριμένη εποχικότητα.

(a)



(b)

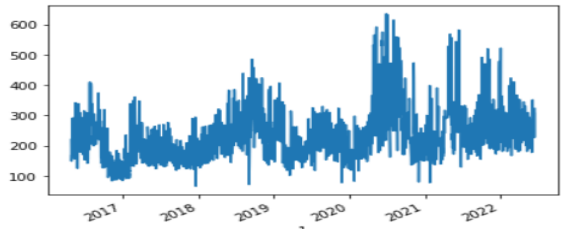
2016-04-26	3901.0	'2016-04-26T00:00:00.0000000'	2.569864
2016-04-27	3901.0	'2016-04-27T00:00:00.0000000'	2.493166
2016-04-28	3901.0	'2016-04-28T00:00:00.0000000'	2.511022
2016-04-29	3901.0	'2016-04-29T00:00:00.0000000'	2.472861
2016-04-30	3901.0	'2016-04-30T00:00:00.0000000'	2.527206
...	...	...	...
2022-06-12	3901.0	'2022-06-12T00:00:00.0000000'	2.203245
2022-06-13	3901.0	'2022-06-13T00:00:00.0000000'	2.110855
2022-06-14	3901.0	'2022-06-14T00:00:00.0000000'	2.056398
2022-06-15	3901.0	'2022-06-15T00:00:00.0000000'	2.26825
2022-06-16	3901.0	'2022-06-16T00:00:00.0000000'	2.161599

Σχήμα 36: Rest + Lighting
(a)Graph (b) Table

9.1.10 Main Electricity Panel > Rest + Lighting > THD Current Neutral

Οι παρακάτω τιμές αφορούν την ολική αρμονική παραμόρφωση γειώσης στο φωτισμό. Όπως και στην προηγούμενη περίπτωση τα δεδομένα μεταβάλλονται χωρίς κάποια τάση ή φανερή εποχικότητα.

(a)



(b)

2016-04-26	3911.0	'2016-04-26T00:00:00.0000000'	220.670911
2016-04-27	3911.0	'2016-04-27T00:00:00.0000000'	148.94297
2016-04-28	3911.0	'2016-04-28T00:00:00.0000000'	168.695489
2016-04-29	3911.0	'2016-04-29T00:00:00.0000000'	167.850512
2016-04-30	3911.0	'2016-04-30T00:00:00.0000000'	244.557575
...	...	...	...
2022-06-12	3911.0	'2022-06-12T00:00:00.0000000'	286.277864
2022-06-13	3911.0	'2022-06-13T00:00:00.0000000'	313.269533
2022-06-14	3911.0	'2022-06-14T00:00:00.0000000'	306.402384
2022-06-15	3911.0	'2022-06-15T00:00:00.0000000'	226.937117
2022-06-16	3911.0	'2022-06-16T00:00:00.0000000'	322.985835

Σχήμα 37: Rest + Lighting
(a)Graph (b) Table

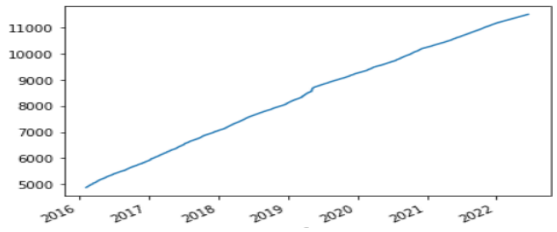
9.2 Κατάστημα 2ο

Το δεύτερο κατάστημα που μελετήθηκε είναι εκείνο με κωδικό 109.

9.2.1 Max calculation Lighting > Active Energy > Lighting Energy

Η μέτρηση αυτή υπολογίζει την μέγιστη φωτεινή ενέργεια. Η χρονοσειρά εμφανίζει αυξητική τάση χωρίς εποχικότητα.

(a)



(b)

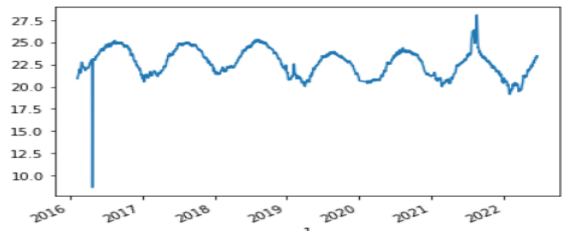
2016-02-10	14409.0	'2016-02-10T00:00:00.0000000'	4878.8
2016-02-11	14409.0	'2016-02-11T00:00:00.0000000'	4884.3
2016-02-12	14409.0	'2016-02-12T00:00:00.0000000'	4892.4
2016-02-13	14409.0	'2016-02-13T00:00:00.0000000'	4901.1
2016-02-14	14409.0	'2016-02-14T00:00:00.0000000'	4903.9
...	...	...	...
2022-06-12	14409.0	'2022-06-12T00:00:00.0000000'	11515.569336
2022-06-13	14409.0	'2022-06-13T00:00:00.0000000'	11517.009766
2022-06-14	14409.0	'2022-06-14T00:00:00.0000000'	11518.439453
2022-06-15	14409.0	'2022-06-15T00:00:00.0000000'	11520.389648
2022-06-16	14409.0	'2022-06-16T00:00:00.0000000'	11521.219727

Σχήμα 38: Max Calculation of Lighting
(a)Graph (b) Table

9.2.2 Controller Siemens > Basement > Temperature

Η μέτρηση αυτή περιγράφει την θερμοκρασία στο υπόγειο και μετρήθηκε με *Controller Siemens*. Η χρονοσειρά δεν παρουσιάζει τάση, όμως έχει εποχικότητα, με κάποιες φανερές διαφοροποιήσεις. Για παράδειγμα, το 2016 προέκυψε μία ξαφνική μείωση της θερμοκρασίας πλησιάζοντας το 0.

(a)



(b)

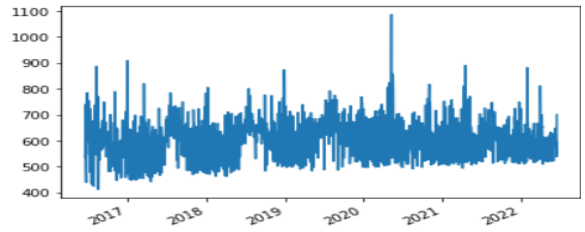
2016-02-10	14489.0	'2016-02-10T00:00:00.0000000'	20.9
2016-02-11	14489.0	'2016-02-11T00:00:00.0000000'	21.033333
2016-02-12	14489.0	'2016-02-12T00:00:00.0000000'	21.089583
2016-02-13	14489.0	'2016-02-13T00:00:00.0000000'	21.227083
2016-02-14	14489.0	'2016-02-14T00:00:00.0000000'	21.2625
...	...	...	...
2022-06-12	14489.0	'2022-06-12T00:00:00.0000000'	23.208229
2022-06-13	14489.0	'2022-06-13T00:00:00.0000000'	23.300104
2022-06-14	14489.0	'2022-06-14T00:00:00.0000000'	23.486771
2022-06-15	14489.0	'2022-06-15T00:00:00.0000000'	23.520416
2022-06-16	14489.0	'2022-06-16T00:00:00.0000000'	23.415946

Σχήμα 39: Temperature of Basement
(a)Graph (b) Table

9.2.3 Controller Siemens > Groundfloor > CO2

Η μέτρηση αυτή αφορά την ποσότητα διοξειδίου του άνθρακα στο υπόγειο και μετρήθηκε με *Controller Siemens*. Το διάγραμμα είναι ιδιαίτερα πυκνό χωρίς τάση και φανερή εποχικότητα.

(a)



(b)

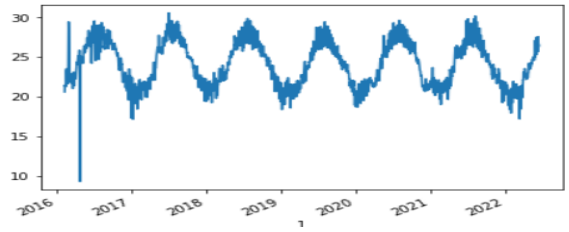
2016-06-15	14984.0	'2016-06-15T00:00:00.0000000'	534.386056
2016-06-16	14984.0	'2016-06-16T00:00:00.0000000'	585.843759
2016-06-17	14984.0	'2016-06-17T00:00:00.0000000'	740.112511
2016-06-18	14984.0	'2016-06-18T00:00:00.0000000'	692.276605
2016-06-19	14984.0	'2016-06-19T00:00:00.0000000'	500.733342
...	...	...	...
2022-06-12	14984.0	'2022-06-12T00:00:00.0000000'	585.725017
2022-06-13	14984.0	'2022-06-13T00:00:00.0000000'	595.904181
2022-06-14	14984.0	'2022-06-14T00:00:00.0000000'	538.333357
2022-06-15	14984.0	'2022-06-15T00:00:00.0000000'	614.893766
2022-06-16	14984.0	'2022-06-16T00:00:00.0000000'	701.140551

Σχήμα 40: CO2 in GroundFloor
(a)Graph (b) Table

9.2.4 Controller Siemens > Groundfloor > Temperature

Η μέτρηση αυτή υπολογίζει τη θερμοκρασία στο ισόγειο. Η μέτρηση έγινε και πάλι με το όργανο *Controller Schneider*. Η χρονοσειρά δεν εμφανίζει τάση αλλά εποχικότητα. Φανερή είναι η εποχικότητα ανά χρόνο. Επιπλέον εμφανίζεται μία απότομη πτώση το 2016.

(a)



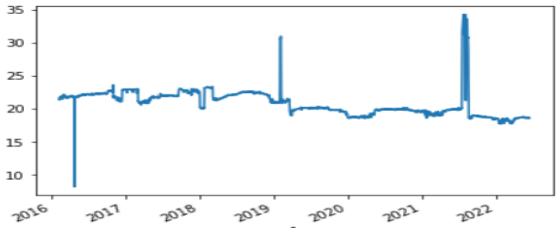
(b)

2016-02-10	15189.0	'2016-02-10T00:00:00.0000000'	20.571429
2016-02-11	15189.0	'2016-02-11T00:00:00.0000000'	21.425
2016-02-12	15189.0	'2016-02-12T00:00:00.0000000'	21.302083
2016-02-13	15189.0	'2016-02-13T00:00:00.0000000'	21.454167
2016-02-14	15189.0	'2016-02-14T00:00:00.0000000'	21.216667
...	...	...	...
2022-06-12	15189.0	'2022-06-12T00:00:00.0000000'	27.149583
2022-06-13	15189.0	'2022-06-13T00:00:00.0000000'	27.554583
2022-06-14	15189.0	'2022-06-14T00:00:00.0000000'	26.488333
2022-06-15	15189.0	'2022-06-15T00:00:00.0000000'	26.240104
2022-06-16	15189.0	'2022-06-16T00:00:00.0000000'	26.492222

Σχήμα 41: Temperature of GroundFloor
(a)Graph (b) Table

9.2.5 Controller Siemens > Data Room > Temperature

Η χρονοσειρά αυτή περιγράφει τη θερμοκρασία στο Data Room και πραγματοποιήθηκε με το όργανο *Controller Schneider*. Δεν παρουσιάζει τάση, αλλά ούτε κάποια φανερή σταθερή εποχικότητα.

(a) 

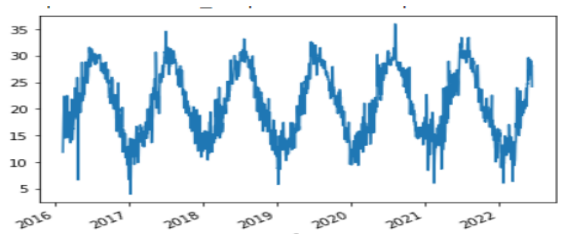
(b)

2016-02-10	15194.0	'2016-02-10T00:00:00.0000000'	21.378571
2016-02-11	15194.0	'2016-02-11T00:00:00.0000000'	21.29375
2016-02-12	15194.0	'2016-02-12T00:00:00.0000000'	21.75625
2016-02-13	15194.0	'2016-02-13T00:00:00.0000000'	21.66875
2016-02-14	15194.0	'2016-02-14T00:00:00.0000000'	21.539583
...	...	...	...
2022-06-12	15194.0	'2022-06-12T00:00:00.0000000'	18.558854
2022-06-13	15194.0	'2022-06-13T00:00:00.0000000'	18.528958
2022-06-14	15194.0	'2022-06-14T00:00:00.0000000'	18.736979
2022-06-15	15194.0	'2022-06-15T00:00:00.0000000'	18.700312
2022-06-16	15194.0	'2022-06-16T00:00:00.0000000'	18.576486

Σχήμα 42: Temperature of Data Room
(a)Graph (b) Table

9.2.6 Controller Siemens > Outdoor > Temperature

Η παρακάτω μέτρηση αφορά την θερμοκρασία εξωτερικά του κτηρίου και υπολογίστηκε και πάλι με *Controller Schneider*. Είναι φανερή η εποχικότητα ανά χρόνο.

(a) 

(b)

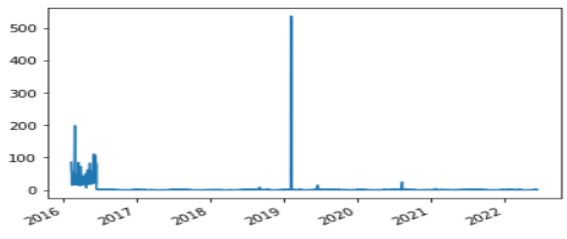
2016-02-10	16269.0	'2016-02-10T00:00:00.0000000'	11.835714
2016-02-11	16269.0	'2016-02-11T00:00:00.0000000'	14.060417
2016-02-12	16269.0	'2016-02-12T00:00:00.0000000'	16.889583
2016-02-13	16269.0	'2016-02-13T00:00:00.0000000'	17.564583
2016-02-14	16269.0	'2016-02-14T00:00:00.0000000'	18.710417
...	...	...	...
2022-06-12	16269.0	'2022-06-12T00:00:00.0000000'	26.424166
2022-06-13	16269.0	'2022-06-13T00:00:00.0000000'	27.372499
2022-06-14	16269.0	'2022-06-14T00:00:00.0000000'	27.908958
2022-06-15	16269.0	'2022-06-15T00:00:00.0000000'	28.251458
2022-06-16	16269.0	'2022-06-16T00:00:00.0000000'	24.181621

Σχήμα 43: Temperature Outdoor
(a)Graph (b) Table

9.2.7 Main Electricity Panel > HVAC > Active Energy

Το διάγραμμα παρουσιάζει την μέγιστη τιμή της *Active Energy* της θέρμανσης, του εξαερισμού και του κλιματισμού από το όργανο *MainElectricityPanel*. Δεν εμφανίζεται τάση ή σταθερή εποχικότητα.

(a)



(b)

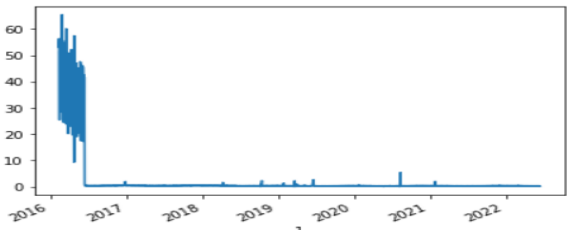
2016-02-11	17819.0	'2016-02-11T00:00:00.0000000'	85.6	NULL	1
2016-02-12	17819.0	'2016-02-12T00:00:00.0000000'	59.4	NULL	1
2016-02-13	17819.0	'2016-02-13T00:00:00.0000000'	59.8	NULL	1
2016-02-14	17819.0	'2016-02-14T00:00:00.0000000'	14.9	NULL	1
2016-02-15	17819.0	'2016-02-15T00:00:00.0000000'	15.0	NULL	1
...	...	...	...	...	...
2022-06-12	17819.0	'2022-06-12T00:00:00.0000000'	0.26709	NULL	96
2022-06-13	17819.0	'2022-06-13T00:00:00.0000000'	0.275391	NULL	96
2022-06-14	17819.0	'2022-06-14T00:00:00.0000000'	1.029134	NULL	96
2022-06-15	17819.0	'2022-06-15T00:00:00.0000000'	1.44043	NULL	96
2022-06-16	17819.0	'2022-06-16T00:00:00.0000000'	0.966283	NULL	38

Σχήμα 44: Active Energy of HVAC
(a)Graph (b) Table

9.2.8 Main Electricity Panel > Lighting Total > Active Energy

Η παρακάτω χρονοσειρά περιγράφει την κατανάλωση ενέργειας από το φωτισμό. Η μέτρηση έγινε με το *Main Electricity Panel*. Όπως και προηγουμένως δεν παρουσιάζεται κάποια τάση ή σταθερή εποχικότητα.

(a)



(b)

2016-02-11	18031.0	'2016-02-11T00:00:00.0000000'	53.1	NULL	2
2016-02-12	18031.0	'2016-02-12T00:00:00.0000000'	54.6	NULL	2
2016-02-13	18031.0	'2016-02-13T00:00:00.0000000'	56.4	NULL	2
2016-02-14	18031.0	'2016-02-14T00:00:00.0000000'	30.1	NULL	2
2016-02-15	18031.0	'2016-02-15T00:00:00.0000000'	25.4	NULL	2
...	...	...	...	...	...
2022-06-12	18031.0	'2022-06-12T00:00:00.0000000'	0.109436	NULL	192
2022-06-13	18031.0	'2022-06-13T00:00:00.0000000'	0.109243	NULL	192
2022-06-14	18031.0	'2022-06-14T00:00:00.0000000'	0.108805	NULL	192
2022-06-15	18031.0	'2022-06-15T00:00:00.0000000'	0.359996	NULL	192
2022-06-16	18031.0	'2022-06-16T00:00:00.0000000'	0.184791	NULL	62

Σχήμα 45: Total Energy of Lighting
(a)Graph (b) Table

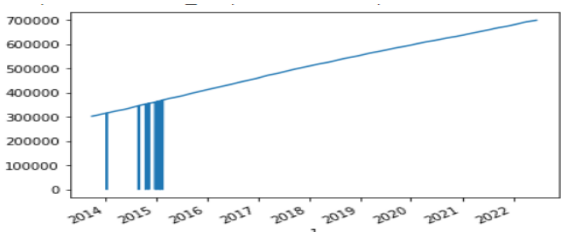
9.3 Κατάστημα 3ο

Το τρίτο κατάστημα που μελετήθηκε είναι εκείνο με κωδικό 62.

9.3.1 Max calculation Main > Active Energy > Main Energy

Η μέτρηση αυτή υπολογίζει την μέγιστη κύρια ενέργεια. Η γραφική παράσταση απεικονίζει αύξουσα τάση χωρίς εποχικότητα. Το 2015 εμφανίζονται ξαφνικές μειώσεις.

(a)



(b)

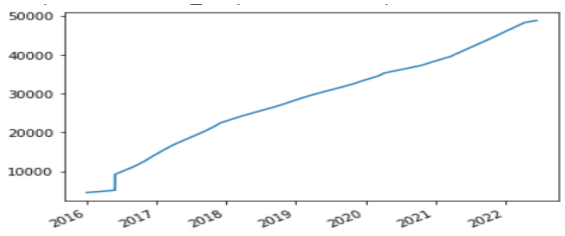
2013-10-03	4499.0	'2013-10-03T00:00:00.0000000'	302430.0
2013-10-04	4499.0	'2013-10-04T00:00:00.0000000'	302563.0
2013-10-05	4499.0	'2013-10-05T00:00:00.0000000'	302635.0
2013-10-08	4499.0	'2013-10-08T00:00:00.0000000'	302906.0
2013-10-10	4499.0	'2013-10-10T00:00:00.0000000'	303214.0
...	...	...	...
2022-06-12	4499.0	'2022-06-12T00:00:00.0000000'	698495.0
2022-06-13	4499.0	'2022-06-13T00:00:00.0000000'	698560.0
2022-06-14	4499.0	'2022-06-14T00:00:00.0000000'	698625.0
2022-06-15	4499.0	'2022-06-15T00:00:00.0000000'	698752.0
2022-06-16	4499.0	'2022-06-16T00:00:00.0000000'	698790.0

Σχήμα 46: Max Calculation of Main Energy
(a)Graph (b) Table

9.3.2 Max Calculation Lighting > Active Energy > Lighting Energy

Η μέτρηση αυτή αφορά τη μέγιστη φωτεινή ενέργεια. Η χρονοσειρά παρουσιάζει αυξητική πρόοδο χωρίς εποχικότητα.

(a)



(b)

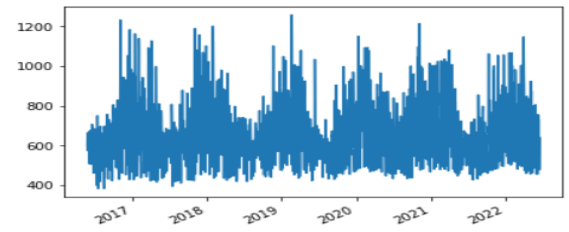
2015-12-29	4509.0	'2015-12-29T00:00:00.0000000'	4580.2
2015-12-30	4509.0	'2015-12-30T00:00:00.0000000'	4585.2
2015-12-31	4509.0	'2015-12-31T00:00:00.0000000'	4589.4
2016-01-01	4509.0	'2016-01-01T00:00:00.0000000'	4593.4
2016-01-02	4509.0	'2016-01-02T00:00:00.0000000'	4596.1
...	...	...	...
2022-06-12	4509.0	'2022-06-12T00:00:00.0000000'	48776.539062
2022-06-13	4509.0	'2022-06-13T00:00:00.0000000'	48779.027344
2022-06-14	4509.0	'2022-06-14T00:00:00.0000000'	48781.539062
2022-06-15	4509.0	'2022-06-15T00:00:00.0000000'	48793.628906
2022-06-16	4509.0	'2022-06-16T00:00:00.0000000'	48797.0

Σχήμα 47: Max Calculation of Lighting Energy
(a)Graph (b) Table

9.3.3 Controller Siemens > Groundfloor > CO2

Η παρακάτω χρονοσειρά απεικονίζει την ποσότητα διοξειδίου του άνθρακα στο υπόγειο και μετρήθηκε με *Controller Siemens*. Στο διάγραμμα δεν γίνεται φανερή κάποια τάση. Όμως, υπάρχει εποχικότητα, με πιο φανερή αυτή ανά χρόνο.

(a)



(b)

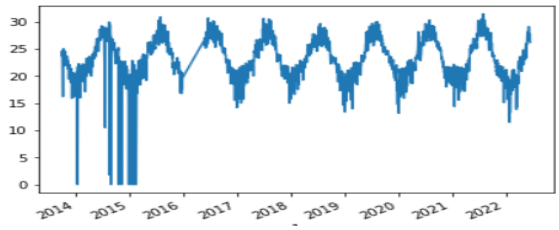
2016-05-27	4654.0	'2016-05-27T00:00:00.0000000'	598.331435
2016-05-28	4654.0	'2016-05-28T00:00:00.0000000'	662.620848
2016-05-29	4654.0	'2016-05-29T00:00:00.0000000'	596.527847
2016-05-30	4654.0	'2016-05-30T00:00:00.0000000'	573.181257
2016-05-31	4654.0	'2016-05-31T00:00:00.0000000'	628.154181
...	...	...	...
2022-06-12	4654.0	'2022-06-12T00:00:00.0000000'	545.935431
2022-06-13	4654.0	'2022-06-13T00:00:00.0000000'	475.825008
2022-06-14	4654.0	'2022-06-14T00:00:00.0000000'	483.677094
2022-06-15	4654.0	'2022-06-15T00:00:00.0000000'	592.856262
2022-06-16	4654.0	'2022-06-16T00:00:00.0000000'	638.502714

Σχήμα 48: CO2 in GroundFloor
(a)Graph (b) Table

9.3.4 Controller Siemens > Groundfloor > Temperature

Η μέτρηση αυτή υπολογίζει τη θερμοκρασία στο ισόγειο. Η μέτρηση έγινε και πάλι με το όργανο *Controller Schneider*. Δεν εμφανίζεται τάση, παρά μόνο εποχικότητα. Επίσης, παρουσιάζονται ορισμένες ξαφνικές πτώσεις το 2015.

(a)



(b)

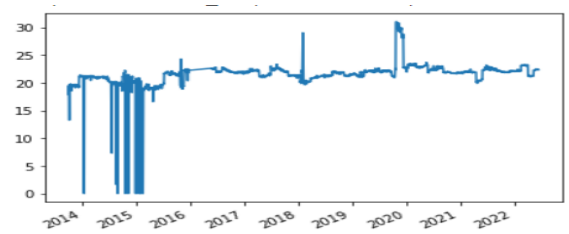
2013-10-03	4749.0	'2013-10-03T00:00:00.0000000'	23.7
2013-10-04	4749.0	'2013-10-04T00:00:00.0000000'	24.615789
2013-10-05	4749.0	'2013-10-05T00:00:00.0000000'	23.066667
2013-10-08	4749.0	'2013-10-08T00:00:00.0000000'	16.166667
2013-10-10	4749.0	'2013-10-10T00:00:00.0000000'	24.940909
...	...	...	...
2022-06-12	4749.0	'2022-06-12T00:00:00.0000000'	27.143229
2022-06-13	4749.0	'2022-06-13T00:00:00.0000000'	27.21177
2022-06-14	4749.0	'2022-06-14T00:00:00.0000000'	27.783541
2022-06-15	4749.0	'2022-06-15T00:00:00.0000000'	26.19802
2022-06-16	4749.0	'2022-06-16T00:00:00.0000000'	26.137567

Σχήμα 49: Temperature of GroundFloor
(a)Graph (b) Table

9.3.5 Controller Siemens > Data Room > Temperature

Η μέτρηση αυτή αφορά τη θερμοκρασία στο Data Room και πραγματοποιήθηκε με το όργανο *Controller Schneider*. Δεν παρατηρείται τάση, ούτε κάποια ορισμένη εποχικότητα.

(a)



(b)

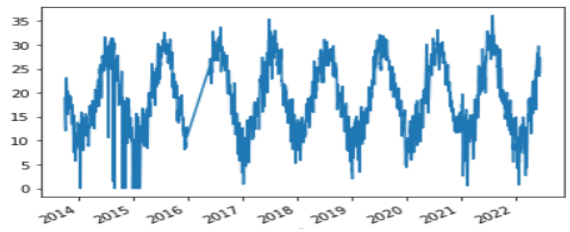
2013-10-03	4744.0	'2013-10-03T00:00:00.0000000'	17.9
2013-10-04	4744.0	'2013-10-04T00:00:00.0000000'	19.542105
2013-10-05	4744.0	'2013-10-05T00:00:00.0000000'	19.455556
2013-10-08	4744.0	'2013-10-08T00:00:00.0000000'	13.266667
2013-10-10	4744.0	'2013-10-10T00:00:00.0000000'	19.718182
...	...	...	...
2022-06-12	4744.0	'2022-06-12T00:00:00.0000000'	22.51552
2022-06-13	4744.0	'2022-06-13T00:00:00.0000000'	22.37302
2022-06-14	4744.0	'2022-06-14T00:00:00.0000000'	22.394166
2022-06-15	4744.0	'2022-06-15T00:00:00.0000000'	22.386041
2022-06-16	4744.0	'2022-06-16T00:00:00.0000000'	22.391621

Σχήμα 50: Temperature of Data Room
(a)Graph (b) Table

9.3.6 Controller Siemens > Outdoor > Temperature

Το διάγραμμα παρουσιάζει τη θερμοκρασία εξωτερικά του κτηρίου και υπολογίστηκε και πάλι με *Controller Schneider*. Η χρονοσειρά εμφανίζει εποχικότητα ανά χρόνο.

(a)



(b)

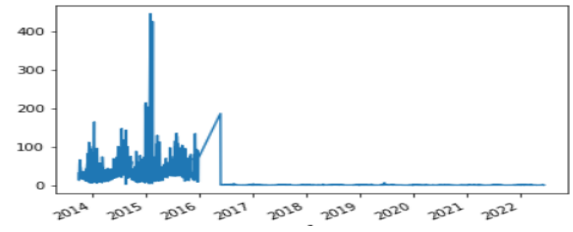
2013-10-03	5339.0	'2013-10-03T00:00:00.0000000'	18.925
2013-10-04	5339.0	'2013-10-04T00:00:00.0000000'	16.726316
2013-10-05	5339.0	'2013-10-05T00:00:00.0000000'	14.466667
2013-10-08	5339.0	'2013-10-08T00:00:00.0000000'	12.033333
2013-10-10	5339.0	'2013-10-10T00:00:00.0000000'	22.072727
...	...	...	...
2022-06-12	5339.0	'2022-06-12T00:00:00.0000000'	23.445364
2022-06-13	5339.0	'2022-06-13T00:00:00.0000000'	24.441145
2022-06-14	5339.0	'2022-06-14T00:00:00.0000000'	26.265937
2022-06-15	5339.0	'2022-06-15T00:00:00.0000000'	27.460416
2022-06-16	5339.0	'2022-06-16T00:00:00.0000000'	24.208513

Σχήμα 51: Temperature of Outdoor
(a)Graph (b) Table

9.3.7 Main Electricity Panel > HVAC > Active Energy

Η παρακάτω μέτρηση δηκλωνει την μέγιστη τιμή της *Active Energy* από το όργανο *HVAC*. Δεν υπάρχει τάση ή κάποια συγκεκριμένη εποχικότητα.

(a)



(b)

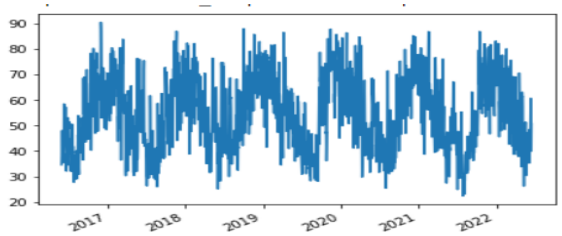
2013-10-04	6981.0	'2013-10-04T00:00:00.0000000'	31.6
2013-10-05	6981.0	'2013-10-05T00:00:00.0000000'	12.4
2013-10-08	6981.0	'2013-10-08T00:00:00.0000000'	55.5
2013-10-10	6981.0	'2013-10-10T00:00:00.0000000'	65.8
2013-10-11	6981.0	'2013-10-11T00:00:00.0000000'	22.6
...	...	...	...
2022-06-12	6981.0	'2022-06-12T00:00:00.0000000'	0.12321
2022-06-13	6981.0	'2022-06-13T00:00:00.0000000'	0.126302
2022-06-14	6981.0	'2022-06-14T00:00:00.0000000'	0.131999
2022-06-15	6981.0	'2022-06-15T00:00:00.0000000'	0.616211
2022-06-16	6981.0	'2022-06-16T00:00:00.0000000'	0.366554

Σχήμα 52: HVAC of Active Energy
(a)Graph (b) Table

9.3.8 Controller Siemens > Outdoor > Humidity

Η χρονοσειρά αυτή αφορά την υγρασία στον εξωτερικό χώρο του καταστήματος. Δεν παρατηρείται ορισμένη τάση, όμως είναι φανερή η εποχικότητα.

(a)



(b)

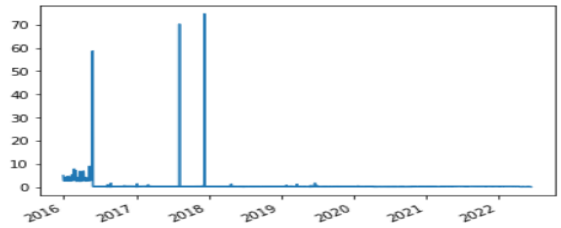
2016-05-27	140792.0	'2016-05-27T00:00:00.0000000'	34.543142
2016-05-28	140792.0	'2016-05-28T00:00:00.0000000'	37.954999
2016-05-29	140792.0	'2016-05-29T00:00:00.0000000'	47.815312
2016-05-30	140792.0	'2016-05-30T00:00:00.0000000'	48.019062
2016-05-31	140792.0	'2016-05-31T00:00:00.0000000'	47.939582
...	...	...	...
2022-06-12	140792.0	'2022-06-12T00:00:00.0000000'	60.648019
2022-06-13	140792.0	'2022-06-13T00:00:00.0000000'	52.758228
2022-06-14	140792.0	'2022-06-14T00:00:00.0000000'	45.972811
2022-06-15	140792.0	'2022-06-15T00:00:00.0000000'	39.900937
2022-06-16	140792.0	'2022-06-16T00:00:00.0000000'	50.679301

Σχήμα 53: Outdoor Temperature
(a)Graph (b) Table

9.3.9 Main Electricity Panel > Lighting Total > Active Energy

Η μέτρηση αυτή αφορά την ολικό φωτισμό. Η χρονοσειρά δεν παρουσιάζει ορισμένη τάση ή φανερή εποχικότητα. Επιπλέον, φαίνεται ορισμένες χρονικές στιγμές να παρουσιάζει ακραίες τιμές.

(a)



(b)

2015-12-30	7081.0	'2015-12-30T00:00:00.0000000'	5.0
2015-12-31	7081.0	'2015-12-31T00:00:00.0000000'	4.2
2016-01-01	7081.0	'2016-01-01T00:00:00.0000000'	4.0
2016-01-02	7081.0	'2016-01-02T00:00:00.0000000'	2.7
2016-01-03	7081.0	'2016-01-03T00:00:00.0000000'	2.7
...	...	...	...
2022-06-12	7081.0	'2022-06-12T00:00:00.0000000'	0.05366
2022-06-13	7081.0	'2022-06-13T00:00:00.0000000'	0.053212
2022-06-14	7081.0	'2022-06-14T00:00:00.0000000'	0.053457
2022-06-15	7081.0	'2022-06-15T00:00:00.0000000'	0.154785
2022-06-16	7081.0	'2022-06-16T00:00:00.0000000'	0.075762

Σχήμα 54: Rest + Lighting
(a)Graph (b) Table

10 Εφαρμογή των αλγορίθμων πρόβλεψης στα δεδομένα από καταστήματα τραπεζών

Στο μέρος αυτό, θα γίνει χρήση ορισμένων δεδομένων από καταστήματα τραπεζών, για να μελετηθεί η συμπεριφορά και η απόδοση των μοντέλων πρόβλεψης που αναλύθηκαν παραπάνω.

10.1 Κατάστημα 1ο

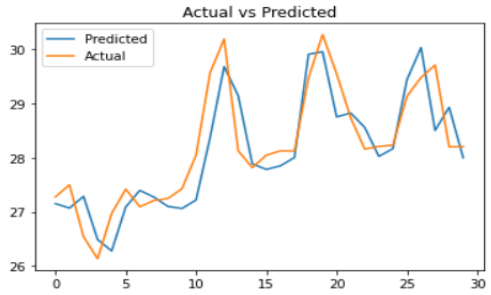
10.1.1 Controller Schneider > Groundfloor > Temperature2

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	4.4244	1.6960	2.1034	0.2789
SARIMA m = 7	0.6268	0.5484	0.7919	0.4332
ARIMA	2.8303	1.4050	1.6824	-1.5594
LSTM	0.3463	0.4559	0.5885	0.6771
CNN	0.3922	0.5121	0.6263	0.4722
LSTM-CNN	0.4534	0.5672	0.6734	0.5800
BILSTM	0.2962	0.4381	0.5442	0.7211
GRU	0.3428	0.4666	0.5855	0.6709
Prophet	1.089	0.7581	1.044	0.015
XGBOOST	0.6972	0.6630	0.8350	0.3696
Random Forest Regression	1.5813	0.9876	1.2575	-0.4298
΄ Random Forest Regression	0.30167	0.4104	0.5492	0.7272

ΣΥΜΠΕΡΑΣΜΑΤΑ

Την καλύτερη πρόβλεψη παρέχουν τα μοντέλα νευρωνικών δικτύων και συγκεκριμένα το $BILSTM$. Αντίθετα την λιγότερο ακριβή πρόβλεψη κάνει το μοντέλο $ARIMA$ με μεγάλη διαφορά από τα υπόλοιπα.



Σχήμα 55: BiLSTM Prediction

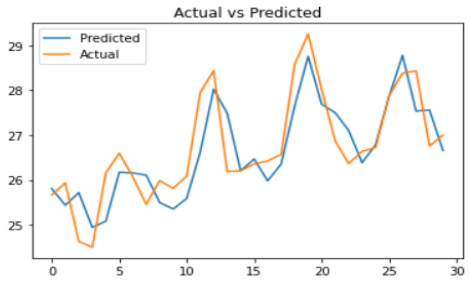
10.1.2 Controller Schneider > Groundfloor > Temperature1

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	2.4599	1.2898	1.5681	-0.8870
SARIMA m = 7	0.6046	0.6064	0.7776	0.5361
ARIMA	2.4588	1.2898	1.5680	-0.8868
LSTM	0.3967	0.5181	0.6299	0.6248
CNN	0.4654	0.5477	0.6822	0.5684
LSTM-CNN	0.5729	0.6483	0.7569	0.3447
BILSTM	0.4323	0.5011	0.6575	0.6223
GRU	0.4011	0.5419	0.6333	0.6418
Prophet	0.6849	0.6583	0.8276	0.4743
XGBOOST	0.6939	0.6765	0.8331	0.4675
Random Forest Regression	0.8847	0.7241	0.9406	0.3211
Παραλλαγή Random Forest Regression	0.5273	0.5594	0.7262	0.5954

ΣΥΜΠΕΡΑΣΜΑΤΑ

Την καλύτερη πρόβλεψη παρέχει το μοντέλο *LSTM*. Αντίθετα την λιγότερο ακριβές είναι το *Triple Exponential Smoothin*



Σχήμα 56: LSTM Prediction

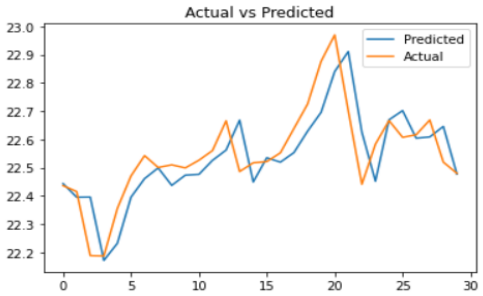
10.1.3 Controller Schneider > Data Room > DataRoom Temp

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	0.0147	0.0901	0.1213	0.4208
SARIMA m = 7	0.0306	0.1268	0.1751	-0.2046
ARIMA	0.0933	0.2667	0.3055	-2.6681
LSTM	0.0107	0.0811	0.1034	0.5404
CNN	0.0157	0.1008	0.1254	0.1590
LSTM-CNN	0.0296	0.1398	0.1721	-0.8002
BILSTM	0.0167	0.0938	0.1294	0.4114
GRU	0.0114	0.0872	0.1068	0.5217
Prophet	0.0229	0.1187	0.1516	0.0969
XGBOOST	0.0213	0.1206	0.1459	0.1633
Random Forest Regression	0.0782	0.2362	0.2797	-2.0743
Παραλλαγή Random Forest Regression	0.0119	0.0845	0.1090	0.5333

ΣΥΜΠΕΡΑΣΜΑΤΑ

Την καλύτερη πρόβλεψη παρέχει το $LSTM$. Αντίθετα την λιγότερο ακριβή πρόβλεψη κάνει το μοντέλο $ARIMA$.



Σχήμα 57: LSTM Prediction

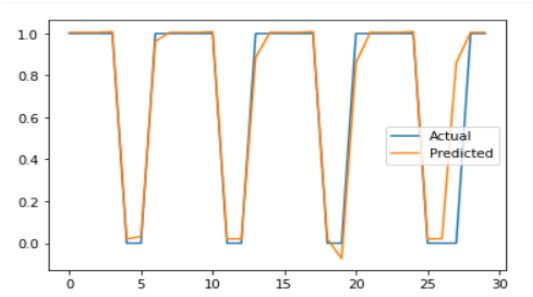
10.1.4 Controller Schneider > Feedback > AC Feedback

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	0.2101	0.4311	0.4583	-0.0003
SARIMA m = 1	0.0582	0.1922	0.2413	0.7223
ARIMA	0.0576	0.1906	0.2401	0.7255
LSTM	0.0290	0.0702	0.1704	0.8326
CNN	0.0286	0.0958	0.1691	0.8177
LSTM-CNN	0.0276	0.0796	0.2660	0.8496
BILSTM	0.0294	0.0801	0.1714	0.8206
GRU	0.0361	0.0769	0.1899	0.7955
Prophet	0.0271	0.0654	0.1647	0.8707
XGBOOST	0.0261	0.0498	0.1616	0.8756
Random Forest Regression	0.0283	0.03800	0.1684	0.8648
Παραλλαγή Random Forest Regression	0.0318	0.0678	0.1784	0.8485

ΣΥΜΠΕΡΑΣΜΑΤΑ

Πιο αποτελεσματικό είναι το μοντέλο *XGBoost*. Αντίθετα την λιγότερο ακριβή πρόβλεψη κάνει το μοντέλο *Triple Exponential Smoothing*.



Σχήμα 58: XGBoost Prediction

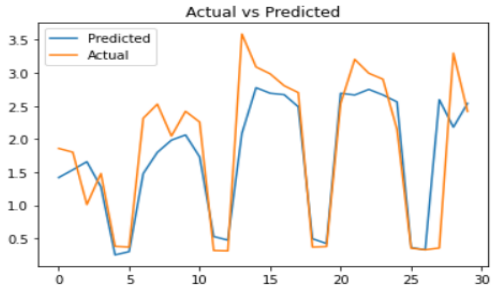
10.1.5 Main Electricity Panel > HVAC > Active Power

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	1.1932	0.9635	1.092	0.0329
SARIMA m = 7	1.1700	0.8850	1.0817	0.0518
ARIMA	1.4253	1.0005	1.1939	-0.1555
LSTM	0.4809	0.4718	0.6935	0.5197
CNN	0.4368	0.4353	0.6609	0.5066
LSTM-CNN	0.4965	0.4389	0.7046	0.4451
BILSTM	0.3767	0.4024	0.6138	0.5398
GRU	0.4651	0.4863	0.6820	0.5362
Prophet	0.4396	0.4737	0.6630	0.6437
XGBOOST	0.3953	0.4141	0.6287	0.6796
Random Forest Regression	0.4043	0.4282	0.6358	0.6723
Παραλλαγή Random Forest Regression	0.4130	0.4096	0.6426	0.6653

ΣΥΜΠΕΡΑΣΜΑΤΑ

Την καλύτερη πρόβλεψη παρέχει το *BILSTM* ενώ τη χειρότερη, το μοντέλο *Triple Exponential Smoothing*.



Σχήμα 59: BiLSTM Prediction

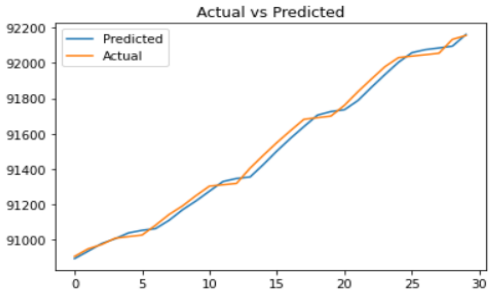
10.1.6 max calculation HVAC > Active Energy > Active Energy

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	63870.4096	199.0674	252.7260	0.6117
SARIMA m = 4	16200.14035	335.1528	405.3486	0.0106
ARIMA	598304.4718	68.6578	773.5014	-2.6377
LSTM	3081.6698	46.1094	55.5128	0.9805
CNN	33121.0244	177.9167	181.9918	0.8059
LSTM-CNN	0.4965	0.4389	0.7046	0.4451
BILSTM	6515.7020	68.6294	80.7199	0.9529
GRU	1002.6683	28.4393	31.6649	0.9939
Prophet	14242.4629	90.7129	119.3418	0.9134
XGBOOST	70952.0271	209.1398	266.3682	0.5686
Random Forest Regression	63209.0454	682.5073	792.8571	-2.8317
Παραλλαγή Random Forest Regression	5625.1231	68.0402	75.0008	0.9657

ΣΥΜΠΕΡΑΣΜΑΤΑ

Την καλύτερη πρόβλεψη παρέχουν τα μοντέλα νευρωνικών δικτύων και συγκεκριμένα το GRU. Το μοντέλο $ARIMA$ δεν παρουσιάζει καλά αποτελέσματα.



Σχήμα 60: GRU Prediction

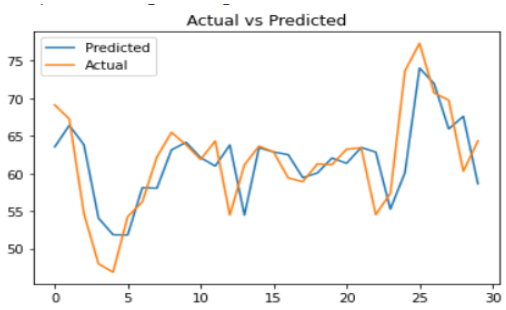
10.1.7 Controller Schneider > Groundfloor > Humidity

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	40.0772	5.1065	6.3307	0.1044
SARIMA m = 1	53.1666	5.8633	7.2915	-0.1881
ARIMA	44.1817	5.0525	6.6469	0.0127
LSTM	24.5818	3.6848	4.9580	0.0223
CNN	33.9976	4.9846	6.3244	-0.0861
LSTM-CNN	39.0759	4.6702	6.2511	-1.669
BILSTM	28.7878	4.041	5.3654	0.1322
GRU	25.1815	3.7937	5.0181	0.0570
Prophet	40.1886	4.9399	6.3394	0.1019
XGBOOST	39.6129	5.1321	6.2939	0.1148
Random Forest Regression	31.6216	4.4694	5.6233	0.2933
Παραλλαγή Random Forest Regression	30.4664	3.8287	5.5196	0.3192

ΣΥΜΠΕΡΑΣΜΑΤΑ

Το *LSTM* φαίνεται να είναι το πιο ακριβές μοντέλο.



Σχήμα 61: LSTM Prediction

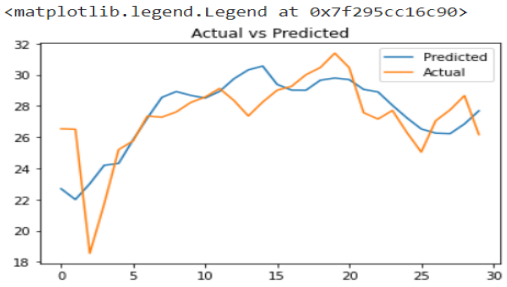
10.1.8 Controller Schneider > Outdoor > Temperature

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	4.4244	1.6596	2.1034	0.2789
SARIMA m = 7	6.6552	2.0860	2.5798	-0.085
ARIMA	6.7884	2.0768	2.6055	-0.1064
LSTM	3.5725	1.3628	1.8901	0.3019
CNN LATHOS	0.0030	0.9846	6.3244	-0.0861
LSTM-CNN	3.4478	1.4237	1.8568	0.3503
BILSTM	3.6155	1.3595	1.9014	0.3018
GRU	3.7572	1.3824	1.9384	0.3291
Prophet	4.9051	1.7909	2.2347	0.2006
XGBOOST	5.9785	1.7982	2.4451	0.0256
Random Forest Regression	13.5058	2.9277	3.6750	-1.2012
Παραλλαγή Random Forest Regression	3.9742	1.2516	1.9935	0.3523

ΣΥΜΠΕΡΑΣΜΑΤΑ

Την καλύτερη πρόβλεψη παρέχει το $LSTM - CNN$, ενώ και πάλι την λιγότερο ακριβή πρόβλεψη κάνει το μοντέλο $ARIMA$.



Σχήμα 62: LSTM-CNN Prediction

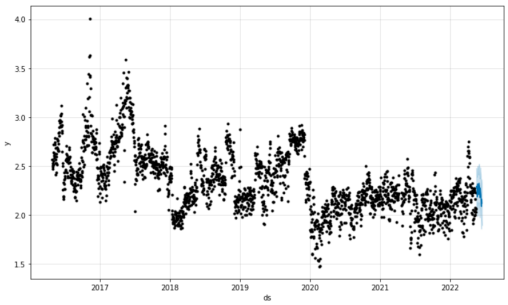
10.1.9 Main Electricity Panel > Rest + Lighting > THD Voltage LN Avg

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	0.0235	0.1335	0.1534	-0.1966
SARIMA m = 7	0.0307	0.1431	0.1751	-0.5594
ARIMA	0.0466	0.1809	0.2159	-1.3721
LSTM	0.0146	0.0998	0.1208	-0.2424
CNN	0.0192	0.1057	0.1386	-1.2776
LSTM-CNN	0.0159	0.0964	0.1259	-1.6642
BILSTM	0.0165	0.1064	0.1283	-0.0264
GRU	0.0163	0.1028	0.1277	-0.7239
Prophet	0.0144	0.0982	0.1201	0.2667
XGBOOST	0.0195	0.1141	0.1398	0.0061
Random Forest Regression	0.0439	0.1678	0.2006	-1.2342
Παραλλαγή Random Forest Regression	0.0184	0.1135	0.1355	0.0653

ΣΥΜΠΕΡΑΣΜΑΤΑ

Την καλύτερη πρόβλεψη προσφέρει *Prophet*.



Σχήμα 63: Prophet Prediction

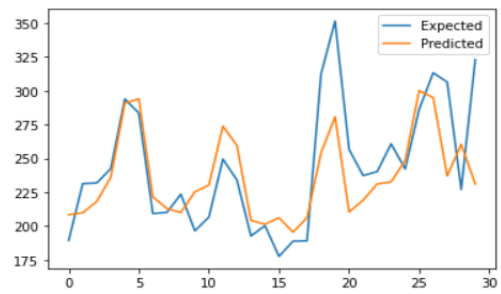
10.1.10 Main Electricity Panel > Rest + Lighting > THD Current Neutral

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	1824.3944	34.6816	42.7124	0.1269
SARIMA m = 1	1498.6334	29.7178	38.7178	0.2828
ARIMA	1498.6334	29.7178	38.7122	0.2828
LSTM	1059.6797	25.4150	35.5527	0.3555
CNN	1118.5280	23.7123	33.4444	-0.1762
LSTM-CNN	1734.7902	31.3287	41.6508	-0.9969
BILSTM	1229.2836	25.6233	35.0611	0.2155
GRU	1076.1362	24.5155	32.8045	0.3640
Prophet	1536.3018	31.1375	39.1957	0.2648
XGBOOST	1636.1383	30.9943	40.4492	0.2170
Random Forest Regression	3441.6987	45.4066	58.6660	-0.6470
Παραλλαγή Random Forest Regression	1071.8109	24.4778	32.7385	0.4871

ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην περίπτωση αυτή, την καλύτερη πρόβλεψη παρέχει το *Random Forest Regression*. Αντίθετα, το μοντέλο *Triple Exponential Smoothing* παρέχει τα χαμηλότερα αποτελέσματα.



Σχήμα 64: Random Forest Prediction

10.2 Κατάστημα 2ο

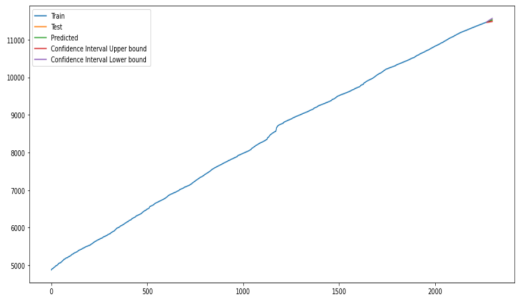
10.2.1 max calculation Lighting > Active Energy > Lighting Energy

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	0.5724	0.6503	0.7566	0.9981
SARIMA m = 12	1.2650	0.9022	1.1247	0.9959
ARIMA	0.5373	0.5918	0.7330	0.9982
LSTM	214.1730	14.5280	14.6346	0.1543
CNN	18.9169	4.2660	4.3494	0.9343
LSTM-CNN	786.7832	28.0091	28.0497	-2.0176
BILSTM	19.0581	4.2601	4.3656	0.9329
GRU	7.0849	2.5440	2.6617	0.9766
Prophet	230.8921	14.6971	15.1951	0.2462
XGBOOST	1334.6084	32.0254	36.5323	-3.3570
Random Forest Regression	1433.6795	33.5924	37.8690	-3.6805
Παραλλαγή Random Forest Regression	10.2871	3.1457	3.2073	0.9664

ΣΥΜΠΕΡΑΣΜΑΤΑ

Το *ARIMA* είναι το πιο ακριβές μοντέλο, σε αντίθεση με το Random Forest που έχει τα υψηλότερα σφάλματα.



Σχήμα 65: ARIMA Prediction

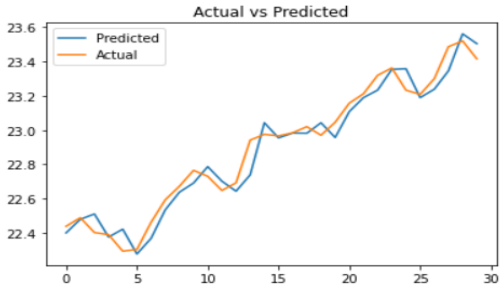
10.2.2 Controller Siemens > Basement > Temperature

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	0.0085	0.0773	0.0919	0.9378
SARIMA m = 7	0.0155	0.1087	0.1246	0.8858
ARIMA	0.6153	0.6915	0.7843	-3.6253
LSTM	0.0082	0.7485	0.0904	0.9388
CNN	0.0096	0.0854	0.0981	0.9190
LSTM-CNN	0.0245	0.1360	0.1567	0.7872
BILSTM	0.0059	0.0624	0.0771	0.9560
GRU	0.0076	0.0725	0.0870	0.9470
Prophet	0.02088	0.1236	0.1445	0.8464
XGBOOST	0.0416	0.1697	0.2040	0.6936
Random Forest Regression	0.0754	0.2322	0.2746	0.4454
Παραλλαγή Random Forest Regression	0.0071	0.0669	0.0842	0.9478

ΣΥΜΠΕΡΑΣΜΑΤΑ

Την καλύτερη πρόβλεψη παρέχουν τα μοντέλα νευρωνικών δικτύων και συγκεκριμένα το $BILSTM$. Αντίθετα την λιγότερο ακριβή πρόβλεψη κάνει το μοντέλο $ARIMA$ με μεγάλη διαφορά από τα υπόλοιπα.



Σχήμα 66: BiLSTM Prediction

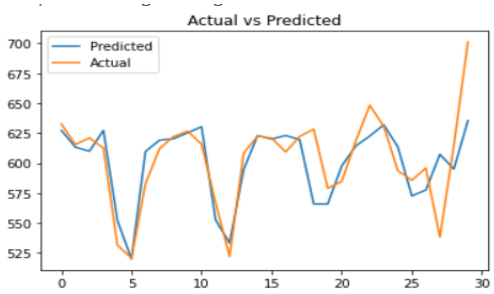
10.2.3 Controller Siemens > Groundfloor > CO2

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	1398.7426	30.6496	37.3998	0.0227
SARIMA m = 7	731.1378	18.5609	27.0396	0.4892
ARIMA	1395.9891	29.9843	37.3621	0.0247
LSTM	592.9086	16.3024	24.3497	0.3773
CNN	926.8541	22.2851	30.4443	-0.1965
LSTM-CNN	773.4193	21.5031	27.8104	0.1397
BILSTM	676.6707	17.5995	26.0129	0.0547
GRU	698.8758	18.8349	26.4363	0.2063
Prophet	861.3719	19.6952	29.3491	0.3981
XGBOOST	767.6205	20.1131	27.7060	0.4637
Random Forest Regression	996.5553	22.5569	31.5682	0.3037
Παραλλαγή Random Forest Regression	737.9471	17.8789	27.1652	0.4844

ΣΥΜΠΕΡΑΣΜΑΤΑ

Στην παραπάνω πρόβλεψη το πιο ακριβές μοντέλο είναι το *LSTM* και το λιγότερο το μοντέλο *Triple Exponential Smoothin*



Σχήμα 67: LSTM Prediction

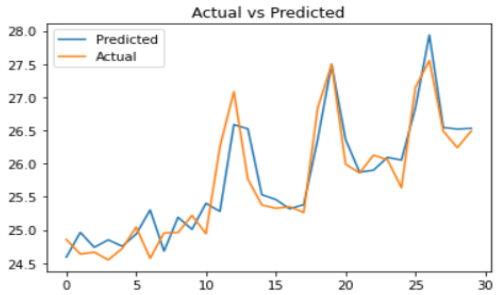
10.2.4 Controller Siemens > Groundfloor > Temperature

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	0.3420	0.4466	0.5848	0.5654
SARIMA m = 7	0.4926	0.4553	0.7018	0.3741
ARIMA	2.6256	1.3463	1.6204	-2.3360
LSTM	0.1958	0.3453	0.4425	0.7340
CNN	0.2644	0.3763	0.5142	0.5704
LSTM-CNN	0.1630	0.2877	0.4038	0.7725
BILSTM	0.1320	0.2768	0.3633	0.8167
GRU	0.2019	0.3379	0.4493	0.7120
Prophet	0.3993	0.5068	0.6319	0.4926
XGBOOST	0.3720	0.5284	0.6099	0.5273
Random Forest Regression	0.4022	0.5417	0.6342	0.4890
Παραλλαγή Random Forest Regression	0.2820	0.3849	0.5310	0.6417

ΣΥΜΠΕΡΑΣΜΑΤΑ

Την καλύτερη πρόβλεψη παρέχουν τα μοντέλα νευρωνικών δικτύων και συγκεκριμένα το $BiLSTM$. Από την άλλη, τα αποτελέσματα της πρόβλεψης του $ARIMA$ αποκλίνουν περισσότερο από τις πραγματικές τιμές.



Σχήμα 68: BiLSTM Prediction

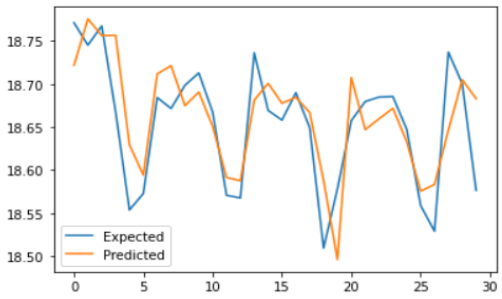
10.2.5 Controller Siemens > Data Room > Temperature

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	0.0044	0.0573	0.0664	0.1195
SARIMA m = 1	0.0159	0.1145	0.1263	-2.1824
ARIMA	0.0066	0.0603	0.0810	-0.3099
LSTM	0.0063	0.0594	0.0796	-0.3682
CNN	0.0057	0.0676	0.0753	-30.1657
LSTM-CNN	773.4193	21.5031	27.8104	0.1397
BILSTM	0.0080	0.0667	0.0893	-0.2628
GRU	0.0066	0.0625	0.0810	-0.2189
Prophet	0.02088	0.1236	0.1445	0.8464
XGBOOST	0.0025	0.0387	0.0502	0.4976
Random Forest Regression	0.0078	0.0772	0.0887	-0.5709
Παραλλαγή Random Forest Regression	0.0023	0.0385	0.0476	0.5484

ΣΥΜΠΕΡΑΣΜΑΤΑ

Το *Random Forest Regression* κάνει την πιο ακριβή πρόβλεψη.



Σχήμα 69: Random Forest Prediction

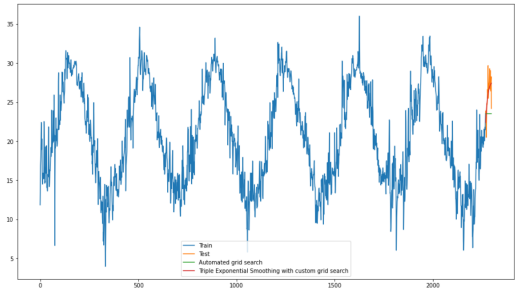
10.2.6 Controller Siemens > Outdoor > Temperature

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	2.0056	1.0507	1.4162	0.5878
SARIMA m = 7	7.1041	2.3376	2.6654	-0.4601
ARIMA	23.1392	4.3070	4.8103	-3.7559
LSTM	2.5182	1.2401	1.5869	0.4566
CNN	3.0342	1.2798	1.7419	0.2115
LSTM-CNN	3.1259	1.4313	1.7680	0.5967
BILSTM	2.4436	1.1393	1.5632	0.4570
GRU	2.5281	1.2372	1.5900	0.4642
Prophet	2.2656	1.1105	1.5052	0.5343
XGBOOST	5.5743	1.8919	2.3609	-0.1457
Random Forest Regression	6.1366	1.9454	2.4772	-0.2613
Παραλλαγή Random Forest Regression	2.3509	1.1547	1.5333	0.5168

ΣΥΜΠΕΡΑΣΜΑΤΑ

Την καλύτερη πρόβλεψη παρέχει το *Triple Exponential Smoothing*. Αντίθετα την λιγότερο ακριβή πρόβλεψη κάνει και πάλι το μοντέλο *ARIMA* με μεγάλη διαφορά από τα υπόλοιπα.



Σχήμα 70: Triple Exponential Smoothing Prediction

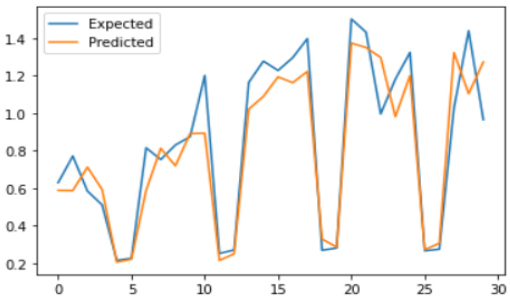
10.2.7 Main Electricity Panel > HVAC > Active Energy

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	0.1621	0.3306	0.4026	0.1359
SARIMA m = 7	0.1886	0.3575	0.4343	-0.0056
ARIMA	0.3518	0.5013	0.5931	-0.8758
LSTM	0.1902	0.3467	0.4361	-21.7659
CNN	0.8975	0.8426	0.9474	0
LSTM-CNN	0.8976	0.8426	0.9474	0
BILSTM	0.1813	0.3719	0.4257	-64.2381
GRU	0.1741	0.3474	0.4172	-13.6069
Prophet	0.2376	0.4059	0.4875	-0.2670
XGBOOST	0.0587	0.1847	0.2422	0.6872
Παραλλαγή Random Forest Regression	0.0619	0.1751	0.2487	0.6701
Random Forest Regression	0.0265	0.1259	0.1629	0.8585

ΣΥΜΠΕΡΑΣΜΑΤΑ

Το πιο ακριβές μοντέλο ήταν η παραλλαγή του *Random Forest Regression*.



Σχήμα 71: Random Forest Prediction

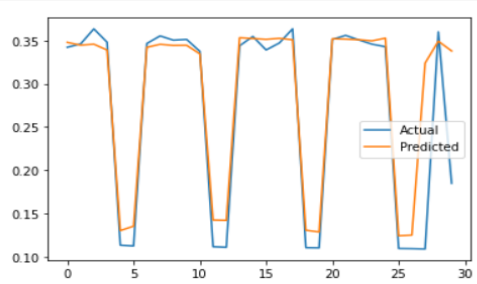
10.2.8 Main Electricity Panel > Lighting Total > Active Energy

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	0.0119	0.1044	0.1094	0.0135
SARIMA m = 1	0.0049	0.0548	0.0704	0.5918
ARIMA	1395.9891	29.9843	37.3621	0.0247
LSTM	0.0163	0.0959	0.1275	-6.1098
CNN	0.0058	0.0554	0.0765	0.0320
LSTM-CNN	0.0140	0.0888	0.1185	-652.7646
BILSTM	0.0126	0.0749	0.1125	-0.4590
GRU	0.0116	0.0920	0.1078	-4.3658
Prophet	0.0043	0.0441	0.0658	0.6425
XGBOOST	0.0025	0.0225	0.0499	0.7941
Random Forest Regression	0.0025	0.0201	0.0497	0.7963
Παραλλαγή Random Forest Regression	0.0028	0.0198	0.0527	0.7709

ΣΥΜΠΕΡΑΣΜΑΤΑ

Το μοντέλο *XGBOOST* είχε τα χαμηλότερα σφάλματα μεταξύ των πραγματικών τιμών της χρονοσειράς και αυτών που προβλέφθηκαν.



Σχήμα 72: ARIMA Prediction

10.3 Κατάστημα 3ο

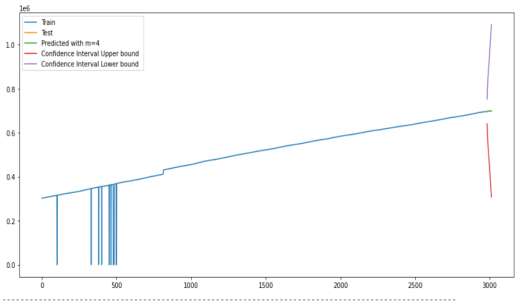
10.3.1 max calculation Main > Active Energy > Main Energy

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	16007.5083	108.5529	126.5208	0.9795
SARIMA m = 4	5901.6417	67.0845	76.8221	0.9924
ARIMA	4602007.9187	2101.5615	2145.2291	-4.8982
LSTM	20660.7435	110.6208	143.7385	0.9671
CNN	148720.5689	382.3563	385.6431	0.8207
LSTM-CNN	105444.4980	320.6188	324.7222	0.8566
BILSTM	6117.9471	54.375	78.2173	0.9931
GRU	38537.6967	193.9229	196.3102	0.9495
Prophet	13337.7901	92.2523	115.4894	0.9829
XGBOOST	2703955.95	1388.7917	1644.9709	-2.4656
Random Forest Regression	3188078.005	1552.4418	1785.5190	-3.0859
Παραλλαγή Random Forest Regression	24278.3731	152.5725	155.8152	0.9689

ΣΥΜΠΕΡΑΣΜΑΤΑ

Την καλύτερη πρόβλεψη παρέχοει το $SARIMA$. Αντίθετα την λιγότερο ακρβή πρόβλεψη κάνει το μοντέλο $XGBOOST$.



Σχήμα 73: SARIMA Prediction

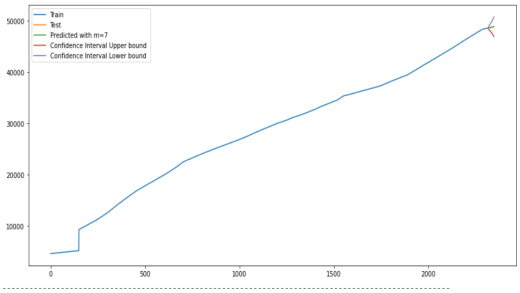
10.3.2 Max Calculation Lighting > Active Energy > Lighting Energy

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	25.1938	4.1111	5.0193	0.9952
SARIMA m = 7	18.0330	3.6121	4.2465	0.9966
ARIMA	35010.4937	162.8087	187.1191	-5.6917
LSTM	37.5111	5.0534	6.1246	0.9928
CNN	2407.0678	48.8083	49.0619	0.5335
LSTM-CNN	77.6245	7.5102	8.8105	0.8730
BILSTM	36.2991	4.9535	6.0249	0.9927
GRU	1297.6150	35.6607	36.0224	0.7462
Prophet	6254.2915	66.8300	79.0841	-0.1954
XGBOOST	21601.6202	127.9441	146.9749	-3.1289
Random Forest Regression	23786.3886	136.2420	154.2283	-3.5464
Παραλλαγή Random Forest Regression	184.0764	12.7823	13.5675	0.9648

ΣΥΜΠΕΡΑΣΜΑΤΑ

Και σε αυτήν την περίπτωση, το καλύτερο μοντέλο ήταν το $SARIMA$.



Σχήμα 74: SARIMA Prediction

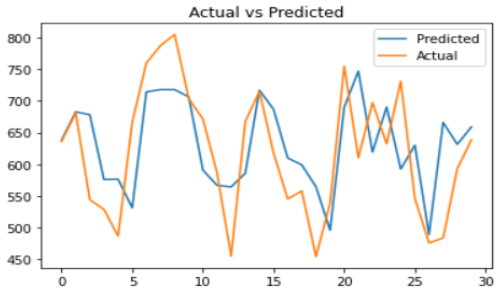
10.3.3 Controller Siemens > Groundfloor > CO2

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	9511.7333	81.9936	97.5281	0.0331
SARIMA m = 7	8132.6992	77.7927	90.1815	0.1733
ARIMA	10954.3629	87.3555	104.6631	-0.1136
LSTM	7085.9387	65.0946	84.1780	-0.2893
CNN	8095.6964	78.4425	89.9761	-1.2908
LSTM-CNN	8143.6337	74.3615	90.2421	-1.6921
BILSTM	6961.7559	72.8437	83.4371	-0.2593
GRU	6835.5251	68.3468	82.6772	-0.4681
Prophet	7335.6628	72.4134	85.6485	0.2543
XGBOOST	7573.8222	74.2318	87.0277	0.2301
Random Forest Regression	7985.1753	72.0629	89.3598	0.1883
Παραλλαγή Random Forest Regression	7672.5172	75.4249	87.5929	0.2201

ΣΥΜΠΕΡΑΣΜΑΤΑ

Την καλύτερη πρόβλεψη παρέχουν τα μοντέλα νευρωνικών δικτύων και συγκεκριμένα το *GRU*. Αντίθετα την λιγότερο ακριβή πρόβλεψη κάνει το μοντέλο *ARIMA*.



Σχήμα 75: GRU Prediction

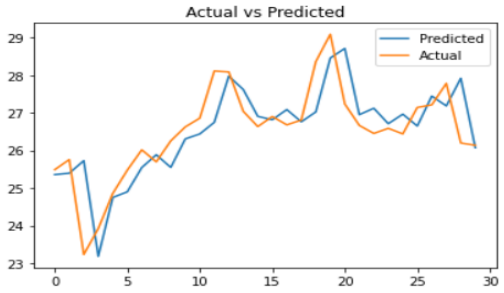
10.3.4 Controller Siemens > Groundfloor > Temperature

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	0.8373	0.7215	0.9151	0.4149
SARIMA m = 12	1.4929	0.9895	1.2218	-0.0431
ARIMA	7.7591	2.5487	2.7855	-4.4218
LSTM	0.6557	0.5915	0.8098	0.5163
CNN	0.7136	0.6638	0.8448	0.2262
LSTM-CNN	0.7415	0.6732	0.8611	0.1746
BILSTM	0.6671	0.5982	0.8168	0.5011
GRU	0.6523	0.5850	0.8077	0.5048
Prophet	1.2912	0.8548	1.1363	0.0978
XGBOOST	0.3720	0.5284	0.6099	0.5273
Random Forest Regression	1.7258	1.0781	1.337	-0.2059
ΠαραλλαγήRandom Forest Regression	0.6810	0.6157	0.8253	0.5241

ΣΥΜΠΕΡΑΣΜΑΤΑ

Την καλύτερη πρόβλεψη παρέχει και πάλι το μοντέλο *GRU*.



Σχήμα 76: GRU Prediction

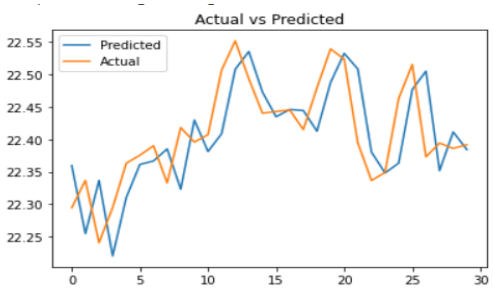
10.3.5 Controller Siemens > Data Room > Temperature

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	0.0038	0.0527	0.0615	0.3352
SARIMA m = 12	0.3793	0.5215	0.0.6158	-65.6495
ARIMA	0.4255	0.6345	0.6523	-73.7798
LSTM	0.6556	0.5914	0.8097	0.5162
CNN	0.0079	0.0759	0.0891	0.4238
LSTM-CNN	0.0106	0.08571	0.1027	0.1397
BILSTM	0.0049	0.0586	0.0704	0.3538
GRU	0.0037	0.0501	0.0610	0.3739
Prophet	0.0312	0.1442	0.1765	-4.4754
XGBOOST	0.0063	0.0663	0.0795	-0.1099
Random Forest Regression	0.0113	0.0889	0.1065	-0.9938
Παραλλαγή Random Forest Regression	0.0049	0.0539	0.0702	0.1352

ΣΥΜΠΕΡΑΣΜΑΤΑ

Την πιο ακριβή πρόβλεψη παρέχει το μοντέλο νευρωνικών δικτύων *GRU*



Σχήμα 77: GRU Prediction

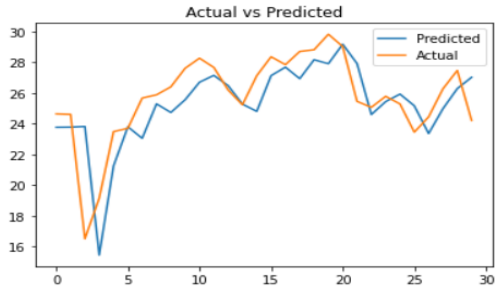
10.3.6 Controller Siemens > Outdoor > Temperature

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	5.7706	1.8586	2.4022	0.2360
SARIMA m = 7	9.0799	2.5071	3.0133	-0.2021
ARIMA	24.3525	4.5350	4.9348	-2.2241
LSTM	3.9671	1.3118	1.9918	0.3863
CNN	5.0708	1.6635	2.2518	0.1084
LSTM-CNN	4.5651	1.5717	2.1366	0.2805
BILSTM	4.5626	1.3667	2.1360	0.3956
GRU	4.2159	1.4851	2.0533	0.3410
Prophet	8.9091	2.6047	2.9848	-0.1795
XGBOOST	7.4407	2.0695	2.7278	0.0149
Random Forest Regression	13.9353	2.9289	3.7330	-0.8449
Παραλλαγή Random Forest Regression	7.1826	1.7105	2.6800	0.0491

ΣΥΜΠΕΡΑΣΜΑΤΑ

Και σε αυτήν την περίπτωση, το *GRU* προέκυψε ως το μοντέλο με τις καλύτερες προβλέψεις.



Σχήμα 78: GRU Prediction

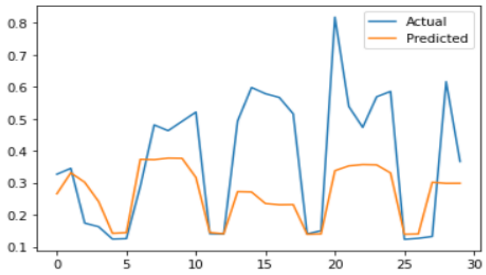
10.3.7 Main Electricity Panel > HVAC > Active Energy

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	0.0369	0.1701	0.1923	0.0832
SARIMA m = 7	0.0459	0.1889	0.2143	-0.1396
ARIMA	0.0663	0.2104	0.2576	-0.6454
LSTM	0.0416	0.1785	0.2038	-5.9506
CNN	0.1790	0.3724	0.4231	0
LSTM-CNN	0.1790	0.3724	0.4231	0
BILSTM	0.0548	0.1865	0.2342	0.1404
GRU	0.0368	0.1466	0.1988	-0.0484
Prophet	21.2538	4.1812	4.6102	-526.1968
XGBOOST	0.0366	0.1425	0.1913	0.0919
Random Forest Regression	0.0393	0.1458	0.1982	0.0257
Παραλλαγή Random Forest Regression	0.0221	0.0970	0.1485	0.4529

ΣΥΜΠΕΡΑΣΜΑΤΑ

Τις καλύτερες προβλέψεις έκανε το *XGBoost* με μικρή διαφορά από το δεύτερο μοντέλο νευρωνικών δικτύων *BILSTM*. Αντίθετα την λιγότερο ακριβή πρόβλεψη κάνει το μοντέλο *PROPHET*.



Σχήμα 79: XGBOOST Prediction

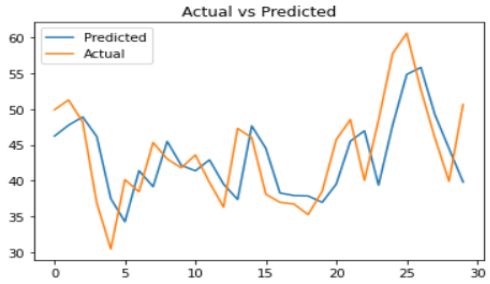
10.3.8 Controller Siemens > Outdoor > Humidity

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	45.4452	5.5982	6.7413	0.0259
SARIMA m = 4	46.2179	5.4770	6.7984	0.0094
ARIMA	51.9612	5.9483	7.2084	-0.1137
LSTM	31.3939	4.7460	5.6030	-0.2287
CNN	45.2402	5.4525	6.7261	-1.7137
LSTM-CNN	43.4939	5.3248	6.5949	-3.4419
BILSTM	30.1117	4.6103	5.4874	-0.1035
GRU	30.8215	4.8565	5.5517	-0.3598
Prophet	48.0619	5.7762	6.9327	-0.0301
XGBOOST	37.9211	5.0944	6.1580	0.1872
Παραλλαγή Random Forest Regression	54.6385	6.2025	7.3918	-0.1711
Random Forest Regression	37.8782	5.2929	6.1545	0.1882

ΣΥΜΠΕΡΑΣΜΑΤΑ

Την καλύτερη πρόβλεψη έκανε το $BILSTM$.



Σχήμα 80: BILSTM Prediction

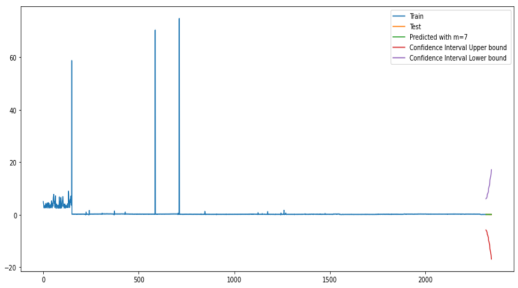
10.3.9 Main Electricity Panel > Lighting Total > Active Energy

Τα σφάλματα που προέκυψαν από τους παραπάνω αλγόριθμους ήταν:

ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΕΩΝ				
Αλγόριθμος Πρόβλεψης	**MSE**	**MAE**	**RMSE**	**R2**
Triple Exponential Smoothing	0.0399	0.0425	0.0686	-1.6102
SARIMA m = 7	0.0005	0.0143	0.02269	0.7139
ARIMA	0.0046	0.0529	0.0678	-1.5562
LSTM	0.0082	0.0798	0.0908	-234.5769
CNN	0.0149	0.1147	0.1223	0
LSTM-CNN	0.0149	0.1147	0.1223	0
BILSTM	0.0115	0.09872	0.1074	-204.68
GRU	0.0043	0.0506	0.0652	-12.2790
Prophet	0.0033	0.0508	0.0571	-0.8138
XGBOOST	0.0016	0.028	0.0394	0.1387
Random Forest Regression	0.0291	0.1679	0.1706	-15.1629
Παραλλαγή Random Forest Regression	0.0005	0.0117	0.0228	0.7105

ΣΥΜΠΕΡΑΣΜΑΤΑ

Το $SARIMA$ είναι το πιο ακριβές μοντέλο σε αυτήν την περίπτωση. Αντίθετα το μοντέλο $LSTMCNN$ και το μοντέλο CNN εμφάνισαν σημαντικά σφάλματα.



Σχήμα 81: SARIMA Prediction

11 Ανάλυση αποτελεσμάτων

Στην ενότητα αυτή θα παρουσιαστούν οι παρατηρήσεις που έγιναν με βάση τα αποτελέσματα των προβλέψεων της προηγούμενης ενότητας. Οι παρατηρήσεις θα γίνουν ως προς τους αλγόριθμους ξεχωριστά και ως προς τα δεδομένα που χρησιμοποιήθηκαν.

11.1 Παρατηρήσεις με βάση τους αλγόριθμους

Όσον αφορά τον κάθε αλγόριθμο ξεχωριστά έγιναν ορατές ορισμένες διαφορές. Ο αλγόριθμος **Τριπλής Εκθετικής Εξομάλυνσης**, γενικά, εφαρμόζεται σε δεδομένα που εμφανίζουν τάση και εποχικότητα και είναι αποτελεσματικότερος όταν τα δεδομένα κινούνται αργά με την πάροδο του χρόνου. Οι χρονοσειρές, όμως, που χρησιμοποιήθηκαν εμφάνιζαν συνήθως παραπάνω από μία εποχικότητα και μεταβάλλονταν με μεγάλη συχνότητα. Γι αυτό το λόγο, συνήθως, ο αλγόριθμος δεν προέβλεπε με μεγάλη ακρίβεια τα δεδομένα. Αντίθετα, στις περιπτώσεις που οι χρονοσειρές εμφάνιζαν ορισμένη τάση και μικρή έως καθόλου εποχικότητα, οι προβλέψεις ήταν αρκετά ακριβείς. Επιπλέον, ακριβώς επειδή έπρεπε να γίνουν οι δοκιμές όλων των συνδυασμών παραμέτρων, οι οποίοι είναι αρκετοί, η διαδικασία ήταν πολύ χρονοβόρα. Παρόλλα αυτά, αν ο συνδυασμός είναι γνωστός, η πρόβλεψη πραγματοποιείται με μεγάλη ταχύτητα.
Ο αλγόριθμος **ARIMA** όπως και ο αλγόριθμος **SARIMA** είναι γραμμικά μοντέλα παλινδόμησης, με αποτέλεσμα να μην εμφανίζουν καλά αποτελέσματα όταν τα δεδομένα δεν εμφανίζουν γραμμική συνάρτηση με τα προηγούμενα. Το μοντέλο ARIMA μάλιστα έκανε τις περισσότερες φορές τη λιγότερο ακριβή πρόβλεψη αφού δεν προβλέπει την ύπαρξη εποχικότητας.
Το πρόβλημα της απαίτησης γραμμικής εξάρτησης μεταξύ των δεδομένων αντιμετώπισαν τα μοντέλα νευρωνικών δικτύων, τα οποία μπορούσαν αποτελεσματικά να εντοπίζουν και εποχικότητα και τάση. Κυρίως το **LSTM**, το **BiLSTM** και το **GRU** ήταν τα μοντέλα που κάναν την πιο ακριβή πρόβλεψη στα περισσότερα σύνολα δεδομένων. Τα τρία αυτά μοντέλα συχνά παρουσίαζαν παρόμοια αποτελέσματα και δεν είναι εύκολο να διακρίνει κανείς σε ποιά περίπτωση είναι κάποιο καλύτερο. Όλα είναι αρκετά καλά. Το κύριο αρνητικό τους είναι ότι απαιτούν έναν έμπειρο forcaster ώστε να επιλεχθούν οι καλύτεροι παράμετροι ενώ επίσης έχει υψηλές απαιτήσεις σε υπολογιστικό χρόνο, αφου πρώτα απαιτείται να εκπαιδευτεί το μοντέλο.
Το **CNN** από την άλλη δεν είχε συνήθως πολύ καλα αποτελέσματα, ιδιαίτερα όταν παρουσίαζε η χρονοσειρά ορισμένη τάση. Σε αντίθεση με ένα LSTM, ένα CNN δεν είναι recurrent, πράγμα που σημαίνει ότι δεν διατηρεί μνήμη προηγούμενων μοτίβων των χρονοσειρών. Το μοντέλο αυτό, εν γένει χρησιμοποιείται περισσότερο ώστε να εκμεταλλεύεται τη «χωρική συσχέτιση» στα δεδομένα, ενώ λειτουργεί πιο αποτελεσματικά σε εικόνες και ομιλίες. Όλα τα νευρωνικά μοντέλα, όμως, έχουν το περιθώριο με διάφορες τεχνικές να εκπαιδευτούν ακόμα καλύτερα και να εμφανίσουν ακόμα καλύτερα αποτελέσματα, κάτι που δεν μπορεί να συμβεί στους στατιστικούς αλγόριθμους.
Ο αλγόριθμος **Prophet** έχει τη δυνατότητα να μπορεί να χειρίζεται ένα σύνολο εποχικοτήτων, γι αυτό τα αποτελέσματα συνολικά ήταν σχετικά καλά σε σχέση με άλλους αλγόριθμους. Μάλιστα, ο forcaster μπορεί να προσθέτει δικές του εποχικότητες για ακόμα πιο ακριβή αποτελέσματα. Επιπλέον, ο χρόνος

που απαιτείται για την πρόβλεψη είναι μικρός σε σχέση με τους προηγούμενους αλγόριθμους που αναφέρθηκαν.
Μικρό χρόνο επίσης απαιτεί και το **XGBOOST**. Οι προβλέψεις στα περισσότερα παραδείγματα ήταν αρκεντά κοντά με το Prophet. Το μοντέλο αυτό όμως δεν αποδίδει τόσο καλά σε μη δομημένα δεδομένα καθώς η ενίσχυση κλίσης είναι πολύ ευαίσθητη σε ακραίες τιμές, αφού κάθε ταξινομητής αναγκάζεται να διορθώσει τα σφάλματα στους learners των προηγούμενων επιπέδων. Επιπλέον, στις περιπτώσεις που οι χρονοσειρές παρουσίαζαν τάση, είχε τα χειρότερα αποτελέσματα.
Η παραλλαγή του **Random Forest** αλγόριθμου εμφάνισε πολύ καλά αποτελέσματα σε σύντομο χρονικό διάστημα. Μάλιστα, τις περισσότερες φορές, ήταν ο δεύτερος πιο ακριβής αλγόριθμος μετά τα μοντέλα νευρωνικών δικτύων. Αντίθετα ο απλός αλγόριθμος δεν είχε συνήθως καλά αποτελέσματα. Παρόλα αυτά ήταν πολύ γρήγορος.

11.2 Παρατηρήσεις με βάση τα δεδομένα

Μία μεγάλη ομάδα των δεδομένων που χρησιμοποιήθηκαν για τη μελέτη ορισμένων αλγορίθμων σχετικά με την πρόβλεψη χρονοσειρών, ήταν η **θερμοκρασία** ορισμένων χώρων στα κτήρια τραπεζών. Πιο συγκεκριμένα, στο πρώτο κατάστημα μελετήθηκαν δύο θερμοκρασίες που αφορούσαν το ισόγειο ,μία για το δωμάτιο των δεδομένων και μία για τον εξωτερικό χώρο. Στο δεύτερο κατάστημα, υπήρχαν δεδομένα για τη θερμοκρασία στο ισόγειο, το υπόγειο, το δωμάτιο δεδομένων και τον εξωτερικό χώρο. Στο τρίτο, χρησιμοποιήθηκαν τα δεδομένα για το ισόγειο, το δωμάτιο δεδομένων και τον εξωτερικό χώρο του κτηρίου. Στο πρώτο, το όργανο που χρησιμοποιήθηκε για τις μετρήσεις ήταν Controller Schneider, ενώ στις άλλες δύο ήταν Controllew Siemens.
Οι χρονοσειρές σε όλες τις μετρήσεις θερμοκρασίας εκτός από αυτές στο δωμάτιο δεδομένων, είχαν πολλές ομοιότητες. Αρχικά δεν εμφάνιζαν τάση και επιπλέον είχαν σταθερή εποχικότητα με πιο φανερή αυτή ανά χρόνο. Το δωμάτιο δεδομένων είχε μία πιο άναρχη μορφή με ορισμένες ακραίες τιμές. Τα νευρωνικά δίκτυα, και κυρίως τα LSTM, BiLSTM και GRU έκαναν τις πιο ακριβείς προβλέψεις. Για το δωμάτιο δεδομένων, που για παράδειγμα στο 2ο κατάστημα δεν εμφάνιζε ορισμένη εποχικότητα, το μοντέλο CNN και CNN-LSTM δεν μπόρεσαν να προβλέψουν με καλή ακρίβεια τις μελλοντικές τιμές της χρονοσειράς. Πολύ κοντά σε αυτά τα αποτελέσματα ήταν οι προβλέψεις της παραλλαγής του αλγόριθμου Random Forest. Το XGBoost και το Prophet είχαν μέτρια αποτελέσματα σε σχέση με τους υπόλοιπους αλγόριθμους. Οι στατιστικοί αλγόριθμοι Triple Exponential Smoothing και SARIMA δεν είχαν κάποια σταθερή θέση. Ορισμένες φορές κάναν πολύ ακριβείς προβλέψεις και άλλοτε προβλέψεις με μεγάλη απόκλιση από τις πραγματικές τιμές. Ο ARIMA ήταν ο χειρότερος με μεγάλη διαφορά τις περισσότερες φορές αφού όλες οι χρονοσειρές θερμοκρασίας παρουσιάζουν εποχικότητα.
Μία άλλη κατηγορία μετρήσεων ήταν οι χρονοσειρές που περιέγραφαν την μέγιστη τιμή της **Active Energy**. Στο πρώτο κατάστημα η ενέργεια αφορούσε τη θέρμανση, τον εξαερισμό και τον κλιματισμό, ενώ στα άλλα δύο την ενέργεια των φώτων και την κύρια ενέργεια. Και στις τέσσερις περιπτώσεις το διάγραμμα είχε φανερή αυξητική τάση. Στην πρώτη περίπτωση υπήρχε μία εποχικότητα, γι αυτό την καλύτερη πρόβλεψη έκανε το μοντέλο GRU. Ακολούθησαν τα υπόλοιπα μοντέλα με λιγότερο ακριβές το ARIMA, που δεν μπορούσε να προβλέψει την εποχικότητα. Στο δεύτερο κατάστημα η μέγιστη ενέργεια

των φώτων εμφάνιζε μηδενική εποχικότητα, οπότε υπήρχε γραμμική εξάρτηση στα δεδομένα. Γι αυτό, τις πιο ακριβές προβλέψεις έκαναν τα στατιστικά μοντέλα με καλύτερο το ARIMA. Τα αποτελέσματα από τα μοντέλα νευρωνικών δικτύων, από την άλλη, ήταν πολύ χαμηλά με χειρότερο το CNN. Παρόμοια αποτελέσματα είχε και η πρόβλεψη της ενέργειας στα φώτα και στην κύρια ενέργεια του κτηρίου του τρίτου καταστήματος. Οι δύο μετρήσεις εμφάνιζαν μία μικρή εποχικότητα. Λόγω αυτής, και στα δύο το μοντέλο ARIMA έκανε τη χειρότερη πρόβλεψη.Τα πιο ακριβή μοντέλα ήταν αυτά των SARIMA και Τριπλής Εκθετικής Εξομάλυνσης.

Επιπλέον στα δύο τελευταία καταστήματα μετρήθηκε η **Active Energy** των φώτων και της θέρμανσης, του εξαερισμού και του κλιματισμού. Και στις δύο περιπτώσεις παρατηρήθηκε ότι τα καλύτερα αποτελέσματα είχαν οι αλγόριθμοι XGBoost,Παραλλαγή του Random Forest, Prophet και SARIMA. Τα νευρωνικά δίκτυα είχαν χαμηλά ποσοστά ακρίβειας. Τα δεδομένα δεν εμφάνιζαν κάποιο φανερό μοτίβο, γεγονός που δυσκόλευε την εκπαίδευση των νευρωνικών μοντέλων.

Στο πρώτο κατάστημα μελετήθηκε η χρονοσειρά που αφορά τη μέτρηση του **feedback**. Οι τιμές εναλάσσονταν από το 1 στο 0 με μεγάλη συχνότητα. Την καλύτερη πρόβλεψη την έκανε το XGBOOST ακολουθώντας οι υπόλοιποι αλγόριθμοι με μικρή διαφορά στην ακρίβεια. Μικρότερη ακρίβεια είχαν οι στατιστικοί αλγόριθμοι με χειρότερο τον αλγόριθμο τριπλής εκθετικής εξομάλυνσης ο οποίος δεν μπόρεσε να βρει έναν καλό συνδυασμό παραμέτρων ώστε να μπορεί να προβλέψει σωστά τη γρήγορη μεταβολή των τιμών.

Στη μέτρηση του πρώτου καταστήματος που αφορά την **Active Power** της θέρμανσης, του εξαερισμου και του κλιματισμού, τα αποτελέσματα ήταν παρόμοια με αυτά στις χρονοσειρές των θερμοκρασιών, αφού και στην περίπτωση αυτή δεν υπάρχει ορισμένα τάση, αλλά εμφανίζοταν εποχικότητα κυρίως ανά χρόνο. Ο αλγόριθμος BILSTM έκανε την καλύτερη πρόβλεψη, με μικρή απόκλιση από τους υπόλοιπους αλγόριθμους. Μικρότερη ακρίβεια είχαν οι στατιστικοί αλγόριθμοι με χειρότερο τον ARIMA.

Στο πρώτο και το τρίτο κατάστημα, επιπλέον, έγινε μελέτη του **ποσοστού υγρασίας** στο ισόγειο και στον εξωτερικό χώρο αντίστοιχα. Η χρονοσειρά είχε σταθερή εποχικότητα με αποτέλεσμα να κάνουν την καλύτερη πρόβλεψη τα νευρωνικά δίκτυα. Ακόμα, παρουσιάστηκε η **ολική αρμονική παραμόρφωση** στην τάση και στη γείωση του φωτισμού, χρονοσειρές με έντονες αλλαγές στις τιμές των δεδομένων. Όλα τα μοντέλα είχαν κοντινά αποτελέσματα.

Τέλος, στο δεύτερο και στο τρίτο κατάστημα μελετήθηκε η ποσότητα διοξειδίου του άνθρακα. Η γραφική παράσταση παρουσίαζε μεταβαλλόμενη εποχικότητα με αποτέλεσμα τα μοντέλα νευρωνικών δικτύων να κάνουν τις πιο ακριβής προβλέψεις και κυρίως τα LSTM και GRU ενώ οι ARIMA και τριπλής εκθετικής εξομάλυνσης παρουσίασαν τα πιο χαμηλά αποτελέσματα.

12 Ensemble μεθόδων για την πρόβλεψη χρονοσειρών

Οι τεχνικές Ensemble που χρησιμοποιήθηκαν στην παρούσα εργασία ήταν το Bagging, η μέθοδος υπολογισμού μέσου όρου και η στοίβαξη.

12.1 Bagging

Η πρώτη μέθοδος που χρησιμοποιήθηκε ήταν το Bagging. Εφαρμόστηκε στον αλγόριθμο XGBOOST.

12.1.1 Κατάστημα 1ο

Αποτελέσματα πριν και μετά tobagging				
Δεδομένα	*RMSE* **πριν**	*RMSE* **μετά**	R^2 **πριν**	R^2 **μετά**
Groundfloor > Temperature2	0.8350	1.1420	0.3696	-0.1792
Groundfloor > Temperature1	0.8331	0.4675	1.0658	0.1283
Data Room > DataRoom Temp	0.1459	0.1438	0.1633	0.1869
Feedback > AC Feedback	0.1616	0.1766	0.8756	0.8514
HVAC > Active Power	0.6287	0.6330	0.6796	0.6753
max HVAC > Active Energy	266.3682	287.1727	0.5686	0.4986
Groundfloor > Humidity	6.2939	7.2947	0.1148	-0.1892
Outdoor > Temperature	2.4451	2.4451	0.0256	0.0256
Rest + Lighting > THD Voltage LN Avg	0.1398	0.1356	0.0061	0.0636
Rest + Lighting > THD Current Neutral	40.4492	39.8043	0.2170	0.2418

12.1.2 Κατάστημα 2ο

Αποτελέσματα πριν και μετά tobagging				
Δεδομένα	*RMSE* **πριν**	*RMSE* **μετά**	R^2 **πριν**	R^2 **μετά**
max Active Energy > Lighting Energy	36.5323	38.0149	-3.3570	-3.7179
Basement > Temperature	0.2040	0.2450	0.6936	0.5583
Groundfloor > CO2	27.7060	35.7991	0.4637	0.1046
Groundfloor > Temperature	0.6099	0.6028	0.5273	0.5384
Data Room > Temperature	0.0502	0.0503	0.4976	0.4944
Outdoor > Temperature	2.3609	2.4584	-0.1457	-0.2422
HVAC > Active Energy	0.2422	0.2430	0.6872	0.6848
Lighting Total > Active Energy	0.0499	0.0524	0.7941	0.7728

12.1.3 Κατάστημα 3ο

Αποτελέσματα πριν και μετά tobagging				
Δεδομένα	*RMSE* πριν	*RMSE* μετά	R^2 πριν	R^2 μετά
max Active Energy > Main Energy	1644.9709	1644.9509	-2.4656	-2.4679
max Active Energy > Lighting Energy	146.9749	147.3865	-3.1289	-3.1519
Groundfloor > CO2	87.0277	97.4751	0.2301	0.0341
Groundfloor > Temperature	1.1363	1.1143	0.0978	0.1324
Data Room > Temperature	0.0795	0.0765	-0.1099	-0.0300
Outdoor > Temperature	2.7278	2.7488	0.0149	-0.0003
HVAC > Active Energy	0.1913	0.1986	0.0919	0.0210
Outdoor > Humidity	6.1580	6.2080	0.1872	0.1740
Lighting Total > Active Energy	0.0394	0.0383	0.1387	0.1839

12.1.4 Παρατηρήσεις

Σύμφωνα με τα παραπάνω αποτελέσματα, η μέθοδος ensamble baging, δεν βελτίωσε την προβλεπτική ικανότητα του αλγορίθμου XGBOOST. Αντιθέτως, τις περισσότερες φορές η ακρίβεια στις προβλέψεις του ήταν λίγο χαμηλότερη από τον απλό αλγόριθμο, με ορισμένες εξαιρέσεις.

12.2 Μέθοδος υπολογισμού μέσου όρου

Η δεύτερη μέθοδος που μελετήθηκε ήταν αυτή του υπολογισμού μέσου όρου. Πραγματοποιήθηκε η εκπαίδευση δύο μοντέλων XGBOOST και της παραλλαγής του Random Forest. Στο τέλος υπολογίστηκε ο μέσος όρος των προβλέψεών τους.

12.2.1 Κατάστημα 1ο

Αποτελέσματα της Μεθόδου Μέσου Όρου			
Δεδομένα	*RMSE XGBOOST*	*RMSE Random Forest*	*RMSE Average Method*
Groundfloor > Temperature2	0.8350	0.5492	0.6121
Groundfloor > Temperature1	0.8331	0.7662	0.6934
Data Room > DataRoom Temp	0.1459	0.1090	0.1136
Feedback > AC Feedback	0.1616	0.1784	0.1666
HVAC > Active Power	0.6287	0.6426	0.6076
max HVAC > Active Energy	266.3682	75.0008	127.9458
Groundfloor > Humidity	6.2939	5.5196	5.1629
Outdoor > Temperature	2.4451	1.9935	1.9009
Rest + Lighting > THD Voltage LN Avg	0.1398	0.1355	0.1192
Rest + Lighting > THD Current Neutral	40.4492	32.7385	33.5377

12.2.2 Κατάστημα 2ο

Αποτελέσματα της Μεθόδου Μέσου Όρου			
Δεδομένα	*RMSE XGBOOST*	*RMSE Random Forest*	*RMSE Average Method*
max Active Energy > Lighting Energy	36.5323	3.20732	19.8392
Basement > Temperature	0.2040	0.0842	0.1467
Groundfloor > CO2	27.7060	27.165	26.6263
Groundfloor > Temperature	0.6099	0.5310	0.4841
Data Room > Temperature	0.0502	0.0476	0.0435
Outdoor > Temperature	2.3609	1.5333	1.4925
HVAC > Active Energy	0. 2422	0.1629	0.1759
Lighting Total > Active Energy	0.0499	0.0527	0.0510

12.2.3 Κατάστημα 3ο

Αποτελέσματα της Μεθόδου Μέσου Όρου			
Δεδομένα	*RMSE XGBOOST*	*RMSE Random Forest*	*RMSE Average Method*
max Active Energy > Main Energy	1644.9709	155.8152	888.6345
max Active Energy > Lighting Energy	146.9749	13.5675	79.0483
Groundfloor > CO2	87.0277	87.5929	86.7706
Groundfloor > Temperature	1.1363	0.8253	0.8490
Data Room > Temperature	0.0795	0.0702	0.0638
Outdoor > Temperature	2.7278	2.6800	2.0504
HVAC > Active Energy	0.1913	0. 1485	0.1428
Outdoor > Humidity	6.1580	6.1817	5.3146
Lighting Total > Active Energy	0.0394	0.0228	0.0275

12.2.4 Παρατηρήσεις

Με βάση τα παραπάνω αποτελέσματα, γίνεται ορατό ότι η μέθοδος ensamble υπολογισμού μέσου όρου λειτούργησε αποτελεσματικά. Τις περισσότερες φορές τα αποτελέσματα που προέκυψαν από τις προβλέψεις ήταν καλύτερα από εκείνα που προκύπτουν από τη δράση του κάθε αλγόριθμου ξεχωριστά. Οι μόνες περιπτώσεις που δεν ίσχυσε αυτό ήταν εκείνες που ένας από τους δύο αλγόριθμους παρουσίαζε μεγάλο σφάλμα. Τέλος, επειδή οι δύο αλγόριθμοι είναι ιδιαίτερα γρήγοροι στις προβλέψεις, ο συνδυασμός τους δεν προσθέτει ορισμένη καθυστέρηση.

12.3 Στοίβαξη

Η στοίβαξη (stacking) που πραγματοποιήθηκε βασίστηκε στα μοντέλα LSTM, Prophet και Random Forest. Στη συνέχεια η πρόβλεψη των μοντέλων αυτών αποτέλεσε είσοδο σε έναν Random Forest estimator ο οποίος πραγματοποίησε τις τελικές προβλέψεις.

12.3.1 Κατάστημα 1ο

Αποτελέσματα της Μεθόδου Στοίβαξης		
Δεδομένα	*RMSE*	R^2
Groundfloor > Temperature2	0.7216	0.5292
Groundfloor > Temperature1	0.6925	0.6319
Data Room > DataRoom Temp	0.1057	0.5611
Feedback > AC Feedback	0.8413	0.1876
HVAC > Active Power	0.6909	0.6469
max HVAC > Active Energy	806.7046	-2.9567
Groundfloor > Humidity	4.9663	0.4488
Outdoor > Temperature	1.9274	0.3946
Rest + Lighting > THD Voltage LN Avg	0.1305	0.1338
Rest + Lighting > THD Current Neutral	30.1818	0.5641

12.3.2 Κατάστημα 2ο

Αποτελέσματα της Μεθόδου Στοίβαξης		
Δεδομένα	*RMSE*	R^2
max Active Energy > Lighting Energy	37.0790	-3.4884
Basement > Temperature	0.0828	0.9495
Groundfloor > CO2	24.0120	0.5972
Groundfloor > Temperature	0.4208	0.7750
Data Room > Temperature	0.0799	-0.2745
Outdoor > Temperature	1.6130	0.4653
HVAC > Active Energy	0. 6000	-0.6834
Lighting Total > Active Energy	0.1758	-1.5467

12.3.3 Κατάστημα 3ο

Αποτελέσματα της Μεθόδου Στοίβαξης		
Δεδομένα	$RMSE$	R^2
max Active Energy > Main Energy	2858072.2540	-2.8850
max Active Energy > Lighting Energy	23645.2854	-3.4526
Groundfloor > CO2	83.1140	0.2978
Groundfloor > Temperature	1.1103	-1.1397
Data Room > Temperature	0.0765	0.5901
Outdoor > Temperature	2.0933	0.4199
HVAC > Active Energy	0.2684	-0.7875
Outdoor > Humidity	5.9265	0.2472
Lighting Total > Active Energy	0.149	0

12.3.4 Παρατηρήσεις

Γίνεται αντιληπτό, ότι η μέθοδος stacking λειτούργησε στις περισσότρες περιπτώσεις ιδιαίτερα αποτελεσματικά. Τα αποτελέσματα, και πάλι, ήταν καλύτερα από εκείνα που προκύπτουν από τη δράση του κάθε αλγόριθμου ξεχωριστά. Μάλιστα τις περισσότερες φορές ήταν ο πιο ακριβής αλγόριθμος σε σύγκριση και με τους υπόλοιπους που μελετήθηκαν στην παρούσα εργασία. Οι μόνες περιπτώσεις που δεν ίσχυσε αυτό ήταν εκείνες που ένας από τους δύο αλγόριθμους παρουσίαζε μεγάλο σφάλμα.

13 Συμπεράσματα

Στα πλαίσια αυτής της εργασίας έγινε εφαρμογή ορισμένων αλγορίθμων για την πρόβλεψη χρονοσειρών σε δεδομένα που έχουν συλλεχθεί από καταστήματα τραπεζών. Με βάση αυτά τα αποτελέσματα επιχειρήθηκε να οριστεί ένας ¨αλγόριθμος' ο οποίος με βάση τη μορφή των δεδομένων, αν για παράδειγμα εμφανίζουν εποχικότητα ή τάση, να γίνεται αυτόματα η επιλογή του κατάλληλου μοντέλου πρόβλεψης. Έγινε χρήση των δεδομένων από τρία διαφορετικά καταστήματα και μελετήθηκαν συνολικά είκοσι επτά σύνολα δεδομένων. Παρατηρήθηκε ότι όταν η χρονοσειρά εμφανίζει σταθερή εποχικότητα αλλά όχι τάση, όπως είναι οι περιπτώσεις μέτρησης θερμοκρασίας ή υγρασίας, τα νευρωνικά δίκτυα κάνουν τις πιο ακριβείς προβλέψεις. Παρόλα αυτά, αντίστοιχα καλές προβλέψεις κάνει και η παραλλαγή του Random Forest το οποίο απαιτεί λιγότερο χρόνο και εμπειρία για την εκπαίδευση του μοντέλου. Συνεπώς το τελευταίο μοντέλο θα μπορούσε να θεωρηθεί κατάλληλο.

Όταν τα δεδομένα εμφανίζουν έντονη τάση και χαμηλή εποχικότητα, όπως για παράδειγμα είναι οι περιπτώσεις που οι μετρήσεις αφορούν τη μέγιστη τιμή της Active Energy ανά ημέρα, το κατάλληλο μοντέλο θα μπορούσε να θεωρηθεί το SARIMA το οποίο είναι ιδιαίτερα αποδοτικό αλλά και γρήγορο.

Όσον αφορά τις χρονοσειρές που δεν έχουν κάποια φανερή εποχικότητα και δεν εμφανίζουν τάση, τα μοντέλα XGBoost, Random Forest, Prophet είναι αποτελεσματικά και ταχύτατα.

Στη συνέχεια δοκιμάστηκαν ορισμένοι μέθοδοι ensemble όπως είναι το baging, η μέθοδος υπολογισμού μέσου όρου και η μέθοδος stacking. Η πρώτη δεν εμφάνισε καλά αποτελέσματα, σε αντίθεση με τις άλλες δύο που βελτίωσαν τα αποτελέσματα σχεδόν σε όλα τα σύνολα δεδομένων. Μάλιστα, στην περίπτωση της εύρεσης μέσου όρου, επειδή τα δύο μοντέλα που χρησιμοποίηθηκαν μπορούν να προβλέπουν τις χρονοσειρές σε μικρό χρονικό διάστημα, η μέθοδος αυτή ήταν εξίσου γρήγορη.

Συνολικά, όμως, δεν μπορεί να εξαχθεί κάποιο συμπέρασμα με απόλυτη βεβαιότητα καθώς έγινε δοκιμή σε ένα μικρό σύνολο από δεδομένα.

14 Μελλοντική Επέκταση

Σε μία επέκταση της παρούσας εργασίας είναι δυνατή η εφαρμογή περισσότερων αλγορίθμων πρόβλεψης όπως είναι οι TCN, deep TCN, nbits, TFT. Ιδιαίτερα στην περίπτωση της στοίβαξης η προσθήκη του μοντλεου TCN στο Επίπεδο-0 θα έφερνε ακόμα καλύτερα αποτελέσματα.
Επιπλέον, τα αποτελέσματα της πρόβλεψης των περισσότερων αλγορίθμων θα ήταν ακόμα πιο ακριβή αν χρησιμοποιούντουσαν μέθοδοι βελτιστοποίησης. Δύο παραδείγματα αποτελούν το grid search(αναζήτηση πλέγματος) και το kfold.
Τέλος, τα αποτελέσματα της εργασίας θα μπορούσαν να οδηγήσουν στη δημιουργία ενός μοντέλου το οποίο με είσοδο τα στοιχεία μίας χρονοσειρά σχετικά με την τάση και την εποχικότητα, να εκπαιδεύεται και να επιλέγει από μόνο του ποιον αλγόριθμο θα χρησιμοποιησει για τη διαδικασία της πρόβλεψης.

Αναφορές

[1] Li, S., Xu, L.D. Zhao, S., "The internet of things: a survey", Inf Syst Front 17, 243–259 (2015). https://doi.org/10.1007/s10796-014-9492-7

[2] Pradyumna Gokhale, Omkar Bhat Sagar Bhat, "Introduction to IOT", IARJSET ISSN (Online) 2393-8021, Vol. 5, Issue 1, January 2018.

[3] A. A. Cook, G. Mısırlı Z. Fan, 'Anomaly Detection for IoT Time-Series Data: A Survey", IEEE Internet of Things Journal, vol. 7, no. 7, pp. 6481-6494, July 2020, doi: 10.1109/JIOT.2019.2958185

[4] https://ismguide.com/the-internet-of-things/

[5] Somayya Madakam, R. Ramaswamy, Siddharth Tripathi (2015) "Internet of Things (IoT): A Literature Review", Journal of Computer and Communications,03,164-173, doi: 10.4236/jcc.2015.35021

[6] M. H. Miraz, M. Ali, P. S. Excell R. Picking, "A review on Internet of Things (IoT), Internet of Everything (IoE) and Internet of Nano Things (IoNT)," 2015 Internet Technologies and Applications (ITA), Wrexham, UK, 2015, pp. 219-224, doi: 10.1109/ITechA.2015.7317398.

[7] Raghavendra Kumar, Pardeep Kumar Yugal Kumar, "Time Series Data Prediction using IoT and Machine Learning Technique", Procedia Computer Science, 2020, Pages 373-381, ISSN 1877-0509, https://doi.org/10.1016/j.procs.2020.03.240.

[8] Chatfield, C. (2000), Time-Series Forecasting (1st ed.), Chapman and Hall/CRC, https://doi.org/10.1201/9781420036206

[9] G.Peter Zhang, "Time series forecasting using a hybrid ARIMA and neural network model", Neurocomputing, Volume 50, 2003, Pages 159-175, ISSN 0925-2312, https://doi.org/10.1016/S0925-2312(01)00702-0.

[10] Lima, Susana Gonçalves, A. Manuela Costa, Marco. (2019), "Time series forecasting using Holt-Winters exponential smoothing: An application to economic data", AIP Conference Proceedings. 2186. 090003, doi: 10.1063/1.5137999.

[11] B V Vishwas Ashish Patel, "Hands-on Time Series Analysis with Python: From Basics to Bleeding Edge Techniques ", https://doi.org/10.1007/978-1-4842-5992-4, 2020

[12] Tao Lin, Tian Guo, Karl Aberer, 2017, "Hybrid neural networks for learning the trend in time series", In Proceedings of the 26th International Joint Conference on Artificial Intelligence (IJCAI'17), AAAI Press, 2273–2279

[13] Jonathan D. Cryer Kung-Sik Chan,"Time Series Analysis With Series Title", Springer Texts in Statistics, https://doi.org/10.1007/978-0-387-75959-3, Springer New York, NY, 2011

[14] Petrakova, Aleksandra Affenzeller, Michael Merkuryeva, Galina, 2015, "Heterogeneous versus Homogeneous Machine Learning Ensembles", Information Technology and Management Science

[15] Oded Maimon Lior Rokach, "Data Mining and Knowledge Discovery Handbook", 2010

[16] Usama Fayyad, Gregory Piatetsky-Shapiro, Padhraic Smyth, 1996, " The KDD process for extracting useful knowledge from volumes of data", Nov. 1996, 27–34. https://doi.org/10.1145/240455.240464

[17] Last, Mark Klein, Yaron Kandel, Abraham, 2001, "Knowledge discovery in time series databases", IEEE transactions on systems, man, and cybernetics, Part B, Cybernetics : a publication of the IEEE Systems, Man, and Cybernetics Society, 31. 160-9. 10.1109/3477.907576.

[18] Alfred, R. Dimitar Kazakov. "Knowledge Discovery : Enhancing Data Mining and Decision Support Integration." (2005).

[19] Holzinger, (2013), "Human-Computer Interaction and Knowledge Discovery (HCI-KDD): What Is the Benefit of Bringing Those Two Fields to Work Together?", In: Cuzzocrea, A., Kittl, C., Simos, D.E., Weippl, E., Xu, L. (eds) Availability, Reliability, and Security in Information Systems and HCI, CD-ARES 2013, Lecture Notes in Computer Science, vol 8127, Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-40511-2_22

[20] S. Siami-Namini, N. Tavakoli and A. Siami Namin, "A Comparison of ARIMA and LSTM in Forecasting Time Series," 2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA), Orlando, FL, USA, 2018, pp. 1394-1401, doi: 10.1109/ICMLA.2018.00227.

[21] Yanrui Ning, Hossein Kazemi, Pejman Tahmasebi, "A comparative machine learning study for time series oil production forecasting: ARIMA, LSTM, and Prophet", Computers Geosciences, Volume 164, 2022, 105126, ISSN 0098-3004, https://doi.org/10.1016/j.cageo.2022.105126

[22] Kane, M.J., Price, N., Scotch, M. et al. "Comparison of ARIMA and Random Forest time series models for prediction of avian influenza H5N1 outbreaks". BMC Bioinformatics 15, 276 (2014). https://doi.org/10.1186/1471-2105-15-276

[23] Brown Robert Goodell, "Smoothing Forecasting and Prediction of Discrete Time Series", Dover Publ 2004

[24] Greg Rafferty, "Forecasting Time Series Data with Facebook Prophet", March 2021, Packt Publishing Ltd.

[25] S.L. Ho, M. Xie, "The use of ARIMA models for reliability forecasting and analysis", Computers Industrial Engineering, 1998, https://doi.org/10.1016/S0360-8352(98)00066-7

[26] Gocheva-Ilieva, Snezhana Ivanov, A. Voynikova, Desislava Boyadzhiev, Doychin. (2013). "Time series analysis and forecasting for air pollution in small urban area: An SARIMA and factor analysis approach". Stochastic Environmental Research and Risk Assessment. 28. 1045-1060. 10.1007/s00477-013-0800-4.

[27] https://www.researchgate.net/figure/The-common-ensemble-architecture$_fig1_293194221Dabral, P.P., Murry, M.$ $//doi.org/10.1007/s40710-017-0226-y$

[28] A. E. Permanasari, I. Hidayah and I. A. Bustoni, "SARIMA (Seasonal ARIMA) implementation on time series to forecast the number of Malaria incidence," 2013 International Conference on Information Technology and Electrical Engineering (ICITEE), Yogyakarta, Indonesia, 2013, pp. 203-207, doi: 10.1109/ICITEED.2013.6676239.

[29] A. Jain, T. Sukhdeve, H. Gadia, S. P. Sahu and S. Verma, "COVID19 Prediction using Time Series Analysis," 2021 International Conference on Artificial Intelligence and Smart Systems (ICAIS), Coimbatore, India, 2021, pp. 1599-1606, doi: 10.1109/ICAIS50930.2021.9395877

[30] Jan G. De Gooijer, Rob J. Hyndman, "25 years of time series forecasting", International Journal of Forecasting, 2006, https://doi.org/10.1016/j.ijforecast.2006.01.001.

[31] Bishop Chris M.,"Neural networks and their applications", https://doi.org/10.1063/1.1144830, 2023/02/16

[32] Sahoo, B.B., Jha, R., Singh, A. et al. Long short-term memory (LSTM) recurrent neural network for low-flow hydrological time series forecasting. Acta Geophys. 67, 1471–1481 (2019). https://doi.org/10.1007/s11600-019-00330-1

[33] Yasi Wang, Hongxun Yao, Sicheng Zhao, "Auto-encoder based dimensionality reduction", Neurocomputing, 2016, https://doi.org/10.1016/j.neucom.2015.08.104.

[34] Huang, Zhiheng and Xu, Wei and Yu, Kai, Bidirectional LSTM-CRF Models for Sequence Tagging, 2015, https://doi.org/10.48550/arxiv.1508.01991,

[35] K. A. Althelaya, E. -S. M. El-Alfy and S. Mohammed, "Evaluation of bidirectional LSTM for short- and long-term stock market prediction," 2018 9th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 2018, pp. 151-156, doi: 10.1109/IACS.2018.8355458.

[36] H. Jahangir, H. Tayarani, S. S. Gougheri, M. A. Golkar, A. Ahmadian and A. Elkamel, "Deep Learning-Based Forecasting Approach in Smart Grids With Microclustering and Bidirectional LSTM Network," in IEEE Transactions on Industrial Electronics, vol. 68, no. 9, pp. 8298-8309, Sept. 2021, doi: 10.1109/TIE.2020.3009604.

[37] Peter T. Yamak, Li Yujian, and Pius K. Gadosey. 2020. A Comparison between ARIMA, LSTM, and GRU for Time Series Forecasting. In Proceedings of the 2019 2nd International Conference on Algorithms, Computing and Artificial Intelligence (ACAI 2019). Association for Computing Machinery, New York, NY, USA, 49–55. https://doi.org/10.1145/3377713.3377722

[38] Petneházi Gábor, Recurrent Neural Networks for Time Series Forecasting, 2019, https://doi.org/10.48550/arxiv.1901.00069

[39] Borovykh, Anastasia and Bohte, Sander and Oosterlee, Cornelis W., Conditional Time Series Forecasting with Convolutional Neural Networks, 2017 https://doi.org/10.48550/arxiv.1703.04691

[40] B. Kumar Jha and S. Pande, "Time Series Forecasting Model for Supermarket Sales using FB-Prophet," 2021 5th International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2021, pp. 547-554, doi: 10.1109/ICCMC51019.2021.9418033.

[41] Hewage, P., Behera, A., Trovati, M. et al. Temporal convolutional neural (TCN) network for an effective weather forecasting using time-series data from the local weather station. Soft Comput 24, 16453–16482 (2020). https://doi.org/10.1007/s00500-020-04954-0

[42] Lara-Benítez, Pedro, Manuel Carranza-García, José M. Luna-Romera, and José C. Riquelme. 2020. "Temporal Convolutional Networks Applied to Energy-Related Time Series Forecasting" Applied Sciences 10, no. 7: 2322. https://doi.org/10.3390/app10072322

[43] Jianping Li, Jun Hao, QianQian Feng, Xiaolei Sun, Mingxi Liu, Optimal selection of heterogeneous ensemble strategies of time series forecasting with multi-objective programming, Expert Systems with Applications, 2021, https://doi.org/10.1016/j.eswa.2020.114091.

[44] Bühlmann, P. (2012). Bagging, Boosting and Ensemble Methods. In: Gentle, J., Härdle, W., Mori, Y. (eds) Handbook of Computational Statistics. Springer Handbooks of Computational Statistics. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-21551-3_33

[45] Joseph Rocca, Ensemble methods: bagging, boosting and stacking, 2019, Towards Data Science

[46] Livieris, Ioannis E., Emmanuel Pintelas, Stavros Stavroyiannis, and Panagiotis Pintelas. 2020. "Ensemble Deep Learning Models for Forecasting Cryptocurrency Time-Series" Algorithms 13, no. 5: 121. https://doi.org/10.3390/a13050121

[47] Lateko, Andi A. H., Hong-Tzer Yang, Chao-Ming Huang, Happy Aprillia, Che-Yuan Hsu, Jie-Lun Zhong, and Nguy˜n H. Phuong. 2021. "Stacking Ensemble Method with the RNN Meta-Learner for Short-Term PV Power Forecasting" Energies 14, no. 16: 4733. https://doi.org/10.3390/en14164733

[48] Rokach, L. (2005). Ensemble Methods for Classifiers. In: Maimon, O., Rokach, L. (eds) Data Mining and Knowledge Discovery Handbook. Springer, Boston, MA. https://doi.org/10.1007/0-387-25465-X_45

[49] B. S. Everitt, "Introduction to Optimization Methods and their Application in Statistics", 06 December 2012, https://doi.org/10.1007/978-94-009-3153-4

[50] Slawomir Koziel, Xin-She Yang, " Computational Optimization, Methods and Algorithms ", 17 June 2011, https://doi.org/10.1007/978-3-642-20859-1

[51] Singiresu S Rao, “Engineering Optimization Theory and Practice”, Fifth Edition, 22 October 2019

[52] Liashchynskyi, Petro and Liashchynskyi, Pavlo, “Grid Search, Random Search, Genetic Algorithm: A Big Comparison for NAS”, 2019, https://doi.org/10.48550/arxiv.1912.06059,

[53] https://maelfabien.github.io/machinelearning/Explorium4/what-is-hyperparameter-optimization

[54] Lyu Z, Yu Y, Samali B, Rashidi M, Mohammadi M, Nguyen TN, Nguyen A. Back-Propagation Neural Network Optimized by K-Fold Cross-Validation for Prediction of Torsional Strength of Reinforced Concrete Beam. Materials (Basel). 2022 Feb 16;15(4):1477. doi: 10.3390/ma15041477.

[55] Ke-Lin Du, M. N. S. Swamy, Neural Networks and Statistical Learning, 25 September 2020, https://doi.org/10.1007/978-1-4471-7452-3

[56] Dhaliwal, Sukhpreet Singh, Abdullah-Al Nahid, and Robert Abbas. 2018. ”Effective Intrusion Detection System Using XGBoost” Information 9, no. 7: 149. https://doi.org/10.3390/info9070149

[57] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD ’16). Association for Computing Machinery, New York, NY, USA, 785–794. https://doi.org/10.1145/2939672.2939785

[58] Zhang, Lingyu, Bian, Wenjie, Qu, Wenyi, Tuo, Liheng, Wang, Yunhai, Time series forecast of sales volume based on XGBoost, Journal of Physics: Conference Series, 2021, https://dx.doi.org/10.1088/1742-6596/1873/1/012067

[59] Tianqi Chen, Tong He, xgboost: eXtreme Gradient Boosting, Package Version: 1.1.1.1, June 11, 2020

[60] Naing, W.Y.N. Htike, Z.Z.. (2015). Forecasting of monthly temperature variations using random forests. 10. 10109-10112.

[61] Adriansson, N., Mattsson, I. (2015). Forecasting GDP Growth, or How Can Random Forests Improve Predictions in Economics? (Dissertation). Retrieved from http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-243028

[62] ROBERT HECHT-NIELSEN, III.3 - Theory of the Backpropagation Neural Network, June 1989, Academic Press, https://doi.org/10.1016/B978-0-12-741252-8.50010-8

[63] Barry J. Wythoff, Backpropagation neural networks: A tutorial, Chemometrics and Intelligent Laboratory Systems, 1993, https://doi.org/10.1016/0169-7439(93)80052-J

[64] A.T.C. Goh, Back-propagation neural networks for modeling complex systems, Artificial Intelligence in Engineering, 1995, https://doi.org/10.1016/0954-1810(94)00011-S

[65] https://medium.com/analytics-vidhya/mae-mse-rmse-coefficient-of-determination-adjusted-r-squared-which-metric-is-better-cd0326a5697e