



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ
ΔΙΑΤΑΞΕΩΝ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

Ανάπτυξη Μοντέλων Μηχανικής Μάθησης για την
Εκτίμηση της Ενεργειακής Εξοικονόμησης από
Δράσεις Ανακαίνισης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Σιαφαρίκας Δ. Χρήστος

Επιβλέπων: Βαγγέλης Μαρινάκης, Επίκουρος Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2023



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ
ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΒΙΟΜΗΧΑΝΙΚΩΝ
ΔΙΑΤΑΞΕΩΝ ΚΑΙ ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

Ανάπτυξη Μοντέλων Μηχανικής Μάθησης για την Εκτίμηση της Ενεργειακής Εξοικονόμησης από Δράσεις Ανακαίνισης

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Σιαφαρίκας Δ. Χρήστος

Επιβλέπων: Βαγγέλης Μαρινάκης, Επίκουρος Καθηγητής Ε.Μ.Π.

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την 10 Μαρτίου 2023.

.....

Β. Μαρινάκης

Επ. Καθηγητής Ε.Μ.Π.

.....

Χ. Δούκας

Αν. Καθηγητής Ε.Μ.Π.

.....

Ι. Ψαρράς

Καθηγητής Ε.Μ.Π.

Αθήνα, Μάρτιος 2023

.....
Σιαφαρίκας Δ. Χρήστος

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

Copyright © Χρήστος Σιαφαρίκας, 2023.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ' ολόκληρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

Πρόλογος

Η διπλωματική αυτή αφιερώνεται στους ανθρώπους που στάθηκαν δίπλα μου από τη πρώτη μέρα που ξεκίνησα την πορεία μου σε αυτή τη σχολή μέχρι και την ολοκλήρωση της διπλωματικής εργασίας.

Αρχικά οφείλω να ευχαριστήσω τον κ. Ελισσαίο Σαρμά για την βοήθεια και την άψογη συνεργασία που είχαμε από την αρχή της διπλωματικής εργασίας και τη συνεχή επικοινωνία που με βοήθησε σε όποιο σημείο αντιμετώπισα κάποια δυσκολία κατά την εκπόνηση της.

Στη συνέχεια θα ήθελα να σταθώ στους ανθρώπους που γνώρισα μέσα στη σχολή αλλά και στους φίλους μου, οι οποίοι με βοήθησαν να αποκτήσω σημαντικές εμπειρίες στα φοιτητικά μου χρόνια ενώ παράλληλα με τον τρόπο σκέψης τους στάθηκαν αρωγοί στη διεύρυνση και του δικού μου τρόπου σκέψης.

Τέλος ιδιαίτερη αναφορά πρέπει να γίνει και στην οικογένεια μου, η οποία με βοήθησε και συνεχίζει να με βοηθά όποτε έχω να αντιμετωπίσω κάποιο πρόβλημα ή δυσκολία.

Σιαφάρικας Χρήστος

Αθήνα, Μάρτιος 2023

Περίληψη

Η παρούσα διπλωματική πραγματεύεται το πρόβλημα της εύρεσης του κέρδους από την εξοικονόμηση ενέργειας σε ένα κτίριο στο οποίο έχουν πραγματοποιηθεί δράσεις ανακαίνισης. Συγκεκριμένα ερευνάται το κομμάτι της βελτιστοποίησης των προβλέψεων της κατανάλωσης που θα είχε το κτίριο αν δεν είχε προηγηθεί η ανακαίνιση αυτή. Εφόσον είναι γνωστές οι καταναλώσεις στο διάστημα πριν την ανακαίνιση όπως επίσης και οι εξωτερικές συνθήκες μετά από αυτήν υπάρχουν τα απαραίτητα δεδομένα ώστε με εφαρμογή σε κατάλληλα μοντέλα να παραχθούν ακριβείς και χρήσιμες προβλέψεις.

Στη παρούσα διπλωματική προτείνονται διαφορά μοντέλα μηχανικής μάθησης τα οποία όμως προκύπτουν από συνδυασμό τριών απλών μοντέλων. Το πρώτο από αυτά τα τρία είναι το XGBOOST δηλαδή ένα μοντέλο που λειτουργεί με βάση τους κανόνες του αλγορίθμου gradient boosting και με αυτό το τρόπο πετυχαίνει αρκετά καλές προβλέψεις ακόμα και μόνο του. Τα άλλα δύο είναι δύο LSTM, δηλαδή νευρωνικά δίκτυα που έχουν τη δυνατότητα να έχουν μια μικρή μνήμη και άρα να μπορούν να εντοπίσουν πολύπλοκες σχέσεις μεταξύ των δεδομένων. Ο λόγος που επιλέχθηκε η χρήση δυο LSTM είναι ότι χρησιμοποιούνται δύο διαφορετικές μέθοδοι κατά την εκπαίδευση τους ώστε να υπάρχει μια πιο ολοκληρωμένη προσέγγιση έχοντας χρησιμοποιήσει και τις δύο διαθέσιμες τεχνικές.

Από το συνδυασμό αυτών των τριών προκύπτουν αλλά τέσσερα μοντέλα τα οποία ουσιαστικά έχουν αποτελέσματα από κάποια από τα προηγούμενα και αυτός ο συνδυασμός είναι ικανός να προσφέρει καλύτερο αποτέλεσμα αφού ένα στοιχείο που κάποιο από τα προηγούμενα δεν εντοπίζει μπορεί να εντοπίζεται από κάποιο άλλο μοντέλο. Όλα τα μοντέλα δοκιμάστηκαν σε 8 σύνολα δεδομένων σε κτίρια με βάση τη Λετονία ενώ τα 2 από τα 8 είχαν ως στόχο την πρόβλεψη της θερμικής ενέργειας που καταναλώθηκε ενώ τα υπόλοιπα της ηλεκτρικής. Ως μοντέλα σύγκρισης επιλέχθηκαν απλοί αλγόριθμοι μηχανικής μάθησης καθώς και μια παραλλαγή του αλγορίθμου των βαθμομερών θέρμανσης και ψύξης.

Λέξεις Κλειδιά: Ενεργειακή Αναβάθμιση, LSTM, XGBOOST, Πρόβλεψη ενεργειακής κατανάλωσης, Νευρωνικά δίκτυα, Μηχανική Μάθηση, Προβλέψεις Χρονοσειρών (Time series forecasting), Πρωτόκολλο IPMVP, Measurement and Verification (M&V).

Abstract

The problem that this Thesis is dealing with is the problem of calculating the accurate profit from the energy savings in a building after renovation actions. Particularly the main scope is the optimization of the forecasting of the energy consumption the building would have if there were never applied any renovation actions. If the consumptions in the period before the renovation are known as well as the external conditions after it, the necessary data are available so that, by application to suitable models, accurate and useful forecasts can be produced.

There are plenty of different machine learning models that are reviewed in this Thesis that are the aftermath of the combination of three simple models. The first of these three models are the XGBOOST, which is a gradient boosting algorithm, and it can produce accurate predictions even on its own. The other two are two LSTM models, which are neural networks with the ability to use a small capacity of memory cells and find some complex relationships between data. The reason that two LSTM are used is that there are two different ways of training that are applied so there is a wholesome result about their capabilities in every of those two ways of training.

From the combination of those three there are four other models which consist of results from some of the simple ones and that combination produces more accurate results because one of them may detect something that the other two have missed. All the models were tested on 8 datasets which consist of data from buildings based in Latvia and 2 of the 8 were aimed at predicting the thermal energy that were consumed and the rest of them the electrical energy that were consumed. Simple machine learning algorithms as well as a variant of the heating and cooling degree-day algorithm were chosen as comparison models.

Keywords: Energy Conservation Measures (ECM), Long Short-Term Memory (LSTM), Extreme Gradient Boosting (XGBOOST), Energy consumption forecasting, Neural Networks, Machine Learning, Time series forecasting, International Performance Measurement and Verification Protocol (IPMVP), Measurement and Verification (M&V).

Περιεχόμενα

Πρόλογος	5
Περίληψη	6
Abstract	8
1. Εισαγωγή	12
1.1 Πρόβλημα	12
1.2 Συμβολή Διπλωματικής	13
1.3 Δομή Διπλωματικής	14
2. Θεωρητικό Υπόβαθρο	15
2.1 Εισαγωγή και ενεργειακή αναβάθμιση	15
2.2 Πρωτόκολλο IPMVP	15
2.3 Measurement and Verification (M&V)	19
3. Βιβλιογραφική Ανασκόπηση	24
3.1 Επισκόπηση του paper: Image from Development and application of a machine learning supported methodology for measurement and verification (M&V) 2.0[8]	24
3.2 Επιπλέον βιβλιογραφία για την βελτιστοποίηση των προβλέψεων	27
4. Προτεινόμενη Λύση	32
4.1 Σύνολα Δεδομένων και Εξαγωγή και επιλογή χαρακτηριστικών	33
4.2 Αλγόριθμοι για σύγκριση και Κύρια Μοντέλα	34
4.3 Προετοιμασία εισόδου και Επιλογή παραμέτρων	39
5. Παρουσίαση Αποτελεσμάτων	42
5.1 Εισαγωγικά στοιχεία	42
5.2 Παρουσίαση Αποτελεσμάτων για το πρώτο σύνολο δεδομένων	45
5.3 Παρουσίαση Αποτελεσμάτων για το δεύτερο σύνολο δεδομένων	64
5.4 Παρουσίαση των αποτελεσμάτων των υπόλοιπων συνόλων δεδομένων	85
5.5 Αποτελέσματα των αλγορίθμων σύγκρισης σε όλα τα σύνολα δεδομένων	88
5.6 Συμπεράσματα για τα αποτελέσματα	89
6. Επίλογος και Μελλοντικές προσεγγίσεις	91
6.1 Επίλογος	91
6.2 Μελλοντικές προσεγγίσεις	92
Παράρτημα Α: Παρουσίαση πηγαίου κώδικα	93
Παράρτημα Β: Διαγράμματα από τα σύνολα δεδομένων του πίνακα 5.25	96
Βιβλιογραφία	110

1. Εισαγωγή

1.1 Πρόβλημα

Η ενεργειακή αναβάθμιση σε κτίρια είναι πάρα πολύ συνηθισμένη καθώς σήμερα τα έξοδα έχουν αυξηθεί εκθετικά και το κόστος ζωής γίνεται όλο και μεγαλύτερο. Ο πόλεμος καθώς και η διαρκής αύξηση των αναγκών σε ηλεκτρική ενέργεια καθιστούν απαγορευτική τη σπατάλη λόγω παρωχημένου εξοπλισμού. Ταυτόχρονα η κλιματική αλλαγή έχει αλλάξει αισθητά τις καιρικές συνθήκες με αποτέλεσμα να μην υπάρχουν οι ίδιες ανάγκες όπως στο παρελθόν. Έτσι κρίνεται επιτακτική ανάγκη να γίνουν οι κατάλληλες αλλαγές στον εξοπλισμό ώστε να εκσυγχρονιστούν τα κτίρια και να αποφεύγουν σπατάλες τόσο για το οικονομικό όσο και το περιβαλλοντικό όφελος.

Με τις ενεργειακές ανακαινίσεις αρκετά κτίρια προσπαθούν να εξοικονομήσουν ενέργεια ώστε να έχουν λιγότερα έξοδα και μικρότερη συμβολή στη ρύπανση. Με το να γίνουν πιο πράσινα βοηθούν ώστε να διατηρείται η φυσική ομορφιά και να μην χαλάει το κλίμα για τις επόμενες γενιές. Οι επιλογές είναι παρά πολλές και τα οφέλη το ίδιο. Οι τόσες διαθέσιμες επιλογές άρχισαν να δημιουργούν το ερώτημα στους ενδιαφερόμενους για το ποια αλλαγή θα είχε τη καλύτερη απόδοση και θα επέφερε το μεγαλύτερο κέρδος.

Μπορεί η μέτρηση της κατανάλωσης με τα νέα δεδομένα να είναι απλή διαδικασία όμως ο υπολογισμός της κατανάλωσης με τα στοιχεία πριν την αναβάθμιση είναι αρκετά πιο πολύπλοκος. Το πρόβλημα προς επίλυση είναι ο υπολογισμός του ρεαλιστικού κέρδους σε ενέργεια που έχει ένα κτίριο μετά τις αλλαγές ενεργειακής εξοικονόμησης καταφέροντας να προβλέψουμε με μεγάλη ακρίβεια την κατανάλωση που θα είχαμε αν δεν είχαν προηγηθεί αλλαγές.

Οι κυριότερες δυσκολίες της ίδιας της πρόβλεψης είναι οι πολλοί παράγοντες που επηρεάζουν την κατανάλωση ενός κτιρίου. Εκτός από τους προφανείς όπως ώρες λειτουργίας όταν μιλάμε για δημόσια ή εμπορικά κτίρια έως και μη προβλέψιμους όπως καιρικές συνθήκες και θερμοκρασία οι συνιστώσες είναι πολλές. Οι περισσότερες προσπάθειες που έχουν ήδη γίνει για την επίλυση του συγκεκριμένου προβλήματος το αντιμετωπίζουν σαν πρόβλημα παλινδρόμησης. Αρκετά λιγότερες είναι οι προτεινόμενες μεθοδολογίες λύσης που περιέχουν βαθιά μάθηση και νευρωνικά δίκτυα με τη χρήση των τελευταίων να περιορίζεται σε πολύ απλά μοντέλα.

Στη παρούσα διπλωματική θα γίνει εκτενής παρουσίαση αρκετών προηγούμενων ερευνών πριν παρουσιαστεί η δική μας προσέγγιση. Με αρκετές καινοτομίες έγινε προσπάθεια εύρεσης καλύτερων μοντέλων τόσο απλών όσο και συνδυαστικών για την επίτευξη μεγαλύτερης ακρίβειας. Τα κυριότερα σημεία διαφοροποίησης από προηγούμενες έρευνες παρουσιάζονται ακολούθως.

1.2 Συμβολή Διπλωματικής

Έχουν γίνει αρκετές προσπάθειες για να βρεθούν οι καλύτερες λύσεις για τον τρόπο με τον οποίο μπορεί να γίνει η πρόβλεψη με μεγαλύτερη ακρίβεια. Στη συγκεκριμένη διπλωματική έγιναν αρκετές προσπάθειες ώστε να διαφοροποιηθεί από τις ήδη υπάρχουσες έρευνες. Οι βασικότερες διαφορές είναι οι παρακάτω:

- Δοκιμάστηκαν τα μοντέλα σε πλήθος Dataset από τα οποία προέκυψαν σημαντικές πληροφορίες για τον τρόπο που αντιμετωπίζονται τα διάφορα αυτά σύνολα από τα μοντέλα που δοκιμάστηκαν. Στα αποτελέσματα είναι ξεκάθαρο ότι το κάθε σύνολο δεδομένων λειτουργεί διαφορετικά και μπορεί δυο σύνολα δεδομένων να μοιάζουν αλλά να μην εκπαιδεύουν το ίδιο καλά τα μοντέλα ενώ κάθε ένα από αυτά μπορεί να ταιριάζει καλύτερα με άλλο μοντέλο.
- Ελέγχθηκε και η βρέθηκαν οι καλύτεροι τρόποι με τους οποίους τα LSTM και το XGBOOST θα μπορούσαν να συνδυαστούν με τον συνδυασμό που ξεχώρισε να είναι αυτός του δεύτερου LSTM με το XGBOOST. Φαινομενικά όπως θα γίνει διακριτό και στη συνέχεια τα αποτελέσματα του XGBOOST είναι αρκετά συντηρητικά παρουσιάζοντας μικρή διακύμανση γύρω από ένα μέσο όρο προβλέψεων. Σε αντίθεση τα LSTM έχουν τη δυνατότητα να παρουσιάζουν μεγάλη διακύμανση και επομένως αποτελούν συμβατά κομμάτια ώστε να αποτελέσουν μέρος μιας ensemble λύσης με το XGBOOST.
- Ως προς τα LSTM που χρησιμοποιήσαμε δημιουργήσαμε δυο μοντέλα, με κοινό χαρακτηριστικό το διπλό επίπεδο LSTM δηλαδή δυο που είναι κολλημένα το ένα μετά το άλλο.
- Η προσπάθεια επεξεργασίας των εισερχομένων στο μοντέλο features από τα οποία εξήχθη μεγάλο πλήθος πληροφορίας προσπαθώντας να εντοπιστεί με ποια από αυτά είναι δυνατόν να επιτευχθούν καλύτερα αποτελέσματα. Μεταξύ αυτών που επιλέχθηκαν είναι συνδυασμοί από χαρακτηριστικά που τροποποιήθηκαν κατάλληλα πχ τριγωνομετρικοί μετασχηματισμοί ώστε να εφαρμόζονται μόνο τα στοιχεία που βοηθούν στην εκπαίδευση και να μην έχουν αρνητικά αποτελέσματα στα μοντέλα. Ενώ άλλα κρίθηκαν ότι δε μπορούν βοηθήσουν όπως πχ η ημέρα της εβδομάδας στην οποία έγινε η μέτρηση
- Χρησιμοποιήθηκε διαφορετικός συνδυασμός χαρακτηριστικών για την εκπαίδευση του XGBOOST και διαφορετικό για την εκπαίδευση του LSTM.

Τα αποτελέσματα αποτελούν διάφορους συνδυασμούς των παραπάνω αποδεικνύοντας ότι το κάθε σύνολο δεδομένων μπορεί να είναι πιο συμβατό με ένα εκ των τριών αλλά σχεδόν πάντα ο συνδυασμός τους φέρνει καλύτερα αποτελέσματα. Τα δεδομένα εισόδου καθορίζουν σε πολύ μεγάλο βαθμό το τελικό αποτέλεσμα αφού υπήρξαν κάποια σύνολα τα οποία κρίθηκαν ανάξια χρήσης καθώς τα δεδομένα τους δε μπορούσαν να προσφέρουν κάποια επιπλέον πληροφορία και η πρόβλεψη τους με ακρίβεια ήταν πρακτικά αδύνατη.

1.3 Δομή Διπλωματικής

Στη συνέχεια θα οριστεί η δομή της διπλωματικής καθώς και τα περιεχόμενα του καθενός.

Το κεφάλαιο 1 είναι μια μικρή εισαγωγή όπου αναφέρεται περιληπτικά το πρόβλημα και ο λόγος που εκπονήθηκε η παρούσα διπλωματική. Αναφέρονται τα σημεία διαφοροποίησης από άλλες προηγούμενες προσεγγίσεις του ίδιου προβλήματος περιληπτικά. Καταλήγει στην παρουσίαση της δομής της διπλωματικής.

Στο κεφάλαιο 2 εδραιώνεται και επεξηγείται το πρόβλημα καθώς και αναλύεται το θεωρητικό του υπόβαθρο. Αρχίζει με μια σύντομη επεξήγηση της έννοιας της ενεργειακής αναβάθμισης και των μέσων που μπορεί να επιτευχθεί. Στη συνέχεια σκιαγραφείται το πρωτόκολλο International Performance Measurement and Verification Protocol (IPMVP) και οι όροι που το απαρτίζουν και ακολουθεί η ανάλυση του Measurement and Verification (M&V) με σκοπό να καλυφθεί όλο το θεωρητικό κομμάτι γύρω από αυτή τη διαδικασία.

Στο κεφάλαιο 3 πραγματοποιείται μια βιβλιογραφική ανασκόπηση με έρευνες και εργασίες που είναι σχετικές. Μέσα από αυτές παρουσιάζονται προηγούμενες μεθοδολογίες και τα αποτελέσματά τους. Ταυτόχρονα εντοπίζονται πολύτιμα στοιχεία που βοήθησαν σημαντικά και στην εκπόνηση της παρούσας διπλωματικής

Στο κεφάλαιο 4 αναλύεται η μεθοδολογία και προσέγγιση της διπλωματικής. Επεξηγούνται τα μοντέλα καθώς και όσες επιλογές οδήγησαν στα καλύτερα δυνατά αποτελέσματα. Παράλληλα επισυνάπτονται διάφορες έννοιες απαραίτητες για την κατανόηση και την επιλογή συγκεκριμένων παραμέτρων.

Στο κεφάλαιο 5 παρουσιάζονται τα αποτελέσματα. Μαζί με γραφικές παραστάσεις και αποτελέσματα συναρτήσεων σφάλματος γίνεται σύγκριση μεταξύ των μοντέλων καθώς και διαφόρων συνδυασμών τους. Αναλυτικά περιγράφονται τα αποτελέσματα για κάποια από τα κτίρια που δοκιμάστηκαν και για διαφορά χρονικά διαστήματα στα οποία έγινε δειγματοληψία των δεδομένων, από τα ωριαία που είναι οι αρχικές μετρήσεις μέχρι και στις μηνιαίες.

Τέλος στο κεφάλαιο 6 διατυπώνονται τα συμπεράσματα της διπλωματικής και προτείνονται κάποιοι τρόποι με τους οποίους θα μπορούσε να επεκταθεί η εργασία στο μέλλον.

Μετά από αυτά ακολουθεί παράρτημα τόσο για τον κώδικα όσο και για τις γραφικές παραστάσεις για τα διάφορα σύνολα δεδομένων που δεν συμπεριλήφθηκαν στο κεφάλαιο με τα αποτελέσματα. Στο κομμάτι του κώδικα εκτός από τον ίδιο το πηγαίο κώδικα υπάρχουν και σχόλια τα οποία εξηγούν διαφορά σημεία του. Στο κομμάτι των γραφικών παραστάσεων παρουσιάζονται τα αποτελέσματα από το κύριο πλήθος των συνόλων δεδομένων καθώς και τα ίδια τα σύνολα δεδομένων παρουσιάζονται σε μορφή διαγράμματος σε σχέση με το χρόνο ώστε να είναι εμφανής η περιοδικότητα και αν αυτή υπάρχει.

2. Θεωρητικό Υπόβαθρο

2.1 Εισαγωγή και ενεργειακή αναβάθμιση

Στο κεφάλαιο αυτό θα αναλυθούν οι έννοιες του IPMVP και του “Measurement and Verification” (M&V) όπως αυτές ορίζονται από τον EVO(Efficiency Valuation Organization) και τα βασικά στοιχεία γύρω από αυτά. Πρώτα όμως θα οριστεί η ενεργειακή αναβάθμιση και κάποιους τρόπους με τους οποίους μπορεί να επιτευχθεί.

Ως ενεργειακή αναβάθμιση ορίζεται η διαδικασία στην οποία προβαίνει κανείς, ώστε να θωρακίσει ένα κτίριο έναντι των απωλειών που έχει και να μειώσει την ενέργεια που αυτό καταναλώνει, ελαχιστοποιώντας την αναγκαία δαπάνη που έχει μετρηθεί σε ένα χρονικό διάστημα για παράδειγμα ένα μήνα [1]. Υπάρχουν αρκετοί τρόποι ώστε να επιτευχθεί η ενεργειακή αναβάθμιση σε ένα κτίριο αλλά οι διάφοροι αυτοί τρόποι μπορούν να χωριστούν σε δύο ομάδες, με την πρώτη να έχει στόχο τη μείωση απωλειών και τη δεύτερη την αναβάθμιση θέρμανσης/ ψύξης του χώρου. Στην πρώτη ομάδα τοποθετούνται οι αλλαγές που αποσκοπούν στη βελτίωση της ικανότητας ενός κτιρίου απέναντι στις απώλειες με μερικά παραδείγματα να είναι η εξωτερική θερμομόνωση τοίχων, αντικατάσταση κουφωμάτων και η μόνωση ταράτσας μεταξύ άλλων. Για την αναβάθμιση του συστήματος θέρμανσης και ψύξης κάποιες λύσεις είναι η εγκατάσταση αντλίας θερμότητας, λέβητα φυσικού αερίου και η εγκατάσταση συστήματος ηλιακής θέρμανσης δηλαδή τεχνολογίες αισθητά νεότερες που υπόσχονται καλύτερα αποτελέσματα με λιγότερη κατανάλωση και απώλειες.

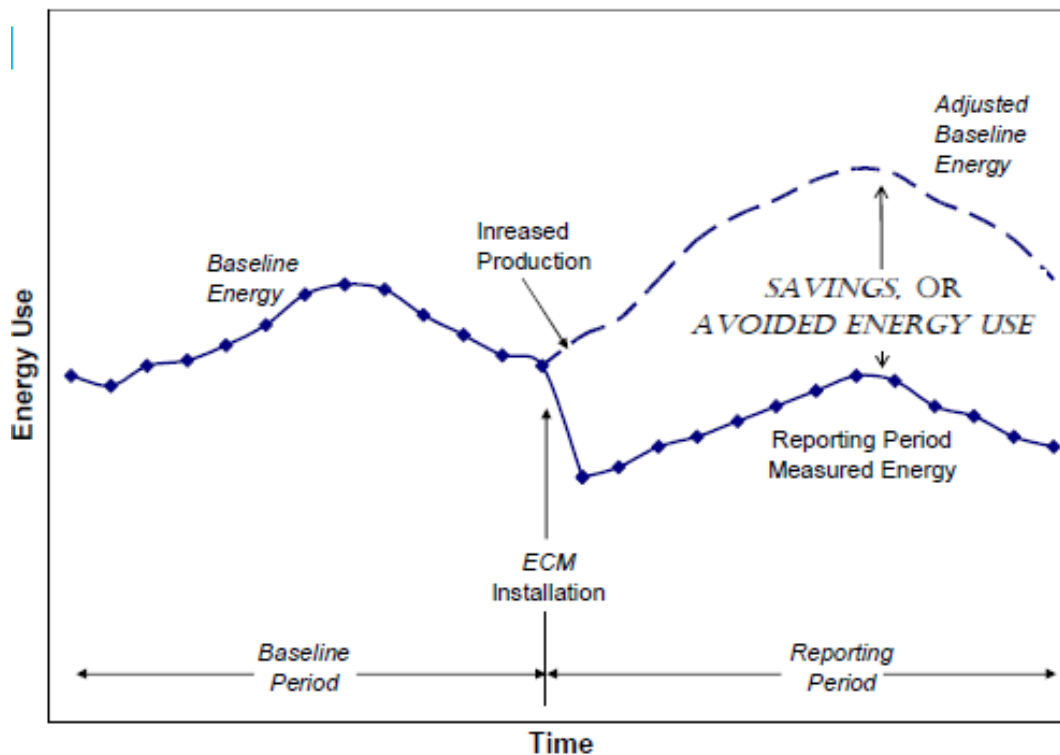
2.2 Πρωτόκολλο IPMVP

Δεν είναι δυνατόν να γίνει λόγος για “Measurement and Verification” χωρίς πρώτα να γίνει εκτενής ανάλυση του πρωτοκόλλου IPMVP [2],[3]. Το πρωτόκολλο αυτό δημιουργήθηκε ώστε να τεθούν κάποια πρότυπα και να αποκτήσει η διαδικασία M&V έναν σταθερό τύπο. Το εν λόγω πρωτόκολλο δημιουργήθηκε το 1997 από τον παγκόσμιο μη κερδοσκοπικό οργανισμό EVO, του οποίου η λειτουργία βασίζεται στη κατασκευή προϊόντων με στόχο την βελτίωση της ενεργειακής αποδοτικότητας και εξελίχθηκε σήμερα να γίνει το εξέχον πρωτόκολλο M&V που χρησιμοποιείται στα νέα έργα ανάπτυξης και αναβάθμισης σε όλο τον κόσμο. Το IPMVP προσφέρει τεχνικές για την επαλήθευση των επιπτώσεων της ενεργειακής αποδοτικότητας, της αποδοτικότητας του νερού, μέχρι και των έργων ανανεώσιμης ενέργειας. Μια συνηθισμένη χρήση του είναι να λειτουργεί συμπληρωματικά σε εθνικές και περιφερειακές κατευθυντήριες γραμμές ενεργειακής αποδοτικότητας όπως της Κατευθυντήριας Γραμμής 14 της ASHRAE[4]. Το πρωτόκολλο απαιτεί ορισμένες δραστηριότητες M&V να γίνονται σε κάθε έργο ώστε να υπάρχει συνέπεια μεταξύ των έργων και οι υπολογισμοί εξοικονόμησης ενέργειας να έχουν μεγάλη ακρίβεια, τόσο σε περιπτώσεις που αφορούν μια ολοκληρωμένη προσέγγιση της εγκατάστασης όσο και για μεμονωμένα μέτρα ενεργειακής αποδοτικότητας.

Το IPMVP διέπεται από συγκεκριμένες αρχές που είναι αποδέκτες από το M&V. Υπάρχουν διάφοροι τύποι χρηστών μεταξύ των οποίων είναι οι εργολάβοι που αναλαμβάνουν να εκπονήσουν έργα με στόχο τη βελτίωση της ενεργειακής αποδοτικότητας σε ένα κτίριο, όπως και εταιρείες παροχής ενεργειακών υπηρεσιών, σύμβουλοι ενέργειας που πραγματοποιούν ποσοτικοποίηση στην εξοικονόμηση ενέργειας και χρήστες ενέργειας οι οποίοι παρακολουθούν την εξοικονόμηση πόρων από τη δική τους αναβάθμιση.

Οι μεθοδολογίες M&V που περιλαμβάνονται στο IPMVP μπορούν να εξηγηθούν με την εξίσωση:

$$\text{Savings} = (\text{Baseline Period Energy} - \text{Reporting Period Energy}) +/- \text{Adjustments}$$



Εικόνα 2.1: Image from watchwire.ai [5]

Για τη μέτρηση της εξοικονόμησης τα δεδομένα οφείλουν να είναι καθορισμένα πριν την εφαρμογή μέτρων ενεργειακής αποδοτικότητας. Κατά τη βασική περίοδο δεν παρακολουθείται μόνο η ενεργειακή κατανάλωση αλλά και άλλοι παράγοντες όπως επίπεδα πληρότητας, απόδοση παραγωγής και άλλα. Οι μετρητές σε όλη την εγκατάσταση συλλέγουν δεδομένα που υποστηρίζονται από προγράμματα διασφάλισης ποιότητας και επαληθεύσεις τρίτων. Αφού έχει μετρηθεί η κατανάλωση ενέργειας τόσο στη βασική περίοδο όσο και στη περίοδο αναφοράς είναι δυνατόν με τη διαφορά τους να βρεθεί η εξοικονόμηση λαμβάνοντας ωστόσο υπόψιν κάποιες προσαρμογές που είναι αναγκαίες για το τελικό αποτέλεσμα. Οι προσαρμογές αυτές επιφέρουν εξισορρόπηση των διακυμάνσεων που συμβαίνουν κατά την περίοδο αναφοράς. Μερικές από αυτές είναι οι καιρικές συνθήκες, ο όγκος παραγωγής ή και οι ανακαινίσεις κτιρίων. Οι προσαρμογές όπως οι καιρικές συνθήκες μπορούν να υπολογιστούν με την κανονικοποίηση των δεδομένων χρήσης σε σχέση με τον καιρό ή αλλά άμεσα διαθέσιμα δεδομένα. Οι προσαρμογές που δεν είναι τακτικές όπως μια ανακαίνιση είναι περίπλοκες και δύσκολο να προβλεφθούν και να επεξεργαστούν όμως πρέπει να υπολογιστούν σε κάθε περίπτωση.

Υπάρχουν τέσσερις γενικές προσεγγίσεις μέσα στο IPMVP πρωτόκολλο οι οποίες θα παρουσιαστούν συνοπτικά παρακάτω. Αυτές μπορούν να προσαρμοστούν στις ανάγκες του κάθε έργου εξοικονόμησης ενέργειας. Κάθε μια από αυτές χρειάζεται δεδομένα ενέργειας και λειτουργικών δεδομένων για τα κτίρια που πραγματεύονται ώστε να προσδιορίσουν την εξοικονόμηση ενέργειας. Για την επιλογή εξετάζονται

διάφορες πτυχές μεταξύ των οποίων και η θέση του ορίου μέτρησης των energy efficient measures (EEM) και της διαθεσιμότητας ή και της αξιοπιστίας των μετρούμενων η εκτιμώμενων δεδομένων.

<p style="writing-mode: vertical-rl; transform: rotate(180deg);">Προσεγγίσεις σε μεμονωμένες αλλαγές</p>	<p>Προσέγγιση 1: "Retrofit Isolation with Key Parameter Measurement"</p>	<p>Προσέγγιση 2: "Retrofit Isolation with All Parameter Measurement"</p>
<p style="writing-mode: vertical-rl; transform: rotate(180deg);">Προσεγγίσεις συνολικά στο κτίριο</p>	<p>Προσέγγιση 3: "Whole Facility Measurement"</p>	<p>Προσέγγιση 4: "Calibrated Simulation Measurement"</p>

Εικόνα 2.2 : Οι τέσσερις προσεγγίσεις του IPMVP πρωτοκόλλου

Προσέγγιση 1: "Retrofit Isolation with Key Parameter Measurement."

Η συγκεκριμένη προσέγγιση αφορά την περίπτωση όπου τα έργα εξοικονόμησης λαμβάνονται ξεχωριστά υπόψιν από το υπόλοιπο κτίριο και είναι η καλύτερη προσέγγιση σε περίπτωση που γίνεται χρήση σχετικά απλών EEM στους οποίους οι συνθήκες της βασικής περιόδου όπως και αυτής μετά την εγκατάσταση εξαρτώνται από κάποιες βασικές παραμέτρους, όπως η ισχύς εισόδου φωτισμού ή η απόδοση του ψυκτικού εξοπλισμού. Απλές τέτοιου είδους εφαρμογές είναι μια αναβάθμιση φωτισμού, στην οποία είναι δυνατόν να παρακολουθηθεί η ισχύς που καταναλώνεται και να εκτιμηθούν οι ώρες λειτουργίας αλλά χωρίς να γίνεται προσμέτρηση των αλληλεπιδράσεων που το μεμονωμένο σύστημα έχει με άλλα συστήματα του κτιρίου. Η συγκεκριμένη προσέγγιση του πρωτοκόλλου έχει το μικρότερο κόστος αφού η χρήση της βασικής περιόδου και η χρήση μετά την εφαρμογή εκτιμώνται χρησιμοποιώντας περιορισμένη συλλογή δεδομένων μαζί με την μηχανική ανάλυση των μετρήσεων των κρίσιμων παραμέτρων του συστήματος. Υπάρχουν διάφοροι

τύποι μετρήσεων με τους πιο συνηθισμένους να είναι οι βραχυπρόθεσμες, οι περιοδικές και οι συνεχείς. Οι μετρήσεις αυτές λαμβάνονται τόσο σε επίπεδο EEM για τη γραμμή βάσης όσο και για τον EEM μετά την εγκατάσταση. Οι παράμετροι που έχουν ενδιαφέρον είναι όσες δεν έχουν ήδη επιλεγεί για τη μέτρηση πεδίου. Όπως αναφέρεται και στην επίσημη ιστοσελίδα του οργανισμού “Οι εκτιμήσεις μπορούν να βασίζονται σε ιστορικά δεδομένα, προδιαγραφές κατασκευαστή ή κρίση μηχανικής. Αναγκαία είναι η τεκμηρίωση της πηγής ή αιτιολόγηση της εκτιμώμενης παραμέτρου”.

Όταν γίνεται χρήση της συγκεκριμένης προσέγγισης είναι σημαντικό να επαληθεύεται η σωστή συνεχής λειτουργία των μέτρων που εφαρμόζονται. Όταν εφαρμόζεται σωστά εξασφαλίζει ότι οι βασικές συνθήκες έχουν καθοριστεί σωστά, επιβεβαιώνει ότι είναι εγκατεστημένος ο κατάλληλος εξοπλισμός / σύστημα και ότι υπάρχει η δυνατότητα να δημιουργήσουν την προβλεπόμενη εξοικονόμηση ενώ επαληθεύει ότι θα συνεχίσουν να αποφέρουν την εξοικονόμηση πόρων που προσδοκείται κατά τη διάρκεια ζωής του εξοπλισμού.

Προσέγγιση 2: “Retrofit Isolation with All Parameter Measurement”

Η προσέγγιση αυτή μοιάζει αρκετά με την πρώτη με τη διαφορά ότι λαμβάνονται υπόψη όλες οι σχετικές επιδόσεις και οι λειτουργικές παράμετροι για την παροχή μακροπρόθεσμων δεδομένων σχετικά με τη χρήση ενέργειας και την εξοικονόμηση των μέτρων. Παράδειγμα τέτοιας εφαρμογής είναι μια αναβάθμιση φωτισμού όπου καταγράφονται τόσο η ισχύς που καταναλώνεται όσο και οι ώρες λειτουργίας. Αυτή η προσέγγιση είναι ιδιαίτερα χρήσιμη όταν υπάρχουν ήδη μετρητές που καταγράφουν τη χρήση ενέργειας των υπό εξέταση υποσυστημάτων.

Η προσέγγιση 2 έχει σίγουρα μεγαλύτερο κόστος σε σχέση με την 1 όμως είναι πιο ακριβής αφού οι μετρήσεις που χρησιμοποιούνται είναι οι ίδιες πιο ακριβείς. Η εξοικονόμηση υπολογίζεται με μέτρηση πεδίου όλων των βασικών παραμέτρων απόδοσης που καθορίζουν τη χρήση ενέργειας του επηρεαζόμενου συστήματος EEM. Η συνεχής παρακολούθηση επενδύσεων μπορεί να βοηθήσει σε περαιτέρω βελτιστοποίηση των συστημάτων και του εξοπλισμού μελλοντικά. Σε περίπτωση αναμονής διακυμάνσεων είναι σημαντικό να μετρούνται διαρκώς οι παράγοντες. Η προσέγγιση 2 είναι καλή σε περιπτώσεις τόσο μεμονωμένων κτιρίων όσο και όταν τα επιχειρησιακά δεδομένα για τον εξοπλισμό είναι άμεσα διαθέσιμα μέσω συστημάτων ελέγχου.

Προσέγγιση 3: “ Whole Facility Measurement”

Με την προσέγγιση αυτή λαμβάνονται υπόψη στοιχεία από ολόκληρο το κτίριο που εξετάζει τα δεδομένα μέτρησης σε επίπεδο ολόκληρης της εγκατάστασης και κάθε μέτρου ξεχωριστά. Με τη μέθοδο αυτή εξετάζεται η εξολοκλήρου χρήση ενέργειας σε μια εγκατάσταση και η αποτελεσματικότητα όλων των μέτρων συλλογικά. Σε περίπτωση πολλών EEM είναι δύσκολη η μεμονωμένη μέτρηση της αποτελεσματικότητας τους αν υπάρχει αλληλεπίδραση μεταξύ τους. Η συγκεκριμένη μέθοδος υπολογίζει την εξοικονόμηση πόρων μετρώντας τη χρήση ενέργειας σε επίπεδο εγκατάστασης. Είναι συνηθισμένο να χρησιμοποιεί τεχνικές όπως πολυμεταβλητή παλινδρόμηση για να συνδυαστούν πληροφορίες που αφορούν τις αλλαγές σε επίπεδο εγκατάστασης στις λειτουργίες, την πληρότητα ή άλλες ανεξάρτητες μεταβλητές όπως η μεταβλητότητα του καιρού και της παραγωγής. Με τη χρήση αλγορίθμων παλινδρόμησης επιτυγχάνεται η εύρεση συσχετίσεων μεταξύ των προηγούμενων στοιχείων και άρα βελτιώνονται οι προβλέψεις. Τέλος σημαντικό είναι να προσδιορίζονται όλες οι ανεξάρτητες μεταβλητές που επηρεάζουν την κατανάλωση ενέργειας, η μορφή και το περιεχόμενο οποιωνδήποτε ξεχωριστών

μοντέλων απόδοσης όπως και η επίδειξη της στατιστικής εγκυρότητας του στα τελικά μοντέλα παλινδρόμησης.

Προσέγγιση 4: “Calibrated Simulation Measurement”

Πάλι έχουμε μια μέθοδο ολοκλήρου κτιρίου αλλά σε αυτή τη περίπτωση μπορεί να χρησιμοποιηθεί και για μεμονωμένα εξαρτήματα και είναι ιδιαίτερα χρήσιμη όταν υπάρχουν λίγα δεδομένα ή πολλές συνδέσεις μεταξύ των EEM. Χρησιμοποιεί λογισμικό για τη δημιουργία βαθμονομημένων μοντέλων προσομοίωσης που μπορούν να προσδιορίσουν την εξοικονόμηση ενέργειας είτε σε επίπεδο εξαρτήματος είτε σε επίπεδο κτιρίου. Σε σχέση με την είσοδο των μοντέλων όπως δεδομένα μόνιμης χρήσης ενέργειας κλπ. είναι δυνατόν να δημιουργήσουν μια καλύτερη βάση προβλέψεων ώστε να πετύχουν τελικά μεγαλύτερη ακρίβεια. Ακόμα είναι εφικτό να συμπεριληφθούν δεδομένα μετρήσεων από μεμονωμένα εξαρτήματα και συστήματα, βελτιώνοντας περαιτέρω αυτά τα μοντέλα. Αρκετά σημαντικό είναι να προσδιοριστούν επαρκώς οι διαδικασίες βαθμονόμησης μοντέλων για τη συχνότητα λήψης των δεδομένων τόσο για τη βασική περίοδο όσο και για την περίοδο αναφοράς και να καθοριστεί το πρόγραμμα της προσομοίωσης όσο και η πηγή των καιρικών δεδομένων.

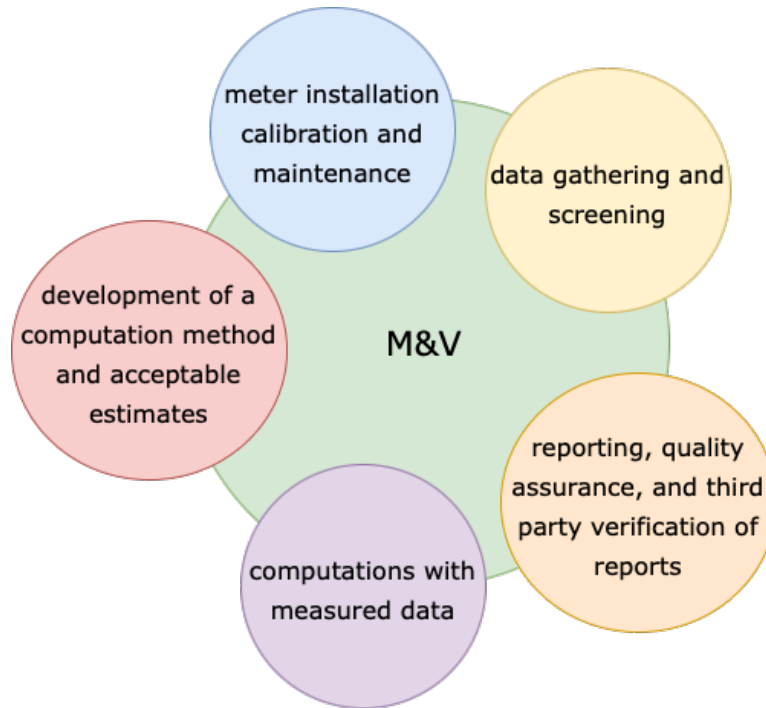
Σημασία του IPMVP

Σε κάθε έργο ενεργειακής αναβάθμισης είναι σημαντικό να παρακολουθείται η εξοικονόμηση ενέργειας και κόστους που είναι το κλειδί για την αναφορά και τη διασφάλιση επιτυχών αποτελεσμάτων. Όσο μεγαλύτερο είναι το έργο τόσο πιο αυστηρά πρέπει να τηρείται το IPMVP από το M&V. Η μέθοδος IPMVP επιτρέπει στο M&V εύκολη κλιμάκωση με βάση το μέγεθος και το εύρος των EEM που εφαρμόζονται. Όταν υπάρχει συμφωνία αποπληρωμής με ποσοστιαία καταβολή μέρους της πραγματικής εξοικονόμησης ενεργειακού κόστους τότε το IPMVP είναι ζωτικής σημασίας. Η εξοικονόμηση πρέπει να παρουσιάζεται με τρόπο διαφανή, να είναι εύκολα ελέγξιμη και αποδεκτή από τον πελάτη ή και ανεξάρτητους επιθεωρητές. Τα ισχυρά πρωτόκολλα M&V κερδίζουν την αποδοχή από το κοινό των έργων ενεργειακής απόδοσης ενώ ενθαρρύνουν τους επενδυτές και άλλους ενδιαφερόμενους φορείς να επιδιώξουν πρόσθετα έργα στο μέλλον.

2.3 Measurement and Verification (M&V)

Μετά από αυτή την εκτενή αναφορά στους όρους γύρω από το πρωτόκολλο IPMVP και τις πολλές αναφορές στο M&V κρίνεται πλέον απαραίτητο να εξηγηθεί εκτενώς τι ακριβώς είναι και ποιοι βασικοί όροι αναφέρονται συχνά μαζί του [6],[7]. Ως “Measurement and Verification” (M&V) ορίζεται η διαδικασία μέτρησης, συλλογής και ανάλυσης δεδομένων με σκοπό την επαλήθευση και την αναφορά της εξοικονόμησης ενέργειας σε μια μεμονωμένη εγκατάσταση που προκύπτει από την εφαρμογή μέτρων εξοικονόμησης ενέργειας Energy Conservation Measures (ECM). Η δυσκολία έγκειται στο γεγονός ότι μετά τις αλλαγές είναι αδύνατο να υπάρχει απευθείας πρόσβαση στην κατανάλωση με τις συνθήκες προ ενεργειακής βελτίωσης και αυτό γιατί το κάθε κτίριο έχει διαφορετικές ανάγκες όποτε και οι τιμές κατανάλωσης διαφέρουν αισθητά, την ώρα που ο προηγούμενος εξοπλισμός έχει ήδη αφαιρεθεί και οι καιρικές συνθήκες δεν είναι ποτέ σταθερές. Είναι απαραίτητο λοιπόν να βρεθεί τρόπος ώστε να υπολογιστεί κατά προσέγγιση με όσο το δυνατόν μικρότερο λάθος ποια θα ήταν η κατανάλωση αν δεν είχαν προηγηθεί αυτές οι αλλαγές. Στη συνέχεια θα εισαχθούν έννοιες γύρω από το M&V. Οι βασικές δραστηριότητες του αναφέρονται στη συνέχεια ενώ μπορεί να το απαρτίζουν μια ή και περισσότερες από αυτές ταυτόχρονα. Αυτές είναι:

A) Η βαθμονόμηση εγκατάστασης μετρητή που είναι η διαδικασία ελέγχου και μέτρησης της ακρίβειας μετρητή σε σύγκριση με ένα υπάρχον πρότυπο. Η συντήρηση είναι η τακτική δοκιμή, βαθμονόμηση, επισκευή ή αντικατάσταση μετρητή για τη διασφάλιση της ακρίβειας και της συμμόρφωσης με τα επίσημα πρότυπα μετρητών.



Εικόνα 2.3: Οι βασικές δραστηριότητες του M&V

B) Η συλλογή δεδομένων και ο έλεγχος τους που προκύπτουν από τις μετρήσεις κατανάλωσης και τη συλλογή καιρικών ή άλλων συνθηκών που την επηρεάζουν. Με τη συλλογή αυτήν και τον έλεγχο της βρίσκεται το απαραίτητο υλικό ώστε να είναι δυνατή τόσο η σύγκριση με απώτερο στόχο την εύρεση του κέρδους ενέργειας όσο και το απαραίτητο υλικό ώστε να είναι επιτυχημένη η πρόβλεψη για το τι θα συνέβαινε αν δεν είχαν προηγηθεί αλλαγές.

Γ) Η ανάπτυξη μιας μεθόδου υπολογισμού και αποδεκτών εκτιμήσεων

Δ) Οι υπολογισμοί με δεδομένα μέτρησης και η υποβολή εκθέσεων

Ε) Η διασφάλιση ποιότητας και η επαλήθευση εκθέσεων από τρίτους

Στόχοι του (M&V)

Πρωταρχικός στόχος του M&V είναι η αύξηση της ενεργειακής εξοικονόμησης. Όπως αναφέραμε υπάρχουν πολλοί τρόποι εξοικονόμησης αλλά το ποσό που τελικά θα εξοικονομηθεί δεν είναι εύκολο να προβλεφθεί. Ο ακριβής προσδιορισμός της εξοικονόμησης ενέργειας παρέχει στους ιδιοκτήτες και τους διαχειριστές εγκαταστάσεων πολύτιμες πληροφορίες σχετικά με τα ECM. Με αυτή την ανατροφοδότηση δίνεται η δυνατότητα στους ενδιαφερόμενους να προσαρμόσουν τη σχεδίαση ή και τις λειτουργίες των ECM για την επίτευξη μεγαλύτερης και

σταθερότερης εξοικονόμησης πόρων, με την πάροδο του χρόνου και με χαμηλότερες διακυμάνσεις.

Αρκετά συνηθισμένο είναι η εξοικονόμηση ενέργειας να αποτελεί τη βάση για πληρωμές βάσει απόδοσης ή και εγγύηση σε μια σύμβαση απόδοσης. Για να υπάρχει διαφάνεια στις συναλλαγές χρειάζεται να υπάρχει μια τεκμηριωμένη και κατανοητή απόδειξη της απόδοσης των ECM με βάση της οποίας θα διευθετηθούν οι αντίστοιχες συμφωνημένες πληρωμές. Ένα καλά καθορισμένο και εφαρμοσμένο σχέδιο M&V δύναται να αποτελέσει τη βάση αυτή για την τεκμηρίωση της απόδοσης με διαφανή τρόπο και να υποβληθεί σε επαλήθευση από τρίτους. Στο ίδιο μήκος κύματος ένα καλό σχέδιο M&V αυξάνει τη διαφάνεια και την αξιοπιστία των αναφορών σχετικά με το αποτέλεσμα των επενδύσεων αποδοτικότητας. Το χτίσιμο αυτό της αξιοπιστίας μεγιστοποιεί την εμπιστοσύνη που έχουν οι επενδυτές και οι χορηγοί στα έργα ενεργειακής απόδοσης, ενισχύοντας σημαντικά τις πιθανότητες τους να χρηματοδοτηθούν.

Πέρα από τα οικονομικά οφέλη και τη χρήση του M&V ως ασφάλεια σε συμφωνίες μπορεί να βοηθήσει σημαντικά στη βελτίωση του μηχανολογικού σχεδιασμού και τις λειτουργίες και συντήρηση των εγκαταστάσεων. Με ένα καλό σχέδιο ο κατασκευαστής μπορεί να γνωρίζει από πριν τα απαιτούμενα έξοδα για την αναβάθμιση με αποτέλεσμα να μπορεί να προγραμματίσει καλύτερα τους τρόπους πληρωμής. Επίσης βοηθά τους διαχειριστές να ανακαλύψουν και να μειώσουν τα προβλήματα συντήρησης και λειτουργίας, με αποτέλεσμα να μπορούν να λειτουργούν αποτελεσματικότερα τις εγκαταστάσεις ενώ ταυτόχρονα να μπορούν να χρησιμοποιηθούν και σαν σκαρίφημα για μελλοντικά σχέδια έργων. Ακόμα είναι δυνατόν να προβλέψει αποκλίσεις στον προϋπολογισμό και να ληφθούν υπόψη από την αρχή του έργου ώστε να μην προκύψουν προβλήματα κατά τη διεκπεραίωση του. Παράλληλα η χρήση ενός σχεδίου M&V για τον προσδιορισμό της εξοικονόμησης ενέργειας βελτιώνει τις αναφορές μείωσης εκπομπών σε σύγκριση με αναφορές χωρίς σχέδιο M&V και παρέχει πρόσθετη αξία στα έργα ενεργειακής αποδοτικότητας.

Εκτός από τα παραπάνω υπάρχουν σημαντικά πλεονεκτήματα στη χρήση του και από την πλευρά της κυβέρνησης και των προγραμμάτων κοινής ωφέλειας για να αξιολογήσουν την εξοικονόμηση σε επιλεγμένες εγκαταστάσεις χρηστών ενέργειας. Χρησιμοποιώντας στατιστικές τεχνικές και άλλες υποθέσεις, η εξοικονόμηση πόρων που προσδιορίζεται από δραστηριότητες M&V σε επιλεγμένες μεμονωμένες εγκαταστάσεις μπορεί να βοηθήσει στην πρόβλεψη εξοικονόμησης σε μη μετρημένες τοποθεσίες, προκειμένου να αναφερθεί η απόδοση ολόκληρου του προγράμματος. Παράλληλα βελτιώνοντας την αξιοπιστία των έργων ενεργειακής διαχείρισης, η M&V αυξάνει την αποδοχή του κοινού για τη σχετική μείωση των εκπομπών. Αυτή η δημόσια αποδοχή ενθαρρύνει τις επενδύσεις σε έργα βελτίωσης της ενεργειακής απόδοσης. Ενισχύοντας την εξοικονόμηση πόρων, οι καλές πρακτικές M&V προβάλλουν τα δημόσια οφέλη που παρέχει η καλή διαχείριση ενέργειας, όπως η βελτίωση της υγείας της κοινότητας, η μειωμένη ρύπανση του περιβάλλοντος και η αύξηση της απασχόλησης.

Αρχές M&V

Για τη τήρηση του IPMVP είναι απαραίτητο να εισάγουμε κάποιες αρχές οι οποίες πρέπει να λαμβάνονται υπόψη και να εφαρμόζονται σε όλη τη διαδικασία M&V. Πρώτη και βασικότερη είναι η ακρίβεια δηλαδή οι εκθέσεις να είναι όσο πιο ακριβείς μπορούν με βάση την αξία και τους στόχους του έργου. Τα κόστη οφείλουν να είναι αρκετά μικρότερα από το προσδοκώμενο κέρδος ενώ πρέπει να λαμβάνονται υπόψη τόσο το πόσο ακριβής είναι μια έκθεση όσο και το ίδιο το κόστος σαν μέρος της

ανάπτυξης του έργου. Η εξέταση όλων των εύλογων παραγόντων που επηρεάζουν την ακρίβεια είναι μια κατευθυντήρια αρχή του IPMVP. Ταυτόχρονα πρέπει να είναι ολοκληρωμένο έργο καθώς θα πρέπει να αξιολογούνται όλες οι επιπτώσεις ενός έργου. Οι δραστηριότητες M&V θα πρέπει να χρησιμοποιούν μετρήσεις για να ποσοτικοποιήσουν τη χρήση ενέργειας εντός του ορίου μέτρησης, να τεκμηριώσουν τους παράγοντες που επηρεάζουν την ενέργεια και να αναφέρουν λεπτομερώς τυχόν εκτιμώμενες τιμές. Εντοπίζοντας βασικούς τομείς στους οποίους απαιτείται κρίση, το IPMVP συμβάλλει στην αποφυγή ασυνεπειών που προκύπτουν από την έλλειψη εξέτασης σημαντικών πτυχών. Παράλληλα πρέπει να διατηρείται μια συντηρητική στάση και όταν γίνονται κρίσεις για αβέβαιες ποσότητες, οι διαδικασίες συγχωνεύσεων και εξαγορών θα πρέπει να σχεδιάζονται για να εκτιμούν εύλογα τις εξοικονομήσεις, έτσι ώστε να μην υπερεκτιμώνται ή να υποτιμώνται. Θα πρέπει να γίνει αξιολόγηση του αντίκτυπου ενός έργου για να διασφαλιστεί ότι τα οφέλη εξοικονόμησης ενέργειας είναι εύλογα και συντηρητικά, λαμβάνοντας δεόντως υπόψη το επίπεδο της στατιστικής εμπιστοσύνης στην εκτίμηση. Η αναφορά της ενεργειακής αποδοτικότητας ενός έργου πρέπει να είναι συνεπής και συγκρίσιμη σε διαφορετικούς τύπους έργων ενεργειακής αποδοτικότητας και επαγγελματίες διαχείρισης ενέργειας για κάθε έργο, διαφορετικές χρονικές περιόδους για το ίδιο έργο και έργα ενεργειακής αποδοτικότητας και νέα έργα ενεργειακού εφοδιασμού. Δε χρειάζεται να είναι ταυτόσημη καθώς αναγνωρίζεται ότι οποιαδήποτε εμπειρικά προερχόμενη αναφορά περιλαμβάνει υποθέσεις που βασίζονται σε ορθή κρίση μηχανικής, η οποία μπορεί να μην γίνεται με τον ίδιο τρόπο από όλους τους παρατηρητές .

Στις αρχές του M&V ανήκει και η σχετικότητα δηλαδή η ανάγκη να γίνεται ο προσδιορισμός της εξοικονόμησης με βάση τόσο τις παρούσες μετρήσεις όσο και πληροφορίες που αφορούν την εγκατάσταση που πραγματοποιείται το έργο. Αυτός ο προσδιορισμός της προσπάθειας εξοικονόμησης πρέπει να μετρήσει τους παράγοντες που επηρεάζουν την ενέργεια και να επαληθεύσει τους δείκτες απόδοσης που προκαλούν συζήτηση σχετικά με τα EEM. Τέλος η διαφάνεια είναι απαραίτητη αφού όλες οι δραστηριότητες M&V είναι απαραίτητο να είναι σαφείς και διαφανείς. Στην ολοκληρωμένη δήλωση θα πρέπει να περιλαμβάνονται η παρουσίαση όλων των στοιχείων ενός σχεδίου M&V και οι εκθέσεις εξοικονόμησης, μαζί με την επιβεβαίωση ότι το σχέδιο M&V είναι συμφωνημένο και κατανοητό από όλα τα ενδιαφερόμενα μέρη. Τα δεδομένα και οι πληροφορίες που συλλέγονται, οι τεχνικές προετοιμασίας δεδομένων, οι αλγόριθμοι, τα υπολογιστικά φύλλα, το λογισμικό, οι υποθέσεις που χρησιμοποιούνται και η ανάλυση θα πρέπει να ακολουθούν όσο το δυνατόν περισσότερο τις τυποποιημένες βέλτιστες πρακτικές του κλάδου, να είναι καλά μορφοποιημένα και τεκμηριωμένες – έτσι ώστε κάθε εμπλεκόμενο μέρος ή ανεξάρτητος επιθεωρητής να μπορεί να κατανοήσει πώς τα δεδομένα και η ανάλυση τους ακολουθεί το σχέδιο M&V και τις διαδικασίες αναφοράς αποταμιεύσεων. Η διαφάνεια σημαίνει επίσης ότι τυχόν πιθανές συγκρούσεις συμφερόντων γνωστοποιούνται σε όλα τα ενδιαφερόμενα μέρη του έργου.

Σε θεωρητικό επίπεδο με όλους αυτούς τους τρόπους μπορούν να επιτευχθούν τα θεμιτά αποτελέσματα και να υπολογιστεί το κέρδος της ενεργειακής αναβάθμισης. Το εύκολο κομμάτι είναι ο υπολογισμός της κατανάλωσης πριν και μετά τις αλλαγές του εξοπλισμού της εκάστοτε περιόδου. Με απλή μέτρηση λαμβάνονται τα ακριβή στοιχεία της κατανάλωσης πριν και μετά τις αλλαγές, για τα μεν πριν με το παλιό και ενεργοβόρο σύστημα ενώ μετά τις αλλαγές οι μετρήσεις θα αντικατοπτρίζουν την κατανάλωση με το νέο και βελτιωμένο σύστημα. Για να υπάρχει όμως περιθώριο σύγκρισης θα πρέπει να υπολογιστεί με κάποιο τρόπο η κατανάλωση που θα υπήρχε αν δεν είχαν προηγηθεί οι αλλαγές. Ερευνάται δηλαδή ο βέλτιστος τρόπος να προβλεφθούν οι τιμές κατανάλωσης των παλαιότερων συστημάτων στο κτίριο ώστε

με μια απλή αφαίρεση να βρεθεί ποια είναι η τιμή της εξοικονόμησης στο χρονικό διάστημα μετά την αναβάθμιση. Επομένως αναζητούνται τρόποι με τους οποίους μπορεί να δημιουργηθεί μια ακριβής προβολή της κατανάλωσης αν δεν είχαν προηγηθεί οι ΕΕΜ, με τις εξωτερικές συνθήκες όπως μετρήθηκαν για το ζητούμενο χρονικό διάστημα. Το δύσκολο κομμάτι σε τέτοιου είδους προβλέψεις είναι ότι δεν είναι δυνατόν να προβλεφθεί με σιγουριά η συμπεριφορά του κτιρίου στις εξωτερικές αλλαγές και οι εκάστοτε απώλειες ενώ δεν είναι λίγες οι φορές που η συμπεριφορά μπορεί να ξεφεύγει του συνηθισμένου και να οφείλεται σε απρόβλεπτους παράγοντες (πχ covid-19). Αλλά ακόμα και χωρίς τέτοιες αναπάντεχες δυσκολίες δεν είναι λίγα τα προβλήματα μιας τέτοιας πρόβλεψης. Από τα σχετικά απλά όπως ελλιπή δεδομένα και δεδομένα με αρκετό θόρυβο μέχρι περιπτώσεις όπου λείπουν δεδομένα για μεγάλα χρονικά διαστήματα αλλά και αλλαγές στην ίδια τη συμπεριφορά χωρίς προφανή λόγο που να προκύπτει από τα δεδομένα. Ακολουθεί μια σύντομη ανασκόπηση των ερευνών που έχουν προηγηθεί αυτής της διπλωματικής και οι λύσεις που προτείνουν στο γενικότερο πρόβλημα όσο και στις δυσκολίες που αντιμετώπισαν. Αρκετές από αυτές προσέφεραν σημαντικά στοιχεία και ιδέες για την εκπόνηση της παρούσας διπλωματικής και για αυτό το λόγο κρίνεται απαραίτητη η συμπερίληψη τους.

3. Βιβλιογραφική Ανασκόπηση

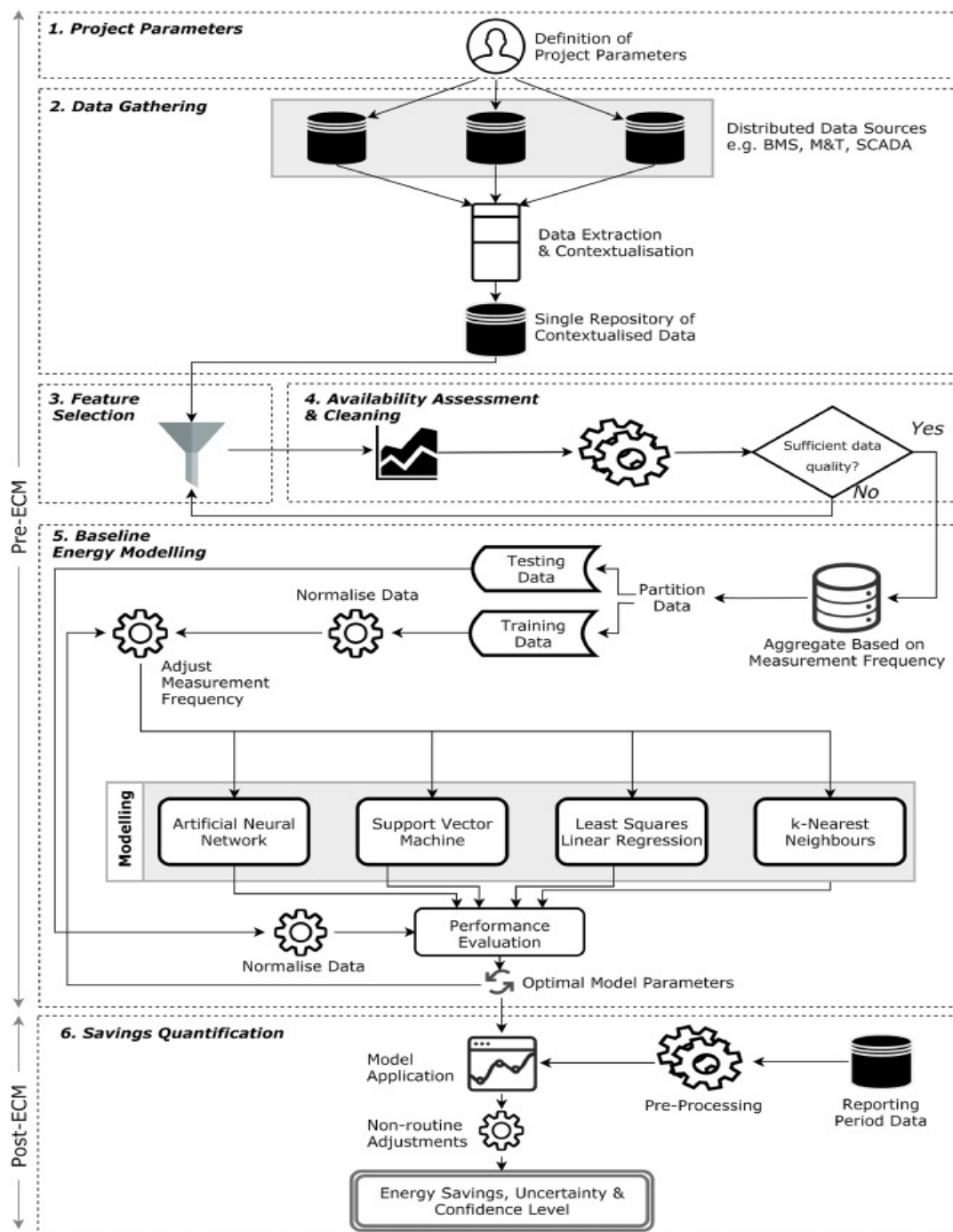
3.1 Επισκόπηση του paper: Image from Development and application of a machine learning supported methodology for measurement and verification (M&V) 2.0[8]

Στα πλαίσια του M&V και της πρόβλεψης της εξοικονόμησης ενέργειας και κατανάλωσης έχουν γίνει αρκετές έρευνες που προσπάθησαν να βελτιώσουν το κομμάτι της πρόβλεψης ώστε τα αποτελέσματα να είναι όσο το δυνατόν πιο κοντά στην πραγματικότητα. Έχουν γίνει αρκετές αξιολογικές προσπάθειες και κάποιες από αυτές θα συμπεριληφθούν στο παρόν κεφάλαιο. Σημαντικό είναι να τονιστεί ότι τα νούμερα και αποτελέσματα της κάθε μιας έχουν μεγάλη σχέση με τη μορφή και τα στοιχεία του συνόλου δεδομένων που χρησιμοποιήθηκαν στην εκάστοτε έρευνα.

Αρχικά οφείλουμε να ξεκινήσουμε από την ανασκόπηση της μεθοδολογίας που ακολουθείται και περιγράφεται στο paper των Gallagher, Leahy, O'Donovan, Bruton και O'Sullivan[4]. Πρώτο κομμάτι που είναι σημαντικό να γίνει λόγος είναι το κομμάτι της συλλογής των δεδομένων και του καθαρισμού τους. Τα δεδομένα αποτελούν σημαντικό μέρος της έρευνας και παίζουν κρίσιμο ρόλο στο τελικό αποτέλεσμα. Είναι λοιπόν απαραίτητο να υπάρχουν τα κατάλληλα δεδομένα ώστε τα αποτελέσματα να είναι ρεαλιστικά και να μην είναι επηρεασμένα σε βαθμό που να θεωρούνται άχρηστα ή προκατειλημμένα. Μερικοί παράγοντες είναι το μέγεθος του συνόλου δεδομένων, οι χαμένες τιμές ή οι περίοδοι στις οποίες δεν έγιναν για οποιοδήποτε λόγο μετρήσεις. Αυτές πρέπει με κάποιο τρόπο να αντιμετωπίζονται από τον εκάστοτε συγγραφέα.

Υπάρχουν διάφοροι τρόποι αντιμετώπισης του συγκεκριμένου προβλήματος αλλά υπάρχει μια ενδεικτική μεθοδολογία που ακολουθείται συνήθως. Αρχικά χρειάζεται έλεγχος και εξέταση των δεδομένων και της μορφής τους ως προς τον τύπο και το ποσοστό των δεδομένων που λείπουν. Τα δεδομένα στο συγκεκριμένο paper είναι από βιομηχανικά κτίρια. Ανάλογα με τον τύπο τους υπάρχουν και διαφορετικοί τρόποι ώστε να επεξεργαστούν και να συμπληρωθούν κατάλληλα. Επίσης ανάλογα με το ποσοστό των δεδομένων που λείπουν εξαρτάται ο τρόπος με τον οποίο θα καλυφθούν οι τιμές αυτές. Αν το ποσοστό αυτό είναι μικρό είναι δυνατόν χωρίς βλάβη να αφαιρεθούν αυτές οι σειρές δεδομένων από το σύνολο χωρίς πρόβλημα ενώ αν είναι μεγάλο πρέπει να βρεθεί τρόπος ώστε να συμπληρωθούν ομαλά σε σχέση με τις υπόλοιπες σειρές ώστε να μην αλλοιωθεί το σύνολο δεδομένων. Μόλις ολοκληρωθεί η διαδικασία αυτή πρέπει να βρεθούν περαιτέρω στοιχεία που μπορούν να εξαχθούν από τα δεδομένα. Για παράδειγμα μια ημερομηνία εκτός από τα εμφανή στοιχεία που δίνει όπως η μέρα, ο μήνας κλπ. δίνει και άλλες πληροφορίες όπως η εποχή και η εβδομάδα του χρόνου. Μετά από αυτή τη διαδικασία το σύνολο δεδομένων είναι έτοιμο για χρήση στο μοντέλο.

Στο παραπάνω paper μετά από τον καθαρισμό των δεδομένων προχωράει στην αξιολόγηση διαφόρων αλγορίθμων μηχανικής μάθησης καθώς και κάποια απλά νευρωνικά δίκτυα. Ακολουθεί εικόνα που δείχνει ακριβώς τη συλλογιστική πορεία και τα βήματα του συγκεκριμένου paper.



Εικόνα 3.1: Image from Development and application of a machine learning supported methodology for measurement and verification (M&V) 2.0[8]

Πρώτα εξετάζει μερικούς απλούς αλγόριθμους μηχανικής μάθησης και συγκεκριμένα τη Διμεταβλητή Γραμμική Παλινδρόμηση (Bi-variable Linear Regression), τη Πολυμεταβλητή Γραμμική Παλινδρόμηση (Multi-variable Linear Regression), τον κ-Κοντινότεροι Γείτονες (k-Nearest Neighbours), το Support Vector Machines (SVM) και το δέντρο αποφάσεων (decision tree). Θα ακολουθήσει σύντομη περιγραφή των παραπάνω αλγορίθμων.

Διμεταβλητή Γραμμική Παλινδρόμηση[9]

Είναι μια γραμμική εξίσωση που περιγράφει τη σχέση μεταξύ μιας μεταβλητής που έχει το ρόλο του ρυθμιστή του αποτελέσματος και μιας μεταβλητής αποτελέσματος, με την τελευταία να μην επηρεάζει τη πρώτη. Όταν υπάρχει μια συνάρτηση της μορφής

$$y_i = b_0 + b_1 x_i + e_i$$

Με το b_1 να είναι η εκτίμηση της κλίσης της γραμμής παλινδρόμησης σε σχέση με το χρόνο της εκπαίδευσης και b_0 η εκτίμηση για την κατακόρυφη τομή της γραμμής παλινδρόμησης. Ο όρος e_i είναι ο υπολειπόμενος όρος ή ο όρος σφάλματος στην παλινδρόμηση. Για την πρόβλεψη τιμής δημιουργείται μια σχέση της μορφής

$$y_i = b_0 + b_1 x_i$$

Που υπολογίζει τους νέους όρους $b_0 + b_1$ με βάση ποιοι προσαρμόζονται καλύτερα στη γραμμή παλινδρόμησης οι οποίοι επιλέγονται έτσι ώστε να ελαχιστοποιείται το άθροισμα του τετραγώνου των διαφορών.

Πολυμεταβλητή Γραμμική Παλινδρόμηση

Είναι αντίστοιχη με τη διμεταβλητή αλλά σε αυτή τη περίπτωση για τον υπολογισμό των προβλέψεων στην παραπάνω εξίσωση προσθέτουμε όσες οι μεταβλητές τόσους όρους της μορφής $b_i x_i$ με τους όρους b να υπολογίζονται εκ νέου από μια συνάρτηση κόστους/σφάλματος όπως και παραπάνω.

κ-Κοντινότεροι Γείτονες[10]

Ο αλγόριθμος αυτός είναι ένας αλγόριθμος ταξινόμησης εποπτευόμενης μάθησης με παραμέτρους, ο οποίος χρησιμοποιεί την απόσταση μεταξύ των δειγμάτων για να κάνει ταξινομήσεις ή προβλέψεις σχετικά με την ομαδοποίηση ενός μεμονωμένου σημείου δεδομένων. Στην περίπτωση της παλινδρόμησης που εξετάζεται επιλέγονται k γείτονες ως κοντινοί με βάση την Ευκλείδεια απόσταση ή κάποιο αντίστοιχο τρόπο και μετά λαμβάνει τον μέσο όρο των τιμών των γειτόνων ώστε να υπολογίσει την τιμή που ψάχνει.

Support Vector Machines(SVM)[11]

Ο στόχος του αλγόριθμου των μηχανών διανυσμάτων υποστήριξης είναι να βρεθεί ένα υπερεπίπεδο σε ένα χώρο N -διάστασης (N — ο αριθμός των χαρακτηριστικών) που ταξινομεί ευδιάκριτα τα σημεία δεδομένων. Τα διανύσματα υποστήριξης είναι σημεία δεδομένων που βρίσκονται πιο κοντά στο υπερεπίπεδο και επηρεάζουν τη θέση και τον προσανατολισμό του υπερεπιπέδου. Χρησιμοποιώντας αυτά τα διανύσματα υποστήριξης, μεγιστοποιούμε το περιθώριο του ταξινομητή. Η διαγραφή των διανυσμάτων υποστήριξης θα αλλάξει τη θέση του υπερεπιπέδου. Αυτά είναι τα σημεία που βοηθούν να δημιουργηθεί το SVM. Ξανά χρησιμοποιούνται συναρτήσεις κόστους/σφάλματος τις οποίες προσπαθούμε να ελαχιστοποιήσουμε ώστε να πετύχουμε την καλύτερη πρόβλεψη.

Δέντρο αποφάσεων[12]

Ο συγκεκριμένος αλγόριθμος είναι ένας αλγόριθμος εποπτευόμενης μάθησης χωρίς παραμέτρους, ο οποίος χρησιμοποιείται τόσο για εργασίες ταξινόμησης όσο και για εργασίες παλινδρόμησης. Έχει τη δομή ενός δέντρου, η οποία αποτελείται από έναν κόμβο ρίζας, κλάδους, εσωτερικούς κόμβους και κόμβους φύλλων. Ένα δέντρο απόφασης ξεκινά με μια ρίζα, η οποία δεν έχει εισερχόμενους κλάδους. Στη συνέχεια, οι εξερχόμενοι κλάδοι από τη ρίζα τροφοδοτούνται στους εσωτερικούς κόμβους, γνωστούς και ως κόμβους απόφασης. Με βάση τα διαθέσιμα χαρακτηριστικά, και οι δύο τύποι κόμβων πραγματοποιούν αξιολογήσεις για να σχηματίσουν ομοιογενή

υποσύνολα, τα οποία υποδηλώνονται με κόμβους φύλλων ή τερματικούς κόμβους. Οι κόμβοι φύλλων αντιπροσωπεύουν όλα τα πιθανά αποτελέσματα μέσα στο σύνολο δεδομένων.

Στη συνέχεια του paper χρησιμοποιείται και ένα απλό νευρωνικό δίκτυο με ένα κρυφό επίπεδο τροφοδοσίας προς τα μπροστά δηλαδή ένα νευρωνικό δίκτυο του οποίου οι συνδέσεις μεταξύ των κόμβων δεν δημιουργούν κύκλο. Μετά από αυτό εισάγουν τις συναρτήσεις σφάλματος CV(RMSE) και NMSE ώστε να αξιολογήσουν τα παραπάνω μοντέλα. Τα αποτελέσματα ως προς το πιο είναι καλύτερο διαφέρουν σημαντικά όταν αλλάζει η περίοδος δειγματοληψίας. Συγκεκριμένα σε μηνιαία βάση το νευρωνικό δίκτυο είναι το καλύτερο ενώ σε ημερήσια βάση είναι το 4 καλύτερο πράγμα που δείχνει ότι η απόδοση τους έχει μεγάλη σχέση με το πλήθος των δειγμάτων. Καταλήγει ότι το πραγματικό πλεονέκτημα της μεθόδου που προτείνεται είναι η περίπτωση όπου είναι αδύνατη η M&V χωρίς την εγκατάσταση πρόσθετου εξοπλισμού μέτρησης και μια περίοδο συλλογής δεδομένων που θα καθυστερούσε την εφαρμογή των ECM. Συνολικά 20 μοντέλα αναπτύχθηκαν κατά την περίοδο αναφοράς χρησιμοποιώντας μια εξαντλητική προσέγγιση. Το βέλτιστο μοντέλο αναγνωρίστηκε ως μοντέλο παλινδρόμησης k-NN που εκπαιδεύτηκε με δεδομένα μετρούμενα ανά ώρα.

3.2 Επιπλέον βιβλιογραφία για την βελτιστοποίηση των προβλέψεων

Οι περισσότερες έρευνες γύρω από το θέμα επικεντρώνονται στο να βρουν πιο αποτελεσματικούς τρόπους ώστε να πετύχουν ακριβείς προβλέψεις με στοιχεία της βασικής περιόδου. Αρχικά οι Granderson et al.[13] χρησιμοποίησαν πληθώρα αλγορίθμων σε εμπορικά κτίρια. Πιο αναλυτικά οι αλγόριθμοι χρησιμοποιήθηκαν σε κτίρια από διάφορες πολιτείες της Αμερικής κυρίως από τις California και Washington, DC. ενώ υπήρξαν και κτίρια από άλλες περιοχές με διαφορετικά κλιματικά χαρακτηριστικά. Όσον αφορά τα μοντέλα που δοκιμάστηκαν είναι 10 και κάποια από αυτά είναι τυχαία δέντρα μαζί με εβδομαδιαία δεδομένα, προηγμένες τεχνικές παλινδρόμησης, προβλέψεις μόνο με βάση τη μέρα και την ώρα, προβλέψεις με την ώρα και τη θερμοκρασία, συνδυασμός κοντινότερων γειτόνων και ενός απλού γραμμικού μοντέλου και προηγμένη παλινδρόμηση κοντινότερων γειτόνων. Στα αποτελέσματα βρέθηκε ότι κάποια κτίρια είναι αδύνατον να προβλεφθούν ενώ σε αυτά που η πρόβλεψη ήταν εφικτή καλύτερα αποτελέσματα παρουσίαζαν τα τυχαία δέντρα σε συνδυασμό με τα εβδομαδιαία δεδομένα και το μοντέλο με τη προηγμένη παλινδρόμηση.

Οι Dong et al.[14] ασχολήθηκαν με την εφαρμογή SVM σε εμπορικά κτίρια σε τροπικές περιοχές. Τα κτίρια είναι γραφεία στη Σιγκαπούρη που επιλέχθηκαν τυχαία μεταξύ άλλων με τα τετραγωνικά μέτρα και την κατανάλωση τους να διαφέρει αντίστοιχα. Ο λόγος που επιλέγεται ο συγκεκριμένος αλγόριθμος είναι ότι αποφεύγονται τα χαρακτηριστικά σε μεγαλύτερα αριθμητικά εύρη που κυριαρχούν σε αυτά σε μικρότερα αριθμητικά εύρη και η αποφυγή αριθμητικών δυσκολιών κατά τον υπολογισμό. Ο πυρήνας που επιλέχθηκε, ο RBF, αντιστοιχίζει μη γραμμικά δείγματα σε χώρο υψηλότερων διαστάσεων, και σε αντίθεση με τον γραμμικό πυρήνα, μπορεί να χειριστεί την περίπτωση όταν η σχέση μεταξύ των ετικετών κλάσεων και των χαρακτηριστικών είναι μη γραμμική ενώ ταυτόχρονα έχει λιγότερες υπέρ-παραμέτρους από τους πολυωνυμικούς πυρήνες. Τα αποτελέσματα έδειξαν μεγάλη ακρίβεια βέβαια με συγκεκριμένους παράγοντες όπως σχετικά μικρό σύνολο δεδομένων, λίγες ελεύθερες παραμέτρους και χρησιμοποιήθηκε η αρχή ελαχιστοποίησης δόκιμου κινδύνου με την υλοποίηση να ελαχιστοποιεί το άνω όριο του σφάλματος γενίκευσης και όχι το σφάλμα εκπαίδευσης. Την ώρα που τα NN δεν μπορούν να αποτελούν γενικευμένη λύση, το SVM έχει μοναδική και βέλτιστη λύση

γιατί είναι σαν να λύνεις ένα πρόβλημα γραμμικά περιορισμένου τετραγωνικού προγραμματισμού.

Οι Fan Tang, Andrew Kusiak και Xiupeng Wei [15] στο δικό τους paper εξερεύνησαν την πρόβλεψη σε μικρή χρονική περίοδο με τη χρήση αλγορίθμου ομαδοποίησης. Αρχικά επέλεξαν να ακολουθήσουν μια διαδικασία επιλογής των καλύτερων χαρακτηριστικών ώστε να έχουν μεγαλύτερη ακρίβεια, στη συνέχεια ομαδοποίησαν τα δεδομένα εισόδου και εφάρμοσαν ένα πολυεπίπεδο perceptron εκπαιδεύοντας το σε κάθε ομάδα. Ξανά το σύνολο δεδομένων είναι από εμπορικά κτίρια με δεδομένα που μάζεψε το πανεπιστήμιο της Αϊόβα. Το πολυεπίπεδο perceptron είναι ένα νευρωνικό δίκτυο όπου η αντιστοίχιση μεταξύ εισόδων και εξόδων είναι μη γραμμική. Ένα Multilayer Perceptron έχει στρώματα εισόδου και εξόδου και ένα ή περισσότερα κρυφά στρώματα με πολλούς νευρώνες στοιβαγμένους μεταξύ τους. Και ενώ στο Perceptron ο νευρώνας πρέπει να έχει μια συνάρτηση ενεργοποίησης που επιβάλλει ένα όριο, όπως το ReLU ή το σιγμοειδές, οι νευρώνες σε ένα Perceptron πολλαπλών επιπέδων μπορούν να χρησιμοποιήσουν οποιαδήποτε αυθαίρετη συνάρτηση ενεργοποίησης. Τα αποτελέσματα από αυτή τη χρήση του συγκεκριμένου μοντέλου έδειξαν ότι το συνδυαστικό μοντέλο MLP λειτουργεί καλύτερα από άλλους παραδοσιακούς αλγόριθμους όπως δέντρα ή SVM. Παράλληλα με την αρχική ομαδοποίηση η πρόβλεψη έγινε ακόμα πιο ακριβής ενώ η διαδικασία επιταχύνθηκε.

Οι B. Yildiz, J.I. Bilbao και A.B. Sproul [16] ασχολήθηκαν με την αξιολόγηση διαφόρων μοντέλων παλινδρόμησης και μηχανικής μάθησης για την ηλεκτρική κατανάλωση εμπορικών κτιρίων. Συγκεκριμένα τα δεδομένα είναι από κτίρια που βρίσκονται σε πανεπιστήμια. Τα μοντέλα αυτόματης παλινδρόμησης αναλύουν σύνολα σημείων δεδομένων σε μια χρονολογική σειρά και συσχετίζουν τη μελλοντική τιμή μιας συγκεκριμένης μεταβλητής με τις προηγούμενες τιμές της. Είναι δυνατό να συσχετιστούν διαφορετικές μεταβλητές εισόδου με την έξοδο έτσι ώστε στην πρόβλεψη φορτίου εμπορικών κτιρίων, το φορτίο μπορεί να είναι συσχετισμένο με άλλες σημαντικές παραμέτρους για τον καιρό και το κτίριο. Άλλο ένα μοντέλο που αναφέρεται είναι το ARIMA που είναι ο συνδυασμός της αυτόματης παλινδρόμησης με τον κινούμενο μέσο. Στη συνέχεια πάνω σε αυτά τα μοντέλα εφαρμόζει δεδομένα ιστορικά, προσομοιωμένα και για την πρόβλεψη φορτίου σε συστήματα θέρμανσης και ψύξης όπως και σε M&V. Τα περισσότερα από τα μοντέλα μηχανικής μάθησης που χρησιμοποιήθηκαν σε αυτή τη μελέτη έδειξαν καλύτερη απόδοση πρόβλεψης από τα μοντέλα MLR, όμως τα δεύτερα έδιναν τη δυνατότητα στο χρήστη να έχει μεγαλύτερη επαφή και έλεγχο ως προς την ανάλυση της πρόβλεψης πράγμα που είναι ένα συγκριτικό πλεονέκτημα. Ένα άλλο πλεονέκτημα της ανάλυσης παλινδρόμησης έναντι της μηχανικής μάθησης είναι η απλή εφαρμογή και χρήσης της πρώτης όμως όταν είναι σημαντική η ακρίβεια προτιμώνται τα δεύτερα.

Σημαντική βοήθεια για την εκπόνηση της διπλωματικής ήταν η έρευνα με όνομα "FORECASTING THE HOURLY ENERGY CONSUMPTION USING XGBOOST" [17]. Τα δεδομένα που χρησιμοποιήθηκαν αποτελούνται από στοιχεία για την ωριαία κατανάλωση ισχύος που κυμαίνονται από το 2002-2018 για ολόκληρη την περιοχή της ανατολικής ΗΠΑ. Πρώτα έγινε η συγκομιδή των δεδομένων, μετά τα επεξεργάστηκαν καταλλήλως ώστε να κερδίσουν ότι περισσότερο με τις συσχετίσεις των δεδομένων. Χρησιμοποίησαν κατά την εξαγωγή των χαρακτηριστικών πολλές έμμεσες πληροφορίες όπως την εβδομάδα του χρόνου στην οποία ανήκε μια μέρα και την εποχή. Στη συνέχεια εφαρμόζει το μοντέλο παλινδρόμησης του XGBOOST σε ένα μέρος του συνόλου δεδομένων και το κρίνει με βάση τη ρίζα του μέσου τετραγωνικού σφάλματος. Στη συνέχεια εκπαιδεύει το μοντέλο του σε όλο το σύνολο δεδομένων και πραγματοποιεί προβλέψεις για το μέλλον με βάση αυτό. Τα αποτελέσματα αυτού το

μοντέλου παλινδρόμησης ήταν πολύ καλά στο συγκεκριμένο πρόβλημα και στη συγκεκριμένη μετρική συνάρτηση.

Οι Hikmet Esen, Mustafa Inalli, Abdulkadir Sengur και Mehmet Esen [18] προσπάθησαν να προβλέψουν την απόδοση αντλίας θερμότητας με γείωση, με τη χρήση νευρωνικού δίκτυα με προ επεξεργασία στάθμισης στατιστικών δεδομένων. Οι αντλίες θερμότητας έχει αποδειχθεί ότι είναι μια εναλλακτική μέθοδος θέρμανσης και ψύξης πιο φιλική στο περιβάλλον που μπορεί να αντικαταστήσει τις παραδοσιακές μεθόδους. Κύριο πλεονέκτημα της είναι ότι οι υπόγειες πηγές θερμότητας έχουν μικρότερες εναλλαγές σε σχέση με τον εξωτερικό άρα και άρα θετικό θερμοδυναμικό πρόσημο. Τα δεδομένα που χρησιμοποιήθηκαν ήταν από ένα μοντέλο με δομή έναν οριζόντιο μετατροπέα θερμότητας εδάφους, εξοπλισμό μονάδας αντλίας θερμότητας και βοηθητικό εξοπλισμό. Το νευρωνικό δίκτυο που επιλέγεται είναι ένα με αλγόριθμο αντίστροφης διάδοσης που μαθαίνει αλλάζοντας τα βάρη σύνδεσης με τι αλλαγές αυτές να αποθηκεύονται ως γνώση. Ο στόχος της έρευνας είναι να απαντηθεί το ερώτημα αν γίνεται να βελτιωθεί σημαντικά το αποτέλεσμα πρόβλεψης ενός νευρωνικού δικτύου με προ επεξεργασία στάθμισης στατιστικών δεδομένων και τελικά αποδεικνύεται ότι όντως έχουν αρκετή βελτίωση. Οι τιμές που προβλέπονται με το SWP-ANN, ειδικά με τον αλγόριθμο εκμάθησης πίσω διάδοσης μαζί με την προώθηση, μπορεί να προβλέψει με μεγάλη ακρίβεια την απόδοση του συστήματος GCHP. Έτσι είναι δυνατόν να επιτευχθούν λύσεις με μεγαλύτερη ακρίβεια απλούστερα και ταχύτερα. Οι ίδιοι δοκίμασαν επίσης την εφαρμογή SVM ξανά με στόχο την πρόβλεψη της απόδοσης μιας αντλίας θερμότητας. Στο πείραμα αυτό απέδειξαν ότι για το συγκεκριμένο σύνολο δεδομένων ένας SVM έχει καλύτερα αποτελέσματα σε σχέση με τα νευρωνικά δίκτυα που χρησιμοποιήθηκαν για τη σύγκριση αποδεικνύοντας έτσι τη δυνατότητα του πρώτου για καλύτερη γενίκευση. Το μεγαλύτερο πλεονέκτημα στη συγκεκριμένη περίπτωση είναι ότι ο SVM χρησιμοποιεί την αρχή της ελαχιστοποίησης δομικού κινδύνου η οποία διαφέρει από τις κλασικές εμπειρικές αρχές ελαχιστοποίησης κινδύνου με την σύγκλιση σε ένα συνολικό μέγιστο και όχι σε διαφορά τοπικά. Οι λιγότερες ελεύθεροι παράμετροι το κάνουν έναν πιο γρήγορο τρόπο πρόβλεψης και καταλήγουν ότι γι' αυτό είναι μια καλή μέθοδος πρόβλεψης της απόδοσης μιας αντλίας θερμότητας.

Οι Harsha Moraliyage, Sanoshi Dahanayake, Daswin De Silva, Nishan Mills, Prabod Rathnayaka, Su Nguyen, Damminda Alahakoon and Andrew Jennings [20] στο paper τους πρότειναν μια ενδεικτική μεθοδολογία τεχνητής νοημοσύνης για την εφαρμογή M&V. Σε αυτή διέκριναν πέντε στάδια εκ των οποίων πρώτο ήταν τα μέσα που παρείχαν τα διαφορά δεδομένα π.χ. μετρητές κλπ., δεύτερο το σύνολο όλων των δεδομένων και ακολουθούσαν αυτά της τεχνητής νοημοσύνης, της ποσοτικοποίησης των ECM και των αποτελεσμάτων. Στο στάδιο της τεχνητής νοημοσύνης χρησιμοποιήθηκε το XGBoost. Το Boosting είναι μια καθιερωμένη έννοια μηχανικής μάθησης με αλγόριθμους, όπως προσαρμοστική ενίσχυση (AdaBoost), δέντρο ενίσχυσης, ενίσχυση κλίσης (GB), ενίσχυση ακραίας κλίσης (XGBoost) και μηχανή ενίσχυσης ελαφριάς κλίσης (lightGBM). Το XGBoost είναι ένας επίσημα καθιερωμένος αλγόριθμος εποπτευόμενης μάθησης για την ανάπτυξη προγνωστικών μοντέλων λόγω της ακρίβειας και της απόδοσής του. Το Boosting εστιάζει σε προβλέψεις με υψηλό σφάλμα στο αρχικό βήμα εκπαίδευσης προσαρμόζοντας την κατανομή του δείγματος για το επόμενο βήμα εκπαίδευσης. Αυτή η διαδικασία συνεχίζεται για πολλαπλές επαναλήψεις έως ότου ο αριθμός των μαθητών φτάσει σε ένα κριτήριο διακοπής. Τα μοντέλα ενίσχυσης δημιουργούν ένα σύνολο αδύναμων μαθητών για να αποκτήσουν δυνατούς μαθητές με καλύτερες επιδόσεις σε σύγκριση με μεμονωμένους αδύναμους μαθητές. Τα δεδομένα που χρησιμοποιήθηκαν ήταν από κτίρια σε πανεπιστημιακές εγκαταστάσεις με διάφορες χρήσεις ενώ ένα κομμάτι των δεδομένων είναι και από την περίοδο της πανδημίας. Σε κάποια από αυτά τα κτίρια

έγινα αναβαθμίσεις αναφορικά με την ενεργειακή κατανάλωση οι οποίες χωρίζονται σε δύο κατηγορίες τις BMS και τις LED, ενώ σε κάποια άλλα κτίρια πραγματοποιήθηκαν και οι δυο. Τελικά το ποσοστό της εξοικονόμησης παρέμεινε σχετικά σταθερό μεταξύ των διαφόρων κτιρίων παρά το γεγονός ότι σε κάποια ακολούθησε καθοδική πορεία με το μέγιστο πάντως να παρατηρείται σε όλες τις περιπτώσεις μόλις οι ECM είχαν ολοκληρωθεί.

Οι Jack Ngarambe, Geun Young Yun, Mat Santamouris [21] ασχολήθηκαν με τον τρόπο που η τεχνητή νοημοσύνη μπορεί να είναι εργαλείο πρόβλεψης της θερμικής άνεσης σε ένα κτίριο. Ως θερμική άνεση ορίζεται η αίσθηση ικανοποίησης που νιώθει κανείς σε ένα χώρο ο οποίος είναι σωστά θερμαινόμενος/κλιματιζόμενος καθώς το ανθρώπινο σώμα έχει μάθει να λειτουργεί κάτω από συγκεκριμένες συνθήκες ανταλλαγής θερμότητας με το περιβάλλον που όταν αυτές δεν ικανοποιούνται προκαλούν δυσφορία στο άτομο. Στη μελέτη αυτή λαμβάνονται υπόψη λεπτομέρειες όπως πχ το ντύσιμο των ανθρώπων μέσα στους χώρους ώστε τα αποτελέσματα των προβλέψεων να είναι πιο ρεαλιστικά. Καταλήγουν ότι ο συνδυασμός μερικών αλγορίθμων μηχανικής μάθησης με δεδομένα για την κατάσταση των ατόμων όπως αυτά μετρούνται από συσκευές συνδεδεμένες με το Ίντερνετ, IoT συσκευές, μπορούν να φέρουν πολύ καλά αποτελέσματα στη μελέτη αυτή. Παρόλα αυτά πρόβλημα είναι ότι δεν υπάρχουν ισχυρές απτές αποδείξεις ότι όντως υπάρχει χρησιμότητα στον έλεγχο της θερμικής άνεσης με συστήματα τεχνητής νοημοσύνης σε πειράματα επί τόπου παρά τις σοβαρές θεωρητικές αποδείξεις.

Οι Alberto Torres-Barrán, Álvaro Alonso, José R. Dorronsoro [22] προσέγγισαν το ζήτημα του συνδυασμού δέντρων παλινδρόμησης και την πρόβλεψη της αιολικής ενέργειας και της ηλιακής ακτινοβολίας. Τα συνδυαστικά μοντέλα είναι αρκετά ελκυστικά αφού έχουν την ικανότητα να διατηρούν την προκατάληψη των μαθητών τους ενώ μειώνουν τις ατομικές τους διακυμάνσεις. Έτσι σε αυτή τη μελέτη ελέγχονται συνδυαστικά μοντέλα και συγκεκριμένα τα μοντέλα της τυχαίας παλινδρόμησης δασών (RFR), της παλινδρόμησης ενισχυμένης κλίσης (GBR) και Extreme Gradient Boosting (XGB). Από αυτά κανένα δεν ήταν καθαρά καλύτερο από τα άλλα αφού στην περίπτωση της αιολικής ενέργειας καλύτερα ήταν τα RFR and XGB ενώ για την ηλιακή ακτινοβολία αρχικά ισοβαθμούν τα GBR and XGB και το RFR έρχεται προ-τελευταίο. Λόγω των πολλών μεταβλητών που παίζουν ρόλο στην επιτυχία των προβλέψεων η καλύτερη λύση είναι η εναλλαγή μεταξύ διαφόρων αλγορίθμων μηχανικής μάθησης καθώς δεν υπάρχει κανένας που να είναι βέβαια καλύτερος από τους υπόλοιπους.

Οι Marc Agenis-Nevers, Yuqi Wang, Muriel Dugachard, Raphael Salvazet, Gwenaelle Becker, Damien Chenu [23] ασχολήθηκαν με την M&V σε διάφορα κτίρια ενώ προτείνουν μια καινοτόμα επιλογή βασικού μοντέλου που εφαρμόζεται στις συμβάσεις πραγματικής ενεργειακής απόδοσης. Χρησιμοποιήθηκαν 11 αλγόριθμοι γραμμικού και μη με στόχο να βρεθεί το πιο σχετικό βασικό μοντέλο ώστε να υπολογιστεί το κέρδος εξοικονόμησης. Τα κτίρια που εξετάστηκαν είναι ξενοδοχεία, πολυκαταστήματα και γραφεία στο Ντουμπάϊ όπου τα έξοδα για κλιματισμό είναι το σημαντικότερο μέρος των λειτουργικών εξόδων ενός κτιρίου. Αρχικά επιδιώχθηκε να βρεθεί το καλύτερο μοντέλο για ένα κτίριο και στη συνέχεια να βρεθεί το κοινό καλύτερο για περισσότερα από ένα κτίρια. Μερικοί από τους αλγόριθμους που χρησιμοποιήθηκαν ήταν οι αλγόριθμοι γραμμικής παλινδρόμησης με τον κανόνα ελαχίστων τετραγώνων, παλινδρόμησης ελάχιστης γωνίας, Μπεϋζιανής γενικευμένης παλινδρόμησης και Cubist που βασίζεται σε δέντρα μεταξύ άλλων. Για την περίπτωση εύρεσης του καλύτερου αλγορίθμου για ένα κτίριο τα αποτελέσματα έδειξαν ότι το GPR ήταν καλύτερο 6 από τις 14 φορές και το BAYES 3 από τις 14. Στην περίπτωση όπου ελέγχονται περισσότερα κτίρια μαζί αυτό που ήταν καλύτερο ήταν

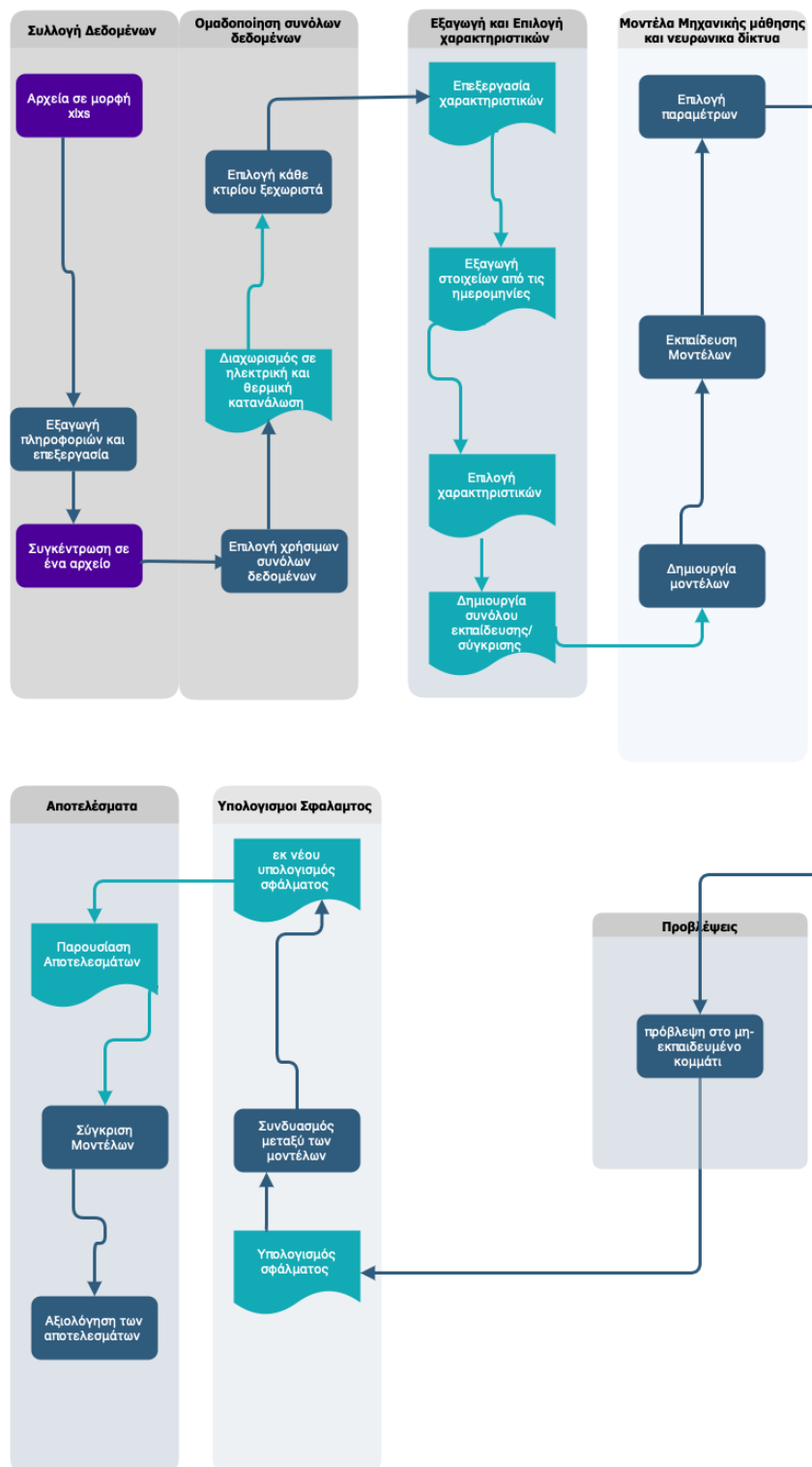
το Gaussian Process αλλά με τόσο μικρή διαφορά από τους άλλους γραμμικούς αλγορίθμους όπου θα μπορούσε κανείς να προβεί σε διαφορετική επιλογή με κριτήριο την ταχύτητα υπολογισμού στην οποία το OLS linear model είναι σαφώς ανώτερο.

Οι Razak Olu-Ajayi, Hafiz Alaka, Ismail Sulaimon, Funlade Sunmola, Saheed Ajayi [24] συνέκριναν διάφορα μοντέλα μηχανικής αλλά και βαθιάς μάθησης σε κατοικίες από το Ηνωμένο Βασίλειο. Στόχοι τους ήταν η σύγκριση απόδοσης μοντέλων ετησίως σε μεγάλο πλήθος κτιρίων, η απόδοση του κάθε αλγορίθμου σε σχέση με την δύναμη υπολογισμού του, η διερεύνηση της επίδρασης των συμπλεγμάτων δόμησης στο επιλεγμένο χαρακτηριστικό και την απόδοση του μοντέλου καθώς και της επίδρασης του μεγέθους των δεδομένων στην απόδοση του μοντέλου. Πέρα από βασικούς αλγόριθμους που έχουν αναφερθεί και νωρίτερα εφαρμόστηκε βαθιά μηχανική μάθηση καθώς και ένα “στοίβαγμα” δηλαδή διαφορετικές τεχνικές μηχανικής εκμάθησης για να φτιαχτεί το μοντέλο. Εκπαιδεύεται ένας αλγόριθμος συνδυασμού για να κάνει τις τελικές προβλέψεις με βάση αυτές των βασικών μοντέλων. Το κύριο όφελος της μεθόδου στοίβαξης είναι η ικανότητα αξιοποίησης των δυνατοτήτων πολλών αλγορίθμων με καλή απόδοση και να πετύχει με το συνδυασμό τους καλύτερα αποτελέσματα με μεγαλύτερη ακρίβεια. Φανερό είναι από την αρχή ότι τα μοντέλα των νευρωνικών δικτύων που χρειάζονται περισσότερα δείγματα από τα απλά μοντέλα μηχανικής μάθησης για να πετύχουν καλά αποτελέσματα αλλά όταν υπάρχουν αυτά τα αρκετά δεδομένα τα αποτελέσματα τους είναι αρκετά καλύτερα. Τα τρία πρώτα κατά αντιστοιχία μοντέλα ήταν τα DNN, ANN και GB με το LR και τον αλγόριθμο στοίβαξης να έρχονται τελευταίοι αλλά να απαιτούν λιγότερο χρόνο για εκπαίδευση. Πιο γρήγορος αλγόριθμος ήταν τα δέντρα αποφάσεων αλλά τα αποτελέσματα τους ήταν αρκετά χειρότερα από τους τρεις πρώτους. Μερικές σημαντικές παρατηρήσεις ήταν ότι η απόδοση του μοντέλου δεν ήταν ευαίσθητη ως προς ένα συγκεκριμένο τύπο κτιρίου και επομένως η ομαδοποίηση κτιρίων δεν είχε κάποιο όφελος.

Οι παραπάνω έρευνες έδωσαν πολύτιμες πτυχές και λύσεις στο πρόβλημα καθώς χρησιμοποιήθηκαν πολλές τεχνικές μηχανικής μάθησης όσο και βαθιάς μάθησης και νευρωνικά δίκτυα. Από αυτά ήταν χαρακτηριστικό ότι σημαντικό ρόλο έπαιζε η μορφή του συνόλου δεδομένων καθώς οι ίδιοι αλγόριθμοι παρουσίαζαν διαφορές στα αποτελέσματα τους όπου κάποιες φορές είχαν μεγαλύτερη ακρίβεια και κάποιες όχι. Άλλη μια παρατήρηση είναι ότι τα νευρωνικά δίκτυα που χρησιμοποιήθηκαν ήταν σχετικά απλά χωρίς να γίνει περαιτέρω εμβάθυνση σε πιο περίπλοκες δομές. Πάνω στη τελευταία παρατήρηση βασίζεται και η λύση που επιλέχθηκε ως η καταλληλότερη ώστε να βασιστεί η παρούσα εργασία.

4. Προτεινόμενη Λύση

Στο κεφάλαιο αυτό θα γίνει εκτενής ανάλυση της προσέγγισης μας στο ζήτημα, θεωρητική επεξήγηση των εργαλείων που χρησιμοποιήθηκαν στον κώδικα[36] καθώς και οι λόγοι επιλογής και διαμόρφωσης τους. Πριν όμως από όλα θα παρουσιαστεί ένα διάγραμμα που θα λειτουργήσει σαν οδηγός για την πορεία της λύσης μας.



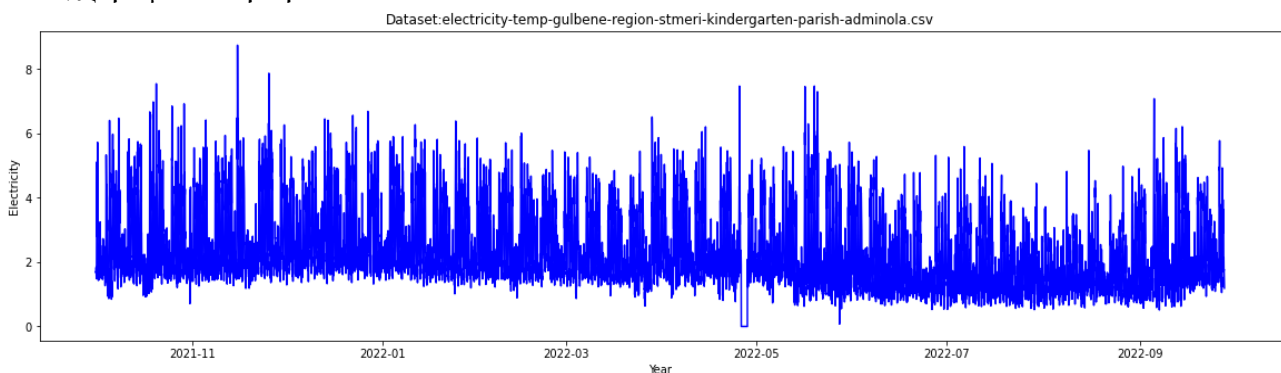
Εικόνα 4.1: Διάγραμμα της προτεινόμενης λύσης της παρούσας διπλωματικής

Αναλυτικά ακολουθεί επεξήγηση για το κάθε μέρος καθώς και για την κάθε ενδιάμεση επιλογή.

4.1 Σύνολα Δεδομένων και Εξαγωγή και επιλογή χαρακτηριστικών

Σύνολα Δεδομένων

Τα δεδομένα που υπάρχουν και πάνω στα οποία δοκιμάστηκαν τα διαφορά μοντέλα είναι από διαφορά κτίρια στη Λετονία. Συγκεκριμένα είναι εμπορικά κτίρια όπως γραφεία, γυμναστήρια και σχολεία. Από αυτά έχουμε την ενεργειακή κατανάλωση αλλά και την κατανάλωση θερμικής ενέργειας. Για το κάθε κτίριο υπάρχουν δεδομένα που έχουν μετρηθεί ανά ώρα και τα περισσότερα από αυτά τις χρονιές 2021-2022. Από τις μετρήσεις αυτές υπήρχαν τιμές που έλλειπαν. Για να μην δημιουργηθεί πρόβλημα με αυτές συμπληρώθηκαν με δύο κυρίως τρόπους. Ο πρώτος όταν υπήρχαν κοντινές τιμές για παράδειγμα αν η τιμή που έλλειπε αντιστοιχούσε στις 15:00 και υπήρχαν οι τιμές στις 16:00 και στις 14:00 χρησιμοποιήθηκε η μέθοδος της παρεμβολής δηλαδή συμπληρωνόταν η τιμή με ένα μέσο όρο των κοντινών της. Στη δεύτερη περίπτωση όπου δηλαδή έλλειπαν αρκετές συνεχόμενες τιμές συμπληρώνονται με βάση την προηγούμενη μέρα ή ένα μέσο όρο της αντίστοιχης ώρας των προηγούμενων ημερών. Η συμπλήρωση όταν έλειπε μια ημερομηνία είναι μια απλή διαδικασία όπου βρίσκεις την προηγούμενη και συμπληρώνεις ανά ώρα μέχρι να βρεθεί η επόμενη που υπάρχει στο σύνολο δεδομένων. Σε κάποια από τα αρχεία δεδομένων υπήρχαν στοιχεία από χρονιές προηγούμενες (2019-2020) όμως ήταν πολύ λίγα και επιλέχθηκε να αφαιρεθούν ώστε να μη δημιουργούν πρόβλημα στην διαδικασία πρόβλεψης. Τα αρχικά δεδομένα ήταν φάκελοι οι οποίοι περιείχαν αρχεία σε μορφή .xlsx με τα στοιχεία για κάθε μήνα. Αφού τα συνενώσαμε για κάθε κτίριο σε ένα αρχείο .csv ακολουθήσαμε την παραπάνω διαδικασία ώστε να “γεμίσουμε” τις τιμές που έλλειπαν. Στη συνέχεια αφού κάναμε την ίδια διαδικασία και για τα αρχεία με τις θερμοκρασίες δημιουργήσαμε δύο ειδών αρχεία, θερμοκρασίας μαζί με κατανάλωση ηλεκτρικής ενέργειας και θερμοκρασίας μαζί με κατανάλωση θερμικής ενέργειας. Σε γενικές γραμμές τα αρχεία αυτά που ήταν τα τελικά και αυτά με τα οποία εκπαιδεύτηκαν τα μοντέλα και περιείχαν τιμές από το Μάιο του 2021 μέχρι και τον Σεπτέμβρη του 2022. Ακολουθεί ένα τυπικό διάγραμμα από ένα από τα αρχεία που χρησιμοποιήθηκαν.



Εικόνα 4.2: Διάγραμμα ενός συνόλου δεδομένων

Στο παραπάνω διάγραμμα παρατηρείται η διακύμανση της ενεργειακής κατανάλωσης στο διάστημα μέχρι και τον Σεπτέμβρη του 2022 για ένα κτίριο διεύθυνσης παιδικού σταθμού. Στο συγκεκριμένο είναι φανερό ότι υπάρχει περιοδικότητα στην κατανάλωση ενώ κάποιες ακραίες τιμές φανερώνουν ότι είτε ο χώρος ήταν κλειστός είτε υπήρχαν ακραίες θερμοκρασίες σε σχέση με τις υπόλοιπες μέρες.

Εξαγωγή και επιλογή χαρακτηριστικών

Από αυτό το σημείο και έπειτα μεταχειριστήκαμε το κάθε αρχείο ξεχωριστά αλλά για όλα ακολουθήθηκε η ίδια διαδικασία. Πρώτα έπρεπε να γίνει εξαγωγή όσων περισσότερων χαρακτηριστικών γίνεται ώστε να υπάρχει μεγαλύτερη πιθανότητα τα μοντέλα μας να έχουν μεγάλη ακρίβεια καθώς και να υπάρχει πλήθος επιλογών ώστε να επιλεχθούν τα καταλληλότερα. Έτσι σε πρώτο χρόνο από την κάθε ημερομηνία διαχωρίστηκαν τα στοιχεία σε διαφορετικές κατηγορίες με βάση τη χρονία, το μήνα, τη μέρα και την ώρα. Αφού οι χρονιές που υπάρχουν στα σύνολα δεδομένων είναι μόλις δυο η αντίστοιχη κατηγορία έγινε σε δυαδική μορφή με το 2021 να λογίζεται ως το 0 και το 2022 ως το 1. Ταυτόχρονα εφόσον υπήρχαν οι ημερομηνίες για αυτά τα δύο χρόνια βρέθηκαν ποιες μέρες αντιστοιχούν σε Σαββατοκύριακα όπως και ποιες αντιστοιχούν σε περίοδο διακοπών ώστε να μπορεί το μοντέλο να διακρίνει τυχόν διαφορετική συμπεριφορά από τα συνηθισμένα και να την αιτιολογεί γνωρίζοντας ότι υπάρχει συγκεκριμένος λόγος και επαναληπτικότητα. Αλλά στοιχεία που εξήχθησαν από τις ημερομηνίες ήταν η ημέρα της εβδομάδας, η εβδομάδα του χρόνου και η ημέρα του χρόνου. Για αυτά τα τρία ισχύει η ίδια λογική δηλαδή γίνεται αρίθμηση των ημερών, εβδομάδων με βάση μια που θεωρείται πρώτη και μετά η κάθε μια αντίστοιχα παίρνει το νούμερο της. Πέρα από τα καθαρά “ημερολογιακά” χαρακτηριστικά που εξήχθησαν υπήρχαν και κάποια αλλά συνδυαστικά. Αρχικά δοκιμάστηκε η μέση θερμοκρασία σε εβδομαδιαίο και μηνιαίο επίπεδο ώστε να γίνει προσπάθεια εύρεσης μιας σχέσης μεταξύ της μέσης θερμοκρασίας με τη διακύμανση της κατανάλωσης. Παράλληλα δοκιμάστηκε η ομαδοποίηση σε κατηγορίες με βάση τη θερμοκρασία με τις ακραίες περιπτώσεις να καταχωρούνται ως την ομάδα 1 ενώ αυτές που ήταν κοντά στο μέσο όρο στην ομάδα 0. Ακόμα επιχειρήθηκε το ίδιο αλλά αντί αυτής της ομαδοποίησης επιλέχθηκε μια καταχώρηση πόντων με βάση ποσό απέχουν οι ακραίες περιπτώσεις από τα όρια που θέσαμε ενώ όσες ήταν μέσα στα όρια έπαιρναν βαθμό 0. Είναι σημαντικό να γίνει γνωστό ότι οι αλγόριθμοι μηχανικής μάθησης αντιμετωπίζουν τη σειρά των αριθμών ως χαρακτηριστικό της σημασίας. Με άλλα λόγια, θα διαβάσουν έναν υψηλότερο αριθμό ως καλύτερο ή πιο σημαντικό από έναν χαμηλότερο αριθμό. Γι’ αυτό το λόγω εφαρμόστηκε ένας τριγωνομετρικός μετασχηματισμός στα χαρακτηριστικά όπου υπήρχε αρίθμηση ώστε να αποφευχθεί τέτοιου είδους σύγχυση στα μοντέλα. Μετά από αρκετές δοκιμές καταλήξαμε σε δυο σύνολα χαρακτηριστικών τα οποία ταίριαζαν καλύτερα στα διαφορά μοντέλα.

4.2 Αλγόριθμοι για σύγκριση και Κύρια Μοντέλα

Αλγόριθμοι για σύγκριση

Αφού ολοκληρώθηκε και η εξαγωγή και επιλογή χαρακτηριστικών ακολούθησε η εφαρμογή απλών αλγορίθμων μηχανικής μάθησης ώστε να υπάρχει μια καλή βάση για σύγκριση της επίδοσης των μοντέλων μας. Οι αλγόριθμοι που επιλέχθηκαν ήταν οι κ- κοντινότεροι γείτονες, τα τυχαία δέντρα και η λογιστική παλινδρόμηση ενώ αναπτύχθηκε και ένας αλγόριθμος παλινδρόμησης στα πρότυπα του Heating-Cooling degree days. Αναλυτικά ο τρόπος λειτουργίας των δύο πρώτων έχει αναφερθεί νωρίτερα. Η λογιστική παλινδρόμηση εκτιμά την πιθανότητα να συμβεί ένα γεγονός, μετά από αποτέλεσμα ψηφοφορίας με βάση ένα δεδομένο σύνολο ανεξάρτητων μεταβλητών. Εφόσον το αποτέλεσμα είναι μια πιθανότητα, η εξαρτημένη μεταβλητή οριοθετείται μεταξύ 0 και 1. Η διαφορά της από τη γραμμική παλινδρόμηση είναι ότι συνήθως εκτιμά κατηγορηματικά δεδομένα και όχι συνεχή. Όσον αφορά το heating-cooling degree days [25] ακολουθούν τη λογική των βαθμοημερών θέρμανσης και ψύξης. Ως βαθμοημέρες θέρμανσης ορίζονται όσες μέρες έχουν μέση θερμοκρασία πάνω από το όριο των 18.3 βαθμών Κελσίου και βασίζεται στη θεωρία ότι όταν επικρατεί αυτή η θερμοκρασία δεν είναι ανάγκη να γίνει χρήση ούτε θέρμανσης ούτε

ψύξης. Αντίστοιχα βαθμοημέρες ψύξης είναι αυτές που η μέση θερμοκρασία του είναι μικρότερη του προαναφερθέντος ορίου. Αφού προσδιοριστούν οι βαθμοημέρες ακολουθείται η παρακάτω διαδικασία παλινδρόμησης:

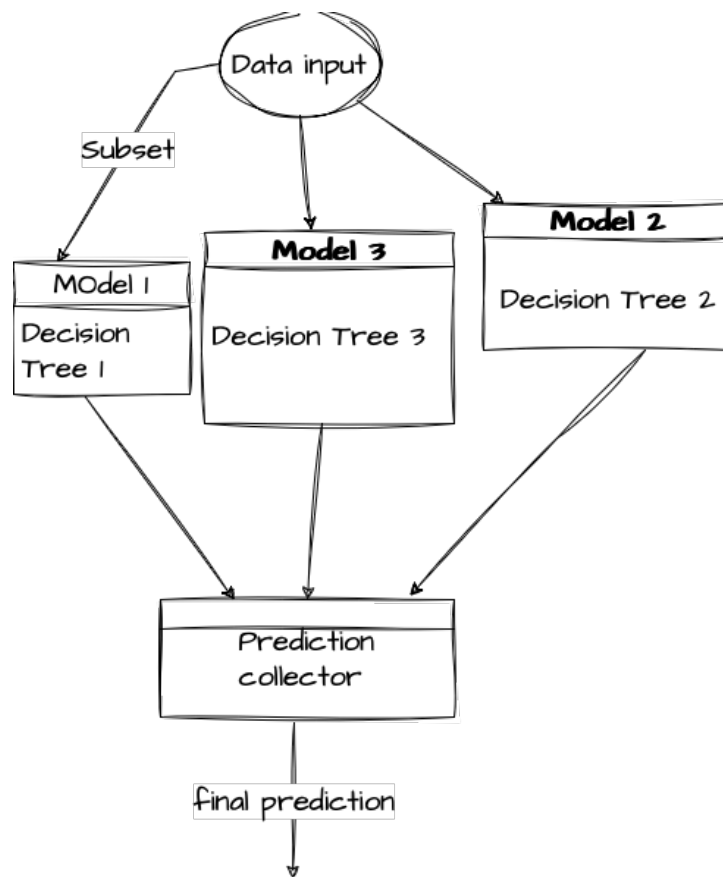
$$E = b \cdot \text{days} + h \cdot \text{HDD} + c \cdot \text{CDD}$$

Όπου χρησιμοποιήθηκε αυτή τη συνάρτηση σαν τον αλγόριθμο της γραμμικής παλινδρόμησης, με τα b , h , c να είναι οι συντελεστές της παλινδρόμησης και οι όροι days , HDD , CDD να είναι το σύνολο των ημερών, οι βαθμοημέρες θέρμανσης και οι βαθμοημέρες ψύξης αντίστοιχα. Τα αποτελέσματα αυτών των αλγορίθμων δεν είναι τα πιο ακριβή ακριβώς γιατί είναι πολύ απλοί αλγόριθμοι χωρίς πολλές δυνατότητες βελτίωσης. Λειτουργούν καθαρά ως μέτρο σύγκρισης για να είναι ξεκάθαρο αν και κατά ποσό τα πιο σύνθετα μοντέλα μπορούν να πετύχουν μεγαλύτερη ακρίβεια. Μεταξύ άλλων υπήρξαν και κάποιοι άλλοι αλγόριθμοι που δοκιμάστηκαν αλλά τα αποτελέσματα ήταν πολύ χειρότερα επόμενων δεν υπάρχει λόγος να γίνει κάποια αναφορά σε αυτά.

Κύρια Μοντέλα

Τα μοντέλα τα οποία εφαρμόστηκαν στα πλαίσια της παρούσας εργασίας είναι όπως προαναφέραμε το XGBoost [26] και τα LSTM [27]. Και τα δυο δοκιμάστηκαν με διάφορες παραμέτρους σε διάφορα σύνολα χαρακτηριστικών και από αυτά επιλέχθηκαν τα πιο ακριβή. Ακολουθεί μια γενική θεωρητική προσέγγιση στον τρόπο λειτουργίας τους και μετά η ανάλυση των επιλογών που έγιναν από την πλευρά μας με στόχο τα καλύτερα αποτελέσματα.

Τα XGB μοντέλα, που το όνομά τους προκύπτει από το **Extreme Gradient Boosting**, αποτελούν μοντέλα που χρησιμοποιούν την τακτική του **ensemble** δηλαδή χρησιμοποιούν περισσότερα από ένα μοντέλα τα οποία δοκιμάζονται και οι τελικές προβλέψεις είναι ο συνδυασμός των επιμέρους που δοκιμάστηκαν. Όταν αυτό συμβαίνει σε περιπτώσεις **Gradient boosting** η διαδικασία που ακολουθείται είναι να εφαρμόζεται πρώτα ένα μοντέλο και στη συνέχεια στο υπόλοιπο του πρώτου μοντέλο να εφαρμόζεται ένα δεύτερο και στο άθροισμα των δύο πρώτων ένα ακόμα. Έτσι στην περίπτωση του XGB έχουμε έναν αλγόριθμο **Gradient boosting** που λειτουργεί έχοντας απλούστερα μοντέλα για τις επιμέρους προβλέψεις, δέντρα απόφασης ή γραμμικής παλινδρόμησης. Επιλέχθηκε η χρήση των δέντρων απόφασης καθώς παρέχουν τη δυνατότητα εύρεσης και αναπαράστασης μη γραμμικών σχέσεων και δεν έχουν σαν αναγκαία συνθήκη τη γραμμική σχέση μεταξύ εισόδου και εξόδου.

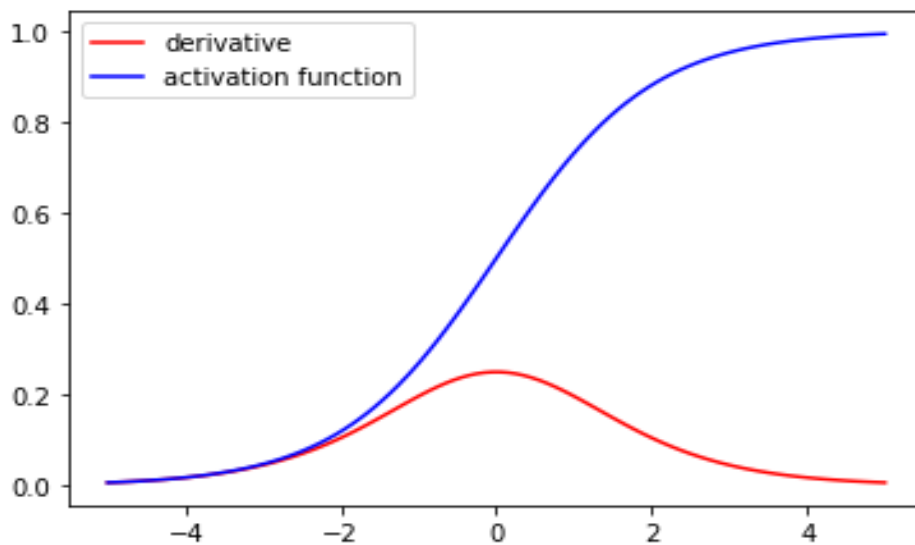


Εικόνα 4.3: Λογική του Gradient boosting αλγορίθμου

Δεν είναι τυχαίο ότι το καλύτερο βάθος συνδέεται με τα χαρακτηριστικά που χρησιμοποιούνται αφού ενώ φαινομενικά τα χαρακτηριστικά πάνω στα οποία εκπαιδεύεται είναι 10 στο πλήθος υπάρχουν τέσσερα που ανά δύο δεν είναι ανεξάρτητα μεταξύ τους και επομένως υπάρχουν 8 ανεξάρτητα χαρακτηριστικά και έτσι το βάθος είναι 8. Επόμενη απαραίτητη μεταβλητή είναι η “objective” η οποία καθορίζει το στόχο του μοντέλου. Επιλέγουμε τη γραμμική παλινδρόμηση ώστε να έχει μεγαλύτερη επιτυχία στις προβλέψεις. Άλλη μια παράμετρος που ορίστηκε ήταν οι φορές που θα επαναληφθούν οι γύροι ενίσχυσης, δηλαδή η διαδικασία του Gradient boosting. Οι 1000 φορές που επιλέχθηκαν μπορεί να είναι ένα μεγάλο νούμερο και να ελλοχεύει κινδύνους και γι’ αυτό χρησιμοποιήθηκε η τεχνική της πρόωρης διακοπής για να αποφευχθεί το overfitting.

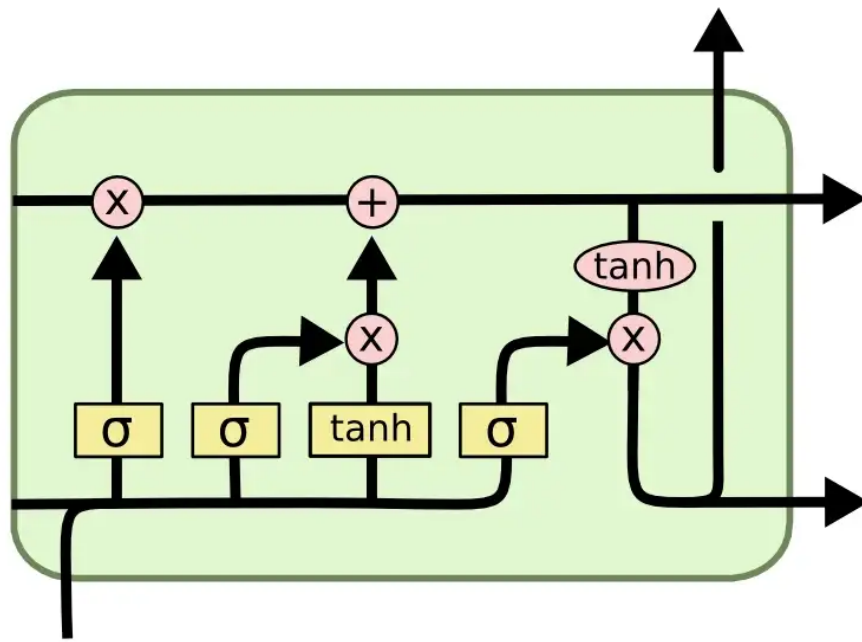
Μετά από αυτό χρησιμοποιήθηκαν τα LSTM και συγκεκριμένα δύο επίπεδα σειριακά ανάλυση του οποίου θα γίνει αμέσως μετά την παρουσίαση του θεωρητικού υποβάθρου των LSTM. Τα long short-term memory (LSTM) networks είναι νευρωνικά δίκτυα τα οποία είναι προηγμένα μοντέλα που αποσκοπούν στη εύρεση λύσης στο πρόβλημα της μακροπρόθεσμης εξάρτησης. Τα συμβατικά μοντέλα recurrent neural network (RNN) αντιμετωπίζουν συχνά το πρόβλημα που στη βιβλιογραφία αναφέρεται ως vanishing gradient problem. Τα RNN είναι νευρωνικά δίκτυα τα οποία διαθέτουν μια σύνδεση μεταξύ εξόδου-εισόδου στους ίδιους νευρώνες των κρυφών.

Η σύνδεση αυτή υπάρχει η περίπτωση να εισέλθει ένα προηγούμενο δεδομένο μαζί με τα νέα στην είσοδο του νευρώνα και άρα να δημιουργηθεί από το δίκτυο μια σχέση μεταξύ προηγούμενων και νεότερων δεδομένων. Μπορεί εκ πρώτης όψευς αυτό να είναι ένα σημαντικό πλεονέκτημα για τη χρήση τους όμως όπως αναφέρθηκε αντιμετωπίζουν συχνά το vanishing gradient problem. Συγκεκριμένα το πρόβλημα αυτό προκύπτει όπου η παραγωγός της συνάρτησης ενεργοποίησης φθίνει τόσο πολύ ώστε δεν μετέχει σε καμία αλλαγή στα βάρη με αποτέλεσμα να μην εκπαιδεύεται το δίκτυο και άρα να μην υπάρχει αύξηση αλλά μείωση στην ακρίβεια του.



Εικόνα 4.4: Vanishing gradient problem.

Η ειδοποιός διαφορά που έχουν τα LSTM από τα απλά RNN είναι ότι διαθέτουν συνδέσεις ανάδρασης οι οποίες πρακτικά επιτρέπουν στα μοντέλα να μπορούν “δουν” και “πίσω” στο χρόνο. Με το πλεονέκτημα αυτό επεξεργάζονται ακολουθίες δεδομένων και εντοπίζουν σχέσεις μεταξύ των σημείων που έχουν εισέλθει και των νέων που εισέρχονται αργότερα. Ταυτόχρονα έχει τη δυνατότητα να επιλέγει ποτέ και αν θα χρησιμοποιηθούν διάφορες πύλες ώστε τα βάρη να ανανεώνονται σωστά και να αποφεύγεται το πρόβλημα που εξηγήθηκε προηγουμένως. Από τη δομή του LSTM φαίνεται και ο τρόπος με τον οποίο επιτυγχάνεται αυτός ο έλεγχος. Ακολουθεί σχήμα που δείχνει ακριβώς τον τρόπο λειτουργίας.



Εικόνα 4.5: LSTM Architecture (source: <https://colah.github.io/>)[28]

Θα ακολουθήσει ανάλυση της λειτουργίας του παραπάνω σχήματος η οποία θα χωριστεί σε τρία βήματα. Πρώτα υπάρχει η πύλη που ονομάζεται forget gate και αναλαμβάνει να αποφασίσει ποια είναι τα χρήσιμα bits της μακροπρόθεσμης μνήμης τόσο από την προηγούμενη κρυφή κατάσταση όσο και από τη νέα εισόδο. Για να επιτευχθεί αυτή η επιλογή όλες οι πληροφορίες εισέρχονται σε ένα άλλο δίκτυο το οποίο ουσιαστικά τις κατανέμει με βάση τη σημασία τους και τις πολλαπλασιάζει με το κατάλληλο βάρος ώστε στη συνέχεια να παίξουν ρόλο ανάλογα με το πόσο σημαντικές είναι. Το δεύτερο μέρος είναι αυτό της νέας μνήμης και της πύλης εισόδου. Είναι υπεύθυνο για την επιλογή των νέων δεδομένων που θα προστεθούν στη μακρά μνήμη σε σχέση με τα ήδη υπάρχοντα στοιχεία του κρυφού επιπέδου αλλά και τα νέα δεδομένα της εισόδου. Όπως φαίνεται και από το σχήμα παραπάνω οι εισόδοι σε αυτό το βήμα είναι ίδιες με το προηγούμενο. Η βασική διαφορά του από το προηγούμενο βήμα είναι η χρήση της συνάρτησης ενεργοποίησης όπου σε αντίθεση με το προηγούμενο βήμα δεν επιλέγεται μόνο η σιγμοειδής αλλά και η tanh η οποία δίνει τη δυνατότητα στα βάρη να κυμαίνονται από το -1 στο 1 άρα να έχουν και αρνητικές τιμές. Τα αποτελέσματα αυτά περνάνε από ένα στάδιο πολλαπλασιασμού μεταξύ τους πριν τροφοδοτηθούν στο επόμενο βήμα. Τρίτο και τελευταίο βήμα είναι η έξοδος όπου λαμβάνεται η απόφαση για την νέα κρυφή κατάσταση. Για αυτή την επιλογή γίνεται χρήση των στοιχείων της εισόδου, τη προηγούμενη κρυφή κατάσταση καθώς και τα ενημερωμένα βάρη τα οποία επεξεργάζεται με τρόπο αντίστοιχο του πρώτου βήματος και καταλήγει στην έξοδο η οποία είναι περιορισμένη στο διάστημα 0-1 ελέω της χρήσης της σιγμοειδούς συνάρτησης ενεργοποίησης. Το τελικό αποτέλεσμα του δικτύου προκύπτει μετά από πολλές επαναλήψεις των προηγούμενων τριών βημάτων. Εκτός από το απλό LSTM έγινε και χρήση ενός αμφίδρομου LSTM το οποίο έχει κάποιες διαφορές από το απλό. Η βασική διαφορά που φαίνεται και από το όνομα είναι ότι η εισόδος κινείται και προς τις δυο πλευρές ενώ και τα στοιχεία που χρησιμοποιεί μπορεί να είναι και από τις δύο πλευρές. Η κατασκευαστική διαφορά είναι η προσθήκη ενός ακόμα επιπέδου που αντιστρέφει τη

ροή των δεδομένων και έτσι δημιουργείται μια ακόμα έξοδος. Οι δυο λοιποί έξοδοι που έχουν δημιουργηθεί συνδυάζονται μεταξύ τους ώστε τελικά να προκύψει μια η οποία θα είναι και η τελική έξοδος του δικτύου. Πρακτικά είναι σαν να υπάρχουν δυο LSTM όπου το ένα κοιτά το άλλο και τα αποτελέσματα τους συνδυάζονται σε μια τελική πύλη. Το πλεονέκτημα του σε σχέση με το απλό είναι ότι μπορεί να εντοπίσει εξαρτήσει προς όλες τις κατευθύνσεις πράγμα που δίνει στο προβλεπτικό μηχανισμό έναν ακόμα άξονα όπου μπορεί να βασίσει τις προβλέψεις του. Πριν το τελικό αποτέλεσμα του μοντέλου μας έχουμε ένα dense layer όπου έρχονται οι έξοδοι από τους προηγούμενους νευρώνες και αναλαμβάνει να πραγματοποιήσει με τις εξόδους αυτές πολλαπλασιασμό πίνακα με διάνυσμα.

Τα LSTM μοντέλα που χρησιμοποιήσαμε ήταν δυο και έγινε προσπάθεια εκτός από σύγκριση μεταξύ τους να συνδυαστούν τα αποτελέσματα τους ώστε να ελεγχθεί και η περίπτωση της “συνεργασίας” των δύο αυτών μοντέλων. Και τα δυο περιέχουν ένα απλό και ένα αμφίδρομο LSTM με το πρώτο να συνδέεται με το δεύτερο σειριακά και μετά να ακολουθούν τα Dense επίπεδα [29],[30]. Η διαφορά που έχουν μεταξύ τους είναι ότι στο ένα από τα δυο χρησιμοποιείται κατά την εκπαίδευση ένα validation set το οποίο λειτουργεί ως μέτρο για τη συχνή αξιολόγηση της εκπαίδευσης. Μπορεί να ελλοχεύει ο κίνδυνος να γίνει το μοντέλο αρκετά προκατειλημμένο αλλά με σωστή χρήση μπορεί να βοηθήσει στην επιλογή των κατάλληλων παραμέτρων. Η διαφορά που δημιουργείται στα δικά μας μοντέλα είναι ότι το μοντέλο χωρίς αυτό το υποσύνολο δεδομένων έχει μεγαλύτερη ελευθερία στις προβλέψεις και αυτές γίνονται με μεγαλύτερη διακύμανση. Αλλη μια μικρή διαφορά είναι η χρήση δεύτερου Dense επιπέδου ουσιαστικά για να έχουμε λίγο μεγαλύτερα οφέλη από τη λειτουργία του.

4.3 Προετοιμασία εισόδου και Επιλογή παραμέτρων

Προετοιμασία εισόδου

Άλλο ένα σημαντικό βήμα για τη χρήση των αλγορίθμων αυτών είναι η προετοιμασία της εισόδου των μοντέλων. Επειδή τα διαφορά μοντέλα που χρησιμοποιούνται έχουν κάποιες ανάγκες για να λειτουργήσουν σωστά επεξεργαζόμαστε τις εισόδους τους καταλλήλως. Αρχικά για τα μοντέλα LSTM οι πίνακες των δεδομένων έρχονται σε μορφή τέτοια ώστε η μια στήλη να έχει μέγεθος όσο το πλήθος των χαρακτηριστικών που θα χρησιμοποιούν. Αυτό χρειάζεται ώστε να δημιουργηθεί σωστά το επίπεδο εισόδου των μοντέλων το οποίο περιέχει τους κατάλληλους νευρώνες εισόδου με σκοπό το πέρασμα των δεδομένων στο κύριο σώμα δηλαδή τα υπόλοιπα επίπεδα για να γίνει η επεξεργασία που αναφέρθηκε νωρίτερα.

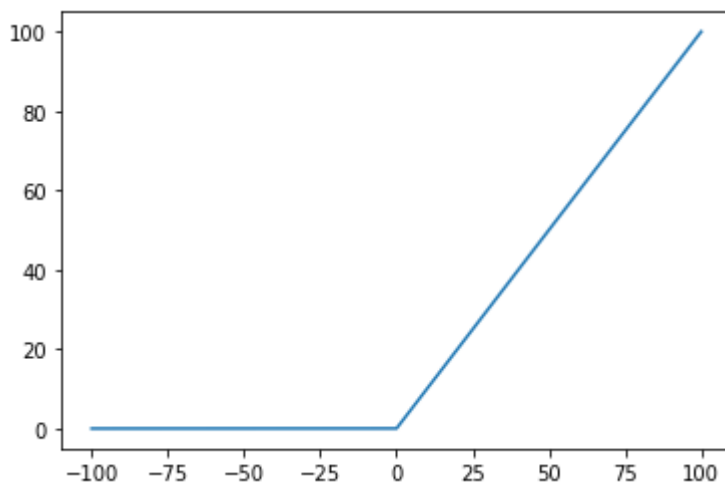
Επιλογή παραμέτρων

Πολύ σημαντικό ρόλο στα αποτελέσματα των μοντέλων παίζει η σωστή επιλογή των παραμέτρων. Σε αυτό το υποκεφάλαιο θα αναλυθεί ο λόγος πίσω από τις επιλογές στην τιμή των παραμέτρων που επιλέξαμε.

Units: Είναι η παράμετρος μέσω της οποίας καθορίζονται τα κρυφά επίπεδα ενός LSTM. Όσο περισσότερα είναι τα κρυφά επίπεδα τόσο πιο πολύπλοκες σχέσεις μεταξύ των δεδομένων θα μπορεί να εντοπίσει το μοντέλο. Επιλέχθηκε η τιμή 80 η οποία ουσιαστικά είναι 10 φορές το πλήθος των χαρακτηριστικών.

Activation functions: Οι συναρτήσεις ενεργοποίησης έχουν ευθύνη για τα αποτελέσματα και τη μορφή της εξόδου ενός δικτύου. Βασικός τους ρόλος είναι να αντιστοιχίζουν τις τιμές της εξόδου σε διαστήματα συνήθως στα 0 έως 1 ή -1 έως 1. Επιλέχθηκαν δυο τέτοιες συναρτήσεις οι linear και η ReLU. Η πρώτη λειτουργεί με βάση τη συνάρτηση της ευθείας και ενώ μπορεί να μην βοηθά με τα πολύπλοκα

δεδομένα είναι χρήσιμη στα dense επίπεδα ώστε να γίνεται η κατάλληλη αντιστοίχιση και να προκύπτουν σωστά οι προβλέψεις. Η δεύτερη συνάρτηση είναι πιο συνθέτη και μη γραμμική. Αρχικά έχει τιμή 0 για όλα τα στοιχεία που είναι μικρότερα του μηδέν και στη συνέχεια παίρνει τιμές με βάση την εξίσωση της ευθείας όπως φαίνεται και στο σχήμα που ακολουθεί.



Εικόνα 4.6: Διάγραμμα της ReLU

Το σημαντικότερο πλεονέκτημα της συγκεκριμένης συνάρτησης ενεργοποίησης είναι ότι με την εξίσωση αυτή λύνει το πρόβλημα του vanishing ή exploding gradient. Αυτό συμβαίνει αφού όλα τα αρνητικά δεδομένα αντιστοιχίζονται στο μηδέν με αποτέλεσμα να μην μπορούν να αλλάξουν τα βάρη και άρα να μην καταστρέφουν την ακρίβεια του μοντέλου.

Adam Optimizer [31]:

Μέσα στα μοντέλα μας επιλέγεται η χρήση του `keras.optimizers.Adam`. Συγκεκριμένα το όνομα Adam προκύπτει από τα αρχικά των λέξεων Adaptive Moment Estimation και είναι ένας αλγόριθμος βελτιστοποίησης για την περίπτωση του gradient descent. Η χρήση του επιλέγεται σε σύγκριση με άλλους καθώς λειτουργεί καλά με μεγάλο πλήθος δεδομένων καθώς δεν έχει μεγάλο υπολογιστικό χρόνο και άρα δε δημιουργεί μεγάλη καθυστέρηση. Αποτελείται από δύο αλγόριθμους τον Momentum και τον Root Mean Square Propagation (RMSP). Ο πρώτος επιταχύνει τον αλγόριθμο gradient descent σύμφωνα με τον «εκθετικά σταθμισμένο μέσο όρο» των κλίσεων. Με τη χρήση μέσων όρων επιτυγχάνεται η σύγκλιση του αλγόριθμου προς τα ελάχιστα με ταχύτερο ρυθμό. Ο δεύτερος είναι μια βελτιωμένη περίπτωση του AdaGrad όπου χρησιμοποιείται ο exponential moving average.

Root Mean Squared Error[32]:

Ως κύρια συνάρτηση σφάλματος επιλέγεται η rmse καθώς είναι ένας κλασικός τρόπος υπολογισμού του σφάλματος στην πρόβλεψη ποσοτικών δεδομένων. Με βάση τον τύπο υπολογισμού της δίνει έμφαση στα μεγάλα λάθη και άρα είναι η καταλληλότερη σε περιπτώσεις που θέλει κανείς να πετυχαίνει προβλέψεις με όσο το δυνατόν μεγαλύτερη ακρίβεια. Η εξίσωση της είναι:

$$RMSE = \sqrt{\sum (\hat{y}_i - y_i)^2 \div n}$$

Όπως φαίνεται όσο μεγαλύτερα είναι τα σφάλματα ακόμα και λίγα στο πλήθος δημιουργούν μεγαλύτερο σφάλμα σε σχέση με πολλά σε πλήθος αλλά μικρά σε τιμή. Για να βρεθεί το ποσοστό αυτού του σφάλματος σε σχέση με το κάθε σύνολο δεδομένων γίνεται διαίρεση του αποτελέσματος της παραπάνω συνάρτησης με το μέσο όρο των τιμών του πραγματικού συνόλου και πολλαπλασιάζεται με το 100.

Mean Bias Error[33]:

Το μέσο σφάλμα προκατάληψης είναι και αυτό μια χρήσιμη μετρική συνάρτηση για τα διάφορα μοντέλα. Μπορεί λόγω της εξίσωσης της να έχει την πιθανότητα τα αρνητικά σφάλματα να εξουδετερώνουν τα θετικά αλλά μπορεί να δώσει πληροφορίες για την κατεύθυνση των σφαλμάτων άρα και προς ποια κατεύθυνση επιδέχεται βελτίωση. Η εξίσωση είναι:

$$MBE = \frac{\sum_{i=1}^n (y_i - \widehat{y}_i)}{n}$$

Στη διπλωματική χρησιμοποιείται η παραλλαγή της, NMBE που σημαίνει ότι το αποτέλεσμα της παραπάνω εξίσωσης διαιρείται με το μέσο όρο των τιμών του test set και πολλαπλασιάζεται με το 100 ώστε να βρεθεί το ποσοστό του σφάλματος σε σχέση με το σύνολο εξέτασης.

Mean Absolute Error[34]:

Για την εκπαίδευση των μοντέλων επιλέγεται η “mae” καθώς βοηθά να επιτευχθεί καλύτερη εκπαίδευση με μικρότερο λάθος. Αυτό συμβαίνει επειδή η συγκεκριμένη συνάρτηση ενώ υπολογίζει το μέγεθος του σφάλματος για όλες τις προβλέψεις δεν επηρεάζεται από την κατεύθυνση του λάθους. Η εξίσωση της είναι:

$$mae = \frac{\sum_{i=1}^n abs(y_i - \lambda(x_i))}{n}$$

Συνοπτικά αποτελεί γραμμική βαθμολογία δηλαδή οι επιμέρους διαφορές σταθμίζονται εξίσου στο μέσο όρο.

Σε άλλες λιγότερο σημαντικές μεταβλητές για το τελικό αποτέλεσμα ανήκουν οι εποχές, το βήμα εκμάθησης και το batch size [35]. Όλες αυτές παίζουν ρόλο στο τελικό αποτέλεσμα αλλά αρκετά λιγότερο από τις προαναφερθείσες. Συνοπτικά το βήμα εκμάθησης “δίνει” αξία στις νέες πληροφορίες που εισέρχονται στο μοντέλο και ρυθμίζει την ταχύτητα με την οποία αυτό μαθαίνει. Ο κίνδυνος που ελλοχεύει είναι ότι ένας μικρός ρυθμός μπορεί να συγκλίνει πολύ αργά και σε τοπικό ελάχιστο αλλά ένας μεγάλος δεν πετυχαίνει πάντα καλά αποτελέσματα. Οι εποχές ορίζουν πόσες φορές θα επαναληφθεί η διαδικασία μάθησης όμως μεγάλος αριθμός εποχών μπορεί να οδηγήσει σε overfitting. Όσον αφορά το batch size είναι υπεύθυνο για τον αριθμό των δειγμάτων εκπαίδευσης που θα επεξεργαστούν πριν γίνει αλλαγή στις εσωτερικές παραμέτρους του μοντέλου.

Μετά από αυτά ακολούθησαν οι γραφικές παραστάσεις των προβλέψεων από αυτά τα μοντέλα όπως και τις προβλέψεις συνδυασμών των δυο LSTM, του καθενός με το XGBoost και όλων μαζί.

Σε αυτό το κεφάλαιο αναλύθηκαν οι επιλογές που έγιναν όπως και τα μοντέλα που χρησιμοποιήθηκαν. Τονίστηκαν οι κυριότερες παράμετροι και ο τρόπος με τον οποίο έγινε η επιλογή στις τιμές τους. Με αυτά τα κριτήρια προέκυψαν τα αποτελέσματα που θα παρουσιαστούν στο επόμενο κεφάλαιο.

5. Παρουσίαση Αποτελεσμάτων

5.1 Εισαγωγικά στοιχεία

Στο παρών κεφάλαιο θα γίνει παρουσίαση των αποτελεσμάτων όπως αυτά προέκυψαν από τα μοντέλα που περιεγράφηκαν νωρίτερα. Τα τελευταία εφαρμόστηκαν σε διάφορα σύνολα δεδομένων και με διαφορετικές δειγματοληψίες, δηλαδή σε ημερήσιες, 8 ωρών και ωριαίες τιμές. Μετά από την παρουσίαση αυτή των αποτελεσμάτων θα τονιστούν τα σημαντικότερα συμπεράσματα όπως και οι λόγοι για τους οποίους κάποια σύνολα δεδομένων έχουν πολύ καλύτερα αποτελέσματα από άλλα.

Πρώτα θα εξηγηθούν οι λόγοι που επιλέχθηκαν αυτές οι διαφορετικές χρονικές στιγμές δειγματοληψίας. Αρχικά τα δεδομένα έτσι όπως έχουν ληφθεί αποτελούν τιμές θερμοκρασίας και κατανάλωσης ανά ώρα. Αυτό σημαίνει ότι υπάρχουν 24 τιμές τη μέρα με ότι αυτό συνεπάγεται για το πλήθος των δεδομένων σε ένα μήνα - χρόνο. Η επιλογή των 8 ωρών γίνεται γιατί τόσες είναι οι ώρες που λειτουργεί συνήθως ένα δημόσιο κτίριο όπως επίσης είναι και το υποχρεωτικό ωράριο εργασίας για τους εργαζόμενους. Έτσι γίνονται δύο ειδών εκ νέου δειγματοληψίες με την πρώτη να προκύπτει από άθροισμα των τιμών ενέργειας και μέσο όρο των τιμών των υπολοίπων χαρακτηριστικών σε αυτό το διάστημα όπως και μέσο όρο όλων των χαρακτηριστικών συμπεριλαμβανομένης και της κατανάλωσης. Η διαδικασία είναι ανάλογη για την ημερήσια δειγματοληψία.

Άλλο ένα σημαντικό στοιχείο είναι ο τρόπος με τον οποίο προέκυψαν τα σύνολα εκπαίδευσης και εξέτασης. Υπάρχουν δύο προσεγγίσεις και σε αυτό το κομμάτι με πρώτη επιλογή να είναι ο σειριακός διαχωρισμός όπου από το τέλος του συνόλου δεδομένων και μέχρι και ένα διάστημα προς την αρχή οριοθετείται το test set και το υπόλοιπο που μένει είναι το training set. Δεύτερος τρόπος είναι αυτός των shuffled δεδομένων δηλαδή τα δυο σετ τιμών δημιουργούνται με τυχαίες τιμές από το σύνολο δεδομένων με ένα ποσοστό συνήθως 80-20.

Ένα ακόμα σημαντικό πράγμα που πρέπει να συμπεριληφθεί πριν την παρουσίαση των αποτελεσμάτων είναι και οι διάφορες μορφές των συνόλων δεδομένων που θα χρησιμοποιηθούν σε αυτό το κεφάλαιο με βάση τη δειγματοληψία τους. Έτσι η μορφή των παρακάτω θα είναι ως εξής. Αρχικά θα γίνεται η παρουσίαση των συνόλων δεδομένων ακολουθούμενη από το διάγραμμα του συνόλου δεδομένων και θα αναφέρεται ο τύπος του κτιρίου και αν υπάρχουν ορισμένες ιδιαιτερότητες του. Στη συνέχεια θα γίνεται γραφική απεικόνιση και για τα τέσσερα διαστήματα δειγματοληψίας ενώ θα ακολουθεί διαχωρισμός τους και σε κάθε ένα θα παρουσιάζεται η τιμή του σφάλματος σε συνδυασμό με τα διαγράμματα για τα επτά μοντέλα(3 βασικά και 4 συνδυασμοί). Η διαδικασία αυτή θα γίνει δυο φορές μια για τη σειριακή περίπτωση και μια για αυτή που ανακατεύονται τα δεδομένα. Ύστερα θα γραφτούν γενικά τα συμπεράσματα που εξάγονται.

ΣΥΝΟΛΑ ΔΕΔΟΜΕΝΩΝ

Σαν πρώτο υποκεφάλαιο θα οριστεί η ανάλυση των συνόλων δεδομένων με διαφορά χαρακτηριστικά από κάποια από τα σύνολα δεδομένων των οποίων τα αποτελέσματα θα αναλυθούν στη συνέχεια. Θα γίνει ένας πίνακας με τα βασικά χαρακτηριστικά των συνόλων δεδομένων που θα παρουσιάσουμε.

Number	Name	Starting Date	End Date
1	Gulbene_region_Stāmerienas_kindergarten_parish_administration	2019-09-01 01:00:00	2022-09-28 00:00:00
2	Gulbene_region_Gulbene_library	2021-05-01 01:00:00	2022-09-28 00:00:00
3	Gulbene_region_Lejasciema_school	2021-05-01 01:00:00	2022-09-28 00:00:00
4	Gulbene_region_Gulbene_secondary_school_nr.2	2021-05-01 01:00:00	2022-09-28 00:00:00
5	Gulbene_region_Stāķi_gym	2021-05-01 01:00:00	2022-09-28 00:00:00
6	Gulbene_region_Jaungulbene_kindergarten_doctorate	2021-10-01 01:00:00	2022-09-28 00:00:00
7	Gulbene_region_Līgo_culture_center_parish_administration_library	2021-05-01 01:00:00	2022-09-28 00:00:00
8	Gulbene_region_Litenes_gym	2022-01-01 00:00:00	2022-09-28 00:00:00

Πίνακας 5.1: Σύνολα Δεδομένων από τα οποία προκύπτουν τα αποτελέσματα

Αυτά θα είναι τα σύνολα δεδομένων που θα χρησιμοποιηθούν ώστε να παρουσιαστούν τα αποτελέσματα των μοντέλων. Τα πρώτα 6 αφορούν την κατανάλωση της ηλεκτρικής ενέργειας ενώ τα δύο τελευταία αφορούν τη θερμική. Για το κάθε ένα θα ακολουθήσει προβολή του σε διάγραμμα καθώς και λεπτομέρειες για τα χαρακτηριστικά του.

Όλα τα σύνολα δεδομένων όπως αναφέρθηκε και νωρίτερα είναι από κτίρια σε περιοχές της Λετονίας. Με βάση το ημερολόγιο της χώρας το πρώτο χαρακτηριστικό που αξίζει να σημειωθεί είναι οι αργίες- διακοπές. Μέσα από χρήση κατάλληλης βιβλιοθήκης προέκυψαν οι ημερομηνίες που αποτελούν αργίες για τα ημερολογιακά έτη 2021, 2022. Ο πίνακας που ακολουθεί περιέχει τόσο την ημερομηνία όσο και την ονομασία της αντίστοιχης αργίας.

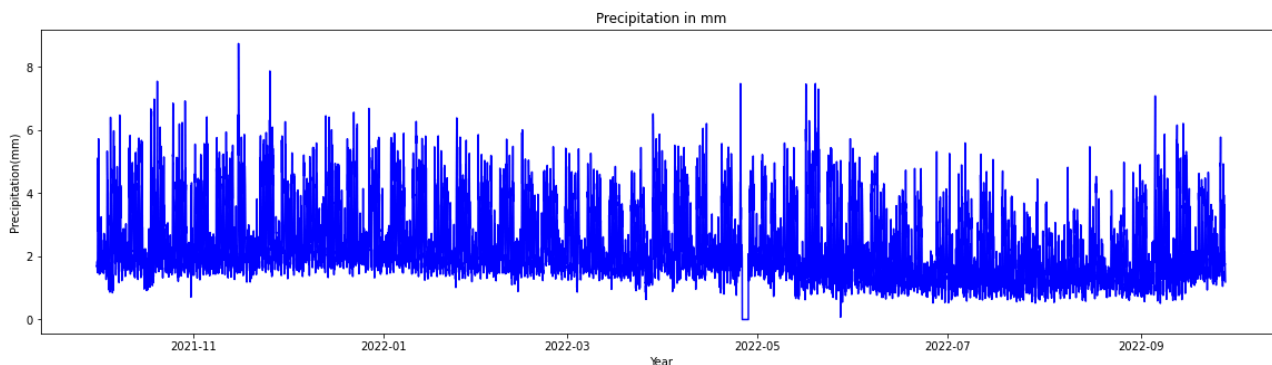
Date	Name of Holiday
01/01/2021	Jaunais gads
02/04/2021	Lielā Piektdiena
04/04/2021	Liieldienas
05/04/2021	Otrās Liieldienas
01/05/2021	Darba svētki
04/05/2021	Latvijas Republikas Neatkarības atjaunošanas diena
23/06/2021	Līgo diena
24/06/2021	Jāņu dienu
18/11/2021	Latvijas Republikas proklamēšanas diena
24/12/2021	Ziemassvētku vakars
25/12/2021	Ziemassvētki
26/12/2021	Otrie Ziemassvētki
31/12/2021	Vecgada vakars
01/01/2022	Jaunais gads
15/04/2022	Lielā Piektdiena
17/04/2022	Liieldienas
18/04/2022	Otrās Liieldienas
01/05/2022	Darba svētki
04/05/2022	Latvijas Republikas Neatkarības atjaunošanas diena
23/06/2022	Līgo diena
24/06/2022	Jāņu dienu
18/11/2022	Latvijas Republikas proklamēšanas diena
24/12/2022	Ziemassvētku vakars
25/12/2022	Ziemassvētki
26/12/2022	Otrie Ziemassvētki
31/12/2022	Vecgada vakars

Πίνακας 5.2: Επίσημες Αργίες στη Λετονία

Έτσι παρατηρούνται οι μέρες οι οποίες έχουν περαστεί ως αργίες στα σύνολα δεδομένων ώστε να δίνουν περαιτέρω πληροφορίες στα μοντέλα. Τώρα θα ξεκινήσει η ανάλυση για το κάθε σύνολο δεδομένων ξεχωριστά.

5.2 Παρουσίαση Αποτελεσμάτων για το πρώτο σύνολο δεδομένων

Αρχικά το πρώτο σύνολο το οποίο χρησιμοποιήθηκε αποτελεί γραφείο αφιερωμένο σε νηπιαγωγείο. Είναι στην περιοχή της Λετονίας που αναγράφεται και στο όνομα του στον παραπάνω πίνακα. Θα παρουσιαστεί το διάγραμμα της κατανάλωσης σε σχέση με το χρόνο και αμέσως μετά κάποια σημαντικά στοιχεία για τα διαφορά χαρακτηριστικά και τις τιμές τους.



Διάγραμμα 5.1 : Διάγραμμα της κατανάλωσης ανά ώρα του πρώτου συνόλου δεδομένων

Όπως φαίνεται έχει μια σταθερή διακύμανση η οποία φαίνεται πως επαναλαμβάνεται σε σχετικά όμοια διαστήματα. Υπάρχουν κάποιες ακραίες τιμές όπως είναι λογικό να συμβαίνει όταν το αποτέλεσμα εξαρτάται από τόσο πολλούς και ευμετάβλητους παράγοντες.

Ακολουθούν πίνακες και εικόνες που περιέχουν σημαντικά στοιχεία για το σύνολο δεδομένων, στην πρώτη παρουσιάζεται λεπτομερώς η μορφή των χαρακτηριστικών και στις υπόλοιπες οι συσχετίσεις μεταξύ τους.

	electricity	meantemp	dayofyear	temp	teams	teams2	hour
Count	8719	8719	8719	8719	8719	8719	8719
Mean	2,385	6,427	182	6,427	0,748	7,872	11,49
Std	1,243	9,302	105,48	9,3	0,433	7,286	6,92
Min	0	-26,0	1	-26	0	0	0
25%	1,516	0	90	0	0	0	5
50%	1,966	5,8	181	5,8	1	7,05	11
75%	3,156	12,9	275	12,9	1	12,854	17
Max	8,742	30,5	365	30,5	1	38,854	23

Πίνακας 5.3: Λεπτομέρειες και στοιχεία για τα χαρακτηριστικά του πρώτου συνόλου δεδομένων

	weekofyear	Quarter	Day	Month	Year	Dayofweek	Weekend
Count	8719	8719	8719	8719	8719	8719	8719
Mean	26,394	2,501	15,657	6,495	0,746	2,993	0,286
Std	15,02	1,121	8,76	3,455	0,435	2	0,451
Min	1	1	1	1	0	0	0
25%	13	1	8	3	0	1	0
50%	26	2	16	6	1	3	0
75%	40	4	23	10	1	5	1
Max	52	4	31	12	1	6	1

	sindayofweek	cosdayofweek	sinmonth	cosmonth	sinhour	coshour
Count	8719	8719	8719	8719	8719	8719
Mean	-0,002086	0,14	5,74E-03	-2,16E-03	0	0,0424
Std	0,654078	0,741	7,036E-01	7,10E-01	0,691	0,72
Min	-0,866025	-1	-1	-1	-0,997	-0,99
25%	-0,866025	-0,5	-0,5	-0,8	-0,63	-0,57
50%	0,000000	1	0	0	0	0,203
75%	0,866025	1	0,5	0,8	0,63	0,854
Max	0,866025	1	1	1	0,997	1

	cosyear	sinyear	cosweekyear	Holiday	sinweekyear
Count	8719	8719	8719	8719	8719
Mean	0	0,999	0,002	0,035	0,008
Std	0,001	0	0,7	0,185	0,7
Min	-0,003	0,999	-1	0	-1
25%	-0,003	0,999	-0,74	0	-0,66
50%	0	1	0	0	0
75%	0	1	0,74	0	0,745
Max	0	1	1	1	1

	dayofweek	weekend	sindayofweek	cosdayofweek	sinmonth
temp	0.01895	0.13326	-4.44794e-02	0.01908	-0.434575
electricity	-0.273452	-0.306577	-1.451927e-01	-0.129845	0.805557
meantemp	0.038095	0.013020	-4.44794e-02	0.019090	-0.434575
dayofyear	-0.000191	-0.010325	-5.252682e-04	-0.012318	-0.758105
teams	-0.000868	0.003993	-4.299015e-03	0.017735	0.331539
teams2	-0.041461	-0.018732	5.604225e-02	0.022690	0.422989
hour	0.000245	0.001168	-7.687540e-04	0.001092	-0.000416
weekofyear	0.004492	0.004671	-2.272040e-03	0.003246	-0.745440
quarter	0.009578	-0.000405	-1.062360e-02	-0.007097	-0.788036
day	-0.004060	0.010533	4.277045e-03	0.024796	0.010911
month	0.000112	-0.011293	-0.170548e-04	-0.014465	-0.769948
year	-0.004205	0.000904	5.473183e-03	0.010721	0.381271
dayofweek	1.000000	0.798442	-5.458651e-01	-0.006265	-0.009476
weekend	0.798442	1.000000	-4.173252e-01	0.010721	-0.000698
sindayofweek	-0.565065	-0.417325	1.000000e+00	0.002999	0.008515
cosdayofweek	-0.006265	0.512552	2.998740e-03	1.000000	-0.007389
sinmonth	-0.009476	-0.000698	0.515062e-03	-0.007389	1.000000
cosmonth	-0.001494	-0.001204	-5.321100e-03	0.000380	-0.002458
sinhour	-0.000220	-0.000232	-4.693980e-07	-0.000169	-0.000208
coshour	-0.000627	-0.000878	6.894427e-04	-0.000837	0.000380
sinyear	-0.004295	0.004904	5.473183e-03	0.010721	0.381271
cosyear	-0.004295	0.004904	5.473183e-03	0.010721	0.381271
sinweekyear	-0.000454	-0.000221	1.000000e+00	-0.000555	0.039168
cosweekyear	-0.000221	-0.000454	1.000000e+00	0.000555	0.039168
Holiday	0.096593	0.041967	-0.749482e-02	-0.028878	0.074202

	cosmonth	sinhour	coshour	sinyear	cosyear
temp	-7.277299e-01	-1.898732e-01	-0.094845	0.352571	0.352571
electricity	2.079867e-01	-1.337370e-01	-0.615436	-0.196689	-0.196689
meantemp	-7.277299e-01	-1.898732e-01	-0.094845	0.352571	0.352571
dayofyear	2.134230e-01	1.339663e-04	0.000021	-0.759734	-0.759734
teams	5.938505e-01	1.476481e-01	0.074121	-0.312365	-0.312365
teams2	7.205265e-01	1.448308e-01	0.067627	-0.336252	-0.336252
hour	-5.842650e-01	-7.265920e-01	-0.019151	0.000068	0.000068
weekofyear	2.347586e-01	1.410640e-04	0.000011	-0.746895	-0.746895
quarter	2.150225e-01	1.969040e-04	-0.000104	-0.778474	-0.778474
day	-4.540335e-03	-1.905036e-04	0.002050	-0.013512	-0.013512
month	2.100010e-01	1.489821e-04	-0.000152	-0.759652	-0.759652
year	-6.479825e-01	-8.576326e-05	-0.000029	1.000000	1.000000
dayofweek	-1.490450e-03	-2.204222e-04	-0.000027	-0.004295	-0.004295
weekend	-1.203933e-03	-2.324673e-04	-0.000078	0.004904	0.004904
sindayofweek	-5.321000e-03	-4.693980e-07	0.000069	0.005473	0.005473
cosdayofweek	3.060354e-04	-1.687945e-04	-0.000037	0.010721	0.010721
sinmonth	-2.458432e-03	-2.079716e-04	0.000380	0.381271	0.381271
cosmonth	1.000000e+00	-4.494761e-07	0.000404	-0.647083	-0.647083
sinhour	-4.494761e-07	1.000000e+00	-0.000085	-0.000086	-0.000086
coshour	4.040166e-04	-8.526970e-05	1.000000	-0.000029	-0.000029
sinyear	-6.479825e-01	-8.576326e-05	-0.000029	1.000000	1.000000
cosyear	-6.479825e-01	-8.576326e-05	-0.000029	1.000000	1.000000
sinweekyear	-2.426996e-01	-2.051891e-04	0.000409	0.521181	0.521181
cosweekyear	9.531493e-01	-2.499680e-05	0.000530	-0.531118	-0.531118
Holiday	1.101643e-02	2.835431e-05	-0.000222	-0.058080	-0.058080

	temp	electricity	meantemp	dayofyear	teams	teams2
temp	1.000000	-0.192681	1.000000	0.101164	-0.758618	-0.951066
electricity	-0.192681	1.000000	0.029938	0.101164	0.104536	0.104536
meantemp	1.000000	-0.192681	1.000000	0.101164	-0.758618	-0.951066
dayofyear	0.101164	0.029938	0.101164	1.000000	-0.074822	-0.096172
teams	-0.758618	0.104536	-0.758618	-0.074822	1.000000	0.626550
teams2	-0.951066	0.104536	-0.951066	-0.096172	0.626550	1.000000
hour	0.137803	0.116140	0.137803	0.000654	-0.112020	-0.104433
weekofyear	0.001997	0.039996	0.001997	0.967516	-0.065291	-0.072713
quarter	0.121214	0.038586	0.121214	0.968667	-0.004843	-0.122862
day	0.064254	-0.032593	0.064254	0.002163	-0.036753	-0.051658
month	0.098693	0.031807	0.098693	0.996545	-0.074074	-0.094785
year	0.352571	-0.196689	0.352571	-0.759734	-0.312365	-0.336252
dayofweek	0.038095	-0.273452	0.038095	-0.000191	-0.000868	-0.041461
weekend	0.038095	-0.273452	0.038095	-0.000191	-0.000868	-0.041461
sindayofweek	-0.044479	0.145193	-0.044479	-0.000525	-0.004299	0.056402
cosdayofweek	-0.019900	-0.129845	-0.019900	0.012318	0.017735	0.022690
sinmonth	-0.434575	0.085557	-0.434575	-0.758618	0.331539	0.422989
cosmonth	-0.727730	0.267997	-0.727730	0.213423	0.593851	0.720526
sinhour	-0.189873	-0.133737	-0.189873	0.000134	0.147646	0.144831
coshour	-0.094845	-0.615436	-0.094845	0.000021	0.074121	0.067627
sinyear	-0.352571	-0.196689	0.352571	-0.759734	-0.312365	-0.336252
cosyear	0.352571	-0.196689	0.352571	-0.759734	-0.312365	-0.336252
sinweekyear	-0.247759	0.021970	-0.247759	-0.778122	0.179473	0.234839
cosweekyear	-0.081578	0.285195	-0.081578	0.022663	0.654076	0.611045
Holiday	-0.068567	-0.078696	-0.068567	0.043958	0.054794	0.077915

	hour	weekofyear	quarter	day	month	year
temp	0.137803	0.001897	0.121214	0.064254	0.098693	0.352571
electricity	0.116140	0.039996	0.038586	-0.032593	0.031807	-0.196689
meantemp	0.137803	0.001897	0.121214	0.064254	0.098693	0.352571
dayofyear	-0.000064	0.967516	0.968667	0.002163	0.996545	-0.074074
teams	-0.112020	-0.065291	-0.004843	-0.036753	-0.074074	-0.312365
teams2	-0.104433	-0.073713	-0.122862	-0.051658	-0.094785	-0.336252
hour	1.000000	-0.000654	0.000175	-0.002596	0.000148	0.000068
weekofyear	-0.000064	1.000000	0.942925	0.051662	0.966347	-0.746895
quarter	0.000175	0.942925	1.000000	0.003447	0.971571	-0.778474
day	-0.002596	0.051662	0.003447	1.000000	-0.000638	-0.013512
month	0.000148	0.966347	0.971571	-0.000638	1.000000	-0.759652
year	0.000068	-0.746895	-0.778474	-0.013512	-0.759652	1.000000
dayofweek	0.000845	0.004492	0.009378	-0.004060	0.000112	-0.004295
weekend	0.001160	0.004671	-0.000405	0.010533	-0.011203	0.004904
sindayofweek	-0.000761	-0.002272	-0.010624	0.004277	-0.000817	0.005473
cosdayofweek	0.001092	0.003246	-0.007897	0.024796	-0.014465	0.010721
sinmonth	-0.000416	-0.745440	-0.788036	0.010911	-0.760948	0.381271
cosmonth	-0.000504	0.247759	0.215923	-0.004540	0.210001	-0.647083
sinhour	-0.725993	0.000141	0.000197	-0.000191	0.000149	-0.000868
coshour	-0.001951	0.000011	-0.000184	0.002050	-0.000152	-0.000029
sinyear	0.000068	-0.746895	-0.778474	-0.013512	-0.759652	1.000000
cosyear	0.000068	-0.746895	-0.778474	-0.013512	-0.759652	1.000000
sinweekyear	-0.000454	-0.778594	-0.806667	0.010998	-0.779988	0.521181
cosweekyear	-0.000655	0.047456	0.017839	-0.005533	0.019027	-0.531118
Holiday	0.000269	0.089651	0.032799	0.039668	0.040982	-0.058080

	sinweekyear	cosweekyear	Holiday
temp	-0.247759	-0.817578	-0.068567
electricity	0.021970	0.285195	-0.070696
meantemp	-0.247759	-0.817578	-0.068567
dayofyear	-0.778122	0.022663	0.043958
teams	0.179473	0.654076	0.054794
teams2	0.234839	0.811045	0.077915
hour	-0.000454	-0.000655	0.000269
weekofyear	-0.778594	0.047456	0.089651
quarter	-0.806667	0.017639	0.032799
day	0.010998	-0.005553	0.039668
month	-0.779988	0.019027	0.040982
year	0.521181	-0.531118	-0.058080
dayofweek	-0.007672	-0.000381	0.096593
weekend	-0.007906	-0.000289	-0.041967
sindayofweek	0.004496	0.000370	-0.097495
cosdayofweek	-0.005595	-0.000135	-0.028878
sinmonth	0.958213	0.239623	0.074202
cosmonth	-0.242370	0.958149	0.011016
sinhour	-0.000205	-0.000025	0.000028
coshour	0.000409	0.000530	-0.000222
sinyear	0.521181	-0.531118	-0.058080
cosyear	0.521181	-0.531118	-0.058080
sinweekyear	1.000000	-0.000221	0.077495
cosweekyear	-0.000221	1.000000	0.039168
Holiday	0.077495	0.039168	1.000000

Εικόνα 5.1: Συσχέτιση των χαρακτηριστικών του πρώτου συνόλου δεδομένων μεταξύ τους

Όπως είναι εμφανές πριν την επιλογή των καλύτερων χαρακτηριστικών έχουν εξαχθεί 25 χαρακτηριστικά για το παρόν σύνολο δεδομένων. Σε αυτά εκτός από τα προφανή όπως μέρα, ώρα και μήνας υπάρχουν και ποιο σύνθετα όπως τα Σαββατοκύριακα ή και οι τριγωνομετρικοί μετασχηματισμοί του συνόλου των ημερών ενός έτους. Στο πρώτο πίνακα φαίνεται ότι για όλα τα χαρακτηριστικά υπάρχει το ίδιο πλήθος δίνοντας έτσι τη δυνατότητα να αξιολογηθούν δίκαια το ένα ως προς το άλλο. Ακόμα διακρίνονται σημαντικά στοιχεία για αυτά όπως ο μέσος όρος των τιμών τους, η διακύμανση τους, η ελάχιστη και η μέγιστη τιμή μαζί με τις ανά 25% τιμές τους. Έτσι μπορεί να δει κανείς το εύρος των τιμών καθώς και τη διασπορά λαμβάνοντας μια πρώτη ιδέα για τον τρόπο με τον οποίο πρέπει να αξιοποιηθεί το κάθε ένα. Τα περισσότερα στοιχεία δείχνουν μικρή διακύμανση πράγμα που σημαίνει ότι οι τιμές τους είναι περιορισμένες σε σχετικά σφιχτά όρια και άρα είναι πιθανό είτε να

επαναλαμβάνονται είτε να δημιουργούνται σχέσεις οι οποίες τα οδηγούν στην εν λόγω συμπεριφορά. Από τις συσχετίσεις προκύπτουν πολύτιμες πληροφορίες για την επιλογή των κατάλληλων χαρακτηριστικών για την εκπαίδευση του μοντέλου. Αρχικά τη μεγαλύτερη συσχέτιση με το στόχο δηλαδή την ενέργεια που καταναλώνεται την έχουν οι διάφοροι συνημιτονοειδής μετασχηματισμοί των ημερών, μηνών, ωρών κτλ. Ακολουθούν τα Σαββατοκύριακα ενώ οι διακοπές έχουν μικρή συσχέτιση με το αποτέλεσμα. Μπορεί σε πρώτη μάλιστα να γεννά την εντύπωση ότι τα καλύτερα αποτελέσματα θα προκύπτουν από εκτενή χρήση τριγωνομετρικών μετασχηματισμών όμως κάτι τέτοιο στην πράξη δεν εξακριβώνεται και μετά από δοκιμές προκύπτει ότι μόνο το coshour μπορεί πραγματικά να ωφελήσει. Αυτό συμβαίνει καθώς ενώ κάποια χαρακτηριστικά έχουν μικρή συσχέτιση με το στόχο εντούτοις παίζουν σημαντικό ρόλο στην εύρεση επαναληπτικών σχέσεων μεταξύ του στόχου και αυτών. Για παράδειγμα όταν είναι περίοδος διακοπών είναι λογικό να υπάρχει μικρή έως καθόλου κατανάλωση σε ένα εμπορικό κτίριο το οποίο δε λειτουργεί τις αργίες.

Μετά από αυτά τα στοιχεία που παρατίθενται από πάνω γίνεται κατανοητή η δομή και η διακύμανση των τιμών στο εν λόγω σύνολο δεδομένων. Οι προσεγγίσεις για τα αποτελέσματα ήταν δυο με τα σύνολα εξέτασης να προκύπτουν είτε μετά από τυχαία επιλογή τιμών του συνόλου δεδομένων είτε με επιλογή βασισμένη στη χρονική σειρά. Η δεύτερη είναι και αυτή που θα εξερευνηθεί πρώτη.

“Σειριακή” προσέγγιση
Ωριαία δειγματοληψία.

Στη συγκεκριμένη περίπτωση τροφοδοτούνται στα μοντέλα τα στοιχεία ακριβώς όπως είναι στα σύνολα δεδομένων δηλαδή ανά ώρα. Αυτή είναι η περίπτωση με τις περισσότερες τιμές και άρα η πιο χρονοβόρα και με μεγαλύτερο περιθώριο λάθους αφού οι διακυμάνσεις των τιμών είναι αρκετά μεγάλες. Μέσα από αυτή τη περίπτωση προέκυψαν τα εξής αποτελέσματα ως προς το σφάλμα στον πίνακα που ακολουθεί:

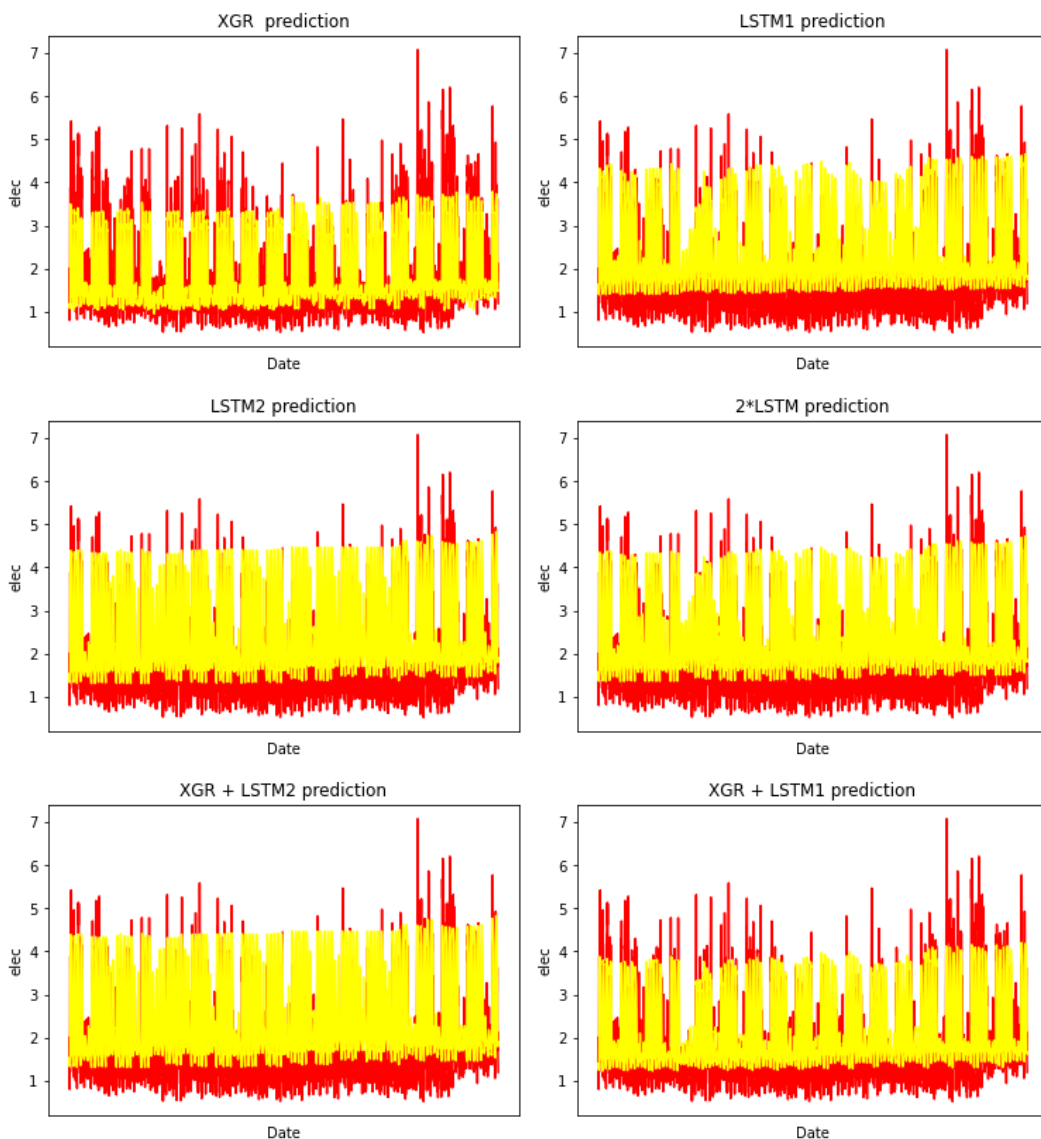
Models	RMSE(%)	NMBE(%)
XGBOOST	30,960	-0,555
LSTM1	41,133	22,699
LSTM2	36,081	6,185
LSTM1 + LSTM2	36,906	14,441
LSTM1 + XGBOOST	33,643	11,071
LSTM2 + XGBOOST	34,995	10,839
LSTM1 + LSTM2 + XGBOOST	33,522	9,443

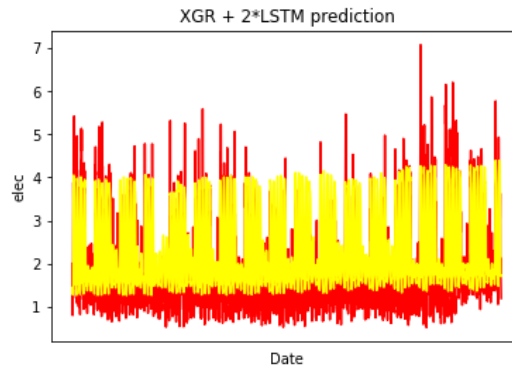
Πίνακας 5.4 : Τιμές Σφαλμάτων του πρώτου συνόλου δεδομένων με ωριαία δειγματοληψία

Παρατηρείται λοιπόν ότι το NMBE που είναι δείκτης μέτρησης της κατεύθυνσης του λάθους τείνει στο 0 στην περίπτωση του XGBOOST σε αντίθεση με τα LSTM και αυτό επιβεβαιώνει την αρχική μας αναφορά ότι το πρώτο μοντέλο έχει μικρότερη

διακύμανση στις προβλέψεις του. Αυτό που φαίνεται είναι ότι με τις προβλέψεις να βρίσκονται κοντά και γύρω από ένα μέσο όρο το λάθος με θετικό πρόσημο αναιρεί το λάθος με αρνητικό και γι' αυτό φτάνει τόσο κοντά στο 0 το σφάλμα. Από όλα έχει μικρότερο RMSE σφάλμα το πρώτο μοντέλο όμως ένας λόγος που το πετυχαίνει αυτό είναι ακριβώς επειδή δεν έχει προβλέψεις που προσπαθούν να πετύχουν το αποτέλεσμα και στις ακραίες συνθήκες και έτσι η συνάρτηση σφάλματος που αντιμετωπίζει το μεγάλο λάθος σε τιμή με μεγαλύτερη βαρύτητα να το θεωρεί καλύτερο πράγμα που στα διαγράμματα που ακολουθούν φαίνεται ότι δεν ανταποκρίνεται πλήρως στην πραγματικότητα.

Διαγράμματα Αποτελεσμάτων

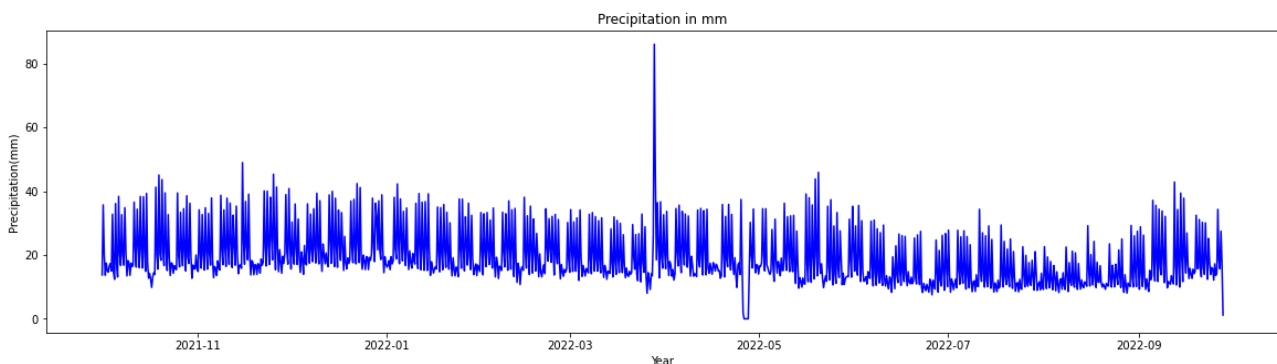




Διαγράμματα αποτελεσμάτων 5.1: Προβλέψεις και πραγματικές τιμές της κατανάλωσης για τα διάφορα μοντέλα του πρώτου συνόλου δεδομένων ανά ώρα

Έτσι από τα διαγράμματα δεν είναι εμφανής μια καλύτερη απόδοση από πλευράς XGBoost απεναντίας το δεύτερο LSTM δείχνει να πετυχαίνει τις καλύτερες προβλέψεις σε σχέση με τα τρία απλά μοντέλα. Από τα συνδυαστικά με πολύ μικρή διαφορά είναι καλύτερο αυτό που συνδυάζονται και τα τρία.

Ανά 8 ώρες δειγματοληψία.



Διάγραμμα 5.2 : Διάγραμμα αθροίσματος της κατανάλωσης ανά 8 ώρες του πρώτου συνόλου δεδομένων

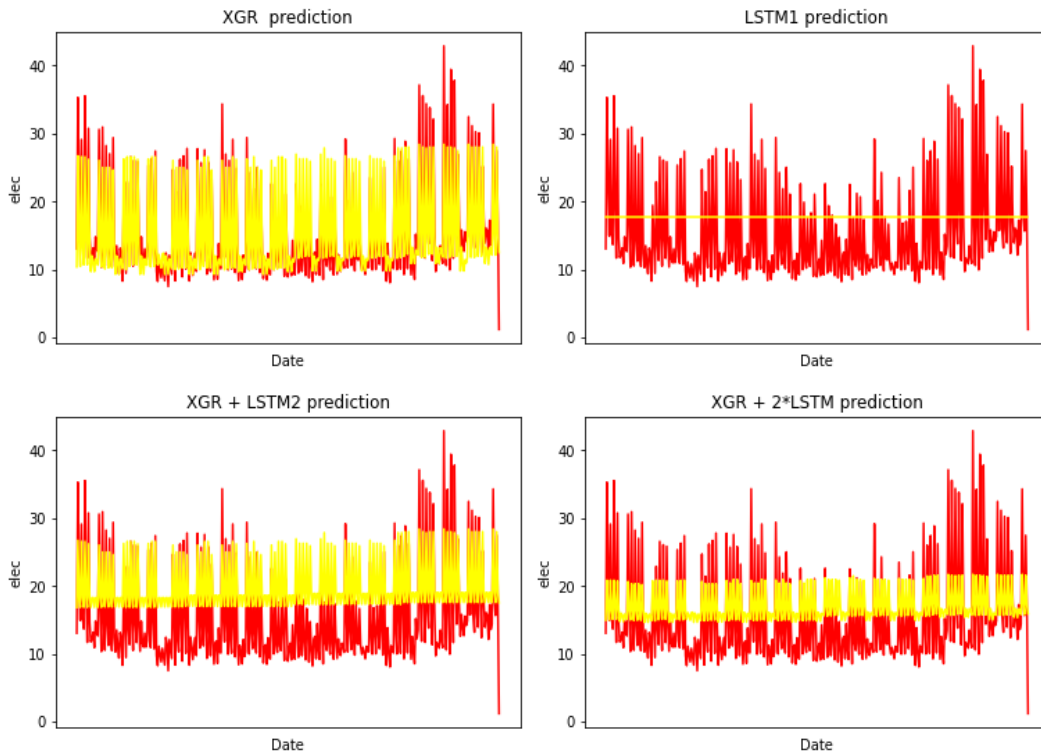
Εύκολα φαίνεται ότι η εικόνα αλλάζει με βάση την δειγματοληψία ανά 8 ώρες και αυτό συμβαίνει γιατί αρκετές ακραίες τιμές μαζί πλέον δημιουργούν εντονότερες διαφορές σε σχέση με τη βάση του συνόλου δεδομένων. Αυτό συμβαίνει καθώς γίνεται αρχικά η επιλογή να αθροίσουμε την ενέργεια που καταναλώνεται σε ένα οκτάωρο. Παρόλα αυτά η μέση διασπορά μειώνεται πράγμα λογικό αφού το πλήθος των τιμών μειώνεται εξίσου. Ακολουθεί ο αντίστοιχος πίνακας αποτελεσμάτων για την περίπτωση που με τη νέα δειγματοληψία στόχος είναι εύρεση του αθροίσματος των τιμών της κατανάλωσης κατά το διάστημα της δειγματοληψίας:

Models	RMSE(%)	NMBE(%)
XGBOOST	20,437	0,467
LSTM1	48,292	15,497

LSTM2	42,896	25,246
LSTM1 + LSTM2	38,8	20,372
LSTM1 + XGBOOST	29,887	7,982
LSTM2 + XGBOOST	29,805	13,737
LSTM1 + LSTM2 + XGBOOST	42,863	25,489

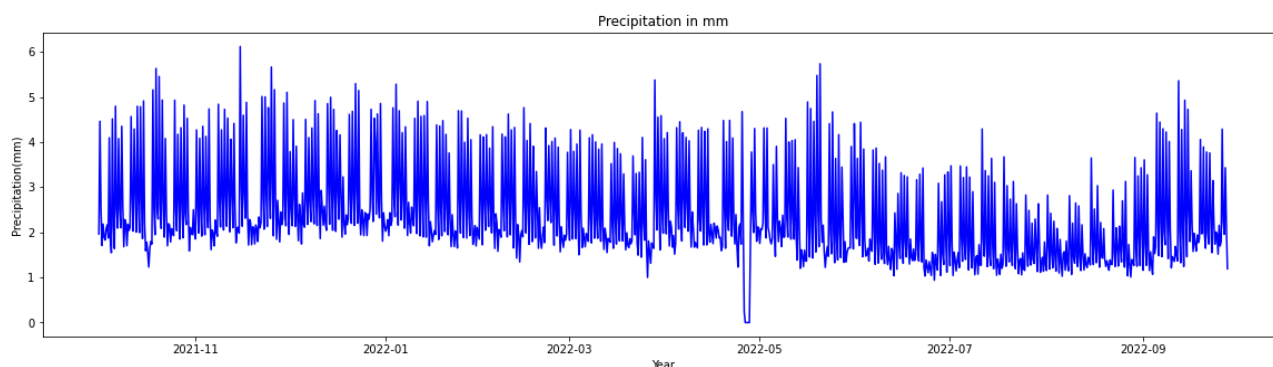
Πίνακας 5.5 : Τιμές Σφαλμάτων του πρώτου συνόλου δεδομένων με ανά 8 ώρες δειγματοληψία για εύρεση του αθροίσματος της κατανάλωσης στο διάστημα αυτό

Φαίνεται ότι τα λιγότερα δεδομένα επηρεάζουν αρνητικά τα LSTM τα οποία έχουν χειρότερα αποτελέσματα αφού δεν μπορούν να εκπαιδευτούν με πολλές τιμές. Το XGBOOST από την άλλη έχει αρκετά καλύτερα αποτελέσματα (10%) αποδεικνύοντας αυτό που ήδη ήταν γνωστό ότι δηλαδή οι αλγόριθμοι μηχανικής μάθησης δεν έχουν πρόβλημα με τα λίγα δεδομένα σε αντίθεση με τη βαθιά μάθηση. Μερικά από τα διαγράμματα που προέκυψαν παρατίθενται στη συνέχεια.



Διαγράμματα αποτελεσμάτων 5.2 : Προβλέψεις και πραγματικές τιμές του αθροίσματος της κατανάλωσης για τα διάφορα μοντέλα του πρώτου συνόλου δεδομένων ανά 8 ώρες

Αν αντί για άθροισμα της κατανάλωσης εντός των 8 ωρών υπολογιστεί ο μέσος όρος της τότε το διάγραμμα του συνόλου δεδομένων δείχνει έτσι.



Διάγραμμα 5.3 : Διάγραμμα μέσου όρου της κατανάλωσης ανά 8 ώρες του πρώτου συνόλου δεδομένων

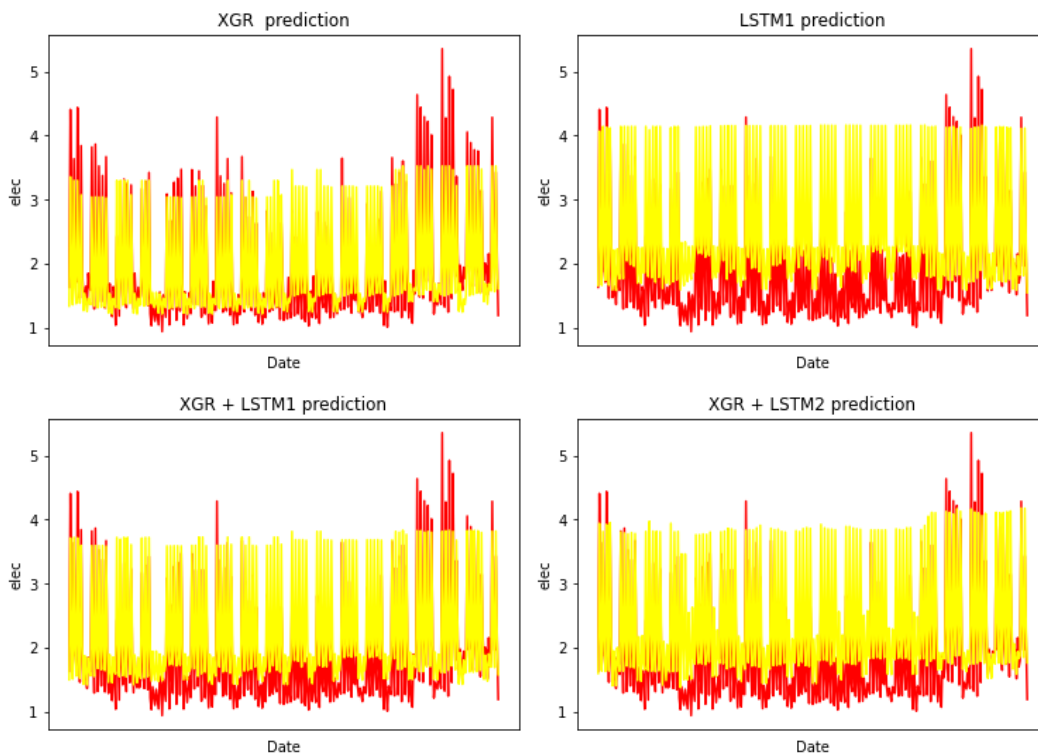
Η διασπορά είναι μεγαλύτερη καθώς πλέον δεν αθροίζεται όπως πριν η ενέργεια αλλά υπολογίζεται ο μέσος όρος ο οποίος μπορεί να απέχει αρκετά από δείγμα σε δείγμα.

Ακολουθεί ο πίνακας όπου αντί για το άθροισμα της κατανάλωσης ανά 8 ώρες γίνεται προσπάθεια εντοπισμού του μέσου όρου της κατανάλωσης σε 8 ώρες.

Models	RMSE(%)	NMBE(%)
XGBOOST	20,029	0,629
LSTM1	50,325	36,227
LSTM2	25,196	3,355
LSTM1 + LSTM2	35,289	19,791
LSTM1 + XGBOOST	30,906	18,428
LSTM2 + XGBOOST	29,805	13,737
LSTM1 + LSTM2 + XGBOOST	42,863	25,489

Πίνακας 5.6: Τιμές Σφαλμάτων του πρώτου συνόλου δεδομένων με ανά 8 ώρες δειγματοληψία για εύρεση του μέσου όρου της κατανάλωσης στο διάστημα αυτό

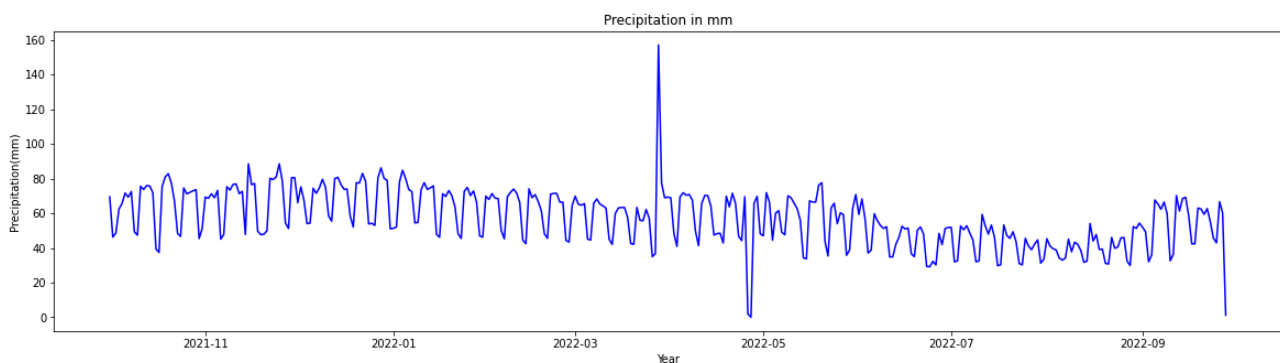
Είναι εμφανής μια βελτίωση στο δεύτερο LSTM δηλαδή σε αυτό που διαθέτει το επίπεδο όπου κάνει μια εκτίμηση της εκπαίδευσης σε ένα άγνωστο για αυτό σετ και μετά συνεχίζει την εκπαίδευση με βάση τη βελτίωση της πρόβλεψης αυτής. Αυτό συμβαίνει καθώς αν και λιγότερα τα νέα δεδομένα και με μεγαλύτερη διασπορά στο σύνολο εκπαίδευσης δεν έχουν τόσο μεγάλη διασπορά στο σύνολο εξέτασης με αποτέλεσμα οι προηγούμενη εκπαίδευση σε συνδυασμό με το μικρό σετ αξιολόγησης να οδηγεί το μοντέλο αυτό σε καλύτερα αποτελέσματα. Το πρώτο LSTM ακριβώς επειδή το σύνολο εξάσκησης έχει μεγαλύτερη διασπορά και μεγαλύτερες τιμές γενικά εκπαιδύεται με αυτό το τρόπο και άρα δε μπορεί να εντοπίσει με ακρίβεια αυτή τη γενικευμένη μείωση των τιμών που παρατηρείται στο σύνολο εξέτασης. Ακολουθούν και τα διαγράμματα που δείχνουν αυτό ακριβώς.



Διαγράμματα αποτελεσμάτων 5.3 : Προβλέψεις και πραγματικές τιμές του μέσου όρου της κατανάλωσης για τα διάφορα μοντέλα του πρώτου συνόλου δεδομένων ανά 8 ώρες

Ημερήσια Δειγματοληψία

Σε αυτό το κομμάτι θα προβούμε αντίστοιχα στον υπολογισμό του μέσου όρου των τιμών για κάθε ημέρα με μοναδική εξαίρεση τον στόχο, δηλαδή την κατανάλωση, της οποίας θα υπολογιστεί εκτός του μέσου όρου ανά ημέρα και το άθροισμα ανά ημέρα. Πρώτα θα είναι η προσπάθεια πρόβλεψης της συνολικής κατανάλωσης. Το διάγραμμα πλέον μοιάζει κάπως έτσι.



Διάγραμμα 5.4 : Διάγραμμα αθροίσματος της κατανάλωσης ανά ημέρα του πρώτου συνόλου δεδομένων

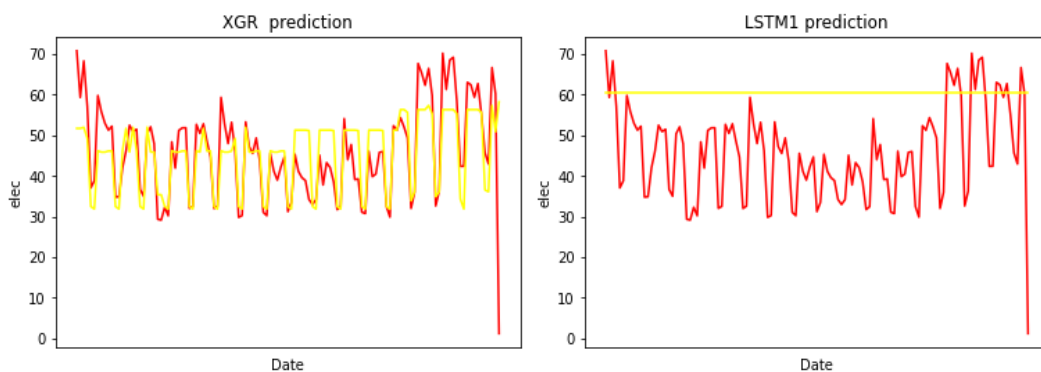
Είναι ξεκάθαρο ότι και πάλι οι φορές που υπήρχαν ακραίες τιμές δημιουργούν μεγαλύτερες ακραίες τιμές αφού αθροίζονται μεταξύ τους σε αντίθεση με τις τιμές που είχαν ομοιόμορφη σχετικά κατανομή και άρα οι θετικές αλληλοεξουδετερώνονται με τις αρνητικές με αποτέλεσμα να υπάρχει και τώρα σε αρκετά σημεία μικρή διασπορά γύρω από ένα μέσο όρο. Επίσης είναι εμφανές ότι ο θόρυβος του αρχικού διαγράμματος έχει πλέον εξαλειφθεί εντελώς με τις διακυμάνσεις να είναι πλέον αρκετά “καθαρές”

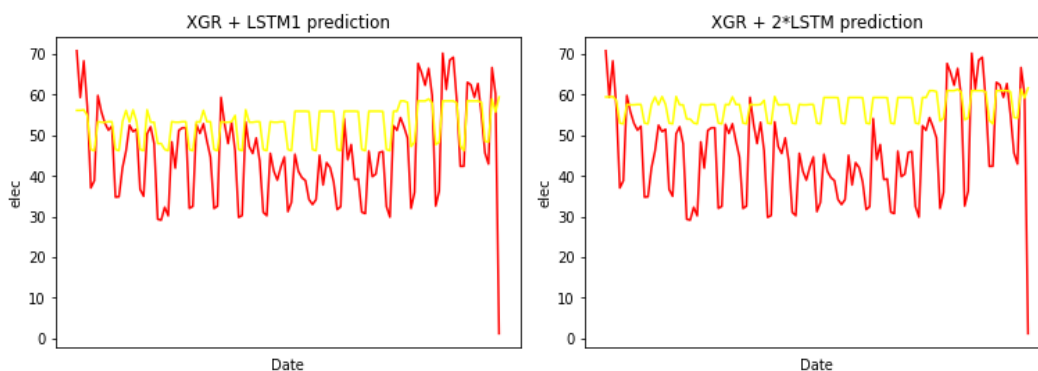
Ακολουθεί ο πίνακας με τα αποτελέσματα των μοντέλων για την παρούσα υποπερίπτωση

Models	RMSE(%)	NMBE(%)
XGBOOST	18,753	-0,996
LSTM1	41,164	32,061
LSTM2	25,83	-0,815
LSTM1 + LSTM2	30,177	15,623
LSTM1 + XGBOOST	25,668	15,532
LSTM2 + XGBOOST	23,282	7,065
LSTM1 + LSTM2 + XGBOOST	24,126	10,083

Πίνακας 5.7: Τιμές Σφαλμάτων του πρώτου συνόλου δεδομένων με ημερήσια δειγματοληψία για εύρεση του αθροίσματος της κατανάλωσης στο διάστημα αυτό

Φαίνεται και πάλι μια μικρή βελτίωση στο πρώτο μοντέλο η οποία ακολουθεί τη βάση όσων ειπώθηκαν προηγουμένως. Βελτίωση έχουν όμως και τα δυο LSTM στην RMSE πράγμα που βασίζεται κυρίως στη μορφή του συνόλου δεδομένων το οποίο δηλαδή πλέον παρουσιάζει αρκετά μικρότερες διακυμάνσεις μιας και οι τιμές προκύπτουν από άθροισμα και μέσους όρους από 24 τιμές. Το πολύ μικρό NMBE λάθος αποδίδεται στη μικρή διακύμανση των τιμών η οποία οδηγεί σε σφάλμα πολύ κοντά στο μέσο όρο και άρα μη αντιληπτό από αυτή τη συνάρτηση σφάλματος. Στην πραγματικότητα όμως δεν υπάρχει βελτίωση απλώς τα μοντέλα βρίσκουν τον καλύτερο μέσο όρο και προβλέπουν την ίδια τιμή για όλα. Αυτό συμβαίνει επειδή τα μοντέλα βαθιάς μάθησης χρειάζονται μεγάλο πλήθος από δεδομένα για να εκπαιδευτούν κατάλληλα πράγμα που όπως αναφέρθηκε και νωρίτερα δε συμβαίνει στην παρούσα κατάσταση. Ακολουθούν τα αντίστοιχα διαγράμματα για κάποια από τα παραπάνω μοντέλα.

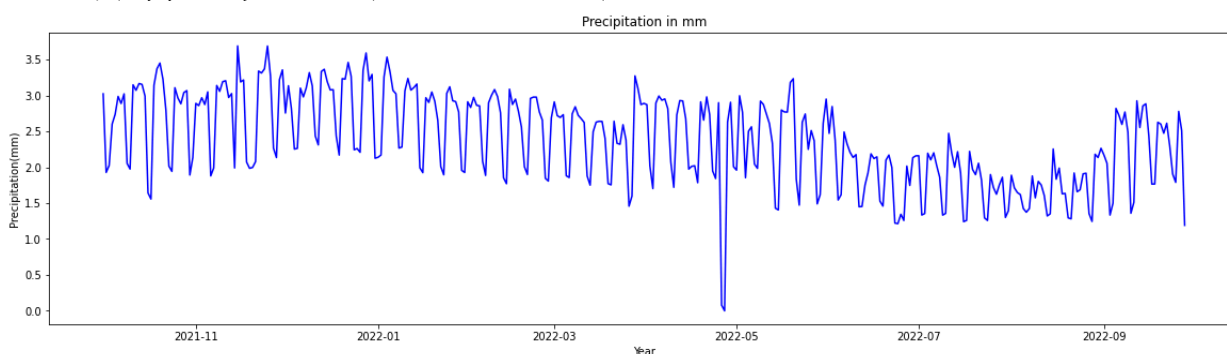




Διαγράμματα αποτελεσμάτων 5.4 : Προβλέψεις και πραγματικές τιμές του αθροίσματος της κατανάλωσης για τα διάφορα μοντέλα του πρώτου συνόλου δεδομένων ανά ημέρα

Αυτό που πάντως φαίνεται επίσης ξεκάθαρα είναι γιατί το δεύτερο έχει μικρότερο σφάλμα από το πρώτο.

Ακολουθεί η περίπτωση πρόβλεψης του μέσου όρου της κατανάλωσης ανά ημέρα με το διάγραμμα της να παρουσιάζεται παρακάτω.



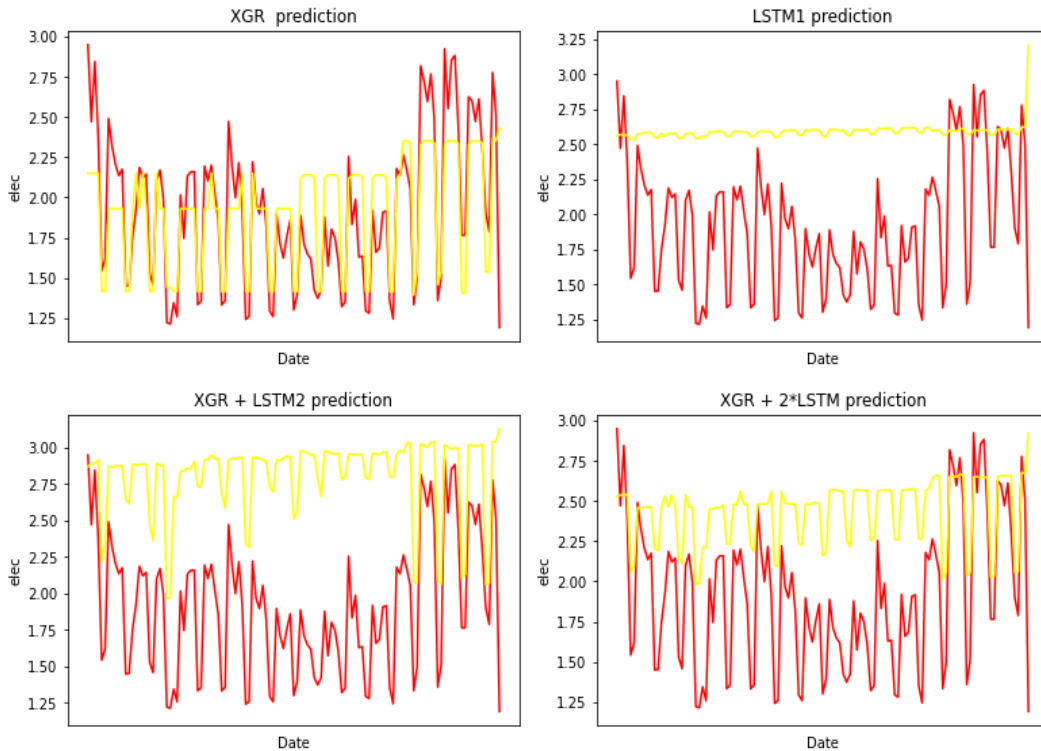
Διάγραμμα 5.5 : Διάγραμμα μέσου όρου της κατανάλωσης ανά ημέρα του πρώτου συνόλου δεδομένων

Και για αυτό ισχύουν τα συμπεράσματα που προέκυψαν από το προηγούμενο αντίστοιχο βήμα δηλαδή ότι παρατηρείται σχετικά μεγαλύτερη διασπορά με λιγότερες ακραίες τιμές. Ακολουθεί ο πίνακας με τα αποτελέσματα των μοντέλων για την παρούσα υποπερίπτωση

Models	RMSE(%)	NMBE(%)
XGBOOST	15,84	-0,245
LSTM1	40,112	31,659
LSTM2	28,109	14,058
LSTM1 + LSTM2	33,461	22,859
LSTM1 + XGBOOST	24,573	15,707
LSTM2 + XGBOOST	27,552	15,539
LSTM1 + LSTM2 + XGBOOST	25,397	15,157

Πίνακας 5.8: Τιμές Σφαλμάτων του πρώτου συνόλου δεδομένων με ημερήσια δειγματοληψία για εύρεση του μέσου όρου της κατανάλωσης στο διάστημα αυτό

Η μικρή βελτίωση που παρουσιάζουν οι τιμές σφάλματος εξηγούνται πλήρως από την παρατήρηση ότι η διακύμανση των τιμών είναι μικρή σε σχέση με το μέσο όρο. Η αύξηση στο δεύτερο LSTM μπορεί να βασίζεται στο γεγονός ότι το σετ αξιολόγησης που χρησιμοποιείται για την εκπαίδευση είναι είτε πολύ μικρό είτε περιέχει κάποια ακραία τιμή που είναι αρκετή να “χαλάσει” την εκπαίδευση. Η υποπερίπτωση ολοκληρώνεται με κάποια διαγράμματα από τα αποτελέσματα.



Διαγράμματα αποτελεσμάτων 5.5: Προβλέψεις και πραγματικές τιμές του μέσου όρου της κατανάλωσης για τα διάφορα μοντέλα του πρώτου συνόλου δεδομένων ανά ημέρα

“Shuffled” προσέγγιση
Ωριαία δειγματοληψία.

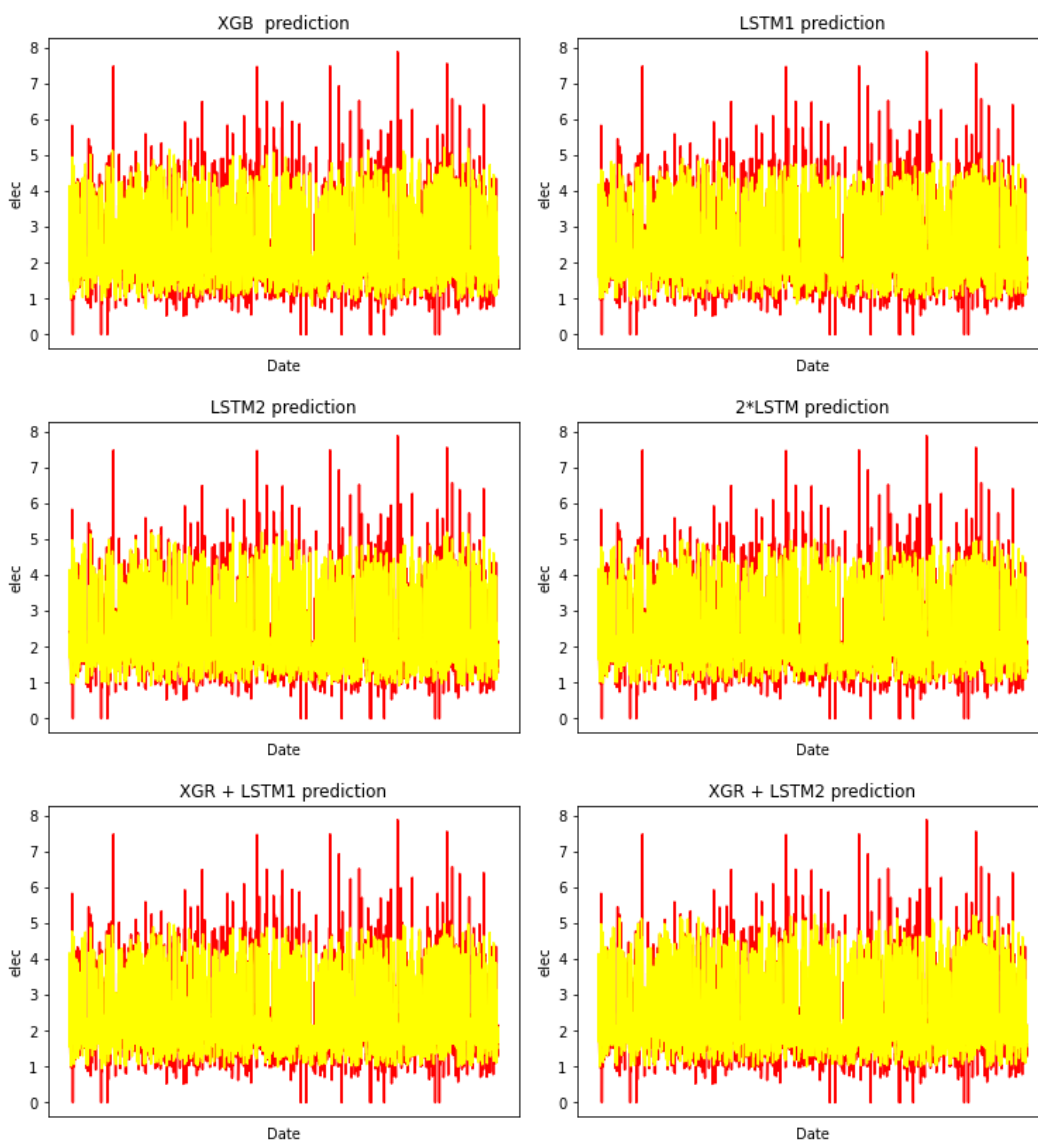
Όλα τα διαγράμματα των συνόλων δεδομένων είναι ίδια με αυτά των αντίστοιχων περιπτώσεων που αναλύθηκαν προηγουμένως και για αυτό δε θα συμπεριληφθούν σε ξανά. Θα παρουσιαστούν και θα σχολιαστούν τα αποτελέσματα και τα διαγράμματα αυτών για την κάθε περίπτωση. Ακολουθεί ο πίνακας των αποτελεσμάτων.

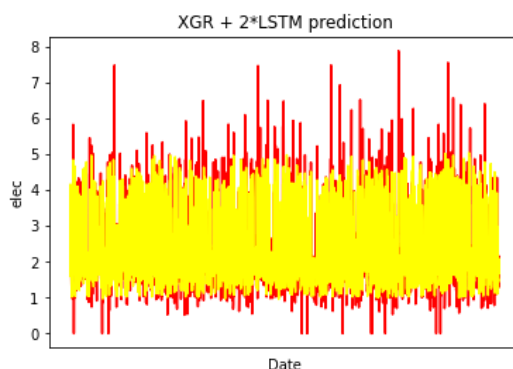
Models	RMSE(%)	NMBE(%)
XGBOOST	22,991	0,011
LSTM1	23,875	-1,203
LSTM2	23,922	-3,038
LSTM1 + LSTM2	23,613	-2,121
LSTM1 + XGBOOST	22,95	-0,596
LSTM2 + XGBOOST	23,220	1,716

LSTM1 + LSTM2 + XGBOOST	23,053	-1,41
-------------------------	--------	-------

Πίνακας 5.9: Τιμές Σφαλμάτων του πρώτου συνόλου δεδομένων με ωριαία δειγματοληψία στην περίπτωση των "Shuffled" δεδομένων

Η πρώτη παρατήρηση είναι ότι τα τρία απλά μοντέλα έχουν πολύ κοντινά αποτελέσματα σε αντίθεση με ότι συνέβαινε στη σειριακή προσέγγιση. Αυτό συμβαίνει επειδή με την τυχαία επιλογή είναι πιο πιθανό να υπάρχει μεγαλύτερο εύρος εκπαίδευσης καθώς και πιο μοιρασμένο σύνολο εξέτασης. Ειδικά στο NMBE είναι εμφανές ότι τέτοιου είδους εκπαίδευση δε δίνει τη δυνατότητα στα μοντέλα να κάνουν λάθος προς μια κατεύθυνση. Θα παρουσιαστούν στη συνέχεια τα διαγράμματα των προβλέψεων των μοντέλων.





Διαγράμματα αποτελεσμάτων 5.6: Προβλέψεις και πραγματικές τιμές της κατανάλωσης για τα διάφορα μοντέλα του πρώτου συνόλου δεδομένων ανά ώρα στην “Shuffled” προσέγγιση

Έτσι όπως φαίνεται και από τα διαγράμματα το καλύτερο μοντέλο είναι ο συνδυασμός του XGBoost με το πρώτο LSTM δηλαδή με το πιο “ελεύθερο” από τα δυο. Στην shuffled προσέγγιση θα χρησιμοποιήσουμε στη συνέχεια scatter plots που αντικατοπτρίζουν την πραγματικότητα των συνόλων εξέτασης και προβλέψεων

Ανά 8 ώρες δειγματοληψία.

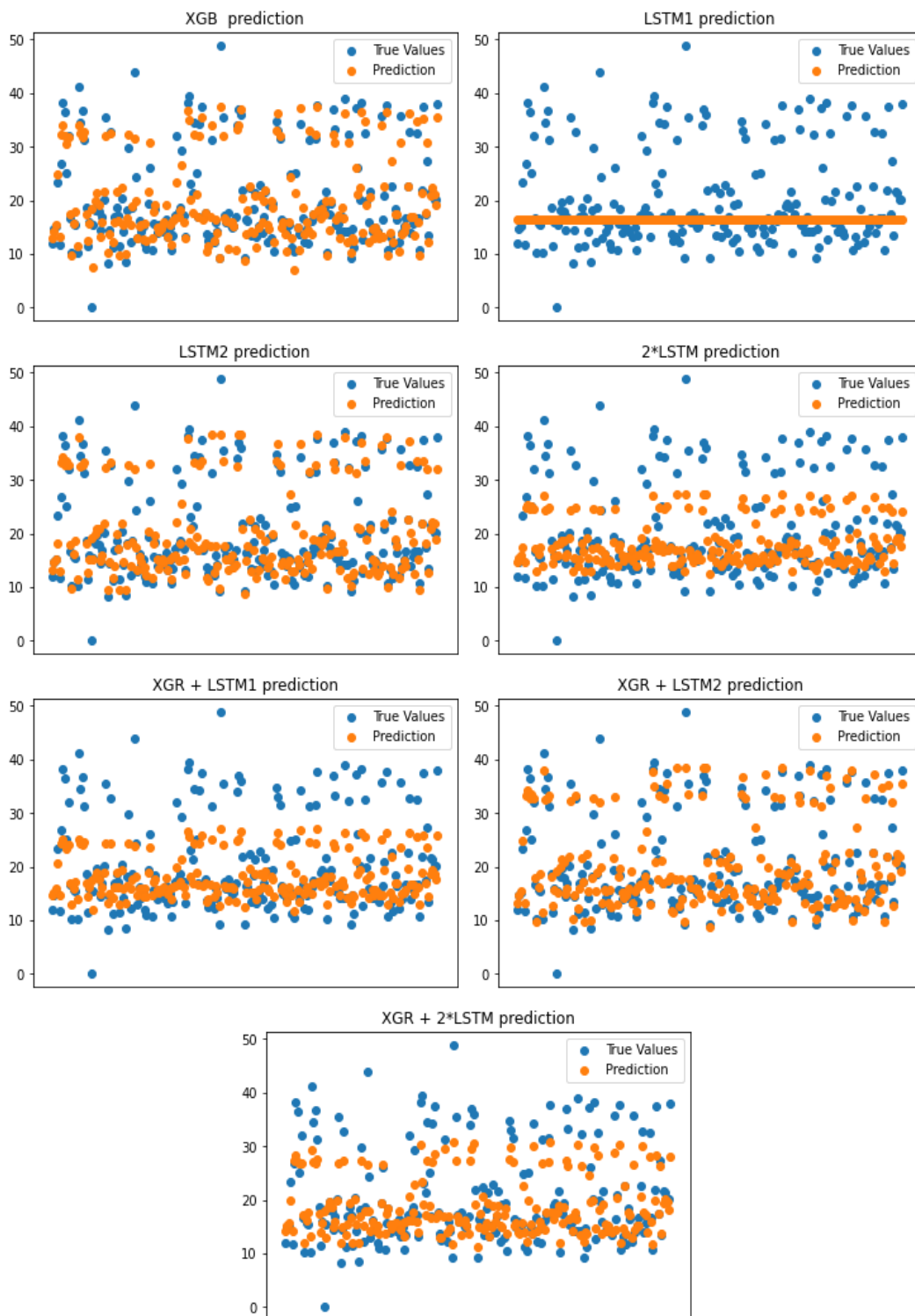
Και πάλι θα παρουσιαστούν οι δυο υποπεριπτώσεις δηλαδή της πρόβλεψης της συνολικής κατανάλωσης ανά 8 ώρες και του μέσου όρου της κατανάλωσης ανά 8 ώρες.

Πρώτα ο πίνακας σφάλματος για το άθροισμα.

Models	RMSE(%)	NMBE(%)
XGBOOST	16,828	-1,58
LSTM1	49,833	-17,438
LSTM2	17,341	-0,853
LSTM1 + LSTM2	29,595	-9,145
LSTM1 + XGBOOST	30,351	-9,509
LSTM2 + XGBOOST	16,956	1,207
LSTM1 + LSTM2 + XGBOOST	24,133	-6,623

Πίνακας 5.10: Πίνακας Σφαλμάτων για την ανά 8 ώρες δειγματοληψία στην περίπτωση του αθροίσματος της κατανάλωσης στο διάστημα αυτό στην περίπτωση των “Shuffled” δεδομένων

Εδώ το δεύτερο LSTM είναι καλύτερο από το πρώτο και αυτό συμβαίνει λόγω του περιορισμού που του βάζει το σετ αξιολόγησης κατά την εκπαίδευση. Ακολουθούν τα αντίστοιχα διαγράμματα.



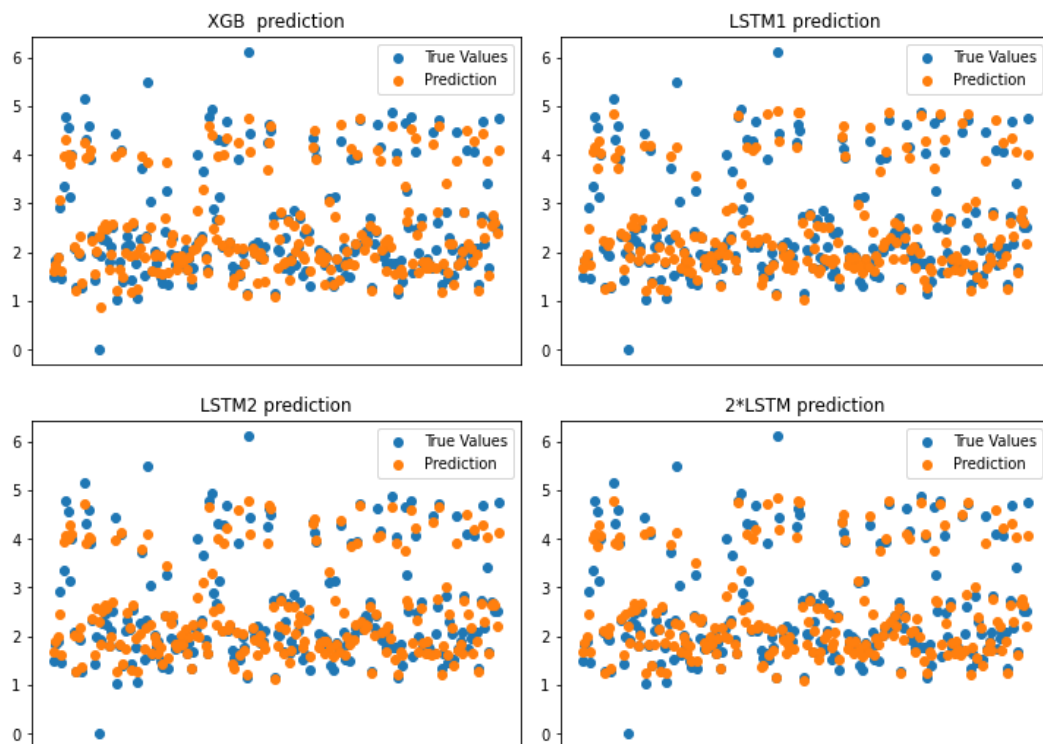
Διαγράμματα αποτελεσμάτων 5.7: Προβλέψεις και πραγματικές τιμές του αθροίσματος της κατανάλωσης για τα διάφορα μοντέλα του πρώτου συνόλου δεδομένων ανά ώρα στην “Shuffled” προσέγγιση

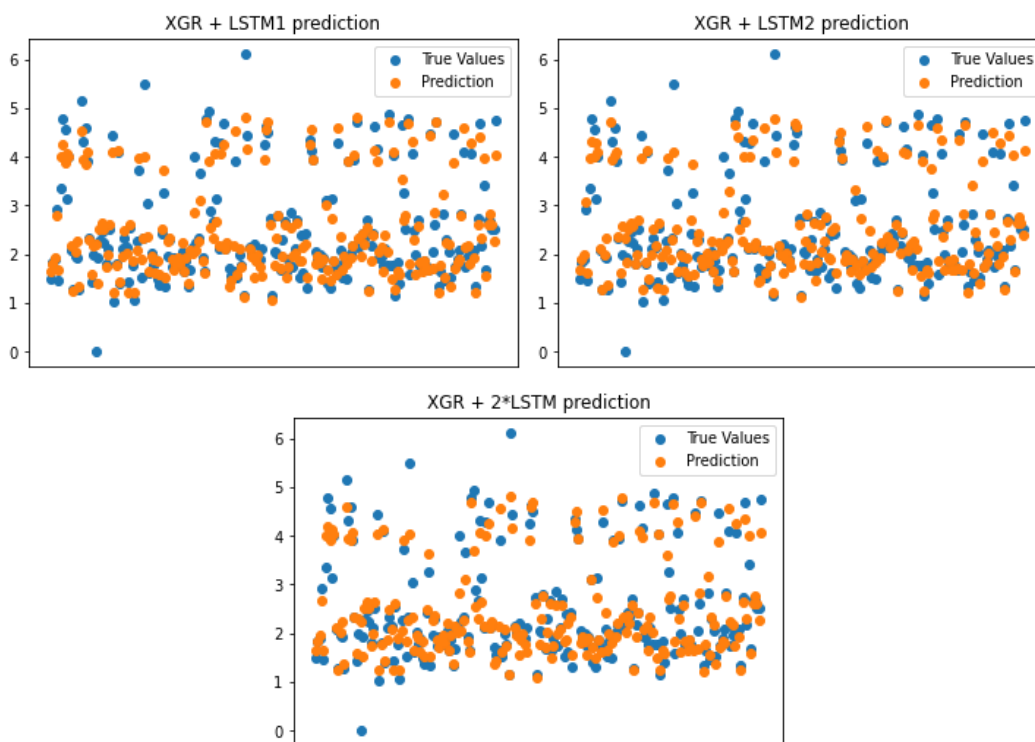
Την ίδια διαδικασία ακολουθούμε και για το μέσο όρο της κατανάλωσης με τον πίνακα σφαλμάτων να ακολουθεί.

Models	RMSE(%)	NMBE(%)
XGBOOST	17,166	-2,387
LSTM1	17,445	-0,878
LSTM2	16,417	0,70
LSTM1 + LSTM2	16.725	-0,086
LSTM1 + XGBOOST	16,855	-1,633
LSTM2 + XGBOOST	16,129	2,124
LSTM1 + LSTM2 + XGBOOST	16,451	-0,853

Πίνακας 5.11 : Πίνακας Σφαλμάτων για την ανά 8 ώρες δειγματοληψία στην περίπτωση της μέσης κατανάλωσης στην περίπτωση των "Shuffled" δεδομένων

Εδώ το καλύτερο όλων είναι το LSTM2 από τα απλά ενώ το καλύτερο συμπεριλαμβανομένων και των συνδυασμών είναι το LSTM2 + XGBOOST. Ακολουθούν τα διαγράμματα των προβλέψεων από όλα τα μοντέλα.





Διαγράμματα αποτελεσμάτων 5.8: Προβλέψεις και πραγματικές τιμές του μέσου όρου της κατανάλωσης για τα διάφορα μοντέλα του πρώτου συνόλου δεδομένων ανά 8 ώρες στην “Shuffled” προσέγγιση

Φαίνεται ξεκάθαρα από τα διαγράμματα ότι τα LSTM καταφέρνουν να πετύχουν αρκετές ακραίες τιμές πράγμα που αποδεικνύει ότι έχουν την δυνατότητα να μην παγιδεύονται γύρω από ένα μέσο όρο αλλά αν έχουν την ελευθερία να εντοπίσουν τιμές που δεν ακολουθούν τη πορεία των περισσοτέρων.

Ημερήσια Δειγματοληψία

Και πάλι θα παρουσιαστούν οι δυο υποπεριπτώσεις δηλαδή της πρόβλεψης της συνολικής κατανάλωσης ανά ημέρα και του μέσου όρου της κατανάλωσης ανά ημέρα.

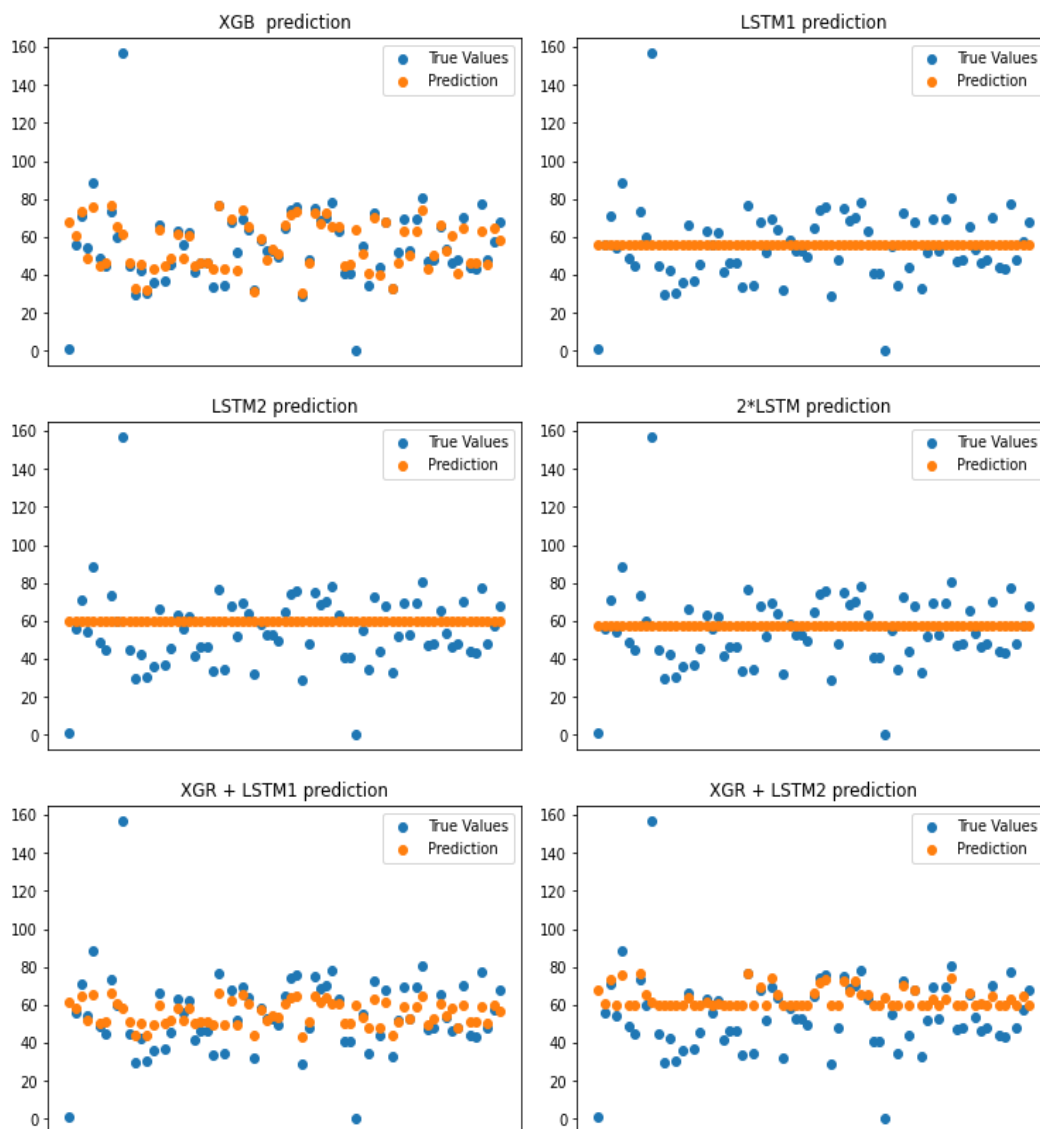
Έτσι προκύπτει ο παρακάτω πίνακας για την περίπτωση της πρόβλεψης της συνολικής κατανάλωσης ανά ημέρα.

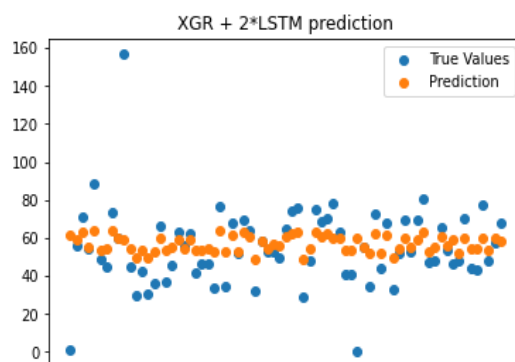
Models	RMSE(%)	NMBE(%)
XGBOOST	40,553	-6,6
LSTM1	43,464	2,962
LSTM2	43,872	-6,668
LSTM1 + LSTM2	43,402	-1,852
LSTM1 + XGBOOST	40,277	-1,819
LSTM2 + XGBOOST	41,12	-3,435

LSTM1 + LSTM2 + XGBOOST	40,782	3,128
-------------------------	--------	-------

Πίνακας 5.12: Πίνακας Σφαλμάτων για την ημερήσια δειγματοληψία στην περίπτωση του αθροίσματος της κατανάλωσης στην περίπτωση των “Shuffled” δεδομένων

Τα αποτελέσματα εδώ είναι χειρότερα και αυτό οφείλεται στο γεγονός ότι τα δείγματα έχουν μειωθεί και δε μπορούν να εντοπίσουν άμεσα τις σχέσεις των τιμών με τα αθροίσματα της κατανάλωσης. Ακολουθούν τα διαγράμματα





Διαγράμματα αποτελεσμάτων 5.9: Προβλέψεις και πραγματικές τιμές του αθροίσματος της κατανάλωσης για τα διάφορα μοντέλα του πρώτου συνόλου δεδομένων ανά ημέρα στην “Shuffled” προσέγγιση

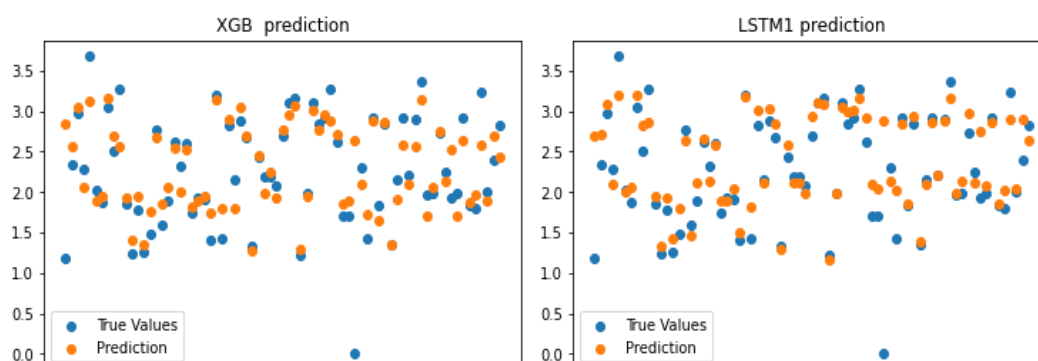
Τώρα ερευνάται η περίπτωση της πρόβλεψης της μέσης ημερήσιας κατανάλωσης.

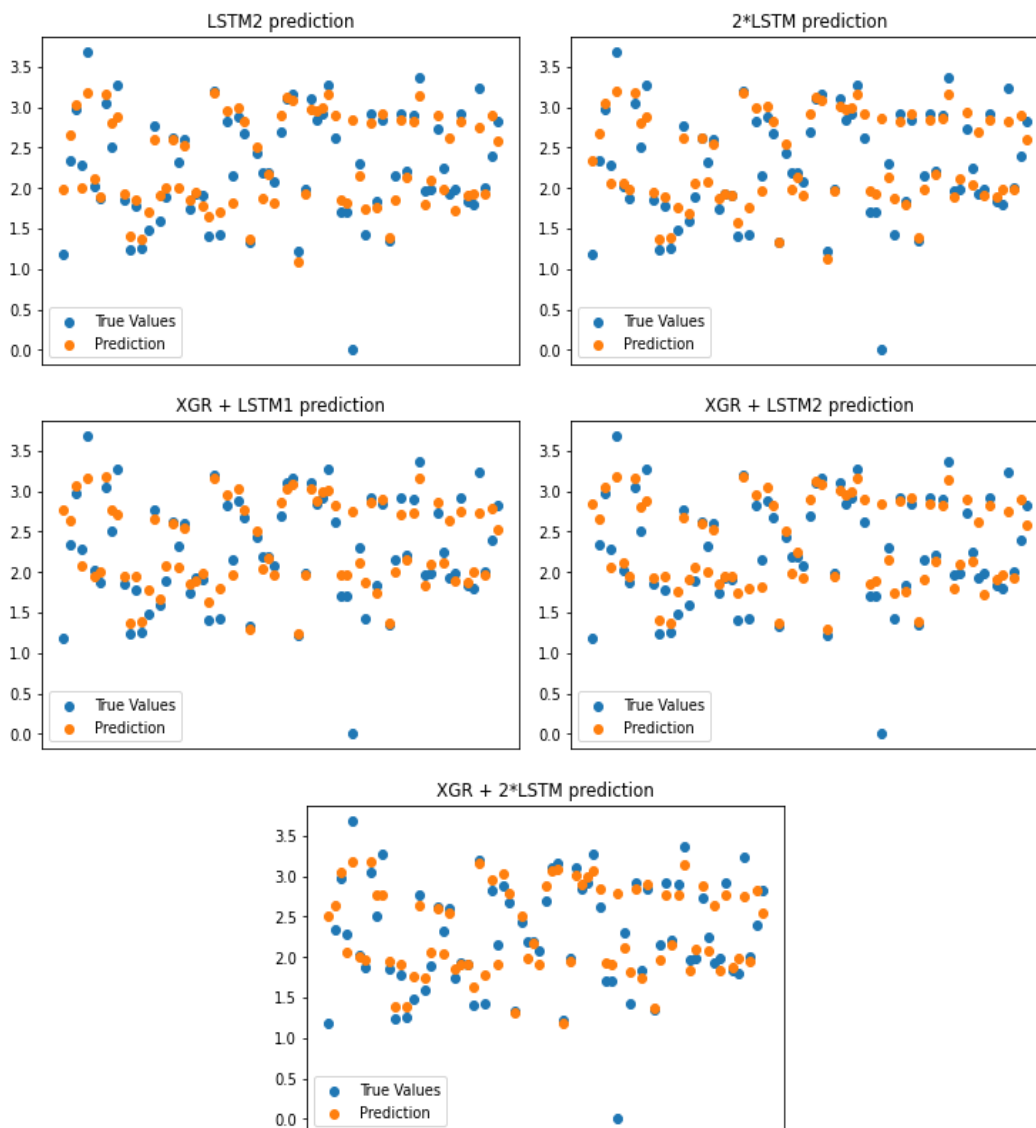
Ακολουθεί ο πίνακας με τα αντίστοιχα σφάλματα.

Models	RMSE(%)	NMBE(%)
XGBOOST	18,704	-3,428
LSTM1	12,936	2,44
LSTM2	13,863	0,439
LSTM1 + LSTM2	13,188	1,441
LSTM1 + XGBOOST	15,084	-0,492
LSTM2 + XGBOOST	14,65	1,673
LSTM1 + LSTM2 + XGBOOST	14,465	-0,181

Πίνακας 5.13 : Πίνακας Σφαλμάτων για την ημερήσια δειγματοληψία στην περίπτωση της μέσης κατανάλωσης στην περίπτωση των “Shuffled” δεδομένων

Εδώ τα αποτελέσματα είναι καλύτερα σε σχέση με πριν καθώς οι σχέσεις μεταξύ των στοιχείων είναι πιο ευθείς. Πλέον δε προσπαθούν τα μοντέλα να προβλέψουν άθροισα αλλά μέσο όρο τιμών που έχει πιο ομοιόμορφη κατανομή με λιγότερες ακραίες τιμές. Τα διαγράμματα αυτών των προβλέψεων είναι:



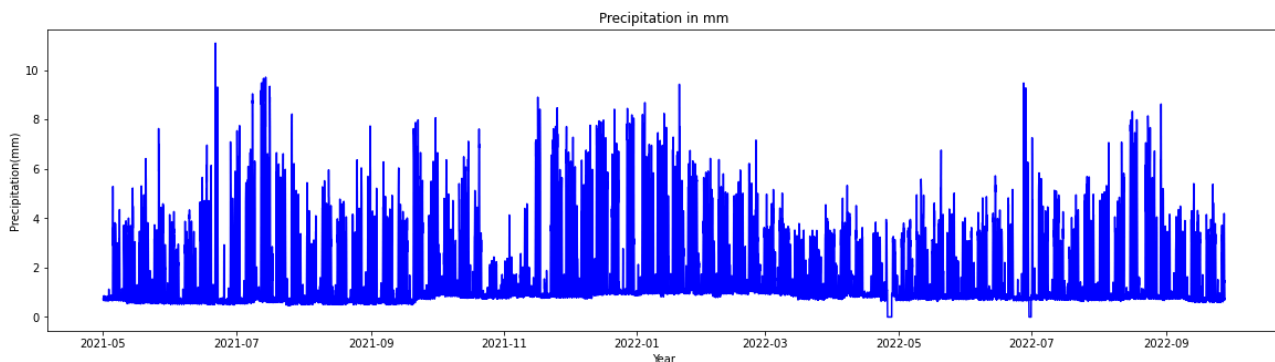


Διαγράμματα αποτελεσμάτων 5.10: Προβλέψεις και πραγματικές τιμές του μέσου όρου της κατανάλωσης για τα διάφορα μοντέλα του πρώτου συνόλου δεδομένων ανά ημέρα στην “Shuffled” προσέγγιση

Εδώ ολοκληρώνεται η παρουσίαση των αποτελεσμάτων για το πρώτο σύνολο δεδομένων από το οποίο ήδη αποδείχθηκε ότι διαφορετικά μοντέλα μπορεί να πετυχαίνουν καλύτερα αποτελέσματα ανάλογα τα δεδομένα εισόδου ακόμα και στο ίδιο σύνολο δεδομένων. Επίσης κάτι άλλο που προκύπτει ως συμπέρασμα από το πρώτο σύνολο δεδομένων είναι ότι όταν γίνει τυχαία δημιουργία των συνόλων εκπαίδευσης και εξέτασης τα αποτελέσματα είναι καλύτερα.

5.3 Παρουσίαση Αποτελεσμάτων για το δεύτερο σύνολο δεδομένων

Το δεύτερο σύνολο δεδομένων είναι από μια βιβλιοθήκη. Οι βιβλιοθήκες όπως είναι γνωστό έχουν ένα ιδιαίτερο ωράριο καθώς δεν είναι τα τυπικά δημόσια κτίρια και επομένως αναμένονται κάποιες διαφοροποιήσεις ως προς την απόδοση αλλά και τα ίδια τα χαρακτηριστικά του συνόλου δεδομένων πιθανόν να περιέχουν ιδιαιτερότητες. Έτσι θα παρουσιαστούν και θα αναλυθούν κάποια στοιχεία για το σύνολο δεδομένων αφού πρώτα παρουσιαστεί το διάγραμμα του.



Διάγραμμα 5.6 : Διάγραμμα της κατανάλωσης ανά ώρα του δεύτερου συνόλου δεδομένων

	temp	electricity	meantemp	dayofyear	teams	teams2	hour
Count	12369	12369	12369	12369	12369	12369	12369
Mean	8,98	1,758	8,979	186,672	0,815	9,771	11,49
Std	9,487	1,638	9,487	91,921	0,388	8,352	6,92
Min	-26	0	-26	1	0	0	0
25%	1,5	0,748	1,5	125	1	1,858	5
50%	9,4	0,933	9,4	189	1	8,558	11
75%	16,1	2,486	16,1	254	1	16,458	17
Max	32,1	11,091	32,1	365	1	43,958	23

	weekofyear	Quarter	Day	Month	Year	Dayofweek	Weekend
Count	12369	12369	12369	12369	12369	12369	12369
Mean	13,081	2,534	15,679	6,651	0,524	2,998	0,287
Std	15,02	0,977	8,77	3,005	0,499	2	0,452
Min	1	1	1	1	0	0	0
25%	18	2	8	5	0	1	0
50%	27	3	16	7	1	3	0
75%	36	3	23	9	1	5	1
Max	52	4	31	12	1	6	1

	sindayofweek	cosdayofweek	sinmonth	cosmonth	sinhour	coshour
Count	12369	12369	12369	12369	12369	12369
Mean	0,000	0,147	-1,084E-02	-2,16E-03	0	0,0424

Std	0,654	0,741	6,847E-01	7,10E-01	0,691	0,72
Min	-0,866	-1	-1	-1	-0,997	-0,99
25%	-0,866	-0,5	-0,866	-0,866	-0,63	-0,57
50%	0	0,5	0	-0,5	0	0,203
75%	0,866	1	0,5	0,5	0,63	0,854
Max	0,866	1	1	1	0,997	1

	sinyear	cosyear	sinweekyear	cosweekyear	Holiday
Count	12369	12369	12369	12369	12369
Mean	0	0,999	-0,052	-0,21	0,038
Std	0,001	0	0,686	0,693	0,178
Min	-0,003	0,999	-1	-1	0
25%	-0,003	0,999	-0,748	-0,822	0
50%	0	1	-0,12	-0,464	0
75%	0	1	0,568	0,464	0
Max	0	1	1	1	1

Πίνακας 5.14: Χαρακτηριστικά του δεύτερου συνόλου δεδομένων

```

temp      temp      electricity  meantemp  dayofyear  teams      teams2
temp      1.000000    0.035282   1.000000  0.092525  -0.665641  -0.977946
electricity 0.035282    1.000000  0.035282  -0.001664 -0.137821  0.021734
meantemp   1.000000    0.035282  1.000000  0.092525  -0.665641  -0.977946
dayofyear  0.092525    -0.001664  0.092525  1.000000  -0.039553  -0.097176
teams      -0.665641    -0.137821  -0.665641  -0.039553  1.000000  0.557049
teams2     -0.977946    0.021734  -0.977946  -0.097176  0.557049  1.000000
hour       0.149801    0.243086  0.149801  -0.000512 -0.174216  -0.122753
weekofyear 0.071052    -0.002387  0.071052  0.969668  -0.031994  -0.074288
quarter    0.115649    0.002361  0.115649  0.959822  -0.055124  -0.120534
day        0.036217    -0.000945  0.036217  0.097436  -0.020686  -0.031202
month      0.092498    -0.002338  0.092498  0.995429  -0.039397  -0.097728
year       -0.068891    0.004692  -0.068891  -0.584320  0.059371  0.064443
dayofweek  0.019368    -0.275504  0.019368  -0.005519  -0.001866  -0.022170
weekend    0.004817    -0.308475  0.004817  -0.013744  0.000046  -0.006810
sindayofweek -0.031886    0.145733  -0.031886  0.002484  -0.000899  0.031990
cosdayofweek -0.017315    -0.140835  -0.017315  -0.011806  0.005491  0.018260
sinmonth   -0.405012    0.014903  -0.405012  -0.757238  0.208359  0.418896
cosmonth   -0.743769    0.135819  -0.743769  0.208722  0.418036  0.759846
sinhour    -0.203842    -0.423958  -0.203842  0.000114  0.235162  0.163836
coshour    -0.102441    -0.469593  -0.102441  0.000385  0.127255  0.079127
sinyear    -0.068891    0.004692  -0.068891  -0.584320  0.059371  0.064443
cosyear    -0.068891    0.004692  -0.068891  -0.584320  0.059371  0.064443
sinweekyear -0.208932    -0.021528  -0.208932  -0.781284  0.099296  0.217278
cosweekyear -0.829115    0.140143  -0.829115  0.018220  0.457947  0.848818
Holiday    -0.063317    -0.101138  -0.063317  0.012107  0.001425  0.078243

```

```

sinweekyear  cosweekyear  Holiday
temp          -0.208932    -0.829115  -0.063317
electricity   -0.021528    0.140143  -0.101138
meantemp      -0.208932    -0.829115  -0.063317
dayofyear     -0.781284    0.018220  0.012107
teams         0.099296    0.457947  0.001425
teams2        0.217278    0.848818  0.078243
hour          0.000345    -0.000357  0.000984
weekofyear    -0.781111    0.045205  0.051022
quarter       -0.790783    0.013323  -0.000665
day           -0.043142    0.002428  0.016194
month         -0.778993    0.013523  0.010855
year          0.415486    -0.014316  -0.019379
dayofweek     0.013012    -0.003507  0.064783
weekend       0.011387    -0.003138  0.026237
sindayofweek -0.007204    0.001928  -0.057060
cosdayofweek  -0.000249    -0.000403  -0.036919
sinmonth      0.955201    0.251737  0.101942
cosmonth      -0.237656    0.957845  0.001504
sinhour       -0.000158    -0.000049  0.000019
coshour       -0.000241    0.000297  -0.000792
sinyear       0.415486    -0.014316  -0.019379
cosyear       0.415486    -0.014316  -0.019379
sinweekyear   1.000000    0.004865  0.104455
cosweekyear   0.004865    1.000000  0.038316
Holiday       0.104455    0.038316  1.000000

```

	cosmonth	sinhour	coshour	sinyear	cosyear
temp	-0.743769	-2.038419e-01	-0.102441	-0.068891	-0.068891
electricity	0.135819	-4.239575e-01	-0.469593	0.004692	0.004692
meantemp	-0.743769	-2.038419e-01	-0.102441	-0.068891	-0.068891
dayofyear	0.208722	1.136249e-04	0.000385	-0.584320	-0.584320
teams	0.418036	2.351622e-01	0.127255	0.059371	0.059371
teams2	0.759846	1.638360e-01	0.070127	0.064443	0.064443
hour	-0.000422	-7.267143e-01	-0.001733	0.000148	0.000148
weekofyear	0.232471	1.105196e-04	0.000387	-0.566401	-0.566401
quarter	0.208665	1.628116e-04	0.000256	-0.577575	-0.577575
day	-0.009338	-1.338052e-04	0.001755	-0.016522	-0.016522
month	0.205660	1.260300e-04	0.000218	-0.585753	-0.585753
year	-0.117118	-9.883656e-05	-0.000097	1.000000	1.000000
dayofweek	-0.001425	-1.553361e-04	-0.000767	0.000193	0.000193
weekend	-0.003528	-1.635164e-04	-0.000994	0.003593	0.003593
sindayofweek	-0.003692	4.440121e-08	0.000569	-0.000449	-0.000449
cosdayofweek	-0.003084	-1.192816e-04	-0.000557	0.005764	0.005764
sinmonth	0.008726	-1.679604e-04	-0.000120	0.409855	0.409855
cosmonth	1.000000	-2.901477e-05	0.000344	-0.117118	-0.117118
sinhour	-0.000029	1.000000e+00	-0.000060	-0.000099	-0.000099
coshour	0.000344	-6.011950e-05	1.000000	-0.000097	-0.000097
sinyear	-0.117118	-9.883656e-05	-0.000097	1.000000	1.000000
cosyear	-0.117118	-9.883656e-05	-0.000097	1.000000	1.000000
sinweekyear	-0.237656	-1.577807e-04	-0.000241	0.415486	0.415486
cosweekyear	0.957845	-4.948821e-05	0.000297	-0.014316	-0.014316
Holiday	0.001504	1.913644e-05	-0.000792	-0.019379	-0.019379

	hour	weekofyear	quarter	day	month	year
temp	0.149801	0.071852	0.115649	0.036217	0.092498	-0.068891
electricity	0.243086	-0.002387	0.002361	-0.000945	-0.002338	0.004692
meantemp	0.149801	0.071852	0.115649	0.036217	0.092498	-0.068891
dayofyear	-0.000512	0.969668	0.959822	0.097436	0.995429	-0.584320
teams	-0.174216	-0.031994	-0.055124	-0.020686	-0.039397	0.059371
teams2	-0.122753	-0.074288	-0.120534	-0.031202	-0.097728	0.064443
hour	1.000000	-0.000516	-0.000365	-0.002153	0.000307	0.000148
weekofyear	-0.000516	1.000000	0.936031	0.072958	0.967128	-0.566401
quarter	-0.000365	0.936031	1.000000	0.007998	0.963750	-0.577575
day	-0.002153	0.072958	0.007998	1.000000	0.002197	-0.016522
month	-0.000307	0.967128	0.963750	0.002197	1.000000	-0.585753
year	0.000148	-0.566401	-0.577575	-0.016522	-0.585753	1.000000
dayofweek	0.001001	-0.000180	0.004066	-0.007284	-0.004812	0.000193
weekend	0.001286	-0.006794	-0.004305	0.007155	-0.014483	0.003593
sindayofweek	-0.000710	0.004680	-0.010193	0.004684	0.002043	-0.000449
cosdayofweek	0.000729	0.000729	-0.006729	0.022675	-0.014062	0.005764
sinmonth	0.000197	-0.744183	-0.772525	0.002791	-0.760423	0.409855
cosmonth	-0.000422	0.232471	0.208665	-0.009338	0.205660	-0.117118
sinhour	0.000119	0.000119	0.000163	-0.000134	0.000126	-0.000099
coshour	-0.072674	-0.001733	0.000387	0.000256	0.001755	-0.000218
sinyear	0.000148	-0.566401	-0.577575	-0.016522	-0.585753	1.000000
cosyear	0.000148	-0.566401	-0.577575	-0.016522	-0.585753	1.000000
sinweekyear	0.000345	-0.781111	-0.790783	-0.043142	-0.778993	0.415486
cosweekyear	-0.000357	0.045205	0.013323	0.002428	0.013523	-0.014316
Holiday	0.000984	0.051022	-0.000665	0.016194	0.010855	-0.019379

	dayofweek	weekend	sindayofweek	cosdayofweek	sinmonth
temp	0.019368	0.004817	-3.188644e-02	-0.017315	-0.405012
electricity	-0.275504	-0.308475	1.457325e-01	-0.140835	0.014903
meantemp	0.019368	0.004817	-3.188644e-02	-0.017315	-0.405012
dayofyear	-0.005519	-0.013744	2.484425e-03	-0.011806	-0.757238
teams	-0.001866	0.000046	-8.988467e-04	0.005491	0.208359
teams2	-0.022170	-0.006810	3.199009e-02	0.018260	0.418896
hour	0.001001	0.001286	-7.099859e-04	0.000729	0.000197
weekofyear	-0.000180	-0.006794	4.680399e-03	0.000729	-0.744183
quarter	0.004066	-0.004305	-1.010336e-02	-0.006729	-0.772525
day	-0.007284	0.007155	4.684309e-03	0.022675	0.002791
month	-0.004812	-0.014483	2.043268e-03	-0.014062	-0.760423
year	0.000193	0.003593	-4.492387e-04	0.005764	0.409855
dayofweek	1.000000	0.791712	-5.666131e-01	-0.000497	0.000188
weekend	0.791712	1.000000	-4.201059e-01	0.515624	0.002028
sindayofweek	-0.566613	-0.420106	1.000000e+00	0.000059	0.001341
cosdayofweek	-0.000497	0.515624	5.930583e-05	1.000000	-0.000696
sinmonth	0.000188	0.002028	1.340829e-03	-0.000696	1.000000
cosmonth	-0.001425	-0.003528	-3.692143e-03	-0.003084	0.008726
sinhour	-0.000155	-0.000164	4.440121e-08	-0.000119	-0.000168
coshour	-0.000767	-0.000994	5.686418e-04	-0.000557	-0.000120
sinyear	0.000193	0.003593	-4.492387e-04	0.005764	0.409855
cosyear	0.000193	0.003593	-4.492387e-04	0.005764	0.409855
sinweekyear	0.013012	0.011387	-7.204030e-03	-0.000249	0.955201
cosweekyear	-0.003507	-0.003138	1.928215e-03	-0.000403	0.251737
Holiday	0.064783	0.026237	-5.705967e-02	-0.036919	0.101942

Εικόνα 5.2: Συσχέτιση των χαρακτηριστικών του δεύτερου συνόλου δεδομένων μεταξύ τους

Από τα στοιχεία αυτά εξάγονται ξανά πολύτιμες πληροφορίες αντίστοιχα με αυτές του πρώτου συνόλου δεδομένων. Υπάρχει ξανά μεγάλη συσχέτιση μεταξύ του στόχου και του coshour αλλά και γενικά με τους συνημιτονοειδής μετασχηματισμούς πράγμα που δείχνει μια πιο γενική σχέση μεταξύ τους. Από αυτό το σύνολο δεδομένων παρατηρώντας ότι ισχύουν όμοια πράγματα με το πρώτο μπορούν εξαχθούν πιο γενικά συμπεράσματα για τα χαρακτηριστικά που παίζουν ρόλο στην εκπαίδευση των μοντέλων και ποσο αυτά βοηθούν. Ακόμα παρατηρούνται κάποιες μικροί διαφορές όπως το μεγαλύτερο πλήθος χαρακτηριστικών σε αυτό το σύνολο δεδομένων (12369, 8719). Ενώ και οι θερμοκρασίες παρουσιάζουν μια μικρή αύξηση με μεγαλύτερη μέγιστη θερμοκρασία καθώς και λίγο μεγαλύτερο μέσο όρο θερμοκρασιών. Με αυτά τα δεδομένα θα παρουσιαστούν τα αποτελέσματα και θα εξηγηθούν τυχόν διαφοροποιήσεις στην απόδοση και οι λόγοι που αυτές προκύπτουν.

“Σειριακή” Προσέγγιση Ωριαία δειγματοληψία

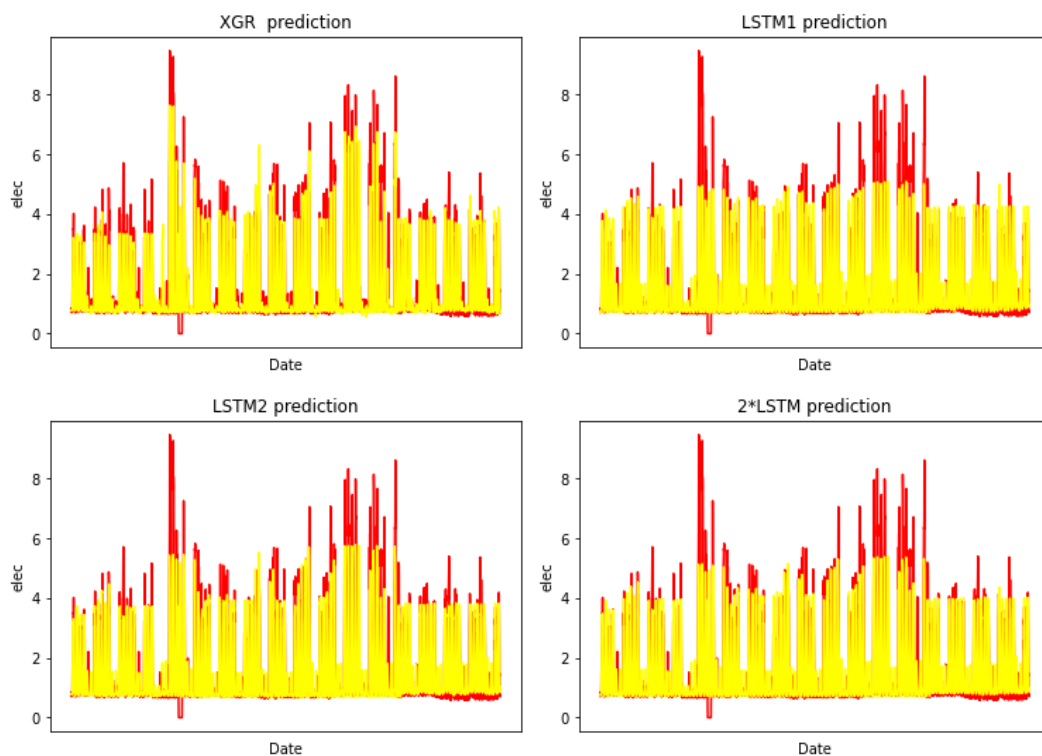
Ξανά πρώτα θα παρουσιαστεί η εφαρμογή των μοντέλων στο σύνολο διατηρώντας την ωριαία μορφή των δεδομένων. Στον πίνακα παρουσιάζονται τα σφάλματα αυτής της περίπτωσης.

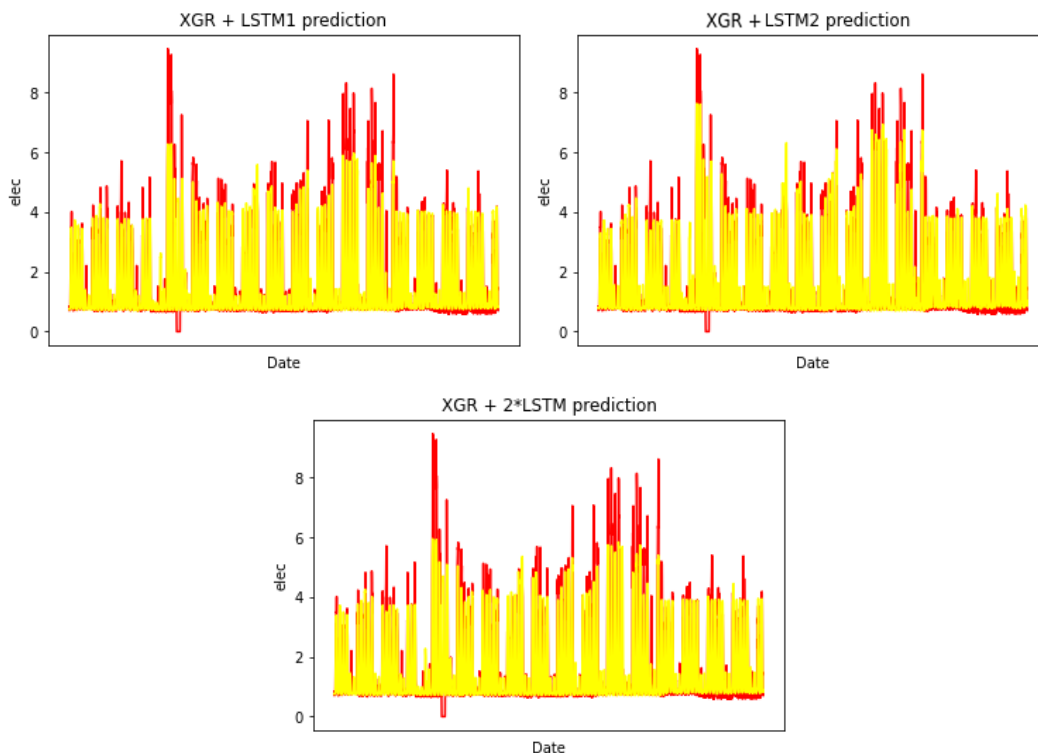
Models	RMSE(%)	NMBE(%)
XGBOOST	37,772	-2,271

LSTM1	46,167	-12,122
LSTM2	39,296	-3,351
LSTM1 + LSTM2	41,811	-7,737
LSTM1 + XGBOOST	39,422	-7,196
LSTM2 + XGBOOST	37,254	-2,840
LSTM1 + LSTM2 + XGBOOST	38,959	-5,911

Πίνακας 5.15: Τιμές Σφαλμάτων του δεύτερου συνόλου δεδομένων με ωριαία δειγματοληψία

Από τα απλά μοντέλα το καλύτερο και πάλι είναι το XGBOOST και δεύτερο έρχεται το LSTM με το σετ αξιολόγησης. Ο συνδυασμός αυτών των μοντέλων είναι και το καλύτερο μοντέλο από όλα με βάση το σφάλμα RMSE ενώ το μικρότερο σφάλμα NMBE το έχει το XGBOOST, πράγμα λογικό αφού όπως έχει αναφερθεί και νωρίτερα είναι αρκετά συντηρητικό στις προβλέψεις του και άρα δεν κάνει μεγάλα λάθη προς μια κατεύθυνση. Τώρα θα ακολουθήσουν τα διαγράμματα των προβλέψεων ώστε να φανεί και πιο είναι το καλύτερο σχηματικά.



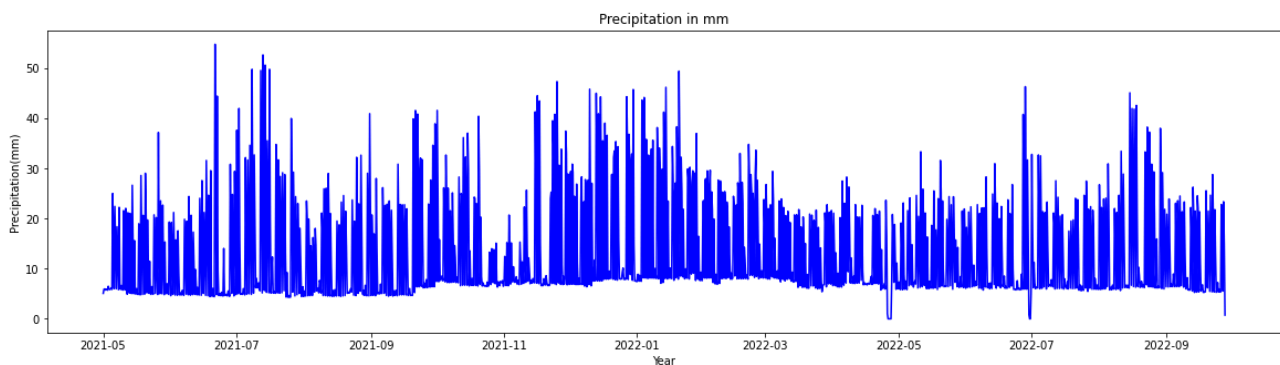


Διαγράμματα αποτελεσμάτων 5.11: Προβλέψεις και πραγματικές τιμές της κατανάλωσης για τα διάφορα μοντέλα του δεύτερου συνόλου δεδομένων ανά ώρα

Όπως φαίνεται και από το σχήμα το καλύτερο είναι ο συνδυασμός του XGBOOST με το δεύτερο LSTM καθώς κατορθώνει να εντοπίσει και να προβλέψει περισσότερες τιμές που ξεφεύγουν από το μέσο όρο. Οι προβλέψεις δείχνουν αρκετά πιο ακριβείς σε σχέση με το πρώτο σύνολο δεδομένων σε αντίθεση με τις τιμές των σφαλμάτων που στο πρώτο σύνολο δεδομένων ήταν αρκετά μικρότερες. Αυτό συμβαίνει επειδή το συγκεκριμένο σύνολο έχει αρκετά υψηλές και χαμηλές τιμές με αποτέλεσμα τα λάθη να είναι μεγαλύτερα σε απόλυτη τιμή όταν αυτές οι τιμές δεν προβλέπονται σωστά. Όπως έχει προαναφερθεί η RMSE “κοστολογεί” πολύ ακριβά τα μεγάλα σε τιμή λάθη με αποτέλεσμα στο παρών σύνολο δεδομένων να εμφανίζει μεγάλο σφάλμα και να μη δικαιολογείται σχηματικά.

Ανά 8 ώρες δειγματοληψία

Και σε αυτό το σύνολο δεδομένων θα διατηρηθεί η συνηθισμένη δομή αφού πρώτα θα παρουσιάζονται τα αποτελέσματα για το άθροισμα της κατανάλωσης ανά 8 ώρες και μετά το μέσο όρο ανά 8 ώρες. Πρώτα παρουσιάζεται το διάγραμμα του αθροίσματος.



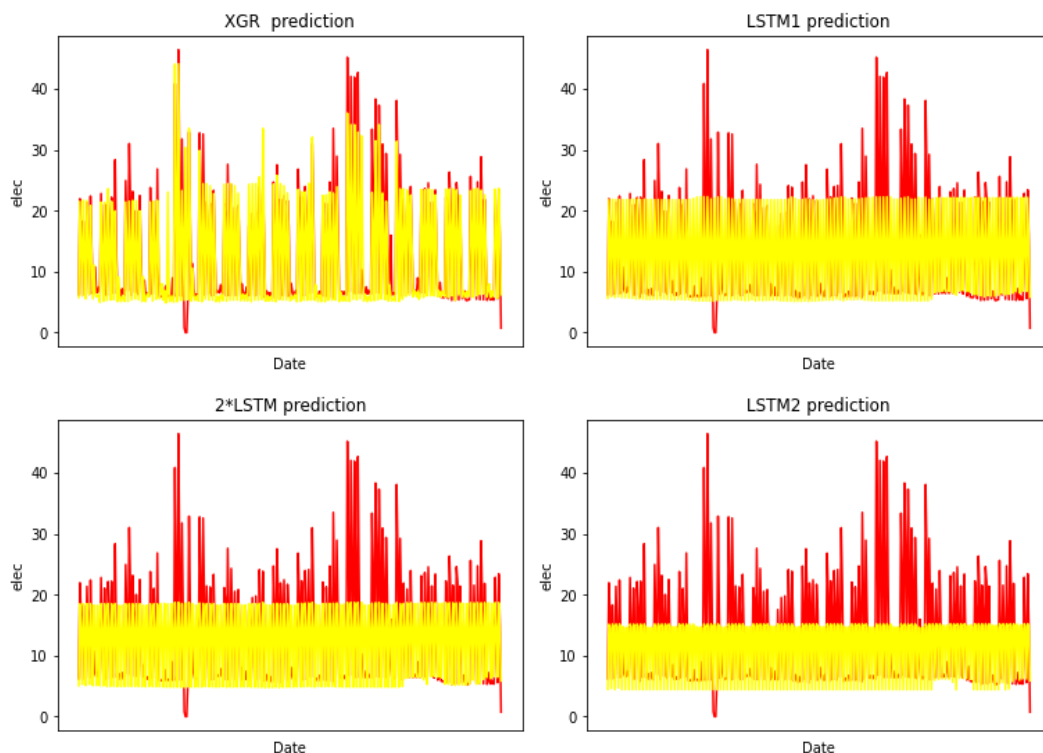
Διάγραμμα 5.7: Διάγραμμα του αθροίσματος της κατανάλωσης ανά 8 ώρες του δεύτερου συνόλου δεδομένων

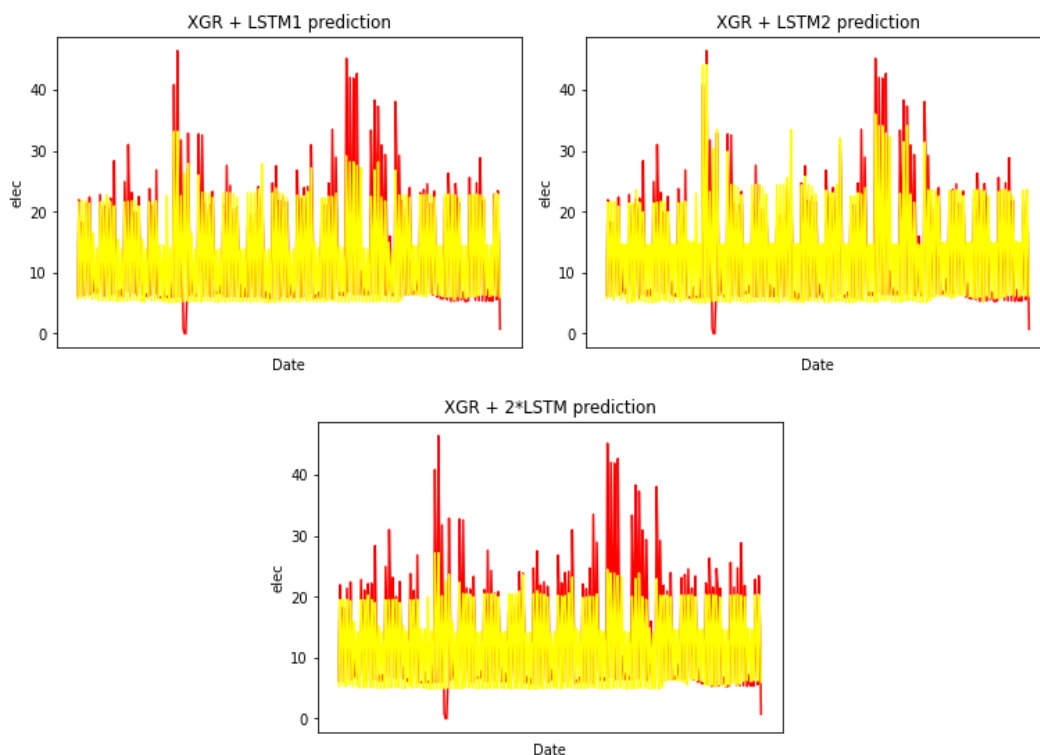
Ακολουθεί ο πίνακας για την πρώτη υποπερίπτωση.

Models	RMSE(%)	NMBE(%)
XGBOOST	28,070	-5,211
LSTM1	32,131	-3,946
LSTM2	33,449	-0,834
LSTM1 + LSTM2	31,960	-2,39
LSTM1 + XGBOOST	28,296	-4,578
LSTM2 + XGBOOST	29,773	3,5495
LSTM1 + LSTM2 + XGBOOST	28,954	-3,330

Πίνακας 5.16: Τιμές Σφαλμάτων του δεύτερου συνόλου δεδομένων με ανά 8 ώρες δειγματοληψία για εύρεση του αθροίσματος της κατανάλωσης στο διάστημα αυτό

Εανά παρατηρείται μισή βελτίωση στα αποτελέσματα με καλύτερο να είναι το XGBOOST. Η σημαντική διαφορά είναι ότι δεύτερο πλέον από τα απλά μοντέλα είναι το LSTM χωρίς το σετ αξιολόγησης και αυτό συμβαίνει γιατί μπορεί πριν να βοηθούσε το σετ αξιολόγησης στην εκπαίδευση όμως τώρα περιορίζει τη εκμάθηση. Ακολουθούν τα διαγράμματα.

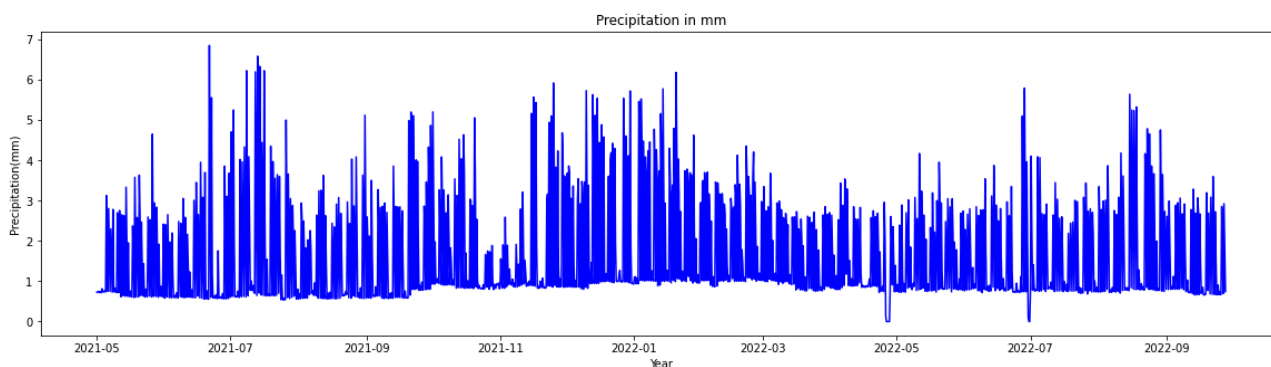




Διαγράμματα αποτελεσμάτων 5.12: Προβλέψεις και πραγματικές τιμές του αθροίσματος της κατανάλωσης για τα διάφορα μοντέλα του δεύτερου συνόλου δεδομένων ανά 8 ώρες

Φαίνεται ότι τα καλύτερο σχηματικά είναι ο συνδυασμός XGBOOST και του δεύτερου LSTM μαζί με το απλό XGBOOST καθώς είναι το μόνο που καταφέρνει να εντοπίσει τα περισσότερα ακραία σημεία.

Η συνέχεια γίνεται με το μέσο όρο της κατανάλωσης ανά 8 ώρες με το διάγραμμα να παρουσιάζεται παρακάτω και μαζί ο πίνακας με τα σφάλματα.



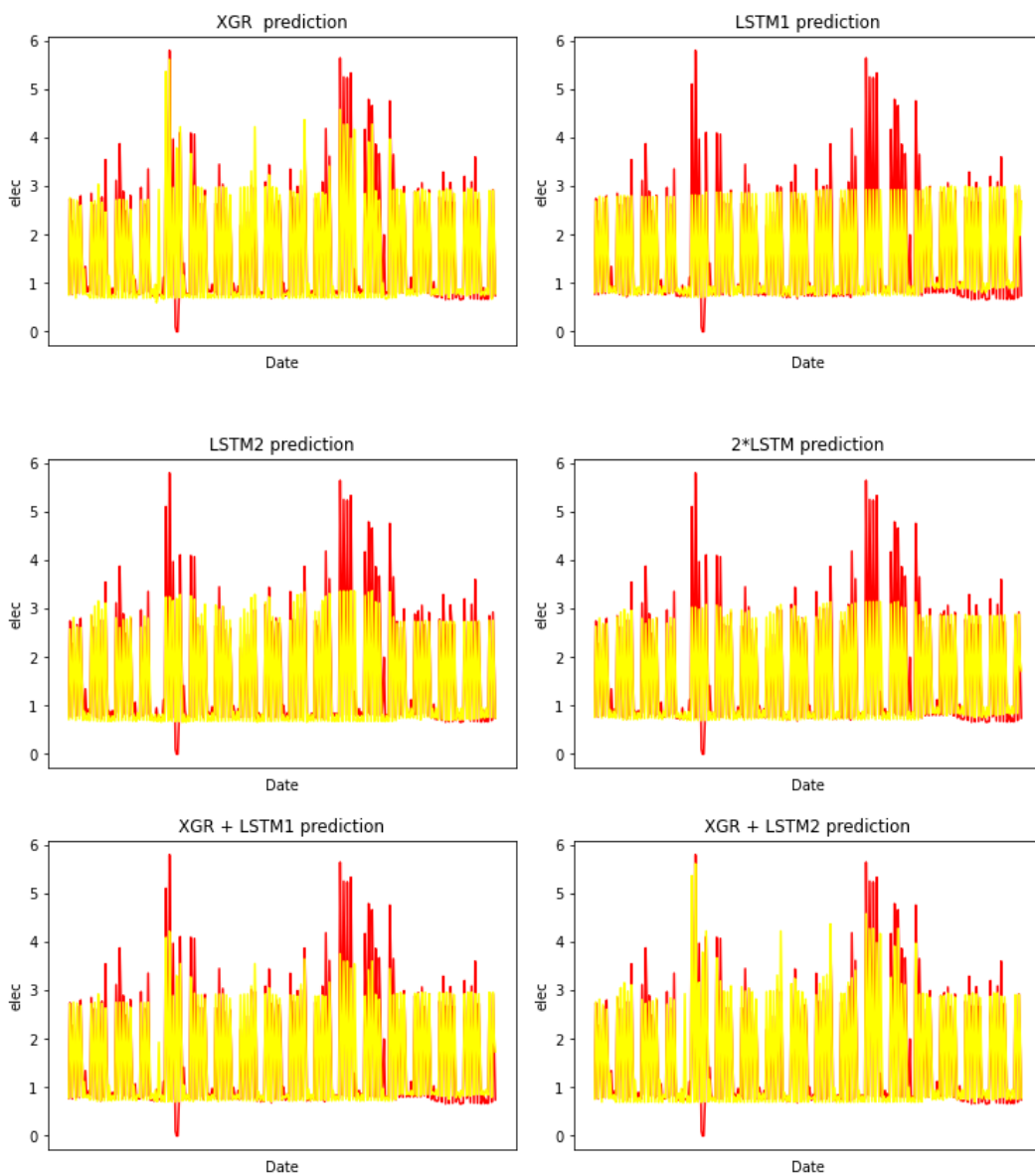
Διάγραμμα 5.8: Διάγραμμα του μέσου όρου της κατανάλωσης ανά 8 ώρες του δεύτερου συνόλου δεδομένων

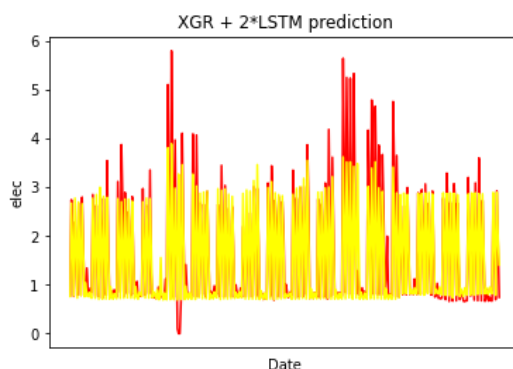
Models	RMSE(%)	NMBE(%)
XGBOOST	27,898	-3,876
LSTM1	30,876	-1,171

LSTM2	34,668	-7,297
LSTM1 + LSTM2	32,411	-4,234
LSTM1 + XGBOOST	27,717	-2,524
LSTM2 + XGBOOST	28,192	0,213
LSTM1 + LSTM2 + XGBOOST	29,324	-4,115

Πίνακας 5.17: Τιμές Σφαλμάτων του δεύτερου συνόλου δεδομένων με ανά 8 ώρες δειγματοληψία για εύρεση του μέσου όρου της κατανάλωσης στο διάστημα αυτό

Υπάρχει μια μικρή βελτίωση στα αποτελέσματα αλλά όχι αρκετά μεγάλη. Αυτό που προκαλεί εντύπωση είναι ότι πλέον καλύτερο μοντέλο είναι το τέταρτο του πίνακα. Ακολουθούν τα διαγράμματα των αποτελεσμάτων.



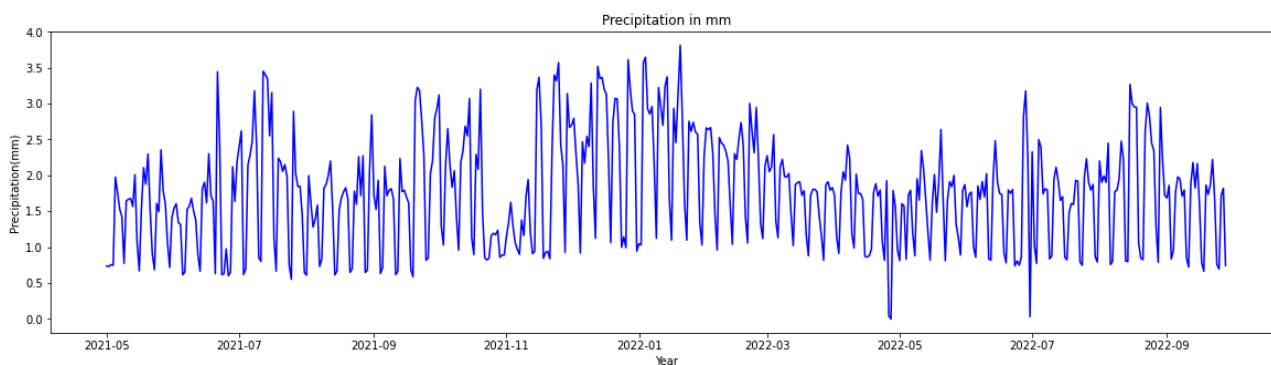


Διαγράμματα αποτελεσμάτων 5.13: Προβλέψεις και πραγματικές τιμές του μέσου όρου της κατανάλωσης για τα διάφορα μοντέλα του δεύτερου συνόλου δεδομένων ανά 8 ώρες

Εάν παρατηρείται ότι τα καλύτερα δυο στις προβλέψεις είναι το XGBoost και ο συνδυασμός του με το LSTM2 αφού πετυχαίνει να εντοπίσει τα ακραία σημεία.

Ημερήσια δειγματοληψία

Παρακάτω θα παρουσιαστούν τα αποτελέσματα για τα στοιχεία σε ημερήσια δειγματοληψία και πρώτα θα παρουσιάζονται τα αποτελέσματα για το άθροισμα της κατανάλωσης ανά ημέρα και μετά το μέσο όρο ανά ημέρα.



Διάγραμμα 5.9: Διάγραμμα του αθροίσματος της κατανάλωσης ανά ημέρα του δεύτερου συνόλου δεδομένων

Φαίνεται ότι το διάγραμμα του συνόλου δεδομένων παρουσιάζει πιο ομοιόμορφη διακύμανση αλλά υπάρχουν εντονότερες ακραίες τιμές αποτέλεσμα του αθροίσματος της κατανάλωσης ανά ημέρα και αναμένονται καλύτερα αποτελέσματα όσον αφορά την απόδοση του XGBOOST.

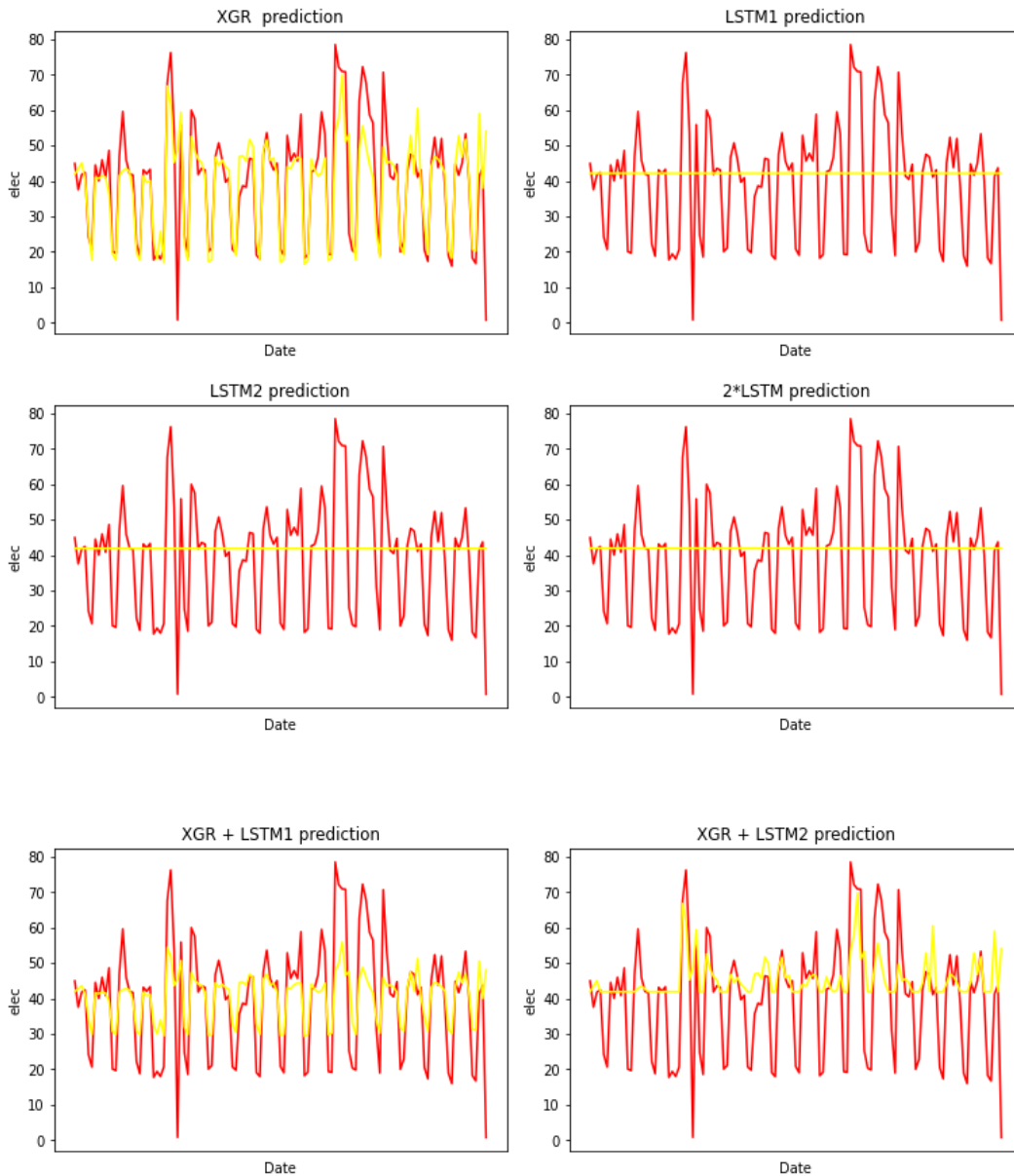
Ακολουθεί ο πίνακας των αποτελεσμάτων.

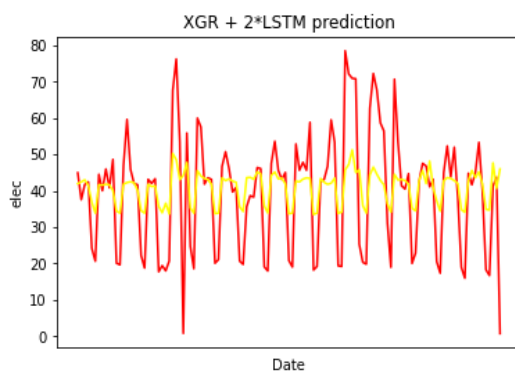
Models	RMSE(%)	NMBE(%)
XGBOOST	25,648	-2,089
LSTM1	42,223	6,452
LSTM2	42,417	7,618
LSTM1 + LSTM2	42,316	7,035
LSTM1 + XGBOOST	30,26	2,181

LSTM2 + XGBOOST	40,009	15,836
LSTM1 + LSTM2 + XGBOOST	33,763	3,993

Πίνακας 5.18: Τιμές Σφαλμάτων του δεύτερου συνόλου δεδομένων με ημερήσια δειγματοληψία για εύρεση του αθροίσματος της κατανάλωσης στο διάστημα αυτό

Τα αποτελέσματα στα LSTM είναι χειρότερα καθώς όπως έχει προαναφερθεί χρειάζονται περισσότερες τιμές για να πετύχουν μικρότερες τιμές σφάλματος. Ακολουθούν τα διαγράμματα των προβλέψεων.

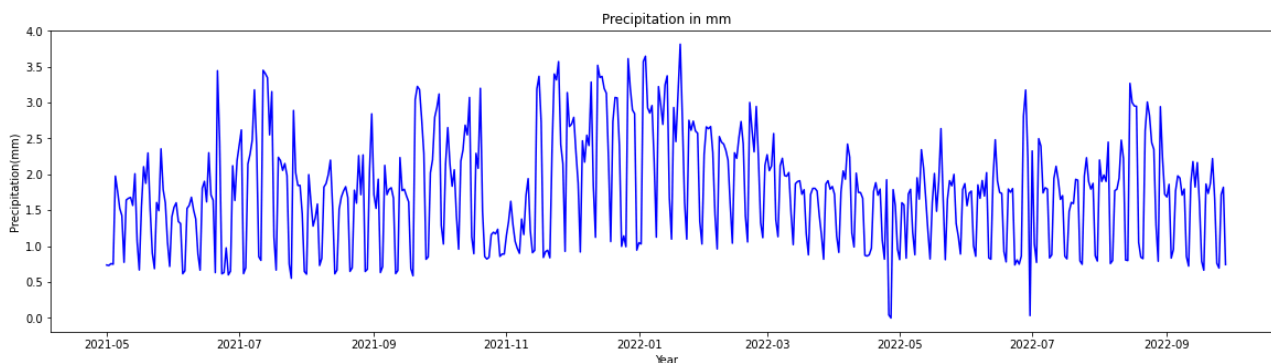




Διαγράμματα αποτελεσμάτων 5.14: Προβλέψεις και πραγματικές τιμές του αθροίσματος της κατανάλωσης για τα διάφορα μοντέλα του δεύτερου συνόλου δεδομένων ανά ημέρα

Αυτό που προκύπτει είναι ότι τελικά τα LSTM δεν προβλέπουν απλά λαμβάνουν μια τιμή και τη διατηρούν κοντά στο μέσο όρο.

Ακολουθεί η περίπτωση του μέσου όρου της κατανάλωσης.



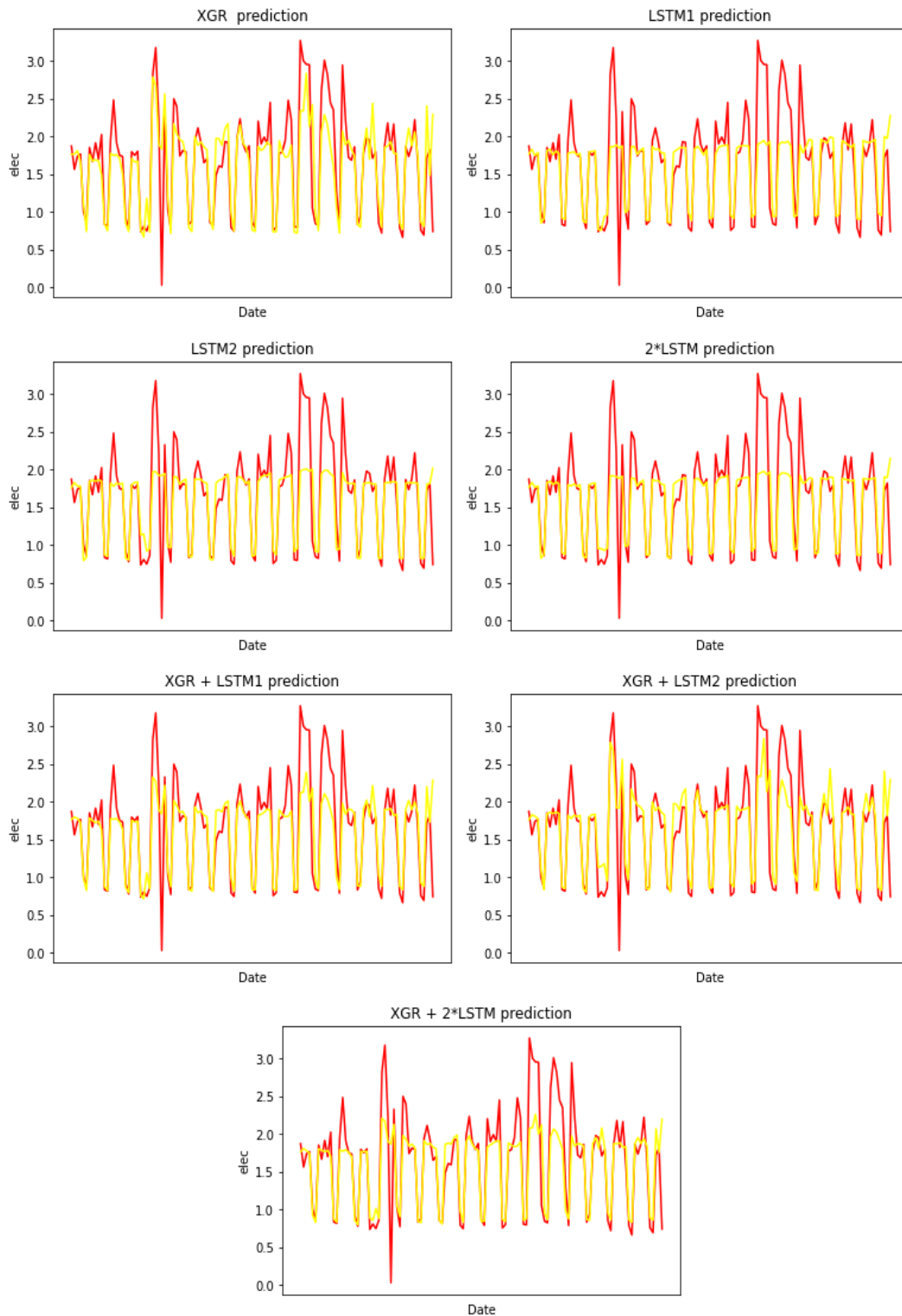
Διάγραμμα 5.10: Διάγραμμα του μέσου όρου της κατανάλωσης ανά ημέρα του δεύτερου συνόλου δεδομένων

Ο πίνακας σφαλμάτων είναι

Models	RMSE(%)	NMBE(%)
XGBOOST	23,769	-2,727
LSTM1	25,901	0,668
LSTM2	27,972	-7,701
LSTM1 + LSTM2	26,577	-3,516
LSTM1 + XGBOOST	23,813	-1,029
LSTM2 + XGBOOST	23,206	0,146
LSTM1 + LSTM2 + XGBOOST	24,755	-3,253

Πίνακας 5.19: Τιμές Σφαλμάτων του δεύτερου συνόλου δεδομένων με ημερήσια δειγματοληψία για εύρεση του μέσου όρου της κατανάλωσης στο διάστημα αυτό

Υπάρχει μια μικρή βελτίωση λόγω της πιο ομοιόμορφης κατανομής των δεδομένων και επειδή το σύνολο εκπαίδευσης μοιάζει αρκετά με το σύνολο εξέτασης. Ακολουθούν τα διαγράμματα.



Διαγράμματα αποτελεσμάτων 5.15: Προβλέψεις και πραγματικές τιμές του μέσου όρου της κατανάλωσης για τα διάφορα μοντέλα του δεύτερου συνόλου δεδομένων ανά ημέρα

Εδώ τα αποτελέσματα των LSTM είναι αρκετά καλύτερα από πριν παρόλο που και εδώ οι τιμές είναι λίγες. Αυτό συμβαίνει γιατί το σύνολο εξέτασης πλέον δεν προκύπτει από άθροισμα τιμών αλλά από τον μέσο όρο τους δίνοντας έτσι τη δυνατότητα να χρειάζονται λιγότερες τιμές για την εκπαίδευση των LSTM.

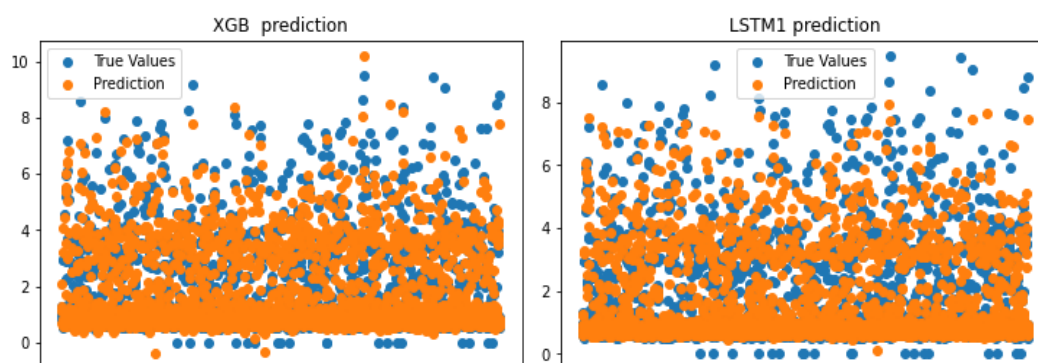
“Shuffled” προσέγγιση
Ωριαία δειγματοληψία.

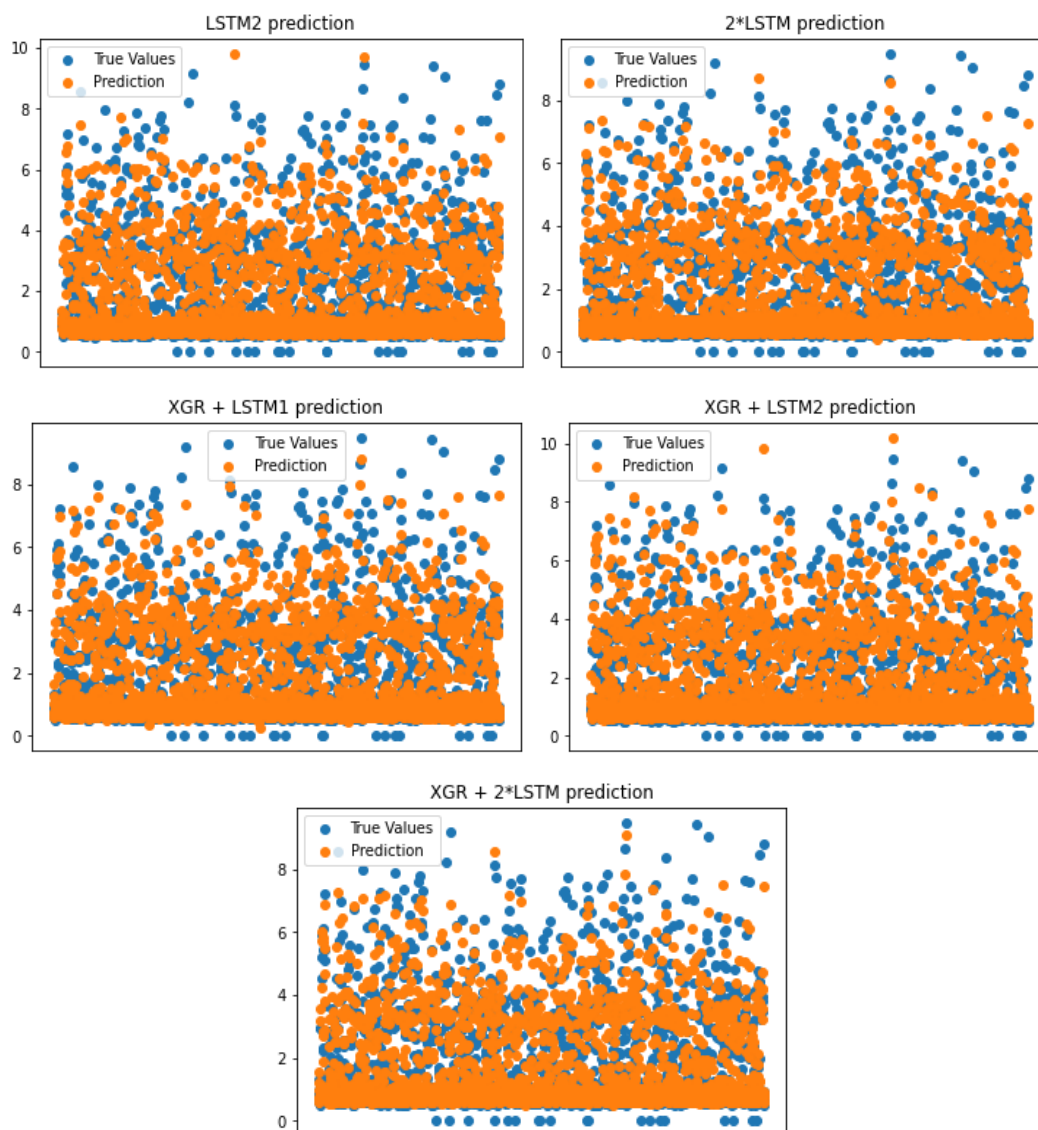
Όλα τα διαγράμματα των συνόλων δεδομένων είναι ίδια με αυτά των αντίστοιχων περιπτώσεων που αναλύθηκαν προηγουμένως και για αυτό δε θα συμπεριληφθούν σε ξανά. Θα παρουσιαστούν και θα σχολιαστούν τα αποτελέσματα και τα διαγράμματα αυτών για την κάθε περίπτωση. Ακολουθεί ο πίνακας των αποτελεσμάτων.

Models	RMSE(%)	NMBE(%)
XGBOOST	34,716	2,464
LSTM1	39,626	-1,952
LSTM2	37,471	-0,91
LSTM1 + LSTM2	37,279	-1,431
LSTM1 + XGBOOST	34,536	0,256
LSTM2 + XGBOOST	37,065	8,183
LSTM1 + LSTM2 + XGBOOST	34,294	-0,132

Πίνακας 5.20: Τιμές Σφαλμάτων του δεύτερου συνόλου δεδομένων με ωριαία δειγματοληψία στην περίπτωση των “Shuffled” δεδομένων

Παρατηρείται και πάλι ότι όταν επιλέγονται τυχαία τα σύνολα εκπαίδευση και εξέτασης τα σφάλματα είναι μικρότερα και η ακρίβεια μεγαλύτερη. Επιβεβαιώνεται λοιπόν ότι η τυχαία επιλογή από όλο το εύρος του συνόλου δεδομένων μπορεί να προσφέρει αρκετά καλύτερη εκπαίδευση στα μοντέλα. Ακολουθούν τα διαγράμματα των αποτελεσμάτων.





Διαγράμματα αποτελεσμάτων 5.16: Προβλέψεις και πραγματικές τιμές της κατανάλωσης για τα διάφορα μοντέλα του δεύτερου συνόλου δεδομένων ανά ώρα στην “Shuffled” προσέγγιση

Και από τα διαγράμματα φαίνεται αυτό που ειπώθηκε νωρίτερα ότι δηλαδή τα αποτελέσματα είναι καλύτερα όταν γίνεται τυχαία επιλογή των δεδομένων για τα σύνολα εκπαίδευσης και εξέτασης.

Ανά 8 ώρες δειγματοληψία

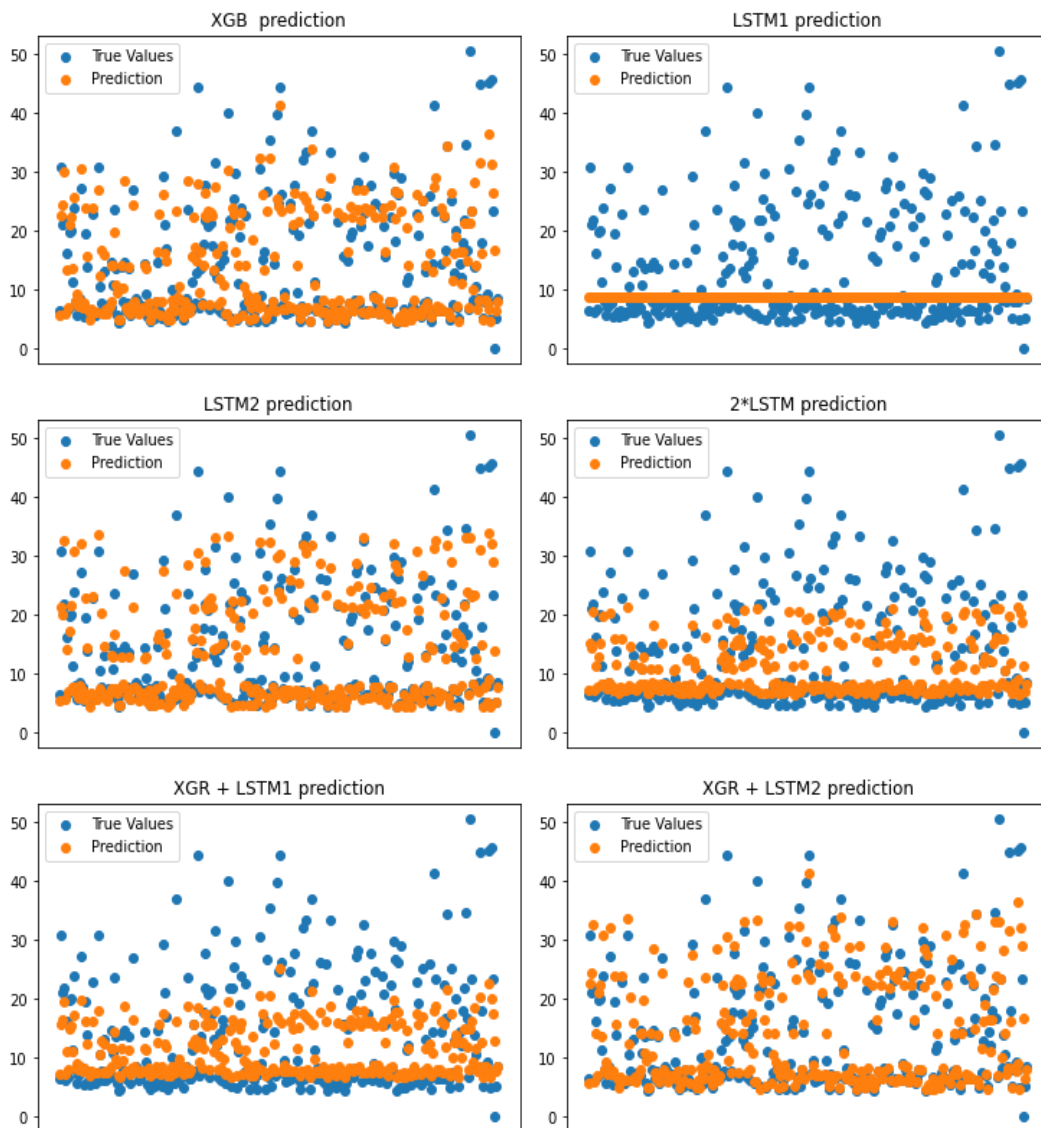
Αρχικά θα παρατεθούν τα αποτελέσματα για την περίπτωση του αθροίσματος στην κατανάλωση και μετά του μέσου όρου της στο διάστημα αυτό. Ακολουθεί ο πίνακας με τα σφάλματα για την πρώτη υποπερίπτωση.

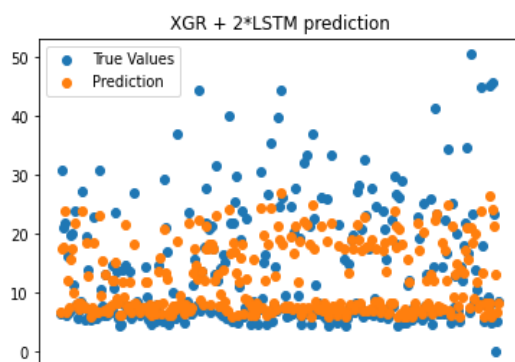
Models	RMSE(%)	NMBE(%)
XGBOOST	32,055	0,505
LSTM1	75,654	-31,072

LSTM2	34,299	-1,885
LSTM1 + LSTM2	47,834	-16,479
LSTM1 + XGBOOST	47,081	-15,283
LSTM2 + XGBOOST	32,099	3,553
LSTM1 + LSTM2 + XGBOOST	39,908	-10,817

Πίνακας 5.21: Τιμές Σφαλμάτων του δεύτερου συνόλου δεδομένων με ανά 8 ώρες δειγματοληψία για την εύρεση του αθροίσματος της κατανάλωσης στην περίπτωση των "Shuffled" δεδομένων

Το γεγονός ότι μειώνονται οι τιμές των χαρακτηριστικών σε πλήθος όσο και ότι αναζητείται η εύρεση του αθροίσματος της κατανάλωσης οδηγεί το πρώτο LSTM σε εντελώς λάθος αποτελέσματα. Ακολουθούν τα διαγράμματα.



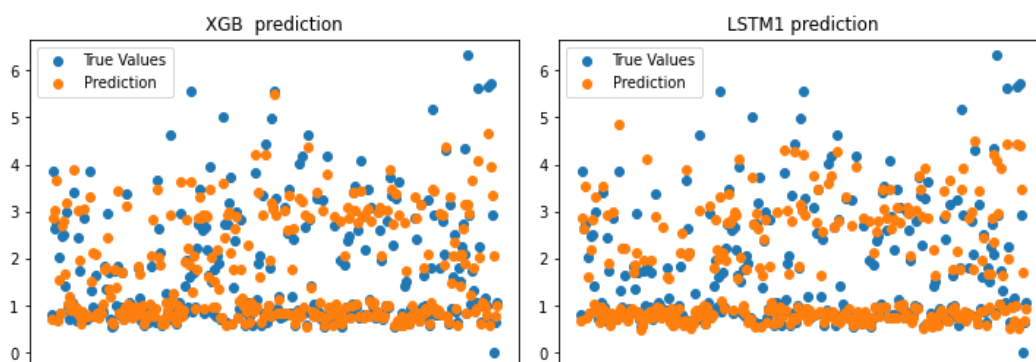


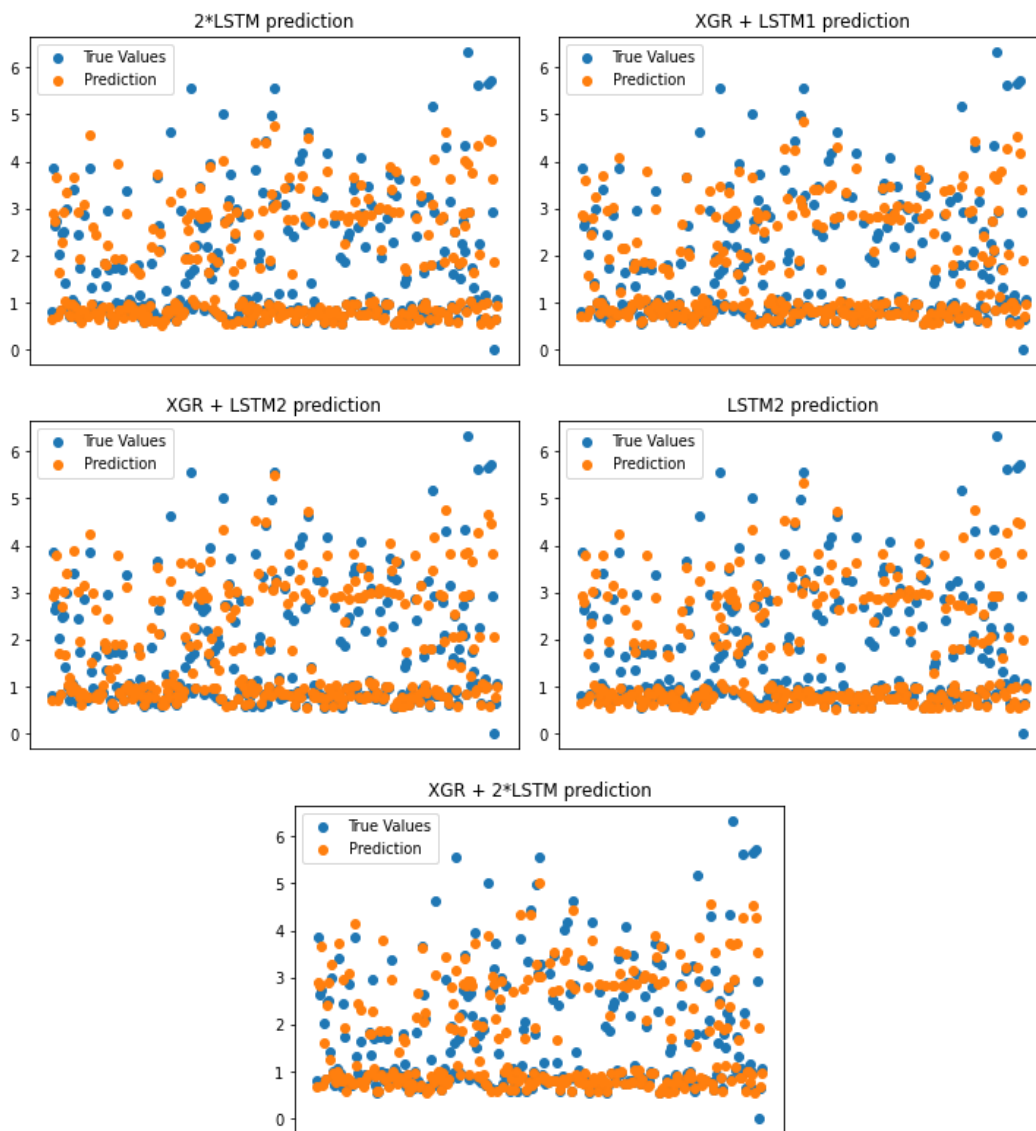
Διαγράμματα αποτελεσμάτων 5.17: Προβλέψεις και πραγματικές τιμές του αθροίσματος της κατανάλωσης για τα διάφορα μοντέλα του δεύτερου συνόλου δεδομένων ανά 8 ώρες στην “Shuffled” προσέγγιση

Έτσι φαίνεται ότι μπορεί το πρώτο LSTM να είναι πρακτικά “άχρηστο” το δεύτερο όμως πετυχαίνει αρκετά καλά αποτελέσματα δεδομένων των συνθηκών. Ακολουθεί η περίπτωση πρόβλεψης του μέσου του της κατανάλωσης με τον πίνακα σφαλμάτων και τα διαγράμματα να ακολουθούν.

Models	RMSE(%)	NMBE(%)
XGBOOST	21,748	0,003
LSTM1	32,638	2,06
LSTM2	30,758	-1,507
LSTM1 + LSTM2	31,1	0,276
LSTM1 + XGBOOST	31,099	0,974
LSTM2 + XGBOOST	30,93	3,04
LSTM1 + LSTM2 + XGBOOST	30,733	0,147

Πίνακας 5.22: Τιμές Σφαλμάτων του δεύτερου συνόλου δεδομένων με ανά 8 ώρες δειγματοληψία για την εύρεση του μέσου όρου της κατανάλωσης στην περίπτωση των “Shuffled” δεδομένων





Διαγράμματα αποτελεσμάτων 5.18: Προβλέψεις και πραγματικές τιμές του μέσου όρου της κατανάλωσης για τα διάφορα μοντέλα του δεύτερου συνόλου δεδομένων ανά 8 ώρες στην “Shuffled” προσέγγιση

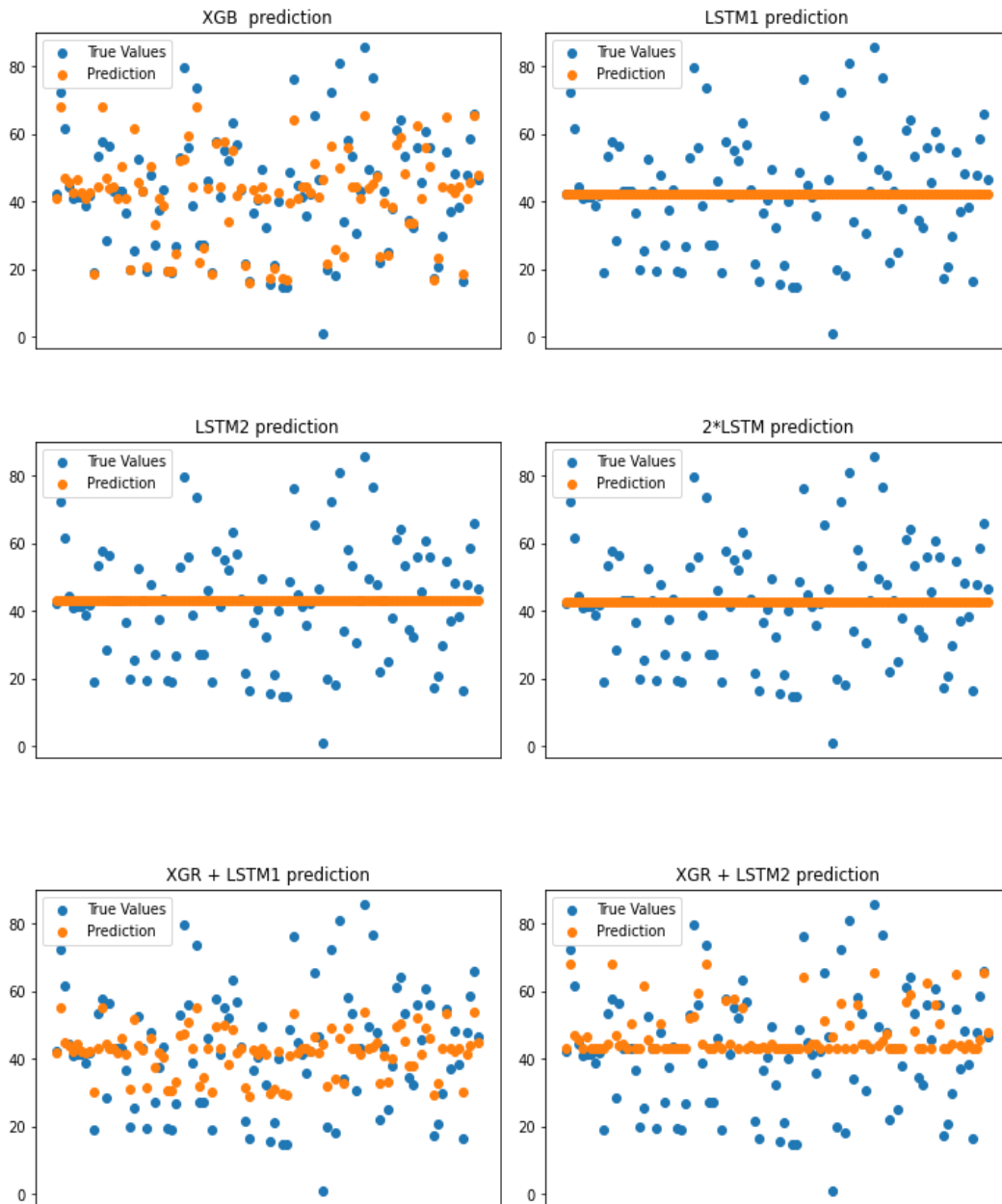
Ανά ημέρα δειγματοληψία

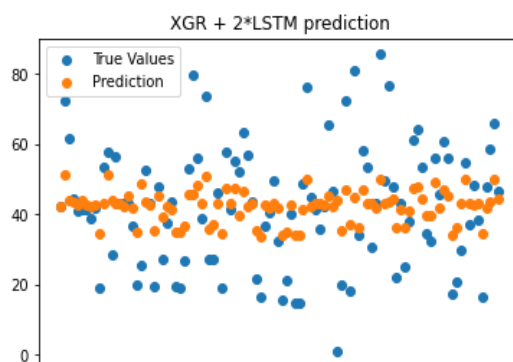
Άθροισμα της κατανάλωσης ανά ημέρα. Για αυτή την υποπερίπτωση θα παρουσιαστεί ο πίνακας με τα σφάλματα και αμέσως μετά τα διαγράμματα.

Models	RMSE(%)	NMBE(%)
XGBOOST	21,748	0,003
LSTM1	36,019	1,565
LSTM2	36,001	1,047
LSTM1 + LSTM2	36,009	1,306
LSTM1 + XGBOOST	25,071	0,78

LSTM2 + XGBOOST	31,766	12,316
LSTM1 + LSTM2 + XGBOOST	28,182	0,872

Πίνακας 5.23: Τιμές Σφαλμάτων του δεύτερου συνόλου δεδομένων με ημερήσια δειγματοληψία για την εύρεση του αθροίσματος της κατανάλωσης στην περίπτωση των "Shuffled" δεδομένων





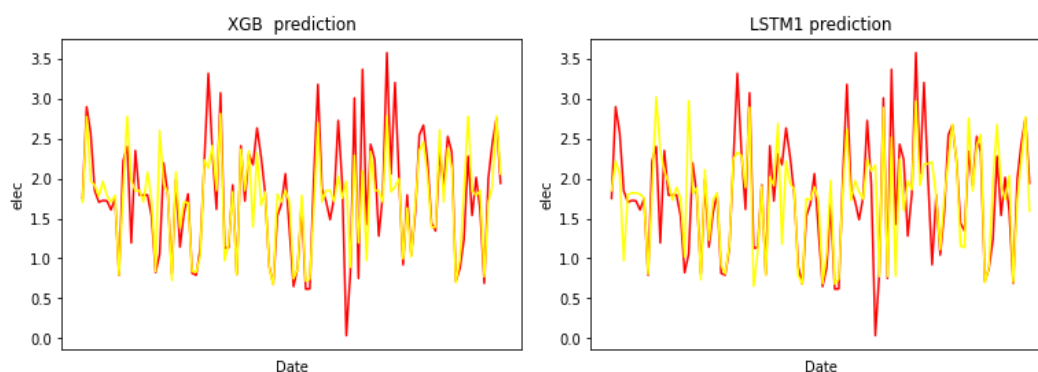
Διαγράμματα αποτελεσμάτων 5.19: Προβλέψεις και πραγματικές τιμές του αθροίσματος της κατανάλωσης για τα διάφορα μοντέλα του δεύτερου συνόλου δεδομένων ανά ημέρα στην “Shuffled” προσέγγιση

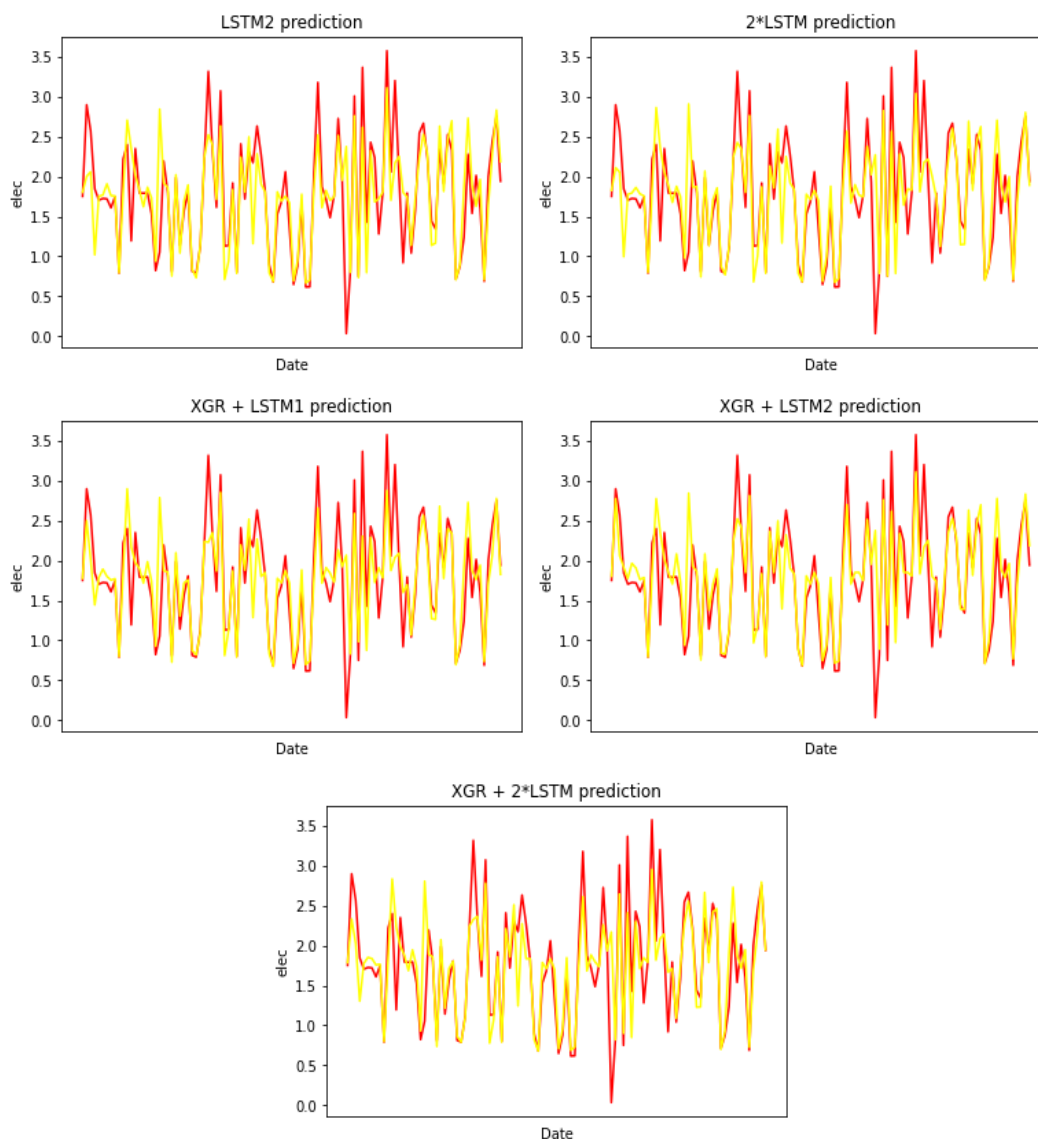
Τα αποτελέσματα των LSTM είναι αρκετά χειρότερα καθώς δεν υπάρχουν επαρκή δεδομένα για καλύτερες προβλέψεις υπό αυτές τις συνθήκες δηλαδή με το άθροισμα της κατανάλωσης.

Μέσος όρος της κατανάλωσης ανά ημέρα. Για αυτή την υποπερίπτωση θα παρουσιαστεί ο πίνακας με τα σφάλματα και αμέσως μετά τα διαγράμματα.

Models	RMSE(%)	NMBE(%)
XGBOOST	20,878	-0,882
LSTM1	25,365	-0,543
LSTM2	26,359	-0,553
LSTM1 + LSTM2	25,725	-0,548
LSTM1 + XGBOOST	22,494	-0,712
LSTM2 + XGBOOST	25,015	3,3
LSTM1 + LSTM2 + XGBOOST	23,620	-0,659

Πίνακας 5.24: Τιμές Σφαλμάτων του δεύτερου συνόλου δεδομένων με ημερήσια δειγματοληψία για την εύρεση του μέσου όρου της κατανάλωσης στην περίπτωση των “Shuffled” δεδομένων





Διαγράμματα αποτελεσμάτων 5.20: Προβλέψεις και πραγματικές τιμές του μέσου όρου της κατανάλωσης για τα διάφορα μοντέλα του δεύτερου συνόλου δεδομένων ανά ημέρα στην “Shuffled” προσέγγιση

Παρατηρείται ότι τα LSTM προβλέπουν εντελώς λάθος δύο κορυφές και αυτό συμβαίνει λόγω του μικρού πλήθους στοιχείων για εκπαίδευση. Επίσης τα αποτελέσματα είναι πολύ καλύτερα σε σχέση με την προηγούμενη περίπτωση.

Σε αυτό το σημείο ολοκληρώνεται η αναλυτική παρουσίαση των πρώτων δύο συνόλων του πίνακα στην αρχή του κεφαλαίου. Για τα υπόλοιπα σύνολα δεδομένων θα παρουσιαστούν μόνο οι πίνακες με τα σφάλματα για τα 7 μοντέλα για όλες τις περιπτώσεις εκτός των δυο που αφορούσαν το άθροισμα της κατανάλωσης καθώς όπως αποδείχθηκε τα LSTM είναι αρκετά αδύναμα στη συγκεκριμένη υποπερίπτωση. Μετά από την παρουσίαση των σφαλμάτων θα ακολουθήσουν τα συμπεράσματα που προέκυψαν από όλα τα πειράματα που διενεργήθηκαν.

5.4 Παρουσίαση των αποτελεσμάτων των υπόλοιπων συνόλων δεδομένων

Η δομή της συνέχειας του κεφαλαίου έχει ως εξής. Πρώτα θα αναγράφεται το μοντέλο και στη συνέχεια θα υπάρχει ο πίνακας με τα σφάλματα για την αντίστοιχη περίπτωση. Στους πίνακες που ακολουθούν θα σημειωθούν οι τιμές σφάλματος για τα υπόλοιπα σύνολα δεδομένων που υπάρχουν στους πίνακες παραπάνω. Η πρώτη στήλη του πίνακα θα έχει το νούμερο που αντιστοιχεί στο νούμερο που έχει το κάθε σύνολο δεδομένων στον πρώτο πίνακα του κεφαλαίου (πίνακας 5.1).

XGBRegressor

	Σειριακή προσέγγιση						"Shuffled" προσέγγιση.					
	Ωριαίο		Ανά 8 ώρες		Ημερήσιο		Ωριαίο		Ανά 8 ώρες		Ημερήσιο	
	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE
3	103,73	-13,68	93,23	-21,52	81,89	-25,27	56,31	2,48	62,58	-6,38	44,83	6,52
4	63,16	-8,86	51,23	-9,49	37,3	0,9	31,32	1,03	26,72	3,7	19,88	0,94
5	48,04	6,39	44,1	4,46	35,47	-2,4	51,48	1,28	50,65	-8,89	22,7	-2,86
6	193,72	26,67	100	20,29	68	-0,09	57,13	-5,06	27,49	-2,44	25,06	-0,54
7	311,55	152,8	300,71	68,2	124,98	45,12	51,01	-1,42	26,48	6,89	14,65	1,27
8	68,08	52,39	69,07	40,16	81,37	-0,34	15,46	0,6	10,61	-0,14	8,33	0

Πίνακας 5.25: Τιμές Σφαλμάτων για τα υπόλοιπα σύνολα δεδομένων για το XGBRegressor

LSTM1

	Σειριακή προσέγγιση						"Shuffled" προσέγγιση					
	Ωριαίο		Ανά 8 ώρες		Ημερήσιο		Ωριαίο		Ανά 8 ώρες		Ημερήσιο	
	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE
3	104,8	-31,74	87,82	-9,02	83,13	-32,25	65,20	-4,25	76,68	-8,63	50,3	10,83
4	71,37	20,52	68,13	-10,12	43,83	-5,04	38,72	-0,55	80,49	-23	46,76	-13,97
5	44,96	-3,34	45,97	35,09	54,47	48,55	65,71	-7,92	28,85	-1,68	21,79	-0,91
6	227,71	85,31	140,91	68,63	106,23	82,67	64,14	-11,07	30,36	1,88	20,81	-0,45
7	215,99	9,7	210,24	78,1	167,22	56,48	53,29	-6,84	50,29	-13,6	56,79	-16,65
8	105,8	84,57	185,53	110,98	237,86	-57,69	18,86	-4,37	27,1	-6,17	26,93	-4,96

Πίνακας 5.26: Τιμές Σφαλμάτων για τα υπόλοιπα σύνολα δεδομένων για το LSTM1

LSTM2

	Σειριακή προσέγγιση						"Shuffled" προσέγγιση.					
	Ωριαίο		Ανά 8 ώρες		Ημερήσιο		Ωριαίο		Ανά 8 ώρες		Ημερήσιο	
	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE
3	66,57	-11,15	121	62	86,81	-41,12	52,41	-1,87	66,80	-9,35	43,01	7,7
4	73,32	5,03	69,5	-19,09	44,39	-9,11	32,56	2,29	31,66	4,11	25,96	7,33
5	46,04	2,15	39,05	11,89	29,42	2,98	53,55	-5,44	28,92	-3,01	20,3	-1,18
6	242,16	-84,94	145,3	57,95	69,96	28,78	60,94	-9,19	28,57	-1,31	30,55	-10,31
7	252,74	-99,91	190,36	80,98	98,87	35,43	72,92	13	39,86	5,96	71,77	-12,98
8	69,02	3,76	69,34	-28,72	267,9	-78,24	46,44	9,59	21,87	8,6	44,09	6,27

Πίνακας 5.27: Τιμές Σφαλμάτων για τα υπόλοιπα σύνολα δεδομένων για το LSTM2

LSTM1 + LSTM2

	Σειριακή προσέγγιση						"Shuffled" προσέγγιση.					
	Ωριαίο		Ανά 8 ώρες		Ημερήσιο		Ωριαίο		Ανά 8 ώρες		Ημερήσιο	
	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE
3	70,16	-9,89	94,74	26,48	84,81	-36,69	54,79	-3,06	65,42	-8,95	44,4	9,27
4	64,82	12,77	68,67	-14,6	44,06	-7,07	33,44	0,86	42,5	-9,46	30,15	-3,32
5	43,46	-0,59	47,54	37,36	36,41	25,76	57,26	-6,68	27,66	-2,35	20,09	-1,04
6	199,29	0,18	140,7	63,29	84,46	55,72	61,49	-10,13	28,85	0,28	20,53	-5,38
7	226,7	-45,1	203,42	76,2	120,95	67,07	55,78	3,08	40,06	-3,82	55,74	-14,82
8	77,72	44,16	95,55	57,89	250,73	-42,78	27,22	2,6	22,37	1,21	28,29	0,65

Πίνακας 5.28: Τιμές Σφαλμάτων για τα υπόλοιπα σύνολα δεδομένων για το LSTM1+ LSTM2

LSTM1 + XGBRegressor

	Σειριακή προσέγγιση						"Shuffled" προσέγγιση.					
	Ωριαίο		Ανά 8 ώρες		Ημερήσιο		Ωριαίο		Ανά 8 ώρες		Ημερήσιο	
	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE
3	67,44	-7,51	89,65	-15,27	81,93	-28,76	57,57	-0,88	62,1	-7,47	45,5	8,68
4	62,49	5,82	53,3	-9,8	39,73	-2,07	33,02	0,23	43,38	-9,66	30,02	-6,51

5	43,35	1,52	45,1	33,65	35,38	23,07	54,6	-3,31	37,1	-5,28	20,41	-1,88
6	202,1	55,99	108,19	44,46	77,16	41,29	56,63	-8,06	25,98	-0,27	21,48	-0,5
7	246,29	81,25	250,69	85,02	134,83	80,71	51,41	-4,13	32,31	-3,35	31,08	-7,69
8	86,04	68,2	122,25	108,02	256,34	-84,92	15,53	-1,88	15,61	-3,15	13,52	-2,48

Πίνακας 5.29: Τιμές Σφαλμάτων για τα υπόλοιπα σύνολα δεδομένων για το LSTM1+ XGBRegressor

LSTM2 + XGBRegressor

	Σειριακή προσέγγιση						"Shuffled" προσέγγιση.					
	Ωριαίο		Ανά 8 ώρες		Ημερήσιο		Ωριαίο		Ανά 8 ώρες		Ημερήσιο	
	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE
3	63,01	-0,6	119,2	71,1	80,99	-23,35	55,14	12,68	64,18	0,4	47,93	16,52
4	64,66	27,37	47,41	6,45	35,87	3,42	32,72	8,24	32,1	10,71	26,02	9,77
5	47,55	16,85	42,78	17,94	32,88	8,92	51,56	7,75	36,87	3,66	20,21	3,55
6	193,72	26,67	147,31	70,9	74,75	36,32	56,18	5,36	27,41	5,87	23,47	6,17
7	311,55	152,8	234,78	86,05	112,65	78,26	69,27	27,98	32,7	15,7	42,78	25,89
8	68,08	52,39	69,07	40,16	259,84	-76,54	41,92	21,05	21,68	10,98	35,44	23,29

Πίνακας 5.30: Τιμές Σφαλμάτων για τα υπόλοιπα σύνολα δεδομένων για το LSTM2+ XGBRegressor

LSTM1 + LSTM2 + XGBRegressor

	Σειριακή προσέγγιση						"Shuffled" προσέγγιση.					
	Ωριαίο		Ανά 8 ώρες		Ημερήσιο		Ωριαίο		Ανά 8 ώρες		Ημερήσιο	
	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE	RMSE	NMBE
3	66,18	-8,72	89,38	10,48	83,2	-32,88	52,58	-1,21	62,94	-8,1	42,82	8,35
4	60,36	5,56	57,5	-12,9	41,02	-4,41	31,22	0,91	32,55	-5,07	25,35	-1,89
5	42,84	1,73	41,01	26,4	31,14	16,37	52,54	-4,02	32,74	-4,53	19,8	-1,65
6	195,61	9,01	116,64	48,96	72,87	37,12	56,84	-8,44	26,15	-0,62	19,3	-3,77
7	231	20,86	210,89	76,09	118,03	67,43	52,96	1,58	32,17	-0,24	38,22	-9,45
8	73,03	46,91	89,02	65,42	251,43	-16,43	21,56	1,94	16,29	0,76	18,98	0,43

Πίνακας 5.31: Τιμές Σφαλμάτων για τα υπόλοιπα σύνολα δεδομένων για το συνδυασμό των τριών μοντέλων

5.5 Αποτελέσματα των αλγορίθμων σύγκρισης σε όλα τα σύνολα δεδομένων

Τέλος ακολουθούν δύο πίνακες που περιέχουν τα αποτελέσματα των ίδιων περιπτώσεων αλλά με τη χρήση απλών αλγορίθμων μηχανικής μάθησης καθώς και ένα αλγόριθμο που βασίζεται στις βαθμομέρες θέρμανσης και ψύξης (ΒΘΨ) υπολογίζοντας μόνο την CVRMSE(%) ώστε να τη συγκρίνουμε με τα από πάνω αποτελέσματα. Πρώτα ο πίνακας με τον τελευταίο αλγόριθμο και το random forest και αμέσως μετά ο πίνακας με το Linear Regression και τους κ-κοντινότερους γείτονες.

	Σειριακή προσέγγιση						"Shuffled" προσέγγιση.					
	Ωριαίο		Ανά 8 ώρες		Ημερήσιο		Ωριαίο		Ανά 8 ώρες		Ημερήσιο	
	RF	ΒΘΨ	RF	ΒΘΨ	RF	ΒΘΨ	RF	ΒΘΨ	RF	ΒΘΨ	RF	ΒΘΨ
1	137,74	57,2	>250	48,3	245,08	25,09	343,02	51,33	>250	45,39	>250	29,08
2	245,47	95,82	>250	72,35	>250	48	>250	88,66	>250	69,19	>250	45,64
3	131,11	117,59	177,2	107,44	170,85	82	268,25	111,74	>250	101,36	>250	65,69
4	160,67	93,26	178,1	88,64	149,72	57,88	>250	82,04	>250	76,11	>250	41,88
5	80,05	101,81	81,74	81,33	79,12	54,5	208,87	102,01	>250	67,76	>250	61,78
6	186,2	135,62	141,2	190,59	95,52	78,02	>250	132,21	>250	100,93	>250	77,45
7	>250	>250	>250	>250	>250	>250	59,36	49,82	>250	51,04	>250	49,31
8	73,38	69,21	113,1	72,89	115,42	72,8	>250	37,02	>250	34,29	>250	39,54

	Σειριακή προσέγγιση						"Shuffled" προσέγγιση.					
	Ωριαίο		Ανά 8 ώρες		Ημερήσιο		Ωριαίο		Ανά 8 ώρες		Ημερήσιο	
	KNN	LR	KNN	LR	KNN	LR	KNN	LR	KNN	LR	KNN	LR
1	86,19	105,47	150,27	>250	81,78	448	201,93	160,14	>250	>250	>250	>250
2	239,32	118,55	>250	>250	>250	>250	161,05	98,83	>250	>250	>250	>250
3	101,07	108,64	101,54	188,3	108,17	176,71	156,22	90,4	>250	>250	>250	>250
4	116,82	103,9	97,54	183,5	88,26	126,9	>250	>250	>250	>250	>250	>250
5	77,09	113,62	91,5	111,8	96,15	69,1	154,29	121,09	>250	>250	>250	>250
6	164,6	180,05	74,1	126,4	85,42	93,51	>250	141,55	>250	>250	>250	>250
7	>250	>250	>250	>250	>250	>250	54,64	73	>250	>250	>250	>250
8	64,18	63,99	94,78	108,2	114,62	111,8	>250	>250	>250	>250	>250	>250

Πίνακας 5.26: Τιμές Σφαλμάτων για τα όλα τα σύνολα δεδομένων στα μοντέλα σύγκρισης.

5.6 Συμπεράσματα για τα αποτελέσματα

Όπως φαίνεται σε σχέση με τους παραπάνω πίνακες τα μοντέλα που επιλέχθηκαν έχουν πολύ καλύτερα αποτελέσματα από τα κλασικά σχεδόν όλες τις φορές. Εκτός από τον αλγόριθμο παλινδρόμησης των βαθμοημερών οι άλλοι τρεις αλγόριθμοι δεν πλησιάζουν ούτε κοντά σε αυτούς που αναλύθηκαν παραπάνω. Ο ένας που κάπως πλησιάζει συνήθως δεν προσπερνά σε απόδοση τους προτεινόμενους αλλά ακόμα και όταν αυτό συμβαίνει υπάρχει πιθανότητα να οφείλεται στην τυχαιότητα της επιλογής των συνόλων εκπαίδευσης και εξέτασης καθώς συμβαίνει ελάχιστες φορές.

Πέρα από αυτό εμφανίζεται μεγάλη διακύμανση στα μοντέλα που πετυχαίνουν το καλύτερο αποτέλεσμα. Από συνδυασμό και των τριών μέχρι και μεμονωμένα όλα τα μοντέλα που προτάθηκαν(εξαιρούνται δηλαδή αυτά που χρησιμοποιήθηκαν μόνο ως μέτρο σύγκρισης) είχαν το μικρότερο σφάλμα σε κάποια περίπτωση. Αυτή η πληθώρα υποδηλώνει ότι το τελικό αποτέλεσμα είναι πάντα πλήρως εξαρτημένο από το σύνολο δεδομένων και το μοντέλο με το οποίο αυτό θα ταιριάξει. Κανένα μοντέλο δεν είχε μόνιμα την μεγαλύτερη ακρίβεια, αφού ακόμα και αν τις περισσότερες φορές το μικρότερο σφάλμα το παρουσίαζε ένας εκ των συνδυασμών, όχι πάντα ο ίδιος, υπήρξαν φορές που και τα απλά μοντέλα είχαν την καλύτερη απόδοση.

Με τα παραπάνω πειράματα προέκυψαν κάποια σημαντικά συμπεράσματα για τα μοντέλα και τον τρόπο με τον οποίο αυτά αλληλεπιδρούν με τα διαφορά σύνολα δεδομένων. Το κάθε σύνολο δεδομένων προσέφερε νέες παρατηρήσεις ή απέδειξε κάποιες που είχαν προκύψει από τα πρώτα σύνολα δεδομένων αλλά τελικά είχαν γενικότερη ισχύ.

Ειδικότερα για τα μοντέλα τα ίδια φάνηκε ότι όταν μειώνεται ο ρυθμός δειγματοληψίας επιτυγχάνεται μεγαλύτερη ακρίβεια στις προβλέψεις παρά το γεγονός ότι μειώνεται το πλήθος των δειγμάτων του συνόλου εκπαίδευσης. Ο λόγος που συμβαίνει αυτό είναι ότι μειώνεται τυχόν θόρυβος των αρχικών δεδομένων και άρα εκπαιδεύονται πολύ καλύτερα τα μοντέλα. Με τα μεγαλύτερα διαστήματα μεταξύ των στοιχείων που τροφοδοτούνται στο εκάστοτε μοντέλο αποφεύγονται τυχόν ακραίες τιμές που έχουν εισχωρήσει στο σύνολο από λάθος καθώς τέτοιες τιμές ουσιαστικά μετριάζονται όταν συνδυάζονται με τις γειτονικές τους.

Για τα LSTM παρατηρήθηκε ακόμα ότι δε μπορούν εύκολα να εντοπίσουν διακυμάνσεις που έχουν προκύψει από πρόσθεση άλλων στοιχείων. Ιδίως το μοντέλο χωρίς το σύνολο αξιολόγησης δεν μπορούσε να προβλέψει καθόλου το άθροισμα της κατανάλωσης σε ένα διάστημα την ώρα που το XGBoost πετύχαινε αρκετές φορές ακόμα και καλά αποτελέσματα.

Ζωτικό ρόλο στα αποτελέσματα έπαιζε και το ίδιο το σύνολο δεδομένων όπως προαναφέραμε καθώς τα σφάλματα που εμφανίστηκαν είχαν τιμές από 25% μέχρι και πάνω από 100% πράγμα που δείχνει ότι οι συναρτήσεις σφάλματος που χρησιμοποιήθηκαν ήταν πολύ ευαίσθητες στο λάθος και κοστολογούσαν πολύ ακριβά μια πρόβλεψη με μεγάλη απόκλιση από την πραγματική τιμή. Τις

περισσότερες φορές παρατηρήθηκε ότι η τυχαία επιλογή δεδομένων για τη δημιουργία των συνόλων εκπαίδευσης και εξέτασης είχε πολύ καλύτερα αποτελέσματα με τα σφάλματα σε ορισμένες περιπτώσεις να πέφτουν ακόμα και πάνω από 2 φορές. Ιδίως στην περίπτωση που στόχος ήταν η πρόβλεψη της θερμικής κατανάλωσης, η σειριακή προσέγγιση δεν είχε καθόλου καλά αποτελέσματα ενώ τις περισσότερες φορές δεν προέβλεπε καθόλου σωστά τις τιμές. Ουσιαστικά ο μόνος τρόπος να υπάρξουν ορθές προβλέψεις για αυτά τα σύνολα δεδομένων ήταν η χρήση της δεύτερης προσέγγισης η οποία μάλιστα σε ορισμένες περιπτώσεις άγγιξε και μονοψήφιο ποσοστό σφάλματος την ώρα που το αποδεκτό όριο είναι στο 15% για την CV(RMSE).

Στα πιο γενικά αποτελέσματα φάνηκε ότι η πρόβλεψη της ενεργειακής κατανάλωσης είναι πιο απλή υπόθεση και τα μοντέλα είχαν αρκετά μεγαλύτερη ακρίβεια σε σχέση με την ακρίβεια που είχαν στην περίπτωση πρόβλεψης της θερμικής κατανάλωσης όταν προσεγγιζόταν το πρόβλημα σειριακά αφού αυτό ήταν αδύνατο για τα δεύτερα. Αυτό ίσως είναι και αποτέλεσμα των περισσότερων παραγόντων που μπορούν να επηρεάσουν τη θερμοκρασία ενός κτιρίου και δεν είναι δυνατόν να προβλεφθούν.

Ταυτόχρονα έγινε διακριτό ότι χαρακτηριστικά με μικρή συσχέτιση μπορούν να παίξουν σημαντικό ρόλο στην ποιότητα μιας πρόβλεψης σε αντίθεση με αλλά που ακόμα και αν έχουν μεγάλη συσχέτιση με τον στόχο δεν μπορούν πραγματικά να εκπαιδεύσουν το μοντέλο. Χαρακτηριστικό παράδειγμα είναι το feature Holiday το οποίο πάντα είχε μικρή συσχέτιση με την κατανάλωση όμως όταν δεν χρησιμοποιούνταν στην εκπαίδευση των μοντέλων τα αποτελέσματα αυτών ήταν από 2-4% χειρότερα, μεγάλη διαφορά αν αναλογιστεί κανείς ότι είναι της τάξεως του 5-10%.

Στο κεφάλαιο αυτό παρουσιάστηκαν τα σημαντικότερα αποτελέσματα των πειραμάτων που έγιναν για την αξιολόγηση των μοντέλων και τον εντοπισμό της δυναμικής τους. Παρουσιάστηκαν τα σύνολα δεδομένων και τα διαγράμματα τους, μετρήθηκαν τα σφάλματα των μοντέλων στις διάφορες περιπτώσεις και παρουσιάστηκαν τα διαγράμματα των προβλέψεων σε σχέση με το σύνολο εξέτασης.

6. Επίλογος και Μελλοντικές προσεγγίσεις

6.1 Επίλογος

Στην παρούσα διπλωματική διερευνήθηκαν λύσεις για το πρόβλημα της ακριβούς πρόβλεψης της ενεργειακής κατανάλωσης των συστημάτων που προϋπήρχαν των δράσεων ενεργειακής ανακαίνισης που πραγματοποιήθηκαν στα κτίρια που πραγματεύεται. Αρχικά παρουσιάστηκαν τα βασικά θεωρητικά κομμάτια για την κατανόηση του προβλήματος και της λύσης. Εξηγήθηκαν τα σημαντικότερα πρωτόκολλα και τονίστηκαν οι βασικές αρχές τους ενώ παρουσιάστηκε και η έννοια της ενεργειακής αναβάθμισης αλλά και οι τρόποι που αυτή μπορεί να επιτευχθεί.

Στη συνέχεια προβλήθηκαν οι λύσεις από άλλες προϋπάρχουσες έρευνες που βοήθησαν ουσιαστικά στην εύρεση και στην διατύπωση της παρούσας λύσης. Με αρκετές που πραγματεύονται αντίστοιχα μοντέλα αλλά και άλλες που απλώς έδιναν μια νέα ιδέα στον τρόπο αξιοποίησης των δεδομένων και στην εξαγωγή χρήσιμων χαρακτηριστικών από αυτά. Μέσα από συνδυασμό αυτών προήλθε και η υλοποίηση της λύσης που πραγματεύεται η παρούσα διπλωματική εργασία

Η λύση αυτή αποτελείται από συνδυασμό βαθιάς μάθησης και ενός αλγορίθμου μηχανικής μάθησης και συγκεκριμένα χρησιμοποιήθηκαν δυο τύποι LSTM και το XGBoost. Αυτά χρησιμοποιήθηκαν τόσο μεμονωμένα όσο και συνδυαστικά με στόχο την εύρεση του καλύτερου μοντέλου. Τέλος παρουσιάστηκαν τα αποτελέσματα για τα διάφορα σύνολα δεδομένων πάνω στα οποία εκπαιδεύτηκαν τα μοντέλα και συγκρίθηκαν με τα αποτελέσματα πιο παραδοσιακών αλγορίθμων που επίσης εκπαιδεύτηκαν στα ίδια σύνολα.

Με τα αποτελέσματα όπως παρουσιάστηκαν προκύπτουν τα εξής συμπεράσματα:

A) Τα LSTM είναι ικανά να πετύχουν τις προβλέψεις με μεγάλη ακρίβεια αρκεί να έχουν μεγάλο πλήθος δεδομένων για εκπαίδευση ή/και αυτό να μοιάζει αρκετά με το σύνολο εξέτασης. Υστερούν σημαντικά όταν χρειάζεται μεγάλη προσαρμογή σε δεδομένα εξέτασης τα οποία απέχουν σημαντικά από το μέσο όρο του συνόλου εκπαίδευσης. Αντίθετα το XGBoost έδειξε ότι μπορεί να ανταπεξέλθει ακόμα και σε αυτές τις συνθήκες.

B) Όταν υπάρχουν πολλά δεδομένα τα οποία είναι χρονικά πολύ κοντά δημιουργούν θόρυβο λόγω των πολύ σύντομων διακυμάνσεων. Αυτός ο θόρυβος αποπροσανατολίζει τις προβλέψεις και για αυτό το λόγο όταν απομακρύνεται με την εκ νέου δειγματοληψία υπάρχουν πολύ καλύτερα αποτελέσματα.

Γ) Η τυχαία επιλογή και διαμόρφωση των συνόλων εκπαίδευσης και εξέτασης είναι απαραίτητη διαδικασία στην προσπάθεια πρόβλεψης της θερμικής κατανάλωσης ενός κτιρίου ενώ είναι αρκετά πιο αποδοτική στην περίπτωση πρόβλεψης της ηλεκτρικής. Αυτό συμβαίνει επειδή η θερμική κατανάλωση κάποιες περιόδους μπορεί να τείνει στο μηδέν για λόγους που δε μπορούν να εντοπιστούν αν δεν γίνει αυτό το

μοίρασμα των τιμών στα δύο υποσύνολα, καθώς δύναται να υπάρξουν περίοδοι όπου ένα κτίριο για κάποιες μέρες δεν καταναλώνει θερμική ενέργεια. Όταν συμβαίνει αυτό είναι αδύνατο να λειτουργήσει σωστά και να πετύχει προβλέψεις η σειριακή προσέγγιση.

Δ) Τα καλύτερα χαρακτηριστικά δεν είναι πάντα αυτά με τη μεγαλύτερη συσχέτιση ούτε αυτά που έχουν πολλές τιμές. Μπορεί ένα χαρακτηριστικό που εμφανίζεται λίγες φορές στο σύνολο δεδομένων (πχ 1/100 ημέρες) να προσφέρει πληροφορίες στο μοντέλο που να αλλάζουν σε μεγάλο βαθμό την εκπαίδευση του και άρα την ικανότητα να επιτύχει καλύτερα αποτελέσματα.

Καταλήγοντας πρέπει να επαναληφθεί ότι δεν μπορεί να υπάρξει ένα μοντέλο σαφώς καλύτερο από τα άλλα καθώς το κάθε ένα από αυτά αλληλεπιδρά καλύτερα με συγκεκριμένα σύνολα δεδομένων. Οι συνδυασμοί πιο απλών μοντέλων βαθιάς μάθησης με αλγορίθμους μηχανικής μάθησης είναι συνήθως πιο αποδοτικοί και χρονικά δεν είναι απαγορευτική η χρήση τους.

6.2 Μελλοντικές προσεγγίσεις

Μελλοντικά υπάρχουν δυο παρακλάδια που αξίζουν εξερεύνησης. Το πρώτο είναι να γίνει η ίδια διαδικασία με μεγαλύτερο πλήθος δεδομένων ή και συνδυασμό πολλών συνόλων δεδομένων. Συγκεκριμένα τα LSTM έχουν δείξει την τάση να εκπαιδεύονται καλύτερα με πολλά δεδομένα εκπαίδευσης καθώς η δομή τους είναι αυτή που χρειάζεται μεγάλο μέγεθος συνόλου εκπαίδευσης για να έχει μεγάλη απόδοση στην επιτυχία των προβλέψεων. Ο συνδυασμός συνόλων δεδομένων εκτός από το γεγονός ότι πετυχαίνει την αύξηση του συνόλου δεδομένων μπορεί να οδηγήσει σε καλύτερη εκπαίδευση και όσων αφορά τις τιμές αφού οι ακραίες τιμές θα είναι πιο διαχειρίσιμες από το σύστημα. Ταυτόχρονα αν τα σύνολα είναι από κτίρια με αντίστοιχη λειτουργία (πχ γραφεία) η εκπαίδευση θα γίνει ακόμα πιο αποτελεσματική αφού η συμπεριφορά τους θα είναι αντίστοιχη ενώ τυχόν ανωμαλίες θα διορθώνονται από τα άλλα κτίρια. Η δεύτερη εκδοχή είναι να δοκιμαστεί άλλος συνδυασμός νευρωνικού δικτύου μαζί με το XGBoost και συγκεκριμένα η χρήση του transformer το οποίο αποτελεί ένα νέο και πιο ικανό νευρωνικό δίκτυο. Με τη σωστή εκπαίδευση του ίσως να καταφέρει να πετύχει καλύτερα αποτελέσματα από το LSTM.

Παράρτημα Α: Παρουσίαση πηγαίου κώδικα

Ακολουθούν τα σημαντικότερα κομμάτια του κώδικα ακολουθούμενα από μια μικρή επεξήγηση για το κάθε ένα. Ολόκληρος ο κώδικας βρίσκεται στο github στη διεύθυνση <https://github.com/chrissiaf/-Energy-saving-profit-forecasting-.git>

Πρώτα θα παρουσιαστεί ο κώδικας που χρησιμοποιήθηκε ώστε να έρθουν τα σύνολα δεδομένων σε κατάλληλη μορφή. Στις επόμενες εικόνες φαίνονται καθαρά τόσο οι διαδικασίες σύνδεσης των διαφόρων αρχείων σε ένα όσο και οι αλλαγές που έγιναν στα δεδομένα τους για να διορθωθούν προβλήματα όπως αυτό των χαμένων δεδομένων χωρίς να διαταραχθεί το γενικότερο πλαίσιο του συνόλου αυτού.

```
def datael(a):
    filesel = os.listdir(path + a)

    #filesel = [f for f in files if f[-3:] == '.xls']

    #print(os.path.join(path + a + "*.xlsx"))
    print(filesel)
    trainel = pd.DataFrame()
    for file in filesel:
        # reading excel files
        k = pd.read_excel(file, names=["electricity", "Date / Time"])
        print(k.head())
        trainel = pd.concat([trainel, k], ignore_index=True)
        print(trainel.tail())
    trainel["Date / Time"] = pd.to_datetime(trainel["Date / Time"])
    trainel["Date / Time"] = trainel["Date / Time"].round('H')

    return trainel
def missingdate(b):
    l = []
    k=0
    for i in b["Date / Time"]:
        if pd.isna(i)==True:
            l.append(k)
        else:
            k=i
    print(l)
    count=[]
    for i in range(0, len(l)):
        if l[i] !=l[i-1]:
            count.append(1)
        elif l[i] ==l[i-1] and i>0:
            count[-1] += 1

    res = []
    [res.append(x) for x in l if x not in res]
    #print(res)
    n =0
    tr=0
    poll=[]
    for p in range(0, len(count)):
        arf = count[p]
        tip = res[tr]
        for i in range(0, arf):
            tip += pd.Timedelta(hours=1)
            poll.append(tip)

        tr +=1

    b = b[b["Date / Time"].notna()]
    df = pd.DataFrame(poll, columns = ["Date / Time"])
    df["electricity"] = np.nan
    b = pd.concat([b, df])
    b = b.sort_values(by="Date / Time").reset_index(drop=True)
    b["electricity"] = b["electricity"].interpolate()
    return b

temperature = temperature.sort_values(by="Date / Time").reset_index(drop=True)
def mixer(a):
    trainelalla = a.join(temperature.set_index("Date / Time", on="Date / Time"))
    trainelalla = trainelalla.iloc[:, [1,2,0]]
    trainelalla.rename(columns = {'avg': 'temp'}, inplace = True)
    trainelalla["temp"] = trainelalla["temp"].interpolate()

    return trainelalla
def savecsv(a, trainelalla):
    trainelalla.to_csv("electricity-temp-" + a + ".csv")

prpr = datael("Gulbene_region_Gulbene_kindergarten_nr.3")
prpr = missingdate(prpr)
prpr.isnull().sum()
all1 = mixer(prpr)

savecsv("gulbene-region-tirzas-primary-school-building1", all1)
```

Εικόνα Α.1 και Α.2

Εδώ φαίνονται ξεκάθαρα οι συναρτήσεις με τις οποίες γινόταν το διάβασμα καθώς και η συμπλήρωση των τιμών που έλειπαν, ενώ στη δεξιά εικόνα φαίνεται και η διαδικασία συγκέντρωσης όλων των πληροφοριών σε ένα αρχείο.

Ακολουθούν οι αλγόριθμοι που χρησιμοποιήθηκαν για τη σύγκριση με τα μοντέλα που προτείνονται.

```

lab = preprocessing.LabelEncoder()
ytrain_transformed = lab.fit_transform(ytrain) #encoding of the dataset to get interger values as labels for all
ytest_transformed = lab.fit_transform(ytest) #the different values of electricity

rf= RandomForestClassifier()
lr = LogisticRegression()
knn = KNeighborsClassifier()
Y = np.mean(ytest)
Y2 = np.mean(ytest_transformed)

rf.fit(xtrain,ytrain_transformed)
ypredrf = rf.predict(xtest)
rmserf = sqrt(mean_squared_error(ypredrf, ytest_transformed))
cvrmsrf= (rmserf/Y2)*100
print("The CV(RMSE) for the rf is ",cvrmsrf,"%")

knn.fit(xtrain,ytrain_transformed)
ypredknn = knn.predict(xtest)
rmseknn = sqrt(mean_squared_error(ypredknn, ytest_transformed))
cvrmsknn = (rmseknn/Y2)*100
print("The CV(RMSE) for the knn is ",cvrmsknn,"%")

lr.fit(xtrain,ytrain_transformed)
ypredlr = lr.predict(xtest)
rmselr = sqrt(mean_squared_error(ypredlr, ytest_transformed))
cvrmslr = (rmselr/Y2)*100
print("The CV(RMSE) for the lr is ",cvrmslr,"%")

```

```

def estimate_coef(x, y):
    # number of observations/points
    n = np.size(x)

    # mean of x and y vector
    m_x = np.mean(x)
    m_y = np.mean(y)

    # calculating cross-deviation and deviation about x
    SS_xy = np.sum(y*x) - n*m_y*m_x
    SS_xx = np.sum(x*x) - n*m_x*m_x

    # calculating regression coefficients
    b_1 = SS_xy / SS_xx
    b_0 = m_y - b_1*m_x

    return (b_0, b_1)

b = estimate_coef(x,y)
def calcdegd(a,b,testx):
    ypred = a+b*testx
    return ypred

```

Εικόνα A.3 και A.4

Εδώ φαίνονται και οι τέσσερις αλγόριθμοι των οποίων τα αποτελέσματα παρουσιάζονται στους πίνακες 5.26

Στη συνέχεια είναι τα τρία βασικά μοντέλα της διπλωματικής και ο τρόπος με τον οποίο αυτά χρησιμοποιήθηκαν και εκπαιδεύτηκαν.

XGBOOST

```

import xgboost as xgb
reg = xgb.XGBRegressor(booster='gbtree',
                       n_estimators=1000,early_stopping_rounds=210,
                       objective='reg:linear', max_depth=8, learning_rate=0.02)
reg.fit(xtrain32, ytrain,
        eval_set=[(xtrain32, ytrain),
                  (xtest32, ytest)], verbose=100)
ap = reg.predict(xtest32)
rmseap = mean_squared_error(ap, ytest,squared=False)
print('Test RMSE: %.3f' % rmseap)
cvrmsap = (rmseap/Y)*100
print("The CV(RMSE) for the XGBRegressor is ",cvrmsap,"%")
print("The NMBE for the XGBRegressor is ",NMBE(ap,ytest),"")
## plot of the test set combing true and predicted values

```

Εικόνα A.5

LSTM1

```
model = Sequential()
model.add(Input(shape=(train_X.shape[1], train_X.shape[2])))# input layer with the shape of the training set
model.add(Bidirectional(LSTM(80, return_sequences=True, name='LSTM4')))# bidirectional lstm with returning sequences so we can add another lstm
model.add(LSTM(40, return_sequences=False, name='LSTM3'))
model.add(Dense(units=1, activation='linear', name='Dense2'))
model.build()
model.summary()
model.compile(
    loss="mae", # error mean absolute
    optimizer=keras.optimizers.Adam(learning_rate=0.01, decay=0.01), #adam optimizer
    metrics= [RootMeanSquaredError()],)

batch_size = [64]
model.fit(train_X,
          train_y,
          epochs=50,
          batch_size=32,
          verbose=1, shuffle=False)

yhat1 = model.predict(xtest)
rmse1 = mean_squared_error(yhat1, ytest, squared=False)
print('Test RMSE: %.3f' % rmse1)
cvrmse1 = (rmse1/Y)*100
print("The CV(RMSE) for the lstm1 is ", cvrmse1, "%")
print("The NMBE for the lstm is ", NMBE(yhat1, ytest), "%")
```

Εικόνα A.6

LSTM2

```
### second lstm model without using metrics on the test set
model2 = Sequential()
model2.add(Input(shape=(train_X.shape[1], train_X.shape[2])))
model2.add(LSTM(80, return_sequences=True, name='LSTM2'))
model2.add(Bidirectional(LSTM(80, return_sequences=False, name='LSTM4'))
model2.add(Dense(units=10, activation='relu', name='Dense2'))
model2.add(Dense(units=1, activation='linear', name='Dense3'))
model2.build()

model2.summary()
model2.compile(
    loss="mae",
    optimizer=keras.optimizers.Adam(learning_rate=0.01, decay=0.001),
    metrics= [RootMeanSquaredError()],)

cp = ModelCheckpoint('model2/', save_best_only=True) # saves the parameter of the best model so you dont have to train it every time
batch_size = [64]
model2.fit(train_X,
          train_y,
          epochs=50,
          batch_size=64,
          validation_data=(test_X, test_y), # validation set so we can train the model in a dataset that is not known to the model
          verbose=1, callbacks=[cp], shuffle=False)

from tensorflow.keras.models import load_model
model2 = load_model('model2/')
print(model2.summary())
print(xtest.shape)
yhat2 = model2.predict(xtest)
print(yhat2.shape)
rmse2 = mean_squared_error(yhat2, ytest, squared=False)
print('Test RMSE: %.3f' % rmse2)
cvrmse2 = (rmse2/Y)*100
print("The CV(RMSE) for the lstm is ", cvrmse2, "%")
print("The NMBE for the lstm is ", NMBE(yhat2, ytest), "%")
```

Εικόνα A.7

Τέλος ακολουθούν κάποιες εντολές που είναι αρκετά σημαντικές για τον υπόλοιπο κώδικα.

```
#xtrain = train[X]
xtraina, xtesta, ytrain, ytest = train_test_split(trainpd, ytrain, test_size=0.2, random_state=25, shuffle = True)
```

Εικόνα A.8

Με αυτή την εντολή γίνεται η δημιουργία των τυχαίων συνόλων εκπαίδευσης και εξέτασης.

Έπεται η εντολή που αυξομειώνει τον ρυθμό των δεδομένων και υπολογίζει το μέσο όρο τους στο επιθυμητό διάστημα.

```
trainpd= trainpd.resample('D').mean()
#trainpd["electricity"] = trainpd2["electricity"]
```

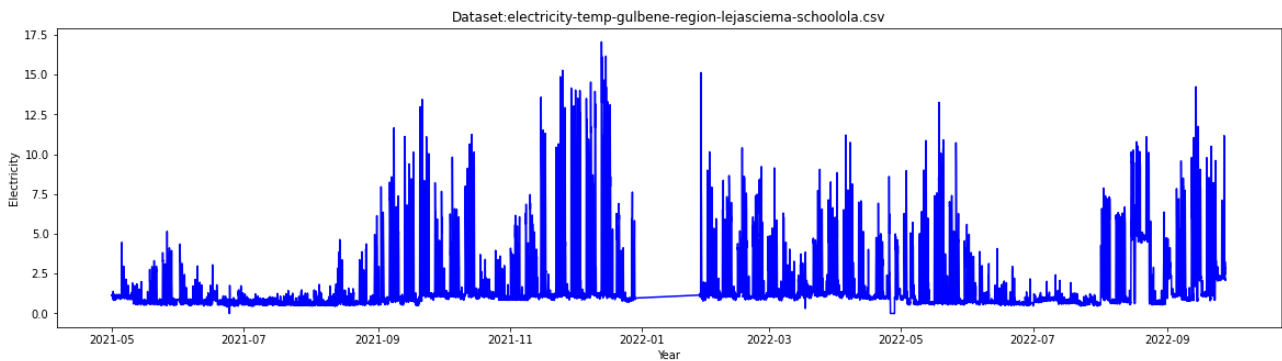
Εικόνα A.9

Παράρτημα Β: Διαγράμματα από τα σύνολα δεδομένων του πίνακα 5.25

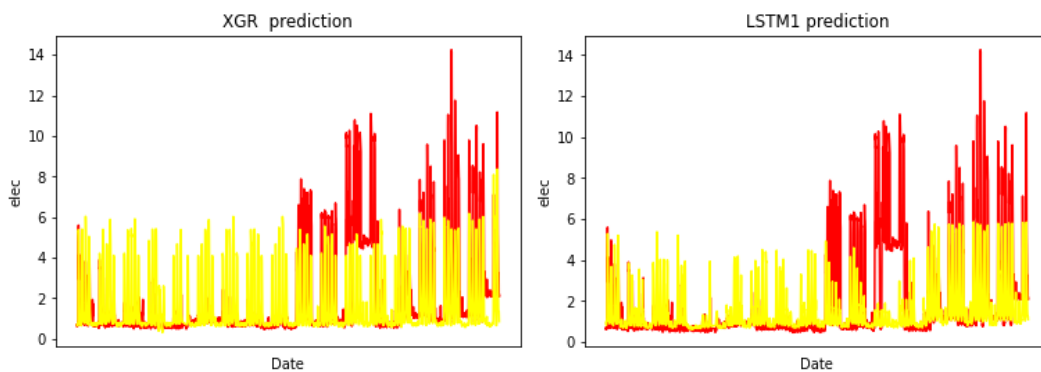
Στο παράρτημα αυτό θα παρουσιαστούν ενδεικτικά κάποια από τα διαγράμματα για τα σύνολα δεδομένων του πίνακα 5.25

Gulbene_region_Lejasciema_school

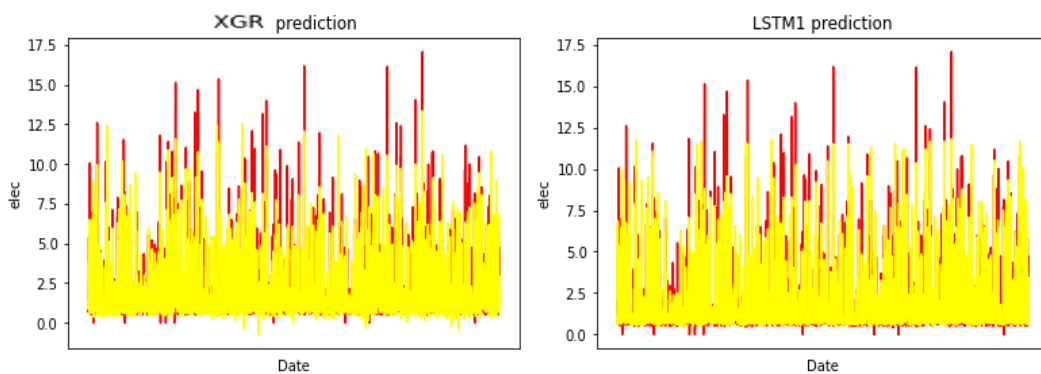
Διάγραμμα συνόλου δεδομένων

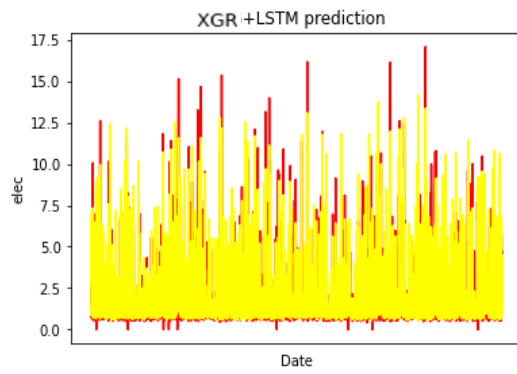


Ωριαία Δειγματοληψία

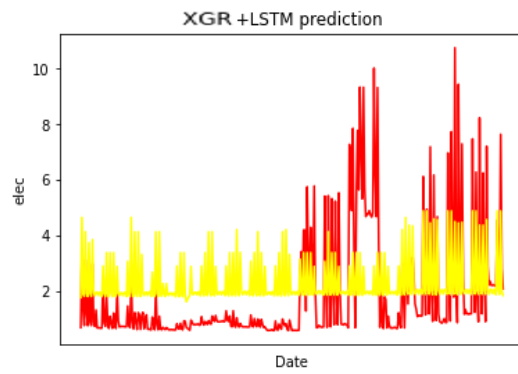
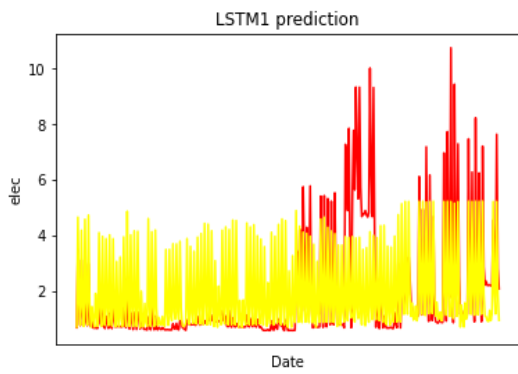
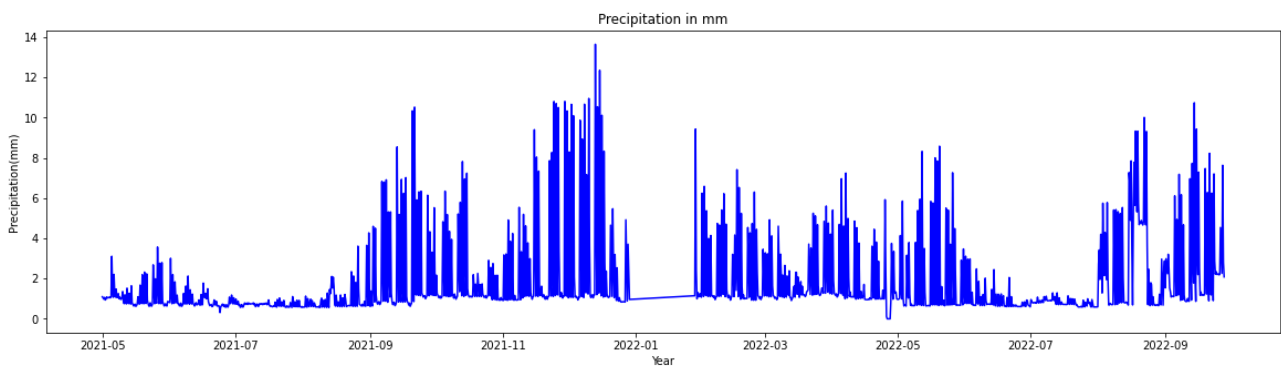


Ωριαία Δειγματοληψία "Shuffled."

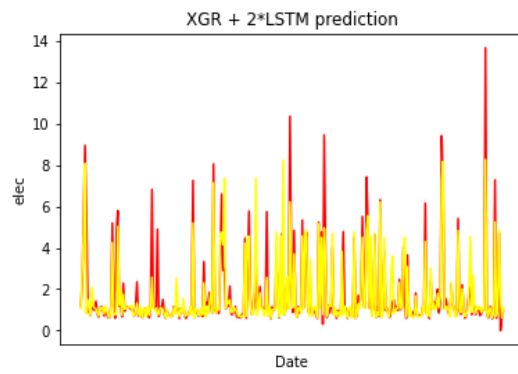
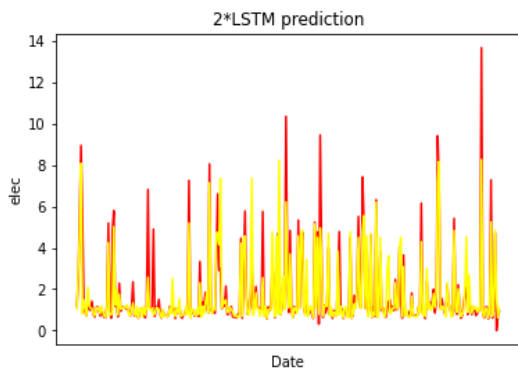




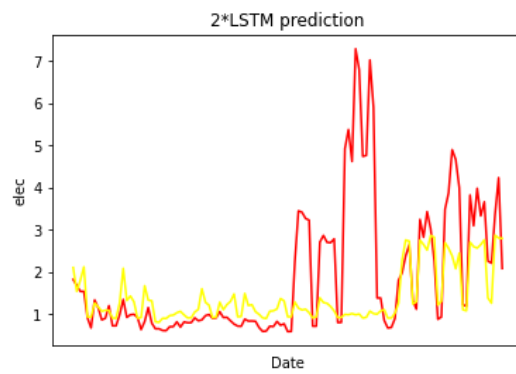
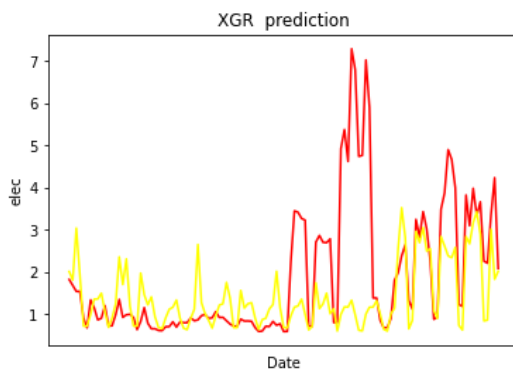
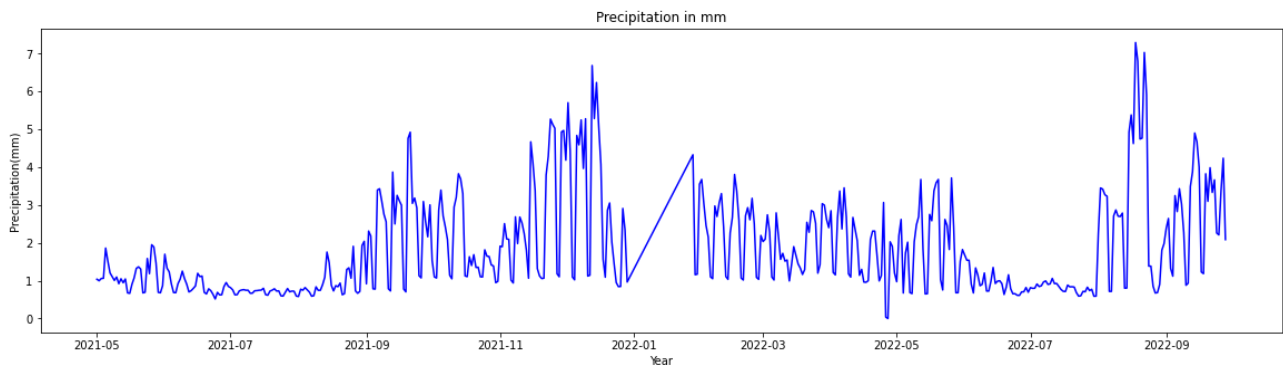
Ανα 8 ώρες δειγματοληψία



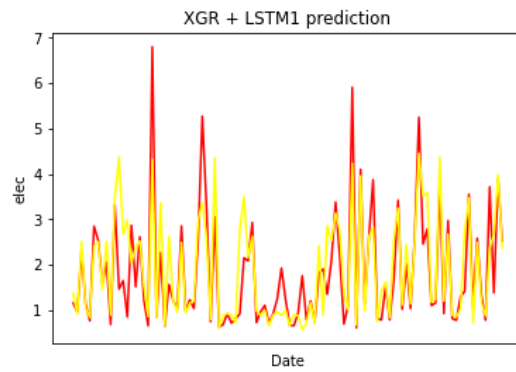
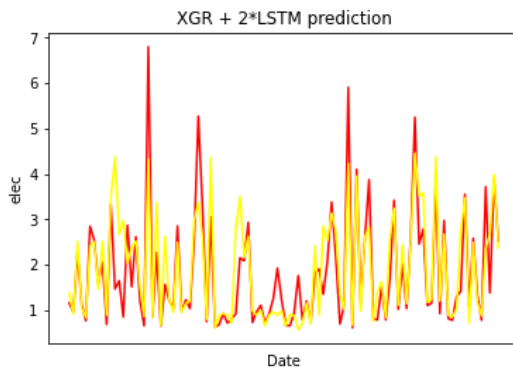
Ανα 8 ώρες δειγματοληψία "Shuffled."



Ημερήσια δειγματοληψία

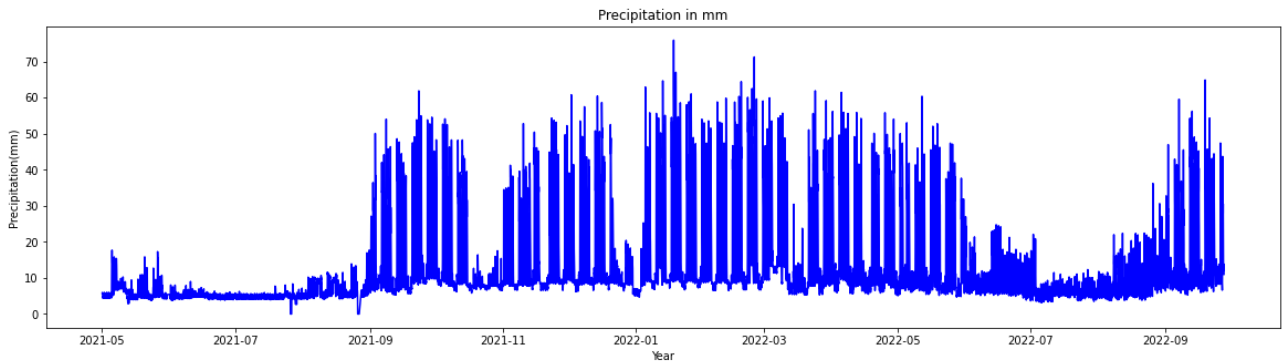


Ημερήσια δειγματοληψία "Shuffled."

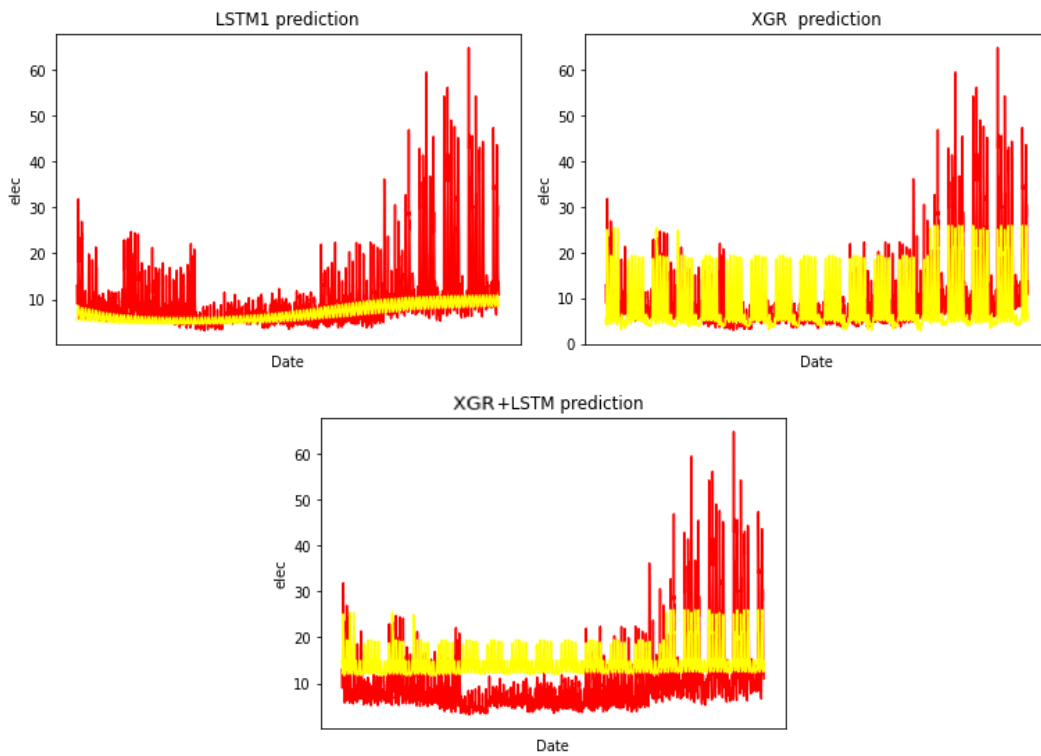


Gulbene_region_Gulbene_secondary_school_nr.2

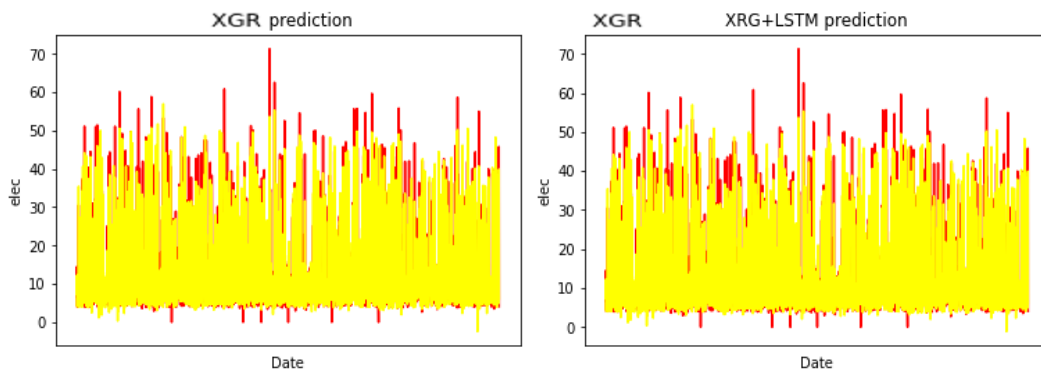
Διάγραμμα συνόλου δεδομένων

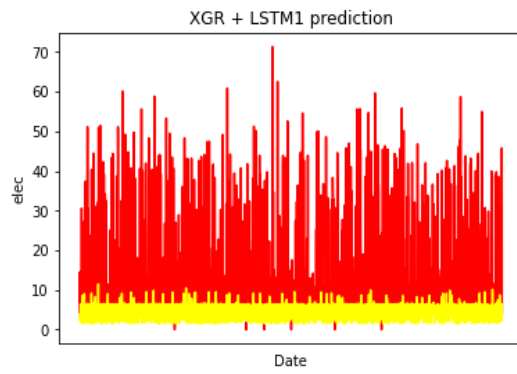


Ωριαία Δειγματοληψία

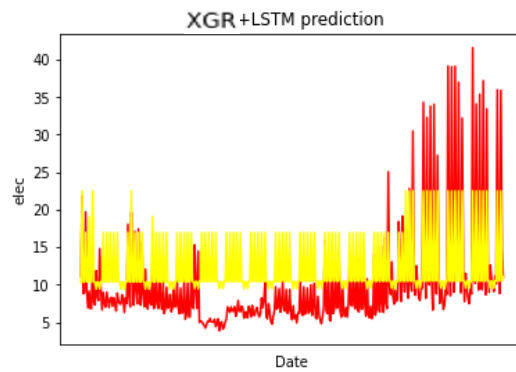
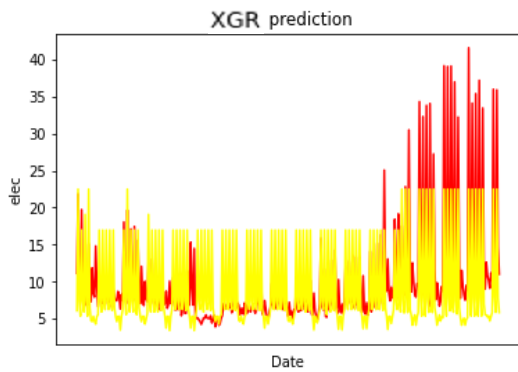
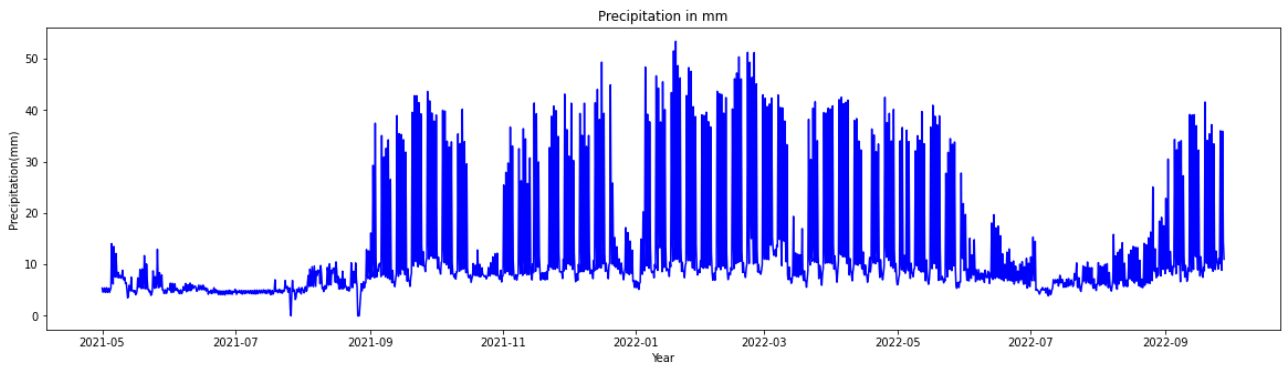


Ωριαία Δειγματοληψία "Shuffled."

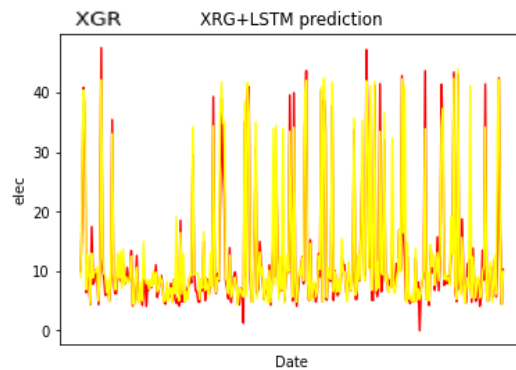
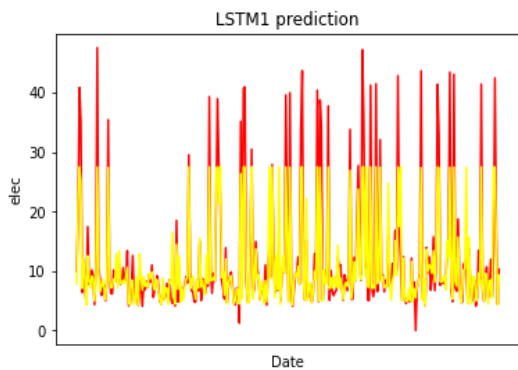




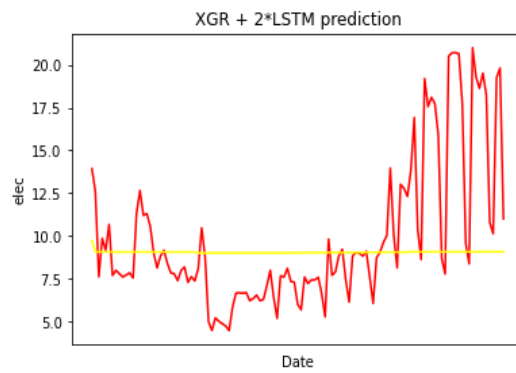
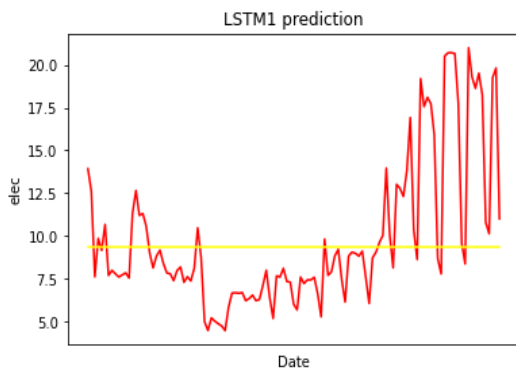
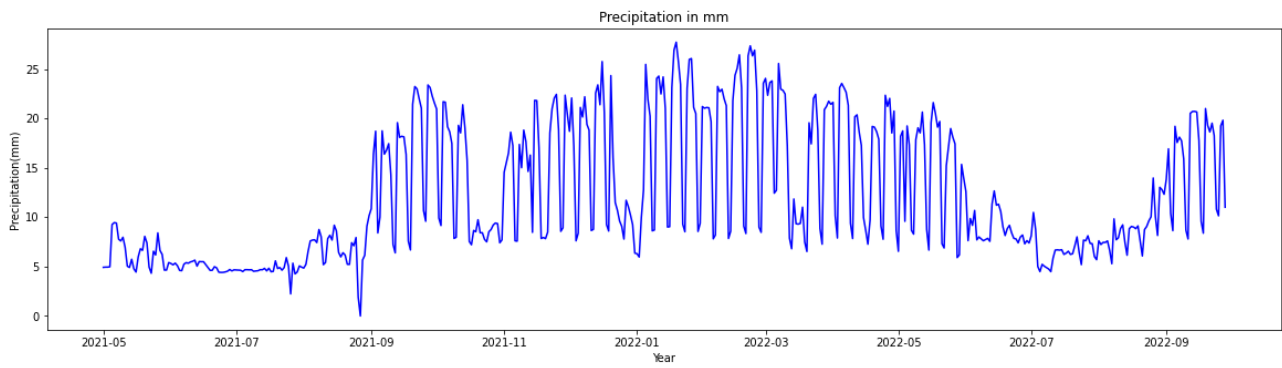
Ανα 8 ώρες δειγματοληψία



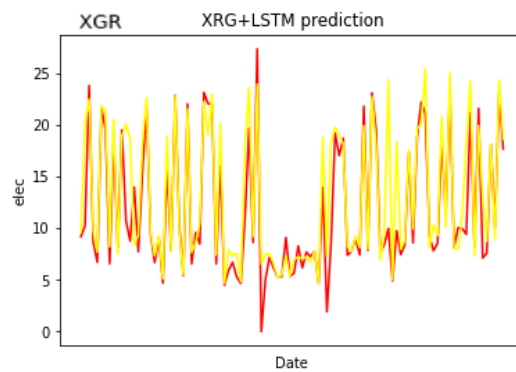
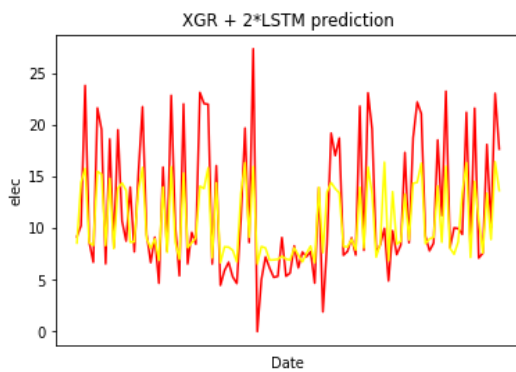
Ανα 8 ώρες δειγματοληψία "Shuffled."



Ημερήσια δειγματοληψία

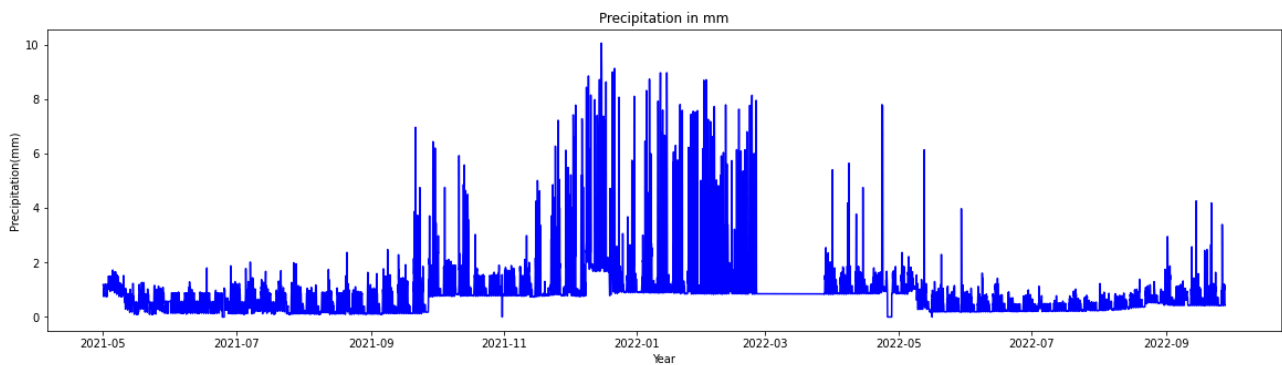


Ημερήσια δειγματοληψία "Shuffled."

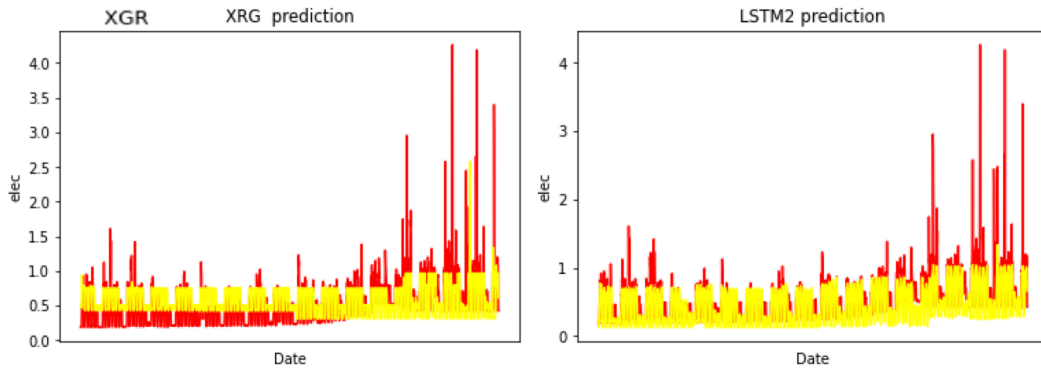


Gulbene_region_Stāki_gym

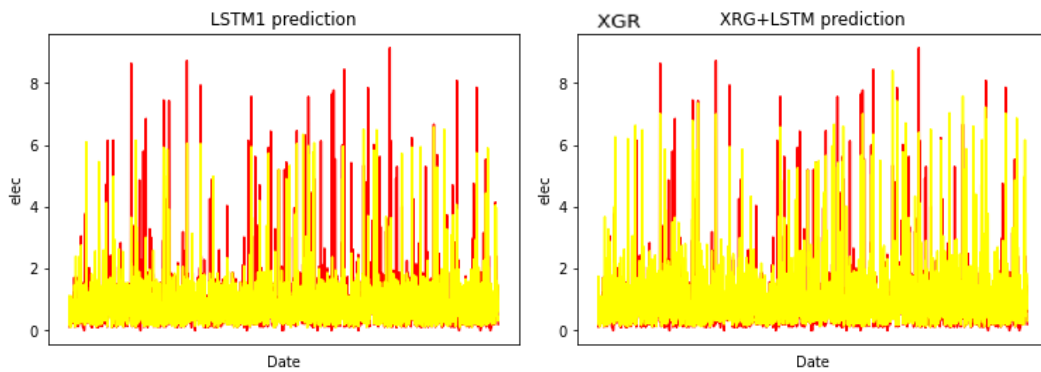
Διάγραμμα συνόλου δεδομένων



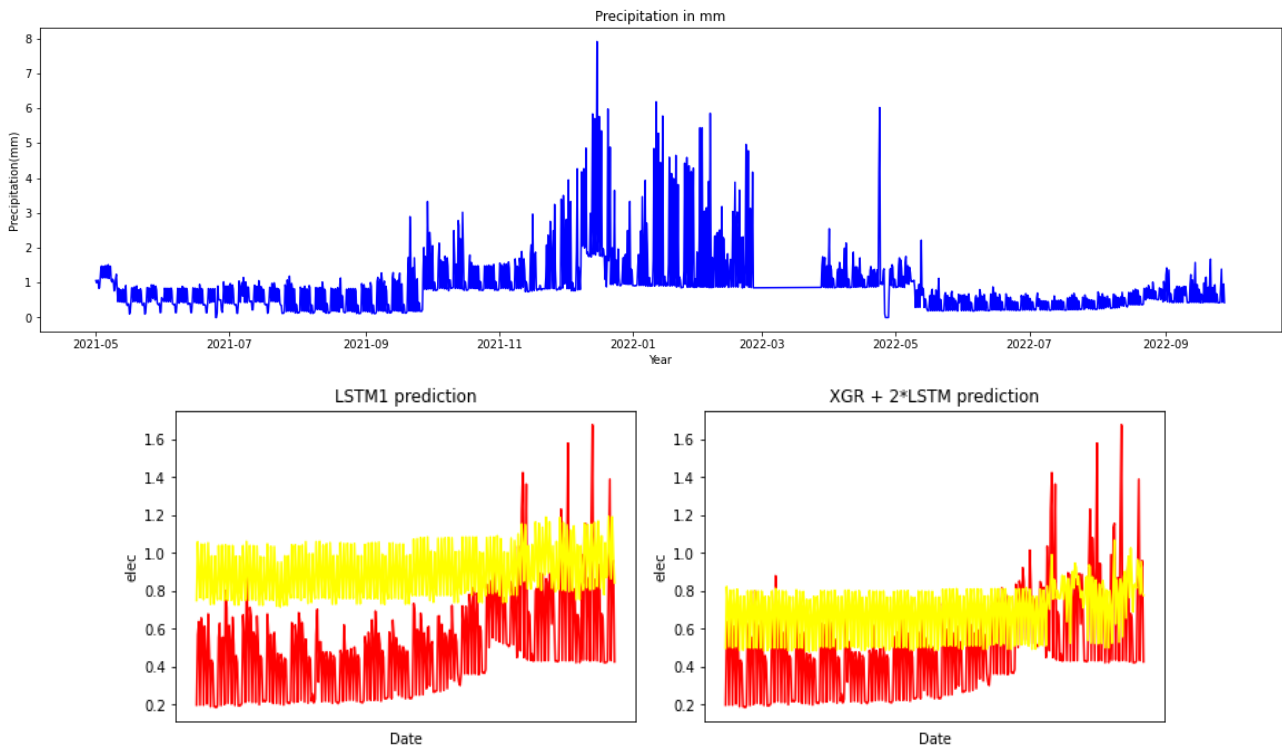
Ωριαία Δειγματοληψία



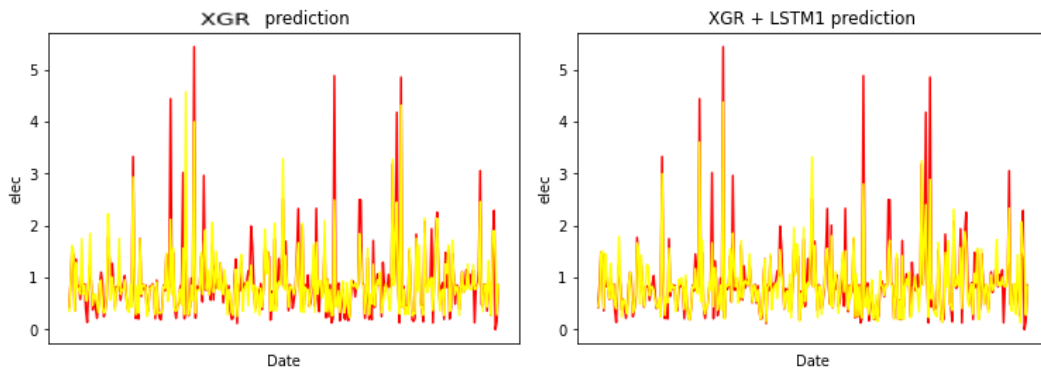
Ωριαία Δειγματοληψία "Shuffled."



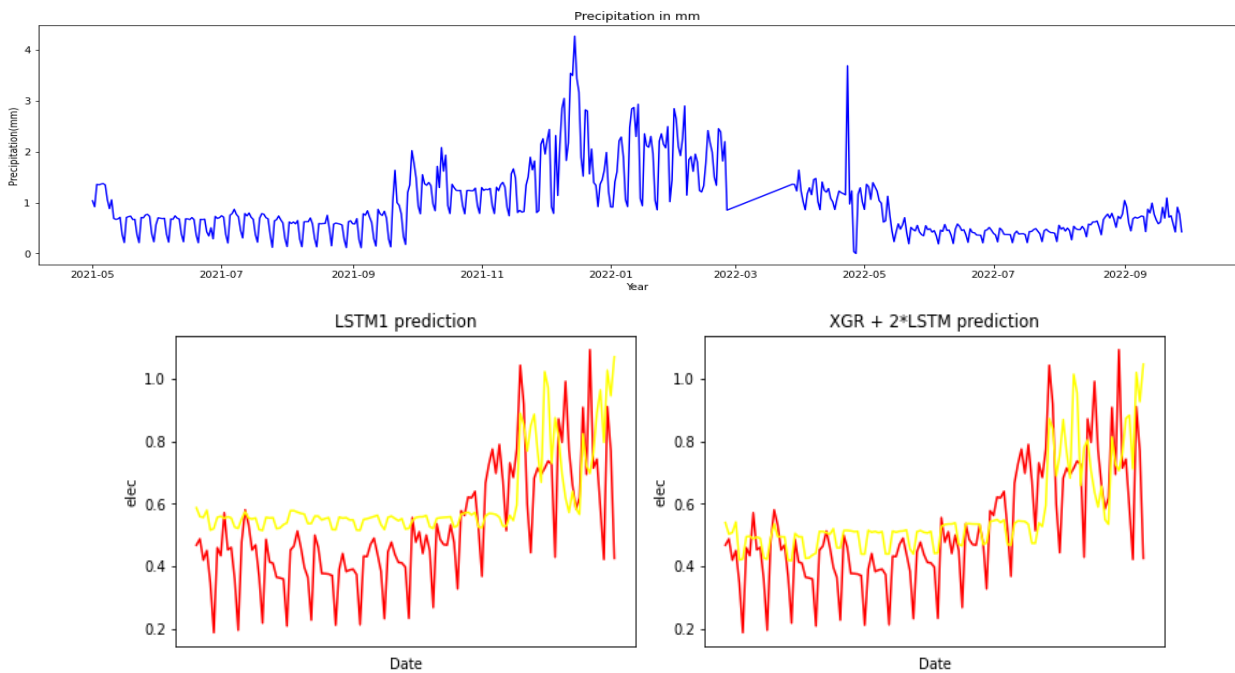
Ανα 8 ώρες δειγματοληψία



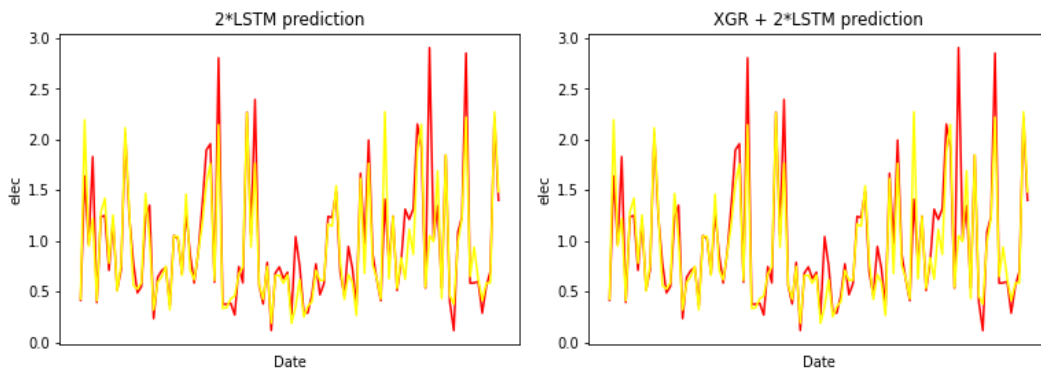
Ανα 8 ώρες δειγματοληψία "Shuffled."



Ημερησια δειγματοληψία

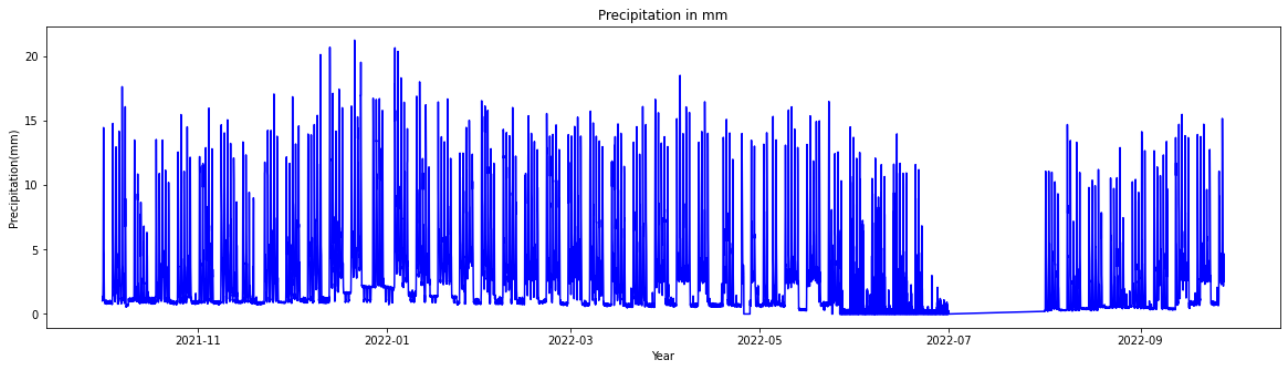


Ημερησια δειγματοληψία "Shuffled."

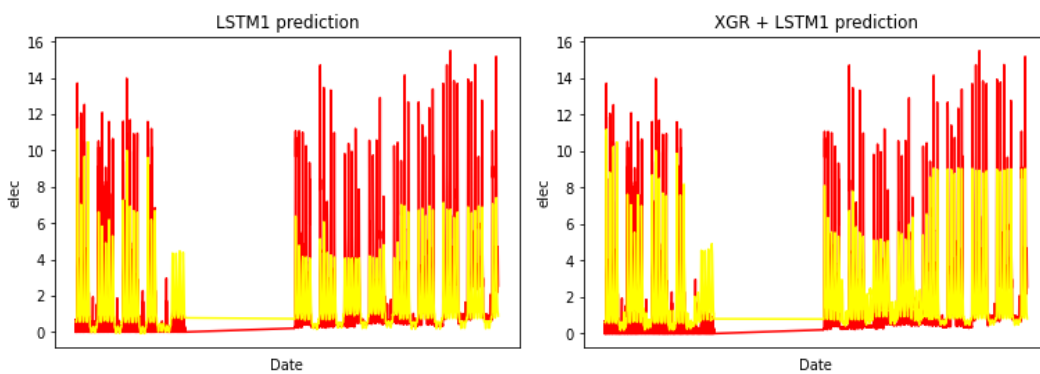


Gulbene_region_Jaungulbene_kindergarten_doctorate

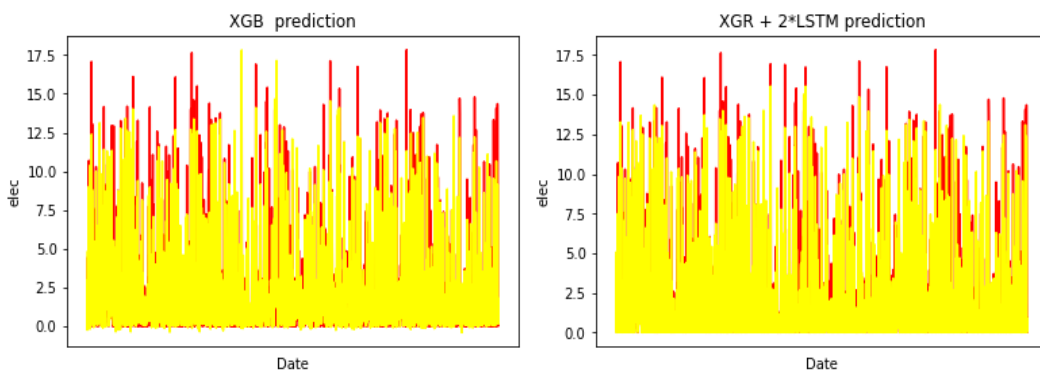
Διάγραμμα συνόλου δεδομένων



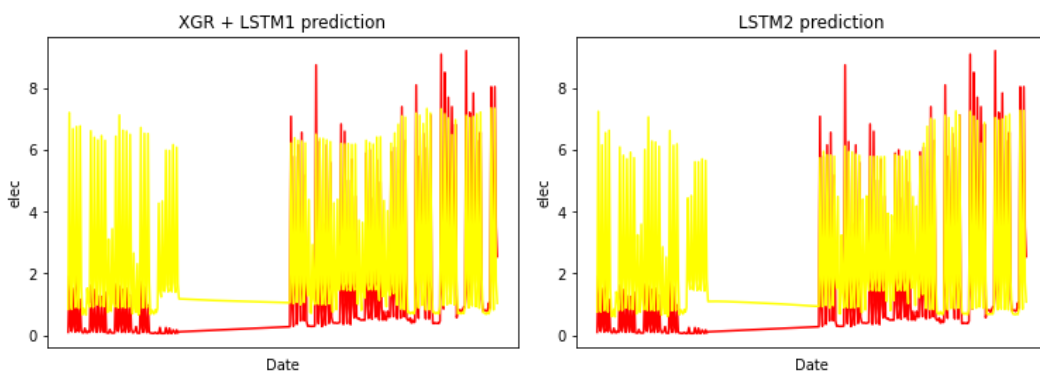
Ωριαία Δειγματοληψία



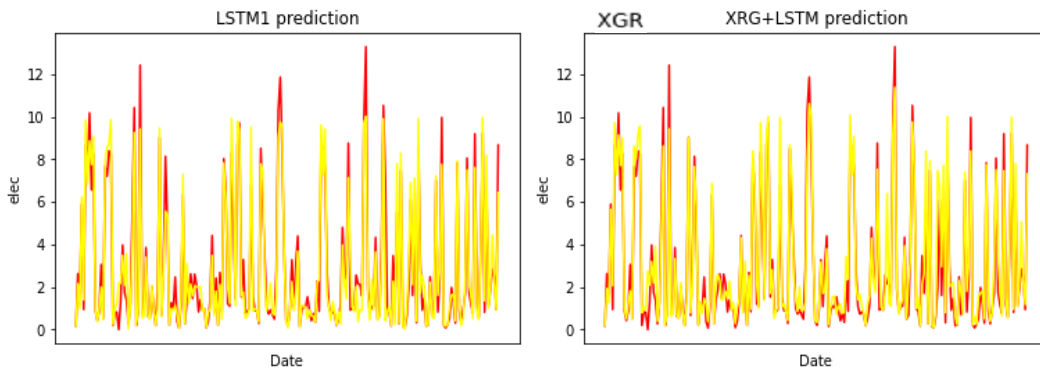
Ωριαία Δειγματοληψία "Shuffled."



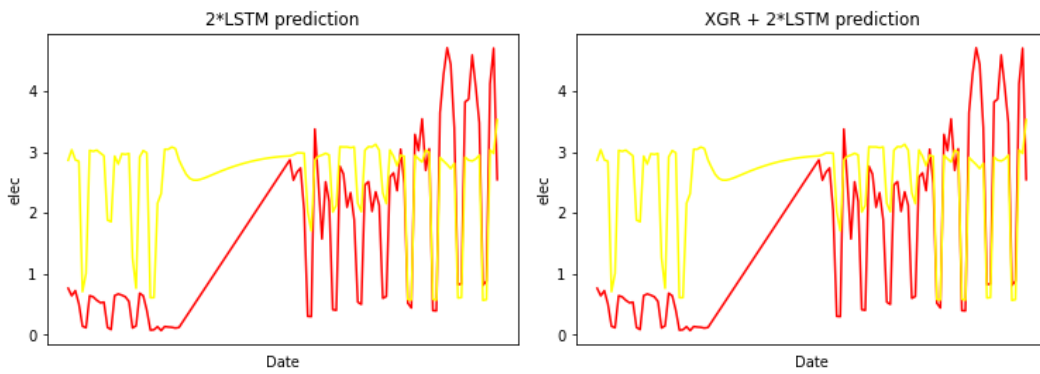
Ανα 8 ώρες δειγματοληψία



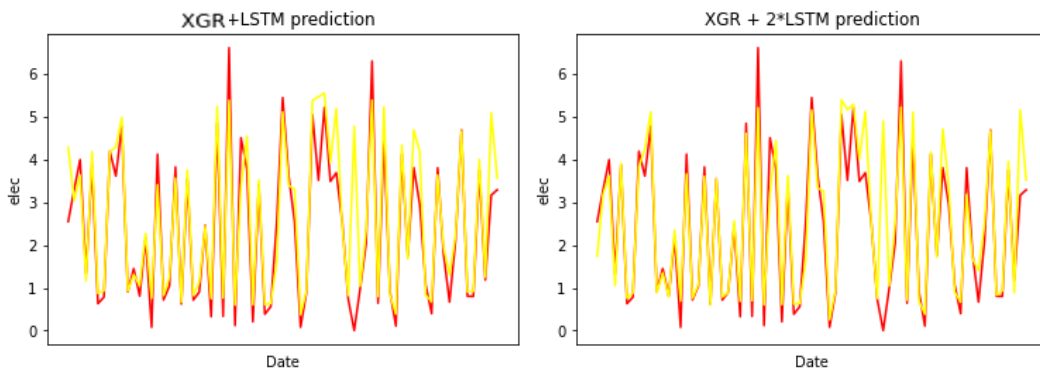
Ανα 8 ώρες δειγματοληψία "Shuffled."



Ημερήσια δειγματοληψία

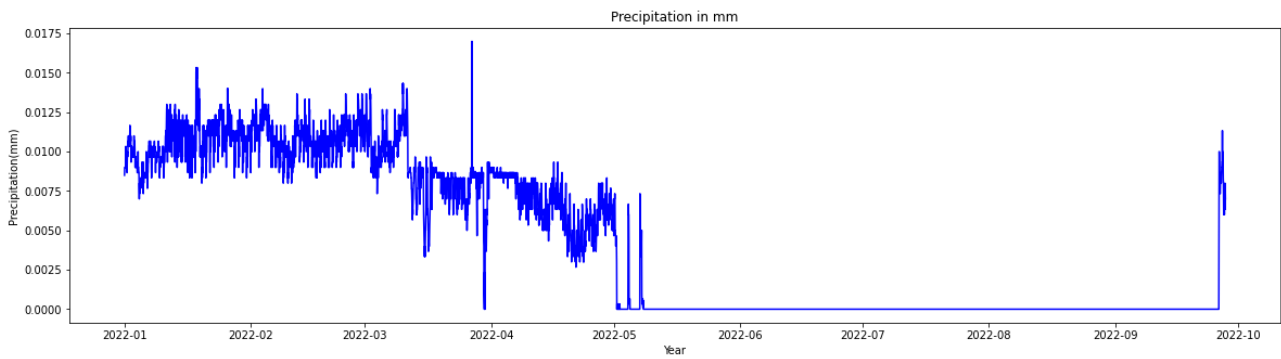


Ημερήσια δειγματοληψία "Shuffled."

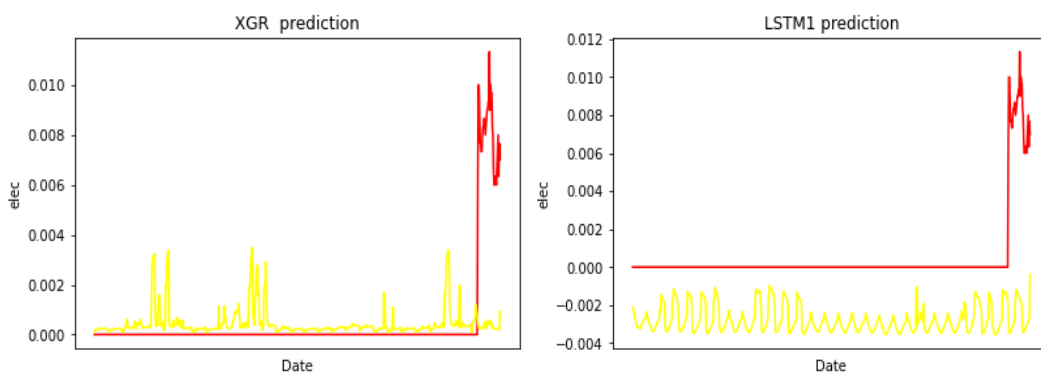


Gulbene_region_Līgo_culture_center_parish_administration_library

Διάγραμμα συνόλου δεδομένων

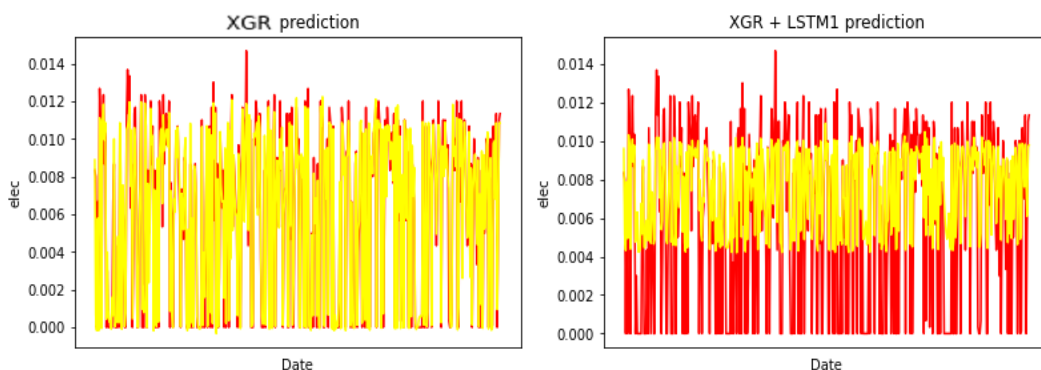


Ωριαία Δειγματοληψία

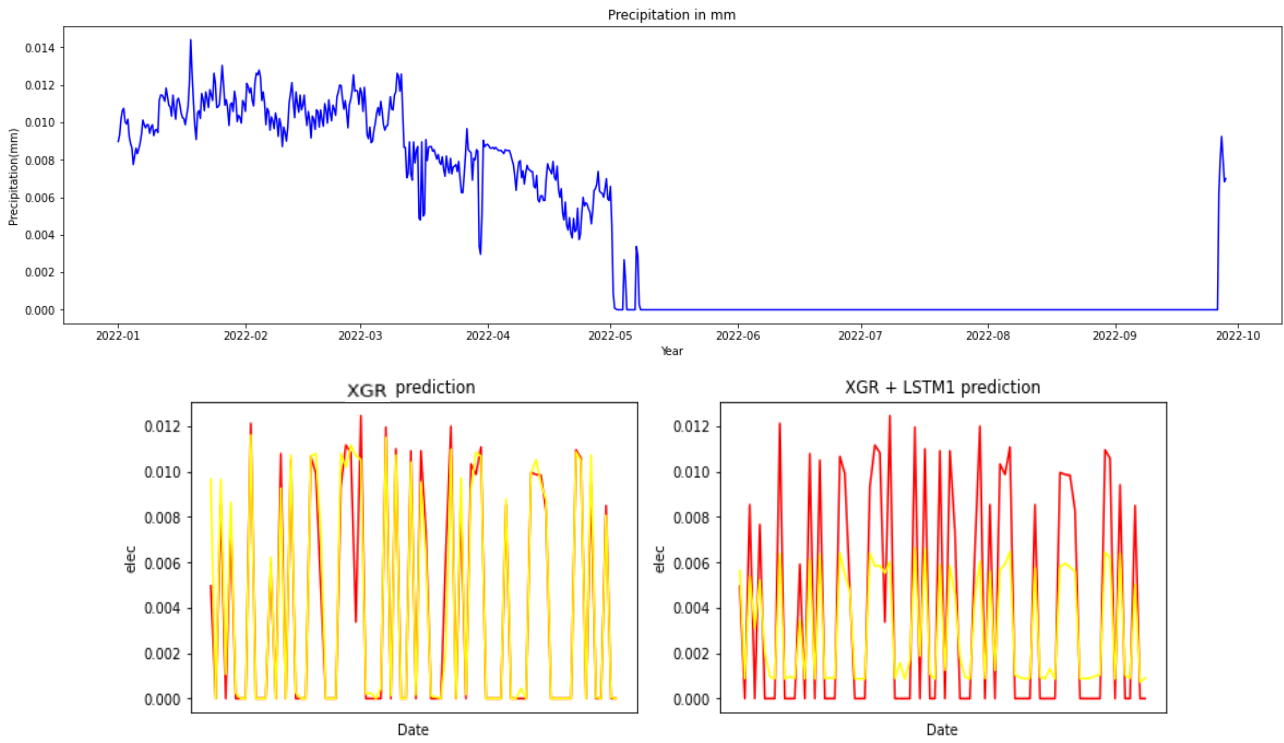


Επειδή όπως περιμέναμε και από τα αποτελέσματα το διάγραμμα της σειριακής περίπτωσης για τη θερμική κατανάλωση δεν αξίζει καθώς δεν υπάρχει καμία ακρίβεια. Έτσι δε θα παρουσιαστεί κάποιο άλλο από τα δυο αυτά σύνολα δεδομένων με σειριακή προσέγγιση

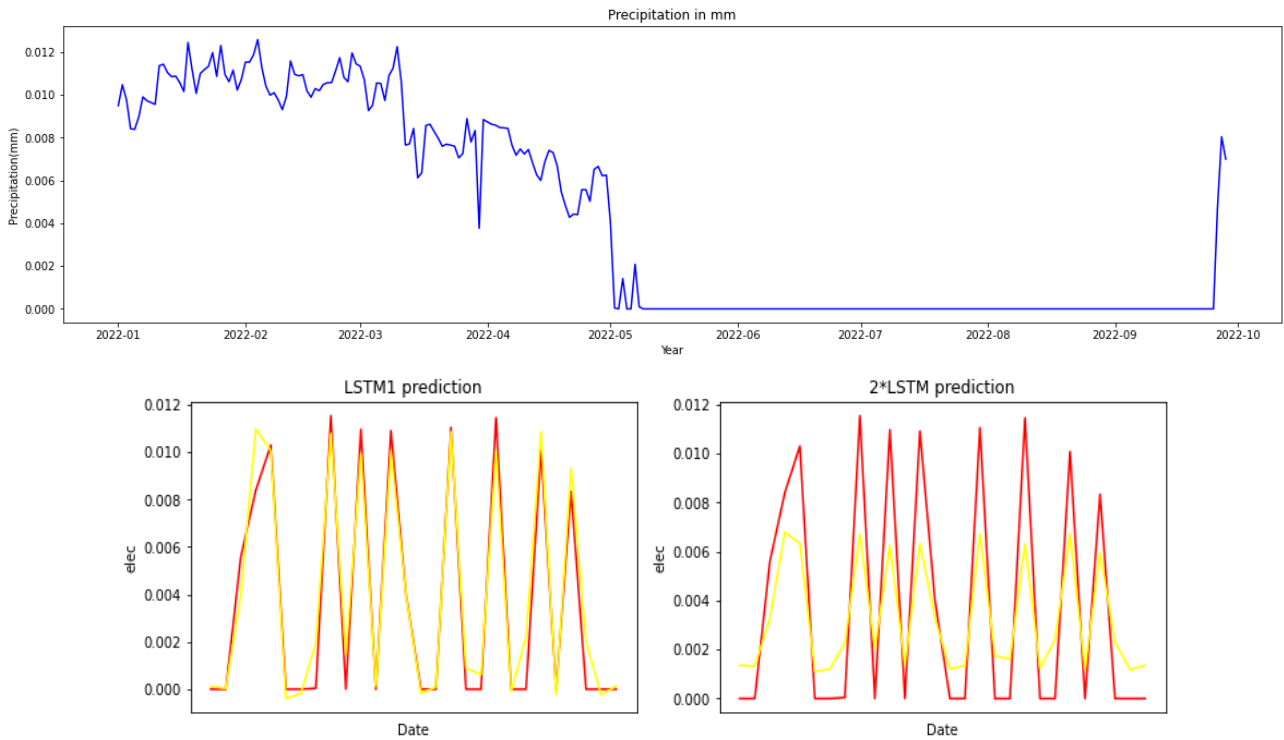
Ωριαία Δειγματοληψία "Shuffled."



Ανα 8 ώρες δειγματοληψία "Shuffled."

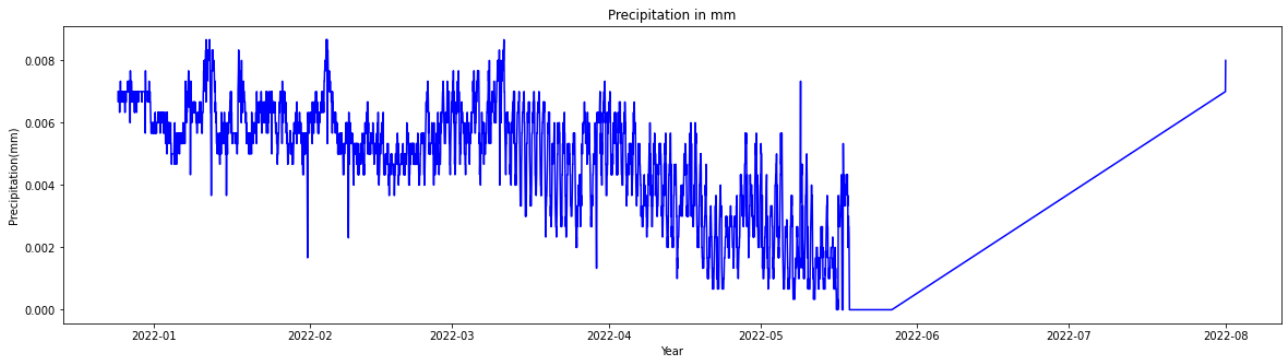


Ημερήσια δειγματοληψία "Shuffled."

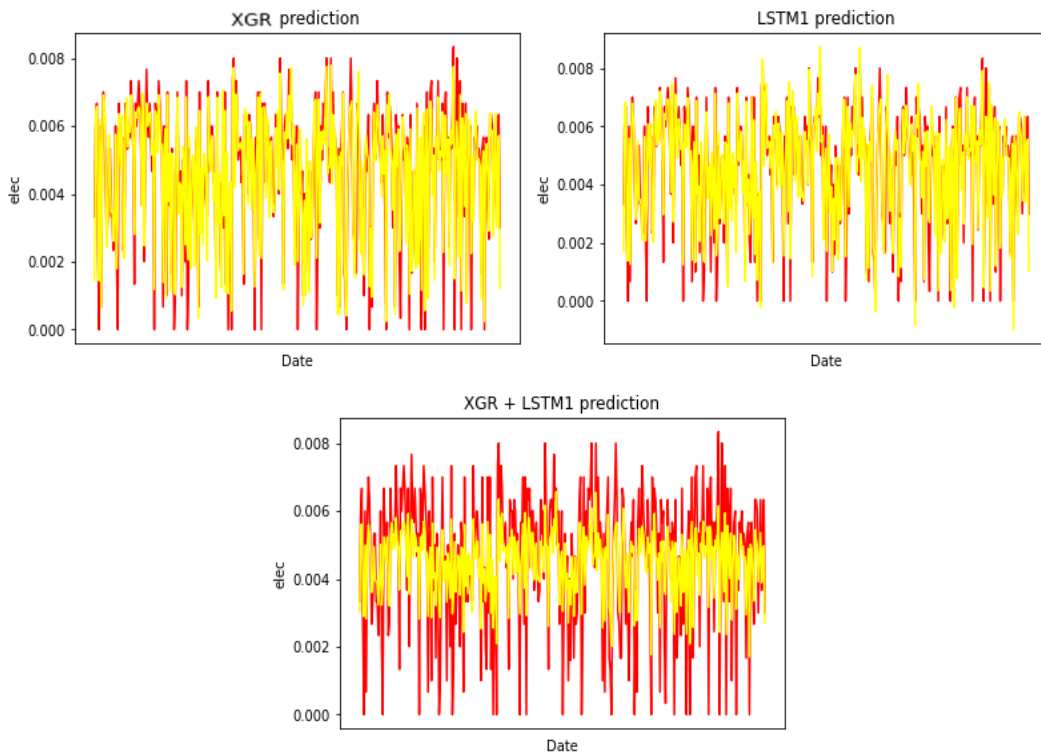


Gulbene_region_Litenes_gym

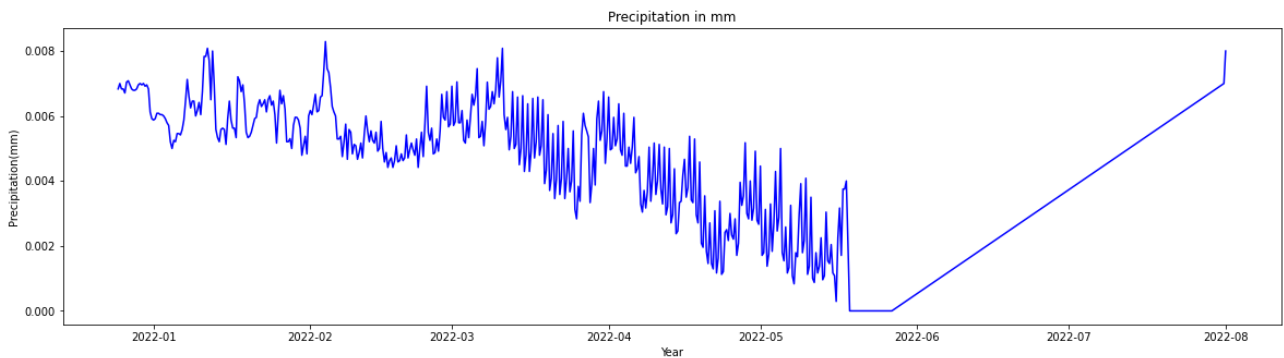
Διάγραμμα συνόλου δεδομένων

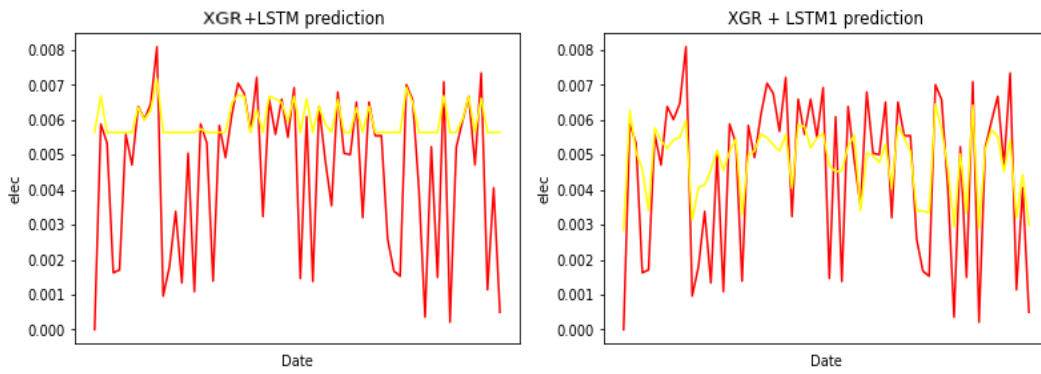


Ωριαία Δειγματοληψία "Shuffled."

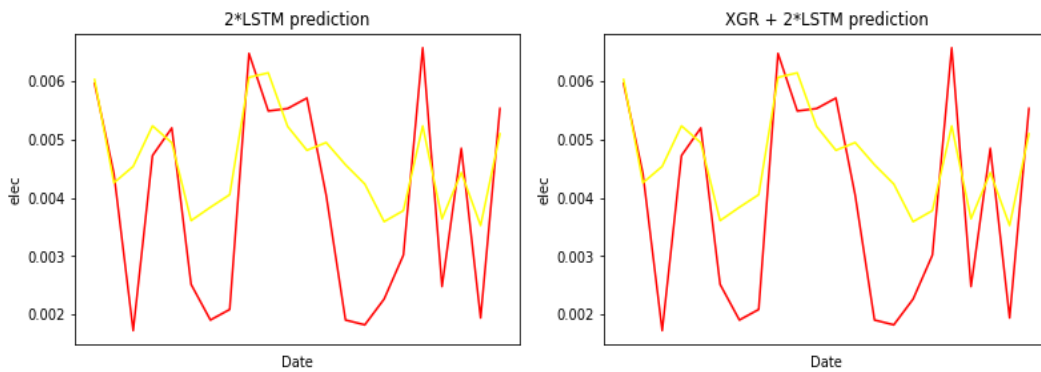
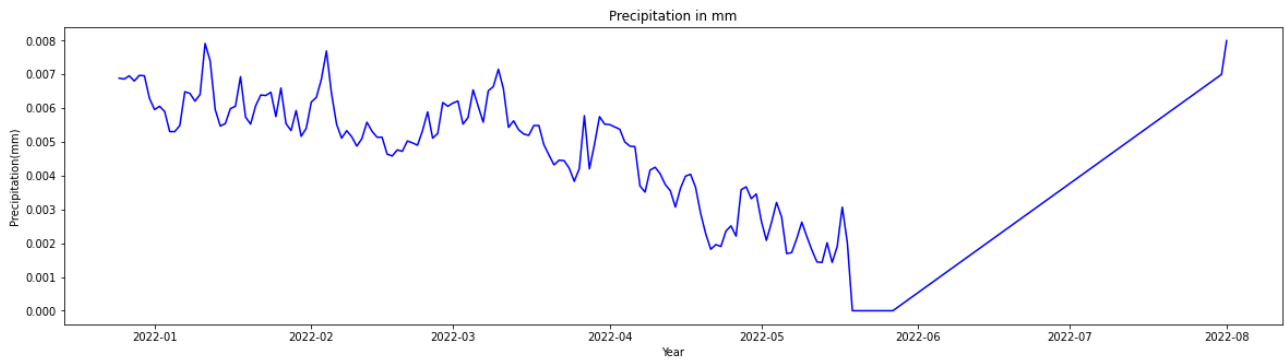


Ανα 8 ώρες δειγματοληψία "Shuffled."





Ημερήσια δειγματοληψία "Shuffled."



Βιβλιογραφία

- [1] http://www.triedrasi.gr/index.php/energeiaki_anavathmisi.html
- [2] Efficiency Valuation Organization, Statistics and Uncertainty for IPMVP, Technical Report, Efficiency Valuation Organization, 2014.
- [3] <https://veregy.com/understanding-the-international-performance-measurement-and-verification-protocol/>
- [4] ASHRAE Guideline 14 (2014). ASHRAE Guideline 14-2014 for Measurement of Energy and Demand Savings, American Society of Heating, Refrigeration and Air Conditioning Engineers, Atlanta, GA.
- [5] <https://watchwire.ai/measuring-verifying-mv-efficiency-projects/>
- [6] Efficiency Valuation Organization, Core Concepts, Tech. Rep. October, Efficiency Valuation Organization (2016).
- [7] <https://evo-world.org/en/m-v/what-is-m-v>
- [8] Gallagher, C.V.; Leahy, K.; O'Donovan, P.; Bruton, K.; O'Sullivan, D.T.J. Development and Application of a Machine Learning Supported Methodology for Measurement and Verification (M&V) 2.0. *Energy Build.* **2018**, *167*, 8–22.
- [9] https://murraylax.org/rtutorials/regression_intro.html
- [10] <https://www.ibm.com/topics/knn>
- [11] <https://towardsdatascience.com/support-vector-machine-introduction-to-machine-learning-algorithms-934a444fca47>
- [12] <https://www.ibm.com/topics/decision-trees>
- [13] J. Granderson, P.N. Price, Development and application of a statistical methodology to evaluate the predictive accuracy of building energy baseline models, *Energy* 66 (1) (2014) 981–990, doi:10.1016/j.energy.2014.01.074.
- [14] B. Dong, C. Cao, S.E. Lee, Applying support vector machines to predict building energy consumption in tropical region, *Energy Build.* 37 (5) (2005) 545–553
- [15] F. Tang, A. Kusiak, X. Wei, Modeling and short-term prediction of HVAC system with a clustering algorithm, *Energy Build.* 82 (2014) 310–321

- [16] B. Yildiz, J.I. Bilbao, A.B. Sproul, A review and analysis of regression and machine learning models on commercial building electricity load forecasting, *Renewable Sustainable Energy Rev.* 73 (2017) 1104–1122
- [17] FORECASTING THE HOURLY ENERGY CONSUMPTION USING XGBOOST
- [18] H. Esen, M. Inalli, A. Sengur, M. Esen, Modeling a ground-coupled heat pump system by a support vector machine, *Renew. Energy* 33 (8) (2008) 1814–1823
- [19] Forecasting of a ground-coupled heat pump performance using neural networks with statistical data weighting pre-processing, *Int. J. Therm. Sci.* 47 (4) (2008) 431–441
- [20] Moraliyage, H.; Dahanayake, S.; De Silva, D.; Mills, N.; Rathnayaka, P.; Nguyen, S.; Alahakoon, D.; Jennings, A. A Robust Artificial Intelligence Approach with Explainability for Measurement and Verification of Energy Efficient Infrastructure for Net Zero Carbon Emissions. *Sensors* **2022**, 22(23), 9503
- [21] J. Ngarambe, G.Y. Yun, M. Santamouris, The use of artificial intelligence (AI) methods in the prediction of thermal comfort in buildings: energy implications of AI-based thermal comfort controls, *Energy and Buildings* 211 (2020) 109807.
- [22] Torres-Barrán, A.; Alonso, Á.; Dorronsoro, J.R. Regression Tree Ensembles for Wind Energy and Solar Radiation Prediction. *Neurocomputing* **2019**, 326–327, 151–160.
- [23] Agenis-Nevers, M.; Wang, Y.; Dugachard, M.; Salvazet, R.; Becker, G.; Chenu, D. Measurement and Verification for Multiple Buildings: An Innovative Baseline Model Selection Framework Applied to Real Energy Performance Contracts. *Energy Build.* **2021**, 249, 111183.
- [24] Olu-Ajayi, R.; Alaka, H.; Sulaimon, I.; Sunmola, F.; Ajayi, S. Building energy consumption prediction for residential buildings using deep learning and other machine learning techniques. [Journal of Building Engineering](#) 2021, 45(103406)
- [25] <https://www.degreedays.net/calculate-energy-savings>
- [26] XGBoost : A scalable tree boosting system, Tianqi Chen, Carlos Guestrin, March 9, 2016, arXiv:1603.02754
- [27] Monner, D., Reggia, J.A., 2012. A generalized LSTM-like training algorithm for second-order recurrent neural networks. *Neural Networks* 25, 70–83. <https://doi.org/10.1016/j.neunet.2011.07.003>.
- [28] <https://amr-khalil.medium.com/what-is-long-short-term-memory-lstm-cdd8c669a73e>
- [29] Lu. Yuwei, Yuan Yuan, Qi Wang, A dense connection-based network for real-time object tracking, *Neurocomputing* 410 (2020) 229–236.

- [30]<https://analyticsindiamag.com/a-complete-understanding-of-dense-layers-in-neural-networks/>
- [31] <https://www.geeksforgeeks.org/intuition-of-adam-optimizer/>
- [32]https://resources.eumetrain.org/data/4/451/english/msg/ver_cont_var/uos3/uos3_ko1.htm
- [33]<https://towardsdatascience.com/common-loss-functions-in-machine-learning-46af0ffc4d23>
- [34]https://medium.com/@20_80_/mean-absolute-error-mae-sample-calculation-6eed6743838a
- [35] <https://machinelearningmastery.com/difference-between-a-batch-and-an-epoch/>
- [36] <https://github.com/chrissiaf/-Energy-saving-profit-forecasting-.git>