



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

UNIVERSITÉ DE LIMOGES



Deep Learning
Εαρινό Εξάμηνο 2022-2023

Εργασία Εξαμήνου

Ονοματεπώνυμο: Χριστοδουλίδης Αναστάσιος
A.M.: aivc22019

Περιεχόμενα

1. Εισαγωγή - Περιγραφή προβλημάτων.....	1
1.1. Μέρος Α: Ταξινομητές για δεδομένα πινάκων.....	1
1.2. Μέρος Β: Δημιουργία, εκπαίδευση, αξιολόγηση και αποθήκευση διαφορετικών αρχιτεκτονικών βαθέων δικτύων σε εικόνες.....	1
1.3. Μέρος Γ: Αξιολόγηση καταλληλότητας διαφορετικών αρχιτεκτονικών αναδρομικών νευρωνικών δικτύων για ενεργειακό επιμερισμό.....	1
2. Μεθόδους.....	2
2.1. Εποπτευόμενη Μάθηση.....	2
2.2. Linear Discriminant Analysis (LDA).....	2
2.3. Logistic Regression.....	2
2.4. Decision Trees.....	2
2.5. K-Nearest Neighbors.....	3
2.6. Naïve Bayes.....	3
2.7. Support Vector Machines.....	3
2.8. Neural Networks.....	4
2.9. Deep Neural Network (DNN).....	4
2.10. Convolutional Neural Network (CNN).....	4
2.11. Recurrent Neural Network (RNN).....	5
2.12. Long Short-Term Memory Network (LSTM).....	5
2.13. Gated Recurrent Unit (GRU).....	5
3. Σύνολα Δεδομένων.....	6
3.1. Μέρος Α: Ελληνικές Εταιρίες.....	6
3.2. Μέρος Β: Σύνολο Δεδομένων mnist.....	6
3.3. Μέρος Γ: Κατανάλωση Ρεύματος Σπιτιού και WaterHeater.....	7
4. Αποτελέσματα-Συμπεράσματα.....	8
4.1. Μέρος Α.....	8
4.2. Μέρος Β.....	9
4.3. Μέρος Γ.....	14

Κατάλογος Πινάκων/Γραφημάτων

Πίνακας	Περιγραφή	Σελίδα
3.1.1.	Δείγμα Δεδομένων-Ελληνικές Εταιρίες	6
3.1.2.	Δείγμα Δεδομένων-Mnist Dataset	7
3.1.3.	Δείγμα I/O κατανάλωσης	7
4.1.1.	Πίνακας Αποτελεσμάτων-Μέρος Α-I	7
4.1.2.	Πίνακας Αποτελεσμάτων-Μέρος Α-II	8
4.2.1.	Αρχιτεκτονική DNN μοντέλου	10
4.2.2.	Αρχιτεκτονική CNN μοντέλου	11
4.2.3.	Πίνακας Αποτελεσμάτων-Μέρος Β	11
4.2.4.	Χάρτες Χαρακτηριστικών	12
4.2.5.	Confusion Matrix-DNN	13
4.2.6.	Confusion Matrix-CNN	13
4.3.1.	MSE Loss of RNN, LSTM & GRU models	14
4.3.2.	RMSE, MAE & Max Error	15
4.3.3.	Σύγκριση Πραγματικών και Προβλεπόμενων Εξόδων	16

1. Εισαγωγή – Περιγραφή Προβλημάτων

1.1. Μέρος Α: Ταξινομητές για δεδομένα πινάκων

Σε ένα χρηματοπιστωτικό ίδρυμα είναι σημαντικό να μπορούν να κάνουν προβλέψεις για εταιρίες, σχετικά με το αν θα πτωχεύσουν ή όχι. Αυτό επειδή, τα χρηματοπιστωτικά ιδρύματα, μπορούν να παρέχουν δάνεια σε εταιρίες και συμβουλευτικές υπηρεσίες σε πελάτες σχετικά με τη διαχείριση των οικονομικών τους και την ανάπτυξη επενδυτικών στρατηγικών, καθώς και άλλες υπηρεσίες που προμηθύνουν οικονομικούς κινδύνους αν οι εταιρίες δεν είναι υγιείς.

Οπότε, έχοντας περάσει από την πρώτη συνέντευξη για μια θέση με αντικείμενο «data analytics & financial risk estimations associate» σε ένα χρηματοπιστωτικό ίδρυμα, ζητείται η δημιουργία ενός ταξινομητή που να εντοπίζει τις εταιρίες που θα χρεωκοπήσουν βάση των δεδομένων που μας παραχώρησαν. Θα γίνει ανάλυση των δεδομένων, τα οποία περιέχουν στοιχεία των επιχειρήσεων (δείκτες απόδοσης και δραστηριοτήτων), που είναι υγιείς ή έχουν χρεωκοπήσει, ώστε να δημιουργήσουμε ένα μοντέλο επιβλεπόμενης μάθησης το οποίο να είναι ικανό να κάνει προβλέψεις για την κατάσταση εταιριών βάση αυτών των δεικτών. Ο στόχος που έχουμε για το μοντέλο είναι να βρίσκει με ποσοστό επιτυχίας τουλάχιστον 62% τις εταιρίες που θα πτωχεύσουν και 70% τις εταιρίες που δεν θα πτωχεύσουν.

Τα μοντέλα που θα χρησιμοποιήσουμε είναι τα Linear Discriminant Analysis, Logistic Regression, Decision Trees, k-Nearest Neighbors, Naïve Bayes, Support Vector Machines και Neural Networks.

1.2. Μέρος Β: Δημιουργία, εκπαίδευση, αξιολόγηση και αποθήκευση διαφορετικών αρχιτεκτονικών βαθέων δικτύων σε εικόνες

Το mnist dataset περιέχει εικόνες από χειρόγραφα νούμερα από το 0 μέχρι το 9. Στόχος είναι η κατηγοριοποίησή αυτών των εικόνων στις σωστές κλάσεις. Αυτό θα γίνει με την δημιουργία δικτύων Deep Neural Network (DNN) και Convolutional Neural Network (CNN). Με την χρήση του KFold θα δημιουργήσουμε 6 μικρότερα σύνολα δεδομένων και ακολούθως θα εκπαιδευτούν 6 DNN και 6 CNN μοντέλα ώστε να επιλέξουμε τα καλύτερα.

1.3. Μέρος Γ: Αξιολόγηση καταλληλότητας διαφορετικών αρχιτεκτονικών αναδρομικών νευρωνικών δικτύων για ενεργειακό επιμερισμό

Για την πρόβλεψη χρονικών δεδομένων γίνεται χρήση αρχιτεκτονικών όπως Recurrent Neural Network (RNN), Long Short-Term memory network (LSTM) και Gated Recurrent Unit (GRU). Στο πρόβλημα μας έχουμε δεδομένα για την κατανάλωση ρεύματος στο σπίτι για περιόδους 120 λεπτών. Επιπρόσθετα, έχοντας δεδομένα κατανάλωσης για την συσκευή WaterHeater, θέλουμε να εκπαιδεύσουμε τα 3 πιο πάνω μοντέλα ώστε να μπορούν βάση της συνολικής κατανάλωσης να προβλέπουν την κατανάλωση της συσκευής.

2. Μεθόδους

2.1. Εποπτευόμενη Μάθηση

Η εποπτευόμενη μάθηση είναι ένας τύπος αλγορίθμου μηχανικής μάθησης όπου ένα μοντέλο εκπαιδεύεται σε ζεύγη εισόδου-εξόδου. Η διαδικασία εκπαίδευσης περιλαμβάνει την παρουσίαση του αλγορίθμου με δεδομένα εισόδου και τις αντίστοιχες σωστές εξόδους. Ο αλγόριθμος στη συνέχεια γενικεύει από αυτά τα παραδείγματα και προσπαθεί να βρει μοτίβα ή σχέσεις μεταξύ των χαρακτηριστικών εισόδου και των εξόδων. Αυτό επιτρέπει στο μοντέλο να κάνει προβλέψεις ή ταξινόμησης σε νέα δεδομένα.

Η εποπτευόμενη μάθηση απαιτεί ένα καλά επισημασμένο σύνολο δεδομένων εκπαίδευσης, όπου κάθε είσοδος συνδέεται με το σωστό αποτέλεσμα. Η ποιότητα και η αντιπροσωπευτικότητα των δεδομένων εκπαίδευσης είναι ζωτικής σημασίας για την απόδοση του μοντέλου και τη γενίκευση σε νέα δεδομένα.

2.2. Linear Discriminant Analysis (LDA)

Ο LDA είναι μια τεχνική μείωσης διαστάσεων και ταξινόμησης. Στοχεύει στο να βρει ένα χώρο μικρότερων διαστάσεων στα δεδομένα, έτσι ώστε να είναι διαχωρίσιμα. Επιπρόσθετα, μπορεί να πετύχει μείωση των διαστάσεων διατηρώντας την πληροφορία των κλάσεων. Παρόλα αυτά ο LDA υποθέτει γραμμικότητα το οποίο μπορεί να προκαλέσει προβλήματα σε μη γραμμικά δεδομένα και είναι ευαίσθητος σε ακραίες τιμές

2.3. Logistic Regression

Η λογιστική παλινδρόμηση λειτουργεί εξαγοντάς ένα σύνολο σταθμισμένων χαρακτηριστικών από την είσοδο και συνδυάζοντας τα γραμμικά. Κάθε χαρακτηριστικό πολλαπλασιάζεται με ένα βάρος και ακολούθως αθροίζονται. Χρησιμοποιείται ευρέως σε διάφορους τομείς, όπως υγείας, οικονομικών και μάρκετινγκ για την απλότητα και την αποτελεσματικότητά του στην ταξινόμηση. Παρόλα αυτά, υποθέτει γραμμικότητα το οποίο μπορεί να προκαλέσει προβλήματα σε δεδομένα που είναι ισχυρά μη-γραμμικά. Επίσης, μπορεί να επηρεαστεί από ακραίες τιμές και μη ισορροπημένα σύνολα δεδομένων.

2.4. Decision Trees

Τα Decision Trees χωρίζουν αναδρομικά τα δεδομένα με βάση τα χαρακτηριστικά τους για να κάνουν προβλέψεις ή ταξινόμησης. Ο αλγόριθμος ξεκινά με ολόκληρο το σύνολο δεδομένων και επιλέγει το καλύτερο χαρακτηριστικό για να χωρίσει τα δεδομένα με βάση ορισμένα κριτήρια. Συνεχίζει να χωρίζει τα δεδομένα σε μικρότερα υποσύνολα σε κάθε κόμβο, δημιουργώντας μια δομή που μοιάζει με δέντρο έως ότου ικανοποιηθεί ένα κριτήριο διακοπής ή να μην μπορεί να γίνει περαιτέρω βελτίωση. Κάθε κόμβος φύλλου αντιπροσωπεύει μια προβλεπόμενη κλάση ή αποτέλεσμα με βάση την πλειονότητα των περιπτώσεων σε αυτό το υποσύνολο. Κατά τη διάρκεια της πρόβλεψης, τα νέα δεδομένα διασχίζουν το δέντρο, ακολουθώντας τη διαδρομή που καθορίζεται από τις συνθήκες χαρακτηριστικών, μέχρι να φτάσουν σε έναν κόμβο φύλλου για την τελική ταξινόμηση.

Τα Δέντρα Απόφασης είναι απλά, ερμηνεύσιμα και μπορούν να χειριστούν μη γραμμικές σχέσεις. Ωστόσο, περιλαμβάνουν την τάση υπερβολικής προσαρμογής εάν δεν ρυθμιστούν σωστά. Είναι ευαίσθητα σε μικρές διακυμάνσεις στα δεδομένα και ασταθής, καθώς μικρές αλλαγές στην είσοδο μπορεί να οδηγήσουν σε διαφορετικές δομές δέντρων.

2.5. k-Nearest Neighbors

Ο k-Nearest Neighbors (k-NN) είναι ένας αλγόριθμος που κάνει προβλέψεις με βάση την ομοιότητα των δεδομένων στον χώρο χαρακτηριστικών. Με δεδομένη μια νέα είσοδο, ο k-NN προσδιορίζει τους k πιο κοντινούς γείτονες στο σύνολο δεδομένων εκπαίδευσης με βάση μια μέτρηση απόστασης (π.χ. Ευκλείδεια απόσταση) και εκχωρεί την ετικέτα κλάσης με βάση την πλειοψηφία ή τον σταθμισμένο μέσο όρο των ετικετών αυτών των γειτόνων. Η τιμή του k, που αντιπροσωπεύει τον αριθμό των πλησιέστερων γειτόνων που εξετάζονται, είναι μια παράμετρος που μπορεί να προσαρμοστεί.

Μπορεί να χειριστεί ταξινόμηση πολλαπλών κλάσεων, μη γραμμικά όρια αποφάσεων και μπορεί να προσαρμοστεί στις μεταβαλλόμενες κατανομές δεδομένων. Ωστόσο, ο k-NN μπορεί να είναι υπολογιστικά ακριβός, ειδικά με μεγάλα σύνολα δεδομένων. Η απόδοση του αλγορίθμου βασίζεται σε μεγάλο βαθμό στην επιλογή του k και της κατάλληλης μέτρησης απόστασης. Είναι επίσης ευαίσθητος σε άσχετα χαρακτηριστικά και στην παρουσία θορυβωδών ή μη ισορροπημένων δεδομένων. Συχνά απαιτούνται βήματα προεπεξεργασίας, όπως η κλιμάκωση χαρακτηριστικών.

2.6. Naïve Bayes

Ο Naïve Bayes είναι ένας πιθανοτικός αλγόριθμος που βασίζεται στο θεώρημα του Bayes και στην υπόθεση της ανεξαρτησίας χαρακτηριστικών. Υπολογίζει την πιθανότητα ενός δεδομένου να ανήκει σε μια συγκεκριμένη κλάση συνδυάζοντας τις πιθανότητες των μεμονωμένων χαρακτηριστικών του να εμφανίζονται σε αυτήν την κλάση. Υποθέτει ότι τα χαρακτηριστικά είναι υπό όρους ανεξάρτητα, που σημαίνει ότι η παρουσία ή η απουσία ενός χαρακτηριστικού δεν επηρεάζει τις πιθανότητες άλλων χαρακτηριστικών.

Χρησιμοποιείται συνήθως στην ταξινόμηση κειμένου, στο φιλτράρισμα ανεπιθύμητων μηνυμάτων, στην ανάλυση συναισθημάτων και στα συστήματα συστάσεων. Τα πλεονεκτήματα του Naïve Bayes περιλαμβάνουν την απλότητα, την αποτελεσματικότητα και την ικανότητά του να χειρίζεται δεδομένα υψηλών διαστάσεων. Μπορεί να λειτουργήσει καλά με μικρά σετ εκπαίδευσης και είναι λιγότερο επιρρεπές σε υπερπροσαρμογή. Ωστόσο, η υπόθεση της ανεξαρτησίας χαρακτηριστικών μπορεί να μην ισχύει σε σενάρια πραγματικού κόσμου, γεγονός που μπορεί να επηρεάσει την ακρίβειά του.

2.7. Support Vector Machines

Τα Support Vector Machines (SVM) λειτουργούν βρίσκοντας ένα βέλτιστο υπερεπίπεδο σε ένα χώρο χαρακτηριστικών υψηλών διαστάσεων που διαχωρίζει στο μέγιστο τα δεδομένα διαφορετικών κλάσεων. Ο αλγόριθμος στοχεύει να βρει το καλύτερο όριο απόφασης μεγιστοποιώντας το περιθώριο, το οποίο είναι η απόσταση μεταξύ του υπερεπίπεδου και των πλησιέστερων σημείων δεδομένων από κάθε κατηγορία. Χρησιμοποιώντας μια συνάρτηση, τα SVM μπορούν να μετατρέψουν τα δεδομένα σε χώρο υψηλότερης διάστασης, επιτρέποντας τον διαχωρισμό μη γραμμικά διαχωρίσιμων κλάσεων. Κατά τη διάρκεια της πρόβλεψης, τα νέα σημεία δεδομένων χαρτογραφούνται στον ίδιο χώρο χαρακτηριστικών και ταξινομούνται με βάση τη θέση τους σε σχέση με το όριο απόφασης.

Τα SVM είναι ικανά να χειρίζονται δεδομένα υψηλών διαστάσεων, να αντιμετωπίζουν αποτελεσματικά τα μη γραμμικά όρια αποφάσεων και να παρέχουν ισχυρή απόδοση γενίκευσης. Τα SVM είναι λιγότερο επιρρεπές σε υπερπροσαρμογή και λειτουργούν καλά με περιορισμένα δεδομένα εκπαίδευσης. Επιπλέον, με τη χρήση διαφορετικών συναρτήσεων, τα SVM μπορούν να καταγράψουν πολύπλοκες σχέσεις και μπορούν να είναι ευέλικτα στη μοντελοποίηση διαφόρων τύπων δεδομένων.

Ωστόσο, η επιλογή της συνάρτησης και των υπερπαραμέτρων του πυρήνα μπορούν να επηρεάσουν σε μεγάλο βαθμό την απόδοσή τους.

2.8. Neural Networks

Τα νευρωνικά δίκτυα, είναι μια κατηγορία μοντέλων μηχανικής μάθησης εμπνευσμένα από τη δομή και τη λειτουργία του ανθρώπινου εγκεφάλου. Αποτελούνται από διασυνδεδεμένους «νευρώνες» οργανωμένους σε στρώματα. Τα βάρη και τα bias που σχετίζονται με τις συνδέσεις μεταξύ των νευρώνων προσαρμόζονται κατά τη διάρκεια της εκπαίδευσης για τη βελτιστοποίηση της απόδοσης του δικτύου. Κατά την εκπαίδευση το δίκτυο προσπαθεί να ελαχιστοποιεί τη διαφορά μεταξύ της προβλεπόμενης εξόδου και της πραγματικής εξόδου προσαρμόζοντας επαναληπτικά τα βάρη.

Τα νευρωνικά δίκτυα είναι ικανά να μαθαίνουν πολύπλοκα μοτίβα, μη γραμμικές σχέσεις και ιεραρχικές αναπαραστάσεις, καθιστώντας τα κατάλληλα για ένα ευρύ φάσμα εργασιών, όπως η αναγνώριση εικόνας και ομιλίας, η επεξεργασία φυσικής γλώσσας και η ανάλυση χρονοσειρών. Μπορούν να χειριστούν αποτελεσματικά μεγάλα σύνολα δεδομένων και χώρους εισόδου υψηλών διαστάσεων. Παρόλα αυτά, μπορεί να είναι υπολογιστικά ακριβά και απαιτούν σημαντικούς υπολογιστικούς πόρους και χρόνο για εκπαίδευση, ειδικά για αρχιτεκτονικές σε βάθος. Συχνά απαιτούν μεγάλο αριθμό σημασμένων δεδομένων εκπαίδευσης για να γενικεύσουν καλά. Επιπλέον, τα νευρωνικά δίκτυα μπορεί να είναι επιρρεπή στην υπερπροσαρμογή αν δεν γίνει σωστή επιλογή των υπερπαραμέτρων.

2.9. Deep Neural Network (DNN)

Τα νευρωνικά δίκτυα βαθιάς μάθησης, έχουν σχεδιαστεί για να μοντελοποιούν πολύπλοκα μοτίβα και ιεραρχικές αναπαραστάσεις ενσωματώνοντας πολλαπλά κρυφά επίπεδα. Αυτά τα δίκτυα αποτελούνται από διασυνδεδεμένους «νευρώνες» οργανωμένους σε επίπεδα. Τα κρυφά επίπεδα επιτρέπουν στο δίκτυο να μαθαίνει όλο και πιο αφηρημένα χαρακτηριστικά σε κάθε επίπεδο βάθους. Κάνουν χρήση του backpropagation για να προσαρμόσουν επαναληπτικά τα βάρη βελτιστοποιώντας την απόδοση του δικτύου.

Τα πλεονεκτήματα των νευρωνικών δικτύων βαθιάς μάθησης περιλαμβάνουν την ικανότητά τους να μαθαίνουν πολύπλοκα μοτίβα, να εξάγουν αυτόματα χαρακτηριστικά και να χειρίζονται μεγάλα και υψηλών διαστάσεων σύνολα δεδομένων. Τα βαθιά δίκτυα μπορούν να καταγράψουν περίπλοκες σχέσεις και να επιτύχουν ανώτερη απόδοση, συχνά ξεπερνώντας τους παραδοσιακούς αλγόριθμους μηχανικής μάθησης. Ωστόσο, τα δίκτυα βαθιάς μάθησης συνοδεύονται από ορισμένες προκλήσεις. Απαιτούν σημαντική ποσότητα σημασμένων δεδομένων εκπαίδευσης, εκτεταμένους υπολογιστικούς πόρους για εκπαίδευση λόγω της πολυπλοκότητάς τους, είναι επιρρεπή σε υπερπροσαρμογή και είναι ευαίσθητα στις επιλογές των υπερπαραμέτρων.

2.10. Convolutional Neural Network (CNN)

Τα συνελκτικά νευρωνικά δίκτυα (CNN) είναι ένας εξειδικευμένος τύπος νευρωνικού δικτύου βαθιάς μάθησης που έχει σχεδιαστεί για την αποτελεσματική επεξεργασία και ανάλυση οπτικών δεδομένων. Τα CNN αξιοποιούν την έννοια της συνέλιξης, όπου εφαρμόζονται φίλτρα στις εικόνες εισόδου για εξαγωγή σχετικών χαρακτηριστικών. Η αρχιτεκτονική του δικτύου αποτελείται από πολλαπλά επίπεδα, συμπεριλαμβανομένων των συνελκτικών επιπέδων, των επιπέδων pooling και των πλήρως συνδεδεμένων επιπέδων. Στα συνελκτικά επίπεδα, τα φίλτρα ολισθαίνουν κατά μήκος της εισόδου, εκτελώντας πολλαπλασιασμούς βάσει στοιχείων και συγκεντρώνοντας τα αποτελέσματα για

τη δημιουργία χαρτών χαρακτηριστικών. Τα επίπεδα pooling μειώνουν τις χωρικές διαστάσεις των χαρτών χαρακτηριστικών και τα πλήρως συνδεδεμένα επίπεδα επεξεργάζονται τα εξαγόμενα χαρακτηριστικά για να κάνουν προβλέψεις.

Ο συνδυασμός συνελκτικών λειτουργιών και της ιεραρχικής εξαγωγής χαρακτηριστικών επιτρέπει στα CNN να μαθαίνουν και να αναγνωρίζουν πολύπλοκα μοτίβα και δομές σε εικόνες, καθιστώντας τα ιδιαίτερα αποτελεσματικά σε εργασίες όπως ταξινόμηση εικόνων, ανίχνευση αντικειμένων και τμηματοποίηση εικόνων. Ωστόσο, τα CNN απαιτούν σημαντικούς υπολογιστικούς πόρους και μπορεί να είναι υπολογιστικά ακριβά, ιδιαίτερα για βαθύτερες αρχιτεκτονικές και η ανάγκη για μεγάλες ποσότητες δεδομένων εκπαίδευσης με ετικέτα είναι επίσης ένας περιορισμός.

2.11. Recurrent Neural Network (RNN)

Τα RNN είναι ένας τύπος νευρωνικού δικτύου βαθιάς μάθησης που έχει σχεδιαστεί για την επεξεργασία χρονικών δεδομένων διατηρώντας μια εσωτερική μνήμη. Σε αντίθεση με τα νευρωνικά δίκτυα, τα RNN έχουν συνδέσεις που σχηματίζουν βρόχους, επιτρέποντας στις πληροφορίες να παραμένουν και να μεταβιβάζονται από το ένα βήμα στο επόμενο. Αυτός ο μηχανισμός μνήμης επιτρέπει στα RNN να συλλαμβάνουν χρονικές εξαρτήσεις και να μοντελοποιούν ακολουθίες διαφορετικών μηκών. Τα RNN επεξεργάζονται διαδοχικά δεδομένα λαμβάνοντας εισαγωγή σε κάθε χρονικό βήμα και ενημερώνοντας την εσωτερική τους κατάσταση με βάση την τρέχουσα είσοδο και την προηγούμενη κατάσταση. Αυτή η δυναμική φύση επιτρέπει στα RNN να χειρίζονται εργασίες όπως η μοντελοποίηση γλώσσας, η αναγνώριση ομιλίας, η μηχανική μετάφραση και η ανάλυση χρονοσειρών.

Τα RNN μπορούν να χειριστούν είσοδο μεταβλητών μηκών και μπορούν να συλλάβουν μακροπρόθεσμες εξαρτήσεις μέσα στα δεδομένα. Έχουν τη δυνατότητα να δημιουργούν προβλέψεις ή εξόδους σε κάθε χρονικό βήμα. Η εκπαίδευση των RNN είναι επιρρεπείς στα προβλήματα του vanishing gradient και exploding gradient τα οποία επηρεάζουν την ικανότητά τους να συλλαμβάνουν μακροπρόθεσμες εξαρτήσεις.

2.12. Long Short-Term Memory Network (LSTM)

Τα δίκτυα LSTM είναι ένας τύπος αρχιτεκτονικής recurrent νευρωνικού δικτύου (RNN) που έχει σχεδιαστεί για την αντιμετώπιση του προβλήματος του vanishing gradient και της καταγραφής μακροπρόθεσμων εξαρτήσεων σε χρονικά δεδομένα. Τα LSTM εισάγουν κελιά μνήμης που επιτρέπουν την αποθήκευση πληροφοριών και την επιλεκτική ενημέρωση με την πάροδο του χρόνου. Κάθε cell έχει τρία κύρια στοιχεία: μια πύλη εισόδου, μια πύλη λήθης και μια πύλη εξόδου. Η πύλη εισόδου καθορίζει πόσες νέες πληροφορίες πρέπει να αποθηκευτούν στο κελί, ενώ η πύλη λήθης ελέγχει την ποσότητα των παλιών πληροφοριών που πρέπει να απορριφθούν. Η πύλη εξόδου καθορίζει πόσες πληροφορίες από το κελί πρέπει να περάσουν στο επόμενο βήμα. Ρυθμίζοντας δυναμικά τη ροή πληροφοριών, τα LSTM μπορούν να διατηρούν και να ενημερώνουν αποτελεσματικά τις σχετικές πληροφορίες σε μεγάλες αλληλουχίες, καθιστώντας τα κατάλληλα για εργασίες που περιλαμβάνουν μακροχρόνιες εξαρτήσεις.

2.13. Gated Recurrent Unit (GRU)

Το Gated Recurrent Unit (GRU) είναι μια παραλλαγή της αρχιτεκτονικής RNN που αντιμετωπίζει το πρόβλημα του vanishing gradient και καταγράφει μακροπρόθεσμες εξαρτήσεις σε χρονικά δεδομένα. Παρόμοια με το LSTM, το GRU εισάγει επίσης μηχανισμούς πύλης για τον έλεγχο

της ροής πληροφοριών. Το GRU έχει δύο πύλες: την πύλη ενημέρωσης και την πύλη επαναφοράς. Η πύλη ενημέρωσης καθορίζει τον όγκο των νέων πληροφοριών που θα αποθηκευτούν στην κρυφή κατάσταση, ενώ η πύλη επαναφοράς ελέγχει πόσο μεγάλο μέρος της προηγούμενης κρυφής κατάστασης θα πρέπει να ξεχαστεί. Με την προσαρμοστική ενημέρωση και επαναφορά της κρυφής κατάστασης, το GRU μπορεί να διατηρήσει αποτελεσματικά πληροφορίες και να μοντελοποιήσει μακροπρόθεσμες εξαρτήσεις σε ακολουθίες. Το GRU προσφέρει μια απλούστερη αρχιτεκτονική σε σύγκριση με το LSTM, με λιγότερες παραμέτρους και υπολογισμούς, καθιστώντας το υπολογιστικά αποδοτικό και ευκολότερο στην εκπαίδευση.

3. Σύνολα Δεδομένων

3.1. Μέρος Α: Ελληνικές Εταιρίες

Το σύνολο δεδομένων περιέχει 13 χαρακτηριστικά για 10716 ελληνικές εταιρίες. Αναλυτικά, περιέχει 8 δείκτες απόδοσης, 3 δυνικούς δείκτες δραστηριοτήτων, το έτος που αφορά τα προηγούμενα μεγέθη και την κατάσταση της εταιρίας (1 όλα καλά, 2 έχει κηρύξει χρεωκοπία). Το χαρακτηριστικό «έτος» αφαιρείται από τα δεδομένα προτού προχωρήσουμε στην δημιουργία των μοντέλων. Οι καταστάσεις 1 και 2 των εταιριών αντιπροσωπεύουν τις κλάσεις των δεδομένων. Χρησιμοποιούνται για την εκπαίδευση και έλεγχο των μοντέλων ως η πραγματική έξοδος των δεδομένων.

365* (Β.Υ/ Κοστ.Πωλ)	Λειτ.Αποτ/ Συν.Ενεργ. (ROA)	ΧΡΗΜ.ΛΑΠΑΝΕΣ / ΠΩΛΗΣΕΙΣ	ΠΡΑΓΜΑΤΙΚΗ ΡΕΥΣΤΟΤΗΤΑ : (ΚΕ- ΑΠΟΘΕΜΑΤΑ) / Β.Υ	(ΑΠΑΙΤ.*365) / ΠΩΛ.	Συν.Υποχρ/ Συν.Ενεργ	Διάρκεια Παραμονής Αποθεμάτων	Λογαριθμός Προσωπικού	• • • • •
1014.128862	0.011751247	0.038505198	0.965479078	567.86655	0.532257874	545.195964	0.301029996	•
561.9746307	0.099662363	0.015775343	1.415466546	335.917481	0.660555911	7.95091011	1.176091259	•
826.9190253	0.009593744	0.093561963	1.575033844	350.6180862	0.550912298	28.60222089	1.204119983	•
493.1769559	0.19628428	0.000580726	1.346779656	414.5087088	0.683716521	1.216887668	1.431363764	•
100.4214655	0.1790376	0.000173132	3.299686929	118.5483807	0.254495823	52.34632125	1.51851394	•

• • • • •	ΕΝΔΕΙΞΗ ΕΞΑΓΩΓΩΝ	ΕΝΔΕΙΞΗ ΕΙΣΑΓΩΓΩΝ	ΕΝΔΕΙΞΗ ΑΝΤΙΠΡΟΣΩΠΕΙΩΝ	ΕΝΔΕΙΞΗ ΑΣΥΝΕΠΕΙΑΣ (=2) (v+1)	ΕΤΟΣ
•	0	1	0	1	2006
•	1	1	1	1	2006
•	0	1	1	1	2006
•	0	0	0	1	2006
•	0	1	1	1	2006

Figure 3.1.1: Δείγμα Δεδομένων – Ελληνικές Εταιρίες

3.2. Μέρος Β: Δημιουργία, εκπαίδευση, αξιολόγηση και αποθήκευση διαφορετικών αρχιτεκτονικών βαθέων δικτύων σε εικόνες

Το mnist dataset περιέχει 60000 εικόνες εκπαίδευσης και 10000 εικόνες ελέγχου. Όλες οι εικόνες είναι grayscale, έχουν διαστάσεις 28*28 pixels και περιέχουν χειρόγραφα νούμερα από το 0 μέχρι το 9. Οι κλάσεις των εικόνων είναι το νούμερο που περιέχουν(0-9). Με KFold δημιουργήσαμε 6

σετ 50000 εικόνων εκπαίδευσης και οι υπόλοιπες 10000 κάθε φορά χρησιμοποιούνται ως σύνολο επικύρωσης. Το σύνολο ελέγχου των 10000 εικόνων παραμένει το ίδιο σε όλα τα μοντέλα.

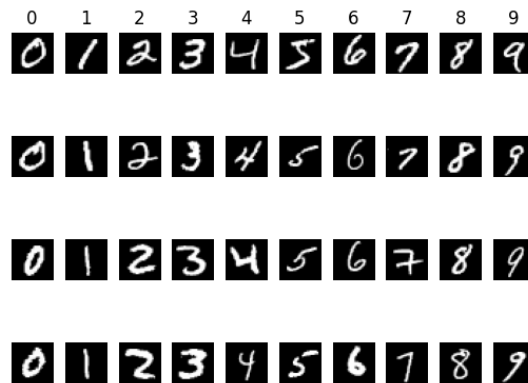


Figure 3.2.1: Mnist Dataset Sample

3.3. Μέρος Γ: Αξιολόγηση καταλληλότητας διαφορετικών αρχιτεκτονικών αναδρομικών νευρωνικών δικτύων για ενεργειακό επιμερισμό

Τα 4 .txt αρχεία που έχουμε, αφορούν την κατανάλωση ρεύματος σε σπίτι και περιέχουν τα εξής δεδομένα:

1. WaterHeatermaxAgg.txt: Μέγιστη τιμή συνολικής κατανάλωσης ρεύματος στο σπίτι σε Watt.
2. WaterHeatermaxApp.txt: Μέγιστη τιμή κατανάλωσης που παρατηρήθηκε στο Water Heater σε Watt.
3. Input_Data.txt: Περιλαμβάνει 50000 γραμμές x 120 τιμές/γραμμή. Κάθε γραμμή είναι η κατανάλωση σε kilowatt, ανά ένα λεπτό, για ένα διάστημα 120 λεπτών. Οι τιμές αφορούν την συνολική κατανάλωση στο σπίτι, σε αυτό το διάστημα.
4. Output_Data.txt: Περιλαμβάνει 50000 γραμμές x 120 τιμές/γραμμή. Κάθε γραμμή είναι η κατανάλωση σε kilowatt, ανά ένα λεπτό, για ένα διάστημα 120 λεπτών. Οι τιμές αφορούν στην κατανάλωση της συσκευής Water Heater, σε αυτό το διάστημα.

Τα WaterHeatermaxAgg και WaterHeatermaxApp μετατράπηκαν από watt σε kW ώστε να έχουμε την ίδια κλίμακα σε όλα τα δεδομένα. Βάση αυτών των 2 τιμών κανονικοποιήθηκαν τα Input_Data.txt και Output_Data.txt αντίστοιχα.

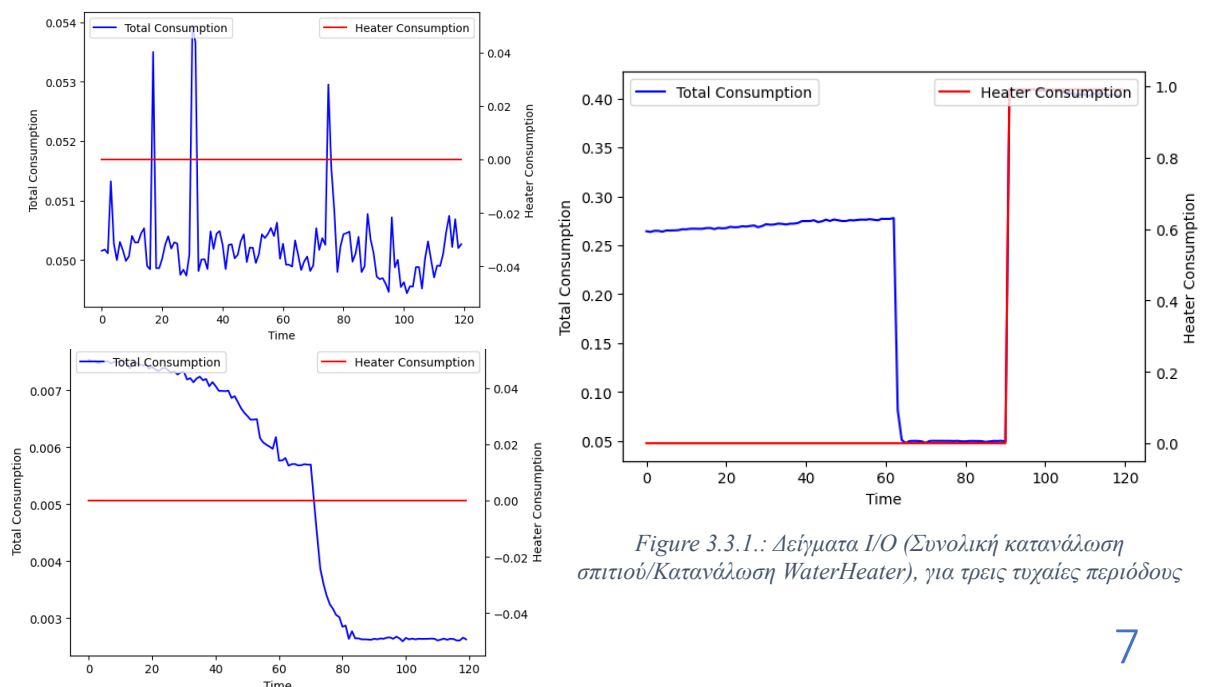


Figure 3.3.1.: Δείγματα I/O (Συνολική κατανάλωση σπιτιού/Κατανάλωση WaterHeater), για τρεις τυχαίες περιόδους

4. Αποτελέσματα-Συμπεράσματα

4.1. Μέρος Α

4.1.1. Αποτελέσματα-Μέρος Α-Ι

Classifier Name	Training or Test Set	Number of Training Samples	Number of Non-Healthy Companies in Training Sample	TP	TN	FP	FN	Precision	Recall	F1 Score	Accuracy
LinearDiscriminantAnalysis	Training	8572	200	8281	14	186	91	0.98	0.99	0.98	Total: 0.97, C1: 0.99, C2: 0.07
LinearDiscriminantAnalysis	Test	8572	200	2080	2	46	16	0.98	0.99	0.98	Total: 0.97, C1: 0.99, C2: 0.04
LogisticRegression	Training	8572	200	8370	1	199	2	0.98	1	0.99	Total: 0.98, C1: 1.0, C2: 0.0
LogisticRegression	Test	8572	200	2096	0	48	0	0.98	1	0.99	Total: 0.98, C1: 1.0, C2: 0.0
LogisticRegression-weighted	Training	8572	200	6484	156	44	###	0.99	0.77	0.87	Total: 0.77, C1: 0.77, C2: 0.78
LogisticRegression-weighted	Test	8572	200	1627	35	13	469	0.99	0.78	0.87	Total: 0.78, C1: 0.78, C2: 0.73
DecisionTree	Training	8572	200	8372	200	0	0	1	1	1	Total: 1.0, C1: 1.0, C2: 1.0
DecisionTree	Test	8572	200	2052	4	44	44	0.98	0.98	0.98	Total: 0.96, C1: 0.98, C2: 0.08
DecisionTree-weighted	Training	8572	200	8372	200	0	0	1	1	1	Total: 1.0, C1: 1.0, C2: 1.0
DecisionTree-weighted	Test	8572	200	2058	6	42	38	0.98	0.98	0.98	Total: 0.96, C1: 0.98, C2: 0.12
KNeighbors	Training	8572	200	8369	21	179	3	0.98	1	0.99	Total: 0.98, C1: 1.0, C2: 0.1
KNeighbors	Test	8572	200	2094	1	47	2	0.98	1	0.99	Total: 0.98, C1: 1.0, C2: 0.02
NaïveBayes	Training	8572	200	7906	49	151	466	0.98	0.94	0.96	Total: 0.93, C1: 0.94, C2: 0.24
NaïveBayes	Test	8572	200	1992	8	40	104	0.98	0.95	0.96	Total: 0.93, C1: 0.95, C2: 0.17
SupportVectorMachines	Training	8572	200	8372	0	200	0	0.98	1	0.99	Total: 0.98, C1: 1.0, C2: 0.0
SupportVectorMachines	Test	8572	200	2096	0	48	0	0.98	1	0.99	Total: 0.98, C1: 1.0, C2: 0.0
SupportVectorMachines-weighted	Training	8572	200	6835	170	30	###	1	0.82	0.9	Total: 0.82, C1: 0.82, C2: 0.85
SupportVectorMachines-weighted	Test	8572	200	1713	28	20	383	0.99	0.82	0.9	Total: 0.81, C1: 0.82, C2: 0.58
NeuralNetwork	Training	8572	200	8371	1	199	1	0.98	1	0.99	Total: 0.98, C1: 1.0, C2: 0.0
NeuralNetwork	Test	8572	200	2094	0	48	2	0.98	1	0.99	Total: 0.98, C1: 1.0, C2: 0.0
NeuralNetwork-weighted	Training	8572	200	6530	167	33	###	0.99	0.78	0.87	Total: 0.78, C1: 0.78, C2: 0.84
NeuralNetwork-weighted	Test	8572	200	1658	33	15	438	0.99	0.79	0.88	Total: 0.79, C1: 0.79, C2: 0.69

Τα μοντέλα που δημιουργήθηκαν δεν καταφέρνουν να πετύχουν τους περιορισμούς επιτυχίας που έχουμε (62% εταιρίες που θα πτωχεύσουν, 70% εταιρίες που δεν θα πτωχεύσουν). Ο λόγος που συμβαίνει αυτό πιθανότατα είναι επειδή το σύνολο δεδομένων μας δεν είναι ισορροπημένο. Δηλαδή, υπάρχουν 10468 εγγραφές με εταιρίες που δεν θα πτωχεύσουν και μόνο 248 που θα πτωχεύσουν. Οπότε τα μοντέλα δεν μπορούν να γενικεύσουν, με αποτέλεσμα κατά τον έλεγχο των μοντέλων να κατηγοριοποιούν όλες τις εταιρίες ως υγιείς.

Για την βελτίωση των αποτελεσμάτων μπορούμε να προσθέσουμε βάρη στις κλάσεις σε μερικά από τα μοντέλα. Όπως φαίνεται και στα αποτελέσματα στον πιο πάνω πίνακα με μπλε χρώμα, αυτά τα μοντέλα αποδίδουν πολύ καλύτερα αποτελέσματα. Επιπρόσθετα, τα μοντέλα LogisticRegression & NeuralNetwork με βάρη πετυχαίνουν και τους περιορισμούς επιτυχίας που είχαν τεθεί.

4.1.2. Αποτελέσματα-Μέρος Α-II

Classifier Name	Training or Test Set	Number of Training Samples	Number of Non-Healthy Companies in Training Sample	TP	TN	FP	FN	Precision	Recall	F1 Score	Accuracy
LinearDiscriminantAnalysis	Training	800	200	563	81	119	37	0.83	0.94	0.88	Total: 0.8, C1: 0.94, C2: 0.4
LinearDiscriminantAnalysis	Test	800	200	1967	15	33	129	0.98	0.94	0.96	Total: 0.92, C1: 0.94, C2: 0.31
LogisticRegression	Training	800	200	568	75	125	32	0.82	0.95	0.88	Total: 0.8, C1: 0.95, C2: 0.38
LogisticRegression	Test	800	200	1981	13	35	115	0.98	0.95	0.96	Total: 0.93, C1: 0.95, C2: 0.27
LogisticRegression-weighted	Training	800	200	457	156	44	143	0.91	0.76	0.83	Total: 0.77, C1: 0.76, C2: 0.78
LogisticRegression-weighted	Test	800	200	1629	35	13	467	0.99	0.78	0.87	Total: 0.78, C1: 0.78, C2: 0.73
DecisionTree	Training	800	200	600	200	0	0	1	1	1	Total: 1.0, C1: 1.0, C2: 1.0
DecisionTree	Test	800	200	1748	27	21	348	0.99	0.83	0.9	Total: 0.83, C1: 0.83, C2: 0.56
DecisionTree-weighted	Training	800	200	600	200	0	0	1	1	1	Total: 1.0, C1: 1.0, C2: 1.0
DecisionTree-weighted	Test	800	200	1737	15	33	359	0.98	0.83	0.9	Total: 0.82, C1: 0.83, C2: 0.31
KNeighbors	Training	800	200	549	130	70	51	0.89	0.92	0.9	Total: 0.85, C1: 0.92, C2: 0.65
KNeighbors	Test	800	200	1869	26	22	227	0.99	0.89	0.94	Total: 0.88, C1: 0.89, C2: 0.54
NaïveBayes	Training	800	200	529	108	92	71	0.85	0.88	0.86	Total: 0.8, C1: 0.88, C2: 0.54
NaïveBayes	Test	800	200	1844	22	26	252	0.99	0.88	0.93	Total: 0.87, C1: 0.88, C2: 0.46
SupportVectorMachines	Training	800	200	578	91	109	22	0.84	0.96	0.9	Total: 0.84, C1: 0.96, C2: 0.46
SupportVectorMachines	Test	800	200	1978	15	33	118	0.98	0.94	0.96	Total: 0.93, C1: 0.94, C2: 0.31
SupportVectorMachines-weighted	Training	800	200	467	163	37	133	0.93	0.78	0.85	Total: 0.79, C1: 0.78, C2: 0.82
SupportVectorMachines-weighted	Test	800	200	1623	35	13	473	0.99	0.77	0.87	Total: 0.77, C1: 0.77, C2: 0.73
NeuralNetwork	Training	800	200	553	90	110	47	0.83	0.92	0.87	Total: 0.8, C1: 0.92, C2: 0.45
NeuralNetwork	Test	800	200	1947	21	27	149	0.99	0.93	0.96	Total: 0.92, C1: 0.93, C2: 0.44
NeuralNetwork-weighted	Training	800	200	459	161	39	141	0.92	0.76	0.83	Total: 0.78, C1: 0.76, C2: 0.8
NeuralNetwork-weighted	Test	800	200	1620	36	12	476	0.99	0.77	0.87	Total: 0.77, C1: 0.77, C2: 0.75

Μια άλλη μέθοδος που μπορούμε να χρησιμοποιήσουμε είναι η υποδειγματοληψία της κλάσης υγιών εταιριών ώστε να έχουμε μια αναλογία 3 υγιείς προς 1 πτωχευμένη εταιρία στο σύνολο εκπαίδευσης. Με αυτή την μέθοδο παρατηρούμε ότι έχουμε γενικά καλύτερα αποτελέσματα σε όλα τα μοντέλα. Παρόλα αυτά για να πετύχουμε τους περιορισμούς επιτυχίας που αναφέρθηκαν πιο πριν πρέπει να χρησιμοποιήσουμε τα μοντέλα με βάρη στις κλάσεις. Εκτός από τα LogisticRegression & NeuralNetwork που πέτυχαν τους περιορισμούς απλά με χρήση βαρών, τώρα παρατηρούμε ότι τους πετυχαίνει και το μοντέλο SupportVectorMachines με βάρη.

4.2. Μέρος Β

Το DNN μοντέλο που δημιουργήσαμε, περιλαμβάνει 2 κρυφά στρώματα με 256 νευρώνες οι οποίοι κάνουν χρήση της συνάρτησης ενεργοποίησης ReLu. Επιπρόσθετα γίνεται Dropout και στα δύο στρώματα με ποσοστό 45%. Τέλος, έχουμε το στρώμα εξόδου που χρησιμοποιούμε την συνάρτηση ενεργοποίησης SoftMax.

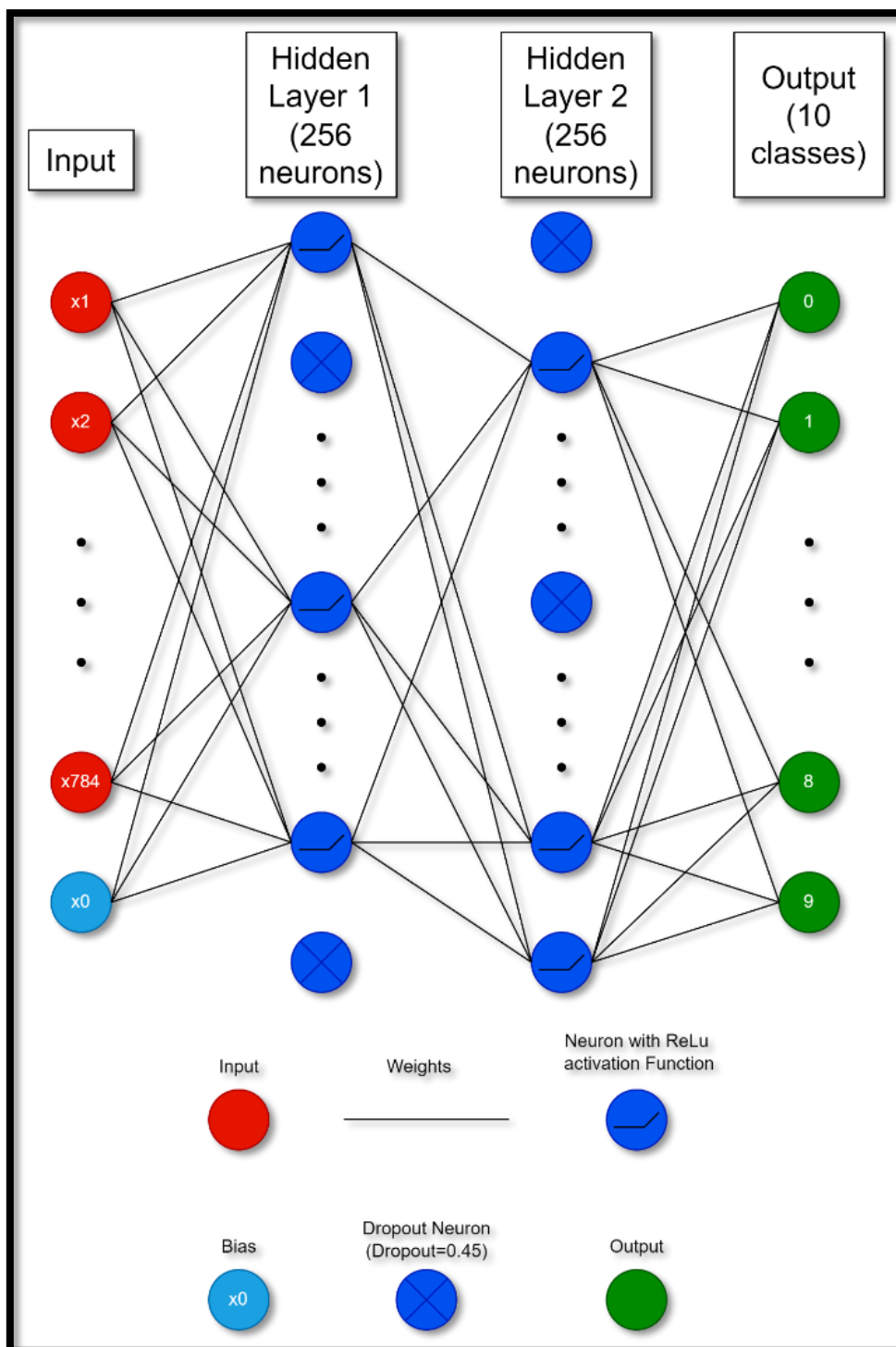


Figure 4.2.1: DNN model Architecture

Το CNN μοντέλο που δημιουργήσαμε δέχεται ως είσοδο εικόνες 28x28 pixels και δημιουργεί στο πρώτο στρώμα συνέλιξης, με την χρήση φίλτρων 3x3, 32 χάρτες χαρακτηριστικών με διαστάσεις 26x26 pixels. Επιπρόσθετα, σε αυτό το στρώμα υπάρχει Dropout ίσο με 25%.

Ακολουθώς, υπάρχει ένα στρώμα MaxPooling όπου μειώνονται οι διαστάσεις των χαρτών σε 13x13 pixels. Το επόμενο στρώμα είναι το 2^ο στρώμα συνέλιξης όπου δημιουργούνται 64 χάρτες χαρακτηριστικών διαστάσεων 11x11 pixels με την χρήση φίλτρων 3x3. Οι νέοι χάρτες περνούν από ένα νέο MaxPooling layer όπου οι διαστάσεις των χαρτών μειώνονται σε 5x5.

Τέλος, υπάρχει ένα Fully Connected στρώμα 128 τιμών με Dropout 50% και συνάρτηση ενεργοποίησης ReLu και το Fully Connected στρώμα με τις 10 εξόδους και συνάρτηση ενεργοποίησης την SoftMax.

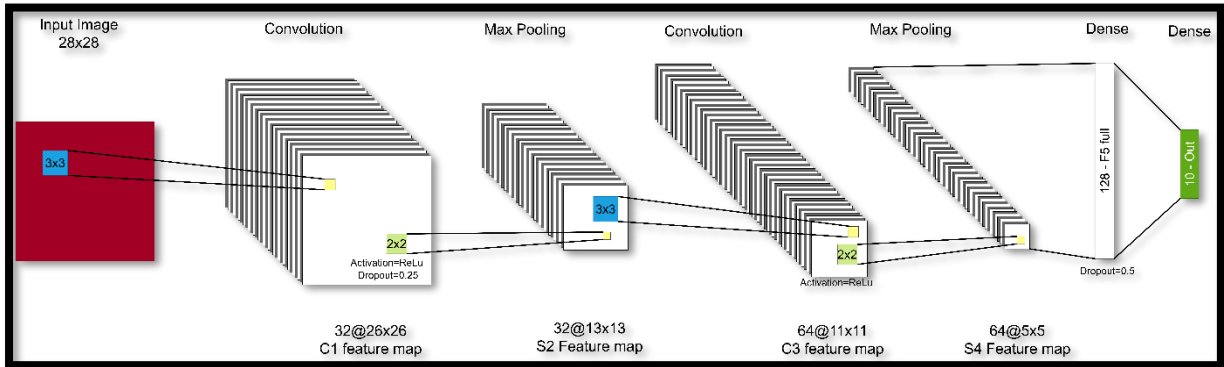


Figure 4.2.2: CNN model Architecture

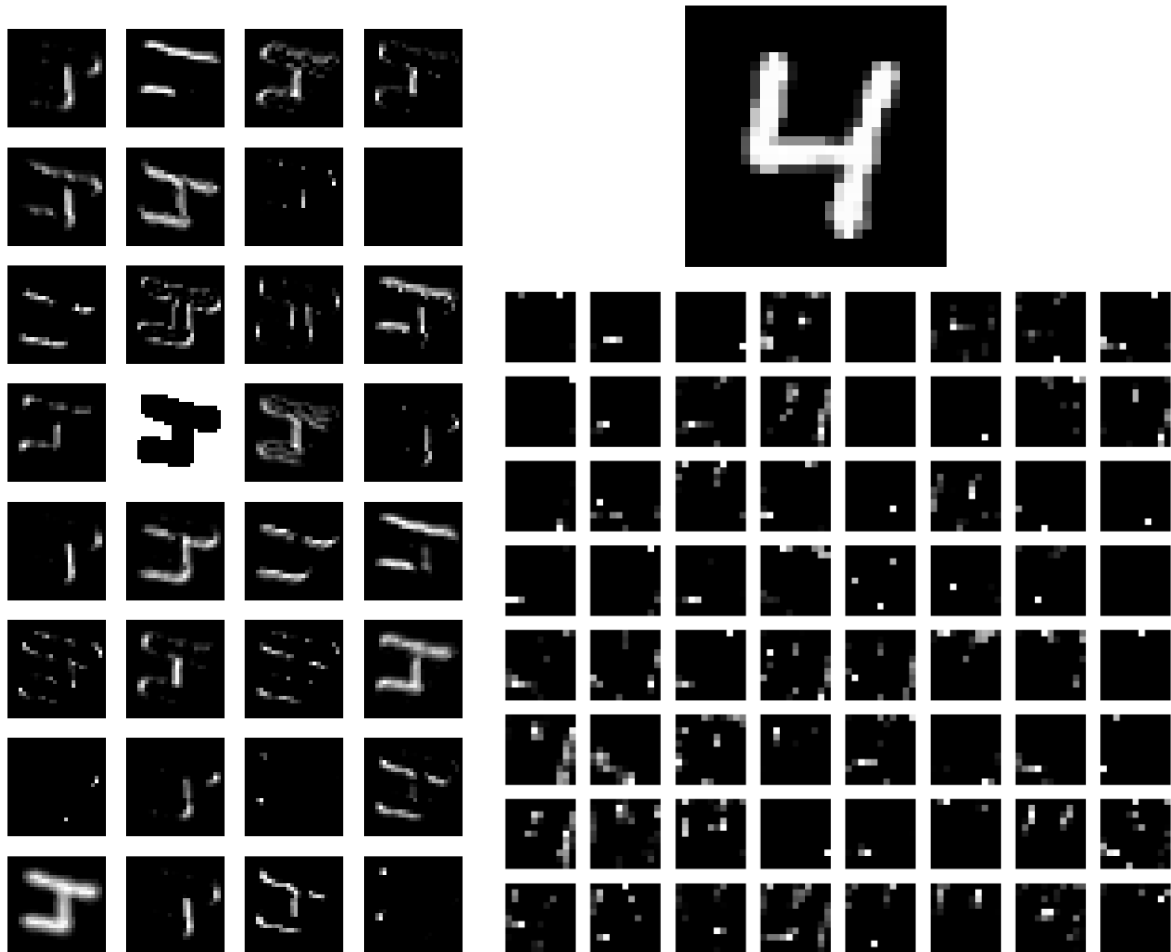


Figure 4.2.3.: Feature maps of top right image. First convolution layer(left), Second convolution layer(right)

Παρατηρούμε ότι στο πρώτο convolution layer (αριστερά εικόνα) δημιουργούνται κάποια χαρακτηριστικά όπως κάθετες και οριζόντιες ακμές ενώ στο 2^ο (δεξιά-κάτω εικόνα) δημιουργούνται πιο αφηρημένα χαρακτηριστικά για την εικόνα με τον αριθμό 4 που φαίνεται στην πάνω δεξιά εικόνα.

4.2.4. Αποτελέσματα-Μέρος Β

Technique Name	Set	Fold number	Accuracy	Precision	Recall	F1 Score
DNN	Train	1	0.9893	0.9891	0.9893	0.9892
DNN	Test	1	0.9753	0.975	0.9752	0.975
DNN	Train	2	0.9943	0.9943	0.9942	0.9942
DNN	Test	2	0.9811	0.9811	0.9808	0.9809
DNN	Train	3	0.9958	0.9958	0.9957	0.9957
DNN	Test	3	0.981	0.981	0.9806	0.9808
DNN	Train	4	0.997	0.9971	0.997	0.997
DNN	Test	4	0.9822	0.9823	0.9821	0.9822
DNN	Train	5	0.9977	0.9977	0.9977	0.9977
DNN	Test	5	0.9832	0.9831	0.9831	0.9831
DNN	Train	6	0.9982	0.9981	0.9982	0.9982
DNN	Test	6	0.9829	0.9828	0.9828	0.9828
CNN	Train	1	0.9981	0.998	0.9981	0.9981
CNN	Test	1	0.9913	0.9912	0.9913	0.9912
CNN	Train	2	0.9992	0.9992	0.9992	0.9992
CNN	Test	2	0.9936	0.9936	0.9935	0.9935
CNN	Train	3	0.9999	0.9999	0.9999	0.9999
CNN	Test	3	0.9941	0.9941	0.994	0.9941
CNN	Train	4	0.9996	0.9996	0.9996	0.9996
CNN	Test	4	0.9923	0.9923	0.9921	0.9922
CNN	Train	5	1	1	1	1
CNN	Test	5	0.9935	0.9935	0.9935	0.9935
CNN	Train	6	0.9998	0.9998	0.9998	0.9998
CNN	Test	6	0.9938	0.9937	0.9937	0.9937

Παρατηρούμε ότι τα CNN μοντέλα που δημιουργήσαμε έχουν καλύτερα αποτελέσματα από τα DNN (περίπου 1% καλύτερη κατηγοριοποίηση στο test set) στο σύνολο δεδομένων mnist. Αυτό συμβαίνει λόγω της ικανότητάς τους να εξάγουν τοπικά χαρακτηριστικά. Τα CNN καταγράφουν αποτελεσματικά τοπικά μοτίβα, μειώνουν τις χωρικές διαστάσεις και μαθαίνουν όλο και πιο περίπλοκες αναπαραστάσεις, καθιστώντας τα κατάλληλα για εργασίες αναγνώρισης ψηφίων και ταξινόμησης εικόνων.

Στα ακόλουθα γραφήματα μπορούμε να δούμε τις σωστά ταξινομημένες εικόνες ανά κλάση για το καλύτερο μοντέλο DNN και CNN που δημιουργήσαμε, καθώς επίσης και πόσες εικόνες κάποιας κλάσης ταξινομήθηκαν λάθος σε κάποιαν άλλη. Όπως αναφέρθηκε και πιο πάνω είναι φανερό ότι το μοντέλο CNN πετυχαίνει καλύτερα αποτελέσματα σε προβλήματα ταξινόμησης εικόνων οπότε και θα επιλεγεί για την ταξινόμηση χειρόγραφων ψηφίων.

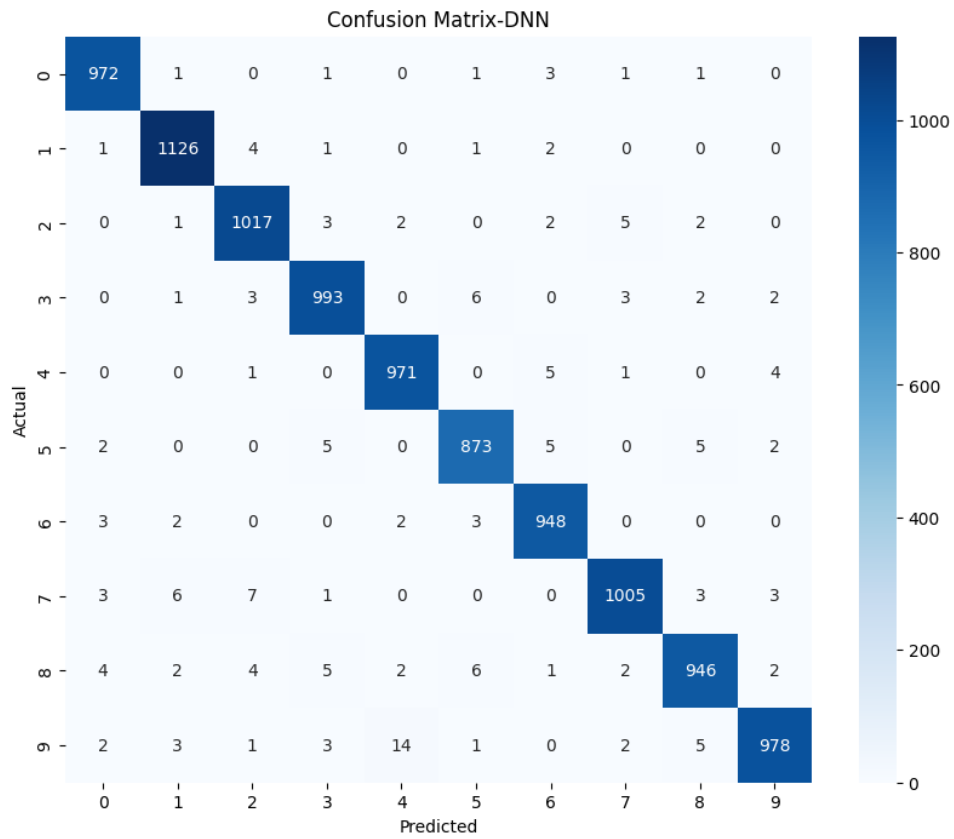


Figure 4.2.5: DNN Confusion Matrix

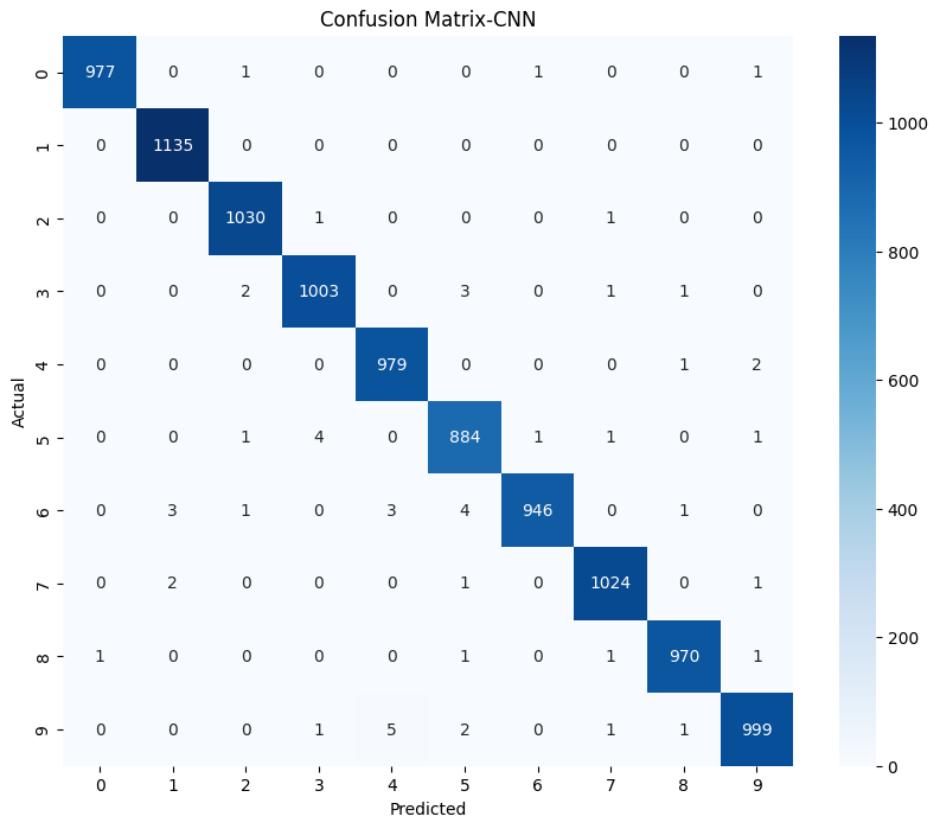


Figure 4.2.6: CNN Confusion Matrix

4.3. Μέρος Γ

Κατά την διάρκεια εκπαίδευσης των μοντέλων θέσαμε ένα περιορισμό ώστε να σταματά η εκπαίδευση αν τυχόν δεν μειωθεί το validation error για 5 συνεχόμενες εποχές. Στα μοντέλα είχαν τεθεί 40 εποχές εκπαίδευσης αλλά η εκπαίδευση του RNN και του LSTM διακόπηκε στις 15 εποχές. Το GRU εκπαιδεύτηκε για 40 εποχές όπως είχαμε θέσει.

Παρατηρούμε ότι το σφάλμα του validation set των μοντέλων RNN και LSTM είναι γύρω στο 0.011 ενώ του GRU είναι στο 0.006. Βάση του MSE το GRU μοντέλο έχει καλύτερα αποτελέσματα στο validation set αλλά και στο training set.

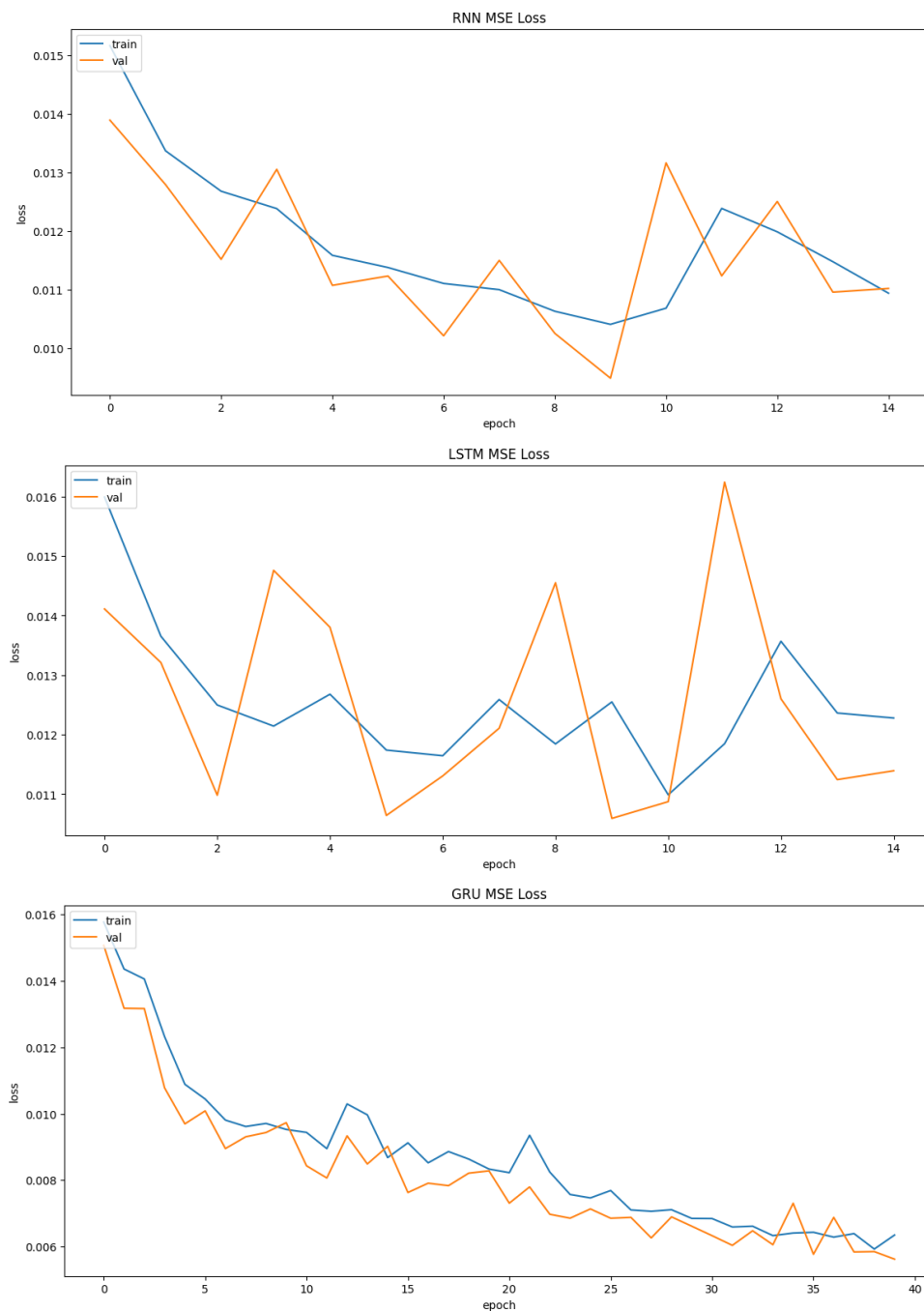


Figure 4.3.1.: MSE Loss (Top:RNN, Middle:LSTM, Bottom:GRU)

Βάση των πιο κάτω διαγραμμάτων παρατηρούμε ότι το μοντέλο GRU, εκτός από τα καλά αποτελέσματα που έχει στο training και validation set, πετυχαίνει και τις καλύτερες προβλέψεις σε νέα δεδομένα(test set). Επίσης, φαίνεται ότι το RNN μοντέλο έχει μικρότερο μέσο σφάλμα σε σύγκριση με το LSTM παρόλα αυτά έχει το μεγαλύτερο max-error που παρατηρήθηκε στην πρόβλεψη μιας περιόδου.

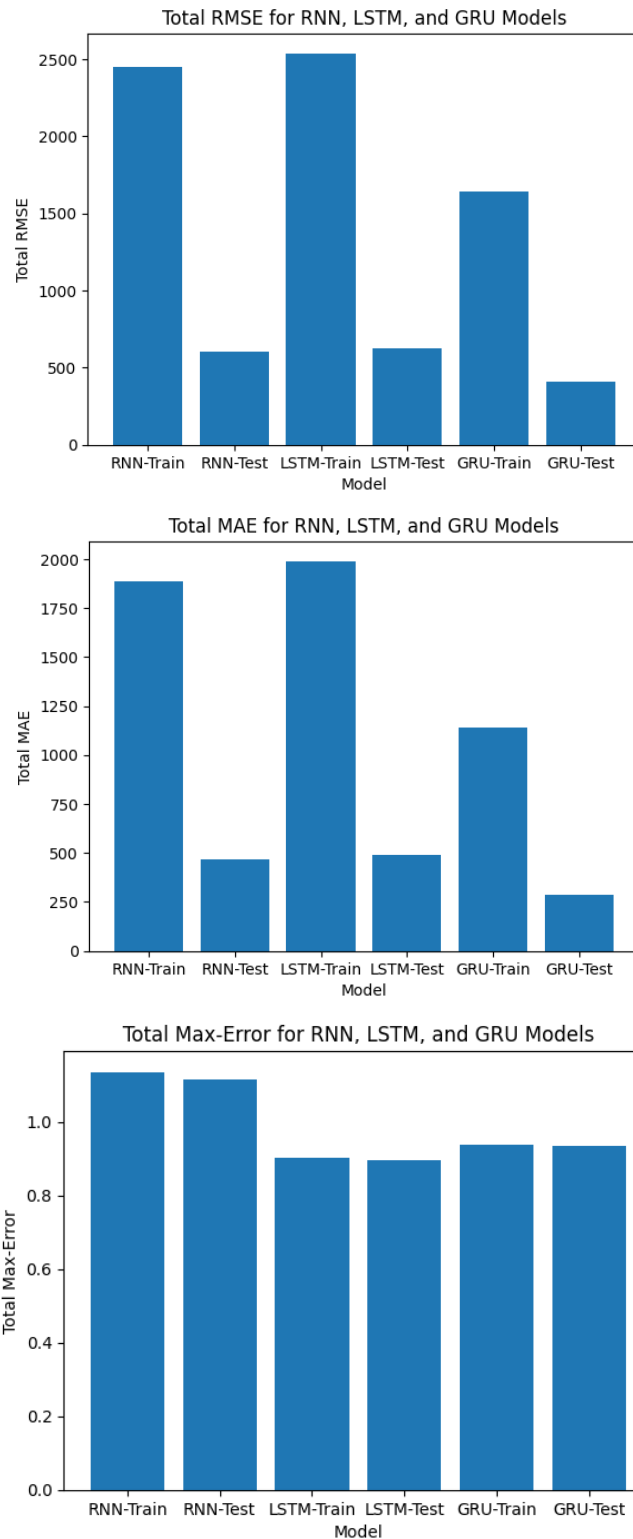


Figure 4.3.2.: RMSE, MAE, Max Error, of Train & Test Set of all models

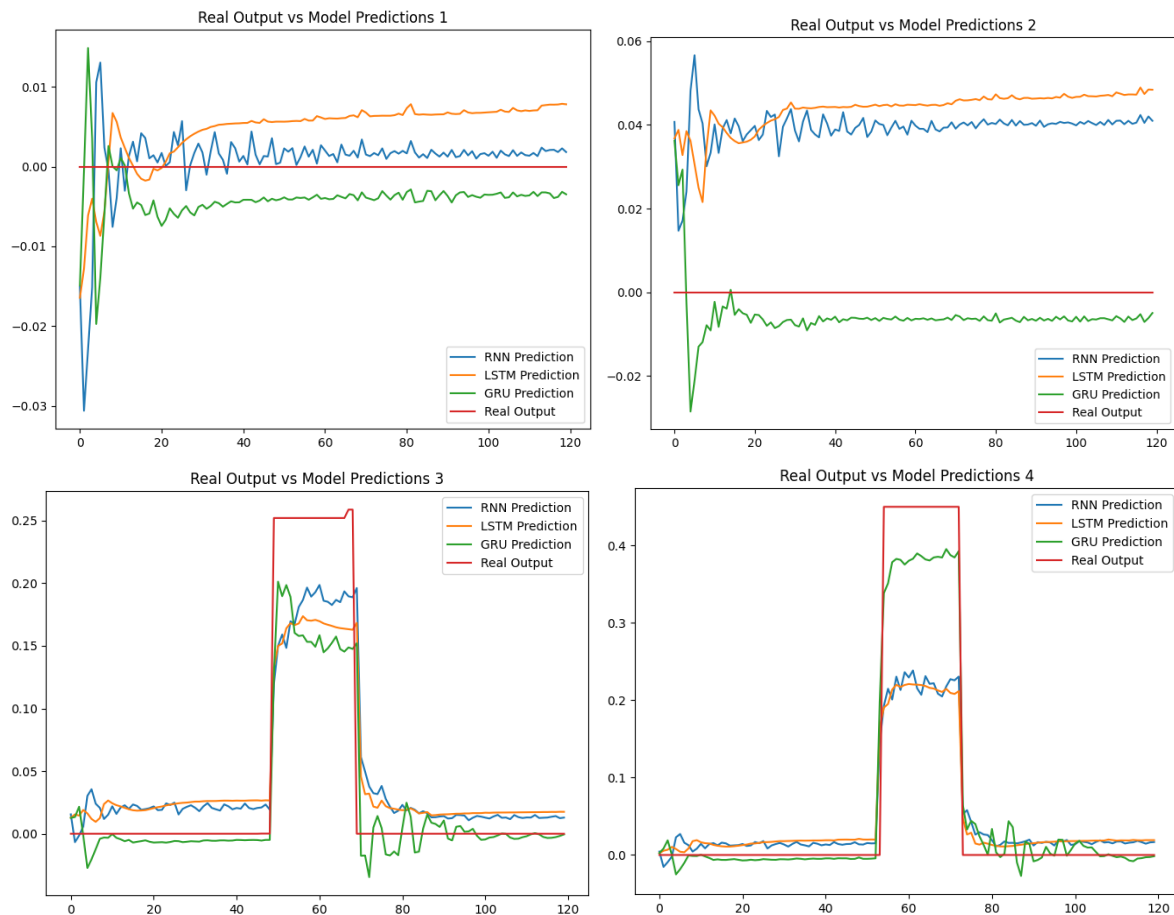


Figure 4.3.3.: Comparison of Real and Predicted Outputs

Στα πιο πάνω διαγράμματα βλέπουμε την πραγματική κατανάλωση της συσκευής WaterHeater με κόκκινη γραμμή και την πρόβλεψη των μοντέλων που δημιουργήσαμε σε τέσσερις περιόδους. Το μοντέλο GRU φαίνεται να προσεγγίζει καλύτερα τις πραγματικές εξόδους. Επιπρόσθετα, φαίνεται να προσεγγίζει καλύτερα την τιμή 0, δηλαδή να κάνει πρόβλεψη ότι η συσκευή μας δεν είναι σε λειτουργία.

Οι χρόνοι εκπαίδευσης των τριών μοντέλων:

- RNN = 44,46 λεπτά
- LSTM = 3.07 λεπτά
- GRU = 7,4 λεπτά

καθιστούν το GRU μοντέλο πολύ γρηγορότερο στην εκπαίδευση σε σύγκριση με το RNN και αναλογικά πιο γρήγορο από το LSTM εφόσον το LSTM εκπαιδεύτηκε για μόνο 15 εποχές ενώ το GRU για 40.

Βάση των όσων αναφέρθηκαν πιο πάνω, το GRU μοντέλο δείχνει να είναι η καλύτερη επιλογή για προβλέψεις στο συγκεκριμένο πρόβλημα.