



**ΠΟΛΥΤΕΧΝΕΙΟ
ΚΡΗΤΗΣ**

ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΑΡΑΓΩΓΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**ΔΙΕΡΕΥΝΗΣΗ ΧΡΗΣΗΣ ΑΛΓΟΡΙΘΜΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ ΓΙΑ
ΠΡΟΒΛΕΨΗ ΛΕΙΤΟΥΡΓΙΑΣ ΑΝΤΛΙΟΣΤΑΣΙΟΥ**

ΑΝΔΡΕΑΔΑΚΗΣ ΑΝΤΩΝΙΟΣ

A.M.: 2013010020

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ: ΓΕΩΡΓΙΟΣ ΣΤΑΥΡΟΥΛΑΚΗΣ

ΧΑΝΙΑ 2022

Περιεχόμενα

Περίληψη	iv
Abstract.....	v
ΚΕΦΑΛΑΙΟ 1: Εισαγωγή	1
ΚΕΦΑΛΑΙΟ 2: Περιγραφή Λογισμικού EPANET	3
2.1 Γενική περιγραφή.....	3
2.2 Δυνατότητες υδραυλικής μοντελοποίησης	3
2.3 Περιγραφή μοντέλου δικτύου	4
ΚΕΦΑΛΑΙΟ 3: Μοντέλα Μηχανικής Μάθησης	6
3.1: Ορισμός και κύριες τεχνικές	6
3.2 Εποπτευομένη μάθηση και Μη εποπτευόμενη μάθηση	7
3.3 Τεχνικές ταξινόμησης (Classification)	9
3.4 Τεχνικές παλινδρόμησης (Regression).....	9
ΚΕΦΑΛΑΙΟ 4: Προσομοίωση Λειτουργίας Αντλιοστασίου με Μηχανική Μάθηση .	20
4.1: Περιγραφή μοντελοποίησης αντλιοστασίου	20
4.2: Διαδικασία παραγωγής δεδομένων	27
4.3 Επιλογή & επικύρωση μοντέλου μηχανικής μάθησης	29
Κεφάλαιο 5: Συμπεράσματα	63
Βιβλιογραφία	64
Παραρτήματα.....	67
Παράρτημα 1 (Δημιουργία αρχείων εκπαίδευσης).....	68
Παράρτημα 2 (Συλλογή τιμών ενέργειας, πίεσης, παροχής)	70
Παράρτημα 3 (Δημιουργία 16 προκαθορισμένων σεναρίων)	72
Παράρτημα 4 (Συλλογή τιμών ενέργειας, πίεσης, παροχής για τα 16 σενάρια)	74

Λίστα Εικόνων

Εικόνα 1 Απεικόνιση δικτύου και φυσικών παραμέτρων [5].....	4
Εικόνα 2 Τεχνικές μηχανικής μάθησης [6]	7
Εικόνα 3 Έννοιες αλγόριθμου γραμμικής παλινδρόμησης [12].....	11
Εικόνα 4 Θετική γραμμική σχέση και η εξίσωση της [12]	11
Εικόνα 5 Αρνητική γραμμική σχέση και η εξίσωση της [12]	12
Εικόνα 6 Διαχωρισμός προβλέψεων κάθε μοντέλου με χρήση πολλαπλών αλγορίθμων [16]	13
Εικόνα 7 Τεχνική bagging για την δημιουργία πρόβλεψης συνδυάζοντας προβλέψεις πολλαπλών μοντέλων [16]	14
Εικόνα 8 Εκμάθηση που γίνεται λαμβάνοντας υπόψιν τα σφάλματα προηγούμενων μοντέλων [16]	15
Εικόνα 9 Τεχνική στοίβαξης για την πρόβλεψη τελικού αποτελέσματος [16]	15
Εικόνα 10 Δομή δέντρου απόφασης [18].....	16
Εικόνα 11 Δομή νευρώνα [19]	18
Εικόνα 12 Δομή νευρωνικού δικτύου [19]	18
Εικόνα 13 Απεικόνιση λειτουργίας νευρωνικού δικτύου [19]	19
Εικόνα 14 Μοντελοποίηση αντλιοστασίου στο λογισμικό EPANET.....	20
Εικόνα 15 Πρώτο μέρος του δικτύου	21
Εικόνα 16 Δεύτερο μέρος του δικτύου	21
Εικόνα 17 Τρίτο μέρος του δικτύου.....	22
Εικόνα 18 Χαρακτηριστική αντλίας.	24
Εικόνα 19 Καμπύλη σχέσης απόδοσης-παροχής.	25
Εικόνα 20 Υδραυλικό δίκτυο.	26
Εικόνα 21 Διαδικασία εισαγωγής δεδομένων	31
Εικόνα 22 Εισαγωγή δεδομένων	32
Εικόνα 23 Επιλογή εκπαίδευσης μοντέλων.....	32
Εικόνα 24 Βέλτιστο μοντέλο βάση RMSE	33
Εικόνα 25 Εισαγωγή δεδομένων δοκιμής	35
Εικόνα 26 Αυξημένη απόκλιση μέσης τετραγωνικής τιμής	35
Εικόνα 27 RMSE, Var8, 100 αρχεία.....	36
Εικόνα 28 Βέλτιστο μοντέλο για την μεταβλητή 9, 1000 αρχεία	46

Λίστα Γραφικών Παραστάσεων

Γραφική Παράσταση 1 Απόκριση τιμών.....	33
Γραφική Παράσταση 2 Πρόβλεψη μοντέλου	34
Γραφική Παράσταση 3 Πρόβλεψη τιμών για τα 16 σενάρια	35
Γραφική Παράσταση 4 Απόκριση τιμών, Var8, 100 αρχεία.	37
Γραφική Παράσταση 5 Πρόβλεψη τιμών, Var8, 100 αρχεία.....	38
Γραφική Παράσταση 6 Πρόβλεψη τιμών, 16 σενάρια, Var8, 100 αρχεία	39
Γραφική Παράσταση 7 Πρόβλεψη τιμών, Var9, 100 αρχεία.....	40
Γραφική Παράσταση 8 Πρόβλεψη για τα 16 σενάρια, Var9, 100 αρχεία.....	41
Γραφική Παράσταση 9 Πρόβλεψη τιμών, Var10, 100 αρχεία	42
Γραφική Παράσταση 10 Πρόβλεψη τιμών, 16 σενάρια, Var10, 100 αρχεία	43
Γραφική Παράσταση 11 Απόκριση τιμών, Var8, 1000 αρχεία	44
Γραφική Παράσταση 12 Πρόβλεψη τιμών, Var8, 1000 αρχεία	45
Γραφική Παράσταση 13 Πρόβλεψη τιμών, 16 σενάρια, Var18, 1000 αρχεία	46
Γραφική Παράσταση 14 Απόκριση τιμών, Var9, 1000 αρχεία	47
Γραφική Παράσταση 15 Πρόβλεψη τιμών, Var9, 1000 αρχεία	48
Γραφική Παράσταση 16 Πρόβλεψη τιμών, 16 σενάρια, Var9, 1000 αρχεία	49
Γραφική Παράσταση 17 Απόκριση τιμών, Var10, 1000 αρχεία	50
Γραφική Παράσταση 18 Πρόβλεψη τιμών, Var10, 1000 αρχεία	51
Γραφική Παράσταση 19 Πρόβλεψη τιμών, Var10, 16 σενάρια, 1000 αρχεία	52
Γραφική Παράσταση 20 Απόκριση τιμών, Var8, 10000 αρχεία	53
Γραφική Παράσταση 21 Πρόβλεψη τιμών, Var8, 10000 αρχεία	54
Γραφική Παράσταση 22 Πρόβλεψη τιμών 16 σενάρια, Var8, 10000 αρχεία	55
Γραφική Παράσταση 23 Απόκρισης τιμών, Var9, 10000 αρχεία	56
Γραφική Παράσταση 24 Πρόβλεψη τιμών, Var9, 10000 αρχεία	57
Γραφική Παράσταση 25 Πρόβλεψη τιμών, 16 σενάρια, Var9, 10000 αρχεία	58
Γραφική Παράσταση 26 Απόκριση τιμών, Var10, 10000 αρχεία	59
Γραφική Παράσταση 27 Πρόβλεψη τιμών, Var10, 10000 αρχεία	60
Γραφική Παράσταση 28 Πρόβλεψη τιμών, 16 σενάρια, Var10, 10000 αρχεία	61

Λίστα Πινάκων

Πίνακας 1 Διαστάσεις σωληνώσεων	23
Πίνακας 2 Συνδεσμολογία δικτύου	26
Πίνακας 3 Κριτήρια επιλογής μοντέλου μηχανικής μάθησης [21]	30
Πίνακας 4 Τιμές RMSE για την κατανάλωση ενέργειας (μεταβλητή 8)	62
Πίνακας 5 Τιμές RMSE για την πίεση (μεταβλητή 9).....	62
Πίνακας 6 Τιμές RMSE για την παροχή (μεταβλητή 10).....	62

Περίληψη

Στην παρούσα εργασία διερευνήθηκε η αποτελεσματικότητα χρήσης αλγορίθμων μηχανικής μάθησης για την πρόβλεψη της πίεσης, παροχής και συνολικής κατανάλωσης ενέργειας σε ένα υδραυλικό δίκτυο. Συγκεκριμένα, το δίκτυο που χρησιμοποιήθηκε ήταν ένα αντλιοστάσιο νερού με δύο ταυτόσημες αντλίες το οποίο καταθλίβει σε δεξαμενή υψηλότερης στάθμης. Μέσω κατάλληλου χειρισμού των βανών στο δίκτυο σωληνώσεων, οι αντλίες μπορούν να λειτουργήσουν ως μεμονωμένες, σε σειρά ή παράλληλα. Με χρήση του αξιόπιστου προγράμματος αριθμητικής προσομοίωσης υδραυλικών δικτύων EPANET και για πολλαπλά τυχαία σενάρια θέσης βανών και ύψους στάθμης της δεξαμενής κατάθλιψης, υπολογίστηκαν οι τιμές πίεσης στους κόμβους, της παροχής στις σωληνώσεις και της συνολικής κατανάλωσης ενέργειας. Οι τιμές αυτές χρησιμοποιήθηκαν για την εκπαίδευση πολλαπλών μοντέλων μηχανικής μάθησης - τύπου Regression - στο πρόγραμμα MATLAB. Το μοντέλο που εμφάνιζε την μικρότερη απόκλιση της μέσης τετραγωνικής τιμής (RMS) χρησιμοποιήθηκε έπειτα για την πρόβλεψη των τιμών πίεσης, παροχής και ενέργειας για 16 προκαθορισμένα σενάρια θέσης βανών και στάθμης δεξαμενής. Μετά από την σύγκριση των προβλεπόμενων τιμών αυτών με τις αντίστοιχες προϋπολογισμένες τιμές από το EPANET, αναδείχθηκε η επιτυχής δυνατότητα χρήσης της μηχανικής μάθησης για την πρόβλεψη λειτουργίας του συγκεκριμένου αντλιοστασίου.

Abstract

In this thesis, the effectiveness of using machine learning algorithms to predict the pressure, flow rate and total energy consumption in a hydraulic network was investigated. Specifically, the network used was a water pumping station with two identical pumps which depresses into a higher level reservoir. Through appropriate manipulation of the valves in the piping network, the pumps can be operated as individual pumps, in series or in parallel. Using the reliable numerical simulation program for hydraulic networks EPANET and for multiple random scenarios of valve position and depression tank level, the values of pressure at the nodes, flow rate in the piping and total energy consumption were calculated. These values were used to train multiple machine learning models - of the regression type - in MATLAB. The model showing the smallest deviation of the root mean square (RMS) was then used to predict the pressure, flow and energy values for 16 predefined scenarios of vane position and reservoir levels. After comparing these predicted values with the corresponding precalculated values from EPANET, the successful feasibility of using machine learning to predict the operation of this pumping station was demonstrated.

ΚΕΦΑΛΑΙΟ 1: Εισαγωγή

Τα λειτουργικά συστήματα ύδρευσης είναι απαραίτητα για ένα κατάλληλο αστικό περιβάλλον καθώς και για τον πληθυσμό που κατοικεί σε αυτό, λόγω του πόσιμου νερού που διανέμουν στους κατοίκους το οποίο είναι σημαντικό για την ικανοποίηση καθημερινών αναγκών αλλά και την κατανάλωση του. Επομένως, το νερό είναι σημαντικό να διανέμεται σε επαρκείς ποσότητες, καθαρό και με τα απαραίτητα χημικά συστατικά σε σωστές ποσότητες έτσι ώστε να αποφευχθούν δηλητηριάσεις και άλλες επιπλοκές στον ανθρώπινο οργανισμό.

Αυτή η επιτακτική ανάγκη καλεί τις πλέον σύγχρονες κοινωνίες, να σχεδιάσουν και να κατασκευάσουν τα κατάλληλα συστήματα παροχής νερού τα οποία θα παρέχουν πόσιμο νερό σε όλη την έκταση του πληθυσμού, σύμφωνα με την ζήτηση και τις απαιτήσεις του. Όμως πριν από την σχεδίαση και την κατασκευή τους, είναι απαραίτητη η κατανόηση και πρόβλεψη τυχόν εμποδίων έτσι ώστε να αποφευχθούν και να υπάρχει μια ομαλή λειτουργία των συστημάτων ύδρευσης. Πιο συγκεκριμένα, το δίκτυο πρέπει να μοντελοποιηθεί, να αναλυθεί και να αξιολογηθεί υπό διάφορες φυσικές και υδραυλικές παραμέτρους ή προϋποθέσεις. Η διαδικασία αυτή ονομάζεται «Προσομοίωση» και μπορεί να επιτευχθεί μέσω σύγχρονων εργαλείων που μας δίνουν αυτή την δυνατότητα. Ένα από τα πιο διαδεδομένα εργαλεία προσομοίωσης υδραυλικών δικτύων είναι το EPANET, το οποίο χρησιμοποιήθηκε για την προσομοίωση του δικτύου στην συγκεκριμένη διπλωματική εργασία.

Το EPANET είναι ένα πρόγραμμα το οποίο εκτελεί προσομοιώσεις δικτύων νερού υπό πίεση παρακολουθώντας και καταγράφοντας την ροή του νερού μέσα στις σωληνώσεις, καθώς και την πίεση στις σωληνώσεις, το ύψος του νερού σε κάθε δεξαμενή και την συγκέντρωση των χημικών συστατικών του νερού σε όλο το μήκος του δικτύου κατά τη διάρκεια μια χρονικής περιόδου, η οποία αποτελείται από πολλαπλά χρονικά στάδια.

Επιπροσθέτως, εκτός από την προσομοίωση των υδραυλικών δικτύων για την αξιολόγηση τους και την κατασκευή τους, θα μπορούσε να συμπεριληφθεί στο κομμάτι της πρόβλεψης και η μηχανική μάθηση με την βοήθεια κατάλληλων μοντέλων. Την συγκεκριμένη δυνατότητα μας την παρέχει το προγραμματιστικό περιβάλλον MATLAB, το οποίο με την βοήθεια του Regression Learner, βοηθάει να εκπαιδευτούν κατάλληλα μοντέλα για την πρόβλεψη και την κατανόηση συγκεκριμένων μεταβλητών για ένα υδραυλικό δίκτυο (κατανάλωση ενέργειας, ροή νερού στα σωληνώσεις, πίεση στους κόμβους του δικτύου. [1] [2] [3]

Στην παρούσα εργασία μοντελοποιήθηκε και προσομοιώθηκε ένα δίκτυο νερού με την βοήθεια του λογισμικού EPANET. Δημιουργήθηκε κατάλληλος κώδικας μέσω του MATLAB, που με την βοήθεια του δημιουργήθηκαν «N» τυχαία αρχεία, καθώς και 16

προκαθορισμένα σενάρια συνδεσμολογίας. Τα συγκεκριμένα αρχεία αντιπροσωπεύουν τους συνδυασμούς συνδεσμολογίας του δικτύου. Μετά την εκτέλεση των αρχείων μέσω DOS, διαμορφώθηκε κατάλληλος κώδικας για την συλλογή των αποτελεσμάτων του κάθε συνδυασμού αλλά και των 16 προκαθορισμένων. Πιο συγκεκριμένα, έγινε συλλογή των τιμών της καταναλισκόμενης ενέργειας, της πίεσης σε έναν κόμβο και της ροής σε μία σωλήνωση. Στη συνέχεια οι τιμές αυτές χρησιμοποιήθηκαν για την εκπαίδευση πολλαπλών μοντέλων μηχανικής μάθησης, τύπου Regression (παλινδρόμησης), μέσω της εφαρμογής του MATLAB, regression learner app. Το μοντέλο που εμφάνιζε την μικρότερη απόκλιση μέσης τετραγωνικής τιμής (RMSE) χρησιμοποιήθηκε για την πρόβλεψη των τιμών της ενέργειας, πίεσης και ροής για τα 16 προκαθορισμένα σενάρια θέσης βανών και στάθμη δεξαμενής.

ΚΕΦΑΛΑΙΟ 2: Περιγραφή Λογισμικού EPANET

2.1 Γενική περιγραφή

Το EPANET v.2.2 είναι λογισμικό σχεδίασης και προσομοίωσης δικτύων νερού που επιτρέπει στον χρήστη να εκτελεί προσομοιώσεις εκτεταμένης διάρκειας παρατηρώντας την συμπεριφορά του νερού αλλά και την ποιότητα του. Ένα δίκτυο αποτελείται από σωληνώσεις, κόμβους (διασταυρώσεις σωληνώσεων), αντλίες, βαλβίδες και δεξαμενές αποθήκευσης νερού. Το EPANET καταγράφει την ροή του νερού κατά μήκος κάθε σωλήνωσης, την πίεση σε κάθε κόμβο, το ύψος του νερού σε κάθε δεξαμενή, καθώς και την συγκέντρωση χημικών στοιχείων καθ' όλη την διάρκεια της προσομοίωσης.

Το συγκεκριμένο λογισμικό είναι σχεδιασμένο έτσι ώστε να βοηθάει τον χρήστη να κατανοήσει καλύτερα την κίνηση αλλά και τα συστατικά που περιέχονται στο νερό σε όλο το μήκος των συστημάτων διανομής του. Μπορεί να χρησιμοποιηθεί για διαφορετικά είδη εφαρμογών, όπως ενδεικτικά:

- Βαθμονόμηση υδραυλικού μοντέλου
- Ανάλυση υπολειμμάτων χλωρίου
- Αξιολόγηση της έκθεσης των καταναλωτών στα χημικά συστατικά

Το EPANET μπορεί να βοηθήσει στην εκτίμηση εναλλακτικών στρατηγικών για την βελτίωση της ποιότητας του νερού, όπως:

- Τροποποίηση χρήσης πηγών σε συστήματα με πολλαπλές πηγές
- Τροποποίηση προγραμμάτων άντλησης και πλήρωσης/εκκένωσης δεξαμενών
- Στοχευμένος καθαρισμός και αντικατάσταση των σωληνώσεων

Λειτουργώντας το EPANET στο περιβάλλον των Windows, παρέχει στον χρήστη ένα ενσωματωμένο περιβάλλον για επεξεργασία δεδομένων εισόδου του δικτύου, εκτέλεση υδραυλικών προσομοιώσεων και εμφάνιση αποτελεσμάτων σε διαφορετικές μορφές, όπως κωδικοποιημένοι χάρτες δικτύου, πίνακες δεδομένων και χρονικά γραφήματα.

2.2 Δυνατότητες υδραυλικής μοντελοποίησης

Η πλήρης και ακριβής μοντελοποίηση είναι απαραίτητη προϋπόθεση για την κατανόηση της συμπεριφοράς και ποιότητας του νερού. Το EPANET περιέχει μια υπερσύγχρονη εργαλειοθήκη που βοηθάει στην υδραυλική ανάλυση, η οποία μας παρέχει μερικές από τις ακόλουθες δυνατότητες:

- Δεν θέτει όρια στον χρήστη όσο αναφορά το μέγεθος του δικτύου που θα αναλυθεί (αριθμός σωληνώσεων, κόμβων, δεξαμενών).

- Περιλαμβάνει μικρές απώλειες οι οποίες οφείλονται στις στροφές, στα εξαρτήματα, στις βαλβίδες.
- Παρέχει μοντέλα αντλιών σταθερής ή μεταβλητής ταχύτητας
- Υπολογίζει με ακρίβεια την ενέργεια και το κόστος άντλησης του νερού.
- Μοντελοποιεί διάφορους τύπους βαλβίδων συμπεριλαμβανομένων βαλβίδες διακοπής, ελέγχου, ρυθμίσεις πίεσης και ελέγχου ροής.
- Επιτρέπει στον χρήστη να σχεδιάζει όπως εκείνος επιθυμεί τις δεξαμενές αποθήκευσης νερού όσον αφορά το μέγεθος και τη διάμετρο, καθώς και το ύψος όπου θα βρίσκεται η δεξαμενή.
- Μπορεί να βασίσει τη λειτουργία του συστήματος τόσο σε απλούς χειρισμούς της στάθμης της δεξαμενής ή με έναν χρονοδιακόπτη, όσο και σε σύνθετους χειρισμούς, οι οποίοι μπορούν να βασίζονται σε μία σειρά από κανόνες (λειτουργία δεξαμενής ή ολόκληρου του συστήματος συγκεκριμένες ώρες της μέρας ή λειτουργία υπό συνθήκες).

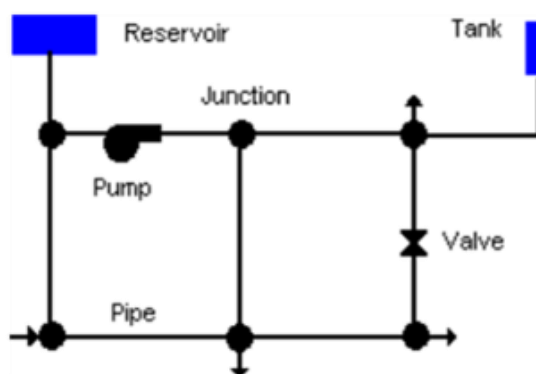
Βάση των παραπάνω δυνατοτήτων, το EPANET βοηθάει τον χρήστη να σχεδιάσει, να προσομοιώσει και να εκτιμήσει απλά ή περίπλοκα δίκτυα ή ήδη υπάρχοντα, ανάλογα με τις ανάγκες που έχει. Στη συνέχεια θα αναφερθούν ορισμένες δυνατότητες μοντελοποίησης για την ποιότητα του νερού. [4]

2.3 Περιγραφή μοντέλου δικτύου

Σε αυτήν την ενότητα θα περιγραφεί ένα μικρό τυπικό δίκτυο, καθώς και οι φυσικές παράμετροι (σωληνώσεις, κόμβοι, δεξαμενές κ.ο.κ.) και οι μη φυσικές παράμετροι (σωληνογραμμές, χαρακτηριστική αντλίας).

2.3.1 Φυσικές παράμετροι

Το EPANET περιγράφει ένα δίκτυο νερού ως ένα σύνολο αντικειμένων που συνδέονται μεταξύ τους όπως σωληνώσεις, αντλίες, βαλβίδες, δεξαμενές, διασταυρώσεις και πηγές. Στην Εικόνα 1, απεικονίζεται ένα απλό δίκτυο, δείχνοντας πως συνδέονται μεταξύ τους οι προαναφερθείσες παράμετροι.



Εικόνα 1 Απεικόνιση δικτύου και φυσικών παραμέτρων [5]

Διακλάδωση (Junctions): Οι διακλαδώσεις είναι τα σημεία του δικτύου μέσω των οποίων ενώνονται μεταξύ τους οι πηγές με τις δεξαμενές νερού και μέσω αυτών το νερό εισέρχεται και εξέρχεται από το δίκτυο.

Δεξαμενή-πηγή (Reservoir): Οι δεξαμενές-πηγές αντιπροσωπεύουν μια ατέρμονη εξωτερική πηγή νερού ή μία καταβόθρα. Χρησιμοποιούνται για να μοντελοποιήσουν λίμνες, ποτάμια, υπόγειους υδροφορείς.

Δεξαμενή Αποθήκευσης (Tanks): Οι δεξαμενές-αποθήκες χρησιμοποιούνται ως αποθηκευτικό μέσο του νερού, των οποίων το μέγεθος και ο όγκος μπορεί να ποικίλει κατά τη διάρκεια της προσομοίωσης.

Σωληνώσεις (Pipes): Οι σωληνώσεις είναι οι σύνδεσμοι του δικτύου που μεταφέρουν νερό από το ένα σημείο του δικτύου στο άλλο. Το EPANET υποθέτει ότι οι σωληνώσεις είναι γεμάτες ανά πάσα στιγμή.

Βαλβίδες (Valves): Οι βαλβίδες είναι τα μέσα που χρησιμοποιούνται σε ένα δίκτυο, τα οποία ελέγχουν την ροή και την πίεση σε συγκεκριμένα σημεία του δικτύου.

2.3.2 Μη φυσικές παράμετροι

Σε αντίθεση με τις φυσικές παραμέτρους, οι μη φυσικές παράμετροι χωρίζονται σε καμπύλες, μοτίβα και ελέγχους, οι οποίες περιγράφουν την συμπεριφορά και την λειτουργία του δικτύου.

Καμπύλες (Curves): Οι καμπύλες είναι παραστάσεις που περιέχουν ζευγάρια δεδομένων τα οποία δείχνουν την συσχέτιση μεταξύ δύο μεταβλητών. Μερικές αναφέρονται ονομαστικά παρακάτω:

- Χαρακτηριστική καμπύλη αντλίας
- Καμπύλη βαθμού απόδοσης
- Καμπύλη όγκου υγρού

Χρονικά μοτίβα (Time patterns): Τα χρονικά μοτίβα αποτελούν ένα σύνολο πολλαπλασιαστών, οι οποίοι εφαρμόζονται σε μία ποσότητα επιτρέποντας να ποικίλει χρονικά.

Έλεγχοι-Ρυθμίσεις (Controls): Οι ρυθμίσεις-έλεγχοι είναι συγκεκριμένες συνθήκες που ορίζονται από τον χρήστη και καθορίζουν την λειτουργία του συστήματος με την πάροδο του χρόνου, καθώς και την στάθμη του νερού στις δεξαμενές, την πίεση σε συγκεκριμένα σημεία του δικτύου. [5]

ΚΕΦΑΛΑΙΟ 3: Μοντέλα Μηχανικής Μάθησης

3.1: Ορισμός και κύριες τεχνικές

Κατά την διάρκεια της εξέλιξης τους, οι άνθρωποι χρησιμοποιούσαν ποικίλους τρόπους και εργαλεία για την ολοκλήρωση των καθημερινών τους αναγκών για να απλοποιήσουν την ζωή τους. Όσο εξελισσόταν ο ανθρώπινος εγκέφαλος τόσο πιο πολύ οδηγούσε τους ανθρώπους να κατασκευάσουν διάφορες μηχανές. Αυτές οι μηχανές απλοποίησαν ακόμα πιο πολύ την ζωή των ανθρώπων δίνοντας τους την ευκαιρία να ανταποκριθούν στις ανάγκες τους που όλο και πιο πολύ πλήθαιναν, μεταξύ των οποίων μέσα εναέριας μεταφοράς (αερόστατα, αεροπλάνα), εργοστάσια και το πιο σημαντικό από όλα τους υπολογιστές. Με την δημιουργία των υπολογιστών ανακαλύφθηκε και αναπτύχθηκε η μηχανική μάθηση.

Σύμφωνα με πρωτοπόρους της μηχανικής μάθησης, όπως τον Arthur Samuel, ο οποίος επινόησε τον όρο «Μηχανική Μάθηση», η μηχανική μάθηση ορίζεται ως ο τομέας των σπουδών που δίνει στους υπολογιστές να «εκπαιδευτούν». Μάλιστα ο Arthur Samuel ήταν γνωστός για το λογισμικό για το παιχνίδι της ντάμας, το οποίο ήταν από τα πρώτα προγράμματα υπολογιστή που μάθαιναν από μόνα τους.

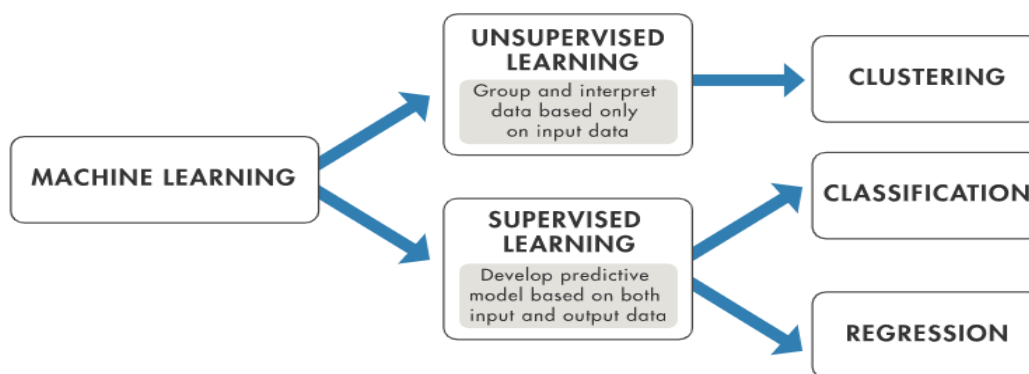
Η μηχανική μάθηση χρησιμοποιείται ως διδακτικό εργαλείο για τις μηχανές «μαθαίνοντας» τους να χειρίζονται δεδομένα περισσότερο αποτελεσματικά. Μερικές φορές δεν μπορούμε να ερμηνεύσουμε με ευκολία τα δεδομένα που χρησιμοποιούμε. Σε αυτή την περίπτωση, εφαρμόζουμε τεχνικές μηχανικής μάθησης. Με τον τόσο μεγάλο όγκο δεδομένων, η ζήτηση για την εφαρμογή της μηχανικής μάθησης αυξάνεται καθημερινά. Πολλές είναι οι βιομηχανίες και εταιρείες που εφαρμόζουν τεχνικές μηχανικής μάθησης για να κατανοήσουν και να αποκτήσουν μεγάλους όγκους δεδομένων, διότι ο σκοπός της μηχανικής μάθησης είναι η απόκτηση γνώσης από τα δεδομένα. Καθημερινά επιστήμονες σε όλους τους τομείς, όπως μαθηματικοί και προγραμματιστές εφαρμόζουν διάφορες προσεγγίσεις για να βρουν την λύση του προβλήματος, το οποίο είναι ο μεγάλος όγκος δεδομένων. Η μηχανική μάθηση έχει γίνει το κλειδί για την επίλυση διάφορων προβλημάτων σε διάφορους τομείς όπως:

- Χρηματοοικονομική μηχανική με την χρήση κατάλληλων αλγορίθμων για την λήψη εμπορικών αποφάσεων
- Επεξεργασία εικόνων για την αναγνώριση προσώπου, κίνησης και αντικειμένων
- Εντοπισμός όγκων και δημιουργία νέων φαρμάκων στην ιατρική

Για την ορθή χρήση των δεδομένων και επίλυση των προβλημάτων, η μηχανική μάθηση βασίζεται σε διάφορους αλγόριθμους. Οι επιστήμονες τονίζουν πως δεν υπάρχει κάποιος αλγόριθμος που να ταιριάζει και να χρησιμοποιείται ως πανάκεια

σε όλους τους τύπους προβλημάτων. Το είδος του αλγόριθμου που θα χρησιμοποιηθεί ποικίλει με το είδος του προβλήματος που πρέπει να επιλυθεί, τον αριθμό των μεταβλητών που το απαρτίζουν καθώς και το είδος του μοντέλου που ταιριάζει καλύτερα.

Στην μηχανική μάθηση χρησιμοποιούνται δύο κύριες τεχνικές: η εποπτευόμενη μάθηση, η οποία εκπαιδεύει ένα μοντέλο σε ήδη υπάρχοντα και γνωστά δεδομένα εισόδου και εξόδου ώστε να μπορούν να προβλεφθούν αποτελέσματα που μπορούν χρησιμοποιηθούν για πρόβλεψη. Η μη εποπτευόμενη μάθηση, η οποία εντοπίζει κρυφά μοτίβα ή εσωτερικές δομές στα δεδομένα που δίνονται ως αρχείο εισόδου, όπως φαίνεται στην Εικόνα 2. [6] [7]



Εικόνα 2 Τεχνικές μηχανικής μάθησης [6]

3.2 Εποπτευόμενη μάθηση και Μη εποπτευόμενη μάθηση

Η εποπτευόμενη μάθηση (Supervised Learning) είναι ο πιο γνωστός και κοινός τύπος αλγορίθμων μηχανικής μάθησης. Χρησιμοποιεί ένα σύνολο δεδομένων εισόδου-εξόδου και ονομάζεται σύνολο δεδομένων εκπαίδευσης, το οποίο είναι γνωστό από τον χρήστη για να εκπαιδεύσει έναν αλγόριθμο με γνωστά δεδομένα εισόδου και εξόδου για να γίνουν οι απαραίτητες προβλέψεις. Το σύνολο δεδομένων εκπαίδευσης περιλαμβάνει δεδομένα εισόδου τα οποία συνδυάζονται με τα δεδομένα εξόδου. Ο σκοπός του αλγορίθμου είναι να δημιουργήσει ένα μοντέλο που θα κάνει προβλέψεις των τιμών απόκρισης για ένα νέο σύνολο δεδομένων, αποκαλύπτοντας τις σχέσεις μεταξύ δεδομένων εισόδου-εξόδου.

Έπειτα, μετά την εκπαίδευση του μοντέλου, ένα σύνολο δεδομένων το οποίο δεν έχει χρησιμοποιηθεί για εκπαίδευση, χρησιμοποιείται για να την πρόβλεψη της απόδοσης του μοντέλου που εκπαιδεύτηκε. Για την ακριβή λήψη των αποτελεσμάτων απόδοσης του μοντέλου, πρέπει το σύνολο δεδομένων που χρησιμοποιήθηκε για εκπαίδευση αλλά και το σύνολο που χρησιμοποιήθηκε για την πρόβλεψη της απόδοσης του μοντέλου, να είναι έγκυρα και να αναφέρονται σε «πραγματικά» στοιχεία.

Η εποπτευομένη μάθηση χρησιμοποιεί τεχνικές ταξινόμησης (classification) και παλινδρόμησης (regression), οι οποίες θα αναλυθούν παρακάτω. [6] [8]

Κατά την επίλυση προβλημάτων, υπάρχουν περιπτώσεις όπου έχουμε δεδομένα που δεν έχουν κατηγοριοποιηθεί ή δεν τους έχει δοθεί κάποιο όνομα ή «ετικέτα» και πρέπει να βρεθούν σχέσεις μεταξύ των δεδομένων που έχουν εισαχθεί, για την επίλυση αυτού του προβλήματος στην μηχανική μάθηση χρησιμοποιείται η τεχνική της μη εποπτευόμενης μάθησης.

Η μη εποπτευόμενη μάθηση (Unsupervised Learning) είναι μία μέθοδος μηχανικής μάθησης κατά την οποία τα μοντέλα όπου χρησιμοποιούνται δεν ελέγχονται με την δεδομένα εκπαίδευσης. Πιο συγκεκριμένα, τα μοντέλα αντί να εκπαιδεύονται από τα δεδομένα εκπαίδευσης όπως είναι τα μοντέλα της εποπτευόμενης μάθησης που τα δεδομένα εκπαίδευσης εκπαιδεύουν τα μοντέλα να προβλέπουν σωστά το αποτέλεσμα, βρίσκουν εκείνα κρυφά μοτίβα και σχέσεις μεταξύ των δεδομένων. Αυτό μπορεί να παρομοιαστεί με τον ανθρώπινο εγκέφαλο που δέχεται πρώτη φορά ερεθίσματα από νέες πληροφορίες.

Η τεχνική της μη εποπτευόμενης μάθησης δεν μπορεί να χρησιμοποιηθεί για την επίλυση προβλημάτων ταξινόμησης (classification) ή παλινδρόμησης (regression), διότι αν και είναι γνωστά τα δεδομένα εισόδου, δεν είναι γνωστά τα αντίστοιχα δεδομένα εξόδου. Σκοπός της συγκεκριμένης μεθόδου είναι μέσω των αλγορίθμων της, να ανακαλύψει την δομή των δεδομένων εισόδου, να τα ομαδοποιήσει σύμφωνα με τα κοινά τους χαρακτηριστικά και να τα παρουσιάσει στην τελική τους μορφή.

Οι λόγοι για τους οποίους χρησιμοποιείται η μη εποπτευόμενη μάθηση είναι οι παρακάτω:

- Η ανακάλυψη και εύρεση σχέσεων και μοτίβων από τα δεδομένα εισόδου.
- Η συμπεριφορά των μοντέλων της μη εποπτευόμενης μάθησης μιμείται την συμπεριφορά του ανθρώπινου εγκέφαλου.
- Όταν δεν υπάρχει κατηγοριοποίηση των δεδομένων ή δεδομένα χωρίς «ετικέτα», η συγκεκριμένη μέθοδος θεωρείται ζωτική για την επίλυση τέτοιων προβλημάτων.
- Στα καθημερινά προβλήματα που καλείται να επιλύσει ο άνθρωπος, δεν υπάρχουν πάντα δεδομένα με γνωστές εισόδους και εξόδους, άρα η χρήση της μη εποπτευόμενης μάθησης είναι απαραίτητη.

Η μέθοδος της μηχανικής μάθησης μπορεί να κατηγοριοποιηθεί περαιτέρω σε δύο ειδών προβλήματα: τα προβλήματα ομαδοποίησης (clustering) και τα προβλήματα συσχέτισης (association). Παρακάτω αναφέρονται ονομαστικά ορισμένοι από τους αλγόριθμους μη εποπτευόμενης μάθησης, όπου και θα αναλυθούν περαιτέρω:

- Ομαδοποίηση K-means (K-means clustering)

- Ιεραρχική Ομαδοποίηση (Hierarchical Clustering)
- Χωρική ομαδοποίηση με βάση την πυκνότητα (Density based spatial clustering, DBSCAN) [9] [10]

Στη παρούσα εργασία χρησιμοποιήθηκε η μέθοδος της εποπτευόμενης μάθησης και οι αλγόριθμοι για την επίλυση προβλημάτων παλινδρόμησης (Regression).

3.3 Τεχνικές ταξινόμησης (Classification)

Οι τεχνικές της ταξινόμησης χρησιμοποιούνται για τιμές απόκρισης που μπορούν να κατηγοριοποιηθούν, όπου τα δεδομένα διαχωρίζονται σε συγκεκριμένες κλάσεις. Οι συγκεκριμένες τεχνικές μπορούν να προβλέψουν διακριτές απαντήσεις, όπως εάν ένα εισερχόμενο μήνυμα ηλεκτρονικού ταχυδρομείου είναι ανεπιθύμητο ή αυθεντικό, στην ιατρική και συγκεκριμένα εάν ένας όγκος είναι καλοήθης ή κακοήθης. Για τον λόγο που αναφέρθηκε, οι συγκεκριμένες τεχνικές για να είναι αποτελεσματικές πρέπει να χρησιμοποιούνται για δεδομένα τα οποία μπορούν να κατηγοριοποιηθούν ή να διαχωριστούν σε συγκεκριμένες ομάδες ή κλάσεις.

Ορισμένοι αλγόριθμοι της συγκεκριμένης κατηγορίας είναι οι παρακάτω, όπου αναφέρονται ονομαστικά και θα αναλυθούν στην συνέχεια:

- Μηχανές Διανυσμάτων Υποστήριξης ΜΔΥ (Support Vector Machines, SVM)
- Διακριτική Ανάλυση (Discriminant Analysis)
- Naïve Bayes
- Πλησιέστερος Γείτονας (Nearest Neighbor) [6] [8]

3.4 Τεχνικές παλινδρόμησης (Regression)

Οι τεχνικές παλινδρόμησης (Regression) είναι στατιστικές τεχνικές, οι οποίες βοηθάνε στην μοντελοποίηση της σχέσης μεταξύ εξαρτημένων (στόχος) και ανεξάρτητων (πρόβλεψη-προγνωστικού) μεταβλητών με μία ή παραπάνω ανεξάρτητες μεταβλητές. Πιο συγκεκριμένα η τεχνική της παλινδρόμησης βοηθάει στην κατανόηση για το πως αλλάζει η τιμή μιας εξαρτημένης μεταβλητής, η οποία αντιστοιχεί σε μία ανεξάρτητη όταν οι υπόλοιπες ανεξάρτητες μεταβλητές παραμένουν σταθερές. Η συγκεκριμένη τεχνική προβλέπει συνεχείς-πραγματικές τιμές όπως είναι η θερμοκρασία, ηλικία, μισθός, ενέργεια, πίεση κ.λπ.. Επειδή χρειάζεται να προβλεφθούν διάφορες τιμές στον πραγματικό κόσμο, όπως είναι οι μισθοί εργαζομένων ή προβλέψεις πωλήσεων, οι λόγοι που εφαρμόζονται οι συγκεκριμένες τεχνικές είναι:

- Εκτίμηση σχέσης μεταξύ στόχου και ανεξάρτητης μεταβλητής.
- Εκτίμηση και εντοπισμός των τάσεων στα δεδομένα.
- Πρόβλεψη πραγματικών τιμών.

- Δυνατότητα προσδιορισμού σημαντικών παραγόντων στα δεδομένα.

Παρακάτω αναφέρονται ονομαστικά ορισμένοι αλγόριθμοι της συγκεκριμένης κατηγορίας, όπου και θα αναλυθούν:

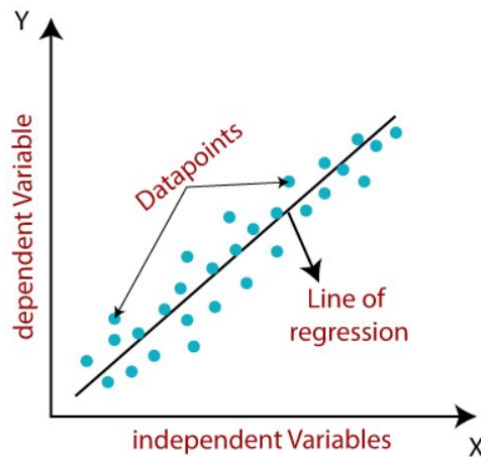
- Γραμμική Παλινδρόμηση (Linear Regression)
- Μοντέλα παλινδρόμησης Gaussian Process (Gaussian Process Regression Models)
- Μοντέλα Μεθόδου Συνόλων (Ensemble Methods)
- Δέντρα Αποφάσεων (Decision Trees)
- Νευρωνικά Δίκτυα (Neural Networks) [6] [11]

3.4.1 Ανάλυση αλγόριθμου: Γραμμική Παλινδρόμηση (Linear Regression)

Η μηχανική μάθηση χρησιμοποιείται σε διάφορους τομείς για την επίλυση προβλημάτων που δεν μπορούν να επιλυθούν εύκολα με απλές προσεγγίσεις του υπολογιστή. Η γραμμική παλινδρόμηση (Linear Regression) αποτελεί ένα από τα πιο απλά και συνηθέστερα μοντέλα μηχανικής μάθησης. Χρησιμοποιείται ως μαθηματική προσέγγιση για προγνωστική ανάλυση. Η γραμμική παλινδρόμηση επιτρέπει την πρόβλεψη συνεχών/πραγματικών μεταβλητών, όπως αριθμός πωλήσεων, τιμές προϊόντων, ηλικία.

Η έννοια της γραμμικής παλινδρόμησης προτάθηκε πρώτη φορά από τον Sir Francis Galton το 1894. Το συγκεκριμένο μοντέλο χρησιμοποιείται για την αξιολόγηση και τον ποσοτικό προσδιορισμό των σχέσεων μεταξύ των μεταβλητών που πρόκειται να εξετασθούν. Η γραμμική παλινδρόμηση χρησιμοποιείται συχνά σε μαθηματικές ερευνητικές μεθόδους, όπου είναι εφικτή η μέτρηση του αποτελέσματος που προβλέφθηκε και η μοντελοποίηση του σε σχέση με πολλαπλές μεταβλητές εισόδου. Αποτελεί μία μέθοδο αξιολόγησης και μοντελοποίησης δεδομένων που δημιουργεί γραμμικές σχέσεις μεταξύ των εξαρτημένων και ανεξάρτητων μεταβλητών.

Το μοντέλο της γραμμικής παλινδρόμησης περιέχει μία γραμμή, η οποία αντιπροσωπεύει την σχέση μεταξύ των εξαρτημένων και ανεξάρτητων μεταβλητών, σύμφωνα με την Εικόνα 3.



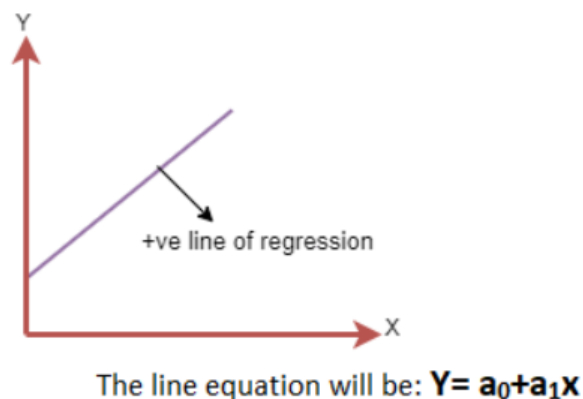
Εικόνα 3 Έννοιες αλγόριθμου γραμμικής παλινδρόμησης [12]

Το μοντέλο της γραμμικής παλινδρόμησης μπορεί να χωριστεί περαιτέρω σε δύο κατηγορίες αλγορίθμων:

- **Απλή Γραμμική Παλινδρόμηση (Simple Linear Regression)**
Χρησιμοποιείται όταν για την πρόβλεψη της τιμής μίας εξαρτημένης μεταβλητής χρησιμοποιείται μια ανεξάρτητη μεταβλητή.
- **Πολλαπλή Γραμμική Παλινδρόμηση (Multiple Linear Regression)**
Χρησιμοποιείται όταν για την πρόβλεψη της τιμής μίας εξαρτημένης μεταβλητής χρησιμοποιούνται πάνω από μία ανεξάρτητες μεταβλητές.

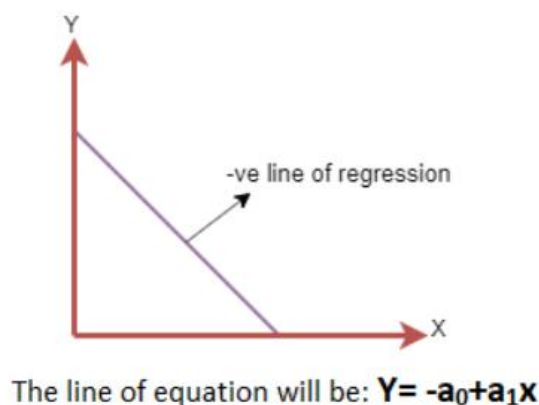
Στο παραπάνω διάγραμμα απεικονιζόταν η σχέση που έχουν οι εξαρτημένες και ανεξάρτητες μεταβλητές με την μορφή γραμμής, η οποία ονομάζεται γραμμή παλινδρόμησης και δείχνει δύο τύπους σχέσεων μεταξύ των μεταβλητών.

Ο πρώτος τύπος σχέσης που μπορεί να εμφανιστεί είναι η θετική γραμμική σχέση, η οποία δημιουργείται όταν η εξαρτημένη μεταβλητή αυξάνεται στον άξονα Y και η ανεξάρτητη μεταβλητή στον άξονα X, Εικόνα 4.



Εικόνα 4 Θετική γραμμική σχέση και η εξίσωση της [12]

Ο δεύτερος τύπος σχέσης είναι η αρνητική γραμμική σχέση που δημιουργείται όταν η εξαρτημένη μεταβλητή μειώνεται στον άξονα Y και η ανεξάρτητη μεταβλητή αυξάνεται στον άξονα X, Εικόνα 5.



Εικόνα 5 Αρνητική γραμμική σχέση και η εξίσωση της [12]

Σκοπός της γραμμικής παλινδρόμησης είναι η εύρεση της καλύτερης γραμμής προσαρμογής που σημαίνει ότι το σφάλμα μεταξύ των προβλεπόμενων τιμών και των πραγματικών τιμών πρέπει να είναι το ελάχιστο. Όσο μικρότερο το σφάλμα, τόσο καλύτερη η γραμμή προσαρμογής. Για την γραμμική παλινδρόμηση χρησιμοποιείται η συνάρτηση κόστους του μέσου τετραγωνικού σφάλματος (Mean Squared Error, MSE), που είναι ο μέσος όρος του τετραγωνικού σφάλματος μεταξύ των πραγματικών και των προβλεπόμενων τιμών. [12] [13]

3.4.2 Ανάλυση αλγόριθμου: Μοντέλα παλινδρόμησης Gaussian Process (Gaussian Process Regression Models, GPR)

Ο αλγόριθμος παλινδρόμησης διαδικασίας Gauss (Gaussian Process Regression, GRP) είναι μία μη παραμετρική, η οποία χρησιμοποιείται συχνά στην μηχανική μάθηση για προβλήματα παλινδρόμησης. Μπορεί να συγκεντρώσει μια μεγάλη ποικιλία σχέσεων μεταξύ των δεδομένων εισόδου και εξόδου, χρησιμοποιώντας έναν άπειρο αριθμό παραμέτρων, αυτό έχει ως αποτέλεσμα τα δεδομένα να προσδιορίζουν το επίπεδο πολυπλοκότητας.

Η συγκεκριμένη μέθοδος έχει αρκετά πλεονεκτήματα, καθώς λειτουργεί καλά σε μικρά σύνολα δεδομένων και παρέχει την δυνατότητα μέτρησης της αβεβαιότητας στις προβλέψεις. Σε σχέση με πολλούς γνωστούς αλγόριθμους εποπτευόμενης μηχανικής μάθησης, οι οποίοι γνωρίζουν ακριβώς τις τιμές για κάθε παράμετρο, αυτή η μη παραμετρική προσέγγιση συγκεντρώνει μια κατανομή πιθανοτήτων για όλες τις πιθανές τιμές. Πέρα από την χρήση του στην εποπτευόμενη μάθηση μπορεί να

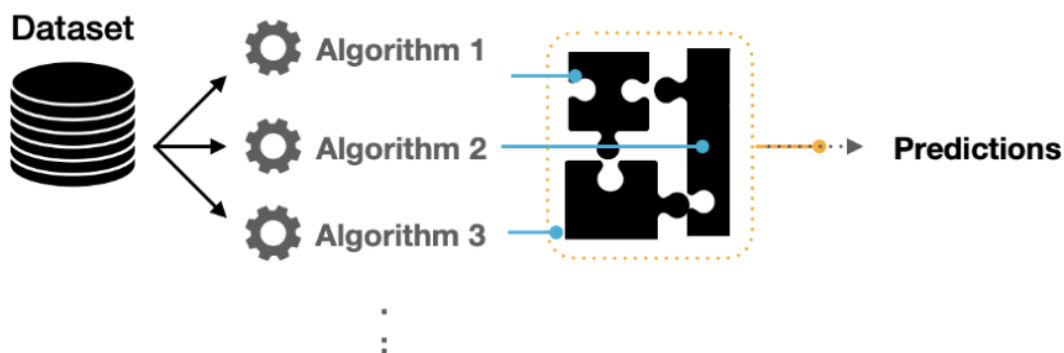
χρησιμοποιηθεί και στην μη εποπτευόμενη μάθηση αλλά και σε άλλες ποικίλες εφαρμογές.

Τέλος, οι μέθοδοι Gauss είναι μία γενίκευση της κατανομής πιθανοτήτων Gauss και χρησιμοποιούνται ως η βασική αρχή για πιο εξελιγμένους, μη παραμετρικούς αλγορίθμους μηχανικής μάθησης για ταξινόμηση και παλινδρόμηση. [14] [15]

3.4.3 Ανάλυση αλγόριθμου: Μέθοδοι Συνόλων (Ensemble Methods Models)

Τα μοντέλα συνόλων (Ensemble methods models) στην μηχανική μάθηση χρησιμοποιούνται προσεγγιστικά, συνδυάζοντας πολλαπλά μοντέλα κατά τη διαδικασία της πρόβλεψης. Τα μοντέλα που χρησιμοποιούνται ονομάζονται εκτιμητές. Το συγκεκριμένο μοντέλο χρησιμοποιείται για να αντιμετωπιστεί μία σειρά προβλημάτων, τα οποία δημιουργούνται κατά την χρήση ενός μόνο εκτιμητή, όπως είναι η υψηλή διακύμανση, το μοντέλο είναι ευαίσθητο στις τιμές εισόδου στα εκπαιδευμένα χαρακτηριστικά, η χαμηλή ακρίβεια που μπορεί να προκύψει κατά την χρήση μόνο ενός μοντέλου, διότι ένας αλγόριθμος μπορεί να μην είναι ικανός να διαχειριστεί μεγάλο όγκο δεδομένων και να ανταποκριθεί στις προσδοκίες και η διάθεση προκατάληψης καθώς το μοντέλο βασίζεται σε ένα βαθμό σε ένα ή περισσότερα χαρακτηριστικά κατά την διαδικασία της πρόβλεψης.

Χρησιμοποιώντας ένα σύνολο δεδομένων και έναν μεμονωμένο αλγόριθμο, το μοντέλο μπορεί να μην καταλήξει στο επιθυμητό αποτέλεσμα. Κάθε αλγόριθμος μηχανικής μάθησης έχει τους περιορισμούς του και αυτό αποτελεί πρόκληση για τον χρήστη, διότι η δημιουργία ενός μοντέλου υψηλής ακρίβειας δεν επιτυγχάνεται με ευκολία. Για να επιτευχθεί μεγαλύτερη ακρίβεια υπάρχει η δυνατότητα συνδυασμού πολλαπλών μοντέλων. Ο συνδυασμός αυτός μπορεί να ολοκληρωθεί, συνδυάζοντας τα αποτελέσματα από το κάθε μοντέλο που χρησιμοποιήθηκε, μειώνοντας το σφάλμα του μοντέλου και διατηρώντας την γενίκευση του. Η συγκεκριμένη ιδέα μπορεί να υλοποιηθεί χρησιμοποιώντας ορισμένες μεθόδους. Η συγκεκριμένη αρχιτεκτονική αναφέρεται και ως meta-algorithm (Εικόνα 6).

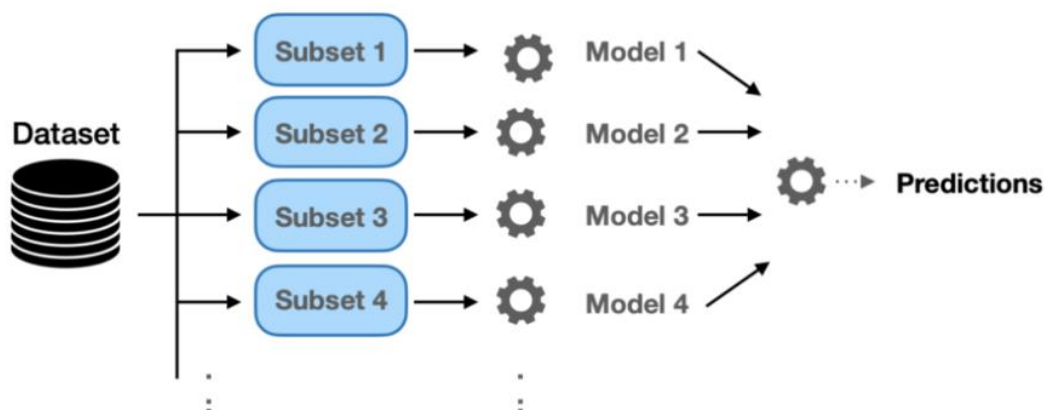


Εικόνα 6 Διαχωρισμός προβλέψεων κάθε μοντέλου με χρήση πολλαπλών αλγορίθμων [16]

Η κατασκευή των μοντέλων συνόλων (Ensemble Methods models) δεν στηρίζεται μόνο στη διακύμανση του αλγορίθμου που χρησιμοποιείται. Θα μπορούσαν να δημιουργηθούν διάφορα μοντέλα, όπου το κάθε ένα ξεχωριστά ειδικεύεται στην πρόβλεψη μιας μεταβλητής. Τα συγκεκριμένα μοντέλα ονομάζονται «αδύναμοι μαθητές» και χρησιμοποιούνται για την δημιουργία του meta-algorithm που αναφέρθηκε παραπάνω. Πιο συγκεκριμένα, τα δεδομένα εισόδου χωρίζονται σε κάθε «αδύναμο» μοντέλο και ταυτόχρονα καταγράφονται οι προβλέψεις τους για κάθε μεταβλητή εισόδου. Ο συνδυασμός των προβλέψεων κάθε μοντέλου δημιουργεί το τελικό μοντέλο συνόλου.

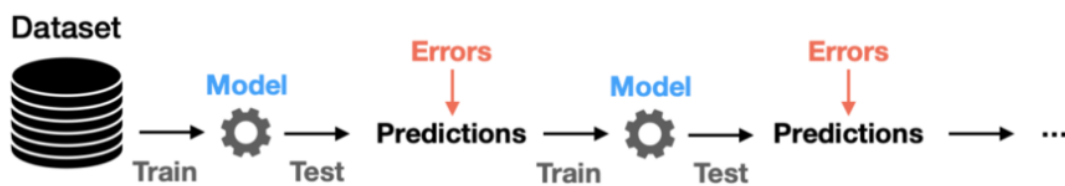
Οι τεχνικές που χρησιμοποιούνται στον συγκεκριμένο αλγόριθμο είναι οι παρακάτω:

Τεχνική bagging: Η συγκεκριμένη τεχνική βασίζεται στα δεδομένα που είναι διαθέσιμα προς εκπαίδευση κατά τη διαδικασία της μάθησης. Κάθε μοντέλο μαθαίνει από τα σφάλματα του προηγούμενου μοντέλου χρησιμοποιώντας ένα ελάχιστο διαφοροποιημένο υποσύνολο των δεδομένων που χρησιμοποιήθηκαν. Με την συγκεκριμένη τεχνική επιτυγχάνεται η μείωση της διακύμανσης και ελαχιστοποίηση της υπερπροσαρμογής (Εικόνα 7).



Εικόνα 7 Τεχνική bagging για την δημιουργία πρόβλεψης συνδυάζοντας προβλέψεις πολλαπλών μοντέλων [16]

Τεχνική boosting: Είναι μια μέθοδος μάθησης συνόλου που συνδυάζει ένα σύνολο «αδύναμων μαθητών», για την δημιουργία ενός πιο ισχυρού μοντέλου για να ελαχιστοποιήσει τα σφάλματα της εκπαίδευσης. Στην συγκεκριμένη τεχνική επιλέγεται ένα τυχαίο δείγμα δεδομένων, προσαρμόζεται σε ένα αρχικό μοντέλο και στην συνέχεια επανεκπαιδεύεται διαδοχικά. Κάθε μοντέλο προσπαθεί να εξαλείψει τα σφάλματα του προηγούμενου. Σκοπός της μεθόδου είναι να συνδυάζονται τα «αδύναμα» μοντέλα για την δημιουργία ενός νέου πιο ισχυρού μοντέλου πρόβλεψης. Ονομαστικά οι τύποι του bagging είναι ο Adaptive Boosting (AdaBoost), ο Gradient Boosting και ο Extreme Gradient Boosting (XGBoost). Στην Εικόνα 8 απεικονίζεται πως λειτουργεί η τεχνική του Boosting.



Εικόνα 8 Εκμάθηση που γίνεται λαμβάνοντας υπόψιν τα σφάλματα προηγούμενων μοντέλων [16]

Τεχνική του Stacking: Η μέθοδος της στοίβαξης (Stacking) είναι παρόμοια με τις προηγούμενες μεθόδους και μπορούν να δημιουργήσουν πιο ισχυρά προγνωστικά μοντέλα. Η συγκεκριμένη μέθοδος χρησιμοποιεί τις προβλέψεις όλων των μοντέλων για να δημιουργήσει μια τελική πιο δυνατή πρόβλεψη. Η διαφορά με τις τεχνικές του boosting και bagging είναι ότι αυτή η τεχνική συνδυάζει πολλά διαφορετικά μοντέλα. Στην Εικόνα 9 απεικονίζεται η λειτουργία της μεθόδου.



Εικόνα 9 Τεχνική στοίβαξης για την πρόβλεψη τελικού αποτελέσματος [16]

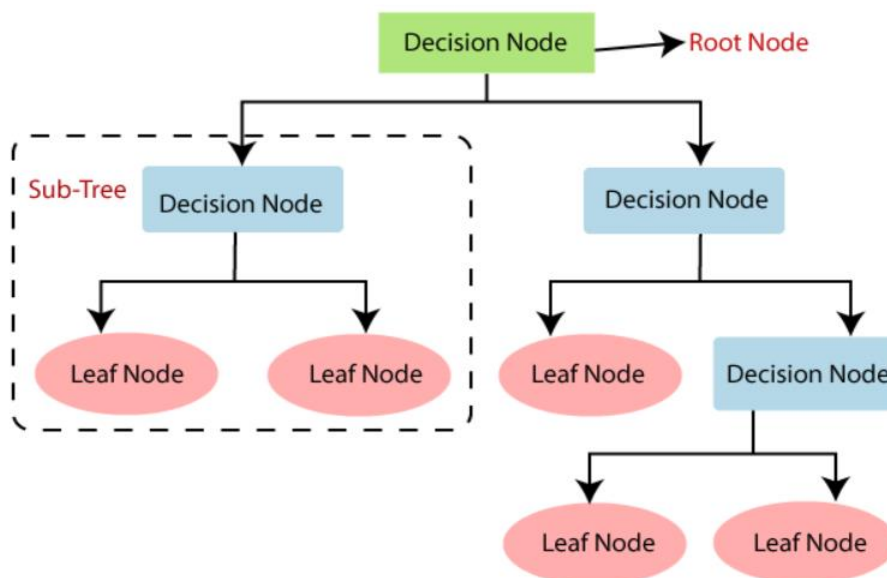
Τεχνική ανάμειξης (Blending): Η τεχνική αυτή είναι παρόμοια με εκείνη της στοίβαξης (stacking) και χρησιμοποιεί ένα μοντέλο μηχανικής μάθησης για να εκπαιδευτεί, έτσι ώστε να χρησιμοποιεί και να συνδυάζει καλύτερα τις προβλέψεις από πολλαπλά μοντέλα. [16] [17]

3.4.4 Ανάλυση αλγόριθμου: Δέντρα αποφάσεων (Decision Trees)

Το μοντέλο των δέντρων αποφάσεων (Decision Trees) είναι μία τεχνική της εποπτευόμενης μάθησης που χρησιμοποιείται για προβλήματα ταξινόμησης και παλινδρόμησης. Είναι μία μέθοδος με δομή δέντρου, όπου οι εσωτερικοί κόμβοι αντικατοπτρίζουν τα χαρακτηριστικά του συνόλου δεδομένων εισόδου, οι κλάδοι τους κανόνες απόφασης και κάθε κόμβος φύλλου αντιπροσωπεύει το αποτέλεσμα. Στα δέντρα αποφάσεων υπάρχουν δύο είδη κόμβων, οι οποίοι είναι οι κόμβοι αποφάσεων (Decision Nodes) και οι κόμβοι φύλλων (Leaf Nodes). Οι αποφάσεις λαμβάνονται από τους κόμβους αποφάσεων, οι οποίοι έχουν πολλαπλούς κόμβους

και τα αποτελέσματα φαίνονται από τους κόμβους φύλλων και δεν περιέχουν περαιτέρω κλάδους. Οι αποφάσεις λαμβάνονται σύμφωνα με τα χαρακτηριστικά των δεδομένων εισόδου και όλες οι πιθανές λύσεις λαμβάνονται σύμφωνα με τις συνθήκες που έχουν δοθεί στο δέντρο.

Ονομάζονται δέντρα αποφάσεων διότι η δομή τους μοιάζει με εκείνη ενός δέντρου, ξεκινώντας από τον αρχικό κόμβο ή αλλιώς τον κόμβο ρίζας και επεκτείνεται σε παρακλάδια. Η Εικόνα 10 αντικατοπτρίζει την δομή ενός δέντρου απόφασης.



Εικόνα 10 Δομή δέντρου απόφασης [18]

Οι όροι που χρησιμοποιούνται στο δέντρο απόφασης είναι:

Κόμβος Ρίζας (Root Node): Είναι ο αρχικός κόμβος όπου ξεκινάει το δέντρο αποφάσεων και αποτελεί το σύνολο των δεδομένων, το οποίο χωρίζεται σταδιακά.

Κόμβος Φύλλου (Leaf Node): Είναι οι κόμβοι όπου εμφανίζεται το τελικό αποτέλεσμα και δεν μπορεί να διαχωριστεί παραπάνω.

Διαίρεση (Splitting): Η διαδικασία κατά την οποία διαιρείται ένας κόμβος απόφασης σε υποκόμβους ανάλογα με τις ορισθείσες συνθήκες.

Κλάδος/Υπόδεντρο (Branch/Sub Tree): Το δέντρο που δημιουργείται από την διαίρεση.

Κλάδεμα (Pruning): Η αφαίρεση των ανεπιθύμητων κόμβων.

Κόμβος Γονέας/Παιδί: Ο κόμβος της ρίζας από τον οποίο ξεκινάει το δέντρο ονομάζεται γονικός κόμβος και οι κόμβοι που δημιουργούνται ονομάζονται θυγατρικοί κόμβοι.

Ο συγκεκριμένος αλγόριθμος ξεκινάει με τον αρχικό κόμβο, τον κόμβο ρίζας. Αρχίζει και συγκρίνει τις τιμές των χαρακτηριστικών που έχουν εισαχθεί από τον κόμβο ρίζας με το πραγματικό σύνολο δεδομένων και βάση τις συγκρίσεις που γίνονται το δέντρο διαιρείται σε κάθε κόμβο. Συνεχίζοντας, συγκρίνονται ξανά τα αποτελέσματα που έχουν προκύψει από τον προηγούμενο κόμβο και συνεχίζει παρακάτω. Η διαδικασία συνεχίζεται μέχρι να φτάσει στον κόμβο φύλλου, όπου είναι ο τελικός κόμβος με το τελικό αποτέλεσμα.

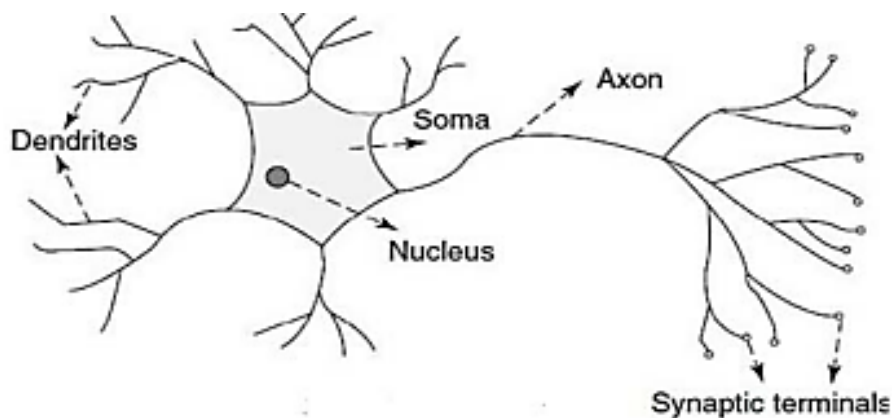
Τέλος τα πλεονεκτήματα και μειονεκτήματα του συγκεκριμένου αλγόριθμου είναι τα ακόλουθα:

- + Μπορεί να γίνει εύκολα κατανοητό καθώς μιμείται την διαδικασία που ακολουθεί ένας άνθρωπος κατά την λήψη των αποφάσεων του.
- + Είναι αρκετά εύχρηστος αλγόριθμος για θέματα αποφάσεων.
- + Βοηθάει στον καθορισμό πολλαπλών πιθανών αποτελεσμάτων.
- + Χρειάζεται λιγότερη προεπεξεργασία δεδομένων.
- Ο συγκεκριμένος αλγόριθμος περιέχει πολλαπλά επίπεδα, ως αποτέλεσμα να γίνεται περίπλοκος.
- Υπάρχει κίνδυνος υπεπροσαρμογής. [18]

3.4.5 Ανάλυση αλγόριθμου: Νευρωνικά Δίκτυα (Neural Networks)

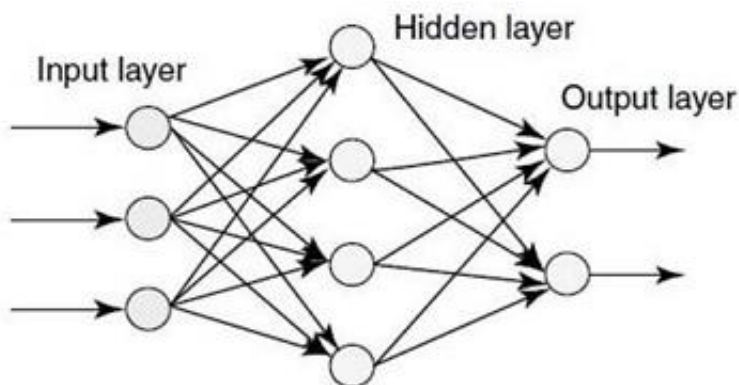
Τα νευρωνικά δίκτυα (Neural Networks) είναι ένας αλγόριθμος μηχανικής μάθησης, τα οποία είναι δομημένα έτσι ώστε να μιμούνται την λειτουργία του ανθρώπινου εγκεφάλου. Είναι μία σειρά αλγορίθμων που αναγνωρίζει τις σχέσεις μεταξύ των δεδομένων ενός συνόλου δεδομένων. Κατά αυτή την έννοια τα νευρωνικά δίκτυα αναφέρονται σε ένα σύστημα νευρώνων οργανικής ή τεχνητής φύσεως. Ο συγκεκριμένος αλγόριθμος μπορούν και προσαρμόζονται στο σύνολο των εισροών και στην μεταβολή τους, παράγοντας το καλύτερο δυνατό αποτέλεσμα χωρίς να είναι απαραίτητος ο επανασχεδιασμός των κριτηρίων που οδηγούν στο αποτέλεσμα. Η προσέγγιση των προβλημάτων γίνεται με την χρήση πολλαπλών επιπέδων, έτσι ο αλγόριθμος μαθαίνει και εξάγει συμπεράσματα με βάση τα παρατηρούμενα δεδομένα.

Τα νευρωνικά δίκτυα ή τεχνητά νευρωνικά δίκτυα (Artificial Neural Networks, ANN) έχουν πάρει την ονομασία τους από την δομή των νευρώνων. Νευρώνας ονομάζεται μία κυτταρική δομή που βρίσκεται στον εγκέφαλο κάθε οργανισμού. Στην Εικόνα 11 απεικονίζεται η δομή ενός νευρώνα.



Εικόνα 11 Δομή νευρώνα [19]

Συγκεκριμένα ένας νευρώνας αποτελείται από τους δενδρίτες (dendrites), το κύριο σώμα του νευρώνα (soma), τον πυρήνα (nucleus), τον άξονα (axon) που συνδέει το σώμα με τις νευρικές απολήξεις (synaptic terminals). Οι δενδρίτες είναι το σημείο εισόδου όπου ο νευρώνας δέχεται ηλεκτρικά σήματα. Κατά την διαδικασία εξόδου, το ηλεκτρικό σήμα μεταφέρεται μέσα στον νευρώνα με την βοήθεια του άξονα και καταλήγει στις νευρικές απολήξεις, όπου και συνεχίζεται η διαδικασία στους επόμενους νευρώνες. Ο πυρήνας αποτελεί την καρδιά του νευρώνα και αποτελεί το κέντρο ελέγχου του. Η συνδεσμολογία των νευρώνων αποτελούν το νευρωνικό δίκτυο μέσω του οποίου ταξιδεύουν τα ηλεκτρικά σήματα που δέχεται ο νευρώνας από τους δενδρίτες και ταξιδεύουν σε όλη την έκταση του εγκεφάλου. Στην Εικόνα 12 απεικονίζεται η δομή ενός νευρωνικού δικτύου που συμπεριφέρεται με παρόμοιο τρόπο.



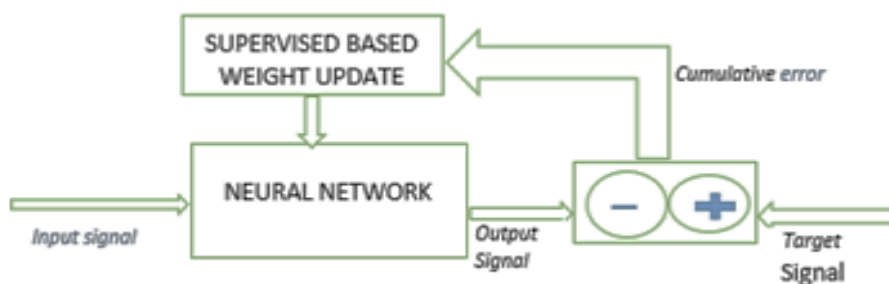
Εικόνα 12 Δομή νευρωνικού δικτύου [19]

Το νευρωνικό δίκτυο έχει παρατηρηθεί ότι έχει παρόμοια συμπεριφορά με έναν νευρώνα. Το νευρωνικό δίκτυο αποτελείται από τρία επίπεδα. Το επίπεδο εισόδου (input layer) όπως οι δενδρίτες, το κρυφό επίπεδο (hidden layer) που αντιστοιχεί στο

σώμα και τον άξονα του νευρώνα και τέλος το επίπεδο εξόδου (output layer) που αντιστοιχεί στις νευρικές απολήξεις. [20]

Ένα νευρωνικό δίκτυο μπορεί να χρησιμοποιηθεί σε δύο βασικές κατηγορίες μηχανικής μάθησης, εποπτευόμενη μάθηση και μη εποπτευόμενη. Στη συγκεκριμένη ενότητα αναφέρεται μόνο το νευρωνικό δίκτυο σε εποπτευόμενη μάθηση.

Κατά την εποπτευόμενη μάθηση, το αποτέλεσμα των δεδομένων εισόδου είναι γνωστό. Το αποτέλεσμα που έχει προβλεφθεί από το νευρωνικό δίκτυο συγκρίνεται με τα πραγματικά αποτελέσματα. Βάση του σφάλματος που έχει προκύψει μεταξύ προβλεπόμενης και πραγματικής εξόδου, οι παράμετροι ή αλλιώς συντελεστές βαρύτητας (weights) του νευρωνικού αναπροσαρμόζονται και στη συνέχεια εισάγονται ξανά στο δίκτυο. Στην Εικόνα 13 απεικονίζεται η διαδικασία που εκτελείται. [19] [20]



Εικόνα 13 Απεικόνιση λειτουργίας νευρωνικού δικτύου [19]

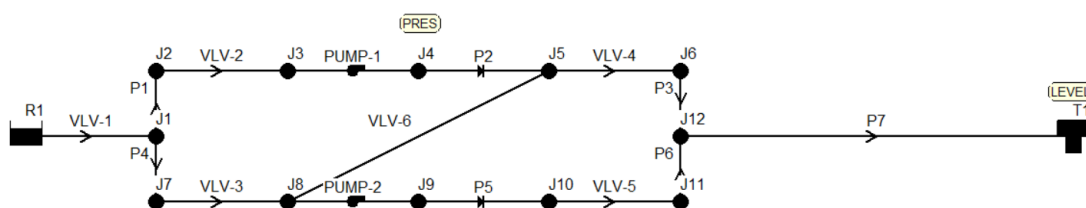
ΚΕΦΑΛΑΙΟ 4: Προσομοίωση Λειτουργίας Αντλιοστασίου με Μηχανική Μάθηση

Στην παρούσα εργασία διερευνάται η αποτελεσματικότητα χρήσης αλγορίθμων μηχανικής μάθησης για την πρόβλεψη της πίεσης, παροχής και συνολικής κατανάλωσης ενέργειας σε ένα αντλιοστάσιο νερού με δύο ταυτόσημες αντλίες το οποίο καταθλίβει σε δεξαμενή υψηλότερης στάθμης. Μέσω κατάλληλου χειρισμού των βανών στο δίκτυο σωληνώσεων του, οι αντλίες μπορούν να λειτουργήσουν ως μεμονωμένες, σε σειρά ή παράλληλα.

Το αντλιοστάσιο μοντελοποιείται στο λογισμικό EPANET το οποίο μπορεί να υπολογίσει με ακρίβεια τις τιμές πίεσης στους κόμβους, της παροχής στις σωληνώσεις και της συνολικής κατανάλωσης ενέργειας. Εκτελώντας πολλαπλά τυχαία σενάρια θέσης βανών και ύψους στάθμης της δεξαμενής κατάθλιψης, δημιουργείται ένα μεγάλο σετ δεδομένων εισόδου και εξόδου. Το σετ αυτό χρησιμοποιείται για την εκπαίδευση πολλαπλών μοντέλων μηχανικής μάθησης - τύπου Regression - στο πρόγραμμα MATLAB όπου και επιλέγεται το καταλληλότερο μοντέλο για την πρόβλεψη της κάθε μεταβλητής (πίεσης, παροχής, ενέργειας).

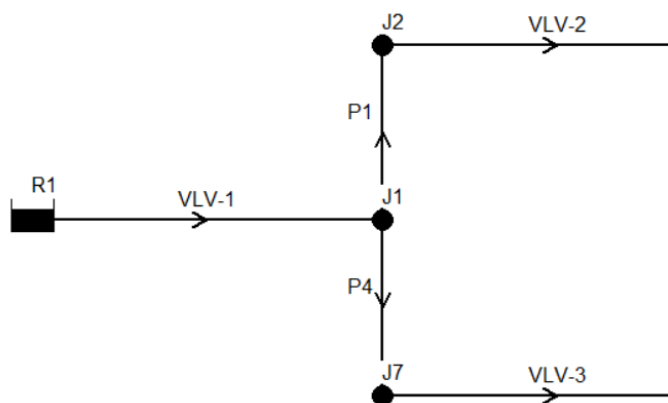
4.1: Περιγραφή μοντελοποίησης αντλιοστασίου

Το υδραυλικό δίκτυο που χρησιμοποιήθηκε για την προσομοίωση των διάφορων σεναρίων (συνδεσμολογία σε σειρά, παράλληλη ή μεμονωμένη λειτουργία των βανών) παρουσιάζεται στην Εικόνα 14 και αναλύεται παρακάτω.



Εικόνα 14 Μοντελοποίηση αντλιοστασίου στο λογισμικό EPANET.

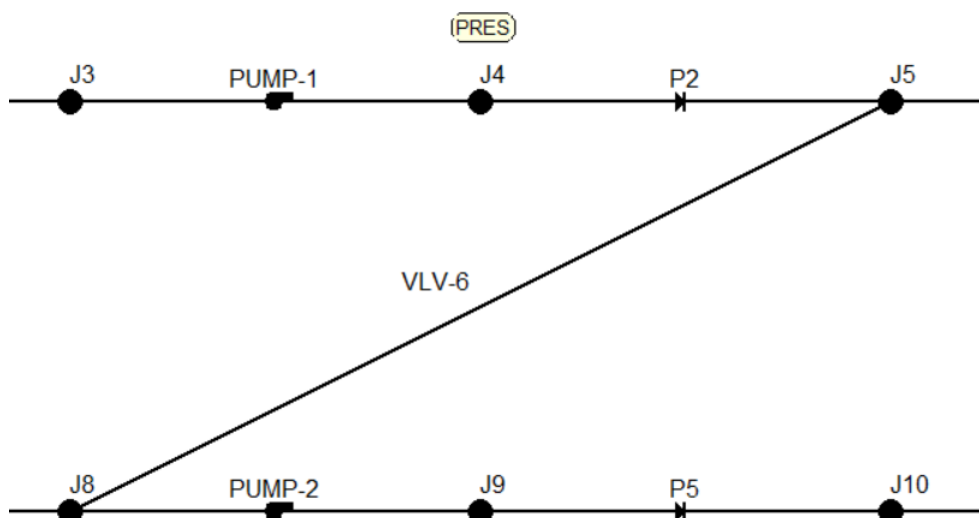
Για μεγαλύτερη ευκολία και ευκρίνεια της εικόνας θα παρουσιαστεί τμηματικά. Το πρώτο μέρος του δικτύου παρουσιάζεται στην Εικόνα 15 καθώς και φυσικές μεταβλητές που το αποτελούν.



Εικόνα 15 Πρώτο μέρος του δικτύου

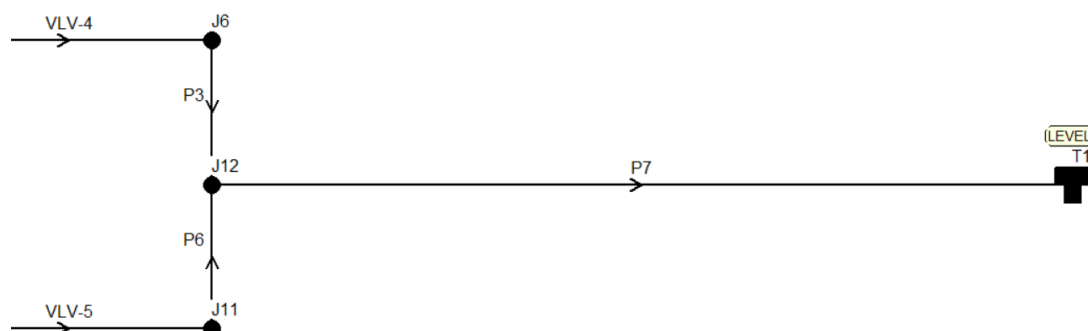
Στο παραπάνω σχήμα απεικονίζεται το πρώτο μέρος του δικτύου και οι μεταβλητές που το αποτελούν είναι:

- R1: αποτελεί το reservoir ή αλλιώς την πηγή από όπου ξεκινάει η ροή του νερού.
- VLV-1 (Pipe VLV-1): η σωλήνωση με βάννα, όπου συνεχίζει το νερό και ελέγχεται η ροή του μέσω της βάννας 1.
- J1 (Junction 1): ο κόμβος-διακλάδωση, στη συγκεκριμένη περίπτωση, στην οποία το νερό διαχωρίζεται και συνεχίζει προς τις σωληνώσεις P1 και P4, όπως δείχνουν και τα βέλη στο σχήμα (στις διακλαδώσεις από μηχανική των Ρευστών ισχύει $Q=Q_1+Q_2$, όπου Q η συνολική παροχή του νερού, Q_1 και Q_2 η παροχές του νερού, αφού εξέλθει της διακλάδωσης).
- P1, P4 (Pipe 1, pipe 4): οι σωληνώσεις του δικτύου χωρίς βάννες.
- J2, J7: οι διακλαδώσεις στις οποίες εισέρχεται και εξέρχεται το νερό και συνεχίζει στις σωληνώσεις VLV-2 και VLV-3 με βάννες.



Εικόνα 16 Δεύτερο μέρος του δικτύου

Στη συνέχεια το νερό συνεχίζει μέσω των VLV-2 και VLV-3 και εισέρχεται στους κόμβους J3 και J8, όπως φαίνεται στην Εικόνα 16. Το πάνω μέρος του δικτύου που αποτελείται από τον κόμβο J3, J4, J5, την αντλία PUMP-1, την σωλήνωση P2 με βαλβίδα ελέγχου. Το νερό είτε συνεχίζει μετά τον κόμβο J5 είτε από την σωλήνωση με βάννα VLV-6, ανάλογα εάν η βάννα που υπάρχει μετά τον κόμβο J5 είναι ανοιχτή ή κλειστή, το οποίο θα αναφερθεί εκ των υστέρων. Το κάτω μέρος του δικτύου αποτελείται από τις ίδιες μεταβλητές J8, J9, J10, την αντλία PUMP-2, την σωλήνωση P5 αντίστοιχα. Ο κόμβος J5 και J8 συνδέονται μεταξύ τους με μία σωλήνωση με βάννα, την βαλβίδα VLV-6, η οποία μπορεί να ανοίγει και να κλείνει αλλάζοντας έτσι την συνδεσμολογία του δικτύου.



Εικόνα 17 Τρίτο μέρος του δικτύου

Η Εικόνα 17 δείχνει το τρίτο μέρος του δικτύου στο οποίο φαίνεται ότι το νερό συνεχίζει είτε μέσω της σωλήνωσης με βαλβίδα VLV-4 είτε μέσω της VLV-5, εισέρχεται στον κόμβο-διακλάδωση J12 από τον οποίο εξέρχεται και καταλήγει στην δεξαμενή αποθήκευσης νερού T1.

Οι παραπάνω εικόνες απεικονίζουν το δίκτυο που κατασκευάστηκε χωρισμένο σε 3 κομμάτια, καθώς και οι μεταβλητές που το αποτελούν.

Κάθε σωλήνωση, δεξαμενή και αντλία χαρακτηρίζονται από συγκεκριμένα χαρακτηριστικά. Παρακάτω αναφέρονται τα κύρια χαρακτηριστικά των σωληνώσεων, πηγών, δεξαμενών, αντλιών, βαλβίδων και διακλαδώσεων/κόμβων που τα δεδομένα που εισάγονται σαν δεδομένα εισόδου.

4.1.1 Σωληνώσεις (Pipes)

Οι σωληνώσεις όπως έχει αναφερθεί σε προηγούμενο κεφάλαιο αποτελούν το μέσο μεταφοράς του νερού κατά μήκος του δικτύου. Τα κύρια δεδομένα εισόδου των σωληνώσεων είναι τα εξής:

- Αναφορά αρχή και τέλους της σωλήνωσης αναφέροντας από ποιον κόμβο ξεκινάει και σε ποιον καταλήγει.

- Διάμετρος και μήκος.
- Τραχύτητα του υλικού για τον υπολογισμό απωλειών.
- Αναφορά των σωληνώσεων (ανοιχτές, κλειστές ή αν φέρουν βαλβίδα ελέγχου)

Οι διαστάσεις των σωληνώσεων που χρησιμοποιήθηκαν στην παρούσα εργασία είναι:

Πίνακας 1 Διαστάσεις σωληνώσεων

	Μήκος (m)	Διάμετρος (mm)	Τραχύτητα (mm)
Pipe VLV-1	15	147.2	0.008116
Pipe VLV-2	0.5	102.2	0.008116
Pipe VLV-3	0.5	102.2	0.008116
Pipe VLV-4	0.5	90	0.008116
Pipe VLV-5	0.5	90	0.008116
Pipe VLV-6	2	90	0.008116
Pipe 1	3	102.2	0.008116
Pipe 2	5	90	0.008116
Pipe 3	3	90	0.008116
Pipe 4	3	102.2	0.008116
Pipe 5	5	90	0.008116
Pipe 6	3	90	0.008116
Pipe 7	500	130.8	0.008116

Το υλικό από το οποίο είναι φτιαγμένες οι σωληνώσεις είναι το πολυαιθυλένιο (πλαστικό) και συγκεκριμένα είναι ένα γραμμικό πολυμερές (LLDPE). Έχει μεγάλη αντοχή στον εφελκυσμό και περιβαλλοντική αντίσταση στις ρωγμές από τάση. [4]

4.1.2 Πηγή (Reservoir)

Το συγκεκριμένο είδος δεξαμενών χρησιμοποιείται για την μοντελοποίηση πηγών, όπως λίμνες, ποτάμια. Το κύριο χαρακτηριστικό εισόδου είναι το μέτρο στήλης υγρού (H) και η αρχική ποιότητα για την ανάλυση ποιότητας νερού. [4]

4.1.3 Δεξαμενές (Tanks)

Οι δεξαμενές είναι τα μέσα που χρησιμοποιούνται κατά την προσομοίωση για την αποθήκευση νερού. Τα δεδομένα εισόδου των δεξαμενών είναι:

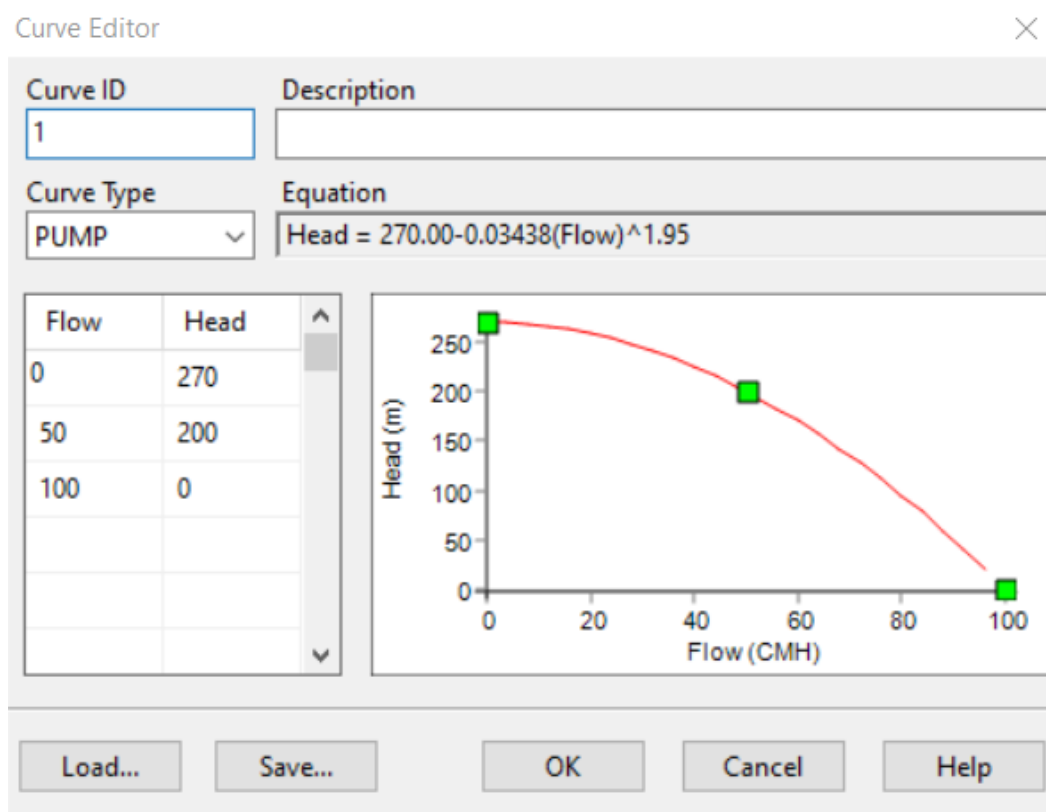
- Το κατώτατο επίπεδο του νερού, όπου το επίπεδο του νερού είναι μηδενικό.

- Διάμετρος ή το σχήμα αν δεν είναι κυλινδρικού σχήματος.
- Αρχικό, ελάχιστο και μέγιστο επίπεδο νερού (initial, minimum, maximum water levels).
- Αρχική ποιότητα νερού

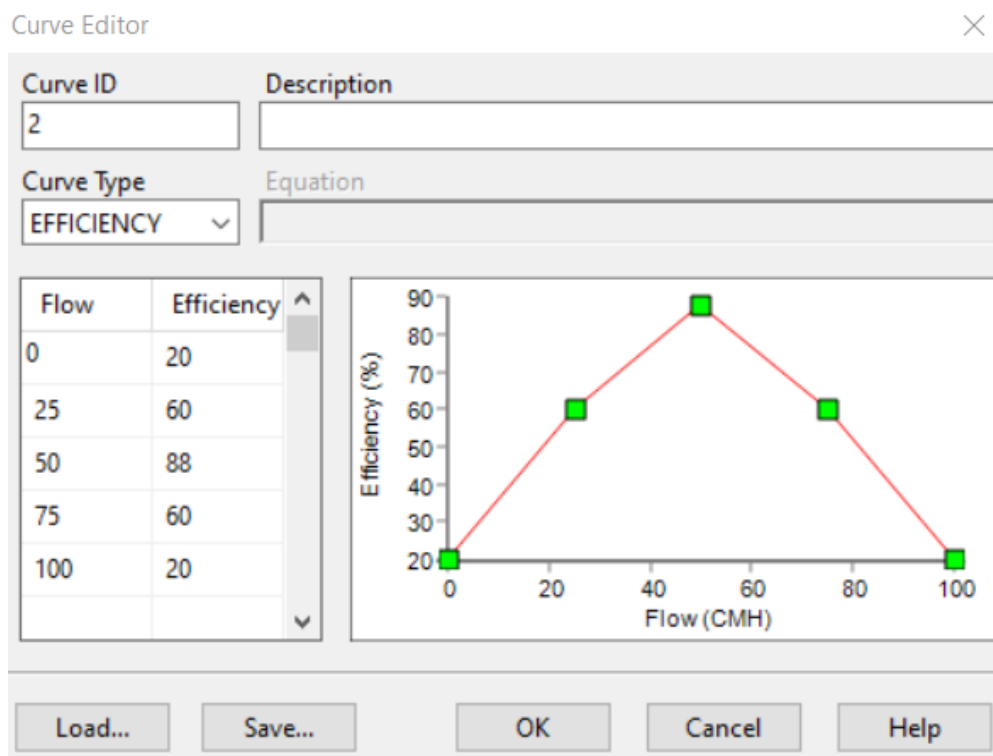
Το EPANET σταματάει την εκροή και την εισροή εάν μία δεξαμενή βρίσκεται στο ελάχιστο και μέγιστο επίπεδο αντίστοιχα. Η διάμετρος της δεξαμενής που χρησιμοποιήθηκε στο δίκτυο είναι 15 m. [4]

4.1.4 Αντλίες (Pumps)

Τα χαρακτηριστικά εισόδου μίας αντλίας θεωρούνται οι κόμβοι αρχής και τέλους και η χαρακτηριστική καμπύλη της αντλίας. Η αντλία μπορεί να θεωρηθεί μια συσκευή, η οποία παρέχει συνεχές ποσό ενέργειας στο νερό για όλους τους δυνατούς συνδυασμούς ροής και μέτρου στήλης υγρού. Στην Εικόνα 18 απεικονίζεται η χαρακτηριστική της αντλίας του δικτύου νερού που μοντελοποιήθηκε.



Εικόνα 18 Χαρακτηριστική αντλίας.



Εικόνα 19 Καμπύλη σχέσης απόδοσης-παροχής.

Στην Εικόνα 19 φαίνεται η χαρακτηριστική της αντλίας και η εξίσωση της που χρησιμοποιήθηκε καθώς και το ζεύγος σημείων ροής-μέτρο στήλης υγρού, βάση των οποίων δημιουργήθηκε η καμπύλη μέσω της εξίσωσης. Στην επόμενη εικόνα, απεικονίζεται η καμπύλη σχέσης απόδοσης-ροής. Τέλος, όπως και στις σωληνώσεις και οι αντλίες μπορούν να ρυθμιστούν να κλείνουν και να ανοίγουν και σε συγκεκριμένα χρονικά διαστήματα με την χρήση συνθηκών. Οι αντλίες που χρησιμοποιήθηκαν είχαν κοινές χαρακτηριστικές καμπύλες και καμπύλες απόδοσης-ροής. [4]

4.1.5 Βαλβίδες (Valves)

Τα δεδομένα εισόδου των βαλβίδων κατά την προσομοίωση του δικτύου είναι τα παρακάτω:

- Αρχικός και τελικός κόμβος.
- Διάμετρος και ρυθμίσεις.
- Η κατάσταση στην οποία βρίσκεται μία βαλβίδα.

Τα δεδομένα εξόδου είναι ο ρυθμός ροής και οι απώλειες. [4]

4.1.6 Κόμβοι (Nodes)

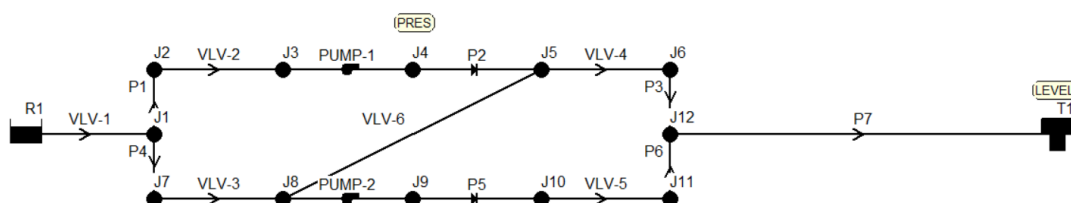
Οι διακλαδώσεις/κόμβοι είναι τα σημεία που ενώνουν σωληνώσεις με αντλίες ή δεξαμενές. Τα βασικά δεδομένα εισόδου που χρειάζονται κατά την προσομοίωση είναι:

- Ανύψωση του νερού, που συνήθως είναι πάνω από το επίπεδο της θάλασσας.
- Η ποσότητα του νερού που απαιτείται.
- Η αρχική ποιότητα του νερού.

Τα χαρακτηριστικά που υπολογίζονται σε έναν κόμβο είναι η ποιότητα νερού, η πίεση που δημιουργείται και η εσωτερική ενέργεια ανά μονάδα βάρους του υγρού. Τέλος στους κόμβους η απαίτηση νερού ποικίλει με τον χρόνο, η οποία μπορεί να έχει και αρνητικές τιμές δείχνοντας την εισαγωγή νερού στο δίκτυο, οι κόμβοι μπορεί να λειτουργούν ως σημεία εισαγωγής χημικών συστατικών. [4]

4.1.7 Συνδεσμολογία δικτύου νερού

Προηγουμένως αναφέρθηκαν ανά κατηγορία, οι φυσικές μεταβλητές του δικτύου νερού καθώς και τα βασικά δεδομένα εισόδου της κάθε μίας μεταβλητής. Το δίκτυο είναι φτιαγμένο έτσι ώστε να υποστηρίξει συνδεσμολογία σε σειρά, παράλληλα και να όλες οι βάνες ανοιχτές ή κλειστές. Στη συνέχεια παρουσιάζεται ολόκληρο το δίκτυο καθώς και ένας πίνακας που δείχνει ποιες βάνες πρέπει να είναι ανοιχτές ή κλειστές, έτσι ώστε να δημιουργηθεί η επιθυμητή συνδεσμολογία.



Εικόνα 20 Υδραυλικό δίκτυο.

Πίνακας 2 Συνδεσμολογία δικτύου

ΣΥΝΔΕΣΜΟΛΟΓΙΑ / ΒΑΛΒΙΔΕΣ	ΠΑΡΑΛΛΗΛΑ	ΣΕ ΣΕΙΡΑ	ΑΝΟΙΚΤΕΣ	ΚΛΕΙΣΤΕΣ
1(VLV-1)	OPEN	OPEN	OPEN	CLOSE
2(VLV-2)	OPEN	OPEN	OPEN	CLOSE
3(VLV-3)	OPEN	CLOSE	OPEN	CLOSE
4(VLV-4)	OPEN	CLOSE	OPEN	CLOSE
5(VLV-5)	OPEN	OPEN	OPEN	CLOSE
6(VLV-6)	CLOSE	OPEN	OPEN	CLOSE

Όπως παρουσιάζεται στην Εικόνα 20 και στον πίνακα, αυτοί πρέπει να είναι οι συνδυασμοί της κατάστασης των βανών για να επιτευχθεί η κάθε συνδεσμολογία.

4.2: Διαδικασία παραγωγής δεδομένων

Η παραγωγή των δεδομένων που χρησιμοποιήθηκαν στην παρούσα εργασία πραγματοποιήθηκε από μία σειρά εκτέλεσης αλγορίθμων, οι οποίοι δημιουργήθηκαν σε μορφή κώδικα στο περιβάλλον του λογισμικού MATLAB. Οι κώδικες και το αρχείο που χρησιμοποιήθηκε ως δεδομένα εισόδου θα παρουσιαστούν σε παράρτημα και θα αναλυθεί επιγραμματικά η λειτουργία τους. Πριν την εκτέλεση του κάθε κώδικα ξεχωριστά, έγινε export το input file (αρχείο εισόδου) του δικτύου που μοντελοποιήθηκε στην εργασία και τροποποιήθηκε κατάλληλα. Έγινε αντικατάσταση της κατάστασης των βανών (open, closed) και του αρχικού επιπέδου του νερού (initial level) με μεταβλητές, οι οποίες θα παίρνουν τυχαίες τιμές για να δημιουργούνται οι τυχαίοι συνδυασμοί.

Όπως φαίνεται στο Παράρτημα 1, δημιουργήθηκε κώδικας με το λογισμικό MATLAB με τον οποίο δημιουργήθηκαν τα αρχεία δεδομένων εισόδου, ένα συγκεντρωτικό αρχείο στο οποίο υπάρχουν τυχαίοι συνδυασμοί συνδεσμολογίας και του αρχικού επιπέδου νερού, καθώς και ένα αρχείο «.bat». Με την βοήθεια του αρχείου εκτελέστηκε κάθε συνδυασμός που δημιουργήθηκε προηγουμένως και δημιουργήθηκαν τα αντίστοιχα report files, τα οποία περιέχουν τα αποτελέσματα από την εκτέλεση του κάθε συνδυασμού μέσω του EPANET.

Πιο συγκεκριμένα, στην αρχή του κώδικα δίνεται από τον χρήστη ο αριθμός των αρχείων που θα παραχθούν, δηλαδή κάθε αρχείο είναι και ένας συνδυασμός συνδεσμολογίας και του αρχικού επιπέδου του νερού. Στη συνέχεια με κατάλληλες εντολές ανοίχθηκε το αρχείο εισόδου (nettemplate.inp) που θα χρησιμοποιηθεί για την δημιουργία των συνδυασμών, δημιουργήθηκαν τα αρχεία του bat αρχείου (eparuns_train.bat) και του αρχείου όπου θα περιέχει τον κάθε συνδυασμό που θα παραχθεί και θα λειτουργήσει αργότερα σαν είσοδος στην εκπαίδευση των μοντέλων μηχανικής μάθησης (neurin_train.txt). Στο κύριο μέρος του κώδικα, δόθηκαν τυχαίες τιμές στις 7 μεταβλητές (6 βάνες και 1 μεταβλητή για το αρχικό επίπεδο του νερού), οι οποίες με κατάλληλες εντολές αντικαταστάθηκαν στο αρχικό αρχείο εισόδου που χρησιμοποιήθηκε για την δημιουργία των συνδυασμών. Τέλος, εισήχθησαν μέσα στο bat αρχείο οι απαραίτητες εντολές για να εκτελεστεί μέσω DOS, αντικαταστάθηκαν οι χαρακτήρες «Open» και «Closed» με «1» και «0» αντίστοιχα για να μπορούν να αναγνωριστούν σαν είσοδος από το κάθε μοντέλο μηχανικής μάθησης.

Όπως φαίνεται στο παράρτημα 2, δημιουργήθηκε ένας κώδικας ο οποίος ύστερα από μια σειρά κατάλληλων εντολών συλλέγει τρεις συγκεκριμένες τιμές (ενέργεια, πίεση και ροή) από τα report files, δηλαδή τα αρχεία που εξόδου κάθε συνδυασμού που δημιουργήθηκαν και εκτελέστηκαν με την βοήθεια του bat αρχείου στον

προηγούμενο κώδικα. Αφού συλλεχθούν οι τρεις μεταβλητές εισάγονται σε ένα αρχείο εξόδου. Τέλος, δημιουργείται ένα συγκεντρωτικό αρχείο που περιέχει όλους τους συνδυασμούς συνδεσμολογίας, το αρχικό επίπεδο του νερού και τις αντίστοιχες τιμές εξόδου.

Στην αρχή του κώδικα ορίζονται οι μεταβλητές της ροής και της πίεσης που θα συλλεχθούν από τα αρχεία εξόδου που δημιουργήθηκαν στον προηγούμενο κώδικα. Πιο συγκεκριμένα, συλλέγονται οι τιμές της ροής στην σωλήνωση «Pipe 7», η πίεση στην διακλάδωση «Junction 12», η οποία είναι η αρχή της σωλήνωσης 7 και η ενέργεια που καταναλώθηκε συνολικά κατά την διάρκεια λειτουργίας του δικτύου. Στην συνέχεια του κώδικα δημιουργούνται δύο αρχεία, το αρχείο που θα συγκεντρωθούν οι τρεις τιμές και το αρχείο που θα περιέχει τον κάθε συνδυασμό και τα δεδομένα εξόδου του κάθε συνδυασμού. Ύστερα με τρεις ξεχωριστές επαναλήψεις, διαβάζονται όλα τα αρχεία εξόδου και συλλέγονται οι τιμές της ενέργειας, πίεσης και ροής και εισάγονται στο αρχείο εξόδου (neurout_train.txt) και κατ' επέκτασιν στο συγκεντρωτικό αρχείο (dataset_train.txt) που θα λειτουργήσει ως το αρχείο εκπαίδευσης των μοντέλων.

Όπως φαίνεται στο παράρτημα 3, δημιουργήθηκε ένας κώδικας με τον οποίο δημιουργήθηκαν οι 16 συνδυασμοί συνδεσμολογίας. Τα σενάρια που δημιουργήθηκαν ήταν προκαθορισμένα με συγκεκριμένες τιμές. Χωρίστηκαν σε 4 κατηγορίες που η κάθε κατηγορία περιέχει μία συγκεκριμένη συνδεσμολογία και η μεταβλητή που αλλάζει είναι η αρχική στάθμη νερού δεξαμενής ως προς τον πυθμένα της. Τα σενάρια που δημιουργήθηκαν είναι τα εξής:

- Βάνες όλες κλειστές, στάθμη νερού δεξαμενής στα 0, 15, 45 και 60 m.
- Βάνες όλες ανοιχτές, στάθμη νερού δεξαμενής στα 0, 15, 45 και 60 m.
- Παράλληλη συνδεσμολογία, στάθμη νερού δεξαμενής στα 0, 15, 45 και 60 m.
- Συνδεσμολογία σε σειρά με στάθμη νερού δεξαμενής στα 0, 15, 45 και 60 m.

Στην αρχή του κώδικα δίνονται συγκεκριμένες τιμές στις 6 βάνες και στην αρχική στάθμη του νερού, ανάλογα με την κατάσταση των βανών (κλειστές, ανοιχτές) και την συνδεσμολογία. Στην συνέχεια ακολουθήθηκε η ίδια διαδικασία με τον κώδικα του παραρτήματος 1, δίνονται οι προκαθορισμένες τιμές στις μεταβλητές των βανών και της αρχικής στάθμης, δημιουργείται το αρχείο bat που έχει την ίδια λειτουργία με εκείνου του πρώτου κώδικα και δίνονται οι τιμές «0» και «1» για τις μεταβλητές χαρακτήρων «Closed» και «Open» αντίστοιχα.

Στο τελευταίο παράρτημα, παράρτημα 4, δημιουργήθηκε παρόμοιος κώδικας με αυτού του παραρτήματος 2. Πιο συγκεκριμένα, με ίδιες εντολές συλλέχθηκαν οι 3 συγκεκριμένες τιμές της ενέργειας, πίεσης και ροής στους ίδιο κόμβο και σωλήνωση και συγκεντρώθηκαν σε ένα αρχείο. Στη συνέχεια συγκεντρώθηκαν οι 16 προκαθορισμένες τιμές των βανών και του αρχικού επιπέδου νερού στην δεξαμενή αποθήκευσης, καθώς και οι αντίστοιχες τιμές εξόδου (ενέργειας, πίεσης, ροής) σε ένα

ξεχωριστό αρχείο, το οποίο χρησιμοποιήθηκε στην συνέχεια σαν test dataset για την επικύρωση της αποτελεσματικότητας του εκπαιδευμένου μοντέλου.

4.3 Επιλογή & επικύρωση μοντέλου μηχανικής μάθησης

Μετά την εκτέλεση του κάθε κώδικα ξεχωριστά και την δημιουργία των απαραίτητων αρχείων εκπαίδευσης των μοντέλων μηχανικής μάθησης επιλέχθηκε το κατάλληλο μοντέλο για κάθε μεταβλητή εισόδου (ενέργεια, πίεση, παροχή). Το επιλεγμένο μοντέλο επικυρώθηκε συγκρίνοντας τις προβλεπόμενες από το μοντέλο τιμές με τις υπολογισμένες τιμές από το EPANET για 16 προκαθορισμένα σενάρια. Συγκεκριμένα, δημιουργήθηκαν δύο αρχεία:

dataset_train.txt – αρχείο με όλα τα τυχαία σενάρια μεταβλητών εισόδου και αντίστοιχα αποτελέσματα για την εκπαίδευση των μοντέλων

dataset_test.txt - αρχείο με τα 16 προκαθορισμένα σενάρια μεταβλητών εισόδου και αντίστοιχα αποτελέσματα για την επικύρωση των μοντέλων

Ως γενικός κανόνας, η επιλογή των μοντέλων μηχανικής μάθησης γίνεται με τα εξής κύρια κριτήρια:

- Τεχνική μηχανικής μάθησης που θα χρησιμοποιηθεί, τύπος προβλήματος (πρόβλημα παλινδρόμησης, regression) και ο απαραίτητος αλγόριθμος που ανήκει σε αυτή την κατηγορία (Εποπτευόμενη μάθηση, Supervised learning)
- Το μέγεθος του αρχείου δεδομένων για κάθε αλγόριθμο.
- Η ταχύτητα με την οποία εκπαιδεύεται ο κάθε αλγόριθμος.
- Κατά πόσο εύκολο ήταν για τον αλγόριθμο να φτάσει σε μία τελική απόφαση.

Τα παραπάνω κριτήρια απεικονίζονται στον Πίνακα 3.

Πίνακας 3 Κριτήρια επιλογής μοντέλου μηχανικής μάθησης [21]

Algorithm	Dataset <i>What is the ideal dataset size for each algorithm?</i>	Training Speed <i>How quickly will the algorithm train without acceleration hardware?</i>	Interpretability <i>How hard is it to see how the algorithm arrived at a decision?</i>	Tuning <i>How much tuning does the algorithm allow?</i>	Comments
Linear models	Small	Very fast	Easy	Minimal	Widely used basic algorithm Linear SVM handles high-dimensional data well
Decision trees	Small	Very fast	Easy	Some	Good generalist algorithm, check for overfitting
(Nonlinear) Support vector machine	Medium sized	Moderately slow	Difficult	Some	Good accuracy
Nearest neighbor	Medium sized	Moderately fast	Moderately easy	Minimal	Lower accuracy, but easy to use and interpret
Naïve Bayes	Medium sized	Very fast	Moderately easy	Some	Widely used for text analytics (e.g., spam filtering); kernel Bayes will run slower
Ensembles	Large	Moderately fast	Difficult	Some	Higher accuracy with a tradeoff of lower interpretability
Neural network (shallow)	Medium sized	Moderately fast	Moderately easy	Some	Still used for signal classification, compression, and forecasting
Deep nets	Large	Very slow	Difficult	A lot	A standard algorithm for image, video, signals, and text

Από τα αριστερά προς τα δεξιά φαίνονται τα ονόματα των αλγορίθμων, το μέγεθος του αρχείου δεδομένων εισόδου, ταχύτητα εκπαίδευσης αλγορίθμων, κατά πόσο ο αλγόριθμος δυσκολεύτηκε για να φτάσει σε μία απόφαση, πόσο συντονισμό επιτρέπει ο αλγόριθμος και τα σχόλια για τον κάθε αλγόριθμο. Η επιλογή του κατάλληλου μοντέλου μηχανικής μάθησης έγινε με την βοήθεια ενός εργαλείου του προγράμματος MATLAB που ονομάζεται Regression Learner App.

Η εφαρμογή Regression Learner App χρησιμοποιείται για την εκπαίδευση μοντέλων παλινδρόμησης. Η διαδικασία χωρίζεται σε 2 μέρη:

- Την εκπαίδευση των μοντέλων μηχανικής μάθησης με τα δεδομένα εκπαίδευσης για την επιλογή του κατάλληλου μοντέλου.

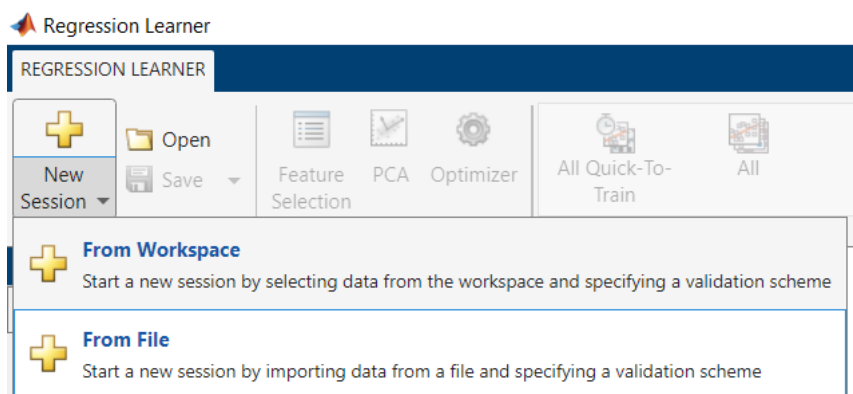
- Επιλογή του μοντέλου με την μικρότερη απόκλιση μέσου τετραγωνικού σφάλματος (Root mean square error, RMSE) και χρήση των δεδομένων δοκιμής για την επικύρωση του μοντέλου.

Υπάρχουν δύο είδη επιλογής και εκπαίδευσης μοντέλων, η αυτοματοποιημένη εκπαίδευση και η χειροκίνητη. Στην παρούσα εργασία χρησιμοποιήθηκε η αυτοματοποιημένη εκπαίδευση. Παρακάτω θα παρουσιαστούν και θα επιλεγθεί το κατάλληλο μοντέλο για 10, 100, 1000 και 10000 συνδυασμούς ως προς κάθε μεταβλητή ξεχωριστά (ενέργεια, πίεση, ροή).

4.3.1 Επιλογή μοντέλου και επικύρωση του για 10 αρχεία

Αρχικά δημιουργήσαμε τα 10 αρχεία με τον κώδικα του παραρτήματος 1, δίνοντας το πλήθος αρχείων ίσο με 10. Στη συνέχεια αφού παράχθηκαν τα 10 αρχεία, εκτελέστηκε το λογισμικό EPANET μέσω DOS με την χρήση του bat αρχείου. Έπειτα δημιουργήθηκαν τα αρχεία εξόδου κάθε συνδυασμού με τα αποτελέσματα τους αντίστοιχα. Με την χρήση του κώδικα του παραρτήματος 2 συγκεντρώθηκαν οι τιμές της ενέργειας, πίεσης και ροής καθώς και οι 10 συνδυασμοί συνδεσμολογίας σε ένα αρχείο που θα χρησιμοποιηθεί σαν αρχείο εκπαίδευσης. Η διαδικασία που ακολουθείται για την εκπαίδευση είναι η εξής:

1. Επιλέγουμε στο λογισμικό MATLAB στην καρτέλα των «APP» την επιλογή «Regression Learner».
2. Μόλις ανοίξει η εφαρμογή εισάγουμε το αρχείο εκπαίδευσης μέσω της επιλογής «New Session > From a file», όπως φαίνεται στην Εικόνα 21.



Εικόνα 21 Διαδικασία εισαγωγής δεδομένων

3. Στη συνέχεια επιλέγουμε «Import Selection», όπως στην Εικόνα 22

Import Text - C:\SENARIA\TRAIN\dataset_train.txt

IMPORT

☒ Delimited ☐ Fixed Width

Column Delimiters Selection/History A1:J10 (init...)

Delimiter Options Variable Names Row 1

☐ No Variable Names

IMPORTED VARIABLE UNIMPORTABLE CELLS

Import Selection IMPORT

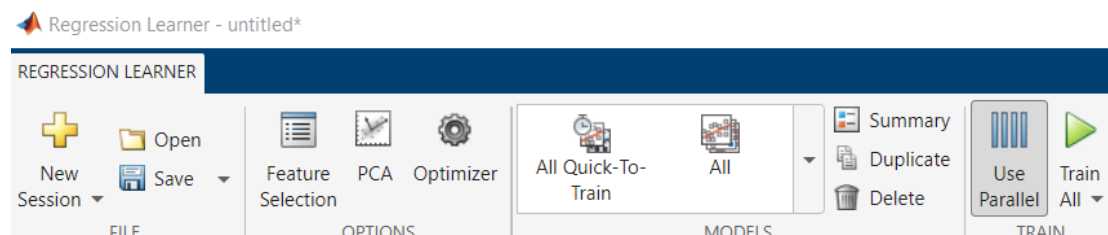
FILE TYPE DELIMITERS SELECTION

dataset_train.txt

	A	B	C	D	E	F	G	H	I	J
	Var1	Var2	Var3	Var4	Var5	Var6	Var7	Var8	Var9	Var10
	Number	Number	Number	Number	Number	Number	Number	Number	Number	Number
1	0	0	1	0	0	1	16.7099	7.36	166.71	0
2	0	0	0	1	0	0	29.1225	7.36	179.12	0
3	0	1	1	0	0	0	39.3444	6.92	189.34	0
4	1	0	0	0	0	0	23.5336	7.36	173.53	0
5	0	1	0	1	1	1	5.8279	3.71	155.83	0
6	0	0	1	0	1	1	22.8935	7.36	172.89	0
7	0	0	1	1	1	0	42.5619	4.56	192.56	0
8	0	1	0	0	1	1	29.9018	7.36	179.9	0
9	0	1	0	1	0	1	30.3574	7.36	180.36	0
10	0	0	0	0	1	1	15.4505	7.36	165.45	0

Εικόνα 22 Εισαγωγή δεδομένων

4. Στην καρτέλα που εμφανίζεται, παρουσιάζονται δύο βασικές επιλογές. Η επιλογή του «Data Set Variable» και του «Response». Η πρώτη είναι τα δεδομένα εισόδου, όπως φαίνεται στην παραπάνω εικόνα, ένας πίνακας 10x10 και η δεύτερη είναι η επιλογή της μεταβλητής εξόδου (ενέργεια, πίεση, ροή) ως προς την οποία θα εκπαιδευτούν τα μοντέλα μηχανικής μάθησης. Κάθε φορά μπορεί να επιλεγεί μία μόνο μεταβλητή εξόδου οπότε η διαδικασία θα επαναληφθεί τρεις φορές. Η μεταβλητές εξόδου είναι οι Var 8, Var 9, Var 10 για την ενέργεια, πίεση και ροή αντίστοιχα.
5. Στη συνέχεια στο παράθυρο της εφαρμογής επιλέγουμε «All» για να επιλεχθούν προς εκπαίδευση όλα τα διαθέσιμα μοντέλα παλινδρόμησης και ύστερα την επιλογή «Train All», όπως στην Εικόνα 23.



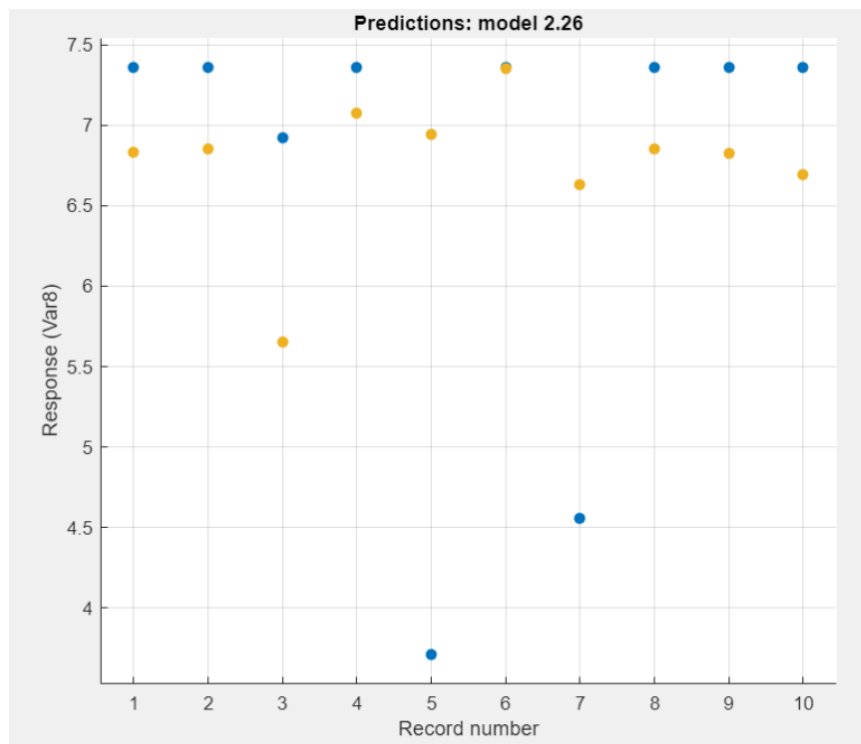
Εικόνα 23 Επιλογή εκπαίδευσης μοντέλων

6. Αφού έγινε η εκπαίδευση των μοντέλων, επιλέγεται εκείνο με τον μικρότερο δείκτη τετραγωνικού σφάλματος. Για τον συγκεκριμένο αριθμό αρχείων είναι το μοντέλο νούμερο 2.26 Kernel, με απόκλιση της μέσης τετραγωνικής τιμής 1.3394

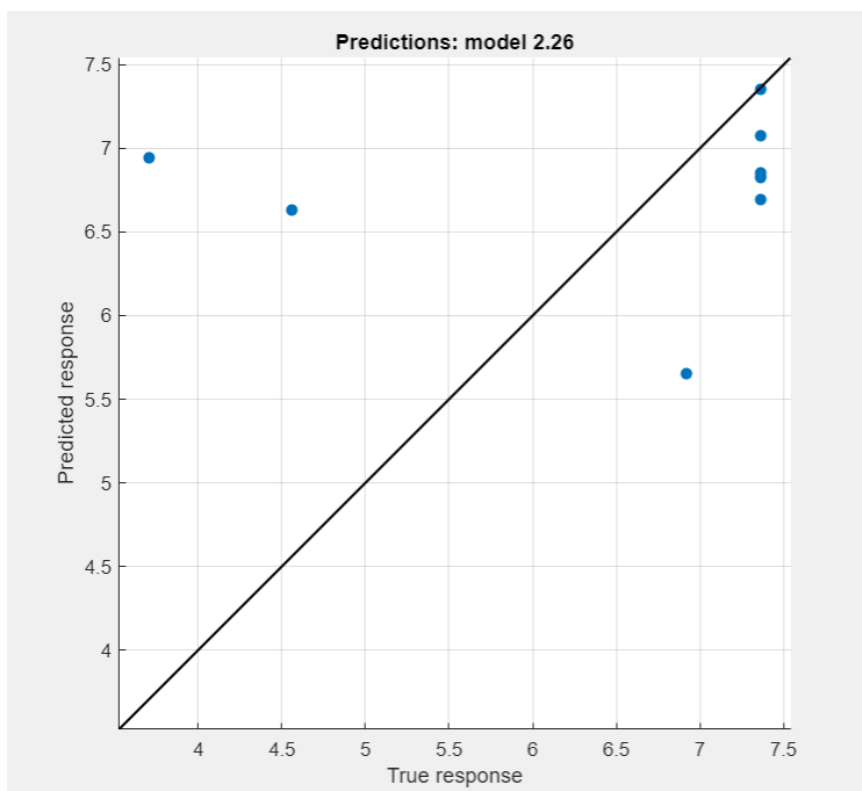
☆	2.23 Neural Network	RMSE (Validation): 2.8164
	Last change: Bilayered Neural Network	9/9 features
☆	2.24 Neural Network	RMSE (Validation): 3.0335
	Last change: Trilayered Neural Network	9/9 features
☆	2.25 Kernel	RMSE (Validation): 1.3757
	Last change: SVM Kernel	9/9 features
☆	2.26 Kernel	RMSE (Validation): 1.3394
	Last change: Least Squares Regression Kernel	9/9 features

Εικόνα 24 Βέλτιστο μοντέλο βάση RMSE

Στη συνέχεια μπορούν να επιλεγθούν μερικές κύριες γραφικές αναπαραστάσεις, όπως παρουσιάζονται στις παρακάτω εικόνες.



Γραφική Παράσταση 1 Απόκριση τιμών



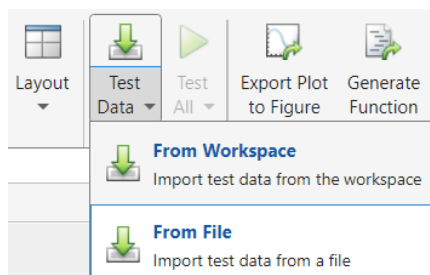
Γραφική Παράσταση 2 Πρόβλεψη μοντέλου

Στην Γραφική Παράσταση 1 «Response Plot» παρουσιάζονται τα αποτελέσματα του μοντέλου. Πιο συγκεκριμένα εμφανίζονται οι προβλεπόμενες αποκρίσεις σε σχέση με τον αριθμό δεδομένων που εισήχθησαν σαν δεδομένα εισόδου (10).

Στην Γραφική Παράσταση 2 παρουσιάζεται κατά πόσο καλά το μοντέλο κάνει προβλέψεις για διάφορες τιμές απόκρισης. Ο άξονας X αντικατοπτρίζει τις πραγματικές αποκρίσεις και ο άξονας Y τις προβλεπόμενες. Το τέλειο μοντέλο θα παρουσίαζε τις τιμές να ταυτίζονται με την γραμμή παλινδρόμησης, και όχι διασκορπισμένες περί της γραμμής αυτής.

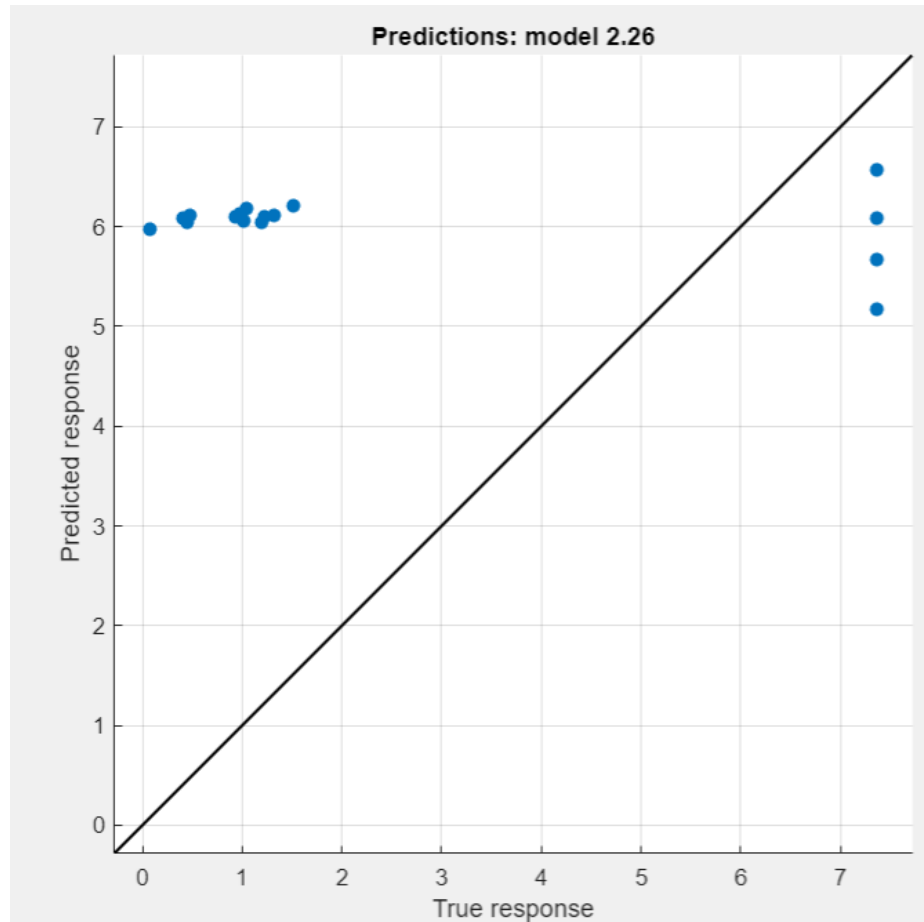
Προφανώς για τον συγκεκριμένο αριθμό των αρχείων (10) που επιλέχθηκαν δεν ήταν δυνατή η σωστή πρόβλεψη, λόγω έλλειψης όγκου δεδομένων.

7. Αφού επιλέξαμε το κατάλληλο μοντέλο, εισάγουμε τα δεδομένα δοκιμής. Επιλέγουμε «Test Data > From File» και στην συνέχεια την επιλογή «Train», όπως στην Εικόνα 25.



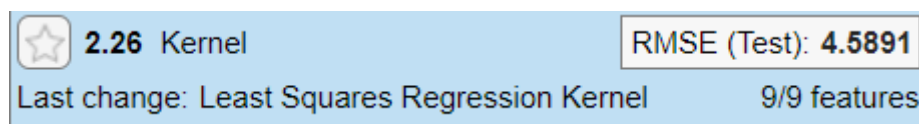
Εικόνα 25 Εισαγωγή δεδομένων δοκιμής

8. Αφού γίνει η εκπαίδευση με το αρχείο δοκιμής μπορούμε να δούμε στην Γραφική Παράσταση 3 που παρουσιάζει τις προβλεπόμενες τιμές από το αρχείο δοκιμής σε σχέση με τις πραγματικές τιμές.



Γραφική Παράσταση 3 Πρόβλεψη τιμών για τα 16 σενάρια

Τέλος παρατηρείται ότι η απόκλιση της μέση τετραγωνικής τιμής αυξάνεται για τις 16 προκαθορισμένες τιμές.



Εικόνα 26 Αυξημένη απόκλιση μέσης τετραγωνικής τιμής

Αυτή ήταν η διαδικασία που ακολουθείται για την εκπαίδευση των μοντέλων. Προφανώς για τον ελάχιστο αριθμό δεδομένων δεν μπορούμε να έχουμε σαφή αποτελέσματα ακόμα και για τις άλλες 2 μεταβλητές. [21]

4.3.2 Επιλογή μοντέλου και επικύρωση του για 100 αρχεία

Μεταβλητή 8 (Var 8, ενέργεια)

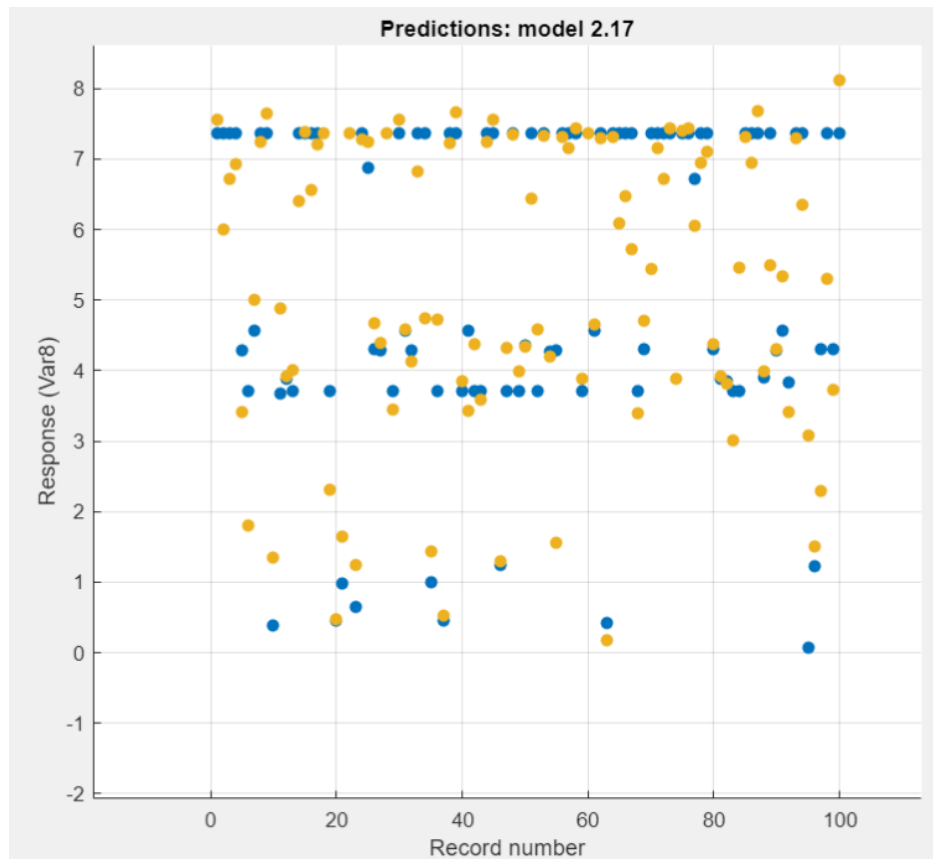
Για την επιλογή του μοντέλου ως όπως την Var 8, που αντιπροσωπεύει την ενέργεια, ακολουθήθηκε η ίδια διαδικασία και τα αποτελέσματα ως όπως την μεταβλητή όπως ενέργειας είναι τα ακόλουθα.

Το μοντέλο με τον καλύτερο δείκτη τετραγωνικού σφάλματος είναι το μοντέλο παλινδρόμησης Gaussian Process, Matern 5/2 GPR (Gaussian Process Regression), με RMSE (Validation): 0.84796, όπως φαίνεται στην Εικόνα 27.

☆ 2.14 Ensemble	RMSE (Validation): 1.092
Last change: Boosted Trees	9/9 features
☆ 2.15 Ensemble	RMSE (Validation): 1.5738
Last change: Bagged Trees	9/9 features
☆ 2.16 Gaussian Process Regr...	RMSE (Validation): 0.89843
Last change: Squared Exponential GPR	9/9 features
☆ 2.17 Gaussian Process Regr...	RMSE (Validation): 0.84796
Last change: Matern 5/2 GPR	9/9 features

Εικόνα 27 RMSE, Var8, 100 αρχεία

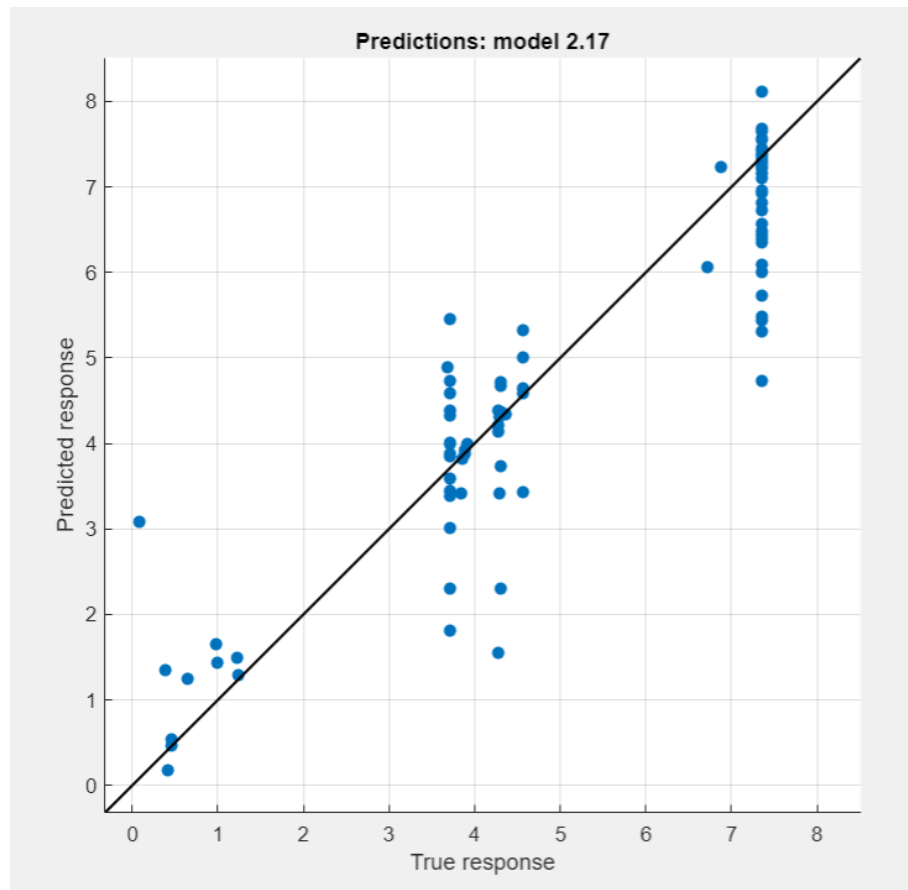
Στην Γραφική Παράσταση 4 φαίνεται η απόκριση των τιμών:



Γραφική Παράσταση 4 Απόκριση τιμών, Var8, 100 αρχεία.

Οι μπλε τιμές είναι οι πραγματικές τιμές και οι κίτρινες οι προβλεπόμενες. Βλέπουμε πως υπάρχει καλύτερη κατανομή και ταύτιση των τιμών όσο αυξάνονται τα δεδομένα.

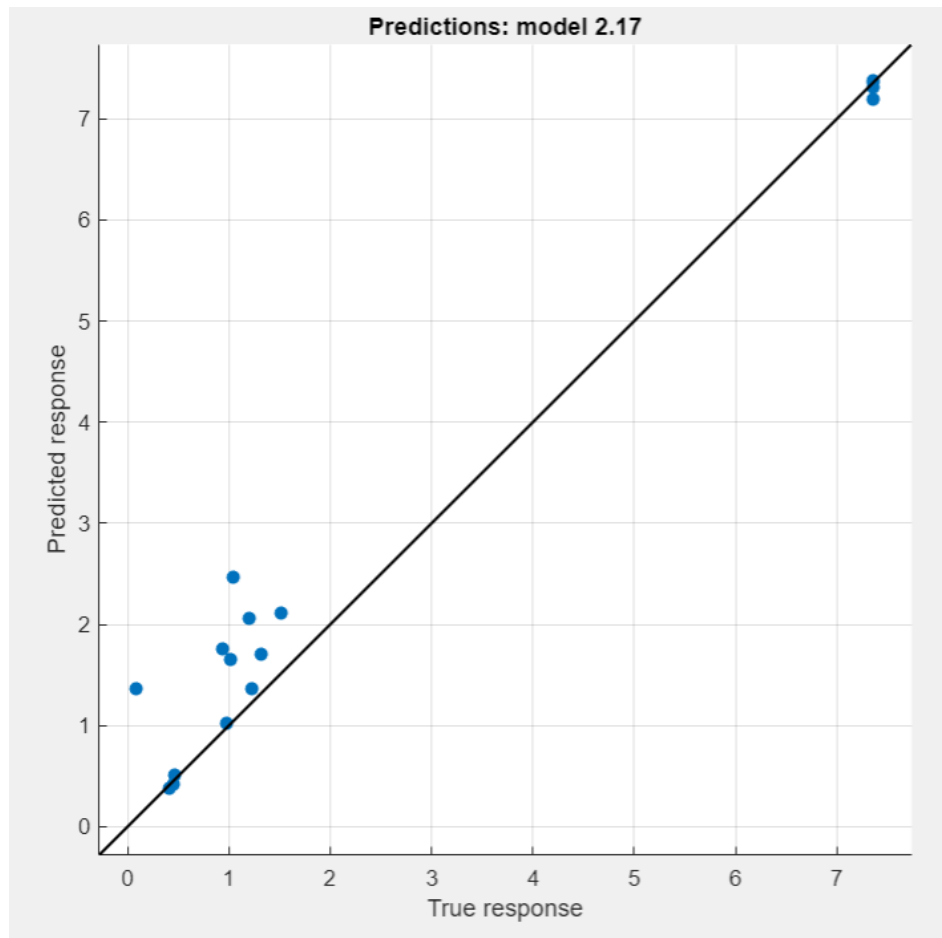
Η γραφική παράσταση για την σχέση προβλεπόμενων αποκρίσεων-πραγματικών παρουσιάζεται παρακάτω στην Γραφική Παράσταση 5 .



Γραφική Παράσταση 5 Πρόβλεψη τιμών, Var8, 100 αρχεία

Εδώ παρατηρείται καλύτερη κατανομή γύρω από την γραμμή παλινδρόμησης. Εκτός από κάποιες τιμές που ξεφεύγουν επειδή δεν υπάρχουν αρκετές παρατηρήσεις.

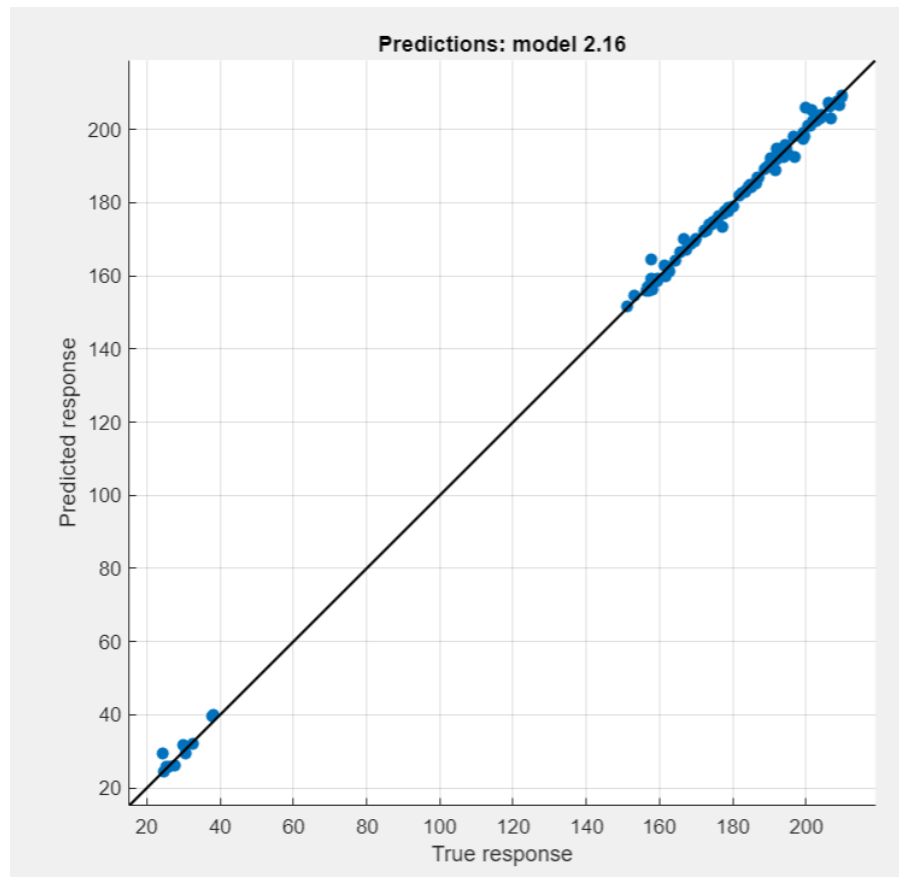
Στη συνέχεια εισαγάγαμε το αρχείο δοκιμής με τα 16 προκαθορισμένα σενάρια και παρατηρήσαμε ότι οι πιο πολλές τιμές πέφτουν ακριβώς πάνω στην γραμμή παλινδρόμησης ή κοντά σε εκείνη εκτός από κάποιες τιμές, όπως στην Γραφική Παράσταση 6. Αυτό σημαίνει ότι το μοντέλο εκπαιδεύεται σωστά και ακόμα πιο καλά όταν δίνονται περισσότερες τιμές εκπαίδευσης. Η απόκλιση μέσης τετραγωνικής τιμής έγινε: 0.61367 [21]



Γραφική Παράσταση 6 Πρόβλεψη τιμών, 16 σενάρια, Var8, 100 αρχεία

Μεταβλητή 9 (Var 9, πίεση)

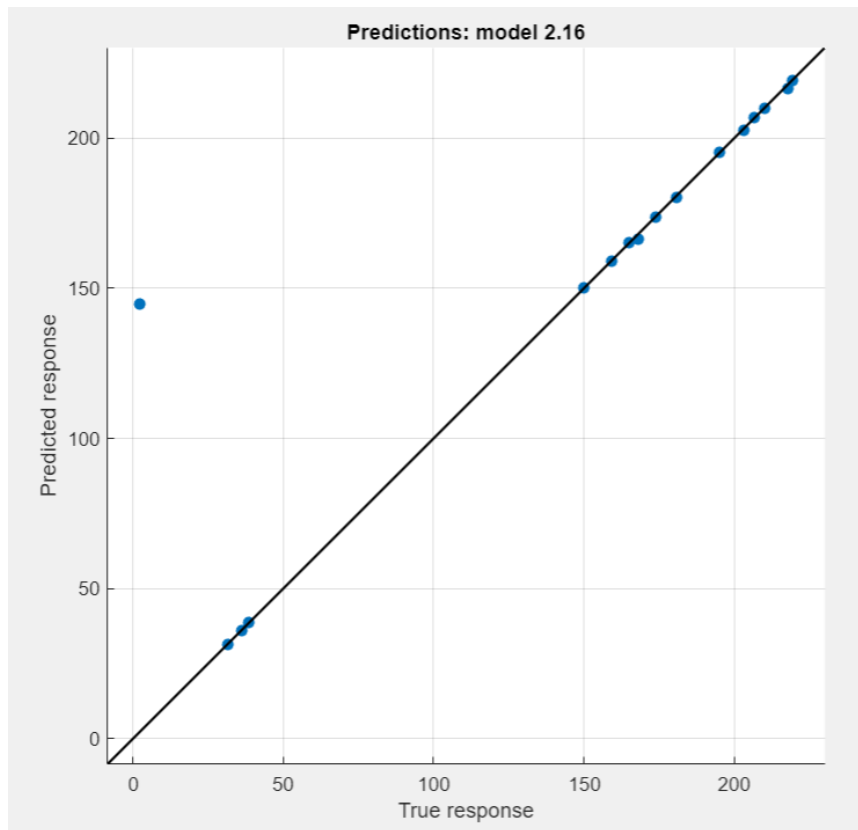
Για την μεταβλητή Var 9, που αντιπροσωπεύει την πίεση έχουμε καλύτερα αποτελέσματα, όπως φαίνονται στην Γραφική Παράσταση 7.



Γραφική Παράσταση 7 Πρόβλεψη τιμών, Var9, 100 αρχεία

Παρατηρείται ότι οι προβλεπόμενες αποκρίσεις είναι σχεδόν ίδιες με τις πραγματικές και αυτό φαίνεται επειδή σχεδόν όλες οι τιμές ταυτίζονται με την γραμμή παλινδρόμησης. Το μοντέλο που επιλέχθηκε για την συγκεκριμένη μεταβλητή είναι το μοντέλο παλινδρόμησης Gaussian Process, Squared Exponential GPR με RMSE 1,6354.

Μετά την επιλογή, έγινε αξιολόγηση του μοντέλου με το αρχείο δοκιμής και είχαμε τα παρακάτω αποτελέσματα στην Γραφική Παράσταση 8.

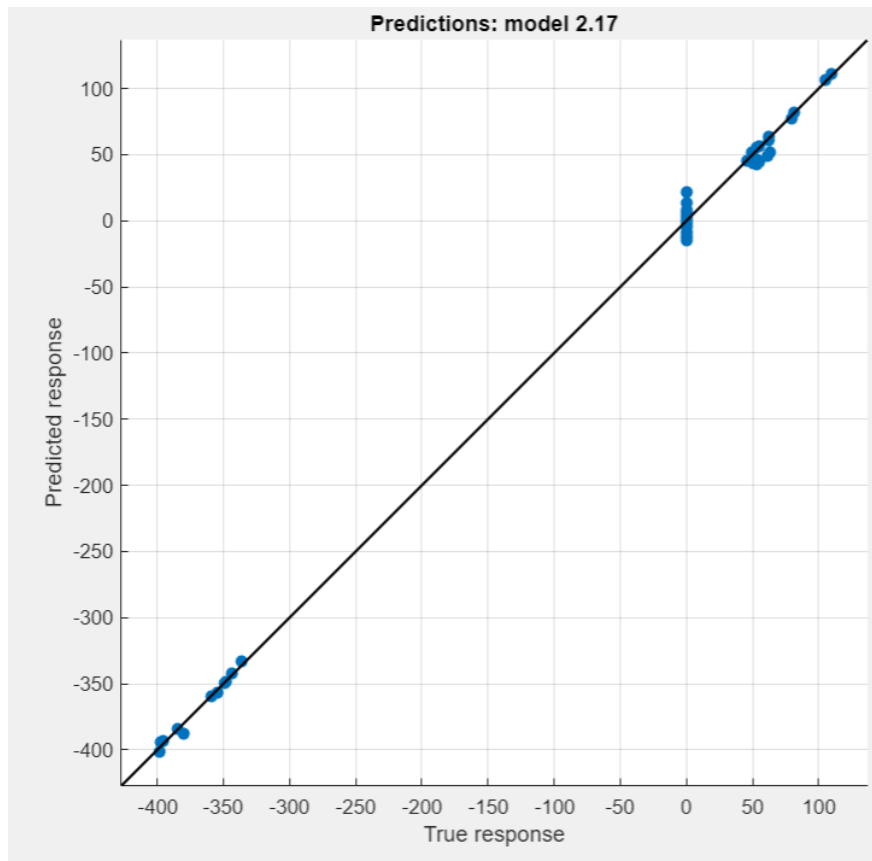


Γραφική Παράσταση 8 Πρόβλεψη για τα 16 σενάρια, Var9, 100 αρχεία

Παρατηρούμε ότι το μοντέλο εκπαιδεύτηκε σωστά και μάλιστα με τις τιμές που του δόθηκαν από το αρχείο δοκιμών, έχουμε μια σχεδόν τέλεια πρόβλεψη. [21]

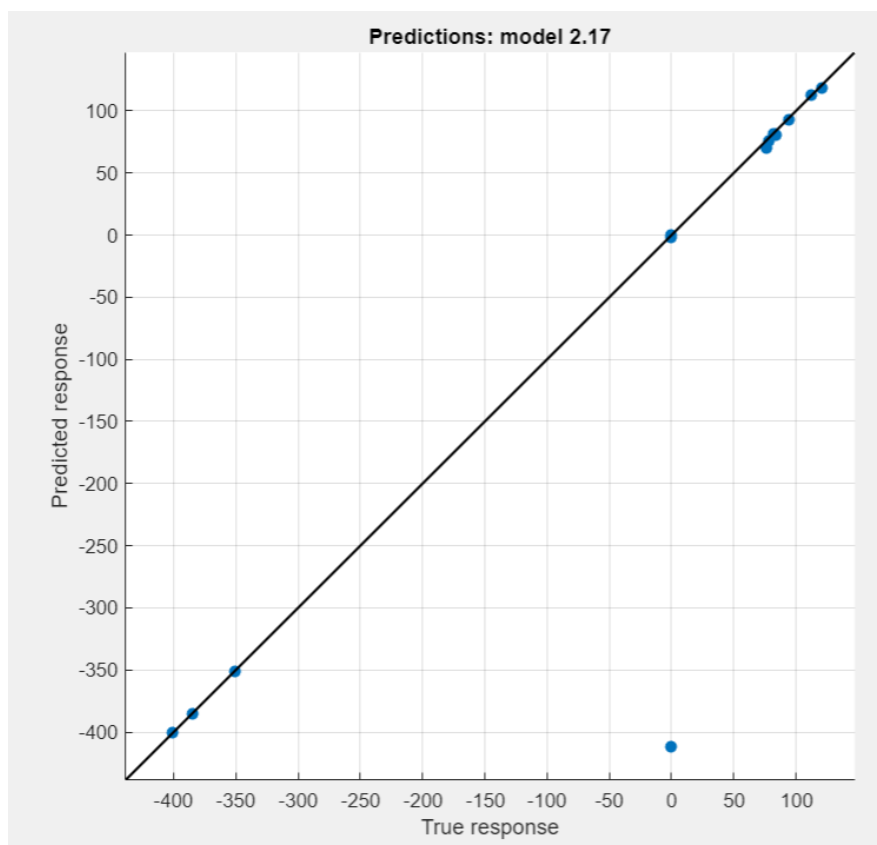
Μεταβλητή 10 (Var10, παροχή)

Ακολουθώντας την ίδια διαδικασία για την μεταβλητή της ροής (Var 10) αποδείχθηκε ότι το καλύτερο μοντέλο είναι το μοντέλο παλινδρόμησης Gaussian Process , Matern 5/2 με RMSE 5.1705 όπως φαίνεται στην Γραφική Παράσταση 9 .



Γραφική Παράσταση 9 Πρόβλεψη τιμών, Var10, 100 αρχεία

Για την συγκεκριμένη μεταβλητή έχουμε ακόμα καλύτερη πρόβλεψη διότι σχεδόν όλες οι τιμές είναι πάνω στη γραμμή παλινδρόμησης. Μετά την επιλογή του μοντέλου, αξιολογήθηκε με το αρχείο δοκιμών και είχαμε την Γραφική Παράσταση 10 με RMSE 103.05.



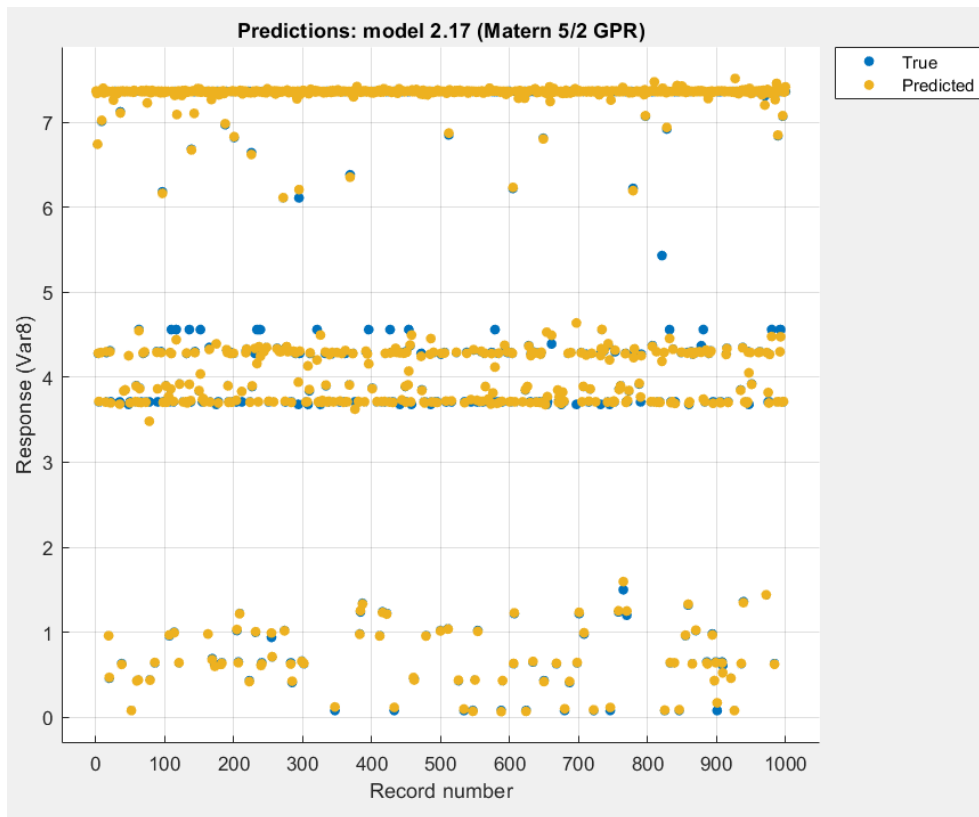
Γραφική Παράσταση 10 Πρόβλεψη τιμών, 16 σενάρια, Var10, 100 αρχεία

Η απόδοση του μοντέλου είναι πολύ καλή διότι βλέπουμε ότι οι τιμές που του δώσαμε για να δούμε κατά πόσο καλά εκπαιδεύτηκε, πέφτουν ακριβώς πάνω στην γραμμή παλινδρόμησης εκτός μίας τιμής γιατί δεν υπήρχαν αρκετές τιμές για αυτή την τιμή κατά την εκπαίδευση του μοντέλου. [21]

4.3.3 Επιλογή μοντέλου και επικύρωση του για 1000 αρχεία

Μεταβλητή 8 (Var 8, ενέργεια)

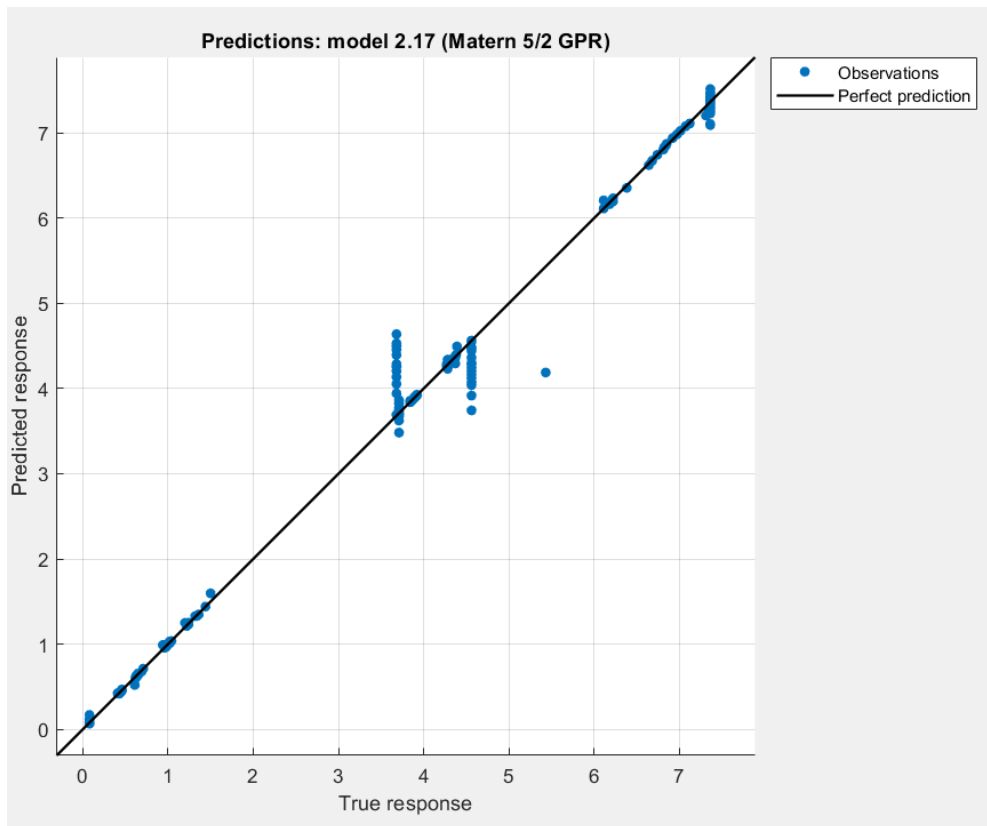
Η αλλαγή που έγινε ήταν η αύξηση των αρχείων στα 1000, άρα έχουμε πιο πολλά δεδομένα. Μετά την εκπαίδευση, το καλύτερο μοντέλο για την μεταβλητή της ροής είναι το μοντέλο παλινδρόμησης Gaussian Process, Matern 5/2 με RMSE 0.10296. Η γραφική παράσταση των αποκρίσεων σε σχέση με τον αριθμό των δεδομένων φαίνεται στην Γραφική Παράσταση 11.



Γραφική Παράσταση 11 Απόκριση τιμών, Var8, 1000 αρχεία

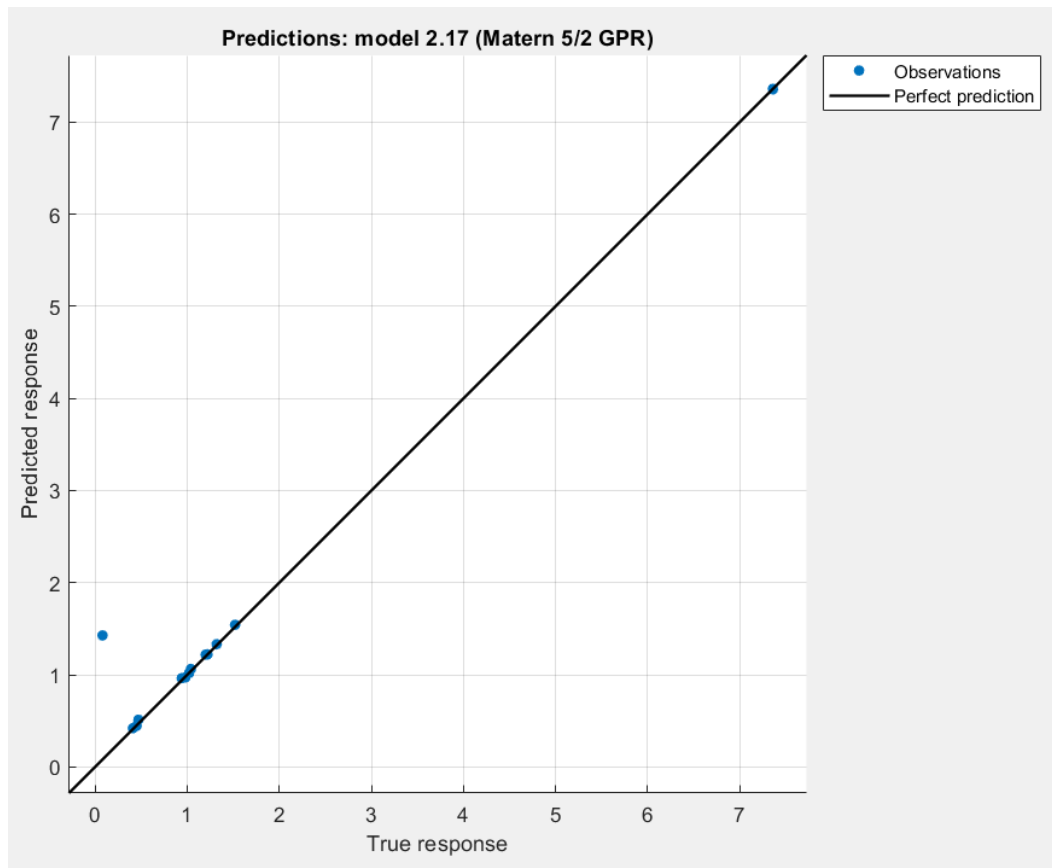
Οι κίτρινες τιμές είναι οι προβλεπόμενες και οι μπλε οι πραγματικές και βλέπουμε ότι ταυτίζονται ως επί το πλείστον.

Παρακάτω παρουσιάζεται η Γραφική Παράσταση 12 των προβλέψεων του μοντέλου.



Γραφική Παράσταση 12 Πρόβλεψη τιμών, Var8, 1000 αρχεία

Για την επικύρωση του μοντέλου χρησιμοποιήσαμε το αρχείο με τα 16 σενάρια και παρατηρήσαμε ότι το μοντέλο εκπαιδεύτηκε σωστά και μας δίνει αξιοπρεπείς προβλέψεις, όπως φαίνεται στην Γραφική Παράσταση 13. [21]



Γραφική Παράσταση 13 Πρόβλεψη τιμών, 16 σενάρια, Var18, 1000 αρχεία

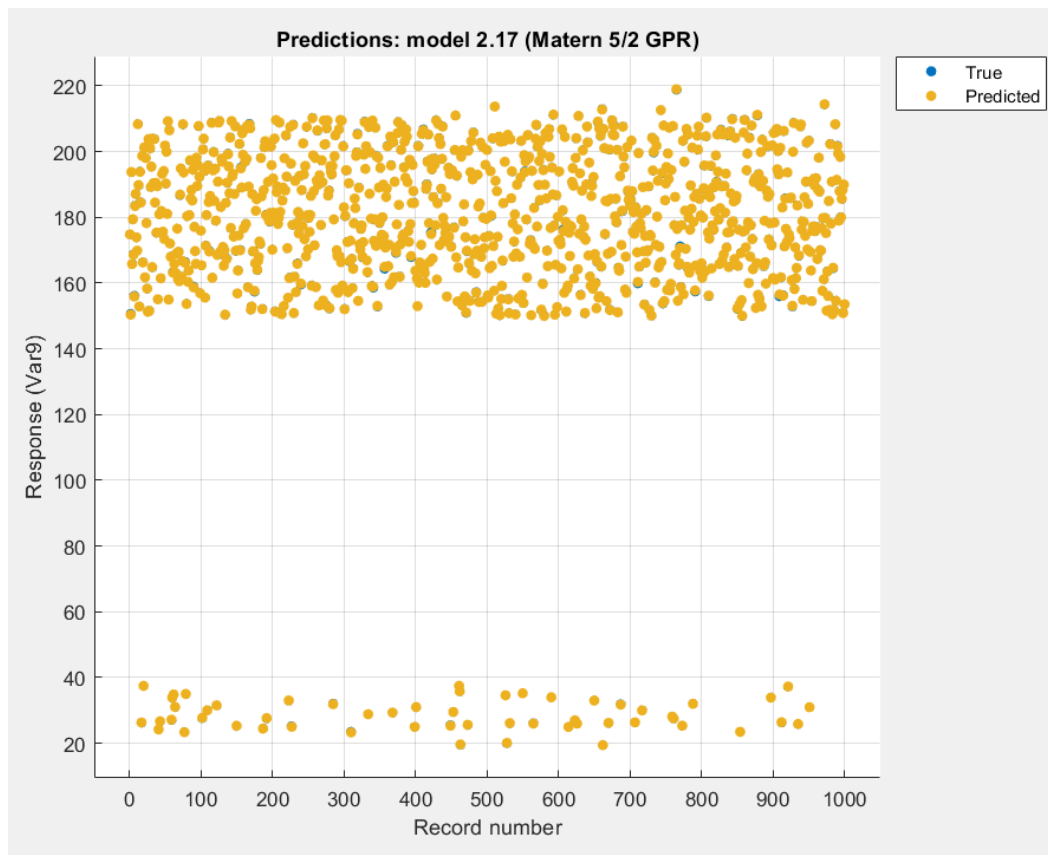
Μεταβλητή 9 (Var 9, πίεση)

Για την μεταβλητή της πίεσης καταλήξαμε ότι το κατάλληλο μοντέλο είναι το μοντέλο παλινδρόμησης Gaussian Process, Matern 5/2 με RMSE 0.059154, όπως φαίνεται στην Εικόνα 28.

☆	2.14 Ensemble	RMSE (Validation): 7.6489
	Last change: Boosted Trees	9/9 features
☆	2.15 Ensemble	RMSE (Validation): 5.4636
	Last change: Bagged Trees	9/9 features
☆	2.16 Gaussian P...	RMSE (Validation): 0.065367
	Last change: Squared Exponential GPR	9/9 features
☆	2.17 Gaussian P...	RMSE (Validation): 0.059154
	Last change: Matern 5/2 GPR	9/9 features

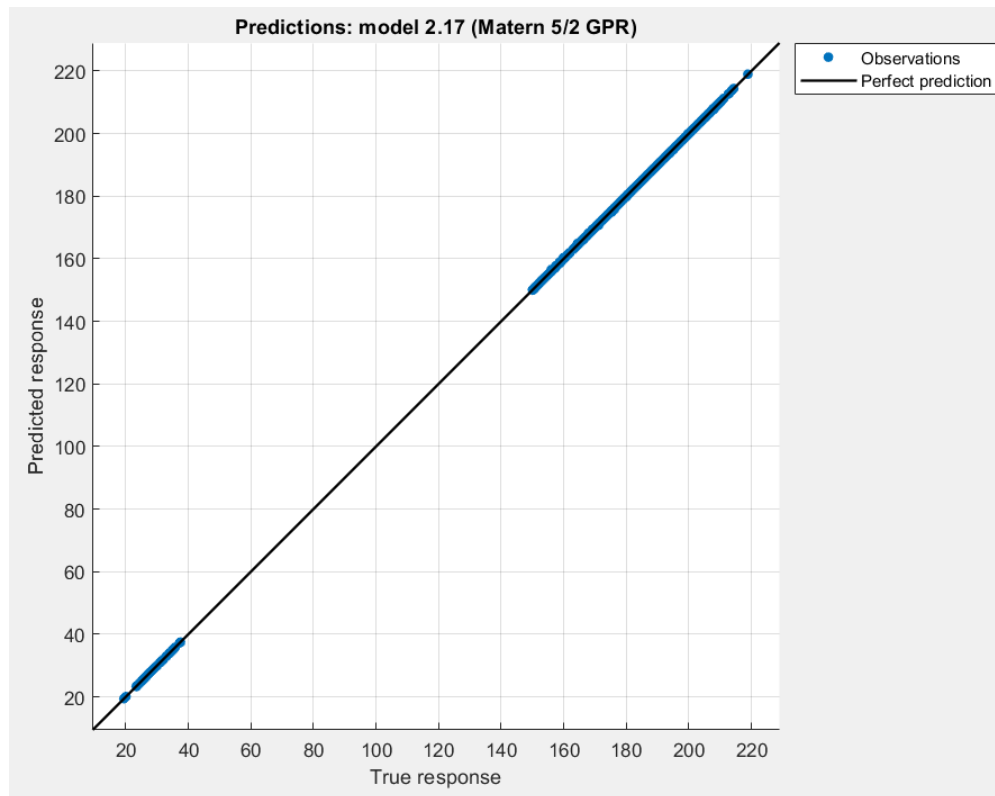
Εικόνα 28 Βέλτιστο μοντέλο για την μεταβλητή 9, 1000 αρχεία

Παρακάτω βλέπουμε την Γραφική Παράσταση 14 των αποκρίσεων σε σχέση με τον αριθμό των δεδομένων και παρατηρούμε ότι οι τιμές συμπίπτουν ακριβώς μετά την εκπαίδευση.



Γραφική Παράσταση 14 Απόκριση τιμών, Var9, 1000 αρχεία

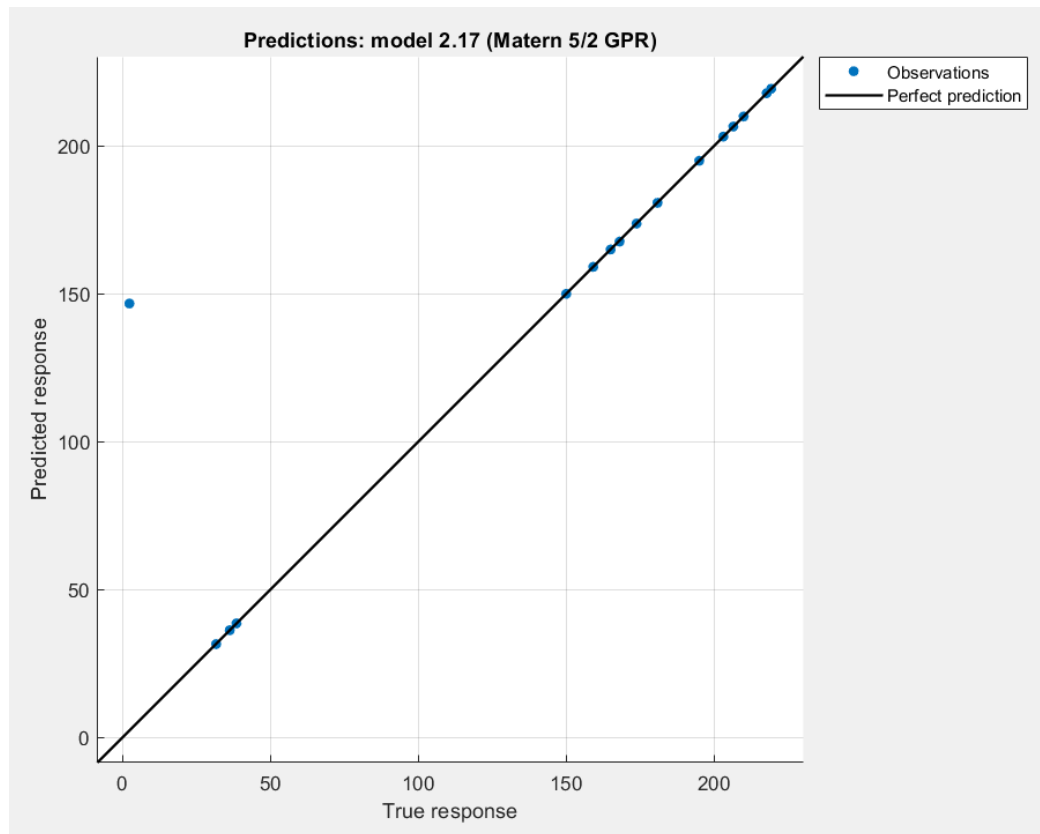
Στη συνέχεια απεικονίζεται η Γραφική Παράσταση 15 των προβλέψεων που έγιναν.



Γραφική Παράσταση 15 Πρόβλεψη τιμών, Var9, 1000 αρχεία

Παρατηρείται ότι το μοντέλο για την συγκεκριμένη μεταβλητή έκανε μια τέλεια πρόβλεψη, καθώς οι προβλεπόμενες αποκρίσεις είναι ίσες με τις πραγματικές αποκρίσεις, δηλαδή όλα τα σημεία είναι διασκορπισμένα πάνω στη γραμμή παλινδρόμησης.

Στη συνέχεια παρουσιάζεται η απόδοση του μοντέλου έχοντας εισάγει στο εκπαιδευμένο μοντέλο το αρχείο με τις 16 προκαθορισμένες τιμές και τα αποτελέσματά τους, όπως φαίνεται στην Γραφική Παράσταση 16.



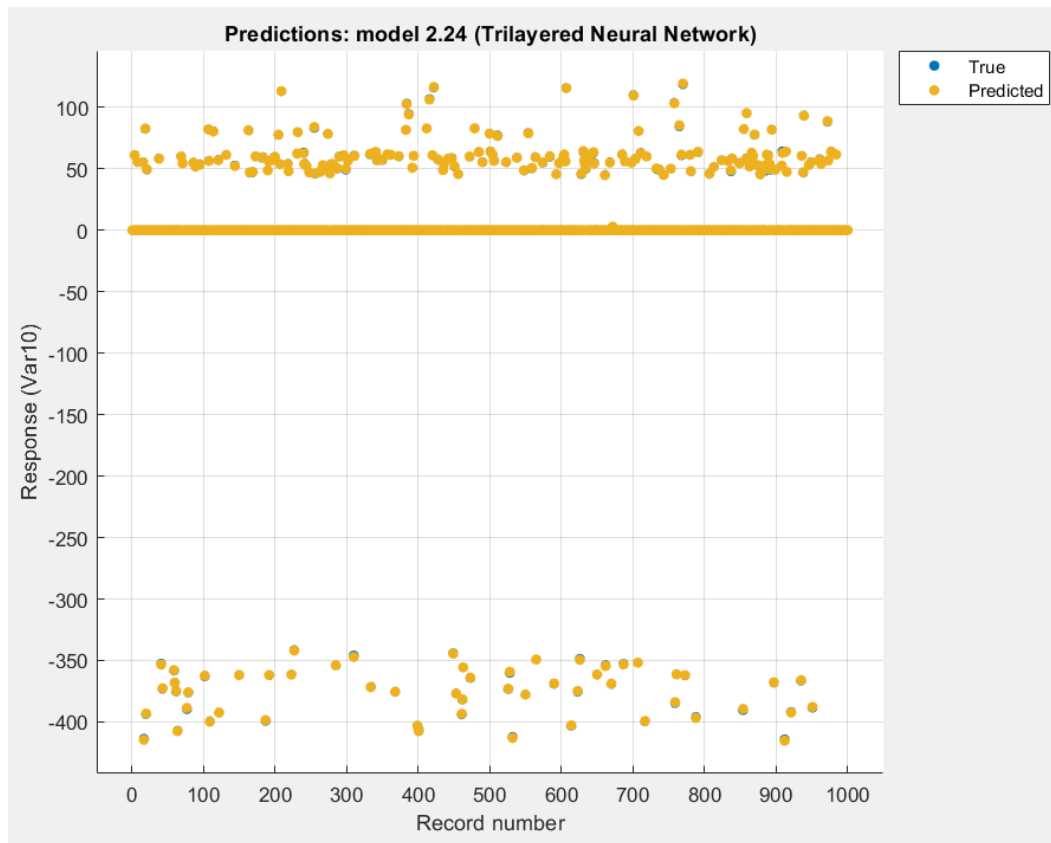
Γραφική Παράσταση 16 Πρόβλεψη τιμών, 16 σενάρια, Var9, 1000 αρχεία

Παρατηρούμε ότι οι τιμές (παρατηρήσεις) είναι ακριβώς πάνω στην γραμμή παλινδρόμησης που υποδεικνύει την τέλεια πρόβλεψη. Άρα το μοντέλο εκπαιδεύτηκε σωστά. [21]

Μεταβλητή 10 (Var 10, παροχή)

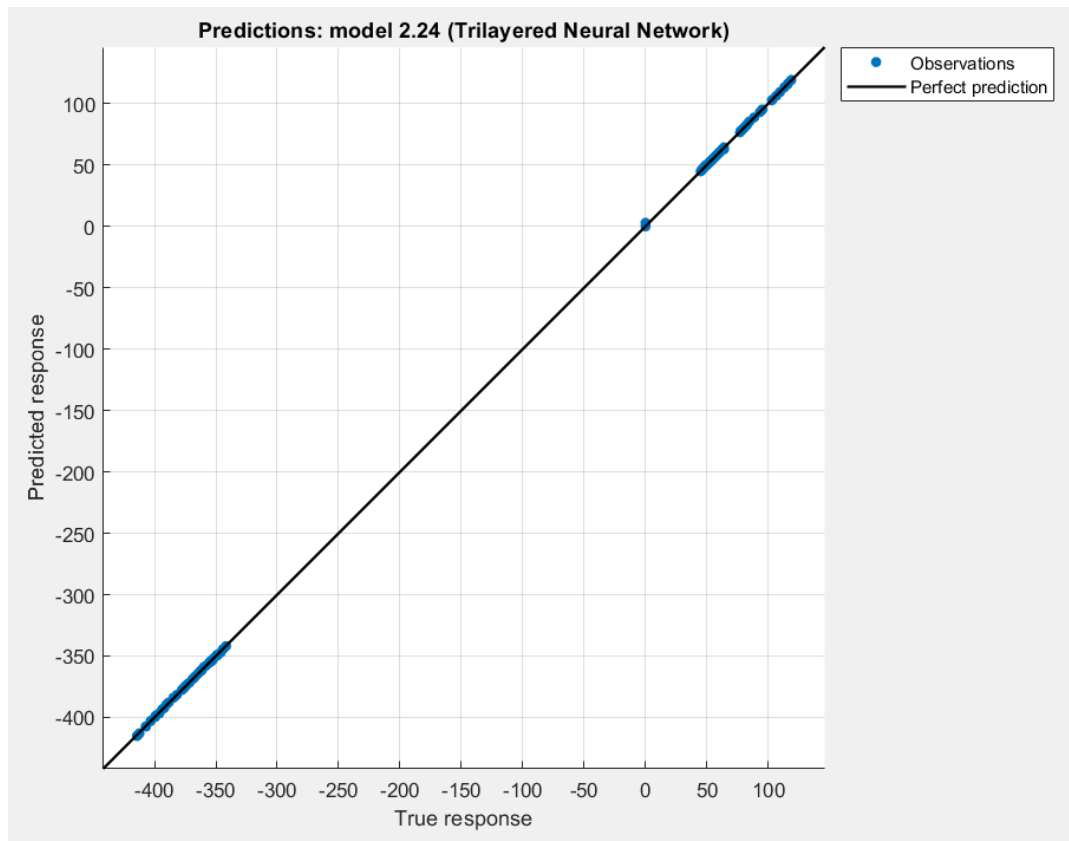
Για την μεταβλητή της παροχής ύστερα από την εκπαίδευση των μοντέλων καταλήξαμε ότι το καλύτερο μοντέλο για αυτή την μεταβλητή είναι το νευρωνικό δίκτυο (Neural Network) , Trilayered Neural Network με RMSE 0.15824.

Η Γραφική Παράσταση 17 που φαίνεται παρακάτω δείχνει ακριβής ταύτιση των προβλεπόμενων αποκρίσεων με των πραγματικών.



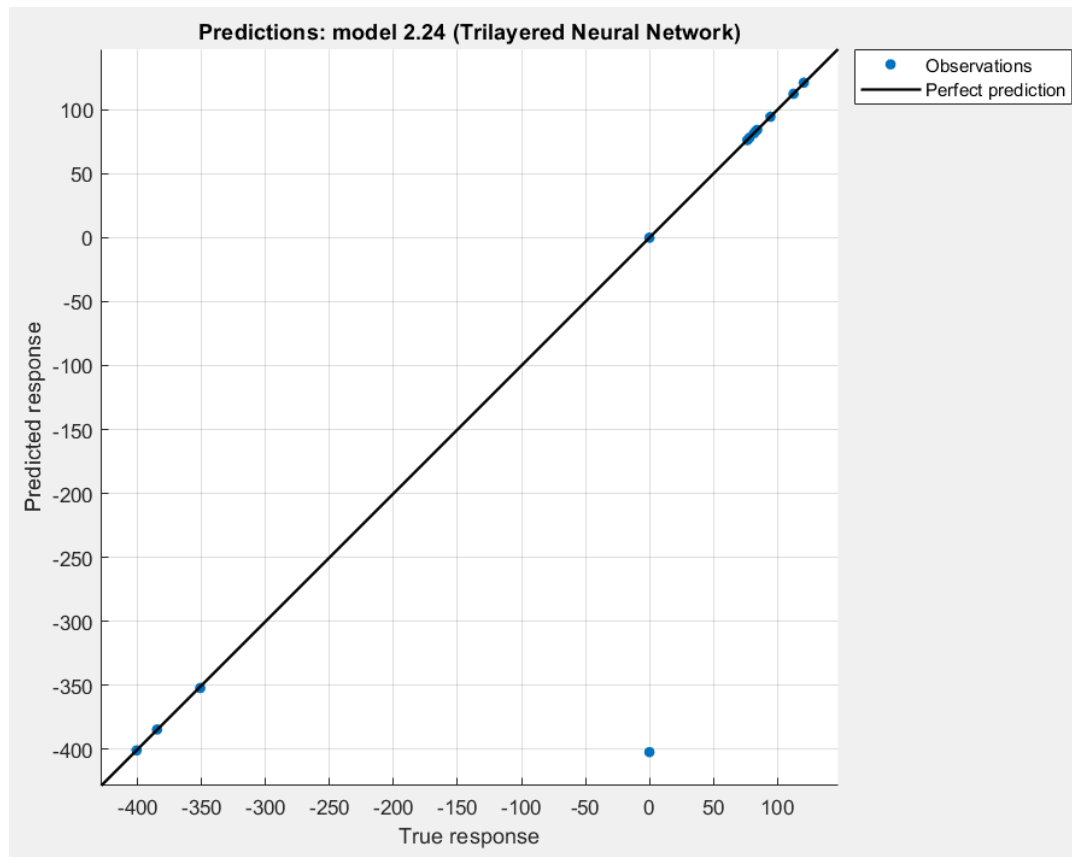
Γραφική Παράσταση 17 Απόκριση τιμών, Var10, 1000 αρχεία

Στην Γραφική Παράσταση 18 βλέπουμε πως και εδώ για την μεταβλητή της παροχής έχουμε μία πολύ καλή πρόβλεψη.



Γραφική Παράσταση 18 Πρόβλεψη τιμών, Var10, 1000 αρχεία

Η παραπάνω πρόβλεψη επιβεβαιώνεται με την Γραφική Παράσταση 19 στην οποία αφού δώσαμε το αρχείο με τα προκαθορισμένα σενάρια, είδαμε ότι το μοντέλο μπορεί και προβλέπει σωστά. [21]

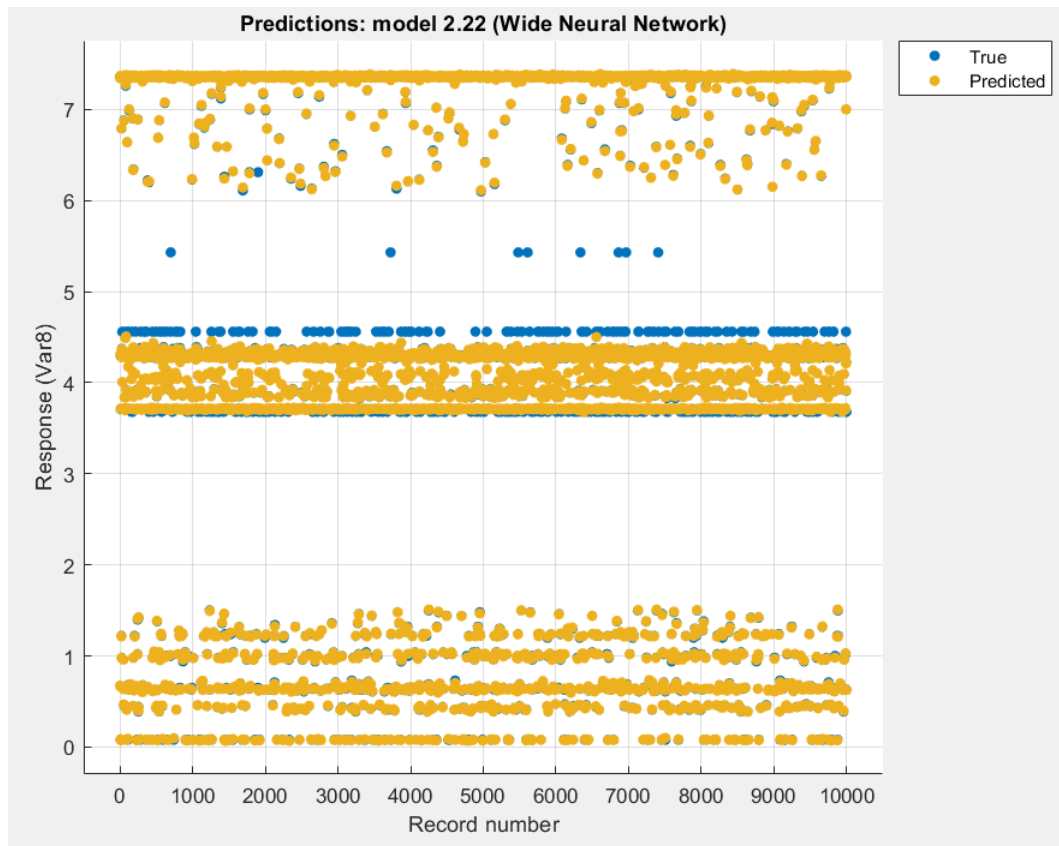


Γραφική Παράσταση 19 Πρόβλεψη τιμών, Var10, 16 σενάρια, 1000 αρχεία

4.3.4 Επιλογή μοντέλου και επικύρωση του για 10000 αρχεία

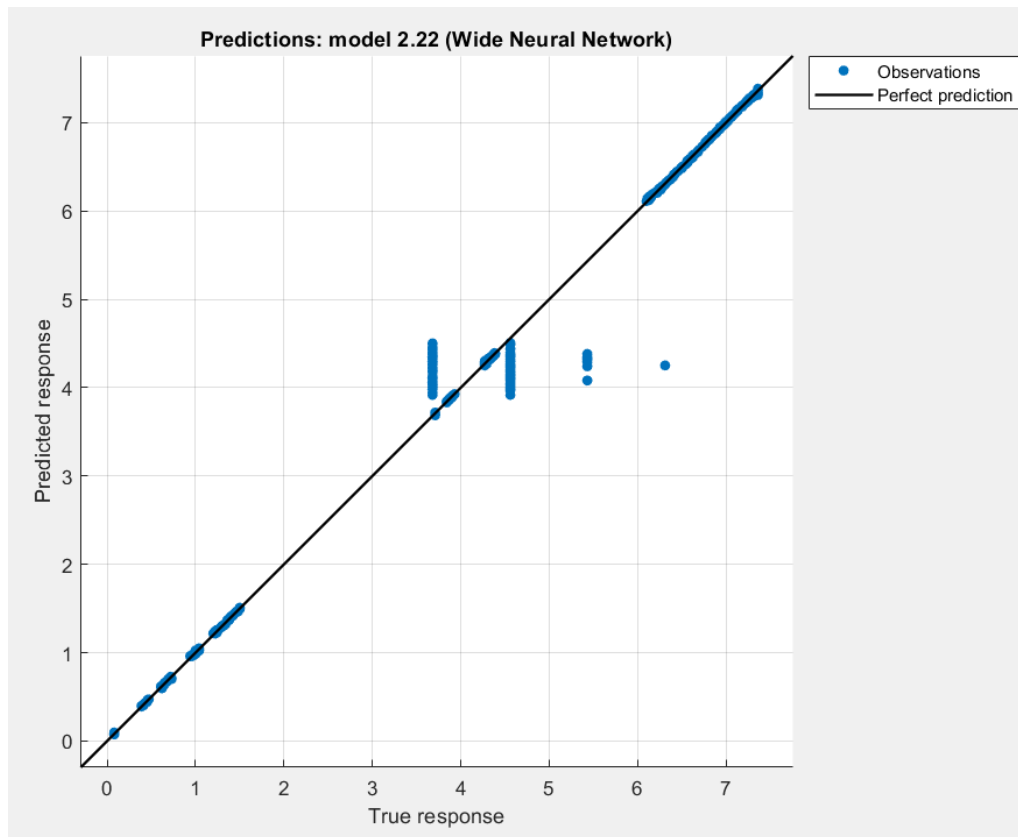
Μεταβλητή 8 (Var 8, ενέργεια)

Αφού ακολουθήθηκε η ίδια διαδικασία και αφού συλλέξαμε τα δεδομένα και δημιουργήσαμε τα αρχεία εκπαίδευσης (train dataset), εκπαιδεύσαμε τα μοντέλα ως προς την μεταβλητή της ενέργειας και παρατηρήσαμε ότι το καλύτερο μοντέλο είναι το μοντέλο του νευρωνικού δικτύου (Neural Network), wide neural network με RMSE 0.090764. Παρακάτω απεικονίζεται η Γραφική Παράσταση 20, οι προβλεπόμενες τιμές σε σχέση με τις πραγματικές.



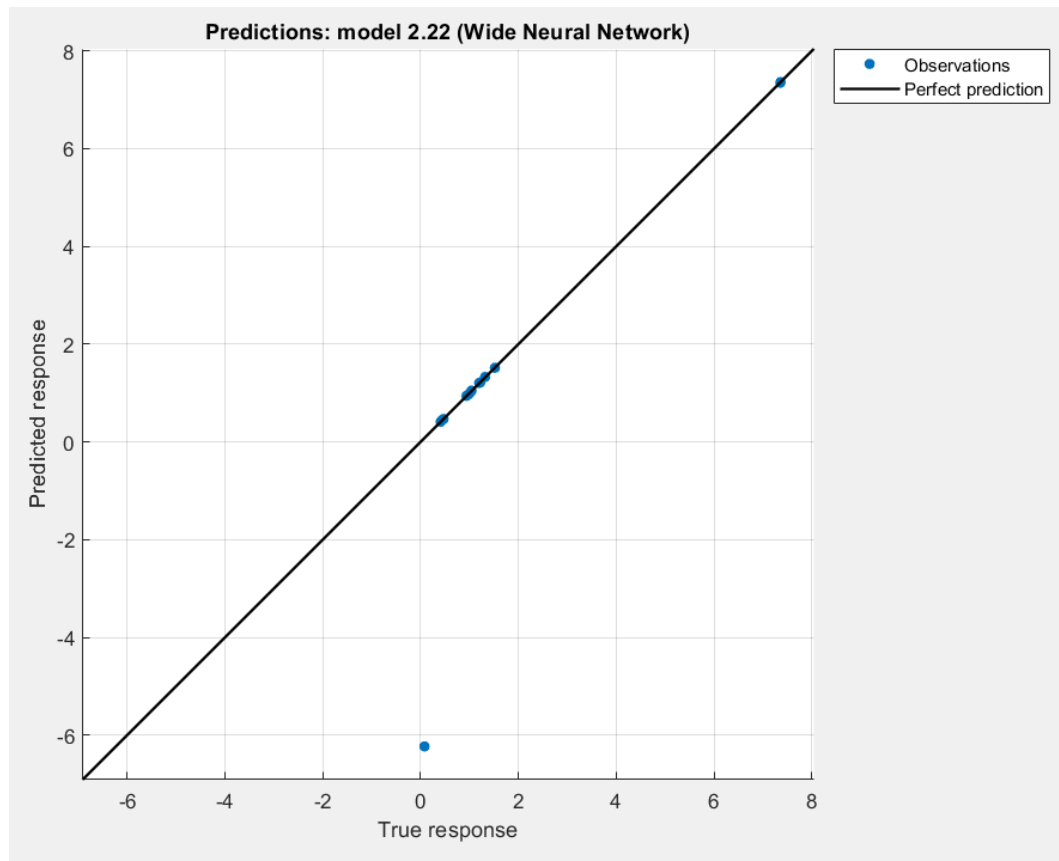
Γραφική Παράσταση 20 Απόκριση τιμών, Var8, 10000 αρχεία

Η Γραφική Παράσταση 21 δείχνει την πρόβλεψη που έκανε το μοντέλο και παρατηρούμε μία ικανοποιητική διασπορά των σημείων πάνω στην γραμμή παλινδρόμησης.



Γραφική Παράσταση 21 Πρόβλεψη τιμών, Var8, 10000 αρχεία

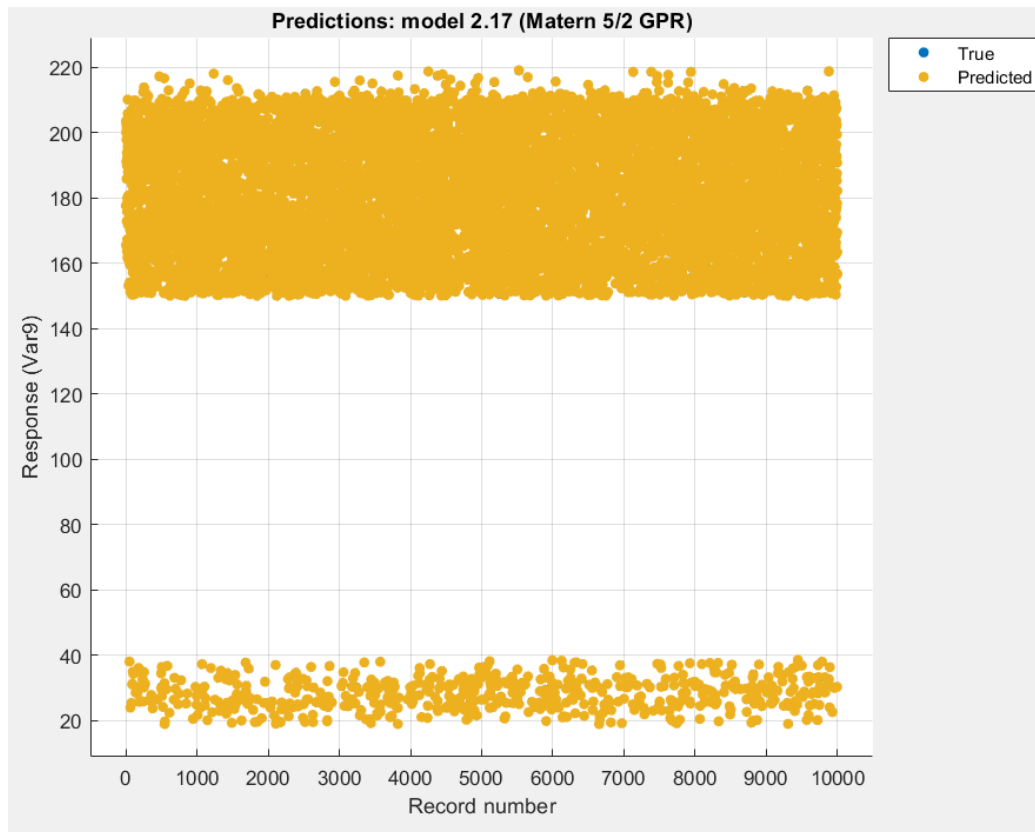
Η επικύρωση του μοντέλου έγινε με την εισαγωγή του αρχείου δοκιμών (test dataset) και παρατηρούμε ότι το μοντέλο εκπαιδεύτηκε σωστά και είναι ικανό για έγκυρες προβλέψεις, το οποίο φαίνεται στην Γραφική Παράσταση 22. [21]



Γραφική Παράσταση 22 Πρόβλεψη τιμών 16 σενάρια, Var8, 10000 αρχεία

Μεταβλητή 9 (Var 9, πίεση)

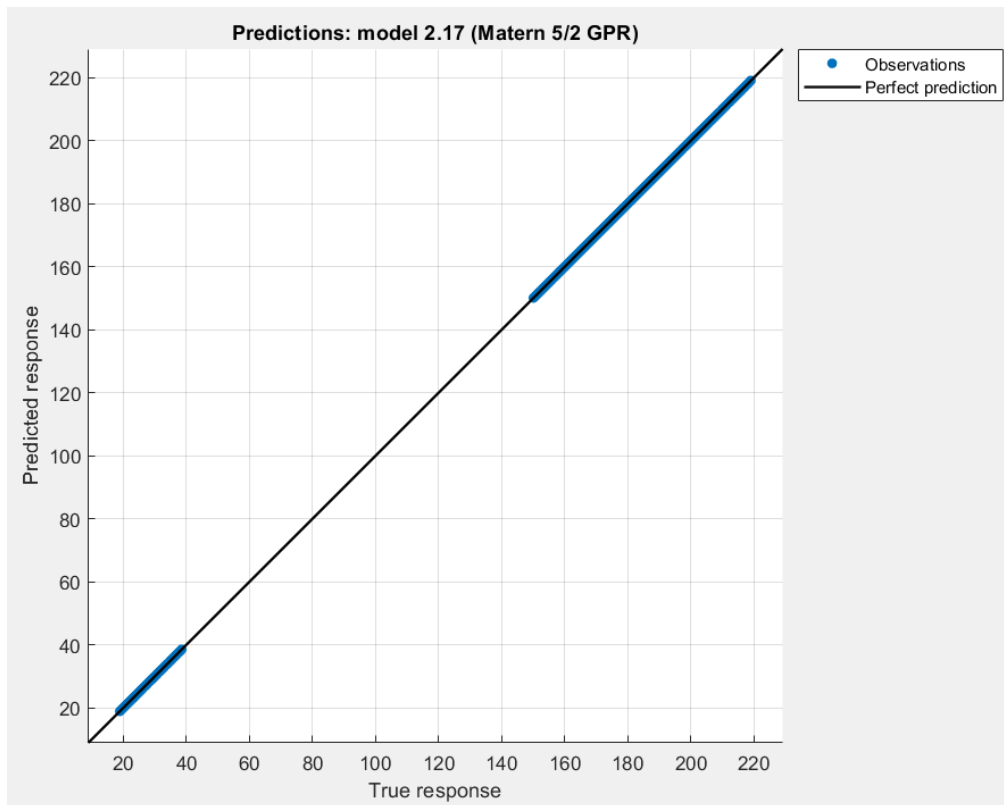
Για την μεταβλητή 9 που αντιπροσωπεύει την πίεση, καταλήξαμε ότι το καλύτερο μοντέλο μηχανικής μάθησης είναι το μοντέλο παλινδρόμησης Gaussian Process , Matern 5/2 (Gaussian Process Regression) με απόκλιση μέσης τετραγωνικής τιμής (RMSE) 0.011882. Στην Γραφική Παράσταση 23 παρουσιάζεται η ταύτιση των προβλεπόμενων και πραγματικών τιμών.



Γραφική Παράσταση 23 Απόκρισης τιμών, Var9, 10000 αρχεία

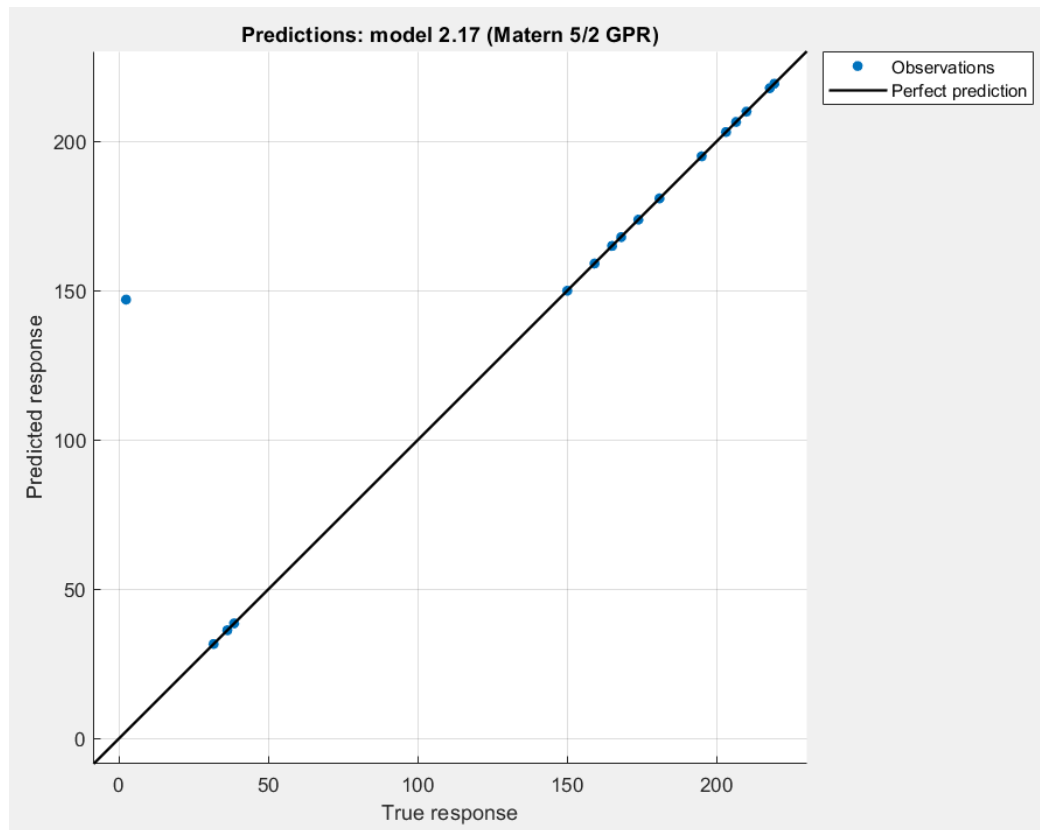
Παρατηρούμε ότι οι προβλεπόμενες τιμές ταυτίζονται με τις πραγματικές κατά την εκπαίδευση του μοντέλου.

Στην Γραφική Παράσταση 24 παρουσιάζεται η πρόβλεψη που έκανε το μοντέλο, παρατηρώντας ότι και εδώ έχουμε μια αψεγάδιαστη πρόβλεψη, καθώς οι προβλεπόμενες αποκρίσεις είναι ίσες με τις πραγματικές και ταυτίζονται στην γραμμή παλινδρόμησης.



Γραφική Παράσταση 24 Πρόβλεψη τιμών, Var9, 10000 αρχεία

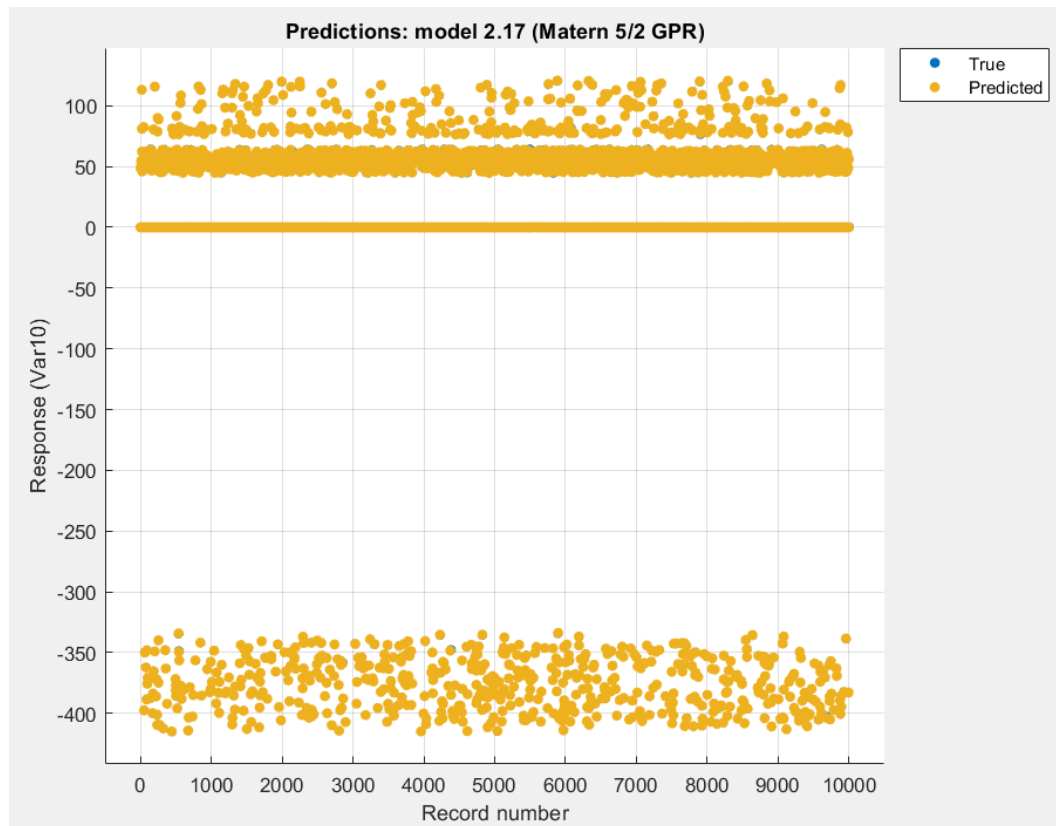
Στην Γραφική Παράσταση 25 παρατηρούμε ότι το μοντέλο η απόδοση του μοντέλου είναι πολύ καλή και εκπαιδεύτηκε σωστά καθώς μόνο μία τιμή από το αρχείο με τα 16 καθορισμένα σενάρια δεν βρίσκεται πάνω στην γραμμή παλινδρόμησης. [21]



Γραφική Παράσταση 25 Πρόβλεψη τιμών, 16 σενάρια, Var9, 10000 αρχεία

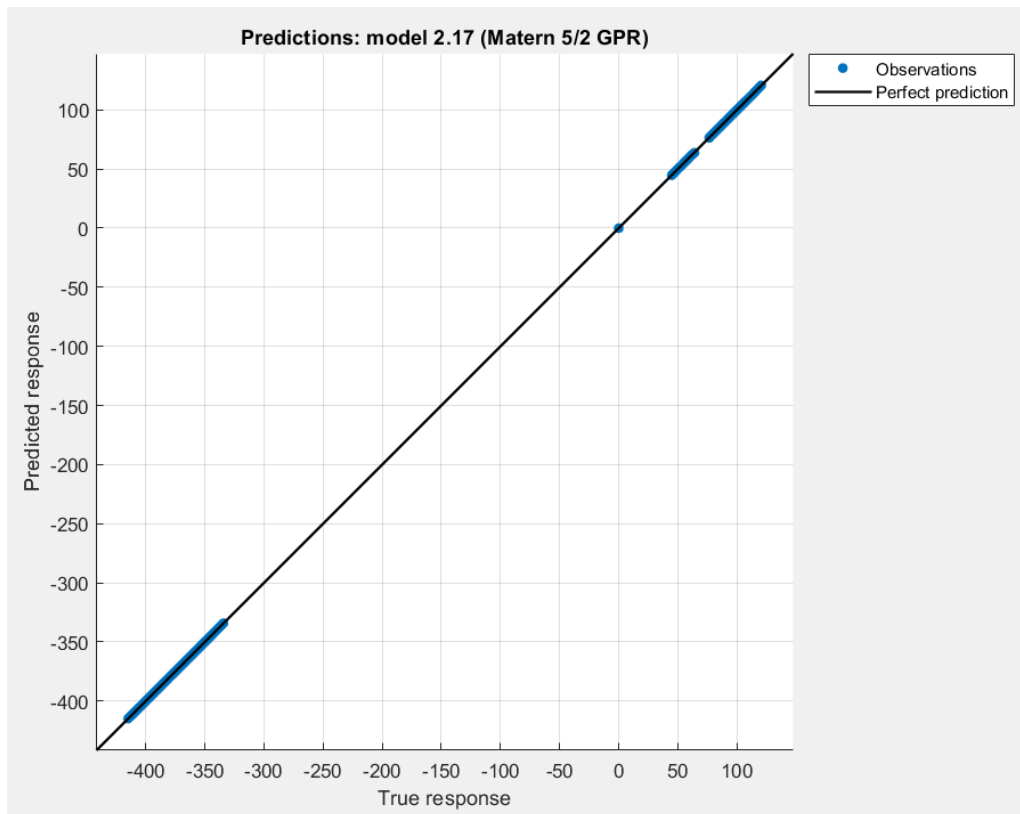
Μεταβλητή 10 (Var 10, παροχή)

Για την μεταβλητή 10 που αντιπροσωπεύει την παροχή, ύστερα από την εκπαίδευση των μοντέλων καταλήξαμε ότι το καλύτερο μοντέλο είναι το μοντέλο παλινδρόμησης Gaussian Process, Matern 5/2 (Gaussian Process Regression) με RMSE 0.029324. Στην Γραφική Παράσταση 26 απεικονίζεται οι αποκρίσεις των τιμών, προβλεπόμενων και πραγματικών, βλέποντας πως οι πραγματικές τιμές επικαλύπτονται από τις προβλεπόμενες.



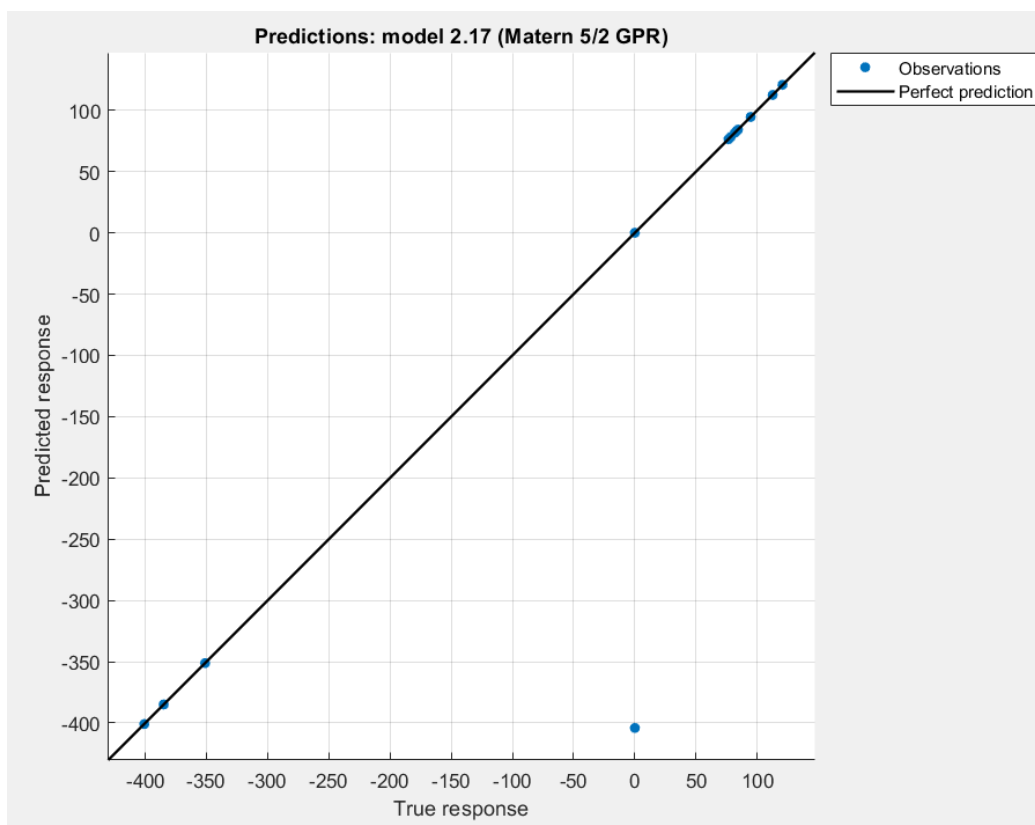
Γραφική Παράσταση 26 Απόκριση τιμών, Var10, 10000 αρχεία

Στη συνέχεια απεικονίζεται στην Γραφική Παράσταση 27 κατά πόσο καλά έγινε η πρόβλεψη με το συγκεκριμένο μοντέλο. Παρατηρούμε ότι το μοντέλο έκανε μία καλή πρόβλεψη με τις τιμές που του δόθηκαν καθώς όλες οι τιμές είναι πάνω στην γραμμή παλινδρόμησης.



Γραφική Παράσταση 27 Πρόβλεψη τιμών, Var10, 10000 αρχεία

Τέλος στην Γραφική Παράσταση 28 βλέπουμε ότι το μοντέλο που επιλέχθηκε εκπαιδεύτηκε σωστά και αυτό φαίνεται από το αρχείο με τις προκαθορισμένες τιμές, καθώς το μοντέλο αναγνωρίζει τι του δόθηκε σαν είσοδος και έκανε σωστές προβλέψεις. [21]



Γραφική Παράσταση 28 Πρόβλεψη τιμών, 16 σενάρια, Var10, 10000 αρχεία

Εν κατακλείδι, παρατηρούμε ότι τα μοντέλα εκπαιδεύτηκαν σωστά και αυτό φαίνεται από τις τελευταίες γραφικές παραστάσεις κάθε μεταβλητής, οι οποίες δείχνουν κατά πόσο καλά αναγνωρίζει το κάθε μοντέλο τις μεταβλητές που του δίνονται σαν είσοδος και τις προβλέψεις που κάνει μετά την εκπαίδευση του. Όσο περισσότερα δεδομένα έχουμε τόσο καλύτερη εκπαίδευση έχει το μοντέλο. Οι τιμές που δεν ήταν κοντά ή πάνω στην γραμμή παλινδρόμησης ήταν τιμές για τις οποίες δεν υπήρχαν αρκετές πληροφορίες έτσι ώστε το μοντέλο να εκπαιδευτεί πάνω σε αυτές.

Παρακάτω παρουσιάζονται συγκετρωτικά οι πίνακες που δείχνουν τα μοντέλα που επλέχθηκαν ως βέλτιστα μετά την εκπαίδευση, οι τιμές εκπαίδευσης και επικύρωσης RMSE για 10, 100, 1000, 10000 αρχεία.

Πίνακας 4 Τιμές RMSE για την κατανάλωση ενέργειας (μεταβλητή 8)

	Πλήθος Σεναρίων			
	10	100	1000	10000
Επιλεχθέν MMM	Kernel, Least Squares Regression Kernel	Gaussian Process Regression, Matern 5/2	GPR, Matern 5/2	Neural Network, Wide NN
RMSE (Εκπαίδευσης)	1.3394	0.84796	0.10296	0.090764
RMSE (Επικύρωσης)	4.5891	0.61367	0.099011	2.9419

Πίνακας 5 Τιμές RMSE για την πίεση (μεταβλητή 9)

	Πλήθος Σεναρίων			
	10	100	1000	10000
Επιλεχθέν MMM	Stepwise Linear Regression	GPR, Squared Exponential GPR	GPR, Matern 5/2	GPR, Matern 5/2
RMSE (Εκπαίδευσης)	0.0076989	1.6354	0.059154	0.011882
RMSE (Επικύρωσης)	77.177	34.458	0.070741	36.167

Πίνακας 6 Τιμές RMSE για την παροχή (μεταβλητή 10)

	Πλήθος Σεναρίων			
	10	100	1000	10000
Επιλεχθέν MMM	Πολλαπλά μοντέλα ως τα βέλτιστα	GPR, Matern 5/2	Neural Network, Trilayered NN	GPR, Matern 5/2
RMSE (Εκπαίδευσης)	0.0	5.1705	0.15824	0.029324
RMSE (Επικύρωσης)	176.93	103.05	97.974	100.99

Κεφάλαιο 5: Συμπεράσματα

Στην παρούσα διπλωματική εργασία μοντελοποιήθηκε και προσομοιώθηκε με επιτυχία ένα δίκτυο ύδρευσης με την βοήθεια του λογισμικού μοντελοποίησης και προσομοίωσης EPANET. Στη συνέχεια, δημιουργήθηκαν κώδικες στο περιβάλλον του MATLAB για την αυτοματοποιημένη δημιουργία πολλαπλών αρχείων καθώς και για την εκτέλεση τους και κατ' επέκταση στην συλλογή των αποτελεσμάτων (συνολική ενέργεια που καταναλώθηκε, πίεσης και παροχής). Στην συνέχεια, δημιουργήθηκαν δύο αρχεία, το ένα για την εκπαίδευση των μοντέλων μηχανικής μάθησης και το άλλο για την επικύρωση των αποτελεσμάτων του μοντέλου που επιλέχθηκε ως το καταλληλότερο.

Αφού εκπαιδεύτηκαν και επιλέχθηκαν τα βέλτιστα μοντέλα για 10, 100, 1000 και 10000 συνδυασμούς συνδεσμολογίας και των αποτελεσμάτων τους, καταλήξαμε στο συμπέρασμα ότι τα μοντέλα εκπαιδεύτηκαν σωστά και ελέγχθηκε η αποτελεσματικότητά τους με το αρχείο δοκιμής για την επικύρωσή τους. Επίσης, παρατηρήσαμε ότι τα μοντέλα με μικρό πλήθος τιμών (10, 100) δεν εκπαιδεύονται σωστά, έτσι όσο αυξάνεται ο αριθμός των δεδομένων τόσο καλύτερη εκπαίδευση υπάρχει. Ωστόσο, πρέπει να σημειωθεί ότι για μεγάλο πλήθος δεδομένων (10000) παρατηρήθηκε μεγάλη υπερδιόρθωση με αποτέλεσμα το RMSE κατά την επικύρωση να γίνεται μεγαλύτερο από αυτό του πλήθους 1000 τιμών.

Επιπρόσθετα, στις γραφικές παραστάσεις για την επικύρωση των αποτελεσμάτων παρατηρήθηκαν ορισμένες τιμές που υπήρχαν στο αρχείο με τα 16 προκαθορισμένα σενάρια, οι οποίες δεν ήταν κοντά στην γραμμή παλινδρόμησης. Αυτό συνέβη διότι δεν υπήρχαν αρκετές τιμές έτσι ώστε να εκπαιδευτεί το μοντέλο και να τις αναγνωρίζει.

Εν κατακλείδι, η διπλωματική εργασία εκτελέστηκε με επιτυχία και η προσέγγιση που ακολουθήθηκε δίδει ικανοποιητικά αποτελέσματα και πιθανώς μπορεί να αξιοποιηθεί σε μελλοντικές μελέτες και εφαρμογές ψηφιακών διδύμων για υδραυλικά δίκτυα.

Βιβλιογραφία

- [1] J.-R. Bermúdez, F.-R. López-Estrada, G. Besançon, G. Valencia-Palomo, L. Torres και H.-R. Hernández, «Modeling and Simulation of a Hydraulic Network for Leak Diagnosis,» *Mathematical and Computational Applications*, vol. 23, no. 4, p. 70, November 2018.
- [2] Gowda, L. G. Ramesh, Jagadeesha και C.J., «Simulation of Hydraulic Parameters in Water Distribution Network Using EPANET and GIS,» 2012.
- [3] L. Grbčić, L. Kranjčević και S. Družeta, «Machine Learning and Simulation-Optimization Coupling for Water Distribution Network Contamination Source Detection,» *Sensors*, vol. 21, no. 4, p. 1157, Feb. 2021.
- [4] L. Rossman, «EPANET,» 2000. [Ηλεκτρονικό]. Available: https://epanet22.readthedocs.io/en/latest/1_introduction.html#what-is-epanet. [Πρόσβαση 2022].
- [5] L. Rossman, «EPANET,» 2000. [Ηλεκτρονικό]. Available: https://epanet22.readthedocs.io/en/latest/3_network_model.html#water-quality-simulation-model. [Πρόσβαση 2022].
- [6] Matlab, «The MathWorks, Inc.,» [Ηλεκτρονικό]. Available: <https://www.mathworks.com/discovery/machine-learning.html>.
- [7] B. Mahesh, «Machine Learning Algorithms -A Review,» 2019.
- [8] Matlab, «The MathWorks, Inc.,» 1994. [Ηλεκτρονικό]. Available: <https://www.mathworks.com/discovery/supervised-learning.html>.
- [9] Matlab, «The MathWorks, Inc.,» 1994. [Ηλεκτρονικό]. Available: <https://www.mathworks.com/discovery/unsupervised-learning.html>.
- [10] javaTpoint, «javaTpoint,» 2011. [Ηλεκτρονικό]. Available: <https://www.javatpoint.com/unsupervised-machine-learning>.
- [11] javaTpoint, «javaTpoint,» 2011. [Ηλεκτρονικό]. Available: <https://www.javatpoint.com/regression-analysis-in-machine-learning>.
- [12] javaTpoint, «javaTpoint,» 2011. [Ηλεκτρονικό]. Available: <https://www.javatpoint.com/linear-regression-in-machine-learning>.

- [13] D. Maulud και A. M. Abdulazeez, «“A Review on Linear Regression Comprehensive in Machine Learning”,» *JASTT*, vol. 1, no. 4, pp.140-147, Dec. 2020.
- [14] H. Sit, «Medium,» 19 June 2019. [Ηλεκτρονικό]. Available: <https://towardsdatascience.com/quick-start-to-gaussian-process-regression-36d838810319>.
- [15] E. Schulz, M. Speekenbrink και A. Krause, «A tutorial on Gaussian process regression: Modelling, exploring, and exploiting functions,» *Journal of Mathematical Psychology*, Volume 85, pp. 1-16, Aug. 2018.
- [16] M. Alhamid, «Medium,» 15 Mar. 2021. [Ηλεκτρονικό]. Available: <https://towardsdatascience.com/ensemble-models-5a62d4f4cb0c>.
- [17] IBM, «IBM Cloud Education, IBM,» 26 May 2021. [Ηλεκτρονικό]. Available: <https://www.ibm.com/cloud/learn/boosting>.
- [18] javaTpoint, «javaTpoint,» 2011. [Ηλεκτρονικό]. Available: <https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm>.
- [19] D. Ayon, «Machine Learning Algorithms : A Review,» 2016.
- [20] Χ. Μακρυγιώργος, «Εφαρμογή αλγορίθμων μηχανικής μάθησης και νευρωνικών δικτύων για την ταξινόμηση υπερηχητικών σημάτων sonar,» 2021.
- [21] Matlab, «The Mathworks, Inc.,» 1994. [Ηλεκτρονικό]. Available: <https://www.mathworks.com/help/stats/assess-model-performance-in-regression-learner.html>.
- [22] D. Petridis, «MULTIPLE DISCRIMINANT ANALYSIS [Chapter],» σε *ANALYSIS OF MULTIVARIATE TECHNIQUES [Undergraduate textbook]*, 2015.
- [23] Π. Ανδριτσάκος, «΄Διατύπωση μετρικών κατανομής πόρων και βέλτιστου τρόπου εκτέλεσης κατανεμημένων αλγορίθμων μηχανικής μάθησης΄,» 2019.
- [24] Ι. Καλατζής, «Ανάλυση Γραμμικής Διάκρισης,» Αθήνα, 2017.
- [25] . Μ. Νάι, «Επιβλεπόμενη Μηχανική Μάθηση και το Πρόβλημα της Ταξινόμησης,» Αθήνα, 2019.
- [26] Ε. ΡΕΘΕΜΙΩΤΑΚΗ, «ΥΠΟΛΟΓΙΣΤΙΚΕΣ ΜΕΘΟΔΟΛΟΓΙΕΣ ΓΙΑ ΤΗΝ ΠΡΟΒΛΕΨΗ ΤΗΣ ΠΤΩΧΕΥΣΗΣ ΤΡΑΠΕΖΩΝ,» ΧΑΝΙΑ, 2016.

- [27] Π. Σουλίδης, «ΜΑΘΗΜΑΤΙΚΗ ΑΝΑΛΥΣΗ ΜΗΧΑΝΩΝ ΔΙΑΝΥΣΜΑΤΩΝ ΥΠΟΣΤΗΡΙΞΗΣ (SUPPORT VECTOR MACHINES) ΚΑΙ ΕΦΑΡΜΟΓΕΣ,» Κατερίνη, 2021.
- [28] S. Yildirim, «Medium,» 23 Apr. 2020. [Ηλεκτρονικό]. Available: <https://towardsdatascience.com/top-machine-learning-algorithms-for-clustering-a09c6771805>.
- [29] A. Khalil, «Medium,» 29 July 2021. [Ηλεκτρονικό]. Available: <https://medium.com/geekculture/dbscan-clustering-algorithm-simply-explained-993c195d1f63>.
- [30] Matlab, «The MathWorks, Inc,» 1994. [Ηλεκτρονικό]. Available: <https://www.mathworks.com/help/stats/support-vector-machines-for-binary-classification.html>.
- [31] javaTpoint, «javaTpoint,» 2011. [Ηλεκτρονικό]. Available: <https://www.javatpoint.com/k-nearest-neighbor-algorithm-for-machine-learning>.
- [32] javaTpoint, «javaTpoint,» 2011. [Ηλεκτρονικό]. Available: <https://www.javatpoint.com/k-means-clustering-algorithm-in-machine-learning>.
- [33] javaTpoint, «javaTpoint,» 2011. [Ηλεκτρονικό]. Available: <https://www.javatpoint.com/hierarchical-clustering-in-machine-learning>.
- [34] javaTpoint, «javaTpoint,» 2011. [Ηλεκτρονικό]. Available: <https://www.javatpoint.com/machine-learning-naive-bayes-classifier>.
- [35] Μ. Παπαδάκη, «Μηχανές διανυσματικής υποστήριξης (SVMs) και εφαρμογές σε πραγματικά σεισμολογικά δεδομένα,» 2012.
- [36] Matlab, «The Mathworks, Inc.,» 1994. [Ηλεκτρονικό]. Available: <https://www.mathworks.com/campaigns/offers/next/choosing-machine-learning-algorithms.html>.

Παραρτήματα

Παράρτημα 1 (Δημιουργία αρχείων εκπαίδευσης)

```

clear;
clc;
plithosfiles=1000;
var1domain={'Open' 'Closed'};
var2domain={'Open' 'Closed'};
var3domain={'Open' 'Closed'};
var4domain={'Open' 'Closed'};
var5domain={'Open' 'Closed'};
var6domain={'Open' 'Closed'};
var7domain=[0 60];

% anoigma template arxeiou
fidinporg=fopen('c:\SENARIA\nettemplate.inp','rt');

%dimiourgia bat arxeiou
fidbat=fopen('c:\SENARIA\TRAIN\eparuns_train.bat', 'wt');

%dimiourgia input arxeiou neuronikou
fidneur=fopen('c:\SENARIA\TRAIN\neurin_train.txt', 'wt');

%dimiourgia input arxeiwn
for fileno=1:plithosfiles
    %Variable replacement
    inpdestfile=['c:\SENARIA\TRAIN\SCEN' num2str(fileno) '.inp'];
    fidinpdes=fopen(inpdestfile,'wt');
    frewind(fidinporg);

    %dinw times stis metavlites
    variable1=char(var1domain(randi(length(var1domain))));
    variable2=char(var2domain(randi(length(var2domain))));
    variable3=char(var3domain(randi(length(var3domain))));
    variable4=char(var4domain(randi(length(var4domain))));
    variable5=char(var5domain(randi(length(var5domain))));
    variable6=char(var6domain(randi(length(var6domain))));
    variable7=num2str(min(var7domain)+(max(var7domain)-
min(var7domain)).*rand(1));

    %kanw antikatastasti tws metavlitwn sto template arxeio
    while 1
        stringline = fgets(fidinporg);
        if ~ischar(stringline), break, end
        %check if string contains a variable that needs to be changed
        stringline = regexprep(stringline, '#VAR1#', variable1);
        stringline = regexprep(stringline, '#VAR2#', variable2);
        stringline = regexprep(stringline, '#VAR3#', variable3);
        stringline = regexprep(stringline, '#VAR4#', variable4);
        stringline = regexprep(stringline, '#VAR5#', variable5);
        stringline = regexprep(stringline, '#VAR6#', variable6);
        stringline = regexprep(stringline, '#VAR7#', variable7);
        %write same or replaced string to inp file
        fprintf(fidinpdes,'%s',stringline);
    end
    fclose(fidinpdes);
end

```

```

    %eggrafi entolis gia epanet sto batfile
    batstring=['C:\Program Files (x86)\EPANET 2.2\runepanet"
"C:\SENARIA\TRAIN\SCEN' num2str(fileno) '.inp" ' '"C:\SENARIA\TRAIN\SCEN'
num2str(fileno) '.rpt"'];
    fprintf(fidbat, '%s\n', batstring);

    %eggrafi input arxeiou neuronikou

variable1_bool=0;variable2_bool=0;variable3_bool=0;variable4_bool=0;variabl
e5_bool=0;variable6_bool=0;
    if strcmp(strtrim(variable1),'Open')==1,variable1_bool=1;end
    if strcmp(strtrim(variable2),'Open')==1,variable2_bool=1;end
    if strcmp(strtrim(variable3),'Open')==1,variable3_bool=1;end
    if strcmp(strtrim(variable4),'Open')==1,variable4_bool=1;end
    if strcmp(strtrim(variable5),'Open')==1,variable5_bool=1;end
    if strcmp(strtrim(variable6),'Open')==1,variable6_bool=1;end
    fprintf(fidneur, '%i;%i;%i;%i;%i;%i;%i;%s\n',
variable1_bool,variable2_bool,variable3_bool,variable4_bool,variable5_bool,
variable6_bool,variable7);
end
fclose(fidinporg);
fclose(fidbat);
fclose(fidneur);

```

Παράρτημα 2 (Συλλογή τιμών ενέργειας, πίεσης, παροχής)

```

clear;
clc;
plithosfiles=1000;
flow_id='P7';
pres_id='J12';

%dimiourgia input arxeiou neuronikou gia parsing paroxis, piesis, energeias
fidneur=fopen('c:\SENARIA\TRAIN\neurout_train.txt', 'wt');

%dimiourgia DATASET TRAIN arxeiou neuronikou
fidneurds=fopen('c:\SENARIA\TRAIN\dataset_train.txt', 'wt');

%anoigma arxeiou me input neuronikou
fidneurin=fopen('c:\SENARIA\TRAIN\neurin_train.txt', 'rt');

%anoigma report arxeiwn
for fileno=1:plithosfiles
    %Variable replacement
    outdestfile=['c:\SENARIA\TRAIN\SCEN' num2str(fileno) '.rpt'];
    fidoutpdes=fopen(outdestfile,'rt');

    %euresi sinolikis katanalosis energeias
    energy=0;
    while 1
        stringline = fgets(fidoutpdes);
        if ~ischar(stringline), break, end

        %euresi energy
        if strcmp(strtrim(stringline),'Energy Usage:')==1
            stringline = fgets(fidoutpdes);
            stringline = fgets(fidoutpdes);
            stringline = fgets(fidoutpdes);
            stringline = fgets(fidoutpdes);
            while 1
                stringline = fgets(fidoutpdes);
                if strcmp(strtrim(stringline),'-----')
                    -----')==1,break,end
                line=textscan(stringline, '%s %f %f %f %f %f %f');
                energy=energy+line{4};
            end
        end
    end
    frewind(fidoutpdes);

    %euresi piesis ston zitoumeno komvo
    while 1
        stringline = fgets(fidoutpdes);
        if ~ischar(stringline), break, end

        %euresi pressure
        if strcmp(strtrim(stringline),'Node Results:')==1
            stringline = fgets(fidoutpdes);
            stringline = fgets(fidoutpdes);
            stringline = fgets(fidoutpdes);
            stringline = fgets(fidoutpdes);

```

```

        while 1
            stringline = fgets(fidoutpdes);
            line=textscan(stringline, '%s %f %f %f %f %s');
            if strcmp(line{1},pres_id)==1
                pressure=line{4};
                break;
            end
        end
    end
end
frewind(fidoutpdes);

%euresi paroxis sto zitoumeno agogo
while 1
    stringline = fgets(fidoutpdes);
    if ~ischar(stringline), break, end

    %euresi flow
    if strcmp(strtrim(stringline),'Link Results:')==1
        stringline = fgets(fidoutpdes);
        stringline = fgets(fidoutpdes);
        stringline = fgets(fidoutpdes);
        stringline = fgets(fidoutpdes);
        while 1
            stringline = fgets(fidoutpdes);
            line=textscan(stringline, '%s %f %f %f %s %s');
            if strcmp(line{1},flow_id)==1
                flow=line{2};
                break;
            end
        end
    end
end
fclose(fidoutpdes);

%eggrafi input arxeiou neuronikou: ENERGEIA, PIESI, PAROXI
fprintf(fidneur, '%f;%f;%f\n', energy, pressure, flow);

%diavasma grammis apo to neurin_train kai eggrafi sto DATASET TRAIN
stringline_old = fgetl(fidneurin);
fprintf(fidneurds, '%s;%f;%f;%f\n',stringline_old, energy, pressure,
flow);

end
fclose(fidneur);
fclose(fidneurin);
fclose(fidneurds);

```

Παράρτημα 3 (Δημιουργία 16 προκαθορισμένων σεναρίων)

```

clear;
clc;

% SENARIA-->VAR#1 VAR#2 VAR#3 VAR#4 VAR#5 VAR#6 VAR#7
% All valves closed operation
senaria{1}={'Closed' 'Closed' 'Closed' 'Closed' 'Closed' 'Closed' '0'};
senaria{2}={'Closed' 'Closed' 'Closed' 'Closed' 'Closed' 'Closed' '15'};
senaria{3}={'Closed' 'Closed' 'Closed' 'Closed' 'Closed' 'Closed' '45'};
senaria{4}={'Closed' 'Closed' 'Closed' 'Closed' 'Closed' 'Closed' '60'};

% All valves open operation
senaria{5}={'Open' 'Open' 'Open' 'Open' 'Open' 'Open' '0'};
senaria{6}={'Open' 'Open' 'Open' 'Open' 'Open' 'Open' '15'};
senaria{7}={'Open' 'Open' 'Open' 'Open' 'Open' 'Open' '45'};
senaria{8}={'Open' 'Open' 'Open' 'Open' 'Open' 'Open' '60'};

% Parallel operation
senaria{9}={'Open' 'Open' 'Open' 'Open' 'Open' 'Closed' '0'};
senaria{10}={'Open' 'Open' 'Open' 'Open' 'Open' 'Closed' '15'};
senaria{11}={'Open' 'Open' 'Open' 'Open' 'Open' 'Closed' '45'};
senaria{12}={'Open' 'Open' 'Open' 'Open' 'Open' 'Closed' '60'};

% Serial operation
senaria{13}={'Open' 'Open' 'Closed' 'Closed' 'Open' 'Open' '0'};
senaria{14}={'Open' 'Open' 'Closed' 'Closed' 'Open' 'Open' '15'};
senaria{15}={'Open' 'Open' 'Closed' 'Closed' 'Open' 'Open' '45'};
senaria{16}={'Open' 'Open' 'Closed' 'Closed' 'Open' 'Open' '60'};

% anoigma template arxeiou
fidinporg=fopen('c:\SENARIA\nettemplate.inp','rt');

%dimiourgia bat arxeiou
fidbat=fopen('c:\SENARIA\TEST\eparuns_test.bat', 'wt');

%dimiourgia input arxeiou neuronikou
fidneur=fopen('c:\SENARIA\TEST\neurin_test.txt', 'wt');

%dimiourgia input arxeiwn
for fileno=1:length(senaria)
    %Variable replacement
    inpdestfile=['c:\SENARIA\TEST\SCEN' num2str(fileno) '.inp'];
    fidinpdes=fopen(inpdestfile,'wt');
    frewind(fidinporg);

    %dinw times stis metavlites
    variable1=char(senaria{fileno}(1));
    variable2=char(senaria{fileno}(2));
    variable3=char(senaria{fileno}(3));
    variable4=char(senaria{fileno}(4));
    variable5=char(senaria{fileno}(5));
    variable6=char(senaria{fileno}(6));
    variable7=char(senaria{fileno}(7));

```

```

%kanw antikatastasti tws metavlitwn sto template arxeio
while 1
    stringline = fgets(fidinporg);
    if ~ischar(stringline), break, end
    %check if string contains a variable that needs to be changed
    stringline = regexprep(stringline, '#VAR1#', variable1);
    stringline = regexprep(stringline, '#VAR2#', variable2);
    stringline = regexprep(stringline, '#VAR3#', variable3);
    stringline = regexprep(stringline, '#VAR4#', variable4);
    stringline = regexprep(stringline, '#VAR5#', variable5);
    stringline = regexprep(stringline, '#VAR6#', variable6);
    stringline = regexprep(stringline, '#VAR7#', variable7);
    %write same or replaced string to inp file
    fprintf(fidinpdes, '%s', stringline);
end
fclose(fidinpdes);

%eggrafi entolis gia epanet sto batfile
batstring=['C:\Program Files (x86)\EPANET 2.2\runepanet'
"C:\SENARIA\TEST\SCEN' num2str(fileno) '.inp" ' ' "C:\SENARIA\TEST\SCEN'
num2str(fileno) '.rpt"'];
fprintf(fidbat, '%s\n', batstring);

%eggrafi input arxeiou neuronikou

variable1_bool=0;variable2_bool=0;variable3_bool=0;variable4_bool=0;variabl
e5_bool=0;variable6_bool=0;
    if strcmp(strtrim(variable1), 'Open')==1,variable1_bool=1;end
    if strcmp(strtrim(variable2), 'Open')==1,variable2_bool=1;end
    if strcmp(strtrim(variable3), 'Open')==1,variable3_bool=1;end
    if strcmp(strtrim(variable4), 'Open')==1,variable4_bool=1;end
    if strcmp(strtrim(variable5), 'Open')==1,variable5_bool=1;end
    if strcmp(strtrim(variable6), 'Open')==1,variable6_bool=1;end
    fprintf(fidneur,
        '%i;%i;%i;%i;%i;%i;%i;%s\n',
variable1_bool,variable2_bool,variable3_bool,variable4_bool,variable5_bool,
variable6_bool,variable7);
end
fclose(fidinporg);
fclose(fidbat);
fclose(fidneur);

```

Παράρτημα 4 (Συλλογή τιμών ενέργειας, πίεσης, παροχής για τα 16 σενάρια)

```

clear;
clc;
plithosfiles=16;
flow_id='P7';
pres_id='J12';

%dimiourgia input arxeiou neuronikou gia parsing paroxis, piesis, energeias
fidneur=fopen('c:\SENARIA\TEST\neurout_test.txt', 'wt');

%dimiourgia DATASET TEST arxeiou neuronikou
fidneurds=fopen('c:\SENARIA\TEST\dataset_test.txt', 'wt');

%anoigma arxeiou me input neuronikou
fidneurin=fopen('c:\SENARIA\TEST\neurin_test.txt', 'rt');

%anoigma report arxeiwn
for fileno=1:plithosfiles
    %Variable replacement
    outdestfile=['c:\SENARIA\TEST\SCEN' num2str(fileno) '.rpt'];
    fidoutpdes=fopen(outdestfile,'rt');

    %euresi sinolikis katanalosis energeias
    energy=0;
    while 1
        stringline = fgets(fidoutpdes);
        if ~ischar(stringline), break, end

        %euresi energy
        if strcmp(strtrim(stringline),'Energy Usage:')==1
            stringline = fgets(fidoutpdes);
            stringline = fgets(fidoutpdes);
            stringline = fgets(fidoutpdes);
            stringline = fgets(fidoutpdes);
            while 1
                stringline = fgets(fidoutpdes);
                if strcmp(strtrim(stringline),'-----')
                    -----')==1,break,end
                line=textscan(stringline, '%s %f %f %f %f %f %f');
                energy=energy+line{4};
            end
        end
    end
    frewind(fidoutpdes);

    %euresi piesis ston zitoumeno komvo
    while 1
        stringline = fgets(fidoutpdes);
        if ~ischar(stringline), break, end

        %euresi pressure
        if strcmp(strtrim(stringline),'Node Results:')==1
            stringline = fgets(fidoutpdes);
            stringline = fgets(fidoutpdes);
            stringline = fgets(fidoutpdes);
            stringline = fgets(fidoutpdes);

```

```

        while 1
            stringline = fgets(fidoutpdes);
            line=textscan(stringline, '%s %f %f %f %f %s');
            if strcmp(line{1},pres_id)==1
                pressure=line{4};
                break;
            end
        end
    end
end
frewind(fidoutpdes);

%euresi paroxis sto zitoumeno agogo
while 1
    stringline = fgets(fidoutpdes);
    if ~ischar(stringline), break, end

    %euresi flow
    if strcmp(strtrim(stringline),'Link Results:')==1
        stringline = fgets(fidoutpdes);
        stringline = fgets(fidoutpdes);
        stringline = fgets(fidoutpdes);
        stringline = fgets(fidoutpdes);
        while 1
            stringline = fgets(fidoutpdes);
            line=textscan(stringline, '%s %f %f %f %s %s');
            if strcmp(line{1},flow_id)==1
                flow=line{2};
                break;
            end
        end
    end
end
fclose(fidoutpdes);

%eggrafi input arxeiou neuronikou: ENERGEIA, PIESI, PAROXI
fprintf(fidneur, '%f;%f;%f\n', energy, pressure, flow);

%diavasma grammis apo to neurin_test kai eggrafi sto DATASET test
stringline_old = fgetl(fidneurin);
fprintf(fidneurds, '%s;%f;%f;%f\n',stringline_old, energy, pressure,
flow);

end
fclose(fidneur);
fclose(fidneurin);
fclose(fidneurds);

```