



Πανεπιστήμιο Δυτικής Μακεδονίας
Τμήμα Μηχανικών Πληροφορικής & Τηλεπικοινωνιών

ΠΡΟΒΛΕΨΗ ΚΑΤΑΝΑΛΩΣΗΣ ΕΝΕΡΓΕΙΑΣ
ΣΕ ΚΕΝΤΡΑ ΔΕΔΟΜΕΝΩΝ
PREDICTION OF ENERGY CONSUMPTION IN DATA CENTERS

Κυριακίδης Αναστάσιος - 863
Επιβλέπων Καθηγητής: Πλόσκας Νικόλαος

10 Οκτωβρίου 2019

Περίληψη

Η εν λόγω διπλωματική εργασία εξετάζει το πρόβλημα της κατανάλωσης ενέργειας σε κέντρα δεδομένων. Ακόμη, θέτει σε εφαρμογή αλγόριθμους μηχανικής μάθησης με σκοπό τη πρόβλεψη κατανάλωσης ενέργειας σε δύο εξαρτήματα του υπολογιστή (επεξεργαστή και κάρτα γραφικών).

Πιο συγκεκριμένα, η εργασία μελετά τα κέντρα δεδομένων μέσω ιστορικών στοιχείων, καθώς και το πρόβλημα της κατανάλωσης ενέργειάς τους, μέσω στατιστικών καταγραφών. Έπειτα, ερευνά το επιστημονικό πεδίο της μηχανικής μάθησης μέσω των εφαρμογών στις οποίες έχει συμβάλει, αναλύει τις δύο κύριες κατηγορίες μηχανικής μάθησης, καθώς και τους αλγόριθμους οι οποίοι χρησιμοποιήθηκαν στην πειραματική διαδικασία.

Επιπλέον, η εργασία παρουσιάζει τη συλλογή και επεξεργασία των δεδομένων, καθώς και το λογισμικό που χρησιμοποιήθηκε. Το επόμενο βήμα της εργασίας είναι η πειραματική διαδικασία, η οποία χωρίζεται σε τρία πειράματα. Στο πρώτο πείραμα επιδιώκεται πρόβλεψη κατανάλωσης ενέργειας στον επεξεργαστή χρησιμοποιώντας τρία και δύο χαρακτηριστικά ως εισόδους, στο δεύτερο πείραμα επιδιώκεται πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών χρησιμοποιώντας τρία και δύο χαρακτηριστικά ως εισόδους και στο τρίτο πείραμα επιδιώκεται πρόβλεψη κατανάλωσης ενέργειας στην κάρτα γραφικών χρησιμοποιώντας τέσσερα και δύο χαρακτηριστικά ως εισόδους. Όλα τα παραπάνω πειράματα πραγματοποιούνται με τη βοήθεια των αλγόριθμων μηχανικής μάθησης.

Το τελευταίο σημείο στο οποίο επικεντρώνεται η εργασία είναι η διεξαγωγή συμπερασμάτων. Παρουσιάζοντας συγκεντρωτικά όλα τα αποτελέσματα από την πειραματική διαδικασία, εξάγονται χρήσιμα συμπεράσματα για τα χαρακτηριστικά που χρησιμοποιήθηκαν ως είσοδοι για τα πειράματα.

Λέξεις κλειδιά: κατανάλωση ενέργειας, κέντρα δεδομένων, επεξεργαστής, κάρτα γραφικών, μηχανική μάθηση, αλγόριθμοι, δεδομένα

Abstract

This thesis examines the power consumption problem in data centers. Furthermore, it harnesses the power of machine learning algorithms in order to predict the power consumption in two PC components (CPU and GPU).

More specifically, the thesis studies data centers through historical data, as well as the power consumption problem through statistical records. Afterwards, it looks into the scientific field of machine learning, it analyzes the main categories of machine learning, and the algorithms that were used in the experimental procedure.

Moreover, the thesis presents how the data were collected and processed, and the software that was used. The next step of the thesis is the experimental procedure, which is separated into three experiments. In the first experiment we try to predict the power consumption of CPU using three and two features as inputs. In the second experiment we try to predict the total power consumption of the CPU and GPU using three and two features as inputs. Finally, in the third experiment we try to predict the power consumption of GPU using four and two features as inputs. The above experiments are materialized with the assistance of the machine learning algorithms.

The last point that this thesis is focused on are the conclusions. Presenting all the results from the experimental procedure, we can draw useful conclusions for the features that were used as inputs for the experiments.

Keywords: power consumption, data centers, CPU, GPU, machine learning, algorithms, data

Δήλωση Πνευματικών Δικαιωμάτων

Copyright (C) Κυριακίδης Αναστάσιος 2019, Κοζάνη

Υπογραφή Φοιτητή

Περιεχόμενα

1	Εισαγωγή	20
1.1	Κίνητρο Εκπόνησης Εργασίας	20
1.2	Σκοπός και Στόχοι	21
1.3	Συνεισφορά στην Επιστημονική Κοινότητα	23
1.4	Δομή Εργασίας	23
2	Κατανάλωση Ενέργειας σε Κέντρα Δεδομένων	25
2.1	Κέντρα Δεδομένων - Ιστορικά Στοιχεία	25
2.2	Το Πρόβλημα της Κατανάλωσης Ενέργειας	27
2.3	Σχετικές Εργασίες	29
3	Αλγόριθμοι Μηχανικής Μάθησης	33
3.1	Εισαγωγή στη Μηχανική Μάθηση	33
3.2	Τύποι Μηχανικής Μάθησης	36
3.2.1	Επιβλεπόμενη Μάθηση	36
3.2.2	Μη-επιβλεπόμενη Μάθηση	38
3.3	Αλγόριθμοι Μηχανικής Μάθησης	38
4	Πειραματική Διαδικασία	51
4.1	Προγράμματα που Χρησιμοποιήθηκαν	51
4.2	Συλλογή και Επεξεργασία Δεδομένων	52
4.3	Πειράματα στον υπολογιστή μεσαίων προδιαγραφών	53
4.4	Πειράματα στον υπολογιστή υψηλών προδιαγραφών	118
5	Συμπεράσματα	185

Κατάλογος σχημάτων

1.1	CPU vs GPU	22
2.1	Ανάλυση κατανάλωση ενέργειας [Dayarathna et al., 2015]	29
3.1	Αναπαράσταση ταξινόμησης (αριστερά) και παλινδρόμησης (δεξιά) στη μηχανική μάθηση	37
3.2	Προβλέψεις μοντέλου k-NN με 1 γείτονα	39
3.3	Προβλέψεις μοντέλου k-NN με 3 γείτονες	40
3.4	Αναπαράσταση Γραμμική Παλινδρόμηση	42
3.5	Το βέλτιστο υπερεπίπεδο μέσω του αλγόριθμου SVM [Pupale, 2018] .	45
3.6	Αλγόριθμος SVM σε μή-γραμμικά δεδομένα [Pupale, 2018]	45
3.7	Δέντρο απόφασης με βάθος 1	46
3.8	Δέντρο απόφασης με βάθος 2 (αριστερά) και με βάθος 9 (δεξιά)	47
3.9	Νευρωνικό δίκτυο με 1 κρυφό επίπεδο	49
3.10	Βαθύ νευρωνικό δίκτυο με 2 κρυφά επίπεδα	50
4.1	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (KNN)	54
4.2	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 3 εισόδους (KNN)	54
4.3	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (KNN)	55
4.4	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (KNN)	55
4.5	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (KNN)	56

4.6	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (KNN)	57
4.7	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (KNN)	57
4.8	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (KNN)	58
4.9	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (KNN)	58
4.10	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 4 εισόδους (KNN)	59
4.11	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (KNN)	59
4.12	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 εισόδους (KNN)	60
4.13	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 3 εισόδους (LR)	61
4.14	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (LR)	62
4.15	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (LR)	63
4.16	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (LR)	64
4.17	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 4 εισόδους (LR)	65
4.18	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 εισόδους (LR)	66

4.19	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (RR)	68
4.20	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 3 εισόδους (RR)	68
4.21	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (RR)	69
4.22	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (RR)	69
4.23	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (RR)	70
4.24	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (RR)	71
4.25	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (RR)	72
4.26	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (RR)	72
4.27	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (RR)	73
4.28	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 4 εισόδους (RR)	74
4.29	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (RR)	75
4.30	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 εισόδους (RR) . . .	75
4.31	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (LR)	77
4.32	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 3 εισόδους (LR)	77
4.33	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (LR)	78

4.34	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (LR)	79
4.35	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (LR)	80
4.36	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (LR)	80
4.37	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (LR)	81
4.38	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (LR)	81
4.39	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (LR)	82
4.40	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 4 εισόδους (LR)	83
4.41	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (LR)	84
4.42	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 εισόδους (LR) . . .	84
4.43	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (SVM)	86
4.44	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 3 εισόδους (SVM)	86
4.45	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (SVM)	87
4.46	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (SVM)	87
4.47	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (SVM)	88

4.48	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (SVM)	88
4.49	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (SVM)	89
4.50	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (SVM)	89
4.51	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (SVM)	90
4.52	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 4 εισόδους (SVM)	90
4.53	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (SVM)	91
4.54	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 εισόδους (SVM)	91
4.55	Δέντρο απόφασης για τη πρόβλεψη κατανάλωσης ενέργειας στον επεξεργαστή με 3 χαρακτηριστικά (DT)	93
4.56	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (DT)	93
4.57	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 3 εισόδους (DT)	94
4.58	Δέντρο απόφασης για τη πρόβλεψη κατανάλωσης ενέργειας στον επεξεργαστή με 2 χαρακτηριστικά (DT)	94
4.59	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (DT)	95
4.60	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (DT)	95
4.61	Δέντρο απόφασης για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (DT)	96

4.62	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (DT)	96
4.63	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (DT)	97
4.64	Δέντρο απόφασης για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (DT)	97
4.65	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (DT)	98
4.66	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (DT)	98
4.67	Δέντρο απόφασης για τη πρόβλεψη κατανάλωσης ενέργειας στην κάρτα γραφικών με 4 χαρακτηριστικά (DT)	99
4.68	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (DT)	99
4.69	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 4 εισόδους (DT)	100
4.70	Δέντρο απόφασης για τη πρόβλεψη κατανάλωσης ενέργειας στην κάρτα γραφικών με 2 χαρακτηριστικά (DT)	100
4.71	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (DT)	101
4.72	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 2 εισόδους (DT)	101
4.73	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (RF)	103
4.74	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 3 εισόδους (RF)	103
4.75	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (RF)	104
4.76	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (RF)	104

4.77	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (RF)	105
4.78	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (RF)	106
4.79	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (RF)	107
4.80	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (RF)	107
4.81	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (RF)	108
4.82	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 4 εισόδους (RF)	108
4.83	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (RF)	109
4.84	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 2 εισόδους (RF)	110
4.85	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (NN)	112
4.86	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 3 εισόδους (NN)	112
4.87	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (NN)	113
4.88	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (NN)	113
4.89	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (NN)	114
4.90	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (NN)	114

4.91	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (NN)	115
4.92	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (NN)	115
4.93	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (NN)	116
4.94	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 4 εισόδους (NN)	116
4.95	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (NN)	117
4.96	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 2 εισόδους (NN)	117
4.97	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (KNN)	119
4.98	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 3 εισόδους (KNN)	119
4.99	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (KNN)	120
4.100	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (KNN)	121
4.101	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (KNN)	121
4.102	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (KNN)	122
4.103	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (KNN)	122

4.104	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (KNN)	123
4.105	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (KNN)	123
4.106	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 4 εισόδους (KNN)	124
4.107	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (KNN)	124
4.108	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 εισόδους (KNN) .	125
4.109	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 3 εισόδους (LR)	126
4.110	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (LR)	127
4.111	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (LR)	128
4.112	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (LR)	129
4.113	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 4 εισόδους (LR)	130
4.114	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 εισόδους (LR) . . .	131
4.115	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (RR)	133
4.116	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 3 εισόδους (RR)	133
4.117	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (RR)	134

4.118	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (RR)	134
4.119	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (RR)	135
4.120	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (RR)	136
4.121	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (RR)	137
4.122	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (RR)	137
4.123	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (RR)	138
4.124	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 4 εισόδους (RR)	139
4.125	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (RR)	140
4.126	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 εισόδους (RR) . . .	140
4.127	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (LR)	142
4.128	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 3 εισόδους (LR)	142
4.129	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (LR)	143
4.130	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (LR)	144
4.131	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (LR)	145

4.132	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (LR)	145
4.133	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (LR)	146
4.134	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (LR)	146
4.135	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (LR)	147
4.136	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 4 εισόδους (LR)	148
4.137	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (LR)	149
4.138	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 εισόδους (LR) . . .	149
4.139	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (SVM)	151
4.140	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 3 εισόδους (SVM)	151
4.141	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (SVM)	152
4.142	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (SVM)	152
4.143	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (SVM)	153
4.144	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (SVM)	153

4.145	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (SVM)	154
4.146	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (SVM)	154
4.147	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (SVM)	155
4.148	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 4 εισόδους (SVM)	155
4.149	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (SVM)	156
4.150	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 εισόδους (SVM) . .	156
4.151	Δέντρο απόφασης για τη πρόβλεψη κατανάλωσης ενέργειας στον επεξεργαστή με 3 χαρακτηριστικά (DT)	158
4.152	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (DT)	158
4.153	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 3 εισόδους (DT)	159
4.154	Δέντρο απόφασης για τη πρόβλεψη κατανάλωσης ενέργειας στον επεξεργαστή με 2 χαρακτηριστικά (DT)	159
4.155	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (DT)	160
4.156	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (DT)	160
4.157	Δέντρο απόφασης για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (DT)	161
4.158	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (DT)	161

4.159	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (DT)	162
4.160	Δέντρο απόφασης για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (DT)	162
4.161	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (DT)	163
4.162	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (DT)	163
4.163	Δέντρο απόφασης για τη πρόβλεψη κατανάλωσης ενέργειας στην κάρτα γραφικών με 4 χαρακτηριστικά (DT)	164
4.164	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (DT)	164
4.165	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 4 εισόδους (DT)	165
4.166	Δέντρο απόφασης για τη πρόβλεψη κατανάλωσης ενέργειας στην κάρτα γραφικών με 2 χαρακτηριστικά (DT)	165
4.167	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (DT)	166
4.168	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 2 εισόδους (DT)	166
4.169	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (RF)	168
4.170	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 3 εισόδους (RF)	168
4.171	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (RF)	169
4.172	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (RF)	169
4.173	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (RF)	170

4.174	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (RF)	171
4.175	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (RF)	172
4.176	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (RF)	172
4.177	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (RF)	173
4.178	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 4 εισόδους (RF)	173
4.179	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (RF)	174
4.180	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 2 εισόδους (RF)	175
4.181	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (NN)	177
4.182	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 3 εισόδους (NN)	177
4.183	Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (NN)	178
4.184	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (NN)	178
4.185	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (NN)	179
4.186	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (NN)	179
4.187	Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (NN)	180

4.188	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (NN)	180
4.189	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (NN)	181
4.190	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 4 εισόδους (NN)	181
4.191	Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (NN)	182
4.192	Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 2 εισόδους (NN)	183

Κεφάλαιο 1

Εισαγωγή

1.1 Κίνητρο Εκπόνησης Εργασίας

Ζούμε σε μια εποχή στην οποία η ανάγκη για αποθήκευση, επεξεργασία και διανομή μεγαλύτερου όγκου δεδομένων καθιστά την ανάγκη δημιουργίας οικονομικά και ενεργειακά αποδοτικότερων κέντρων δεδομένων. Είναι ευρέως γνωστό πως τα κέντρα δεδομένων καταναλώνουν μεγάλο ποσό ενέργειας. Σύμφωνα με την Ευρωπαϊκή Επιτροπή [Garris, 2018], μέχρι το 2020, τα κέντρα δεδομένων στην Ευρώπη θα χρησιμοποιούν 259 TWh ηλεκτρικής ενέργειας, το οποίο αντιστοιχεί στο 1.7% της παγκόσμιας κατανάλωσης ενέργειας.

Σύμφωνα με έρευνα του 2016 [Garris, 2018] για τα Ευρωπαϊκά κέντρα Δεδομένων, το 43% των κέντρων δεδομένων δεν πληρούν τους στόχους ενεργειακής απόδοσης. Μόνο το 29% των κέντρων δεδομένων που πήραν μέρος στην έρευνα μπορούν να ποσοτικοποιήσουν την περιβαλλοντική τους επίπτωση. Ακόμη, το 97% παρατήρησε διάφορους τομείς, συμπεριλαμβανομένης και της ενεργειακής αποδοτικότητας, οι οποίοι θα μπορούσαν να βελτιωθούν.

Ουσιαστικά, οι εγκαταστάσεις αυτές καταναλώνουν πολύ περισσότερη ενέργεια, απ'ότι θα περίμεναν οι ιδιοκτήτες τους. Οι Ευρωπαϊκοί φορείς χάραξης πολιτικής έχουν αναγνωρίσει τα κέντρα δεδομένων ως έναν από τους πιο αναπτυσσόμενους τομείς, όσον αφορά την κατανάλωση ενέργειας.

Μέσα στην επόμενη δεκαετία, εκτιμάται πως θα υπάρχουν 30 φορές περισσότερα δεδομένα και 1000 φορές περισσότεροι διακομιστές. Ως αποτέλεσμα, ο αριθμός των κέντρων δεδομένων (αλλά και η ποσότητα ενέργειας που καταναλώνουν) θα αυξάνονται συνεχώς [Garris, 2018].

Ένα υποπεδίο της επιστήμης των υπολογιστών, η μηχανική μάθηση, η οποία εστιάζει στην ανάλυση και διερμηνία μοτίβων σε δεδομένα, με σκοπό ο υπολογιστής να μάθει να αποφασίζει και να προβλέπει χωρίς την ανθρώπινη παρέμβαση, έρχεται να βοηθήσει στην επίλυση του παραπάνω προβλήματος. Αφού συλλεχθούν προσεκτικά (ποσοτικά και ποιοτικά) τα απαραίτητα δεδομένα, επεξεργάζονται από τους αλγορίθμους μηχανικής μάθησης και εξάγουν τις επιθυμητές προβλέψεις.

Όλα τα παραπάνω αποτέλεσαν κίνητρο για τη διεξαγωγή και εκπόνηση της συγκεκριμένης διπλωματικής εργασίας, η οποία στοχεύει στη μελέτη και πρόβλεψη κατανάλωσης ενέργειας συγκεκριμένων πολυπύρηνων υπολογιστών.

1.2 Σκοπός και Στόχοι

Ο σκοπός της εν λόγω εργασίας, είναι να μελετηθεί το πρόβλημα της κατανάλωσης ενέργειας σε κέντρα δεδομένων, καθώς και η εφαρμογή αλγορίθμων μηχανικής μάθησης, οι οποίοι, λαμβάνοντας υπόψιν συγκεκριμένες μετρήσεις σε μεμονομένους υπολογιστές, θα προβλέπουν την κατανάλωση ενέργειας στον επεξεργαστή, καθώς και στην κάρτα γραφικών.

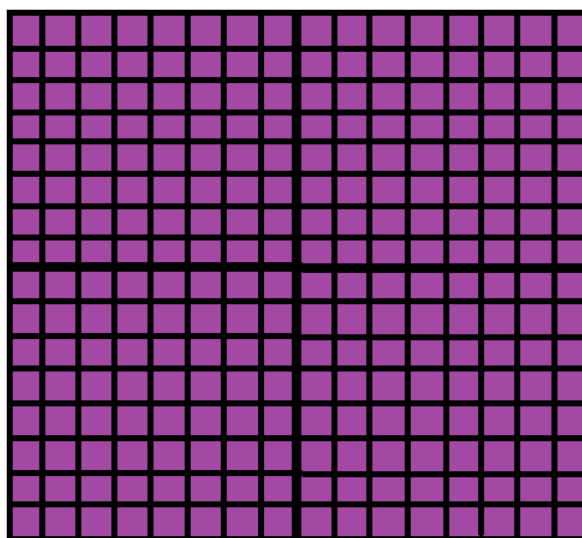
Παρόλο που οι κάρτες γραφικών θεωρούνταν ως μια τεχνολογία αποκλειστικά για τον κλάδο των παιχνιδιών, πλέον έχουν κάνει την εμφάνισή τους στα κέντρα δεδομένων, εγείροντας νέες πρωτοβουλίες σε επιστήμες όπως η μηχανική μάθηση, η τεχνητή νοημοσύνη κα. Έχοντας ως κύριο σκοπό δημιουργίας την ικανοποίηση αναγκών απαιτητικών παιχνιδιών, οι κάρτες γραφικών βρίσκουν πλέον μια νέα εφαρμογή, μέσα στα κέντρα δεδομένων. Η παράλληλη αρχιτεκτονική τους τις κάνει κατάλληλες για πολλές παραδοσιακές εργασίες, οι οποίες μπορούν να ωφεληθούν από την υπολογιστική τους επιτάχυνση. Επειδή οι κάρτες γραφικών διαθέτουν χιλιάδες πυρήνες και έχουν σχεδιαστεί για να χειρίζονται κίνηση γραφικών σε κλάσμα δευτερολέπτου, αν τους ανατεθεί μεγάλος όγκος δεδομένων για επεξεργασία, μπορούν να τον διευθετήσουν εξαιρετικά γρήγορα.

Σχήμα 1.1: CPU vs GPU

Επεξεργαστής



Κάρτα Γραφικών



Όπως φαίνεται και στο σχήμα 1.1, οι επεξεργαστές αποτελούνται από μερικούς πυρήνες, τα τελευταία μοντέλα διαθέτουν μέχρι 32. Κάθε πυρήνας εκτελεί έναν υπολογισμό τη φορά, ενώ η ενσωματωμένη μνήμη στο κύκλωμα εξασφαλίζει ότι τα δεδομένα βρίσκονται κοντά στους πυρήνες και είναι διαθέσιμα για γρήγορη πρόσβαση. Σε αντίθεση, οι σύγχρονες κάρτες γραφικών περιέχουν 3000 - 5000 πυρήνες (εικόνα 1.1) οι οποίοι επεξεργάζονται ξεχωριστά νήματα ταυτόχρονα [Martin, 2017].

Οι στόχοι της εν λόγω εργασίας συνεχίζονται με την ανάλυση και χρησιμοποίηση των δημοφιλέστερων αλγορίθμων μηχανικής μάθησης. Αφού λήφθηκαν προσεκτικά οι μετρήσεις (ποσοστό χρήσης, κατανάλωση ενέργειας σε Watt, θερμοκρασία κ.ά. σε επεξεργαστή και κάρτα γραφικών) σε 2 διαφορετικούς πολυπύρηνους υπολογιστές, οι παραπάνω αλγόριθμοι επεξεργάζονται τα δεδομένα τα οποία συλλέχθηκαν και προβλέπουν με μεγάλη ακρίβεια την κατανάλωση ενέργειας σε επεξεργαστή και κάρτα γραφικών αντίστοιχα.

Μετά τον σχεδιασμό και την χρήση, εκτός από την πρόβλεψη κατανάλωσης ενέργειας, οι παραπάνω αλγόριθμοι θα δώσουν εικόνα για το κατά πόσο ένα χαρακτηριστικό (από τα δεδομένα) είναι σημαντικό. Η απόδοση κάθε αλγορίθμου θα αξιολογηθεί και θα συγκριθεί με τους υπόλοιπους για τη διεξαγωγή χρήσιμων συμπερασμάτων.

1.3 Συνεισφορά στην Επιστημονική Κοινότητα

Η εν λόγω διπλωματική εργασία αποτελεί προϊόν πρωτότυπου συνθετικού έργου, με απώτερο σκοπό τη συνεισφορά στην επιστημονική κοινότητα, αναλύοντας το πρόβλημα της κατανάλωσης ενέργειας σε κέντρα δεδομένων, καθώς και το κατά πόσο και πώς μπορεί να βοηθήσει η μηχανική μάθηση στη επίλυση αυτού, τόσο σε κέντρα δεδομένων, όσο και σε μεμονομένους υπολογιστές. Μέσα από τη πειραματική διαδικασία, τίθενται σε εφαρμογή οι αλγόριθμοι μηχανικής μάθησης, μελετώντας τη πρόβλεψη κατανάλωσης ενέργειας σε επεξεργαστή και κάρτα γραφικών, δύο διαφορετικών πολυπύρηνων υπολογιστών. Ακόμη, υπάρχει δυνατότητα για περαιτέρω έρευνα πάνω στο συγκεκριμένο πρόβλημα, λαμβάνοντας υπόψιν και άλλα μέρη του υπολογιστή (όπως ο σκληρός δίσκος, η μνήμη τυχαίας προσπέλασης, κ.ά.).

1.4 Δομή Εργασίας

Η συγκεκριμένη εργασία χωρίζεται σε πέντε κεφάλαια. Το πρώτο κεφάλαιο είναι εισαγωγικό και παρουσιάζει το κίνητρο, τους σκοπούς και τους στόχους την εν λόγω εργασίας. Ακόμη αναφέρεται η συνεισφορά της εργασίας στην επιστημονική κοινότητα καθώς και η δομή που ακολουθείται.

Στο δεύτερο κεφάλαιο παρουσιάζεται μια ιστορική αναδρομή των κέντρων δεδομένων, αναφέροντας σημαντικά επιτεύγματα των τελευταίων δεκαετιών. Ακολουθεί ανάλυση του προβλήματος της κατανάλωσης ενέργειας στα κέντρα δεδομένων, καθώς και παρουσίαση εργασιών με σχετικά θέματα.

Στο τρίτο κεφάλαιο παρουσιάζεται το πεδίο έρευνας της μηχανικής μάθησης, καθώς και τεχνολογικές ανακαλύψεις στις οποίες έχει συμβάλει. Έπειτα αναλύονται οι δύο τύποι της μηχανικής μάθησης και παρουσιάζονται αναλυτικά οι αλγόριθμοι που θα χρησιμοποιηθούν.

Το τέταρτο κεφάλαιο περιγράφει την πειραματική διαδικασία. Αρχικά παρουσιάζει τα λογισμικά που χρησιμοποιήθηκαν και εξηγεί τον τρόπο συλλογής των δεδομένων. Έπειτα παρουσιάζονται τα αποτελέσματα των πειραμάτων όλων των αλγορίθμων, σε δύο υπολογιστές διαφορετικών προδιαγραφών.

Στο πέμπτο κεφάλαιο παρουσιάζονται τα συμπεράσματα. Συνοψίζοντας συγκεκριμένα τα αποτελέσματα της πειραματικής διαδικασίας, λαμβάνουμε χρήσιμες

πληροφορίες για την συμβολή των χαρακτηριστικών που χρησιμοποιήθηκαν στα πειράματα.

Κεφάλαιο 2

Κατανάλωση Ενέργειας σε Κέντρα Δεδομένων

2.1 Κέντρα Δεδομένων - Ιστορικά Στοιχεία

Η ιδέα των κέντρων δεδομένων δεν είναι καινούργια, καθώς είχε αρχίσει να αναπτύσσεται πολλές δεκαετίες πριν. Τη δεκαετία του 1960, με την έννοια 'υπολογιστής' εννοούσαμε ένα μεγάλο σύστημα υπολογιστών τοποθετημένο σε δωμάτια, αυτό που σήμερα αποκαλούμε 'κέντρο δεδομένων'. Τα κέντρα δεδομένων ήταν αρκετά δαπανηρά, για αυτό επιχειρήσεις μπορούσαν να δανειστούν μέρος του όλου συστήματος, για να πραγματοποιήσουν συγκεκριμένες λειτουργίες.

Στα τέλη του 1950 [sho, 2019] οι εταιρίες American Airlines και IBM συνεργάστηκαν ώστε να δημιουργήσουν το σύστημα κρατήσεων επιβατών SABRE (Semi-Automated Business Research Environment), αυτοματοποιώντας γρήγορα και αξιόπιστα τις αεροπορικές κρατήσεις. Η ιδέα ενός συστήματος που θα επεξεργάζεται δεδομένα, το οποίο θα μπορεί να δημιουργεί και να διαχειρίζεται αεροπορικές κρατήσεις θέσεων και να διαθέτει άμεσα τα συγκεκριμένα δεδομένα ηλεκτρονικά σε οποιονδήποτε πράκτορα και σε οποιαδήποτε τοποθεσία παγκοσμίως, έγινε πραγματικότητα το 1960 [sho, 2019], ανοίγωντας έτσι την πόρτα στα πρώτα επιχειρησιακής κλίμακας κέντρα δεδομένων και πιο πρόσφατα ιντερνετικής κλίμακας κέντρα δεδομένων. Για πρώτη φορά, υπολογιστές συνδέθηκαν μαζί μέσω ενός δικτύου, το οποίο επέτρεψε σε ανθρώπους από όλο τον κόσμο να εισάγουν δεδομένα, καθώς και να επεξεργάζονται αιτήσεις και να ανταλλάσσουν πληροφορίες. Την ίδια εποχή, η κεντρική μονάδα ενός υπολογιστή ήταν πολύ μεγάλη και καταλάμβανε χώρο όσο

ένα δωμάτιο. Κύριος ιδιοκτήτης των περισσότερων κέντρων δεδομένων εκείνη την εποχή ήταν η κυβέρνηση, με τις επιχειρήσεις να αρχίζουν δειλά δειλά να επενδύουν σε υπολογιστές [Harvey, 2017].

Κατά την δεκαετία του 1980 [Harvey, 2017], οι υπολογιστές άρχισαν να αναπτύσσονται με ραγδαίους ρυθμούς εξαιτίας της άνθησης της βιομηχανίας των μικροϋπολογιστών. Αυτή η εποχή χαρακτηρίστηκε από την καινοτόμο εισαγωγή της IBM, τον προσωπικό υπολογιστή. Κατά συνέπεια, η IBM [Mitchell, 2018] καθιέρωσε εγκαταστάσεις υπολογιστών με κόστος 30 εκατομμύρια δολάρια στο πανεπιστήμιο του Cornell, με σκοπό να χρησιμοποιούνται ως κέντρο δεδομένων για τους προσωπικούς της υπολογιστές. Η πρωτοφανής ανάπτυξη της πληροφοριακής τεχνολογίας ώθησε να γίνει αντιληπτή η σημασία της ίδρυσης και συντήρησης ισχυρών κέντρων δεδομένων.

Τη δεκαετία του 1990 [Mitchell, 2018], οι μικρο-υπολογιστές άρχισαν να κάνουν την εμφάνισή τους, παίρνοντας τα ηνία από τα μεγάλα συστήματα υπολογιστών. Ολοένα και περισσότερες εταιρίες εγκαθιστούσαν δωμάτια διακομιστών στο κτήριό τους, λόγω του σχετικά χαμηλού κόστους του δικτυακού εξοπλισμού. Τα κέντρα δεδομένων ήρθαν περισσότερο στην επιφάνεια στις αρχές του 2000 [Harvey, 2017], κατά τη περίοδο του φαινομένου dot com bubble, αφού οι επιχειρήσεις χρειαζόταν γρήγορες ταχύτητες ίντερνετ για να καθιερώσουν τη θέση τους στο διαδίκτυο. Κατά συνέπεια, οι περισσότερες επιχειρήσεις άρχισαν να δημιουργούν μεγάλες εγκαταστάσεις ώστε να παρέχουν αντίγραφα ασφαλείας των δεδομένων τους, ενώ έννοιες όπως η κατανάλωση ενέργειας και ο κλιματισμός άρχισαν να απασχολούν ανησυχητικά.

Τη περίοδο 2000 - 2018 [Pramatarov, 2018], η ενεργειακή αποδοτικότητα άρχισε να παρουσιάζει προβλήματα συντήρησης. Τα κέντρα δεδομένων εκείνης της εποχής κατανάλωναν απίστευτη ενέργεια, με αποτέλεσμα η προσοχή να στραφεί στο σχεδιασμό αποδοτικότερων συστημάτων ψύξης και στη μείωση της κατανάλωσης ενέργειας. Στις μέρες μας [Mitchell, 2018] βιώνουμε το πικ της υπολογιστικής τεχνολογίας. Κάθε γνωστή εταιρία είναι πολύ πιθανό να έχει τουλάχιστον ένα κέντρο δεδομένων. Μεγάλα κέντρα δεδομένων δημιουργούνται καθημερινά για να εκπληρώσουν τις ανάγκες της βιομηχανίας. Τα τελευταία χρόνια, τα κέντρα δεδομένων έχουν εικονικοποιηθεί λόγω της άφιξης των τεχνολογιών υπολογιστικής

νέφους. Σήμερα, όλες οι πληροφορίες που λαμβάνουμε από μεγάλους κολοσσούς (Google, Amazon, Microsoft) περνάνε πρώτα από τα κέντρα δεδομένων τους, πριν διανεμηθούν σε μας μέσω του διαδικτύου.

2.2 Το Πρόβλημα της Κατανάλωσης Ενέργειας

Τα κέντρα δεδομένων, ως μεγάλης κλίμακας υπολογιστικές υποδομές που λειτουργούν όλο το εικοσιτετράωρο, προωθούν τη γρήγορη ανάπτυξη της πληροφοριακής τεχνολογίας και έχουν μεγάλο αντίκτυπο στην οικονομία. Η κρισιμότητα των κέντρων δεδομένων βασίζεται σε δύο φαινόμενα. Πρώτον, η αυξανόμενη ανάπτυξη στη ζήτηση για υπολογισμό, επεξεργασία και αποθήκευση δεδομένων από μια μεγάλη κλίμακα υπηρεσιών cloud, όπως η Google και το Facebook, από φορείς τηλεπικοινωνίας, όπως η British Telecom, από τράπεζες και άλλες υπηρεσίες, οδήγησαν στον πολλαπλασιασμό των κέντρων δεδομένων. Δεύτερον, η ανάγκη υποστήριξης μιας μεγάλης ποικιλίας εφαρμογών, οι οποίες μπορεί να εκτείνονται από μερικά δευτερόλεπτα έως εκείνες που λειτουργούν αδιάκοπα, έχει ωθήσει στο να χτιστούν μεγάλης κλίμακας υπολογιστικές υποδομές. Όντας μεγάλης κλίμακας υπολογιστικές εγκαταστάσεις, τα κέντρα δεδομένων, φέρουν τεράστιους ενεργειακούς προϋπολογισμούς, οι οποίοι προκαλούν διάφορα ενεργειακά θέματα απόδοσης [Dayarathna et al., 2015].

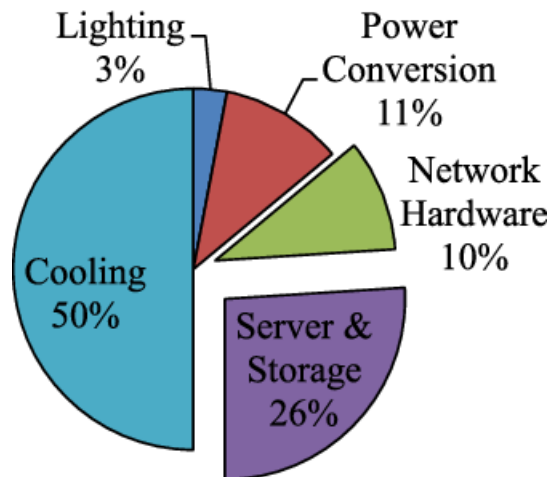
Η ενεργειακή αποδοτικότητα των κέντρων δεδομένων έχει κατορθώσει να γίνει παράγοντας - κλειδί τα τελευταία χρόνια εξαιτίας της οικονομικής και της περιβαλλοντικής επίδρασης. Τα κέντρα δεδομένων συμβάλλουν στις παγκόσμιες εκπομπές αερίων του θερμοκηπίου [Supermicro, 2018], με ένα ποσοστό της τάξης του 2%. Ένα ακόμη αρνητικό είναι τα ηλεκτρονικά απόβλητα, γνωστά και με τον όρο e-waste, είναι υποπροϊόν της λειτουργίας των κέντρων δεδομένων. Σύμφωνα με έρευνα [Supermicro, 2018], τα ηλεκτρονικά απόβλητα των κέντρων δεδομένων αντιστοιχούν το 2% σε στερεά απόβλητα και το 70% σε τοξικά απόβλητα. Τα κέντρα δεδομένων έχουν μεγάλη οικονομική επίδραση για πολλούς λόγους. Ένα τυπικό κέντρο δεδομένων [Dayarathna et al., 2015] καταναλώνει ενέργεια όσο 25.000 νοικοκυριά. Οι εγκαταστάσεις των κέντρων αυτών καταναλώνουν έως και 100 – 200 φορές ηλεκτρική ενέργεια περισσότερο από έναν τυπικό χώρο γραφείων. Σύμφωνα με έρευνα του 2016 [Ross, 2018], το 2015 η συνολική ηλεκτρική ενέργεια που κατανάλωσαν τα

κέντρα δεδομένων παγκοσμίως ήταν 416, 2 TWh, πολύ περισσότερη από τη συνολική κατανάλωση ενέργειας του Ηνωμένου Βασιλείου. Επιπλέον, το ενεργειακό κόστος για την τροφοδότηση ενός τυπικού κέντρου δεδομένων διπλασιάζεται κάθε τέσσερα χρόνια. Για αυτό, η απότομη αύξηση ανάγκης της ηλεκτρικής ενέργειας αλλά και του κόστους της, κάνει τον λογαριασμό ρεύματος ένα πολύ σημαντικό έξοδο για τα σημερινά κέντρα δεδομένων. Σε μερικές περιπτώσεις, τα έξοδα ρεύματος ξεπερνάνε τα έξοδα αγοράς εξοπλισμού.

Η ενεργειακή χρήση των κέντρων δεδομένων δημιουργεί πολλά περιβαλλοντικά προβλήματα. Για παράδειγμα, το 2005 [Dayarathna et al., 2015], η συνολική κατανάλωση ενέργειας των κέντρων δεδομένων ήταν το 1% της συνολικής κατανάλωσης ενέργειας της Αμερικής, και δημιουργούσε εκπομπές αερίων όσο και μια μεσαίου μεγέθους χώρα, όπως η Αργεντινή. Το 2010 [Dayarathna et al., 2015] η παγκόσμια ηλεκτρική ενέργεια που απαιτούνταν για τα κέντρα δεδομένων εκτιμώνταν ανάμεσα στο 1.1% – 1.5% της παγκόσμιας συνολικής ηλεκτρικής ενέργειας, ενώ στην Αμερική τα κέντρα δεδομένων κατανάλωναν 1.7% – 2.2% της συνολικής ηλεκτρικής ενέργειας της χώρας. Εν έτει 2019, η παγκόσμια ηλεκτρική ενέργεια που απαιτείται για την λειτουργία των κέντρων δεδομένων έχει φτάσει το 3% [McNerney, 2019]. Έρευνα του 2012 [Dayarathna et al., 2015] απέδειξε πως τα κέντρα δεδομένων κατανάλωναν παγκοσμίως 270TWh, και πως αυτή η κατανάλωση είχε σύνθετο ρυθμό αιτήσιας ανάπτυξης (Compound Annual Growth Rate) της τάξης του 4.4% από το 2007 μέχρι το 2012. Σύμφωνα με την IDC [McNerney, 2019], ο αριθμός των κέντρων δεδομένων από το 2012 μέχρι σήμερα μεγάλωσε κατά 16 φορές (από 500.000 σε 8.000.000). Κάθε χρόνο, εκατομμύρια κέντρα δεδομένων χρησιμοποιούν τόνους εξαρτημάτων υλικού, καταναλώνοντας ηλεκτρική ενέργεια όσο μία χώρα και εκπέμποντας διοξείδιο του άνθρακα όσο η παγκόσμια αεροπορική βιομηχανία. Αν και οι τεχνολογικές εξελίξεις είναι δύσκολο να προβλεφτούν με ακρίβεια, αρκετά μοντέλα προέβλεψαν πως η κατανάλωση ενέργειας των κέντρων δεδομένων μπορεί να ξεπεράσει το 10% της παγκόσμιας ηλεκτρικής ενέργειας μέχρι το 2030 [McNerney, 2019]. Για τους παραπάνω λόγους, η ενεργειακή αποδοτικότητα των κέντρων δεδομένων λαμβάνεται πολύ σοβαρά υπόψιν στις μέρες μας, ίσως πιο ψηλά και από έννοιες όπως η ασφάλεια. Ακόμη και όταν οι διακομιστές (servers) βρίσκονται σε αδρανή κατάσταση, καταναλώνουν σημαντική ποσότητα ενέργειας, η οποία μπορεί να μειωθεί

κλείνοντας τους αδρανείς διακομιστές. Η προαναφερθείσα λύση, καθώς και άλλες όπως η ενοποίηση του φόρτου εργασίας, πρέπει να λυφθούν ώστε να μειωθεί η κατανάλωση ηλεκτρικής ενέργειας των κέντρων δεδομένων. Από την άλλη μεριά, αυτές οι τεχνικές μείωσης ενεργειακού κόστους, μειώνουν την απόδοση του συστήματος, καταλλήγοντας σε μια σύνθετη ισορροπία μεταξύ εξοικονόμησης ενέργειας και υψηλής απόδοσης.

Σχήμα 2.1: Ανάλυση κατανάλωση ενέργειας [Dayarathna et al., 2015]



Η ενέργεια η οποία καταναλώνεται από ένα κέντρο δεδομένων μπορεί να κατηγοριοποιηθεί σε δύο μέρη, την ενέργεια η οποία χρησιμοποιείται από τον εξοπλισμό πληροφορικής τεχνολογίας (διακομιστές, δίκτυα, αποθηκευτικός χώρος, κα.) και την ενέργεια από τις εγκαταστάσεις υποδομής (συστήματα ψύξης, κλιματισμού, κα.)

Το ποσό ενέργειας που καταναλώνεται από τα δύο αυτά μέρη εξαρτάται από το σχεδιασμό του κέντρου δεδομένων, καθώς και από την αποδοτικότητα του εξοπλισμού. Για παράδειγμα, σύμφωνα με τη στατιστική που παρουσίασε η Infotech group (σχήμα 2.1), την περισσότερη ενέργεια σε ένα κέντρο δεδομένων καταναλώνει το σύστημα ψύξης (50%), ενώ οι διακομιστές και οι αποθηκευτικοί χώροι (26%) έρχονται δεύτεροι στην ιεραρχία [Dayarathna et al., 2015].

2.3 Σχετικές Εργασίες

Η αρχιτεκτονική ενός κέντρου δεδομένων είναι αρκετά πολύπλοκη, αφού δεν αποτελείται μόνο από το υλικό (hardware), αλλά και από το λογισμικό που τρέχει στην υποδομή του. Σύμφωνα με τον M. Dayarathma et al [Dayarathna et al., 2015], μπορούμε να κατηγοριοποιήσουμε τα στοιχεία του σε δύο επίπεδα, το υλικό και το

λογισμικό. Το επίπεδο υλικού αποτελείται από αρκετά στοιχεία όπως συστήματα φύξης, μονάδες διανομής ενέργειας, εξοπλισμός φωτισμού, διακομιστές και δικτυακός εξοπλισμός. Το επίπεδο λογισμικού χωρίζεται και αυτό σε δύο επίπεδα, το επίπεδο λειτουργικού συστήματος και το επίπεδο εφαρμογών. Το πρώτο αναφέρεται κυρίως στο λειτουργικό σύστημα που είναι εγκατεστημένο στους διακομιστές, ενώ το δεύτερο αναφέρεται στις ποικίλες εφαρμογές που τρέχουν στους διακομιστές. [Dayarathna et al., 2015]

Αρκετή έρευνα έχει υπάρξει πάνω στη πρόβλεψη κατανάλωσης ενέργειας σε κέντρα δεδομένων. Διάφορες προσπάθειες έχουν γίνει για να μειωθεί το κόστος που σχετίζεται με τη διανομή ενέργειας αλλά και τη φύξη του εξοπλισμού. Ως εκ τούτου, η διαχείριση ενέργειας έχει γίνει σημαντικό θέμα στις επιχειρήσεις. Αρκετές έρευνες [Bircher and John, 2011] [Economou et al., 2006] αξιολογούν την κατανάλωση ενέργειας σε επίπεδο συστήματος και προτείνουν μοντέλα τα οποία προβλέπουν τη κατανάλωση ενέργειας, λαμβάνοντας υπόψιν διάφορα εξαρτήματα του διακομιστή. Άλλες έρευνες [Contreras and Martonosi, 2005] [Joseph and Martonosi, 2001] [Rodrigues et al., 2013] [Singh et al., 2009] εστιάζουν περισσότερο στην κατανάλωση ενέργειας στον επεξεργαστή, καθώς είναι το εξάρτημα μέσα στον διακομιστή το οποίο καταναλώνει την περισσότερη ενέργεια.

Ο W L Bircher et al στην έρευνά του [Bircher and John, 2011] ασχολήθηκε με τη πρόβλεψη κατανάλωσης ενέργειας χρησιμοποιώντας γεγονότα απόδοσης του επεξεργαστή. Τα μοντέλα του εκμεταλλεύονται το φαινόμενο της κατάρευσης των παραπάνω γεγονότων. Τα γεγονότα αυτά, τα οποία είναι ορατά στην επεξεργαστική μονάδα, σχετίζονται με την κατανάλωση ενέργειας σε υποσυστήματα συμπεριλαμβανομένης της μνήμης, του κυλώματος εισόδου/εξόδου, του δίσκου και του μικροεπεξεργαστή. Υποσυστήματα τα οποία βρίσκονται αρκετά μακριά από τον μικροεπεξεργαστή απαιτούν γεγονότα τα οποία είναι σε μεγαλύτερη συσχέτιση με το υποσύστημα, όπως συσκευή διακόπτη εισόδου/εξόδου ή κατάσταση ρολογιού πύλης. Σύμφωνα με τον Bircher, τα μοντέλα μνήμης πρέπει να λαμβάνουν υπόψιν τους δραστηριότητες οι οποίες δεν προέρχονται από τον μικροεπεξεργαστή. Σε αυτή την περίπτωση, τα γεγονότα άμεσης πρόσβασης μνήμης (Direct Memory Access) έδειξαν ότι έχουν μεγάλη συσχέτιση με την ενέργεια στην μνήμη. Αποδείχθηκε πως μπορούμε να προβλέψουμε την συνολική ενέργεια του συστήματος με ένα μέσο σφάλμα μι-

κρότερο του 9% για κάθε υποσύστημα που χρησιμοποιεί γεγονότα απόδοσης τα οποία καταρρέουν από την επεξεργαστική μονάδα.

Ο D Economou et al στην έρευνά του [Economou et al., 2006] παρουσίασε το Mantis, μια παρεμβατική μέθοδο για μοντελοποίηση της κατανάλωσης ενέργειας ολόκληρου του συστήματος, η οποία μπορεί να χρησιμοποιηθεί εύκολα στην έρευνα της ενέργειας. Το Mantis βασίζεται σε μετρήσεις χρήσης των εξαρτημάτων, οι οποίες συλλέγονται μέσω του λειτουργικού συστήματος ή των μετρητών πρότυπου υλικού. Κατά τη διάρκεια μια φάσης βαθμονόμησης, τα εξαρτήματα στρεσάζονται μεμονωμένα μέσω συνθετικών φόρτων εργασίας και δημιουργούνται μοντέλα για να συσχετίσουν τις μετρήσεις χρήσης των εξαρτημάτων με την ενέργεια που μετρήθηκε. Έπειτα τα μοντέλα αυτά ενσωματώνονται στο Mantis για να προβλέψουν τη μέση κατανάλωση ενέργειας στο συνολικό σύστημα. Ο D Economou αναλύει το σχεδιασμό του Mantis, εστιάζοντας στην μοντελική προσέγγιση για τα κύρια εξαρτήματα του συστήματος και την υπόστασή τους σε δύο διαφορετικές κατηγορίες διακομιστών, ένα σύστημα χαμηλών δυνατοτήτων και ένα σύστημα υψηλών δυνατοτήτων. Για την επικύρωση του μοντέλου μέτρησε την κατανάλωση εναλλασσόμενου ρεύματος των συστημάτων αυτών και τα σύγκρισε με τα αποτελέσματα από το Mantis. Παρατηρήθηκε πως το Mantis προέβλεπε με μέσο σφάλμα κοντά στο 10% των κανονικών μετρήσεων ενέργειας.

Ο R Rodrigues et al [Rodrigues et al., 2013] στην έρευνά του χρησιμοποίησε μετρητές απόδοσης για τη πρόβλεψη κατανάλωσης ενέργειας σε μικροεπεξεργαστές. Βρήκε ένα υποσύνολο μετρητών οι οποίοι ήταν κατάλληλοι για πρόβλεψη κατανάλωσης ενέργειας δια μέσου πολλαπλών αρχιτεκτονικών, με σχετικό σφάλμα 5%. Αυτό επιτεύχθηκε χρησιμοποιώντας έναν υψηλής απόδοσης (HPerf) και ένα χαμηλής απόδοσης (LP) πυρήνα. Με το σχετικό σφάλμα στο 5%, ο προτινόμενος εκτιμητής είναι αρκετά βελτιωμένος σε σχέση με άλλων ερευνών με αντίστοιχα σφάλματα 9% [Singh et al., 2009] και 11.5% [Li and John, 2003]. Κατάφερε να δείξει πως για μικρές με μεσαίες παραλλαγές στην αρχιτεκτονική, ο ίδιος εκτιμητής ενέργειας μπορεί να χρησιμοποιηθεί σε διάφορες αρχιτεκτονικές, με μια μικρή απώλεια ακρίβειας της τάξης του 3%.

Ο K. Singh [Singh et al., 2009] χρησιμοποίησε ένα γραμμικό μοντέλο το οποίο χαρτογραφεί μετρητές απόδοσης για τη κατανάλωση ενέργειας. Κατάφερε να προ-

βλέπει με ακρίβεια την κατανάλωση ενέργειας γράφοντας και χρησιμοποιώντας δικά του benchmarks, ώστε να παράγει δεδομένα για να δημιουργήσει το μοντέλο πρόβλεψης. Τα benchmarks στρεσάρουν τέσσερις μετρητές, οι οποίοι επιλέγονται βάσει της συσχετιζόμενης τιμής τους. Εκμεταλεύτηκε το ενεργειακό του μοντέλο για να εκτελέσει ένα προγραμματισμό νήματος. Σταματούσε και συνέχιζε διαδικασίες με βάση τη κατανάλωση ενέργειας τους, εξασφαλίζοντας πως το περιβάλλον ενέργειας δεν υπερβαίνεται. Η έρευνά του είναι εξαιρετικά χρήσιμη για ενοποιημένα κέντρα δεδομένων, στα οποία η εικονοποίηση οδηγεί σε πολλαπλούς διακομιστές. Χρησιμοποιώντας την εκτιμώμενη μεθοδολογία του, προβλέπει την κατανάλωση ενέργειας με ακρίβεια, επιτρέποντας ακριβή τιμολόγηση της χρήσης ενέργειας και των κόστων ψύξης. Παρατηρήθηκε πως η πρόβλεψη κατανάλωσης ενέργειας ανά πυρήνα είναι δύσκολη, αφού κάποιοι πόροι διαμοιράζονται ανάμεσα στους πυρήνες όπως οι κρυφές μνήμες και ο ελεγκτής δυναμικής μνήμης τυχαίας προσπέλασης.

Στην έρευνά του, ο M Dayarathma et al [Dayarathna et al., 2015], ανέλυσε την ενεργειακή ιεραρχία ενός κέντρου δεδομένων σε επίπεδα. Αφού χώρισε τα στοιχεία ενός κέντρου δεδομένων σε υλικό και λογισμικό, ανέλυσε πάνω από 200 σύγχρονα μοντέλα ενέργειας σε διαφορετικά επίπεδα, ιεραρχικά από κάτω προς τα πάνω. Παρατηρήθηκε πως υπήρχε πολύ μεγάλος αριθμός ερευνών στην κατανάλωση ενέργειας σε χαμηλότερα ιεραρχικά επίπεδα, ενώ πολύ λιγότερες στα υψηλότερα επίπεδα.

Ο M Rodriguez et al [Rodriguez-Martinez et al., 2011] ασχολήθηκε με τη πρόβλεψη κατανάλωσης ενέργειας των queries σε βάσεις δεδομένων. Η μεθοδολογία που χρησιμοποιήθηκε βασίστηκε σε πολλαπλή γραμμική παλινδρόμηση και παραγοντικά πειράματα. Το σημαντικό κομμάτι της μεθοδολογίας είναι ότι δεν χρειάζεται να παρθούν μετρήσεις σε ξεχωριστά εξαρτήματα υλικού. Αντί αυτού, τα δεδομένα αντλήθηκαν από ενεργειακές μετρήσεις από εσωτερικούς σένσορες, καθώς και από έτοιμα στοιχεία όπως το μέγεθος των πλειάδων, οι αριθμοί των στηλών και ο αριθμός των διακομιστών. Παρατηρήθηκε πως οι προβλέψεις ήταν πιο ακριβείς από άλλα εναλλακτικά μοντέλα και πως η μέγιστη ενέργεια επηρεάζεται από την εκλεκτικότητα, τον αριθμό των στηλών στον πίνακα και τον αριθμό των διακομιστών στο σύστημα.

Κεφάλαιο 3

Αλγόριθμοι Μηχανικής Μάθησης

3.1 Εισαγωγή στη Μηχανική Μάθηση

Η μηχανική μάθηση είναι ένα πεδίο έρευνας το οποίο διασταυρώνει κλάδους όπως η επιστήμη των υπολογιστών, η στατιστική και η τεχνητή νοημοσύνη και ασχολείται κυρίως με την εξαγωγή γνώσεων από δεδομένα. Η εφαρμογή της μηχανικής μάθησης έχει εξαπλωθεί ευρέως τα τελευταία χρόνια στην καθημερινή ζωή. Από αυτοματοποιημένες προτάσεις για το τί ταινία να δει κάποιος, τί φαί να παραγγείλει ή τί προϊόντα να αγοράσει, μέχρι την αναγνώριση των γνωστών του προσώπων στις φωτογραφίες, πάρα πολλές σύγχρονες ιστοσελίδες και συσκευές έχουν στον πυρήνα τους αλγόριθμους βασισμένους στη μηχανική μάθηση. Σύνθετες ιστοσελίδες όπως το Facebook, η Google, το Amazon και το Netflix περιέχουν πολύπλοκα μοντέλα μηχανικής μάθησης [Müller et al., 2016].

Τις τελευταίες δύο δεκαετίες, η μηχανική μάθηση έχει γίνει ένα κοινό εργαλείο σε κάθε σχεδόν εργασία, η οποία απαιτεί εξαγωγή πληροφορίας από μεγάλα σύνολα δεδομένων. Καθημερινά είμαστε περικυκλωμένοι από τεχνολογία βασισμένη στην μηχανική μάθηση. Μηχανές αναζήτησης μαθαίνουν το πώς να μας παρέχουν το βέλτιστο αποτέλεσμα, λογισμικό μαθαίνει να φιλτράρει την αλληλογραφία και να ξεχωρίζει την ανεπιθύμητη και οι συναλλαγές μέσω πιστωτικών καρτών προστατεύονται από λογισμικό το οποίο μαθαίνει να ξεχωρίζει απάτες. Ψηφιακές κάμερες μαθαίνουν να ξεχωρίζουν πρόσωπα, εφαρμογές έξυπνων προσωπικών βοηθών στα κινητά μαθαίνουν να αναγνωρίζουν ηχητικές εντολές και αυτοκίνητα έχουν εξοπλιστεί με συστήματα πρόληψης ατυχημάτων, τα οποία βασίζονται σε αλγόριθμους μηχανικής μάθησης. Εκτός από τις εμπορικές εφαρμογές, η μηχανική μάθηση έχει

επηρεάσει σημαντικά επιστημονικές εφαρμογές όπως η βιοπληροφορική, η ιατρική και η αστρονομία [Shalev-Shwartz and Ben-David, 2014].

Ένα κοινό στοιχείο των παραπάνω εφαρμογών είναι ότι, σε αντίθεση με τις περισσότερες παραδοσιακές χρήσεις των υπολογιστών, σε αυτές τις περιπτώσεις, εξαιτίας της πολυπλοκότητας των μοτίβων που χρειάζεται να ανιχνευθούν, ένας προγραμματιστής δεν μπορεί να προσδιορίσει με σαφήνεια και λεπτομέρεια το πώς θα εκτελεστούν. Παίρνοντας παράδειγμα από τους ανθρώπους, πολλές από τις δεξιότητές μας αποκτώνται μέσω εμπειρίας και όχι ακολουθώντας σαφείς οδηγίες που μας δόθηκαν [Shalev-Shwartz and Ben-David, 2014].

Ιστορικά, η πρώτη αναφορά νευρωνικού δικτύου έγινε το 1943 [Mac, 2019], όταν ο νευροφυσιολόγος Warren McCulloch και ο μαθηματικός Walter Pitts έγραψαν ένα άρθρο για τους νευρώνες και το πώς λειτουργούν. Αποφάσισαν να δημιουργήσουν ένα μοντέλο χρησιμοποιώντας ένα ηλεκτρικό κύκλωμα, και έτσι γεννήθηκε το πρώτο νευρωνικό δίκτυο. Το 1950 [Mac, 2019], ο Alan Turing δημιούργησε το διάσημο τεστ του Turing. Για να περάσει επιτυχώς ένας υπολογιστής το παραπάνω τεστ, θα έπρεπε να είναι ικανός να πείσει έναν άνθρωπο πως ο ίδιος ο υπολογιστής είναι άνθρωπος. Το 1952 [Mac, 2019], ένα υπολογιστικό πρόγραμμα από τον Arthur Samuel το οποίο έπαιζε ντάμα, ήταν το πρώτο πρόγραμμα το οποίο μπορούσε να μαθαίνει ενόσω έτρεχε. Το 1958 [Mac, 2019] ο Frank Rosenblatt σχεδίασε το πρώτο τεχνητό νευρωνικό δίκτυο, το Perceptron, του οποίου κύριως στόχος ήταν αναγνώριση μοτίβων και σχημάτων. Ένα ακόμη παράδειγμα πρώιμου νευρωνικού δικτύου ήρθε το 1959 [Mac, 2019], όταν οι Bernard Widrow και Marcian Hoff δημιούργησαν δύο μοντέλα στο πανεπιστήμιο του Stanford. Το πρώτο λεγόταν ADELIN και μπορούσε να ανιχνεύσει διαδικά μοτίβα. Για παράδειγμα, σε μια ροή από bits, μπορούσε να προβλέψει πιο θα είναι το επόμενο. Το δεύτερο μοντέλο λεγόταν MADELINE και μπορούσε να εξαλείψει την ηχώ στις τηλεφωνικές γραμμές, πράγμα που έκανε την εφαρμογή του εξαιρετικά χρήσιμη. Παρά την επιτυχία του μοντέλου MADELINE, δεν υπήρξε πολύ μεγάλη πρόοδος μέχρι τα τέλη του 1970 [Mac, 2019], για πολλούς λόγους, κυρίως λόγω της δημοτικότητας της αρχιτεκτονικής Von Neumann. Στην εν λόγω αρχιτεκτονική οι εντολές και τα δεδομένα αποθηκεύονται στην ίδια μνήμη, το οποίο είναι ασφαλώς πιο κατανοητό από ένα νευρωνικό δίκτυο, έτσι πολλοί προγραμματίζαν βασιζόμενοι σε αυτή.

Το ενδιαφέρον για τα νευρωνικά δίκτυα αναζωπυρώθηκε ξανά το 1982 [Mac, 2019], όταν ο John Hopfield πρότεινε ένα δίκτυο το οποίο είχε αμφίδρομες γραμμές, παρόμοια με το πως λειτουργούν πραγματικά οι νευρώνες. Επιπλέον, την ίδια χρονιά, η Ιαπωνία ανακοίνωσε πως επικεντρωνώταν σε πιο προχωρημένα νευρωνικά δίκτυα, πράγμα που ώθησε και την Αμερική να μπει και αυτή δυναμικά στην έρευνα.

Τα νευρωνικά δίκτυα χρησιμοποιούν την προς τα πίσω διάδοση (back propagation), η οποία ήρθε στο προσκήνιο το 1986 [Mac, 2019], όταν τρεις επιστήμονες από το τμήμα ψυχολογίας του Stanford αποφάσισαν να επεκτείνουν έναν αλγόριθμο των Widrow και Hoff από το 1962.

Στα τέλη του 1980 και το 1990 [Mac, 2019] δεν υπήρξε κάποια σημαντική προσθήκη στην έρευνα. Ωστόσο, το 1997 [Mac, 2019], ο υπολογιστής της IBM με όνομα Deep Blue, ο οποίος ήταν ένας υπολογιστής που έπαιζε σκάκι, νίκησε τον παγκόσμιο πρωταθλητή. Από τότε, έχει γίνει αρκετή πρόοδος στο πεδίο, όπως το 1998, όταν έρευνα στα εργαστήρια AT&T Bell πάνω στην αναγνώριση ψηφίου ήταν αρκετά αποτελεσματική στο να αναγνωρίζει χειρόγραφους ταχυδρομικούς κώδικες.

Από την αρχή του εικοστού πρώτου αιώνα, πολλές επιχειρήσεις συνειδητοποίησαν πως η μηχανική μάθηση θα αυξήσει την υπολογιστική δυνατότητα. Το 2010 [Marr, 2016] η Microsoft κυκλοφορεί το Kinect, το οποίο μπορεί να ανιχνεύσει 20 ανθρώπινα χαρακτηριστικά 30 φορές το δευτερόλεπτο, επιτρέποντας στους ανθρώπους να αλληλεπιδρούν με τον υπολογιστή μέσω κινήσεων και χειρονομιών. Το 2012 [Marr, 2016] η Google αναπτύσσει το GoogleBrain, ένα βαθύ νευρωνικό δίκτυο το οποίο στοχεύει σε ανίχνευση μοτίβων σε εικόνες και βίντεο. Το 2014 [Marr, 2016] το Facebook ανέπτυξε το Deepface, ένα βαθύ νευρωνικό δίκτυο το οποίο μπορεί και αναγνωρίζει άτομα σε φωτογραφίες, με την ίδια ευκολία όπως ένας άνθρωπος. Την ίδια χρονιά [BBC, 2018], η εταιρία DeepMind, έγινε γνωστή με την ανάπτυξη ενός νευρωνικού δικτύου το οποίο μπορούσε να παίξει βιντεοπαιχνίδια, αναλύοντας τη συμπεριφορά των πιξελς στην οθόνη. Ένα χρόνο αργότερα [BBC, 2018], η ίδια εταιρία αναπτύσσει το AlphaGo, ένα λογισμικό το οποίο μπορεί να παίξει το κινέζικο επιτραπέζιο παιχνίδι γκο (θεωρείται από πολλούς από τα πιο πολύπλοκα επιτραπέζια), νικώντας τον καλύτερο παίκτη στον κόσμο.

3.2 Τύποι Μηχανικής Μάθησης

Εντός του επιστημονικού πεδίου του μηχανικής μάθησης υπάρχουν δύο κύριες κατηγορίες, η επιβλεπόμενη μάθηση και η μη-επιβλεπόμενη μάθηση. Στην επιβλεπόμενη μάθηση, ο χρήστης παρέχει στον αλγόριθμο ζευγάρια δεδομένων εισόδου/εξόδου, και ο αλγόριθμος βρίσκει τρόπο να παράξει την επιθυμητή έξοδο, έχοντας ως δεδομένο μια είσοδο. Συγκεκριμένα, ο αλγόριθμος είναι ικανός να δημιουργήσει την επιθυμητή έξοδο από μια είσοδο που δεν έχει ξαναδεί, χωρίς καμιά βοήθεια από τον άνθρωπο. Αντίθετα, στη μη-επιβλεπόμενη μάθηση, δεν είναι γνωστά τα δεδομένα εξόδου, αλλά μόνο τα δεδομένα εισόδου. Αν και υπάρχουν πολλές επιτυχημένες εφαρμογές της μη-επιβλεπόμενης μάθησης, συνήθως, φέρουν δυσκολία στην κατανόηση και την αξιολόγηση [Müller et al., 2016].

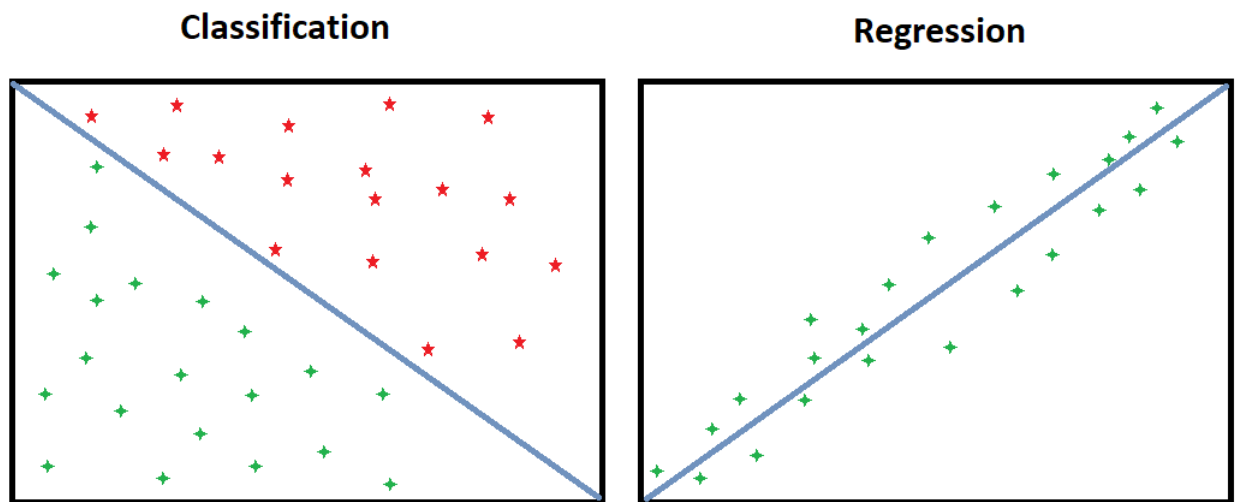
3.2.1 Επιβλεπόμενη Μάθηση

Η επιβλεπόμενη μάθηση εφαρμόζεται με την ταξινόμηση (σχήμα 3.1), όταν θέλουμε να χαρτογραφήσουμε δεδομένα εισόδου σε ετικέτες δεδομένων εξόδου, ή την παλινδρόμηση (σχήμα 3.1), όταν θέλουμε να χαρτογραφήσουμε δεδομένα εισόδου σε μια συνεχή έξοδο. Γνωστοί αλγόριθμοι που συναντάμε στην επιβλεπόμενη μάθηση είναι η γραμμική παλινδρόμηση (linear regression), ο αλγόριθμος του Naive Bayes, τα support vector machines, τα τεχνητά νευρωνικά δίκτυα (artificial neural networks), τα random forests κα. Στη ταξινόμηση καθώς και στη παλινδρόμηση, ο στόχος είναι να βρεθούν συγκεκριμένες σχέσεις ή δομές στα δεδομένα εισόδου, οι οποίες θα μας βοηθήσουν να παράγουμε αποτελεσματικά τα σωστά δεδομένα εξόδου. Πρέπει να σημειωθεί πως τα σωστά δεδομένα εξόδου προσδιορίζονται σε μεγάλο βαθμό από τα δεδομένα εκπαίδευσης (training data), συμπερασματικά θορυβώδη ή λάθος δεδομένα θα μειώσουν την αποτελεσματικότητα του μοντέλου [Soni, 2018].

Όταν χρησιμοποιούμε την επιβλεπόμενη μάθηση, οι κύριοι παράγοντες, οι οποίοι συσχετίζονται μεταξύ τους, είναι η πολυπλοκότητα του μοντέλου και η αντιστάθμιση κλίσης-διακύμανσης.

Η πολυπλοκότητα του μοντέλου αναφέρεται στη πολυπλοκότητα της συνάρτησης που προσπαθεί να μάθει ο αλγόριθμος. Ο ειδικός βαθμός της πολυπλοκότητας μοντέλου προσδιορίζεται από τη φύση των δεδομένων εκπαίδευσης. Αν το μέγεθος

Σχήμα 3.1: Αναπαράσταση ταξινόμησης (αριστερά) και παλινδρόμησης (δεξιά) στη μηχανική μάθηση



των δεδομένων είναι μικρό, ή αν τα δεδομένα δεν καλύπτουν μεγάλο εύρος, θα πρέπει να επιλεχθεί ένα μοντέλο χαμηλής πολυπλοκότητας, διότι σε διαφορετική περίπτωση θα υπάρξει υπερφόρτωση (overfit). Η υπερφόρτωση αναφέρεται στο να μαθαίνεις μια λειτουργία η οποία ταιριάζει πολύ καλά στα δεδομένα εκπαίδευσης, αλλά δεν γενικοποιεί σε άλλα δεδομένα [Soni, 2018].

Η αντιστάθμιση κλίσης-διακύμανσης σχετίζεται άρρηκτα με τη γενικοποίηση του μοντέλου. Σε κάθε μοντέλο, υπάρχει μια ισορροπία μεταξύ της κλίσης, η οποία είναι το σταθερό σφάλμα, και της διακύμανσης, η οποία είναι το ποσό που το σφάλμα μπορεί να διαφέρει μεταξύ διαφορετικών σετ δεδομένων εκπαίδευσης. Άρα, ένα μοντέλο με υψηλή κλίση και χαμηλή διακύμανση θα προέβλεπε λάθος το 20% των περιπτώσεων, ενώ ένα μοντέλο με χαμηλή κλίση και υψηλή διακύμανση θα προέβλεπε λάθος 5% - 50% των περιπτώσεων, με μεγάλη εξάρτηση βέβαια από τα δεδομένα που χρησιμοποιήθηκαν για την εκπαίδευση. Πρέπει να σημειωθεί πως γενικά η κλίση και η διακύμανση κινούνται σε αντίθετες κατευθύνσεις. Αυξάνοντας την κλίση συνήθως οδηγεί σε μικρότερη διακύμανση και το αντίστροφο. Καθώς δημιουργείται ένα μοντέλο, η φύση του προβλήματος καθώς και η φύση των δεδομένων είναι αυτές που θα προσδιορίσουν το πού θα εστιάσει στο φάσμα της κλίσης-διακύμανσης. Γενικά, αυξάνοντας την κλίση δημιουργούνται μοντέλα με σχετικά εγγυημένη απόδοση, το οποίο σε συγκεκριμένες εργασίες είναι πολύ σημαντικό. Επιπροσθέτως, για να δημιουργηθούν μοντέλα τα οποία γενικοποιούν σε καλό βαθμό, η διακύμανση του μοντέλου θα πρέπει να κλιμακώνεται μαζί με το μέγεθος και τη πολυπλοκότητα

των δεδομένων εκπαίδευσης. Μικρά και απλά σετ δεδομένων θα πρέπει να εφαρμόζονται σε μικρής διακύμανσης μοντέλα, ενώ μεγάλα και σύνθετα σετ δεδομένων συνήθως απαιτούν υψηλής διακύμανσης μοντέλα [Soni, 2018].

3.2.2 Μη-επιβλεπόμενη Μάθηση

Η πιο συνηθισμένες λειτουργίες της μη-επιβλεπόμενης μάθησης είναι η ομαδοποίηση (clustering), η αναπαραστασιακή μάθηση (representation learning), και η εκτίμηση πυκνότητας (density estimation). Από τους πιο γνωστούς αλγόριθμους ξεχωρίζουν ο αλγόριθμος ομαδοποίησης k-means, η ανάλυση κύριου στοιχείου (principal component analysis) και οι autoencoders. Από τη στιγμή που τα δεδομένα δεν έχουν ετικέτες, στις περισσότερες μη-επιβλεπόμενες μεθόδους, δεν υπάρχει συγκεκριμένη μέθοδος για να συγκρίνουμε την απόδοση του μοντέλου [Soni, 2018].

Δύο συνήθης χρήσεις της μη-επιβλεπόμενης μάθησης είναι η διερευνητική ανάλυση και η μείωση των διαστάσεων.

Η μη-επιβλεπόμενη μάθηση είναι αρκετά χρήσιμη στην διερευνητική ανάλυση επειδή μπορεί να αναγνωρίσει αυτόματα μία δομή σε δεδομένα. Σε καταστάσεις όπου είναι αδύνατο ή μη πρακτικό για έναν άνθρωπο να διακρίνει μοτίβα σε δεδομένα, η μη-επιβλεπόμενη μάθηση μπορεί να παρέχει πληροφορίες οι οποίες μπορούν να χρησιμοποιηθούν για να τεστάρουν ξεχωριστές υποθέσεις [Soni, 2018].

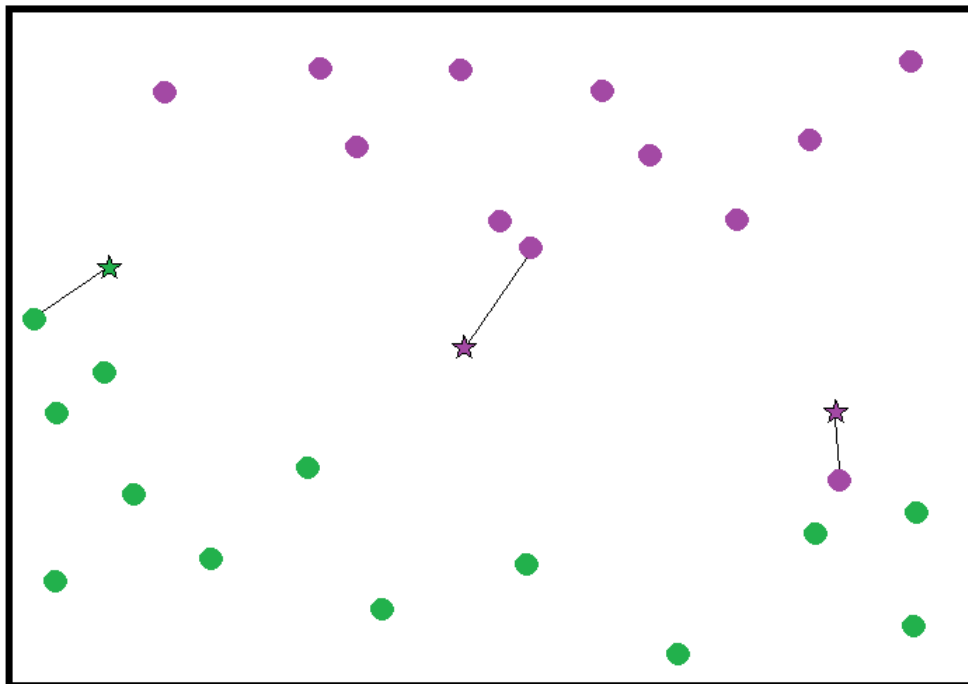
Η μείωση των διαστάσεων, η οποία αναφέρεται στις μεθόδους που χρησιμοποιούνται για την παρουσίαση δεδομένων χρησιμοποιώντας λιγότερες στήλες ή χαρακτηριστικά, μπορεί να πραγματοποιηθεί μέσω μη-επιβλεπόμενων μεθόδων. Στην αναπαραστασιακή μάθηση, προσπαθούμε να μάθουμε σχέσεις μεταξύ ξεχωριστών χαρακτηριστικών, επιτρέποντας να παρουσιάσουμε τα δεδομένα μας χρησιμοποιώντας τα λανθάνων χαρακτηριστικά, τα οποία αλληλοσυνδέονται με τα αρχικά. Αυτή η σπάνια λανθάνων δομή συχνά αναπαριστάται χρησιμοποιώντας αρκετά λιγότερα χαρακτηριστικά από όσα ήταν αρχικά, ώστε να επεξεργάζεται περαιτέρω δεδομένα πιο εύκολα, και να μπορεί να εξαλείφει περιττά χαρακτηριστικά [Soni, 2018].

3.3 Αλγόριθμοι Μηχανικής Μάθησης

Κ-Πλησιέστεροι Γείτονες

Ο αλγόριθμος K-Πλησιέστερων Γειτόνων είναι αναμφισβήτητα ο πιο απλός αλγόριθμος μηχανικής μάθησης. Το χτίσιμο του μοντέλου αποτελείται μόνο την εισαγωγή των δεδομένων εκπαίδευσης. Για να γίνει μια πρόβλεψη για ένα νέο δεδομένο εισόδου, ο αλγόριθμος βρίσκει τους κοντινότερους του 'γείτονες' στο σετ των δεδομένων εκπαίδευσης. Στην πιο απλή του μορφή, ο αλγόριθμος k-NN εξετάζει μόνο ένα γείτονα, αυτόν τον οποίο είναι πιο κοντά στο καινούργιο δεδομένο εισόδου [Müller et al., 2016].

Σχήμα 3.2: Προβλέψεις μοντέλου k-NN με 1 γείτονα

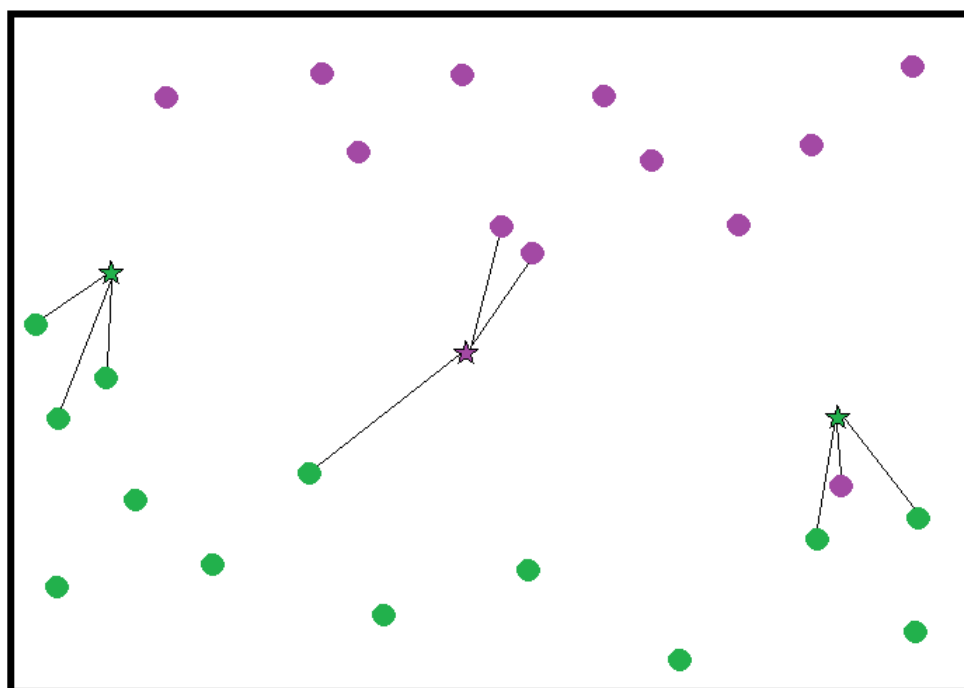


Στο σχήμα 3.2, τα αστέρια απεικονίζουν τρία νέα δεδομένα εισόδου. Κάθε ένα από αυτά ενώνεται με τον πλησιέστερό γείτονά του με μια γραμμή. Η πρόβλεψη των νέων δεδομένων εισόδου από τον k-NN αλγόριθμο με έναν γείτονα απεικονίζεται από τα αντίστοιχα χρώματα των αστεριών.

Αντί να εξεταστεί μόνο ο πλησιέστερος γείτονας, μπορεί να εξεταστεί αυθαίρετος αριθμός γειτόνων. Αν ληφθεί υπόψιν παραπάνω από ένας γείτονας, για κάθε νέο δεδομένο εισόδου, γίνεται καταμέτρηση των γειτόνων που ανήκουν στην πράσινη και στη μωβ κλάση. Έπειτα αναθέτουμε το νέο δεδομένο στη κλάση που εμφανίζεται πιο συχνά.

Όπως παρατηρείται στο σχήμα 3.3, χρησιμοποιώντας τρεις πλησιέστερους γείτονες, η πρόβλεψη για το κάτω δεξιά νέο δεδομένο δεν είναι η ίδια με την πρόβλεψη

Σχήμα 3.3: Προβλέψεις μοντέλου k-NN με 3 γείτονες



η οποία έγινε λαμβάνοντας υπόψιν ένα γείτονα.

Δύο είναι οι σημαντικότερες παράμετροι στον αλγόριθμο k-NN, ο αριθμός των γειτόνων με τη μεταβλητή $n_{neighbors}$ και το πώς γίνεται η μέτρηση της απόστασης μεταξύ των δεδομένων. Στην πράξη, χρησιμοποιώντας ένα μικρό αριθμό πλησιέστερων γειτόνων (3-5) συχνά λειτουργεί καλά, αλλά θα πρέπει σίγουρα για κάθε πρόβλημα να ελέγχεται εμπειρικά. Η απόσταση των σημείων μετριέται εξ ορισμού με την Ευκλείδεια απόσταση, η οποία λειτουργεί καλά στα περισσότερα προβλήματα [Müller et al., 2016].

Ένα από τα πλεονεκτήματα του k-NN είναι ότι το μοντέλο του γίνεται πολύ εύκολα αντιληπτό και συχνά προσφέρει λογική επίδοση, χωρίς πολλές προσαρμογές. Η χρήση του συγκεκριμένου αλγορίθμου είναι μια αρχή, πριν εφαρμοστούν πιο προχωρημένες τεχνικές. Το χτίσιμο του μοντέλου των πλησιέστερων γειτόνων είναι συνήθως αρκετά γρήγορο, αλλά όταν το σετ δεδομένων εκπαίδευσης είναι πολύ μεγάλο (αριθμός των χαρακτηριστικών ή αριθμός δειγμάτων) η πρόβλεψη μπορεί να είναι αργή [Müller et al., 2016].

Γραμμική Παλινδρόμηση

Τα γραμμικά μοντέλα είναι μια κατηγορία μοντέλων τα οποία χρησιμοποιούνται

ευρέως στη πράξη και έχουν μελετηθεί εκτεταμένα τις τελευταίες δεκαετίες, με τις ρίζες τους να φθάνουν μέχρι και 100 χρόνια πίσω. Τα γραμμικά μοντέλα κάνουν μια πρόβλεψη με τη βοήθεια μιας γραμμικής συνάρτησης [Müller et al., 2016].

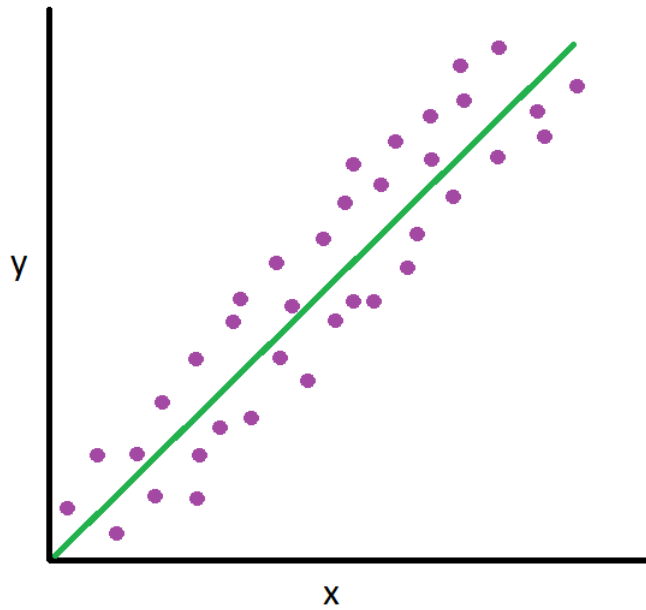
Για τη παλινδρόμηση, η γενική πρόβλεψη για ένα γραμμικό μοντέλο ακολουθεί την εξίσωση:

$$y = w[0] * x[0] + w[1] * x[1] + \dots + w[p] * x[p] + b$$

Γενικά, ένα γραμμικό μοντέλο κάνει μια πρόβλεψη υπολογίζοντας ένα σταθμισμένο άθροισμα χαρακτηριστικών εισόδου, συν μια σταθερά που ονομάζεται κλίση [Géron, 2017]. Στην παραπάνω εξίσωση, τα $x[0] - x[p]$ υποδηλώνουν τα χαρακτηριστικά ενός δεδομένου, τα w και b είναι παράμετροι του μοντέλου και το y είναι η πρόβλεψη που πραγματοποιεί το μοντέλο. Για ένα σετ δεδομένων το οποίο αποτελείται από ένα χαρακτηριστικό, η εξίσωση γίνεται:

$$y = w[0] * x[0] + b$$

Σχήμα 3.4: Αναπαράσταση Γραμμική Παλινδρόμηση



Η Γραμμική Παλινδρόμηση (σχήμα 3.4), ή αλλιώς μέθοδος ελαχίστων τετραγώνων, είναι η πιο απλή και κλασσική γραμμική μέθοδος για παλινδρόμηση. Για να εκπαιδευτεί ένα μοντέλο γραμμικής παλινδρόμησης, πρέπει να βρεθούν οι τιμές των w και b , οι οποίες ελαχιστοποιούν το μέσο τετραγωνικό σφάλμα μεταξύ των προβλέψεων και των πραγματικών τιμών. Το μέσο τετραγωνικό σφάλμα είναι το άθροισμα των τετραγωνικών διαφορών μεταξύ των προβλέψεων και των πραγματικών τιμών [Géron, 2017].

$$\sum_1^n (pred_i - y_i)^2$$

Η γραμμική παλινδρόμηση δε διαθέτει παραμέτρους, πράγμα το οποίο την κάνει πιο εύκολη στην κατανόηση, αλλά παράλληλα δεν υπάρχει τρόπος να ελεγχθεί η πολυπλοκότητα του μοντέλου [Müller et al., 2016].

Παλινδρόμηση Ridge

Ένα ακόμη μοντέλο για παλινδρόμηση είναι το Ridge Regression (γνωστό και ως κανονικοποίηση Tikhonov), το οποίο είναι μια κανονικοποιημένη έκδοση των ελαχίστων τετραγώνων, με τη διαφορά ότι εισάγεται ένας παράγοντας κανονικοποίησης στη συνάρτηση κόστους. Στη παλινδρόμηση ridge, οι συντελεστές w επιλέγονται όχι μόνο για να προβλέπουν καλά στα δεδομένα εκπαίδευσης, αλλά και για να ταιριάζουν

ξουν έναν επιπλέον περιορισμό. Θέλουμε το μέγεθος των συντελεστών να είναι όσο μικρότερο δυνατόν [Géron, 2017]. Δηλαδή, όλες οι τιμές των w θα πρέπει να είναι κοντά στο μηδέν. Αυτό σημαίνει πως κάθε χαρακτηριστικό θα πρέπει να έχει όσο μικρότερη επιρροή γίνεται στο αποτέλεσμα, ενώ το μοντέλο θα πρέπει να προβλέπει ικανοποιητικά. Αυτός ο περιορισμός είναι ένα παράδειγμα της κανονικοποίησης, η οποία πρακτικά σημαίνει ρητός περιορισμός του μοντέλου για την αποφυγή υπερφόρτωσης. Η συγκεκριμένη κανονικοποίηση η οποία χρησιμοποιείται από την παλινδρόμηση ridge είναι γνωστή ως L2 κανονικοποίηση [Müller et al., 2016].

Το μοντέλο παλινδρόμησης ridge διατηρεί την αναλογία μεταξύ της απλότητας του μοντέλου και της επίδοσής του στο σετ δεδομένων εκπαίδευσης. Η σημασία που θέτει το μοντέλο στην απλότητα ή την επίδοσή του προσδιορίζεται από την παράμετρο $alpha$. Η βέλτιστη τροποποίηση της παραμέτρου αυτής εξαρτάται από το σετ δεδομένων που χρησιμοποιείται. Αυξάνοντας τη παράμετρος $alpha$, οι συντελεστές ωθούνται στο μηδέν, το οποίο μειώνει την επίδοση στο σετ δεδομένων εκπαίδευσης, αλλά μπορεί να βοηθήσει στην γενίκευση. Μειώνοντας τη παράμετρο $alpha$, οι συντελεστές υφίστανται μικρότερο περιορισμό. Για πολύ μικρές τιμές της παραμέτρου $alpha$, οι συντελεστές μετά βίας περιορίζονται, οπότε καταλήγουμε σε μοντέλο αρκετά παρόμοιο με αυτό των ελαχίστων τετραγώνων [Müller et al., 2016].

Παλινδρόμηση Lasso

Μια εναλλακτική κανονικοποιημένη εκδοχή της γραμμικής παλινδρόμησης είναι η παλινδρόμηση lasso (Least Absolute Shrinkage Operator Regression). Όπως και με τη παλινδρόμηση ridge, χρησιμοποιώντας τη παλινδρόμηση lasso επίσης περιορίζει τους συντελεστές κοντά στο μηδέν, αλλά με ένα διαφορετικό τρόπο, ο οποίος λέγεται κανονικοποίηση L1. Ένα σημαντικό χαρακτηριστικό της L1 κανονικοποίησης είναι πως τείνει να εξαλείψει εντελώς τα βάρη των λιγότερο σημαντικών χαρακτηριστικών, θέτωντάς τα μηδέν. Αυτό έχει ως αποτέλεσμα κάποια χαρακτηριστικά να αγνοούνται εντελώς από το μοντέλο, ερμηνεύοντας το παραπάνω φαινόμενο ως ένα τρόπο αυτόματης επιλογής χαρακτηριστικών και δίνοντας ως έξοδο ένα πιο αραιό μοντέλο [Géron, 2017]. Όταν κάποια χαρακτηριστικά έχουν την τιμή μηδέν συχνά το μοντέλο είναι ευκολότερο να ερμηνευτεί, και μπορεί να αποκαλήψει τα πιο σημαντικά χαρακτηριστικά του. Ομοίως με την παλινδρόμηση ridge, η παλινδρόμηση

lasso έχει μια παράμετρο κανονικοποίησης, την α , η οποία ελέγχει το ρυθμό με τον οποίο οι συντελεστές τείνουν προς το μηδέν.

Συμπερασματικά, η παλινδρόμηση ridge είναι συνήθως η πρώτη επιλογή από τα δύο τελευταία μοντέλα. Παρόλα αυτά, αν υπάρχει μεγάλο πλήθος χαρακτηριστικών και εικάζεται πως λίγα από αυτά θα είναι σημαντικά, η παλινδρόμηση lasso ίσως είναι καλύτερη επιλογή. Με την ίδια σκέψη, αν θέλουμε το μοντέλο να είναι εύκολο στην ερμηνεία, η παλινδρόμηση lasso θα παρέχει ένα μοντέλο το οποίο θα είναι εύκολο στην κατανόηση, αφού θα επιλέξει μόνο ένα υποσύνολο από τα χαρακτηριστικά. Στην πράξη, ο συνδιασμός των δύο παλινδρομήσεων lasso και ridge δουλεύει καλύτερα, με το τίμημα πως θα πρέπει να ρυθμιστούν δύο παράμετροι, ένας για την L1 κανονικοποίηση και ένας για την L2 κανονικοποίηση [Müller et al., 2016].

Μηχανές Διανυσμάτων Υποστήριξης

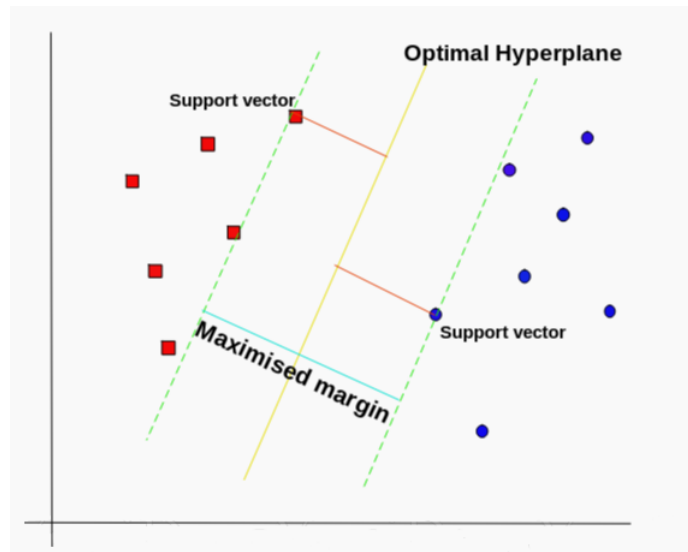
Το support vector machine είναι ένα γραμμικό μοντέλο το οποίο χρησιμοποιείται για προβλήματα ταξινόμησης και παλινδρόμησης, ενώ μπορεί να λύσει γραμμικά και μή-γραμμικά προβλήματα. Η γενική ιδέα του SVM είναι απλή, όπως φαίνεται στην σχήμα 3.5, ο αλγόριθμος δημιουργεί μια γραμμή ή ένα υπερεπίπεδο (hyperplane), το οποίο διαχωρίζει τα δεδομένα σε κλάσεις. Με τον όρο υπερεπίπεδο, σε έναν Ευκλειδιανό χώρο n -διαστάσεων, ορίζεται ένα επίπεδο υποσύνολο $n-1$ διαστάσεων του χώρου αυτού, το οποίο χωρίζει το χώρο σε δύο ασύνδετα μέρη [Pupale, 2018].

Με μια αρχική προσέγγιση, ο αλγόριθμος SVM λαμβάνει τα δεδομένα ως είσοδο και παράγει ως έξοδο μια γραμμή (ή υπερεπίπεδο) η οποία διαχωρίζει τις δύο κλάσεις, αν αυτό είναι δυνατό.

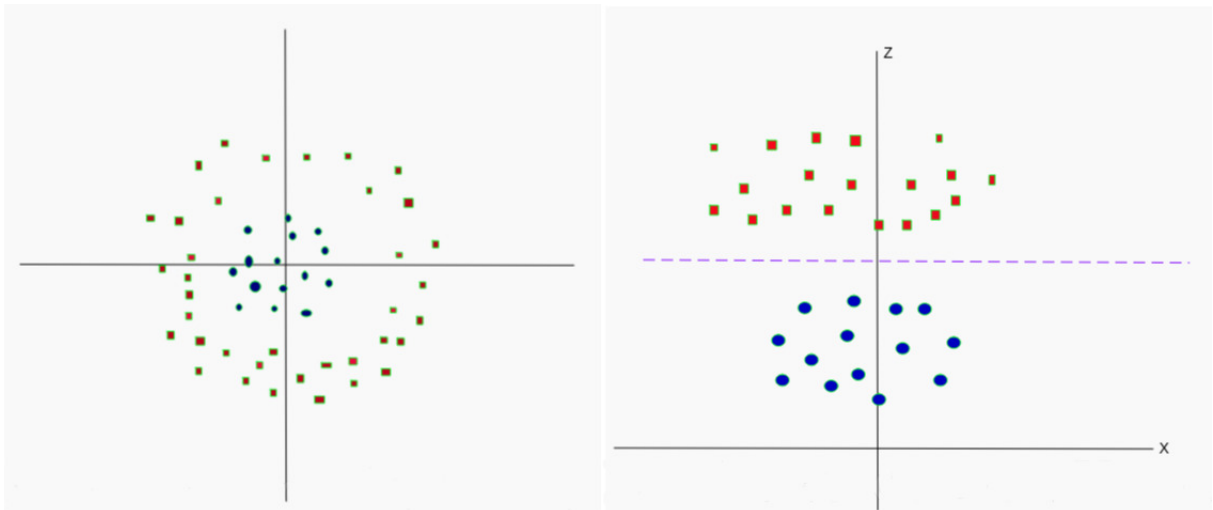
Σύμφωνα με τον αλγόριθμο SVM, βρίσκουμε τα σημεία των κλάσεων τα οποία είναι πιο κοντά στη γραμμή. Αυτά τα σημεία λέγονται βοηθητικοί δείκτες και η απόσταση αυτών λέγεται περιθώριο. Το βέλτιστο υπερεπίπεδο είναι αυτό του οποίου το περιθώριο είναι το μέγιστο δυνατό.

Γίνεται αντιληπτό πως τα δεδομένα στο (αριστερό) σχήμα 3.6 δεν είναι γραμμικά διαχωρίσιμα. Όπως είναι εμφανές, δεν είναι δυνατόν να υπάρξει μια ευθεία γραμμή η οποία θα μπορεί να ταξινομή τα δεδομένα. Παρόλα αυτά, τα δεδομένα μπορούν να μετατραπούν σε γραμμικά διαχωρίσιμα, προσθέτοντας μια ακόμη διάσταση (δεξιό σχήμα 3.6) [Pupale, 2018].

Σχήμα 3.5: Το βέλτιστο υπερεπίπεδο μέσω του αλγόριθμου SVM [Pupale, 2018]



Σχήμα 3.6: Αλγόριθμος SVM σε μη-γραμμικά δεδομένα [Pupale, 2018]



Έτσι, μπορούμε να ταξινομήσουμε μη-γραμμικά δεδομένα προσθέτοντας μια επιπλέον διάσταση ώστε να γίνουν γραμμικά διαχωρίσιμα, και έπειτα να τα προβάλουμε πίσω στην αρχική τους διάσταση χρησιμοποιώντας μαθηματικούς μετασχηματισμούς.

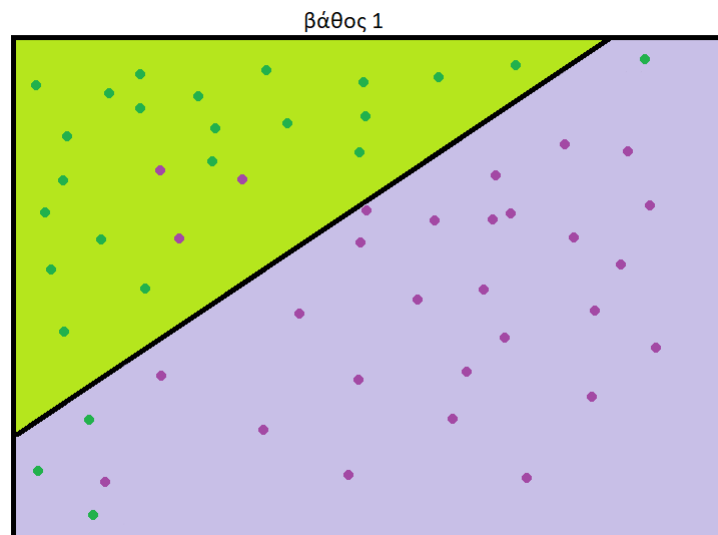
Ο αλγόριθμος SVM έχει δύο σημαντικές παραμέτρους, τις C και Γ . Η C μια παράμετρος κανονικοποίησης, παρόμοια με αυτές που χρησιμοποιούν τα γραμμικά μοντέλα, περιορίζει τη σημαντικότητα κάθε δεδομένου. Μια μικρή τιμή στη παράμετρο C σημαίνει πως το μοντέλο θα είναι πολύ περιορισμένο, δηλαδή το κάθε δεδομένο θα έχει πολύ μικρή επιρροή. Αυξάνοντας την τιμή της παραμέτρου C επιτρέπει στα δεδομένα να έχουν μεγαλύτερη επιρροή στο μοντέλο και κάμπει

το όριο απόφασης έτσι ώστε να τα ταξινομήσει όλα σωστά. Η παράμετρος *Gamma* ελέγχει το εύρος του Γκαουσιανού πυρήνα. Μια μικρή τιμή στη παράμετρο *Gamma* σημαίνει πως το όριο απόφασης θα είναι ομαλό, άρα το μοντέλο θα είναι σχετικά απλό. Αυξάνοντας τη παράμετρο *Gamma*, τα όρια προσπαθούν να συμπεριλάβουν όλα τα σημεία, δημιουργώντας ένα πιο σύνθετο μοντέλο [Müller et al., 2016].

Δέντρα Απόφασης

Τα δέντρα απόφασης είναι ευρέως χρησιμοποιήσιμα μοντέλα για προβλήματα ταξινόμησης και παλινδρόμησης. Ουσιαστικά, μαθαίνουν μια ιεραρχία αν/τότε ερωτήσεων, οι οποίες οδηγούν στη σωστή απάντηση, με τον πιο γρήγορο τρόπο. Κάθε κόμβος στο δέντρο αναπαριστά μια ερώτηση ή ένα τερματικό κόμβο (γνωστό και ως φύλλο) που περιέχει την απάντηση. Στη διάλεκτο της μηχανικής μάθησης, οι παραπάνω ερωτήσεις αποκαλούνται δοκιμές (δε θα πρέπει να μπερδεύονται με τα δεδομένα δοκιμών, τα οποία είναι τα δεδομένα τα οποία χρησιμοποιούμε για να δούμε πόσο καλά γενικεύει το μοντέλο) [Müller et al., 2016].

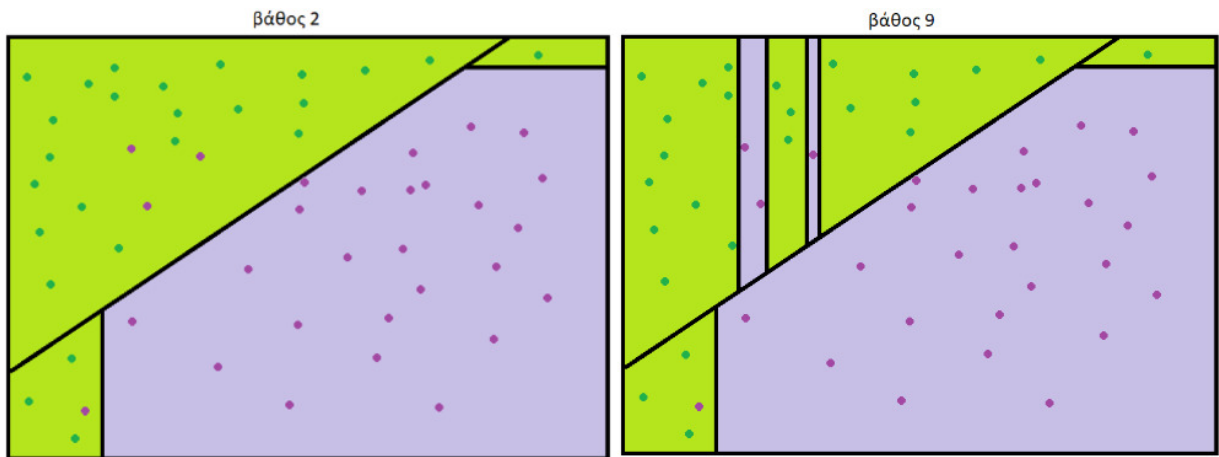
Σχήμα 3.7: Δέντρο απόφασης με βάθος 1



Για τη δημιουργία ενός δέντρου, ο αλγόριθμος ελέγχει όλες τις πιθανές δοκιμές και βρίσκει αυτή την οποία δίνει περισσότερες πληροφορίες για τη μεταβλητή στόχου. Το σχήμα 3.7 δείχνει τη πρώτη δοκιμή που επιλέχθηκε, η οποία διαχωρίζει σχετικά καλά τα δεδομένα της μωβ κλάσης από αυτά της πράσινης.

Αν και η πρώτη δοκιμή έκανε ικανοποιητική δουλειά στον διαχωρισμό των δύο κλάσεων, υπάρχουν ακόμη αρκετά δεδομένα τα οποία δεν ταξινομήθηκαν σωστά.

Σχήμα 3.8: Δέντρο απόφασης με βάθος 2 (αριστερά) και με βάθος 9 (δεξιά)



Όπως γίνεται αντιληπτό και στο (αριστερό) σχήμα 3.8, μπορεί να επιτευχθεί ένα πιο ακριβές μοντέλο επαναλαμβάνοντας τη διαδικασία αναζήτησης της καλύτερης δοκιμής. Αυτή η επαναλαμβανόμενη διαμέριση των δεδομένων συνεχίζεται μέχρι κάθε κλάση να περιλαμβάνει αποκλειστικά δικά της στοιχεία (δεξιό σχήμα 3.8). Κάθε φύλλο του δέντρου το οποίο περιέχει δεδομένα τα οποία μοιράζονται την ίδια μεταβλητή στόχου, ονομάζεται αγνό [Müller et al., 2016].

Συνήθως, χτίζοντας ένα δέντρο και συνεχίζοντας μέχρι όλα του τα φύλλα να είναι αγνά, οδηγεί σε πολύπλοκα μοντέλα και υπερφόρτωση. Η παρουσία αγνών φύλλων σημαίνει πως το δέντρο είναι 100% ακριβές στο σετ δεδομένων εκπαίδευσης. Υπάρχουν δύο γνωστές στρατηγικές για να αποτραπεί η υπερφόρτωση, το σταμάτημα της δημιουργίας του δέντρου από νωρίς (γνωστό και ως pre-pruning) ή να δημιουργηθεί το δέντρο και έπειτα να αφαιρεθούν οι καταρρέων κόμβοι (γνωστό και ως post-pruning) [Müller et al., 2016].

Τυχαία Δάση

Όπως παρατηρήθηκε, ένα σημαντικό μειονέκτημα των δέντρων αποφάσεων είναι ότι τείνουν να υπερφορτώνουν τα δεδομένα. Μία λύση σε αυτό το πρόβλημα μπορούν να δώσουν τα τυχαία δάση. Ένα τυχαίο δάσος είναι ένα σύνολο από δέντρα απόφασης, όπου κάθε δέντρο είναι ελαφρώς διαφορετικό από το άλλο. Η γενική ιδέα πίσω από τα τυχαία δάση είναι πως κάθε δέντρο μπορεί να προβλέψει σχετικά καλά, αλλά είναι πολύ πιθανό να υπερφορτώσει μέρος των δεδομένων. Αν δημιουργηθούν πολλά δέντρα, το καθένα από τα οποία θα προβλέπει καλά και θα

υπερφορτώνει με διαφορετικούς τρόπους, είναι δυνατό να μειωθεί η υπερφόρτωση παίρνοντας τον μέσο όρο των αποτελεσμάτων τους. Αυτή η μείωση στην υπερφόρτωση, διατηρώντας βέβαια την καλή πρόβλεψη των δέντρων, μπορεί να απεικονιστεί με τη χρήση αυστηρών μαθηματικών [Géron, 2017] [Müller et al., 2016].

Υπάρχουν δύο μέθοδοι κατα τις οποίες τα δέντρα σε ένα τυχαίο δάσος τυχαιοποιούνται, η επιλογή των δεδομένων τα οποία χρησιμοποιούνται για τη δημιουργία ενός δάσους και η επιλογή των χαρακτηριστικών για κάθε δοκιμή. Για το χτίσιμο ενός τυχαίου δάσους πρέπει πρώτα να αποφασιστεί ο αριθμός των δέντρων απόφασης που θα περιέχει (καθορίζεται από την παράμετρο $n_{estimators}$). Για το χτίσιμο ενός δέντρου, δημιουργείται νέο σετ δεδομένων το οποίο περιέχει κάποια ίδια στοιχεία και έχει το ίδιο μέγεθος με το αρχικό. Γίνεται αντιληπτό πως με την επανάληψη κάποιων στοιχείων στο καινούργιο σετ δεδομένων, κάποια άλλα θα λείπουν από αυτό (περίπου το $\frac{1}{3}$). Έπειτα το δέντρο απόφασης δημιουργείται βασιζόμενο στο καινούργιο σετ δεδομένων. Ωστόσο, ο αλγόριθμος, αντί να ψάχνει την καλύτερη δοκιμή για κάθε κόμβο, επιλέγει σε κάθε κόμβο τυχαία ένα υποσύνολο των χαρακτηριστικών, και ψάχνει την καλύτερη δυνατή δοκιμή στην οποία συμμετέχει ένα από τα παραπάνω χαρακτηριστικά. Ο αριθμός των χαρακτηριστικών που επιλέγονται ελέγχεται από την παράμετρο $max_features$. Αυτή η επιλογή υποσυνόλου χαρακτηριστικών επαναλαμβάνεται ξεχωριστά σε κάθε κόμβο, έτσι ώστε κάθε κόμβος σε ένα δέντρο απόφασης να προβλέπει χρησιμοποιώντας διαφορετικό υποσύνολο [Müller et al., 2016].

Η παραπάνω δειγματοληψία δεδομένων οδηγεί κάθε δέντρο απόφασης στο τυχαίο δάσος να έχει δημιουργηθεί πάνω σε διαφορετικό σετ δεδομένων. Εξαιτίας της επιλογής των χαρακτηριστικών σε κάθε κόμβο, κάθε διαχωρισμός σε κάθε δέντρο απόφασης λειτουργεί σε ξεχωριστό υποσύνολο χαρακτηριστικών. Μαζί, οι δύο παραπάνω μηχανισμοί εξασφαλίζουν πως όλα τα δέντρα απόφασης σε ένα τυχαίο δάσος είναι διαφορετικά [Müller et al., 2016].

Βαθιά Νευρωνικά Δίκτυα

Ένα νευρωνικό δίκτυο είναι μια σειρά αλγορίθμων η οποία προσπαθεί να αναγνωρίσει υποκείμενες σχέσεις σε ένα σετ δεδομένων, μέσω μιας διαδικασίας η οποία μιμείται τον τρόπο που λειτουργεί ο ανθρώπινος εγκέφαλος. Τα νευρωνικά δίκτυα μπορούν να θεωρηθούν ως μια γενίκευση των γραμμικών μοντέλων, πραγ-

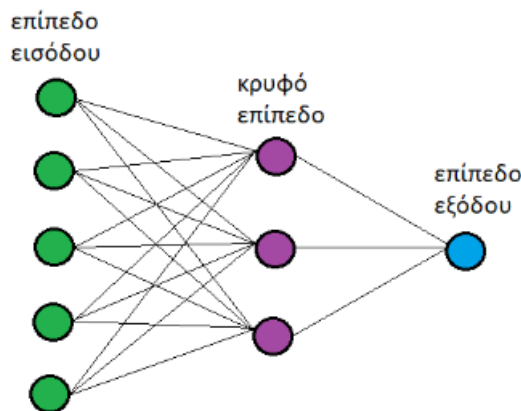
ματοποιώντας πολλαπλά στάδια επεξεργασίας για να φτάσουν σε μια απόφαση [Müller et al., 2016].

Όπως ειπώθηκε και παραπάνω, η πρόβλεψη από ένα γραμμικό μοντέλο δίνεται από την εξίσωση:

$$y = w[0] * x[0] + w[1] * x[1] + \dots + w[p] * x[p] + b$$

Στη παραπάνω εξίσωση, το y συμβολίζει ένα σταθμισμένο άθροισμα των χαρακτηριστικών εισόδου $x[0] - x[p]$, πολλαπλασιαζόμενα με τους συντελεστές $w[0] - w[p]$. Στα νευρωνικά δίκτυα, αυτή η διαδικασία του υπολογισμού σταθμισμένων αθροισμάτων επαναλαμβάνεται πολλές φορές, αρχικά υπολογίζοντας τις κρυφές μονάδες (hidden units), οι οποίες αντιπροσωπεύουν ένα ενδιάμεσο βήμα, και συνδιάζονται ξανά χρησιμοποιώντας σταθμισμένα αθροίσματα για να επιτευχθεί το τελικό αποτέλεσμα [Müller et al., 2016].

Σχήμα 3.9: Νευρωνικό δίκτυο με 1 κρυφό επίπεδο

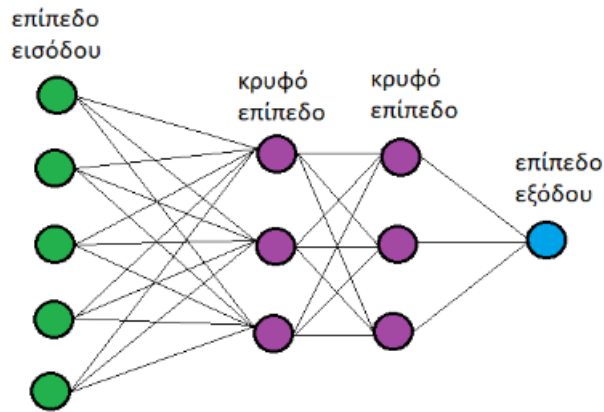


Όπως γίνεται αντιληπτό από το σχήμα 3.9, το συγκεκριμένο μοντέλο έχει πολλούς συντελεστές, γνωστοί και ως βάρη. Υπάρχει ένας συντελεστής μεταξύ κάθε εισόδου και κρυφής μονάδας (το οποίο δημιουργεί το κρυφό επίπεδο), και ένας μεταξύ κάθε κρυφής μονάδας (στο κρυφό επίπεδο) και την έξοδο.

Υπολογίζοντας μια σειρά από σταθμισμένα αθροίσματα, από μαθηματικής απόψεως, είναι το ίδιο με το να υπολογίζεται ένα σταθμισμένο άθροισμα. Για αυτό ο αλγόριθμος, αφού υπολογίσει ένα σταθμισμένο άθροισμα για κάθε κρυφή μονάδα, εφαρμόζει μια μη-γραμμική συνάρτηση στο αποτέλεσμα. Το αποτέλεσμα αυτής της συνάρτησης χρησιμοποιείται στον υπολογισμό του σταθμισμένου αθροίσματος για

τη τελική έξοδο y . Οι πιο γνωστές μη-γραμμικές συναρτήσεις που χρησιμοποιούνται είναι η relu , η οποία αποκόπτει τις αρνητικές τιμές, και η tanh , η οποία διακυμαίνεται στο -1 για χαμηλές τιμές και στο $+1$ για υψηλές τιμές [Müller et al., 2016].

Σχήμα 3.10: Βαθύ νευρωνικό δίκτυο με 2 κρυφά επίπεδα



Μια σημαντική παράμετρος η οποία πρέπει να ορίζεται από τον χρήστη είναι ο αριθμός των κρυφών μονάδων σε ένα κρυφό επίπεδο. Ο αριθμός αυτός μπορεί να είναι μικρός (άπό 10 για πολύ μικρά ή απλά σετ δεδομένων) ή και μεγάλος (10.000 για πολύ μεγάλα και πολύπλοκα σετ δεδομένων). Σημαντική παράμετρος είναι και ο αριθμός των κρυφών επιπέδων. Όπως παρατηρείται και στην εικόνα 3.10, οποιοδήποτε νευρωνικό δίκτυο διαθέτει πάνω από ένα κρυφό επίπεδο, ανήκει στην κατηγορία βαθιά νευρωνικά δίκτυα (deep neural networks) [Müller et al., 2016].

Τα νευρωνικά δίκτυα έχουν βοηθήσει σε πολλές εφαρμογές και προβλήματα στη μηχανική μάθηση. Ένα από τα κύρια πλεονεκτήματά τους είναι ότι μπορούν να συλλάβουν πληροφορίες οι οποίες εμπεριέχονται σε μεγάλα σετ δεδομένων και να φτιάξουν αρκετά πολύπλοκα μοντέλα. Αν τους δωθεί αρκετός υπολογιστικός χρόνος, δεδομένα και προσεκτικός ορισμός των παραμέτρων, τα νευρωνικά δίκτυα συνήθως υπερνικούν τους υπόλοιπους αλγόριθμους μηχανικής μάθησης. Από την άλλη μεριά, τα μεγάλα και πολύπλοκα νευρωνικά δίκτυα, συνήθως απαιτούν πολύ χρόνο για να εκπαιδευτούν. Ακόμη απαιτείται προσεκτική προεπεξεργασία των δεδομένων. Τέλος ο ορισμός των παραμέτρων είναι κάθε άλλο παρά εύκολος, αφού τις περισσότερες φορές πραγματοποιείται εμπειρικά για κάθε πρόβλημα [Müller et al., 2016].

Κεφάλαιο 4

Πειραματική Διαδικασία

4.1 Προγράμματα που Χρησιμοποιήθηκαν

Το πειραματικό μέρος της εργασίας εκτελέστηκε στη γλώσσα προγραμματισμού Python. Πιο συγκεκριμένα, για την εκτέλεση των αλγορίθμων, χρησιμοποιήθηκε η έκδοση Python 3.7 και ο IDE JetBrains Pycharm Community Edition. Για να ελέγξουμε τεχνικά τη χρήση του επεξεργαστή και της κάρτας γραφικών χρησιμοποιήθηκαν τα προγράμματα HeavyLoad και 3DMark αντίστοιχα. Το HeavyLoad [Hea, 2019] είναι ένα λογισμικό το οποίο τροποποιεί τη χρήση διάφορων εξαρτημάτων ενός υπολογιστή (επεξεργαστής, σκληρός δίσκος, μνήμη τυχαίας προσπέλασης, κα). Επιλέχθηκε διότι δίνει τη δυνατότητα για επιλογή του αριθμού των πυρήνων των οποίων η χρήση θα τροποποιηθεί, κάθε φορά στον επεξεργαστή. Το 3DMark [3DM, 2019] είναι ένα λογισμικό συγκριτικής αξιολόγησης το οποίο εφαρμόζει τεστ στρεσαρίσματος με τη μορφή βίντεο. Επιλέχθηκε λόγω της μεγάλης ποικιλίας του στα παραπάνω τεστ, με αποτέλεσμα να δίνει αρκετά μεγάλο εύρος μετρήσεων της κάρτα γραφικών. Για τη καταμέτρηση των μετρήσεων χρησιμοποιήθηκαν τα προγράμματα AIDA64 Extreme και TechPowerUp GPU-Z. Το AIDA64 Extreme [AID, 2019] προσφέρει μεγάλο εύρος καταγραφής μετρήσεων και είναι το κύριο λογισμικό που χρησιμοποιήθηκε για τη καταμέτρηση. Το TechPowerUp GPU-Z [gpru, 2019] είναι ένα ελαφρύ λογισμικό το οποίο δίνει πληροφορίες για τη κάρτα γραφικών, καθώς καταγράφει και αρκετές μετρήσεις για αυτή. Επιλέχθηκε λόγω της μεγαλύτερης ακρίβειας στη καταγραφή κατανάλωσης ενέργειας της κάρτας γραφικών.

4.2 Συλλογή και Επεξεργασία Δεδομένων

Για τη πειραματική διαδικασία χρησιμοποιήθηκε ένας υπολογιστής μεσαίων προδιαγραφών και ένας υπολογιστής υψηλών προδιαγραφών. Ο υπολογιστής μεσαίων προδιαγραφών αποτελείται από τον επεξεργαστή intel core i5-4690k και από την κάρτα γραφικών nvidia GeForce GTX 970. Ο i5-4690k [Int, 2019a] είναι ένας τετραπύρηνος επεξεργαστής, χρονισμένος στα 3.5 GHz, με δυνατότητα overclocking. Η GeForce GTX 970 [GTX, 2019] είναι μια κάρτα γραφικών τύπου GDDR5, διαθέτει 1664 πυρήνες, καθώς και μέγεθος μνήμης 4 GB. Ο υπολογιστής υψηλών προδιαγραφών αποτελείται από τον επεξεργαστή intel core i7-3770 και από την κάρτα γραφικών nvidia Titan X (Pascal). Ο i7-3770 [Int, 2019b] είναι ένας τετραπύρηνος επεξεργαστής ο οποίος με τη τεχνολογία της υπερνημάτωσης καταφέρνει να έχει 8 πυρήνες (2 νήματα ανά πυρήνα) και είναι χρονισμένος στα 3.4 GHz. Η Titan X (Pascal) [Tit, 2019] είναι μια κάρτα γραφικών τύπου GDDR5X, διαθέτει 3584 πυρήνες, καθώς και μέγεθος μνήμης 12 GB.

Τα χαρακτηριστικά τα οποία μετρήθηκαν για τη κατανάλωση ενέργειας στον επεξεργαστή είναι η χρήση του επεξεργαστή, η χρήση της κάρτας γραφικών, η θερμοκρασία στον επεξεργαστή και η κατανάλωση στην κάρτα γραφικών. Για την κατανάλωση στην κάρτα γραφικών μετρήθηκαν τα χαρακτηριστικά χρήση του επεξεργαστή, χρήση της κάρτας γραφικών, θερμοκρασία στην κάρτα γραφικών και θερμική σχεδιαστική ισχύ (TDP) στην κάρτα γραφικών.

Στον υπολογιστή μεσαίων προδιαγραφών, για τη συλλογή των χαρακτηριστικών για τη κατανάλωση ενέργειας του επεξεργαστή, τέθηκαν τρία σενάρια στρεσαρίσματος (για 1, 2 και 3 πυρήνες αντίστοιχα) από το λογισμικό HeavyLoad, και σε καθένα από αυτά το λογισμικό AIDA64 Extreme κατέγραφε (ανά 3 δευτερόλεπτα) τις απαραίτητες μετρήσεις. Για τη συλλογή των χαρακτηριστικών για τη κατανάλωση ενέργειας στην κάρτα γραφικών, τέθηκαν 5 σενάρια στρεσαρίσματος (Cloud Gate, Fire Strike, Ice Storm, Ice Storm Extreme και Sky Diver) από το λογισμικό 3DMark, και καταγράφηκαν οι απαραίτητες μετρήσεις. Αντίστοιχα και στον υπολογιστή υψηλών προδιαγραφών ακολουθήθηκε η ίδια διαδικασία, με τη διαφορά πως έγινε χρήση απομακρυσμένης πρόσβασης. Για τη κατανάλωση ενέργειας στον επεξεργαστή συγκεντρώθηκαν 100 δείγματα και για την κατανάλωση ενέργειας στην

κάρτα γραφικών συγκεντρώθηκαν 200 δείγματα. Μετά τη συλλογή των δεδομένων, ακολούθησε προσεκτική επεξεργασία, αφαιρώντας τις ακραίες τιμές.

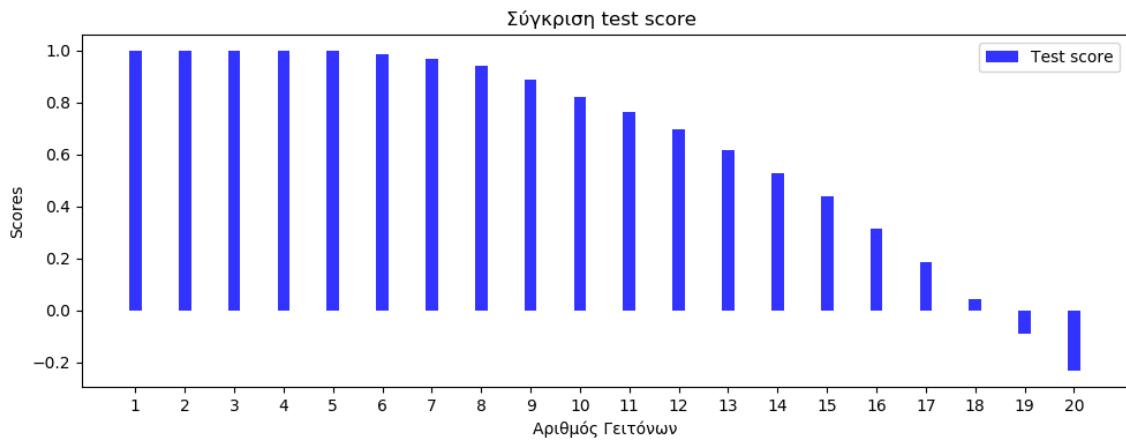
Ουσιαστικά, για κάθε αλγόριθμο, θέλουμε να διεξάγουμε τρία πειράματα. Το πρώτο πείραμα αφορά την πρόβλεψη κατανάλωσης ενέργειας στον επεξεργαστή και αποτελείται από δύο σενάρια. Αρχικά λαμβάνουμε ως εισόδους τη χρήση του επεξεργαστή και της κάρτας γραφικών καθώς και τη θερμοκρασία του επεξεργαστή και προβλέπουμε την κατανάλωση ενέργειας στον επεξεργαστή, έπειτα λαμβάνουμε τη χρήση του επεξεργαστή και της κάρτας γραφικών ως εισόδους και προβλέπουμε την κατανάλωση ενέργειας του επεξεργαστή. Το δεύτερο πείραμα αφορά τη συνολική πρόβλεψη ενέργειας σε επεξεργαστή και κάρτα γραφικών και αποτελείται από δύο σενάρια. Αρχικά λαμβάνουμε ως εισόδους τη χρήση του επεξεργαστή και της κάρτας γραφικών καθώς και τη θερμοκρασία του επεξεργαστή και προβλέπουμε το άθροισμα της κατανάλωσης ενέργειας των δύο εξαρτημάτων, έπειτα λαμβάνουμε τη χρήση του επεξεργαστή και της κάρτας γραφικών ως εισόδους και προβλέπουμε άθροισμα της κατανάλωσης ενέργειας των δύο εξαρτημάτων. Το τρίτο πείραμα αφορά την πρόβλεψη κατανάλωση ενέργειας στην κάρτα γραφικών και αποτελείται από δύο σενάρια. Αρχικά λαμβάνουμε ως εισόδους τη χρήση της κάρτας γραφικών και του επεξεργαστή καθώς και τη θερμοκρασία αλλά και τη θερμική σχεδιαστική ισχύ της κάρτας γραφικών και προβλέπουμε την κατανάλωση ενέργειας στην κάρτα γραφικών, έπειτα λαμβάνουμε ως εισόδους τη χρήση του επεξεργαστή και της κάρτας γραφικών και προβλέπουμε την κατανάλωση ενέργειας στην κάρτα γραφικών.

4.3 Πειράματα στον υπολογιστή μεσαίων προδιαγραφών

Κ-Πλησιέστεροι Γείτονες

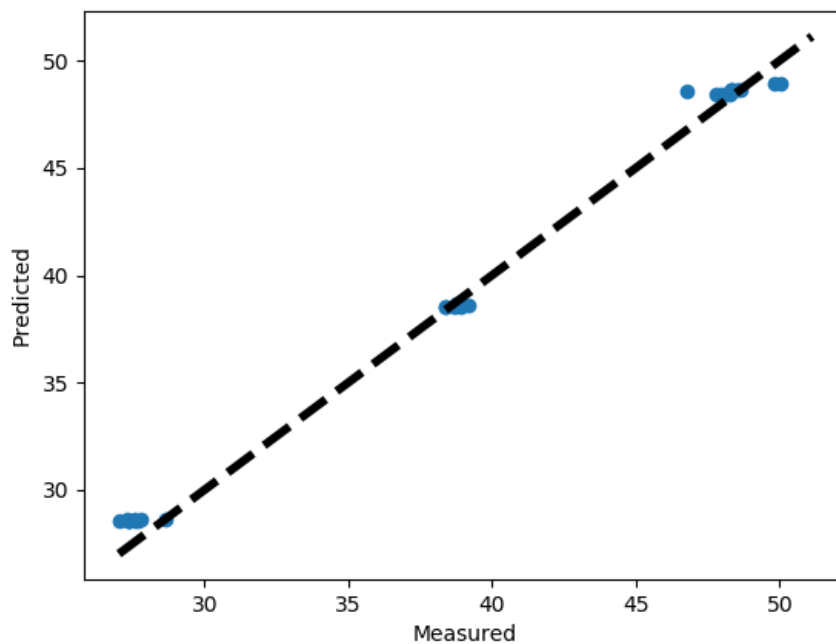
Ο αλγόριθμος περιέχει μια επανάληψη (από 1 μέχρι 20 με βήμα 1), όπου η μεταβλητή *n_neighbors* (αριθμός γειτόνων) λαμβάνει την τιμή της επανάληψης, για καλύτερη διεξαγωγή συμπερασμάτων. Στο πρώτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στον επεξεργαστή. Χρησιμοποιώντας τρία χαρακτηριστικά ως εισόδους, η βέλτιστη τιμή της μεταβλητής *n_neighbors* (αριθμός γειτόνων) είναι 1 (σύμφωνα με το παρακάτω διάγραμμα). Ορίζοντας τη μεταβλητή *n_neighbors* στη τιμή 1, λαμβάνουμε το καλύτερο score στο test σετ, το οποίο είναι 0.99.

Σχήμα 4.1: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (KNN)



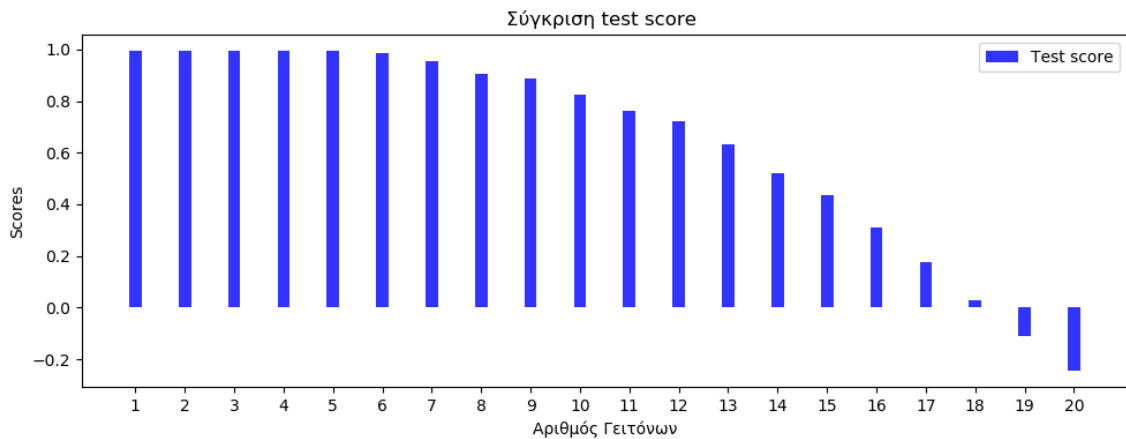
Όπως φαίνεται στο σχήμα 4.1, για τις πρώτες δεκαοχτώ τιμές της μεταβλητής $n_neighbors$, τα test scores έχουν θετική τιμή. Καθώς η μεταβλητή $n_neighbors$ αυξάνεται, παρατηρείται αισθητή μείωση του test score.

Σχήμα 4.2: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 3 εισόδους (KNN)



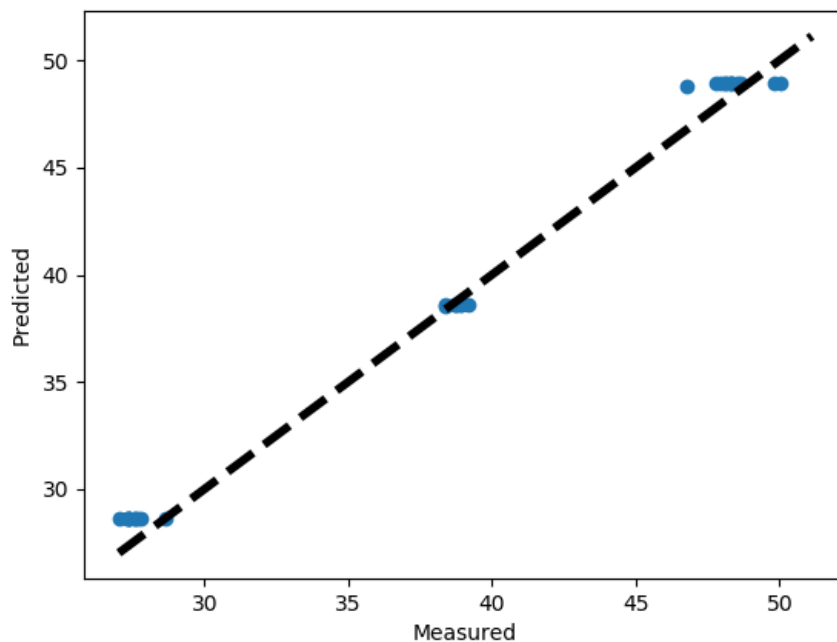
Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση του επεξεργαστή, η βέλτιστη τιμή της μεταβλητής $n_neighbors$ (αριθμός γειτόνων) είναι 5 (σύμφωνα με το σχήμα 4.3). Ορίζοντας τη μεταβλητή $n_neighbors$ στη τιμή 5, λαμβάνουμε το καλύτερο score στο test σετ, το οποίο είναι 0.99.

Σχήμα 4.3: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (KNN)



Όπως φαίνεται και στο σχήμα 4.3, για τις πρώτες δεκαεφτά τιμές της μεταβλητής $n_neighbors$, τα test scores έχουν θετικές τιμές. Καθώς η μεταβλητή $n_neighbors$ αυξάνεται, παρατηρείται αισθητή μείωση του test score.

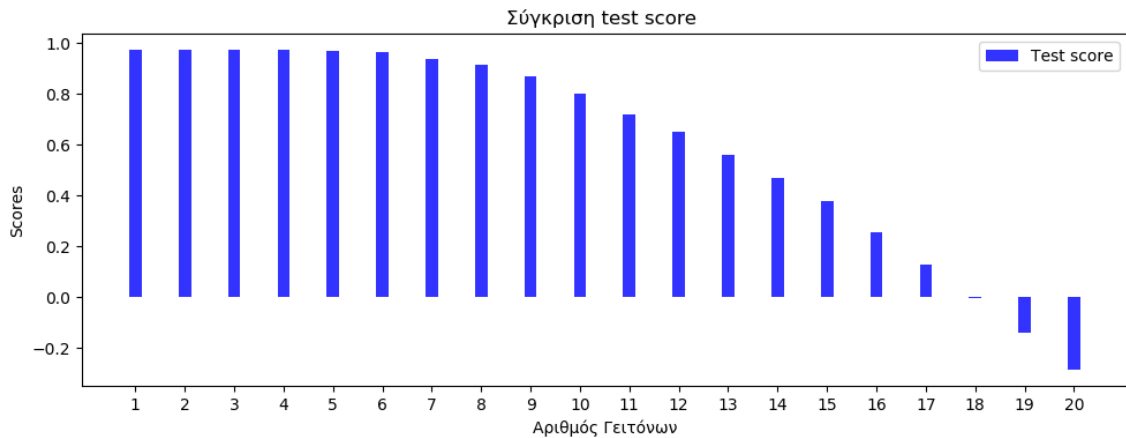
Σχήμα 4.4: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (KNN)



Στο δεύτερο πείραμα κάνουμε πρόβλεψη του αθροίσματος κατανάλωσης ενέρ-

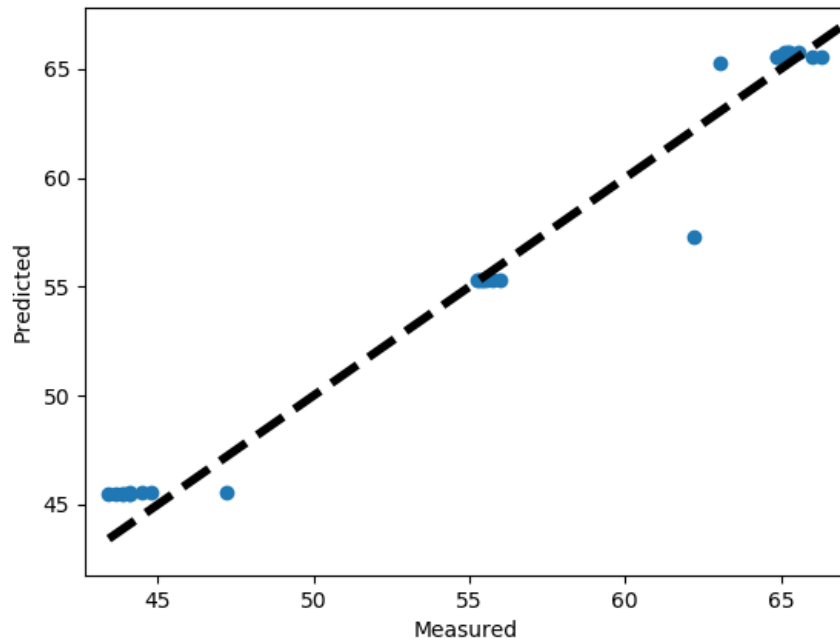
γείας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 3 χαρακτηριστικά ως εισόδους. Ορίζοντας τη μεταβλητή $n_neighbors$ (αριθμός γειτώνων) στη βέλτιστη τιμή 3, σύμφωνα με το σχήμα 4.5, λαμβάνουμε το καλύτερο score στο test σετ, το οποίο είναι 0.97.

Σχήμα 4.5: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (KNN)



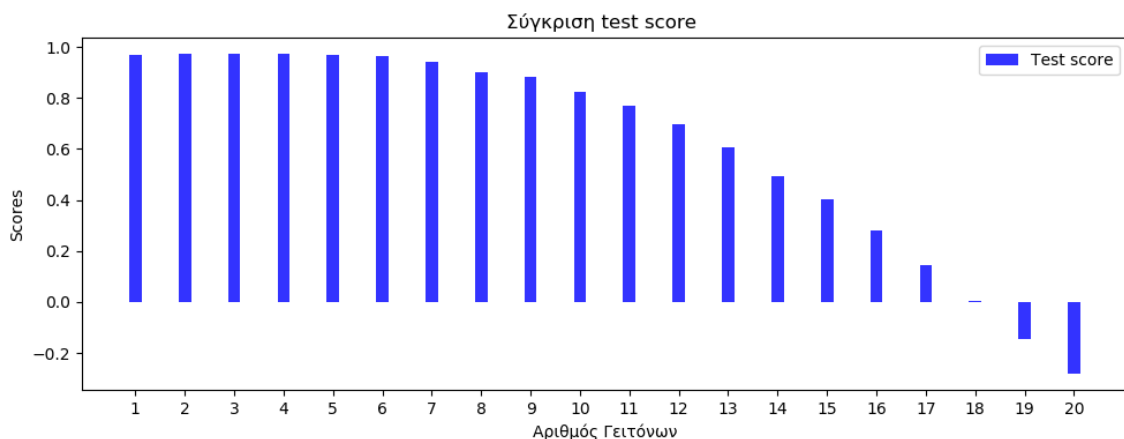
Όπως φαίνεται στο σχήμα 4.5, για τις πρώτες δεκαεφτά τιμές της μεταβλητής $n_neighbors$, τα test scores έχουν θετική τιμή. Καθώς η μεταβλητή $n_neighbors$ αυξάνεται, παρατηρείται αισθητή μείωση του test score.

Σχήμα 4.6: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (KNN)



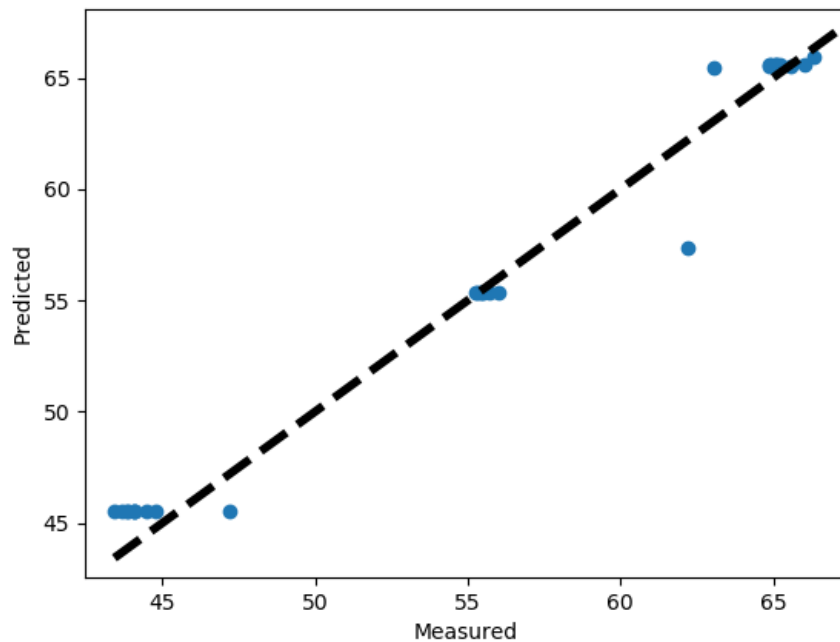
Για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 2 χαρακτηριστικά ως εισόδους, ορίζουμε τη μεταβλητή $n_neighbors$ (αριθμός γειτώνων) στη βέλτιστη τιμή 2, σύμφωνα με το σχήμα 4.7. Ορίζοντας τη μεταβλητή $n_neighbors$ στη τιμή 2, λαμβάνουμε το καλύτερο score στο test σετ, το οποίο είναι 0.97.

Σχήμα 4.7: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (KNN)



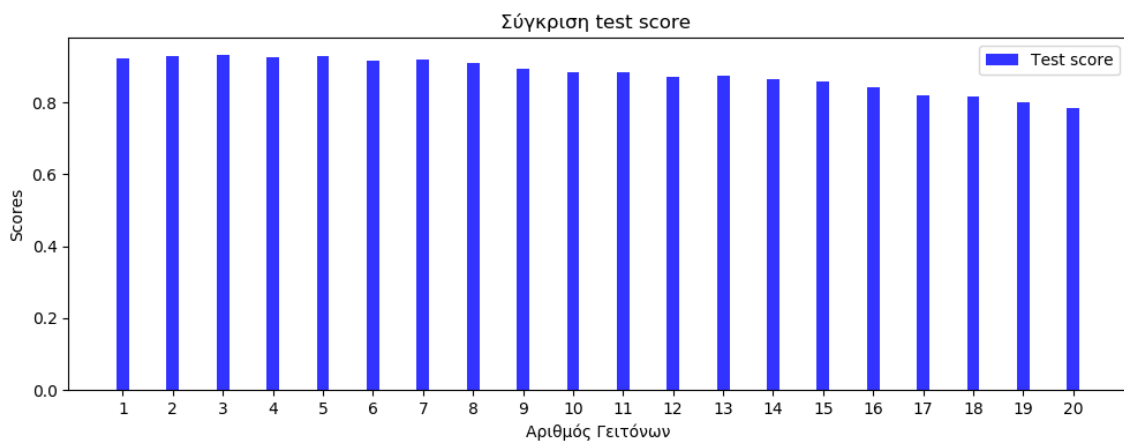
Όπως φαίνεται και στο σχήμα 4.7, για τις πρώτες δεκαεφτά τιμές της μεταβλητής $n_neighbors$, τα test scores έχουν θετικές τιμές. Καθώς η μεταβλητή $n_neighbors$ αυξάνεται, παρατηρείται αισθητή μείωση του test score.

Σχήμα 4.8: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (KNN)



Στο τρίτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στην κάρτα γραφικών. Χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους, η βέλτιστη τιμή της μεταβλητής $n_neighbors$ (αριθμός γειτόνων) είναι 2 (σύμφωνα με το σχήμα 4.9) με test score 0.92.

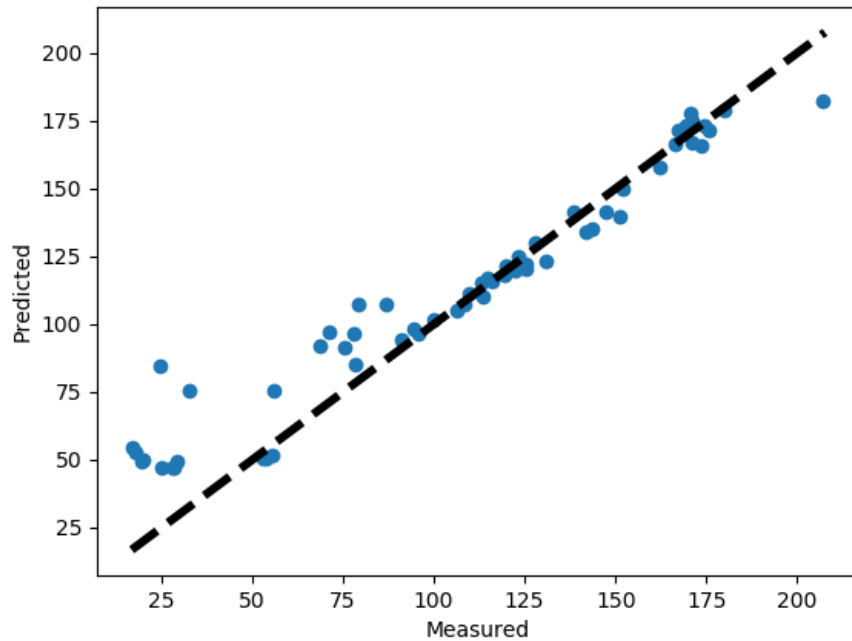
Σχήμα 4.9: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (KNN)



Όπως φαίνεται στο σχήμα 4.9, το test score παίρνει τη βέλτιστη τιμή του όταν η μεταβλητή $n_neighbors$ παίρνει την τιμή 2. Καθώς η μεταβλητή $n_neighbors$ αυξάνεται, το test score μειώνεται.

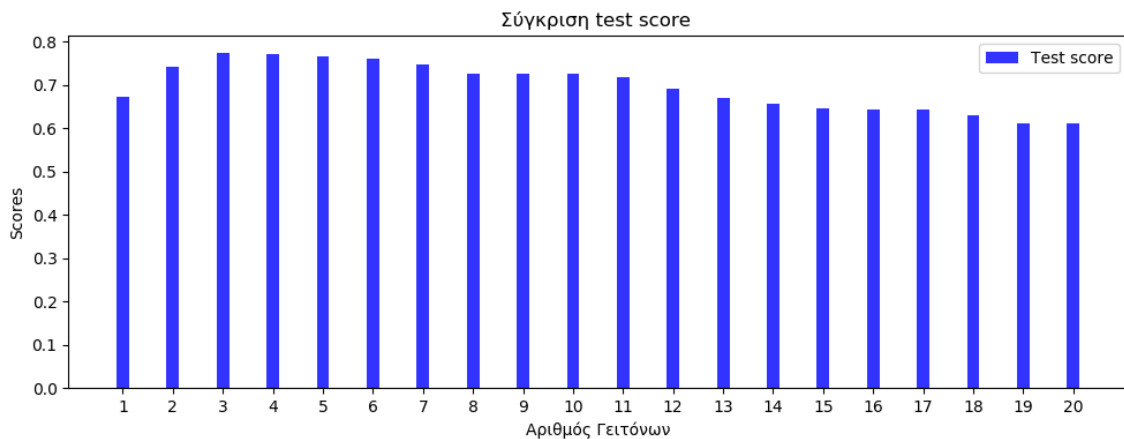
Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατα-

Σχήμα 4.10: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 4 εισόδους (KNN)



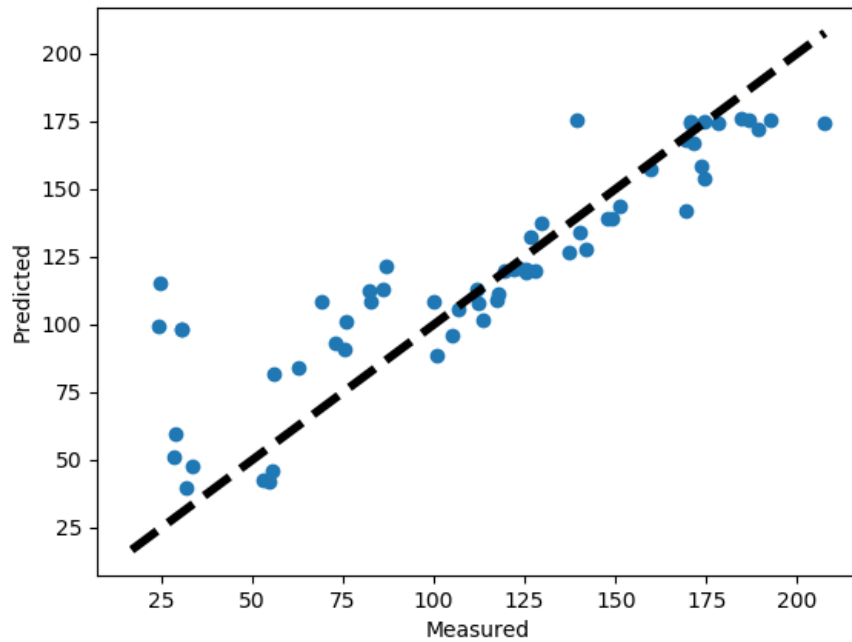
νάλωση της κάρτας γραφικών, η βέλτιστη τιμή της μεταβλητής $n_neighbors$ (αριθμός γειτόνων) είναι 3 (σύμφωνα με το σχήμα 4.11) με test score 0.77.

Σχήμα 4.11: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (KNN)

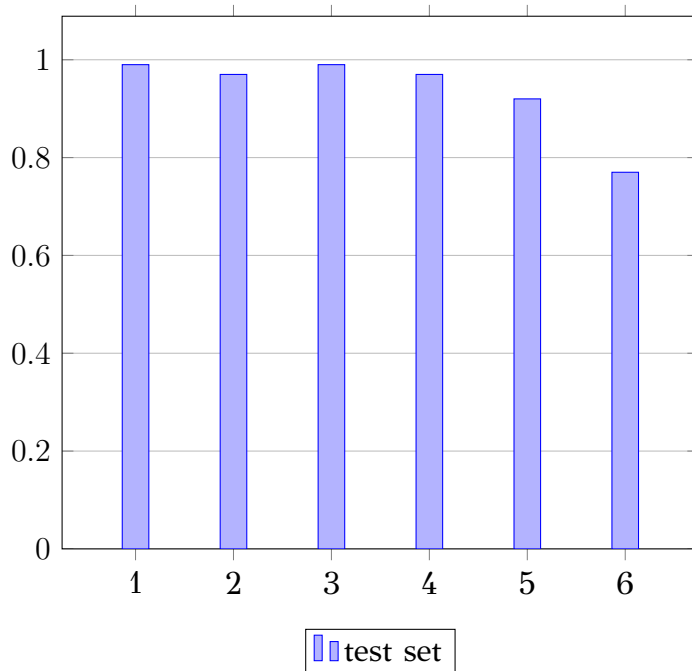


Όπως φαίνεται στο σχήμα 4.11, το test score παίρνει τη βέλτιστη τιμή του όταν η μεταβλητή $n_neighbors$ παίρνει την τιμή 3. Καθώς η μεταβλητή $n_neighbors$ αυξάνεται, το test score αυξομειώνεται.

Σχήμα 4.12: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 εισόδους (KNN)



Τα σχήματα 4.2, 4.4, 4.6, 4.8, 4.10 και 4.12 απεικονίζουν διαγράμματα πραγματικών με προβλεπόμενων τιμών για κάθε μία από τις παραπάνω περιπτώσεις. Για κάθε σημείο, στον άξονα X αναπαριστάται η πραγματική τιμή του, ενώ στον άξονα Y η προβλεπόμενη τιμή. Όσο πιο κοντά είναι κάποιο σημείο στη γραμμή (μοντέλο), τόσο πιο σωστά πραγματοποιήθηκε η πρόβλεψη.



Στο παραπάνω σχήμα συνοψίζονται συγκεντρωτικά τα scores των test σετ για

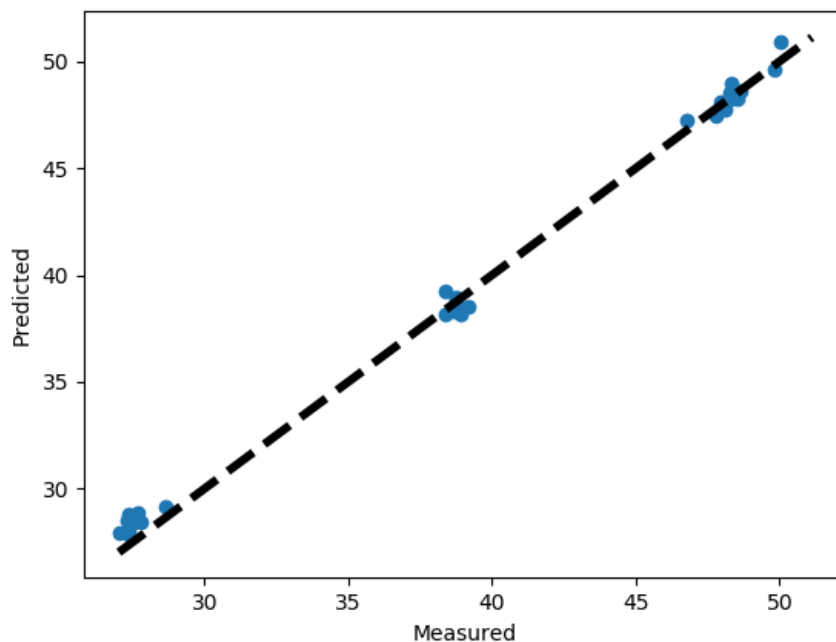
όλα τα παραπάνω σενάρια με τη σειρά που πραγματοποιήθηκαν. Παρατηρούμε ότι η πρόβλεψη κατανάλωσης ενέργειας της κάρτας γραφικών είχε μεγαλύτερη επιτυχία χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους. Η πρόβλεψη κατανάλωσης ενέργειας του επεξεργαστή είχε τα ίδια test score χρησιμοποιώντας τρία και δύο χαρακτηριστικά ως εισόδους. Ομοίως, η πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών είχε τα ίδια test score χρησιμοποιώντας τρία και δύο χαρακτηριστικά ως εισόδους.

Γραμμική Παλινδρόμηση

Στο πρώτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στον επεξεργαστή. Χρησιμοποιώντας τρία χαρακτηριστικά ως εισόδους, το μοντέλο της γραμμικής παλινδρόμησης δίνεται από την παρακάτω εξίσωση:

$$y = 0.34cpu_{\%} - 0.07gpu_{\%} + 0.23cpu_c + 9.72$$

Σχήμα 4.13: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 3 εισόδους (LR)

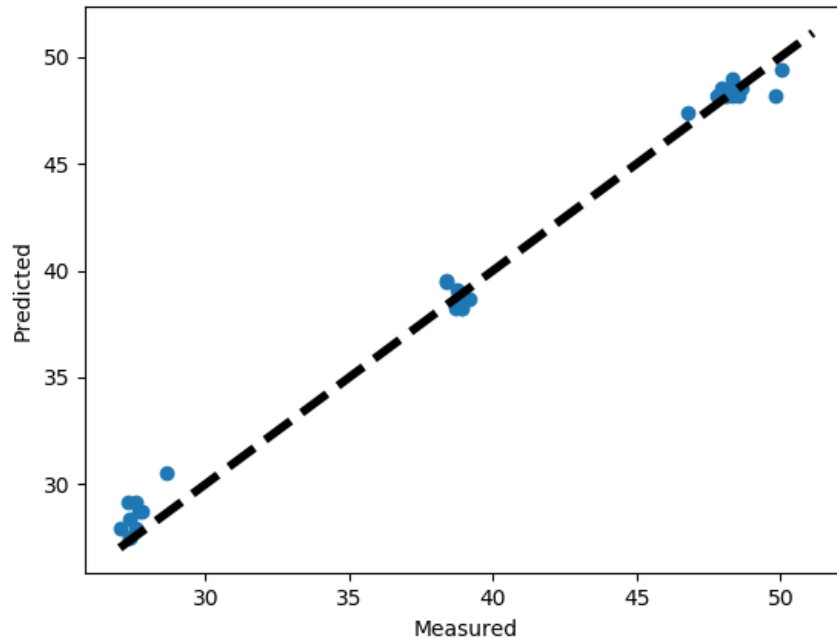


Όπως φαίνεται και στη παραπάνω εξίσωση, η χρήση και η θερμοκρασία του επεξεργαστή έχουν θετικούς συντελεστές. Αντίθετα, η χρήση της κάρτας γραφικών έχει αρνητικό συντελεστή, συνεισφέροντας λιγότερο στο μοντέλο. Η σταθερά του μοντέλου φέρει τη τιμή 9.72, ενώ το score στο test σετ είναι 0.99. Χρησιμοποιώντας δύο

χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση του επεξεργαστή, το μοντέλο της γραμμικής παλινδρόμησης δίνεται από την παρακάτω εξίσωση:

$$y = 0.41cpu_{\%} + 0.01gpu_{\%} + 17.11$$

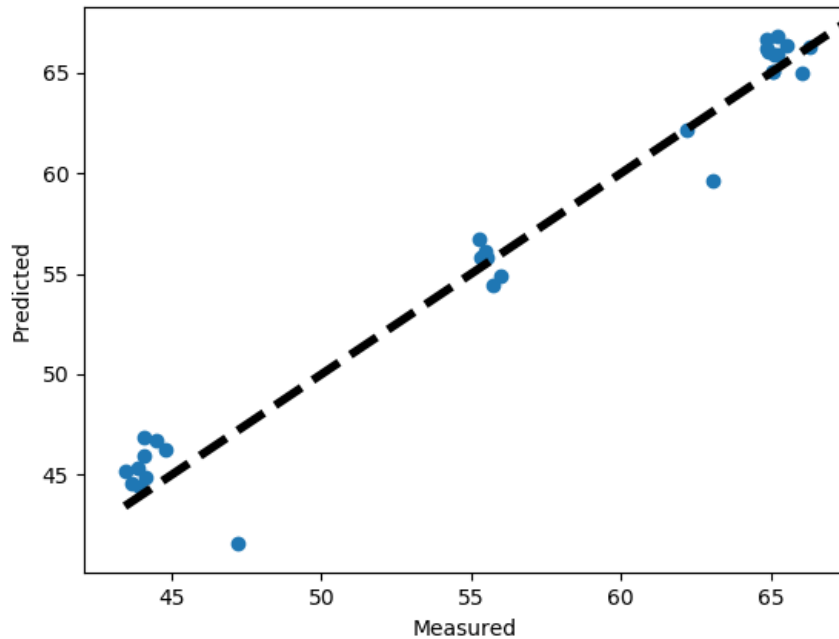
Σχήμα 4.14: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (LR)



Όπως φαίνεται στη παραπάνω εξίσωση, η χρήση του επεξεργαστή και της κάρτας γραφικών έχουν θετικούς συντελεστές, με τη χρήση της κάρτας γραφικών να μην επηρεάζει σημαντικά το μοντέλο. Η σταθερά του μοντέλου έχει τιμή 17.11, ενώ το score στο test σετ είναι 0.98. Στο δεύτερο πείραμα κάνουμε πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών. Χρησιμοποιώντας 3 χαρακτηριστικά ως εισόδους, το μοντέλο της γραμμικής παλινδρόμησης δίνεται από την παρακάτω εξίσωση:

$$y = 0.46cpu_{\%} - 1.14gpu_{\%} - 0.15cpu_c + 46.03$$

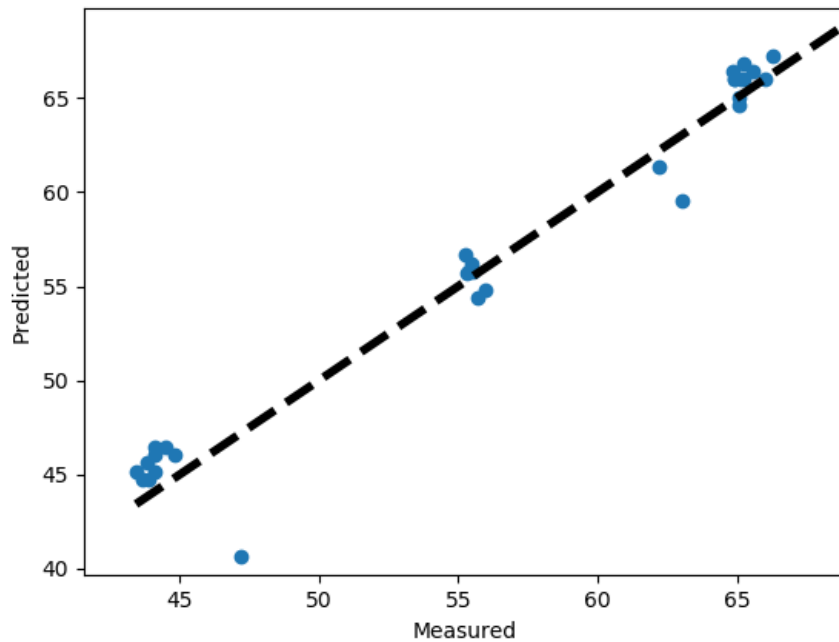
Σχήμα 4.15: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (LR)



Όπως γίνεται αντιληπτό στη παραπάνω εξίσωση, η χρήση του επεξεργαστή έχει θετικό συντελεστή. Αντίθετα, η χρήση της κάρτας γραφικών και η θερμοκρασία του επεξεργαστή έχουν αρνητικούς συντελεστές, με τη θερμοκρασία του επεξεργαστή να συμβάλει λιγότερο στο μοντέλο. Η σταθερά του μοντέλου φέρει την τιμή 46.03, ενώ το score στο test σετ είναι 0.96. Για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 2 χαρακτηριστικά ως εισόδους, το μοντέλο της γραμμικής παλινδρόμησης δίνεται από την παρακάτω εξίσωση:

$$y = 0.42cpu\% - 1.40gpu\% + 41.19$$

Σχήμα 4.16: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (LR)

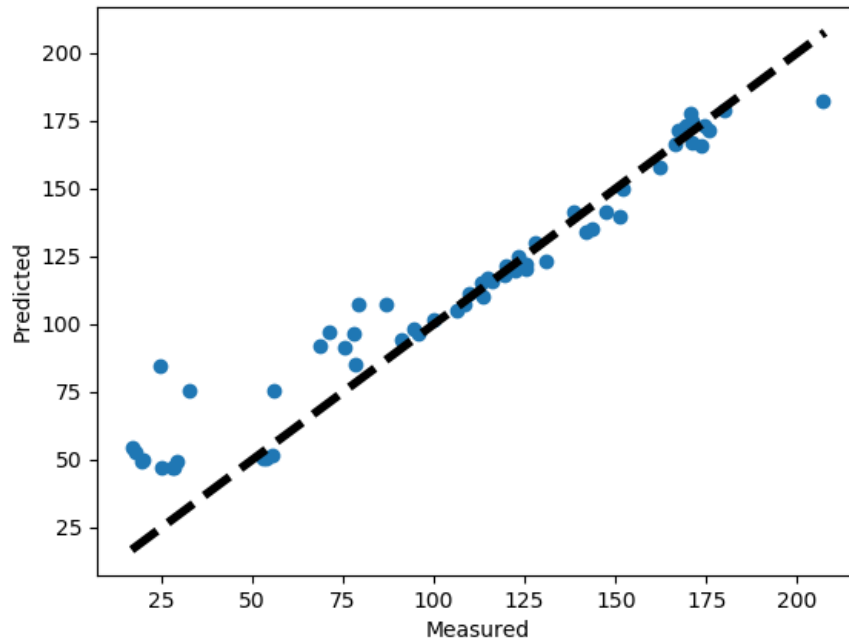


Παρατηρώντας την παραπάνω εξίσωση βλέπουμε πως, η χρήση του επεξεργαστή έχει θετικό συντελεστή ενώ η χρήση της κάρτας γραφικών έχει αρνητικό συντελεστή, με τη χρήση της κάρτας γραφικών να έχει κύριο ρόλο στο μοντέλο, ενώ η χρήση του επεξεργαστή το επηρεάζει λιγότερο. Η σταθερά του μοντέλου έχει τιμή 41.19, ενώ το score στο test σετ είναι 0.97.

Στο τρίτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στην κάρτα γραφικών. Χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους, το μοντέλο της γραμμικής παλινδρόμησης δίνεται από την παρακάτω εξίσωση:

$$y = 0.005cpu_{\%} - 0.002gpu_{\%} - 0.004gpu_c + 2.49gpu_{tdp} + 0.12$$

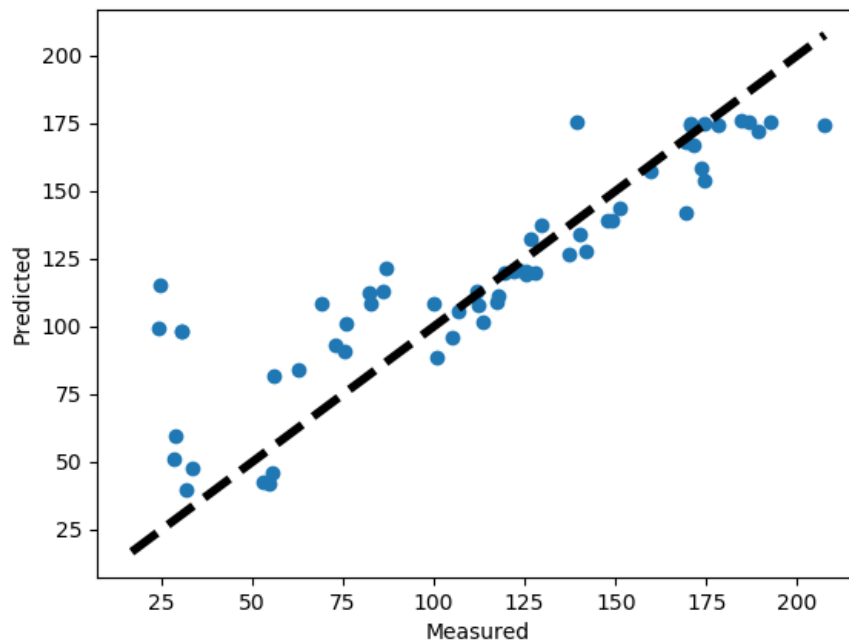
Σχήμα 4.17: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 4 εισόδους (LR)



Όπως φαίνεται και στην παραπάνω εξίσωση, η χρήση του επεξεργαστή και η θερμική σχεδιαστική ισχύ της κάρτας γραφικών έχουν θετικούς συντελεστές, ενώ η χρήση και η θερμοκρασία της κάρτας γραφικών έχουν αρνητικούς συντελεστές. Η θερμική σχεδιαστική ισχύ της κάρτας γραφικών έχει κύριο ρόλο στο μοντέλο, ενώ τα υπόλοιπα χαρακτηριστικά δε το επηρεάζουν σχεδόν καθόλου, ως αμελητέες ποσότητες. Η σταθερά του μοντέλου έχει τιμή 0.12, ενώ το score στο test σετ είναι 0.99. Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση της κάρτας γραφικών, το μοντέλο της γραμμικής παλινδρόμησης δίνεται από την παρακάτω εξίσωση:

$$y = -0.02cpu_{\%} + 1.59gpu_{\%} + 41.44$$

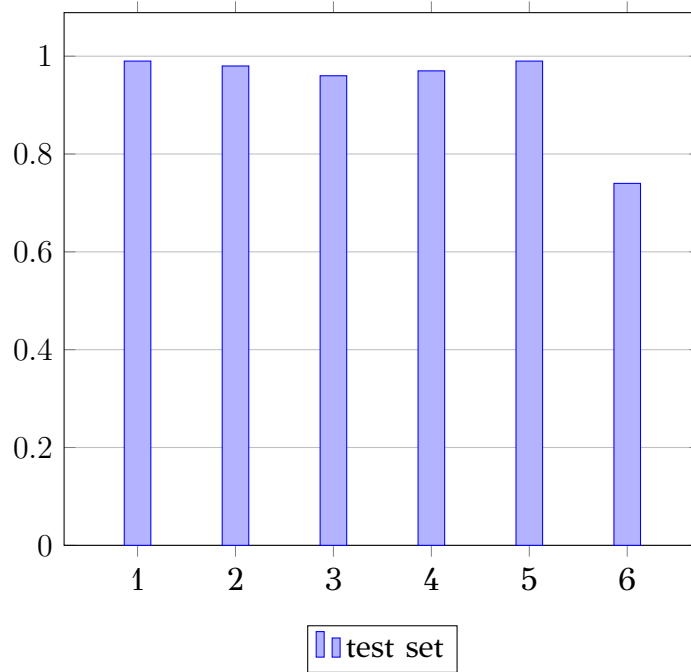
Σχήμα 4.18: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 εισόδους (LR)



Όπως γίνεται αντιληπτό στην παραπάνω εξίσωση, η χρήση του επεξεργαστή έχει αρνητικό συντελεστή, μην επηρεάζοντας σημαντικά το μοντέλο. Η χρήση της κάρτας γραφικών έχει θετικό συντελεστή, έχοντας κύριο ρόλο στο μοντέλο. Η σταθερά του μοντέλου έχει τιμή 41.44, ενώ το score στο test σετ είναι 0.74.

Τα σχήματα 4.13, 4.14, 4.15, 4.16, 4.17 και 4.18 απεικονίζουν διαγράμματα πραγματικών με προβλεπόμενων τιμών για κάθε μία από τις παραπάνω περιπτώσεις. Για κάθε σημείο, στον άξονα X αναπαριστάται η πραγματική τιμή του, ενώ στον άξονα Y η προβλεπόμενη τιμή. Όσο πιο κοντά είναι κάποιο σημείο στη γραμμή (μοντέλο), τόσο πιο σωστά πραγματοποιήθηκε η πρόβλεψη.

Στο παρακάτω σχήμα συνοψίζονται συγκεντρωτικά τα scores των train και test σετ για όλες τις παραπάνω περιπτώσεις με τη σειρά που πραγματοποιήθηκαν. Παρατηρούμε ότι η πρόβλεψη κατανάλωσης ενέργειας της κάρτας γραφικών είχε μεγαλύτερη επιτυχία χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους. Η πρόβλεψη κατανάλωσης ενέργειας του επεξεργαστή ήταν ελάχιστα καλύτερη χρησιμοποιώντας τρία χαρακτηριστικά ως εισόδους. Η πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών ήταν ελάχιστα καλύτερη χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους.



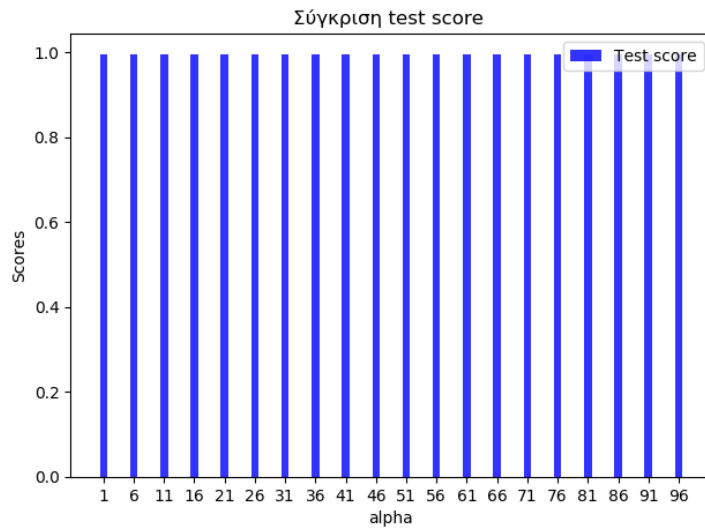
Παλινδρόμηση Ridge

Ο αλγόριθμος περιέχει μια επανάληψη (από 1 μέχρι 100 με βήμα 5), όπου η μεταβλητή $alpha$ λαμβάνει την τιμή της επανάληψης, για καλύτερη διεξαγωγή συμπερασμάτων. Στο πρώτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στον επεξεργαστή. Χρησιμοποιώντας τρία χαρακτηριστικά ως εισόδους, το μοντέλο της παλινδρόμησης ridge με τη μεταβλητή $alpha$ να έχει τη μέγιστη τιμή 11 δίνεται από την παρακάτω εξίσωση:

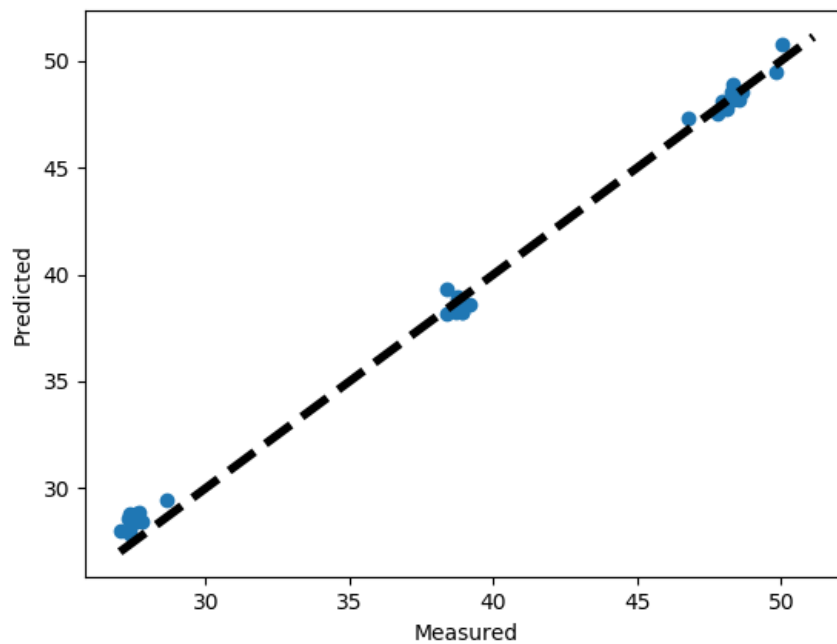
$$y = 0.34cpu_{\%} - 0.07gpu_{\%} + 0.23cpu_c + 9.77$$

Όπως φαίνεται και στη παραπάνω εξίσωση, η χρήση και η θερμοκρασία του επεξεργαστή έχουν θετικούς συντελεστές. Αντίθετα, η χρήση της κάρτας γραφικών έχει αρνητικό συντελεστή, συνεισφέροντας λιγότερο στο μοντέλο. Η σταθερά του μοντέλου φέρει τη τιμή 9.77, ενώ το score στο test set είναι 0.99. Καθώς αυξάνεται η μεταβλητή $alpha$ παρατηρείται μεταβολή στο 4ο δεκαδικό ψηφίο των test score (σχήμα 4.19).

Σχήμα 4.19: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (RR)



Σχήμα 4.20: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 3 εισόδους (RR)



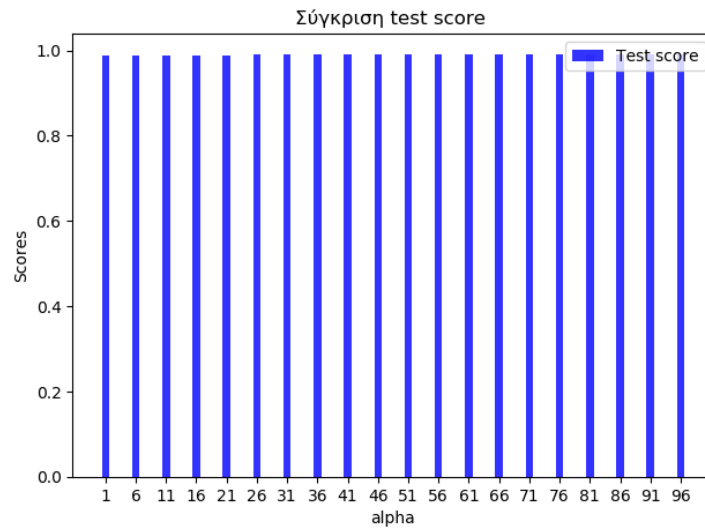
Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση του επεξεργαστή, το μοντέλο της παλινδρόμησης ridge με τη μεταβλητή $alpha$ να έχει τη μέγιστη τιμή 96 δίνεται από την παρακάτω εξίσωση:

$$y = 0.41cpu\% + 0.01gpu\% + 17.22$$

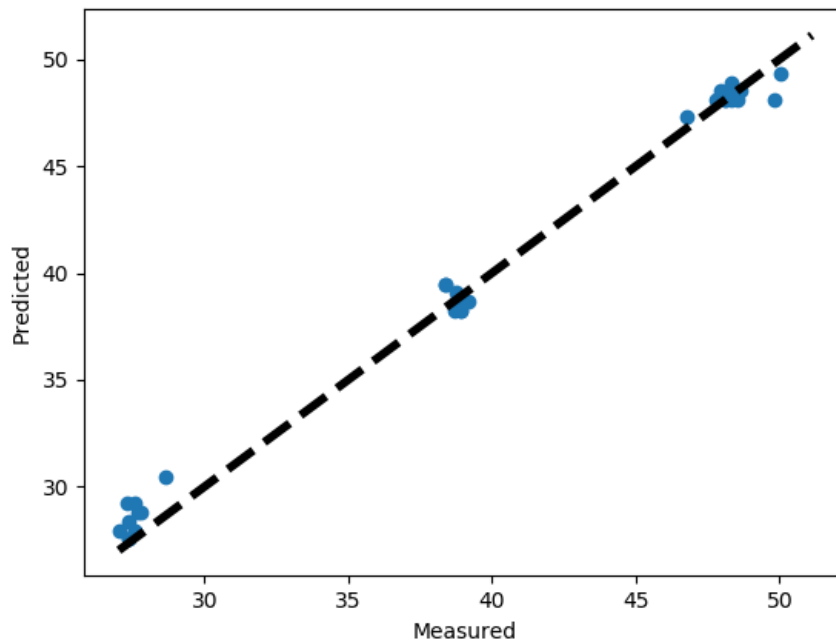
Όπως φαίνεται στη παραπάνω εξίσωση, η χρήση του επεξεργαστή και της κάρτας

γραφικών έχουν θετικούς συντελεστές, με τη χρήση της κάρτας γραφικών να μην επηρεάζει σημαντικά το μοντέλο. Η σταθερά του μοντέλου έχει τιμή 17.22, ενώ το score στο test σετ είναι 0.98. Καθώς αυξάνεται η μεταβλητή α παρατηρείται μεταβολή στο 4ο δεκαδικό ψηφίο των test score (σχήμα 4.21).

Σχήμα 4.21: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (RR)



Σχήμα 4.22: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (RR)



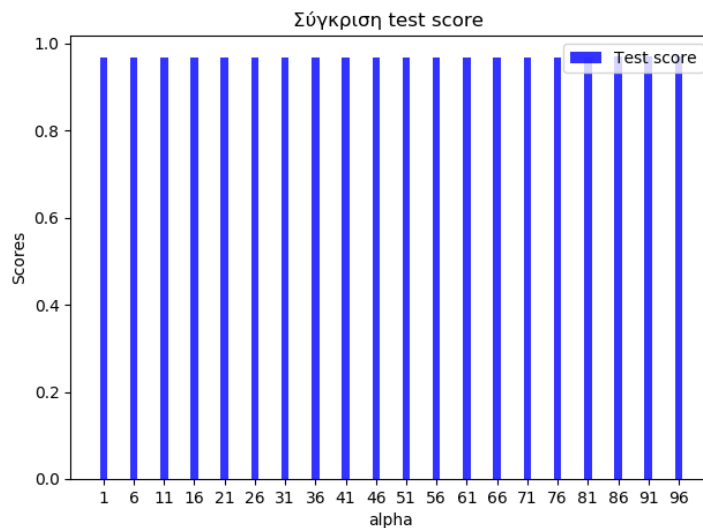
Στο δεύτερο πείραμα κάνουμε πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 3 χαρακτηρι-

στικά ως εισόδους, το μοντέλο της παλινδρόμησης ridge με τη μεταβλητή $alpha$ να έχει τη μέγιστη τιμή 96 δίνεται από την παρακάτω εξίσωση:

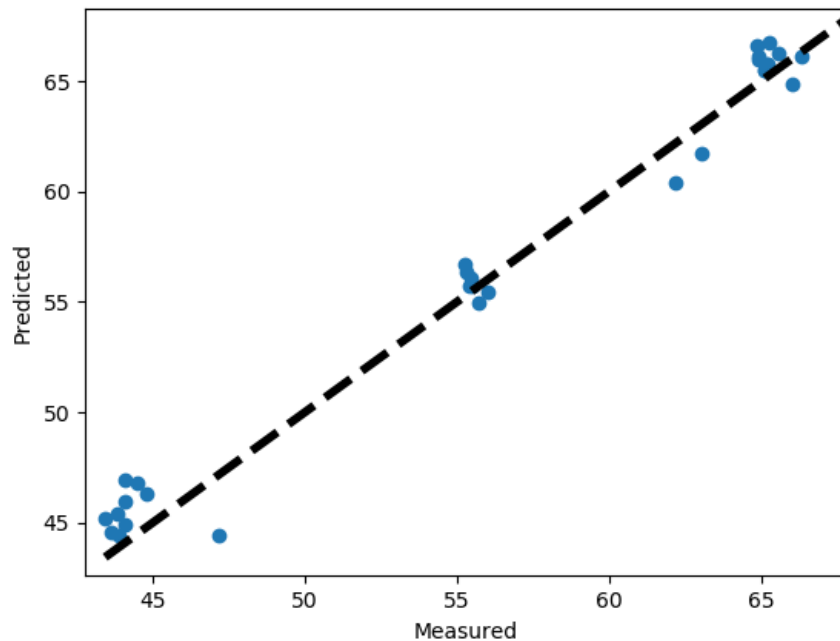
$$y = 0.46cpu_{\%} - 0.78gpu_{\%} - 0.15cpu_c + 43.50$$

Όπως γίνεται αντιληπτό στη παραπάνω εξίσωση, η χρήση του επεξεργαστή έχει θετικό συντελεστή. Αντίθετα, η χρήση της κάρτας γραφικών και η θερμοκρασία του επεξεργαστή έχουν αρνητικούς συντελεστές, με τη θερμοκρασία του επεξεργαστή να συμβάλει λιγότερο στο μοντέλο. Η σταθερά του μοντέλου φέρει την τιμή 46.50, ενώ το score στο test σετ είναι 0.96. Καθώς αυξάνεται η μεταβλητή $alpha$ παρατηρείται μεταβολή στο 3ο δεκαδικό ψηφίο των test score (σχήμα 4.23).

Σχήμα 4.23: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (RR)



Σχήμα 4.24: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (RR)

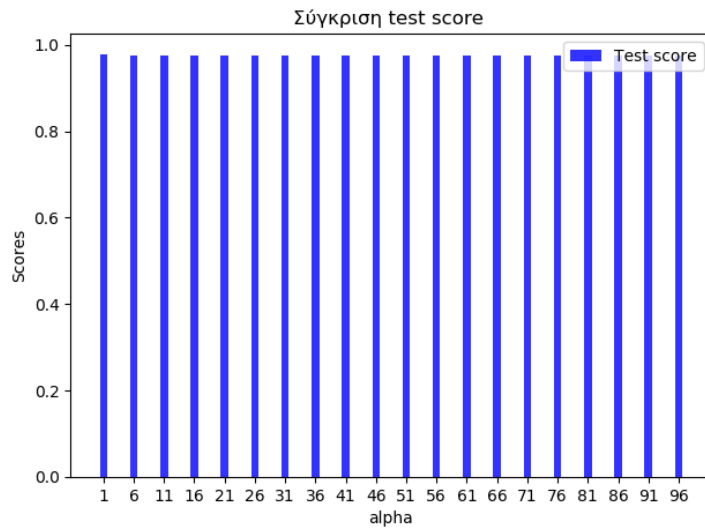


Για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 2 χαρακτηριστικά ως εισόδους, το μοντέλο της παλινδρόμησης ridge με τη μεταβλητή α να έχει τη μέγιστη τιμή 1 δίνεται από την παρακάτω εξίσωση:

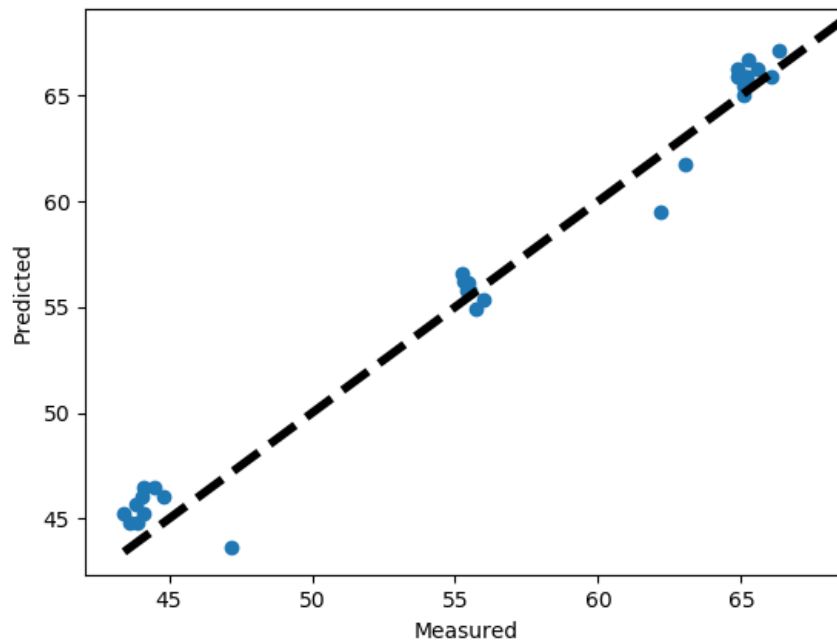
$$y = 0.42cpu\% - 1.39gpu\% + 41.14$$

Παρατηρώντας την παραπάνω εξίσωση βλέπουμε πως, η χρήση του επεξεργαστή έχει θετικό συντελεστή ενώ η χρήση της κάρτας γραφικών έχει αρνητικό συντελεστή, με τη χρήση της κάρτας γραφικών να έχει κύριο ρόλο στο μοντέλο, ενώ η χρήση του επεξεργαστή το επηρεάζει λιγότερο. Η σταθερά του μοντέλου έχει τιμή 41.14, ενώ το score στο test σετ είναι 0.97. Καθώς αυξάνεται η μεταβλητή α παρατηρείται μεταβολή στο 3ο δεκαδικό ψηφίο των test score (σχήμα 4.25).

Σχήμα 4.25: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (RR)



Σχήμα 4.26: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (RR)

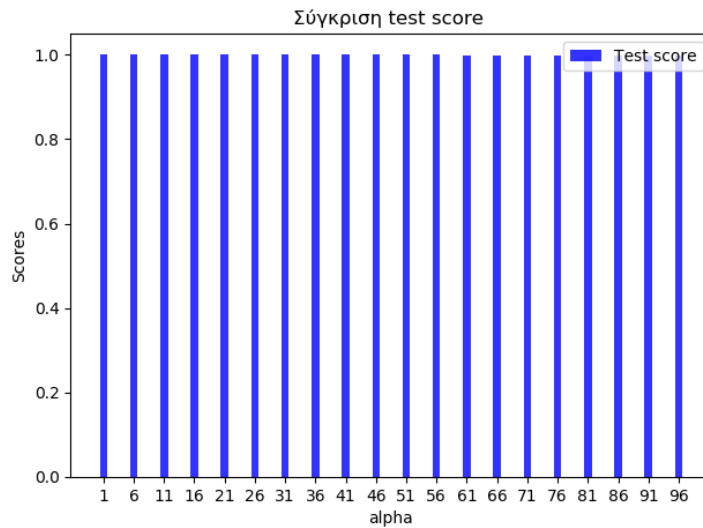


Στο τρίτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στην κάρτα γραφικών. Χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους, το μοντέλο της παλινδρόμησης ridge με τη μεταβλητή $alpha$ να έχει τη μέγιστη τιμή 1 δίνεται από την παρακάτω εξίσωση:

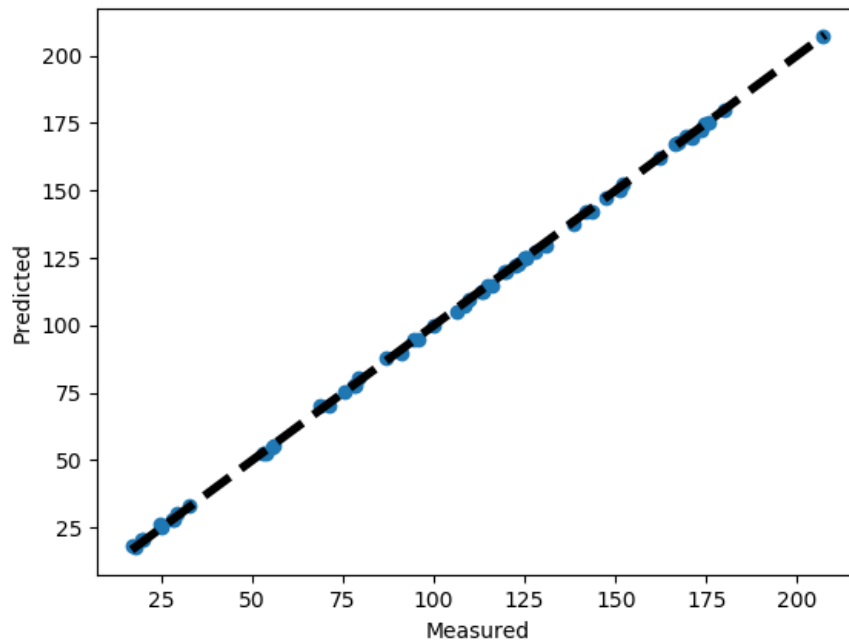
$$y = 0.005cpu_{\%} - 0.002gpu_{\%} - 0.004gpu_c + 2.49gpu_{tdp} + 0.12$$

Όπως φαίνεται και στην παραπάνω εξίσωση, η χρήση του επεξεργαστή και η θερμική σχεδιαστική ισχύ της κάρτας γραφικών έχουν θετικούς συντελεστές, ενώ η χρήση και η θερμοκρασία της κάρτας γραφικών έχουν αρνητικούς συντελεστές. Η θερμική σχεδιαστική ισχύ της κάρτας γραφικών έχει κύριο ρόλο στο μοντέλο, ενώ τα υπόλοιπα χαρακτηριστικά δε το επηρεάζουν σχεδόν καθόλου, ως αμελητέες ποσότητες. Η σταθερά του μοντέλου έχει τιμή 0.12, ενώ το score στο test σετ είναι 0.99. Καθώς αυξάνεται η μεταβλητή α παρατηρείται μεταβολή στο 4ο δεκαδικό ψηφίο των test score (σχήμα 4.27).

Σχήμα 4.27: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (RR)



Σχήμα 4.28: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατά-νάλωσης του επεξεργαστή με 4 εισόδους (RR)

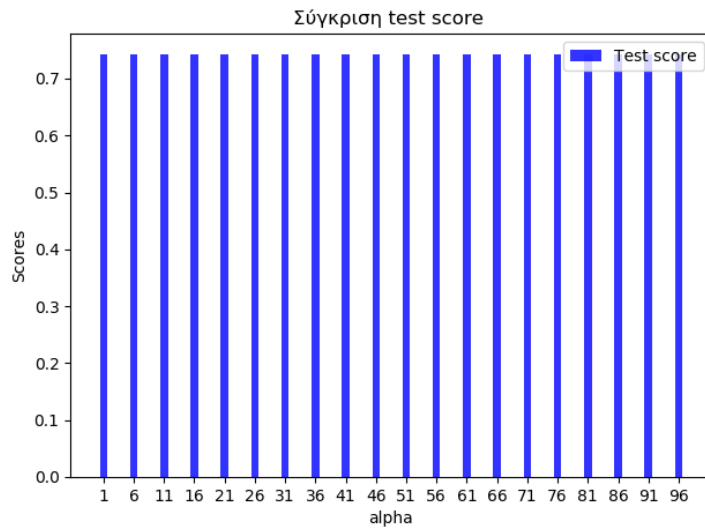


Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση της κάρτας γραφικών, το μοντέλο της παλινδρόμησης ridge με τη μεταβλητή $alpha$ να έχει τη μέγιστη τιμή 96 δίνεται από την παρακάτω εξίσωση:

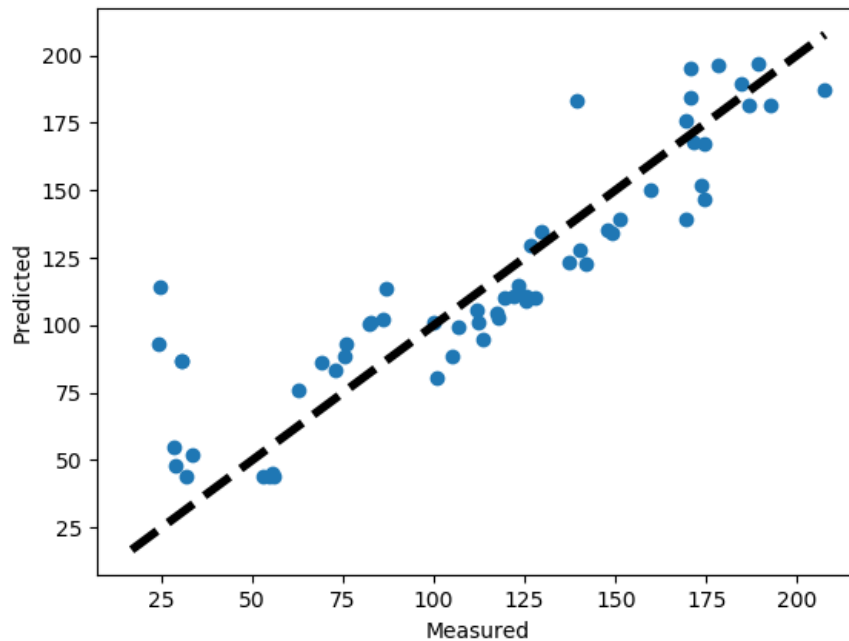
$$y = -0.02cpu\% + 1.59gru\% + 41.49$$

Όπως γίνεται αντιληπτό στην παραπάνω εξίσωση, η χρήση του επεξεργαστή έχει αρνητικό συντελεστή, μην επηρεάζοντας σημαντικά το μοντέλο. Η χρήση της κάρτας γραφικών έχει θετικό συντελεστή, έχοντας κύριο ρόλο στο μοντέλο. Η σταθερά του μοντέλου έχει τιμή 41.49, ενώ το score στο test σετ είναι 0.74. Καθώς αυξάνεται η μεταβλητή $alpha$ παρατηρείται μεταβολή στο 5ο δεκαδικό ψηφίο των test score (σχήμα 4.29).

Σχήμα 4.29: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (RR)



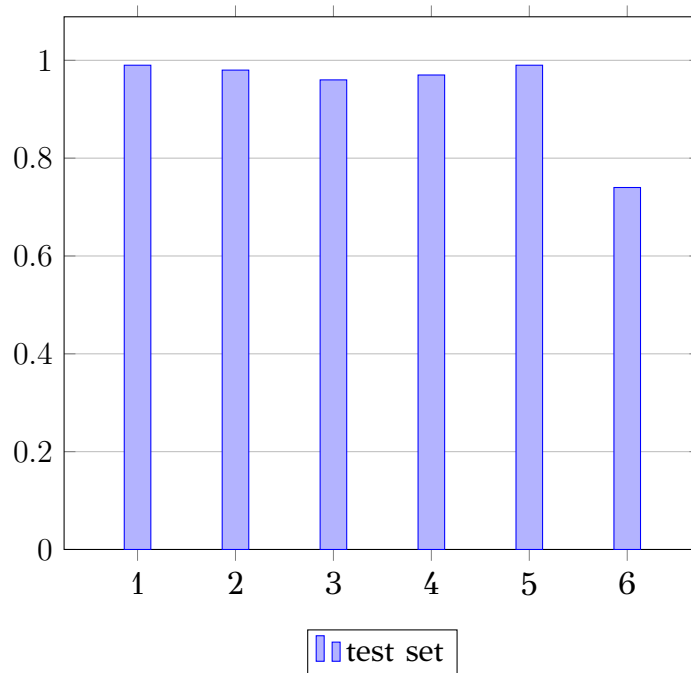
Σχήμα 4.30: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 εισόδους (RR)



Τα σχήματα 4.20, 4.22, 4.24, 4.26, 4.28 και 4.30 απεικονίζουν διαγράμματα πραγματικών με προβλεπόμενων τιμών για κάθε μία από τις παραπάνω περιπτώσεις. Για κάθε σημείο, στον άξονα X αναπαριστάται η πραγματική τιμή του, ενώ στον άξονα Y η προβλεπόμενη τιμή. Όσο πιο κοντά είναι κάποιο σημείο στη γραμμή (μοντέλο), τόσο πιο σωστά πραγματοποιήθηκε η πρόβλεψη.

Στο παρακάτω σχήμα συνοψίζονται συγκεντρωτικά τα scores των train και test

σετ για όλες τις παραπάνω περιπτώσεις με τη σειρά που πραγματοποιήθηκαν. Παρατηρούμε ότι η πρόβλεψη κατανάλωσης ενέργειας της κάρτας γραφικών είχε μεγαλύτερη επιτυχία χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους. Η πρόβλεψη κατανάλωσης ενέργειας του επεξεργαστή ήταν ελάχιστα καλύτερη χρησιμοποιώντας τρία χαρακτηριστικά ως εισόδους. Η πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών ήταν ελάχιστα καλύτερη χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους.



Παλινδρόμηση Lasso

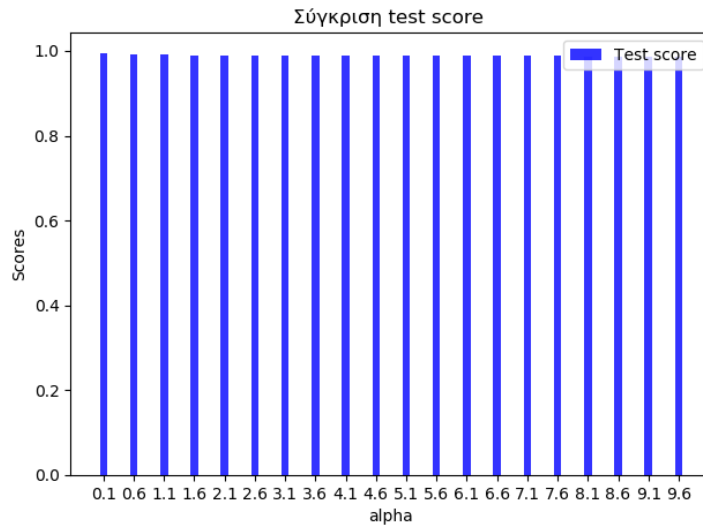
Ο αλγόριθμος περιέχει μια επανάληψη (από 0.1 μέχρι 10 με βήμα 0.5), όπου η μεταβλητή $alpha$ λαμβάνει την τιμή της επανάληψης, για καλύτερη διεξαγωγή συμπερασμάτων. Στο πρώτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στον επεξεργαστή. Χρησιμοποιώντας τρία χαρακτηριστικά ως εισόδους, το μοντέλο της παλινδρόμησης lasso με τη μεταβλητή $alpha$ να έχει τη βέλτιστη τιμή 0.1 δίνεται από την παρακάτω εξίσωση:

$$y = 0.34cpu_{\%} - 0.02gpu_{\%} + 0.22cpu_c + 9.81$$

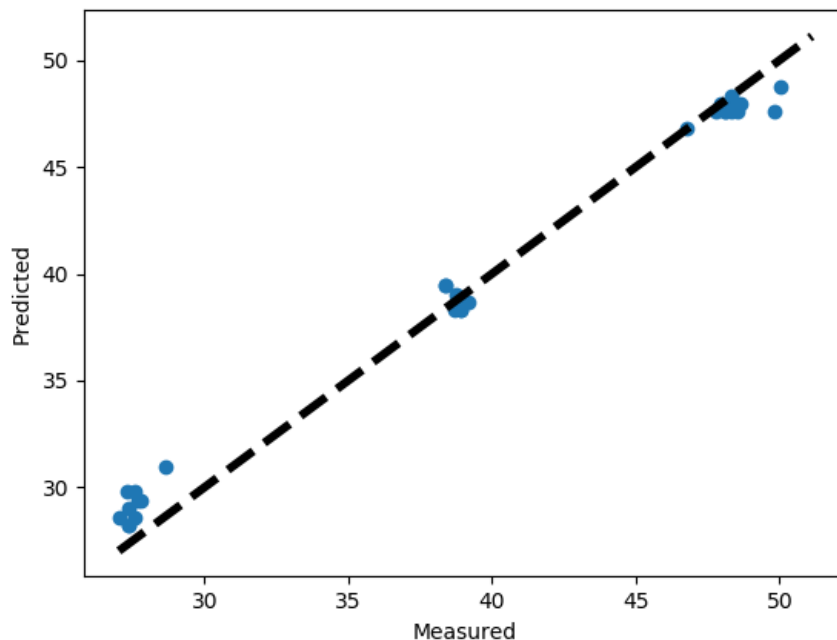
Όπως φαίνεται και στη παραπάνω εξίσωση, η χρήση και η θερμοκρασία του επεξεργαστή έχουν θετικούς συντελεστές. Αντίθετα, η χρήση της κάρτας γραφικών

έχει αρνητικό συντελεστή, συνεισφέροντας λιγότερο στο μοντέλο. Η σταθερά του μοντέλου φέρει τη τιμή 9.81, ενώ το score στο test σετ είναι 0.99. Καθώς αυξάνεται η μεταβλητή $alpha$ παρατηρείται μικρή μεταβολή στο test score (σχήμα 4.31).

Σχήμα 4.31: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (LR)



Σχήμα 4.32: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 3 εισόδους (LR)

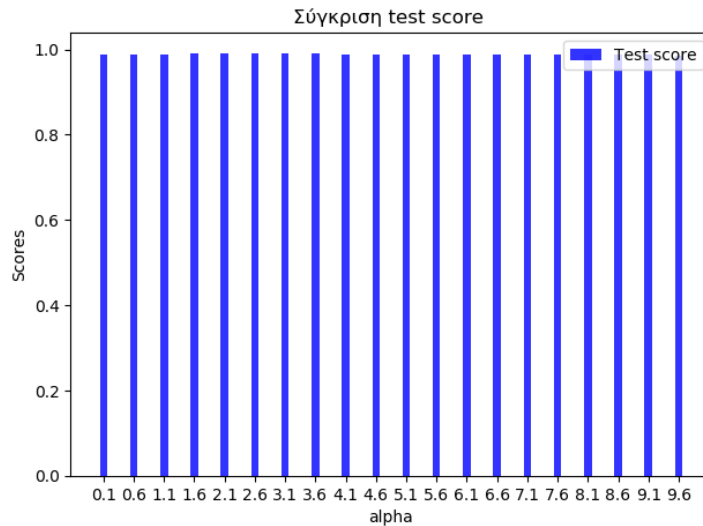


Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση του επεξεργαστή, το μοντέλο της παλινδρόμησης lasso με τη μεταβλητή $alpha$ να έχει τη βέλτιστη τιμή 2.6 δίνεται από την παρακάτω εξίσωση:

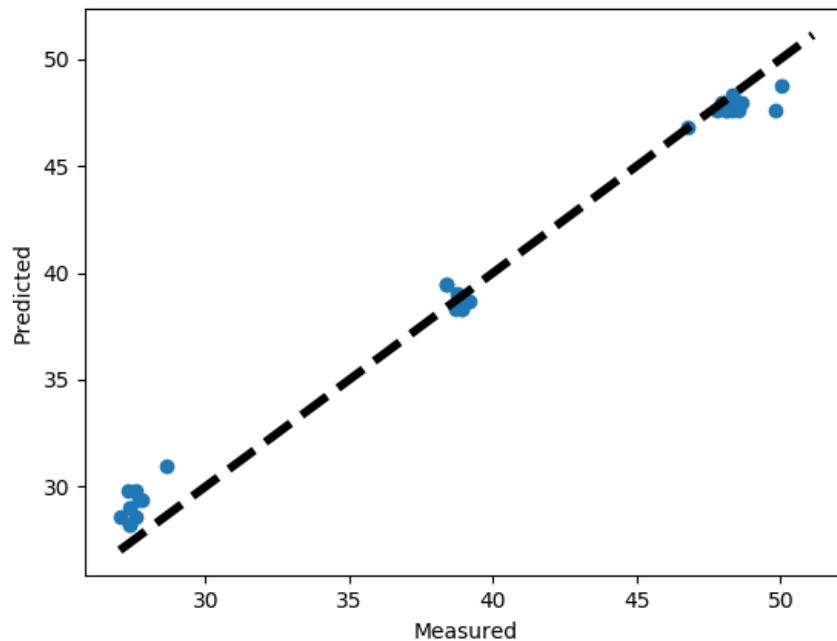
$$y = 0.40cpu\% + 0gru\% + 17.56$$

Όπως φαίνεται στη παραπάνω εξίσωση, η χρήση του επεξεργαστή έχει θετικό συντελεστή ενώ η χρήση της κάρτας γραφικών έχει μηδενικό συντελεστή, με τη χρήση της κάρτας γραφικών να μην επηρεάζει καθόλου το μοντέλο. Η σταθερά του μοντέλου έχει τιμή 17.56, ενώ το score στο test σετ είναι 0.98. Καθώς αυξάνεται η μεταβλητή *alpha* παρατηρείται μεταβολή στο 3ο δεκαδικό ψηφίο των test score (σχήμα 4.33).

Σχήμα 4.33: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (LR)



Σχήμα 4.34: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (LR)

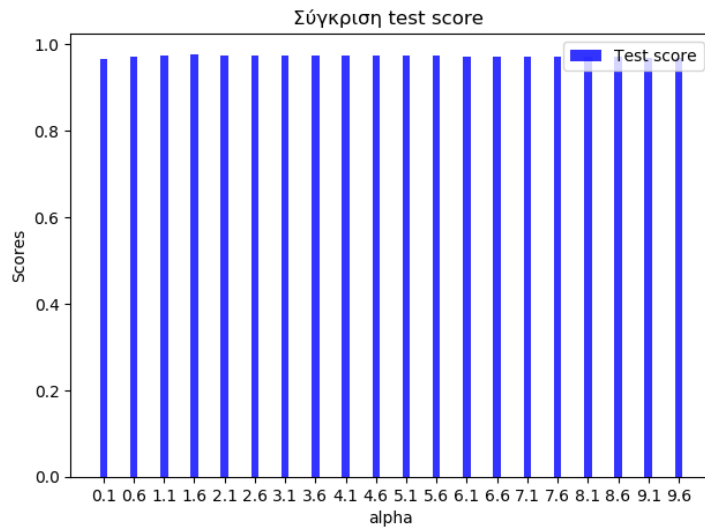


Στο δεύτερο πείραμα κάνουμε πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 3 χαρακτηριστικά ως εισόδους. Το μοντέλο της παλινδρόμησης lasso με τη μεταβλητή $alpha$ να έχει τη βέλτιστη τιμή 1.6 δίνεται από την παρακάτω εξίσωση:

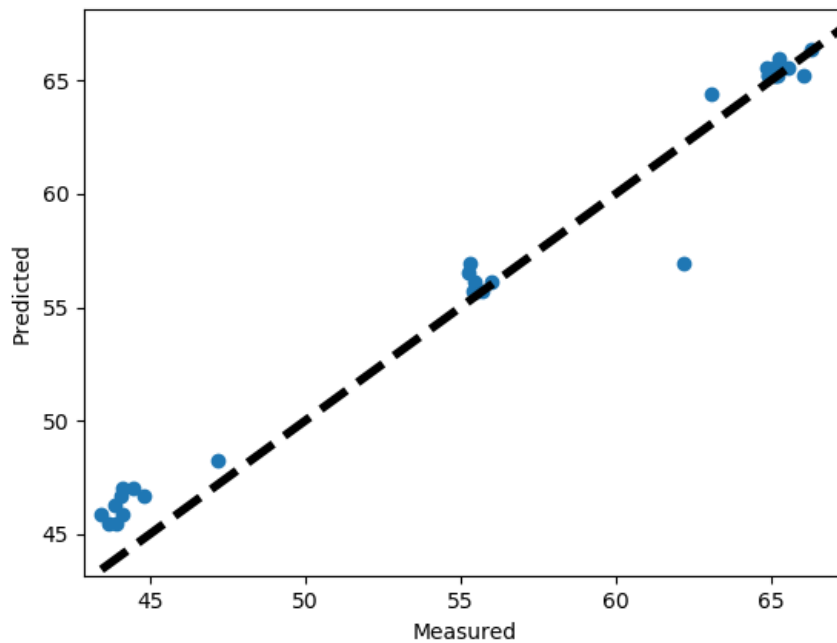
$$y = 0.42cpu_{\%} - 0.55gpu_{\%} - 0.02cpu_c + 37.97$$

Όπως γίνεται αντιληπτό στη παραπάνω εξίσωση, η χρήση του επεξεργαστή έχει θετικό συντελεστή. Αντίθετα, η χρήση της κάρτας γραφικών και η θερμοκρασία του επεξεργαστή έχουν αρνητικούς συντελεστές, με τη θερμοκρασία του επεξεργαστή να επηρεάζει ελάχιστα το μοντέλο. Η σταθερά του μοντέλου φέρει την τιμή 37.97, ενώ το score στο test σετ είναι 0.97. Καθώς αυξάνεται η μεταβλητή $alpha$ παρατηρείται μικρή μεταβολή στο test score (σχήμα 4.35).

Σχήμα 4.35: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (LR)



Σχήμα 4.36: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (LR)

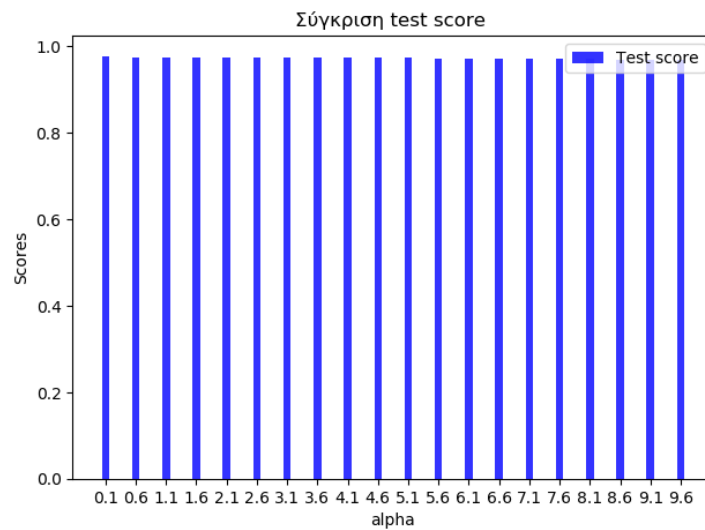


Για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 2 χαρακτηριστικά ως εισόδους, το μοντέλο της παλινδρόμησης lasso με τη μεταβλητή α να έχει τη βέλτιστη τιμή 0.1 δίνεται από την παρακάτω εξίσωση:

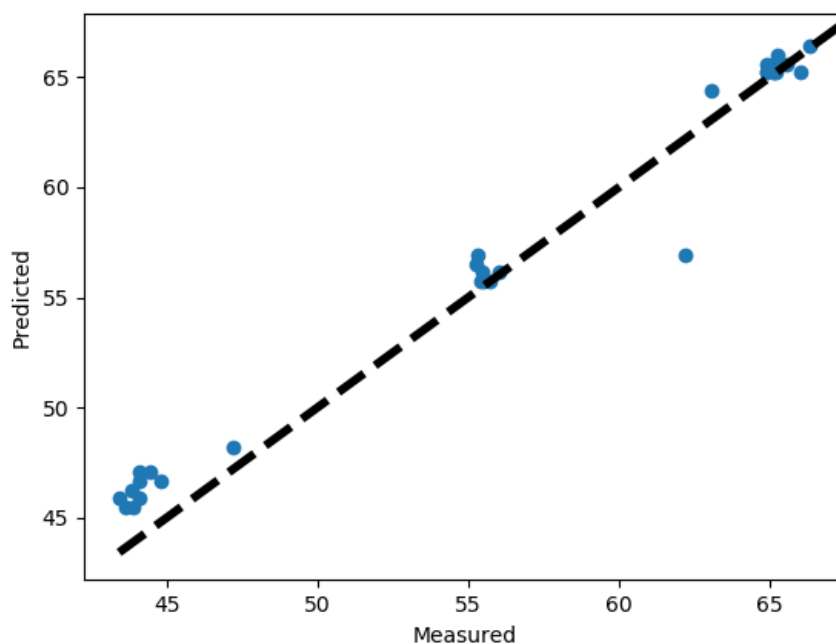
$$y = 0.42cpu\% - 1.35gpu\% + 40.95$$

Παρατηρώντας την παραπάνω εξίσωση βλέπουμε πως, η χρήση του επεξεργαστή έχει θετικό συντελεστή ενώ η χρήση της κάρτας γραφικών έχει αρνητικό συντελεστή, με τη χρήση της κάρτας γραφικών να έχει κύριο ρόλο στο μοντέλο, ενώ η χρήση του επεξεργαστή το επηρεάζει λιγότερο. Η σταθερά του μοντέλου έχει τιμή 40.95, ενώ το score στο test σετ είναι 0.97. Καθώς αυξάνεται η μεταβλητή α παρατηρείται μικρή μεταβολή στο test score (σχήμα 4.37).

Σχήμα 4.37: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (LR)



Σχήμα 4.38: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (LR)

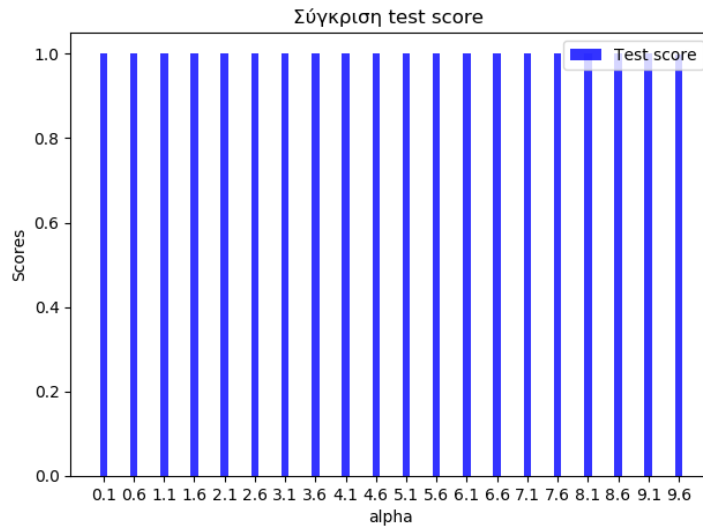


Στο τρίτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στην κάρτα γραφικών. Χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους, το μοντέλο της παλινδρόμησης lasso με τη μεταβλητή $alpha$ να έχει τη βέλτιστη τιμή 0.1 δίνεται από την παρακάτω εξίσωση:

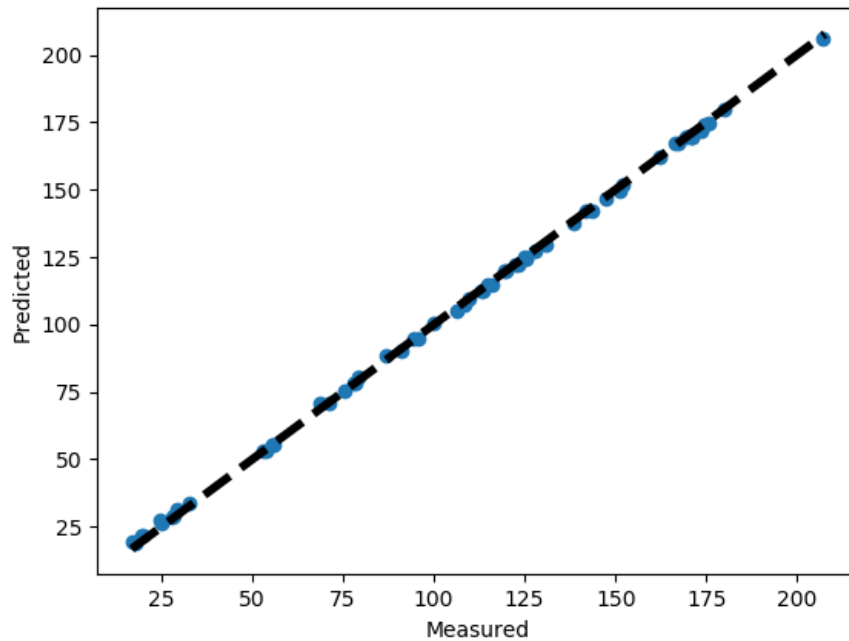
$$y = 0.004cpu\% - 0gpru\% - 0gpru_c + 2.49gpru_{tdp} + 0.017$$

Όπως φαίνεται και στην παραπάνω εξίσωση, η χρήση του επεξεργαστή και η θερμική σχεδιαστική ισχύ της κάρτας γραφικών έχουν θετικούς συντελεστές, ενώ η χρήση και η θερμοκρασία της κάρτας γραφικών έχουν μηδενικούς συντελεστές. Η θερμική σχεδιαστική ισχύ της κάρτας γραφικών έχει κύριο ρόλο στο μοντέλο, ενώ τα υπόλοιπα χαρακτηριστικά δε το επηρεάζουν σχεδόν καθόλου. Η σταθερά του μοντέλου έχει τιμή 0.017, ενώ το score στο test σετ είναι 0.99. Καθώς αυξάνεται η μεταβλητή $alpha$ παρατηρείται μεταβολή στο 4ο δεκαδικό ψηφίο των test score (σχήμα 4.39).

Σχήμα 4.39: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (LR)



Σχήμα 4.40: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 4 εισόδους (LR)

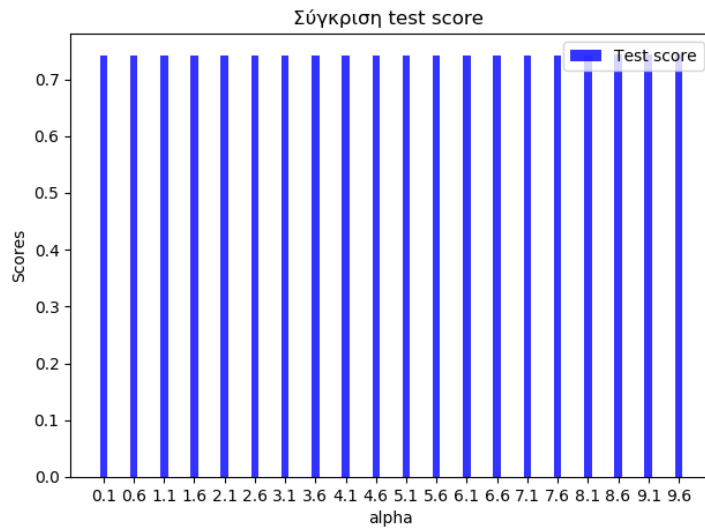


Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση της κάρτας γραφικών, το μοντέλο της παλινδρόμησης lasso με τη μεταβλητή $alpha$ να έχει τη βέλτιστη τιμή 9.6 δίνεται από την παρακάτω εξίσωση:

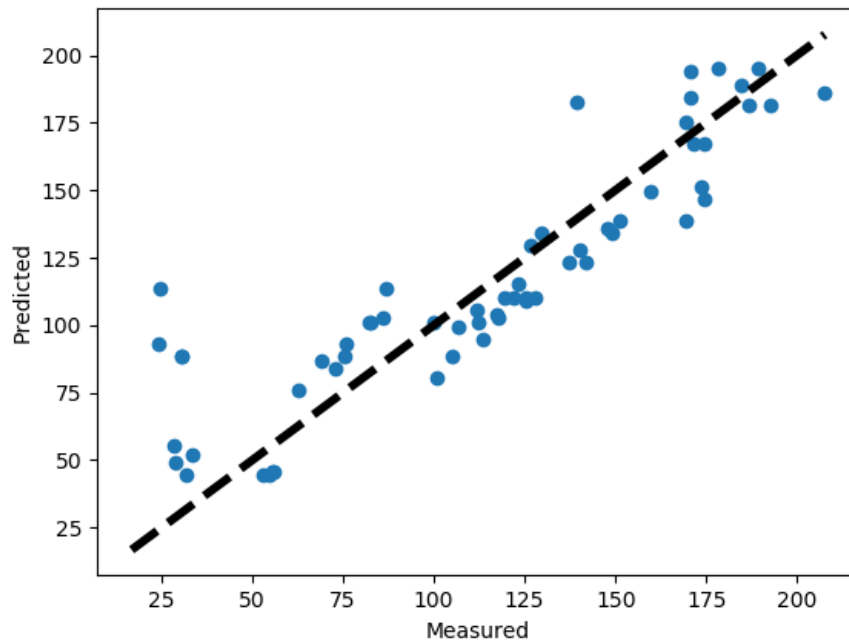
$$y = -0.37u_{\%} + 1.57g_{\%} + 41.16$$

Όπως γίνεται αντιληπτό στην παραπάνω εξίσωση, η χρήση του επεξεργαστή έχει μηδενικό συντελεστή, μην επηρεάζοντας το μοντέλο. Η χρήση της κάρτας γραφικών έχει θετικό συντελεστή, έχοντας κύριο ρόλο στο μοντέλο. Η σταθερά του μοντέλου έχει τιμή 41.16, ενώ το score στο test σετ είναι 0.74. Καθώς αυξάνεται η μεταβλητή $alpha$ παρατηρείται μεταβολή στο 3ο δεκαδικό ψηφίο των test score (σχήμα 4.41).

Σχήμα 4.41: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (LR)



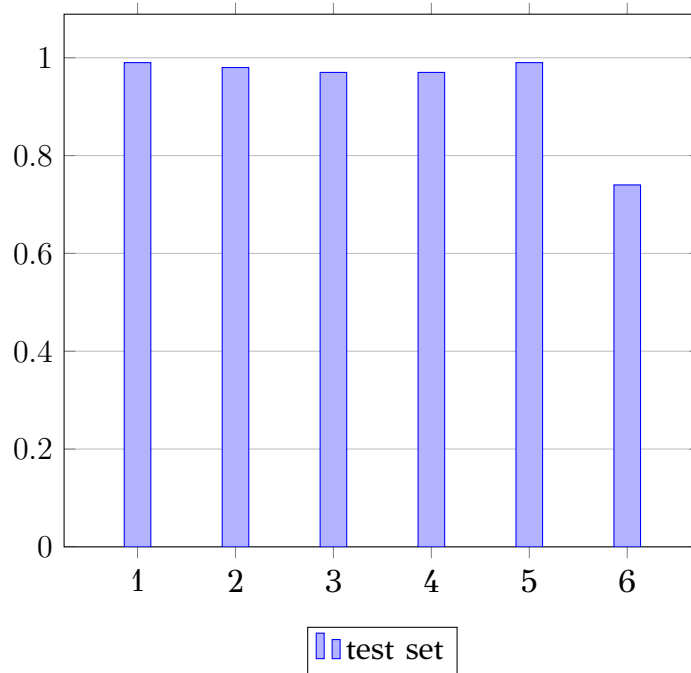
Σχήμα 4.42: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 εισόδους (LR)



Τα σχήματα 4.32, 4.34, 4.36, 4.38, 4.40 και 4.42 απεικονίζουν διαγράμματα πραγματικών με προβλεπόμενων τιμών για κάθε μία από τις παραπάνω περιπτώσεις. Για κάθε σημείο, στον άξονα X αναπαριστάται η πραγματική τιμή του, ενώ στον άξονα Y η προβλεπόμενη τιμή. Όσο πιο κοντά είναι κάποιο σημείο στη γραμμή (μοντέλο), τόσο πιο σωστά πραγματοποιήθηκε η πρόβλεψη.

Στο παρακάτω σχήμα συνοψίζονται συγκεντρωτικά τα scores των train και test

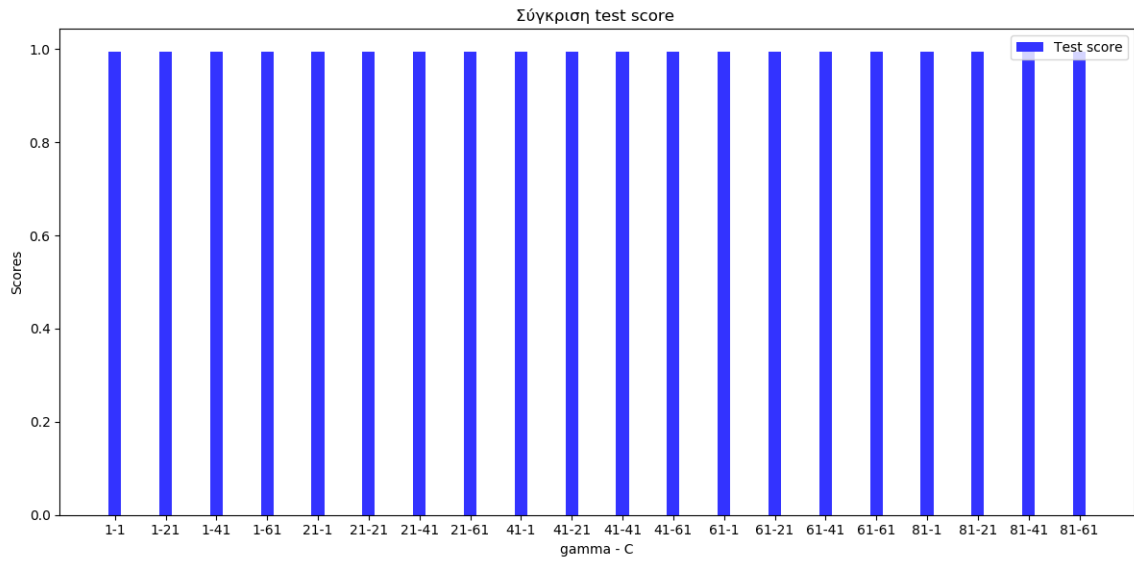
σετ για όλες τις παραπάνω περιπτώσεις με τη σειρά που πραγματοποιήθηκαν. Παρατηρούμε ότι η πρόβλεψη κατανάλωσης ενέργειας της κάρτας γραφικών είχε μεγαλύτερη επιτυχία χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους. Η πρόβλεψη κατανάλωσης ενέργειας του επεξεργαστή ήταν ελάχιστα καλύτερη χρησιμοποιώντας τρία χαρακτηριστικά ως εισόδους. Η πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών είχε το ίδιο αποτέλεσμα τόσο για δύο όσο και για τρία χαρακτηριστικά ως εισόδους.



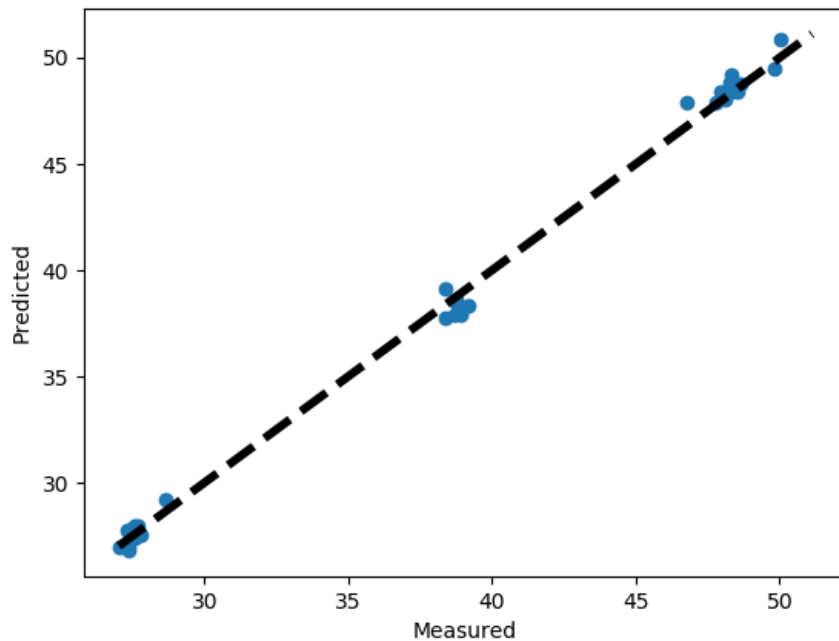
Μηχανές Διανυσμάτων Υποστήριξης

Ο αλγόριθμος περιέχει δύο επαναλήψεις (από 1 μέχρι 100 με βήμα 20 για τη μεταβλητή γ και από 1 μέχρι 80 με βήμα 20 για τη μεταβλητή C), όπου οι μεταβλητές γ και C λαμβάνουν την τιμή των επαναλήψεων, για καλύτερη διεξαγωγή συμπερασμάτων. Στο πρώτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στον επεξεργαστή. Χρησιμοποιώντας τρία χαρακτηριστικά ως εισόδους, ορίζουμε τις μεταβλητές γ και C στη τιμή 1 και λαμβάνουμε test score 0.99 (σχήμα 4.43).

Σχήμα 4.43: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (SVM)

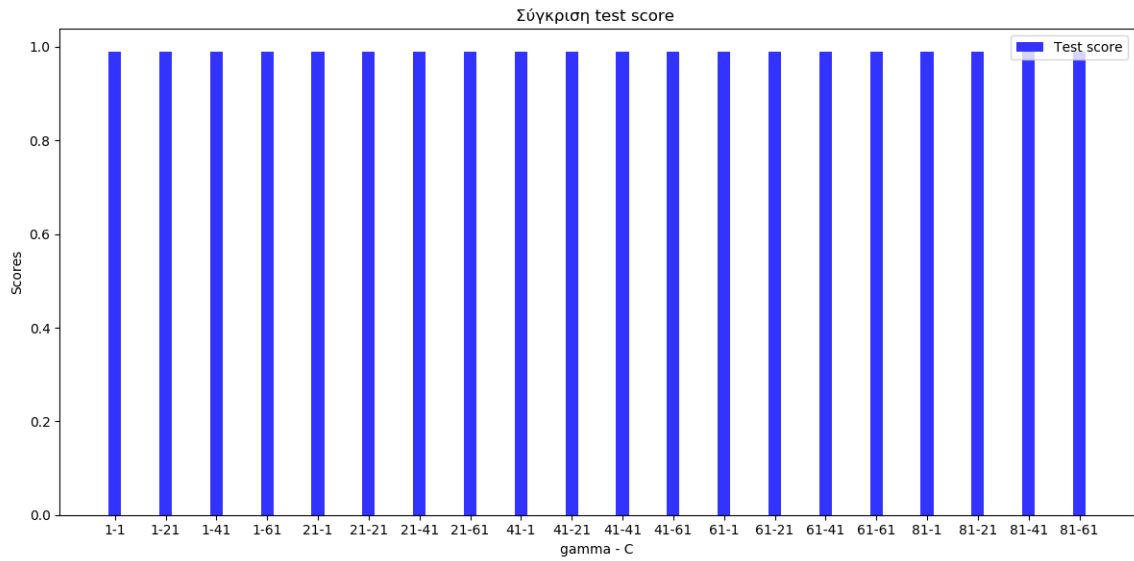


Σχήμα 4.44: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 3 εισόδους (SVM)

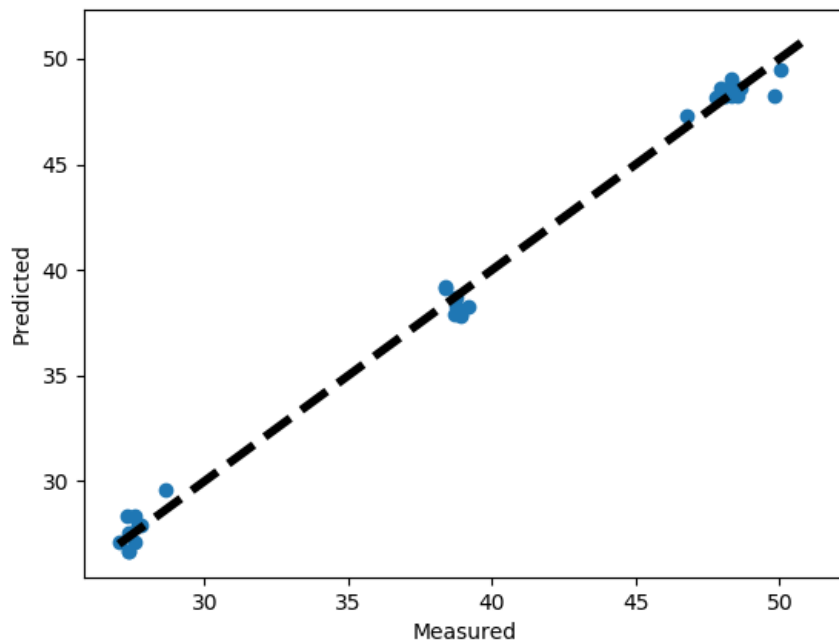


Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση του επεξεργαστή, ορίζουμε τις μεταβλητές $gamma$ και C στη τιμή 1 και λαμβάνουμε test score 0.98 (σχήμα 4.45).

Σχήμα 4.45: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (SVM)

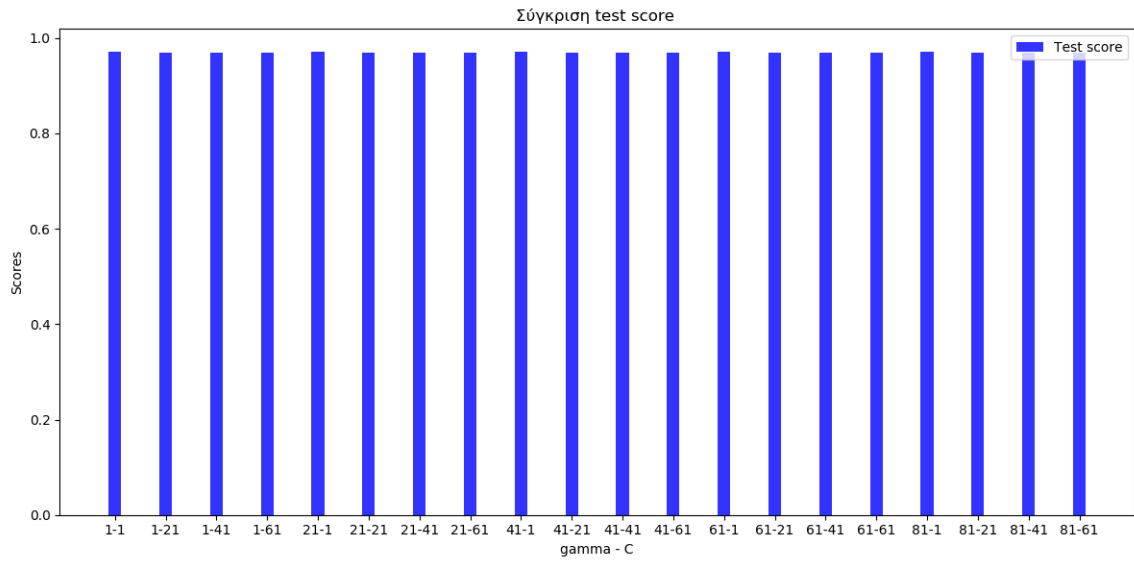


Σχήμα 4.46: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (SVM)

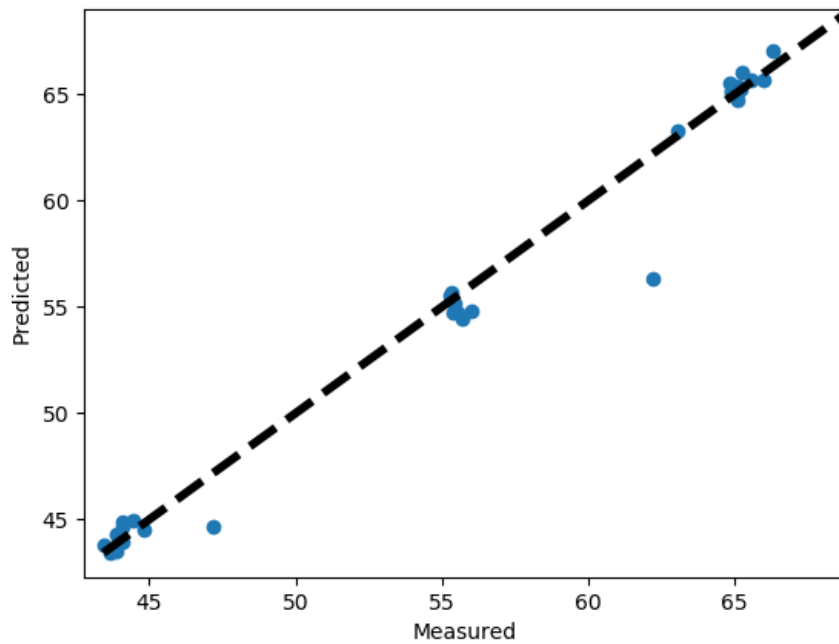


Στο δεύτερο πείραμα κάνουμε πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιούμε 3 χαρακτηριστικά ως εισόδους. Ορίζουμε τις μεταβλητές $gamma$ και C στη τιμή 1 και λαμβάνουμε test score 0.97 (σχήμα 4.47).

Σχήμα 4.47: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (SVM)

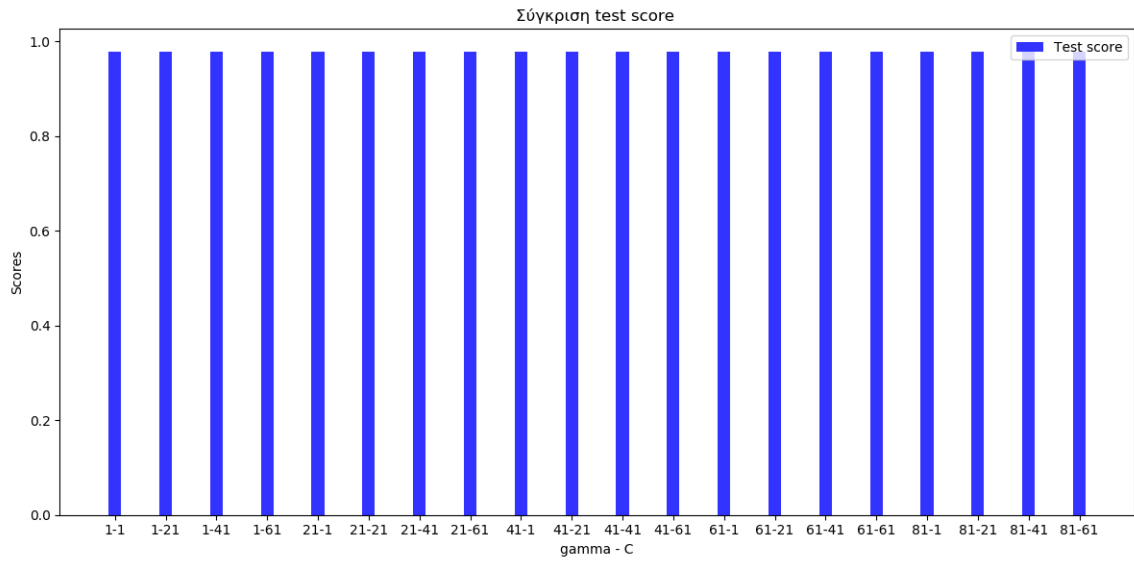


Σχήμα 4.48: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (SVM)

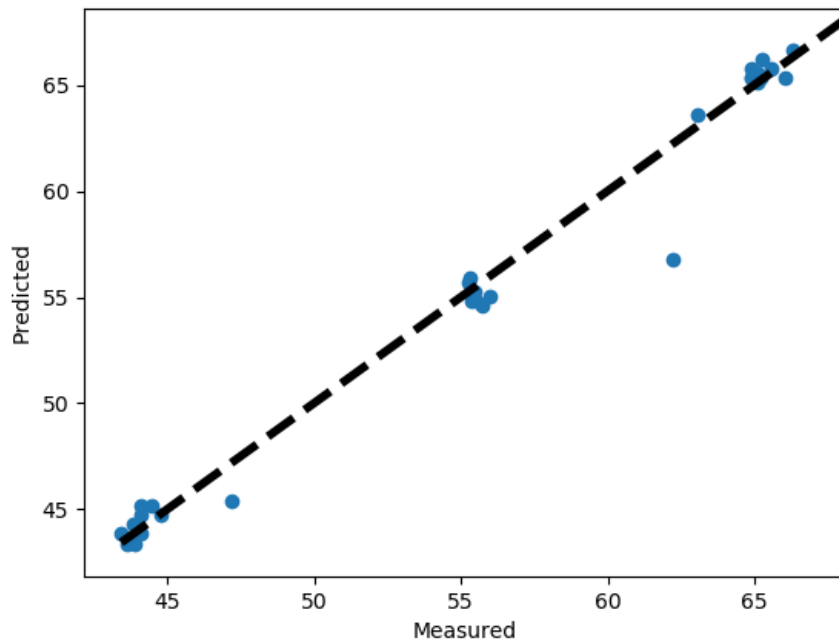


Για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιούμε 2 χαρακτηριστικά ως εισόδους. Ορίζουμε τις μεταβλητές $gamma$ και C στη τιμή 1 και λαμβάνουμε test score 0.97 (σχήμα 4.49).

Σχήμα 4.49: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (SVM)

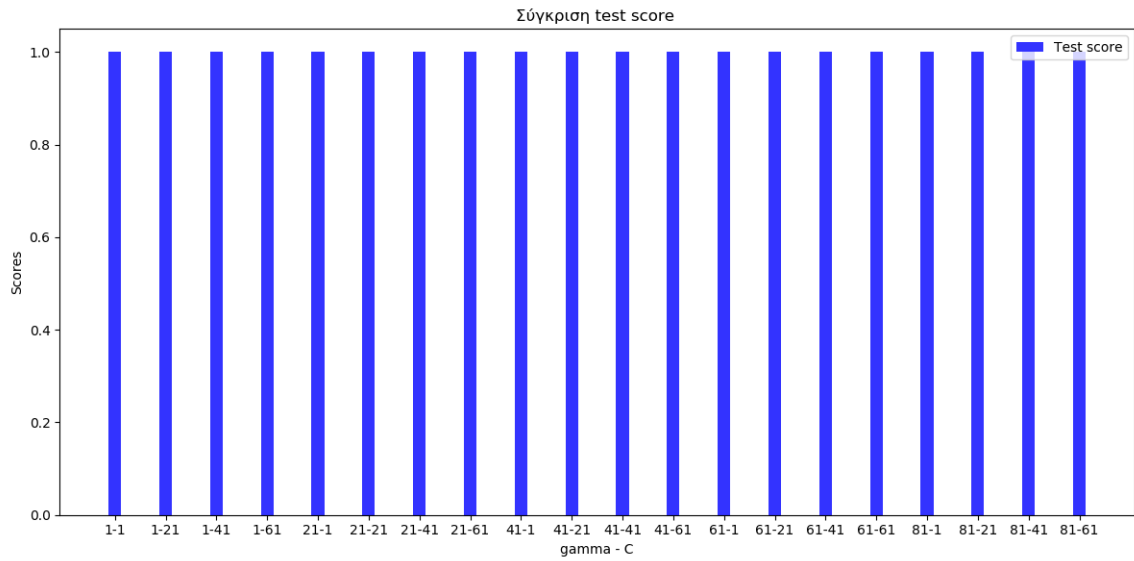


Σχήμα 4.50: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (SVM)

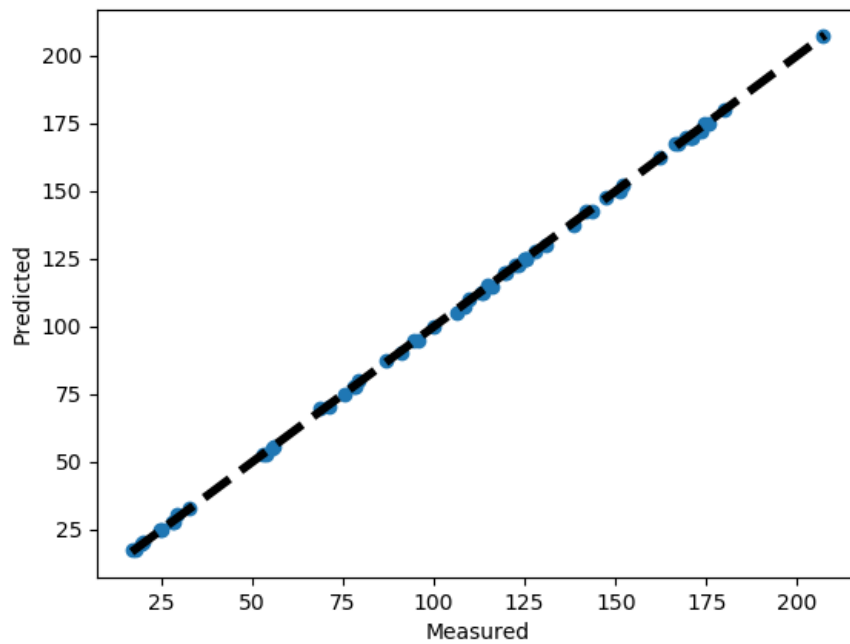


Στο τρίτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στην κάρτα γραφικών. Χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους, ορίζουμε τις μεταβλητές $gamma$ και C στη τιμή 1 και λαμβάνουμε test score 0.99 (σχήμα 4.51).

Σχήμα 4.51: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (SVM)

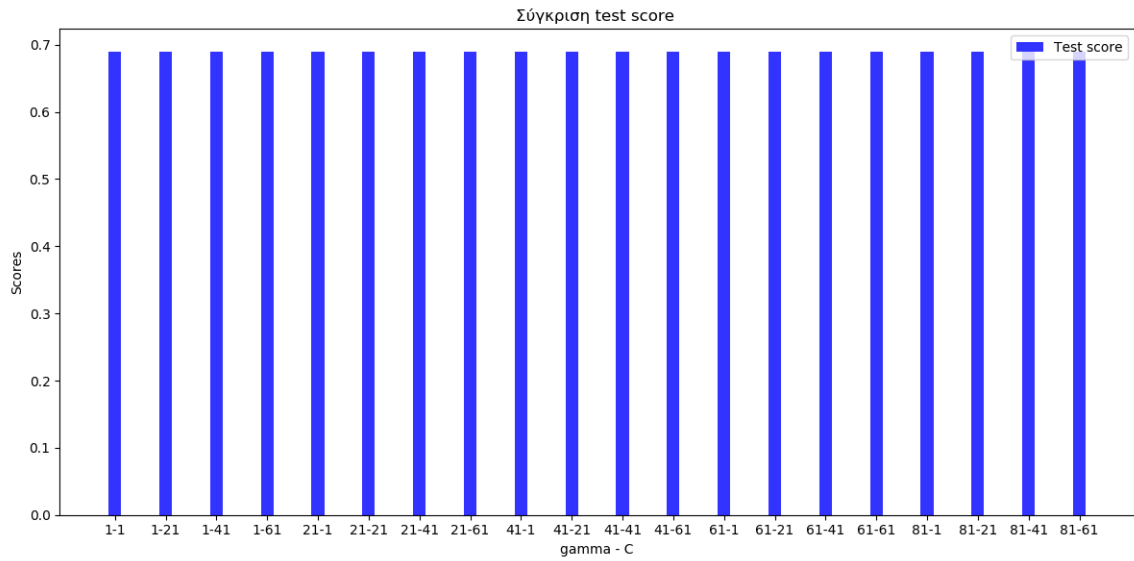


Σχήμα 4.52: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 4 εισόδους (SVM)

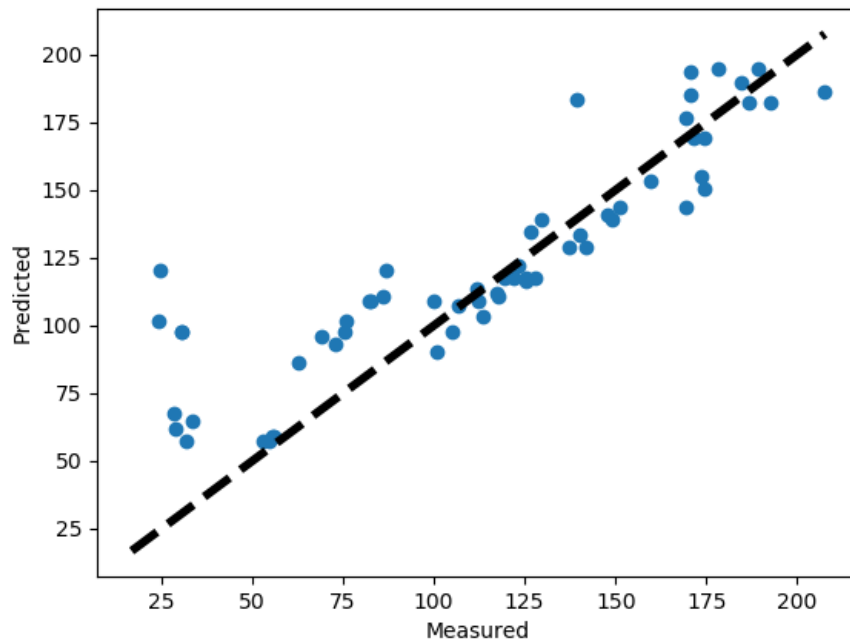


Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση της κάρτας γραφικών, ορίζουμε τις μεταβλητές $gamma$ και C στη τιμή 1 και λαμβάνουμε test score 0.68 (σχήμα 4.53).

Σχήμα 4.53: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (SVM)



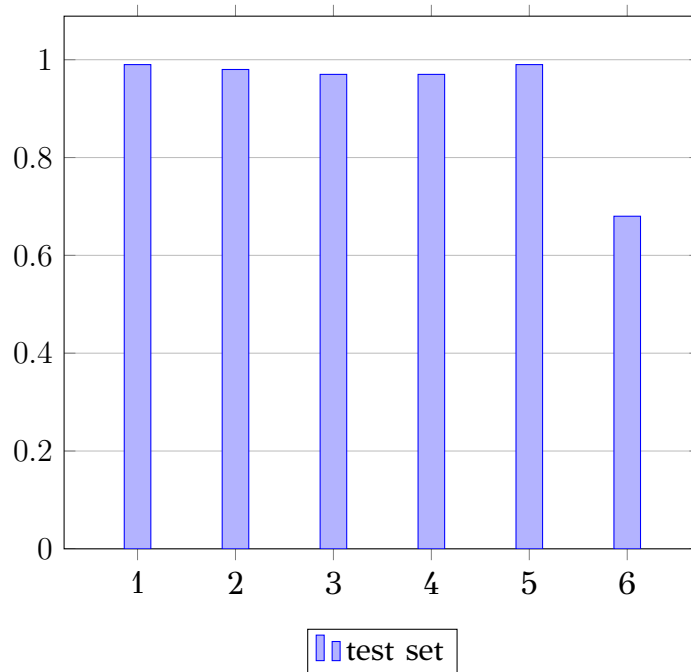
Σχήμα 4.54: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 εισόδους (SVM)



Τα σχήματα 4.32, 4.34, 4.36, 4.38, 4.40 και 4.42 απεικονίζουν διαγράμματα πραγματικών με προβλεπόμενων τιμών για κάθε μία από τις παραπάνω περιπτώσεις. Για κάθε σημείο, στον άξονα X αναπαριστάται η πραγματική τιμή του, ενώ στον άξονα Y η προβλεπόμενη τιμή. Όσο πιο κοντά είναι κάποιο σημείο στη γραμμή (μοντέλο), τόσο πιο σωστά πραγματοποιήθηκε η πρόβλεψη.

Στο παρακάτω σχήμα συνοψίζονται συγκεντρωτικά τα scores των train και test

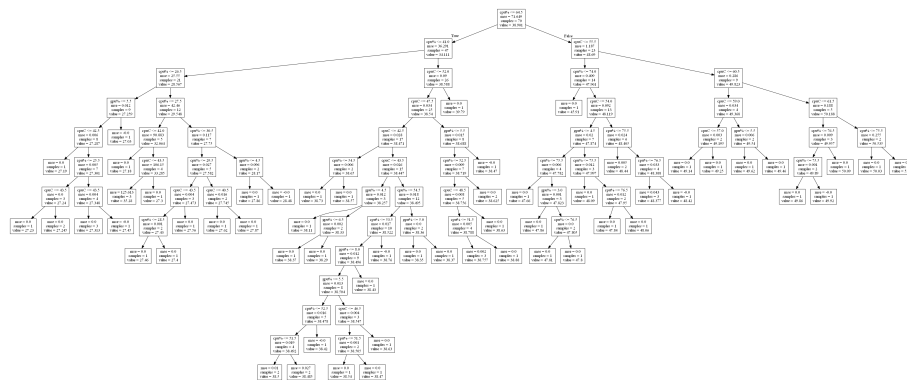
σετ για όλες τις παραπάνω περιπτώσεις. Παρατηρούμε ότι η πρόβλεψη κατανάλωσης ενέργειας της κάρτας γραφικών είχε μεγαλύτερη επιτυχία χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους. Η πρόβλεψη κατανάλωσης ενέργειας του επεξεργαστή ήταν ελάχιστα καλύτερη χρησιμοποιώντας τρία χαρακτηριστικά ως εισόδους. Η πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών είχε το ίδιο αποτέλεσμα τόσο για δύο όσο και για τρία χαρακτηριστικά ως εισόδους.



Δέντρα Απόφασης

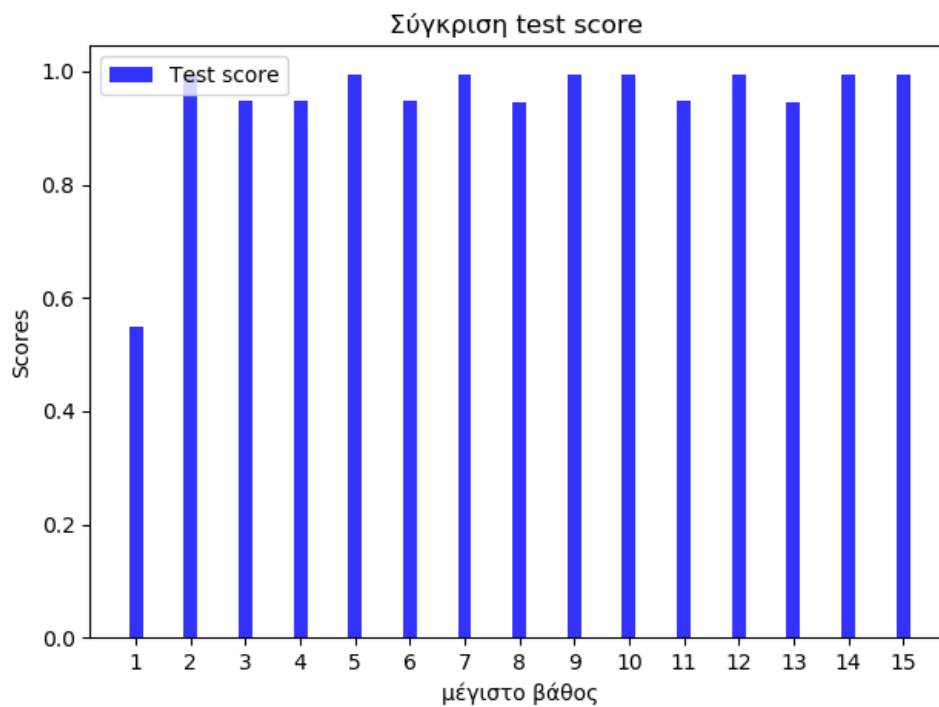
Ο αλγόριθμος περιέχει μια επανάληψη (από 1 μέχρι 15 με βήμα 1), η οποία προσδιορίζει το βάθος το οποίο θα φτάσει το δέντρο απόφασης (η μεταβλητή *max_depth* λαμβάνει την τιμή της επανάληψης), για την καλύτερη διεξαγωγή συμπερασμάτων. Στο πρώτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στον επεξεργαστή. Χρησιμοποιώντας τρία χαρακτηριστικά ως εισόδους, το δέντρο απόφασης φαίνεται στο σχήμα 4.55.

Σχήμα 4.55: Δέντρο απόφασης για τη πρόβλεψη κατανάλωσης ενέργειας στον επεξεργαστή με 3 χαρακτηριστικά (DT)

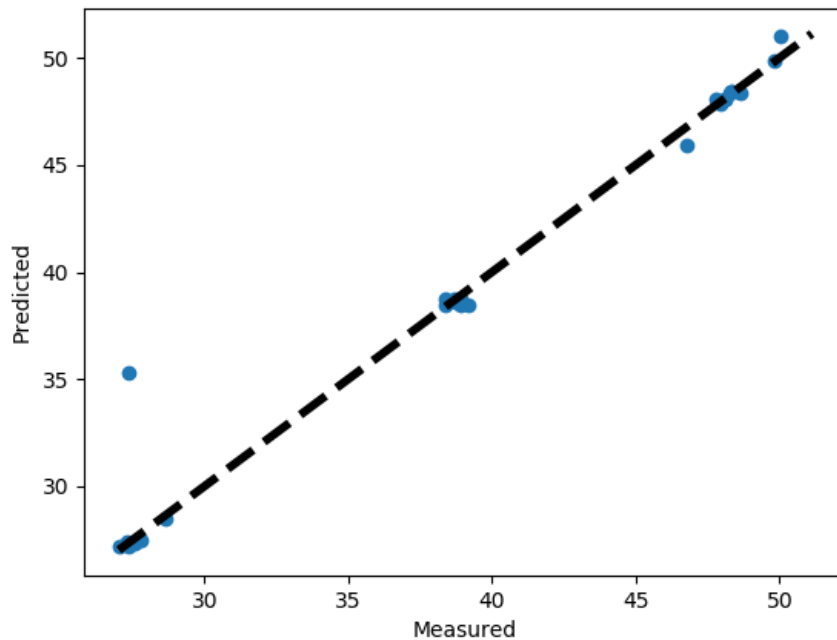


Όπως γίνεται αντιληπτό στο σχήμα 4.56, καθώς αυξάνεται το βάθος του δέντρου απόφασης το test score αυξομειώνεται. Για βάθος δέντρου 12 το test score έχει τη βέλτιστη τιμή του (0.99).

Σχήμα 4.56: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (DT)

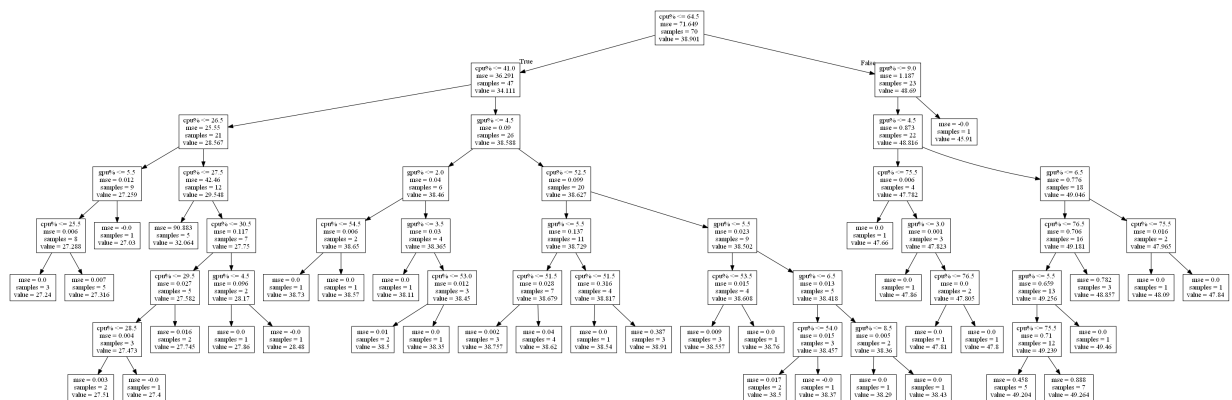


Σχήμα 4.57: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 3 εισόδους (DT)



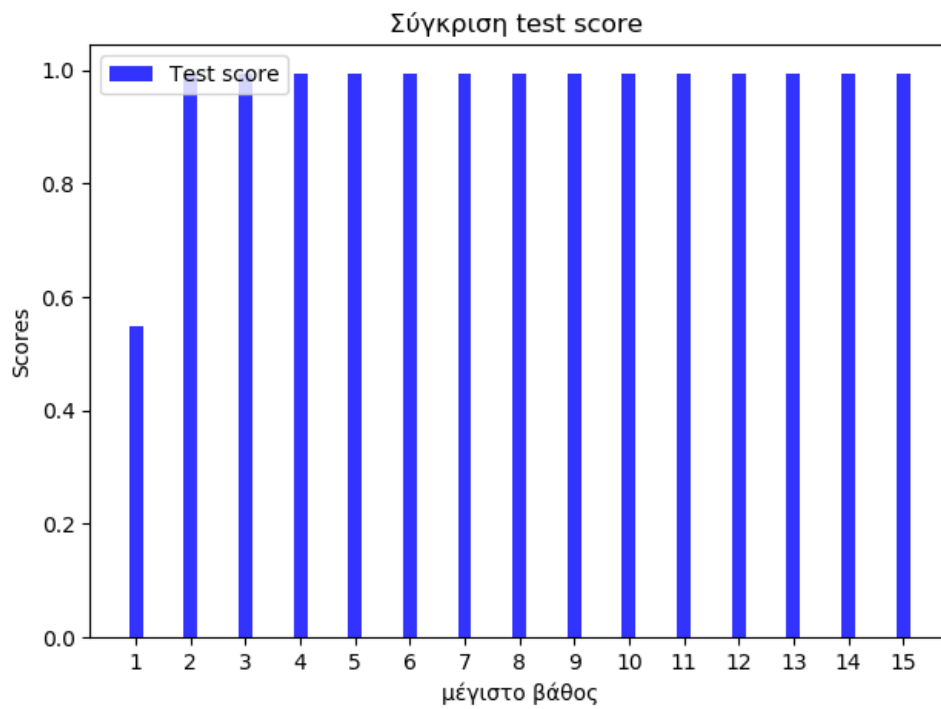
Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση του επεξεργαστή, το δέντρο απόφασης για τη πρόβλεψη κατανάλωσης ενέργειας στον επεξεργαστή φαίνεται στο σχήμα 4.58.

Σχήμα 4.58: Δέντρο απόφασης για τη πρόβλεψη κατανάλωσης ενέργειας στον επεξεργαστή με 2 χαρακτηριστικά (DT)

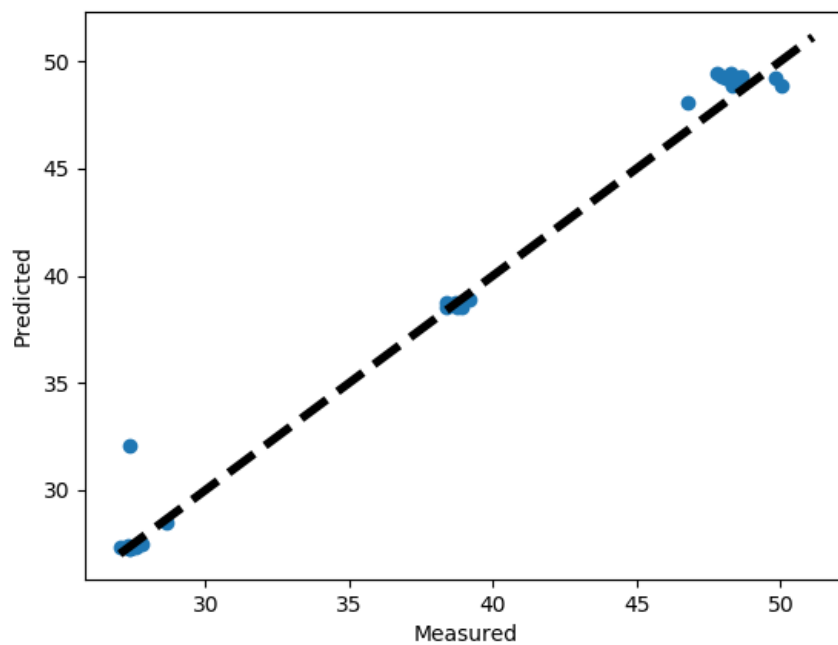


Όπως φαίνεται και στο σχήμα 4.59, καθώς αυξάνεται το βάθος του δέντρου απόφασης το test score αυξομειώνεται. Για βάθος δέντρου 7 το test score έχει τη βέλτιστη τιμή του (0.99).

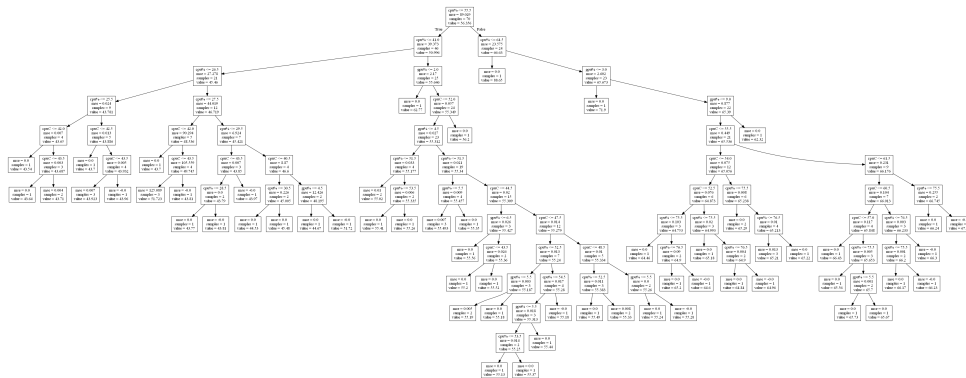
Σχήμα 4.59: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (DT)



Σχήμα 4.60: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (DT)

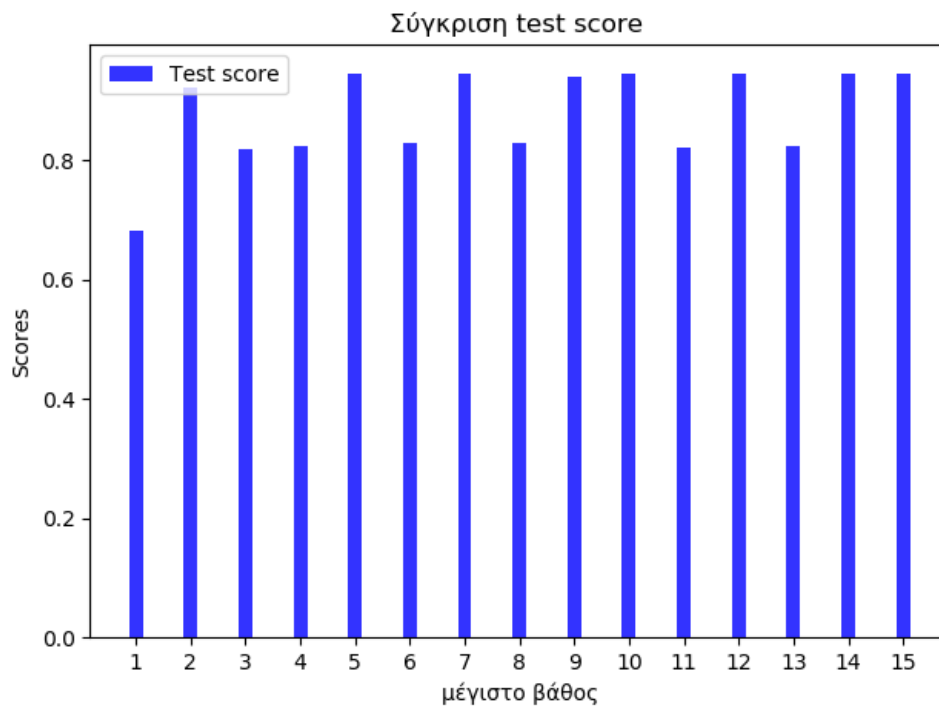


Σχήμα 4.61: Δέντρο απόφασης για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (DT)

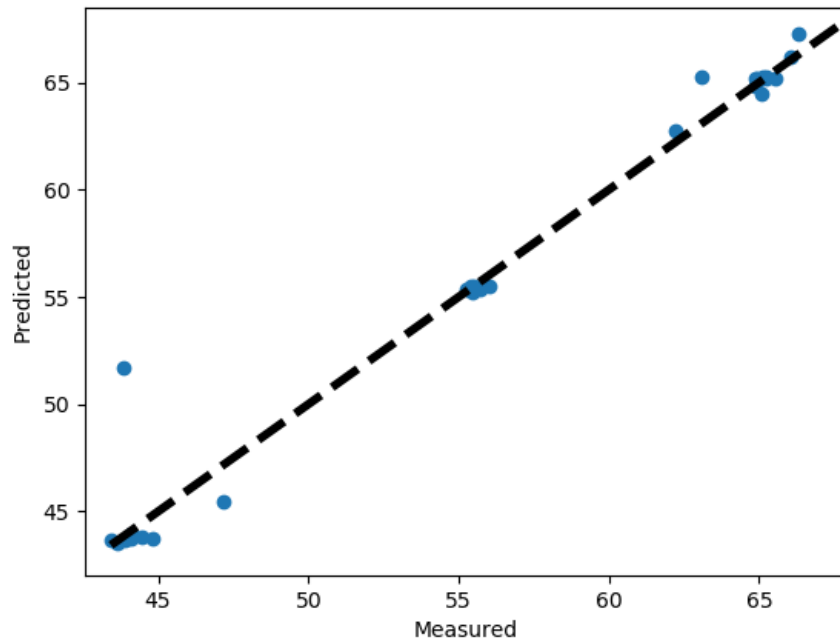


Στο δεύτερο πείραμα κάνουμε πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 3 χαρακτηριστικά ως εισόδους, το δέντρο απόφασης δίνεται από το σχήμα 4.61. Όπως είναι εμφανές στο σχήμα 4.62, καθώς αυξάνεται το βάθος του δέντρου απόφασης, το test score αυξομειώνεται. Για βάθος δέντρου 12 το test score έχει τη βέλτιστη τιμή του (0.94).

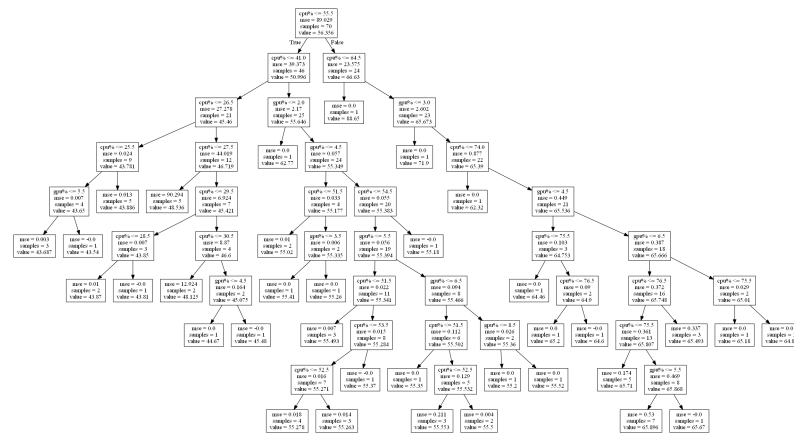
Σχήμα 4.62: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (DT)



Σχήμα 4.63: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (DT)

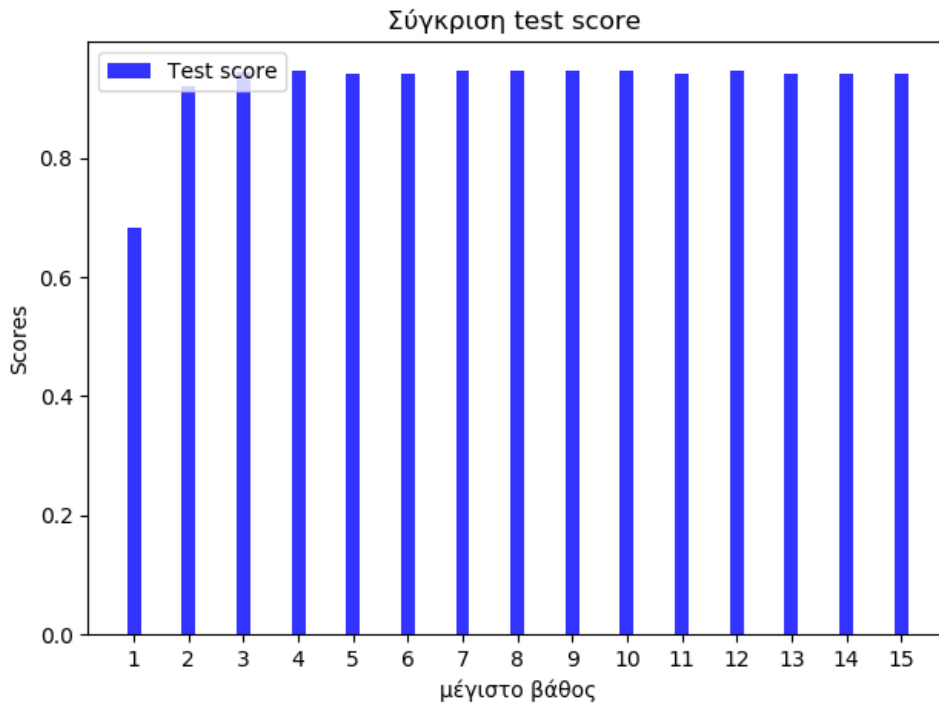


Σχήμα 4.64: Δέντρο απόφασης για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (DT)

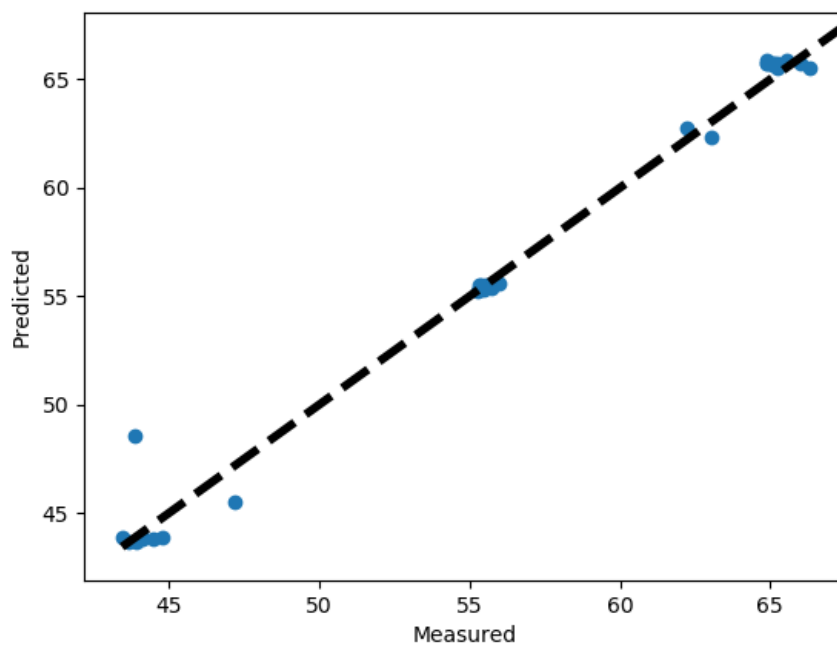


Για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 2 χαρακτηριστικά ως εισόδους, το δέντρο απόφασης δίνεται από το σχήμα 4.64. Όπως παρατηρείται και στο σχήμα 4.65, καθώς αυξάνεται το βάθος του δέντρου απόφασης, το test score αυξομειώνονται. Η βέλτιστη τιμή του test score βρίσκεται σε βάθος 9 με τιμή 0.94.

Σχήμα 4.65: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (DT)



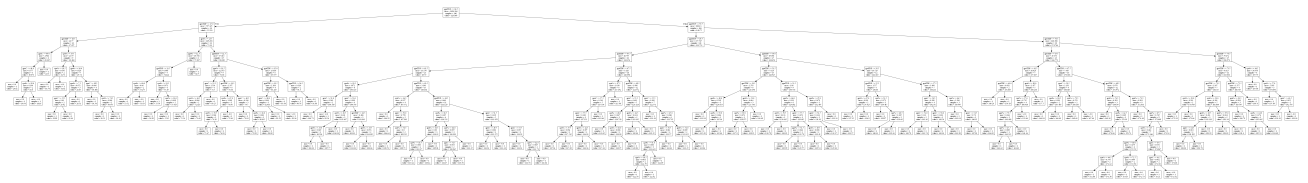
Σχήμα 4.66: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (DT)



Στο τρίτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στην κάρτα γραφικών. Χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους, το δέντρο απόφασης για τη πρόβλεψη κατανάλωσης ενέργειας στον επεξεργαστή φαίνεται στο

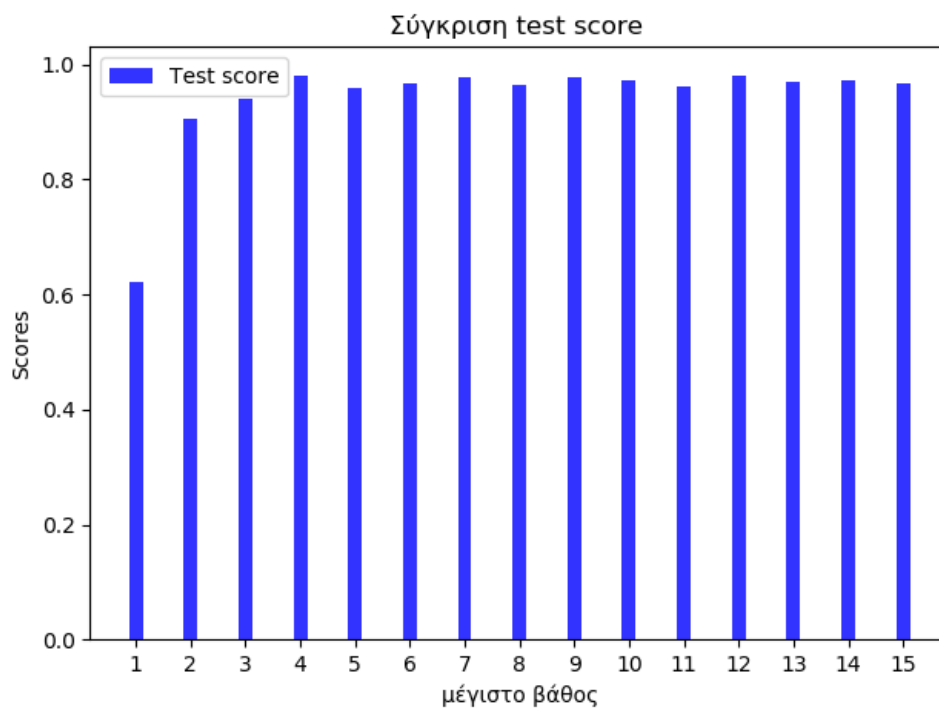
σχήμα 4.67.

Σχήμα 4.67: Δέντρο απόφασης για τη πρόβλεψη κατανάλωσης ενέργειας στην κάρτα γραφικών με 4 χαρακτηριστικά (DT)

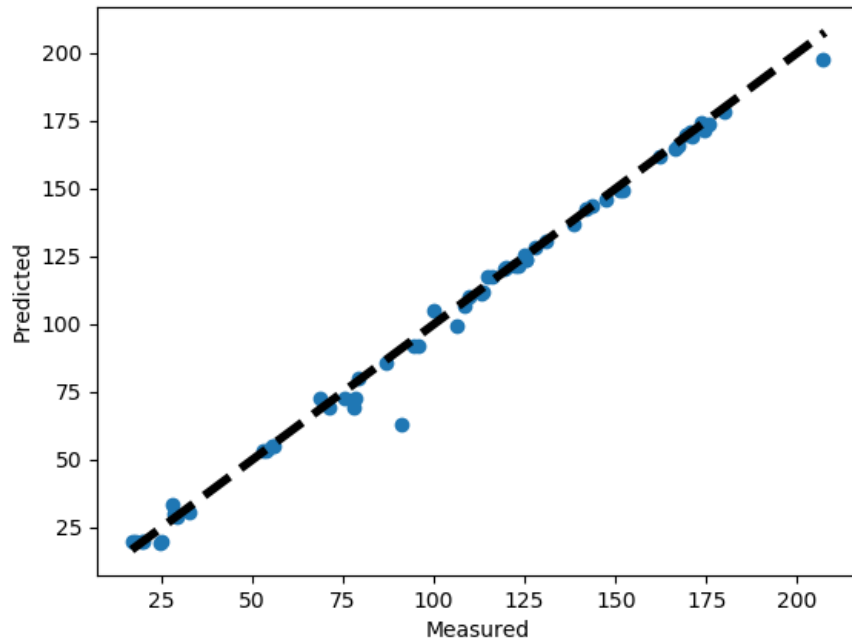


Όπως φαίνεται και στο σχήμα 4.68, καθώς αυξάνεται το βάθος του δέντρου απόφασης το test score αυξομειώνεται. Για βάθος δέντρου 12 το test score έχει τη βέλτιστη τιμή του (0.98).

Σχήμα 4.68: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (DT)

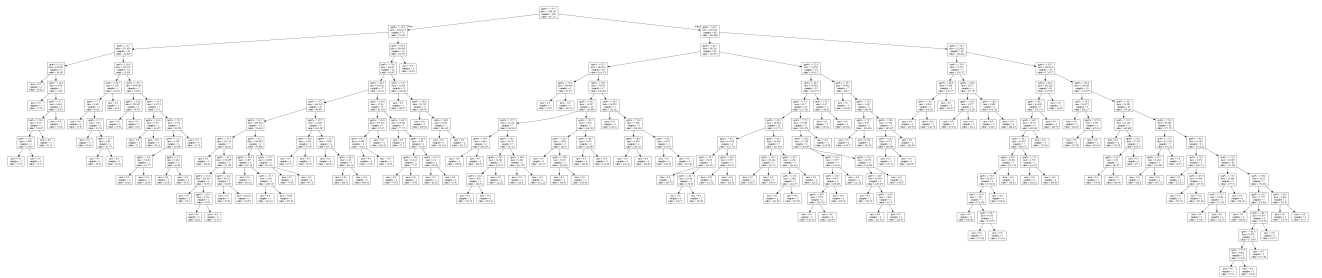


Σχήμα 4.69: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 4 εισόδους (DT)



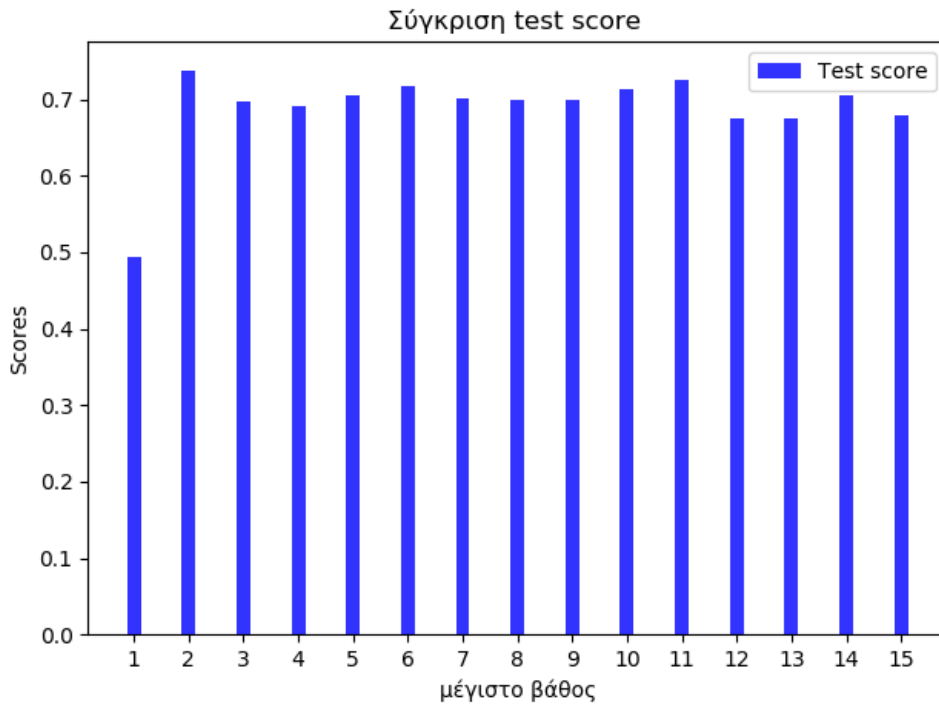
Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση της κάρτας γραφικών, το δέντρο απόφασης φαίνεται στο σχήμα 4.70.

Σχήμα 4.70: Δέντρο απόφασης για τη πρόβλεψη κατανάλωσης ενέργειας στην κάρτα γραφικών με 2 χαρακτηριστικά (DT)

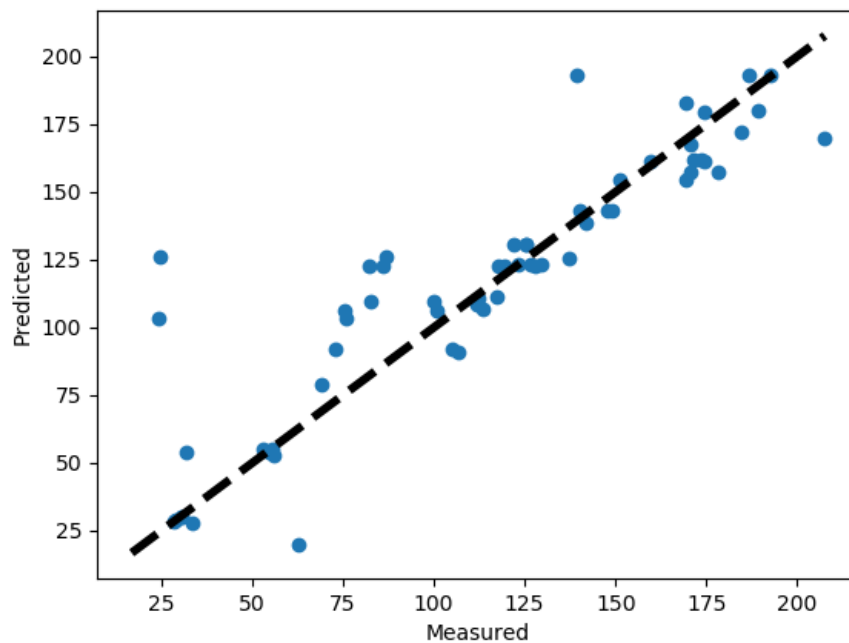


Όπως φαίνεται και στο σχήμα 4.71, καθώς αυξάνεται το βάθος του δέντρου απόφασης το test score αυξομειώνεται. Για βάθος δέντρου 2 το test score έχει τη βέλτιστη τιμή του (0.73).

Σχήμα 4.71: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (DT)



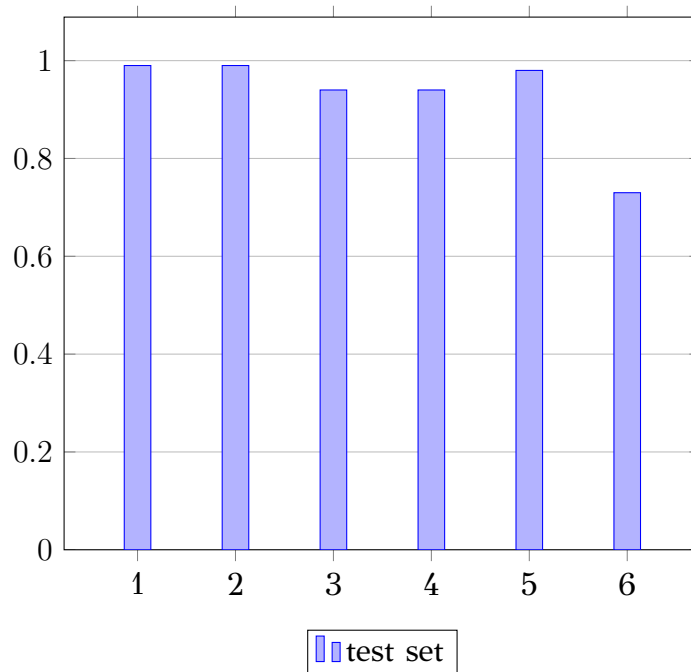
Σχήμα 4.72: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 2 εισόδους (DT)



Τα σχήματα 4.57, 4.60, 4.63, 4.66, 4.69 και 4.72 απεικονίζουν διαγράμματα πραγματικών με προβλεπόμενων τιμών για κάθε μία από τις παραπάνω περιπτώσεις. Για κάθε σημείο, στον άξονα X αναπαριστάται η πραγματική τιμή του, ενώ στον

άξονα Y η προβλεπόμενη τιμή. Όσο πιο κοντά είναι κάποιο σημείο στη γραμμή (μοντέλο), τόσο πιο σωστά πραγματοποιήθηκε η πρόβλεψη.

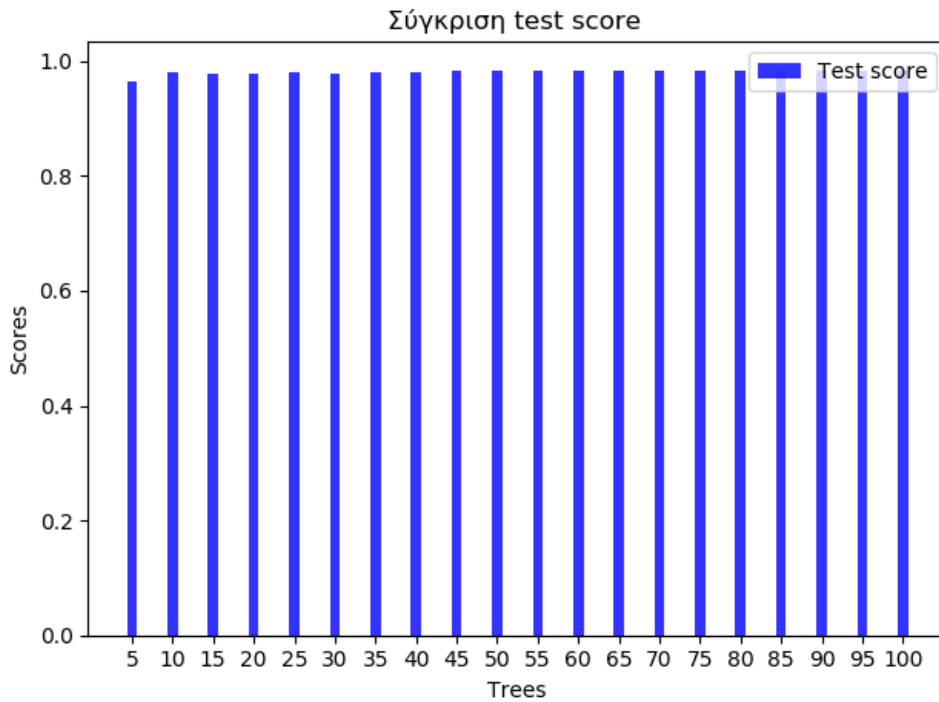
Στο παρακάτω σχήμα συνοφίζονται συγκεντρωτικά τα scores των train και test σετ για όλες τις παραπάνω περιπτώσεις. Παρατηρούμε ότι η πρόβλεψη κατανάλωσης ενέργειας στην κάρτα γραφικών είχε μεγαλύτερη επιτυχία χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους. Η πρόβλεψη κατανάλωσης ενέργειας του επεξεργαστή έχει το ίδιο αποτέλεσμα χρησιμοποιώντας τρία και δύο χαρακτηριστικά ως εισόδους. Η πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών είχε το ίδιο αποτέλεσμα τόσο για δύο όσο και για τρία χαρακτηριστικά ως εισόδους.



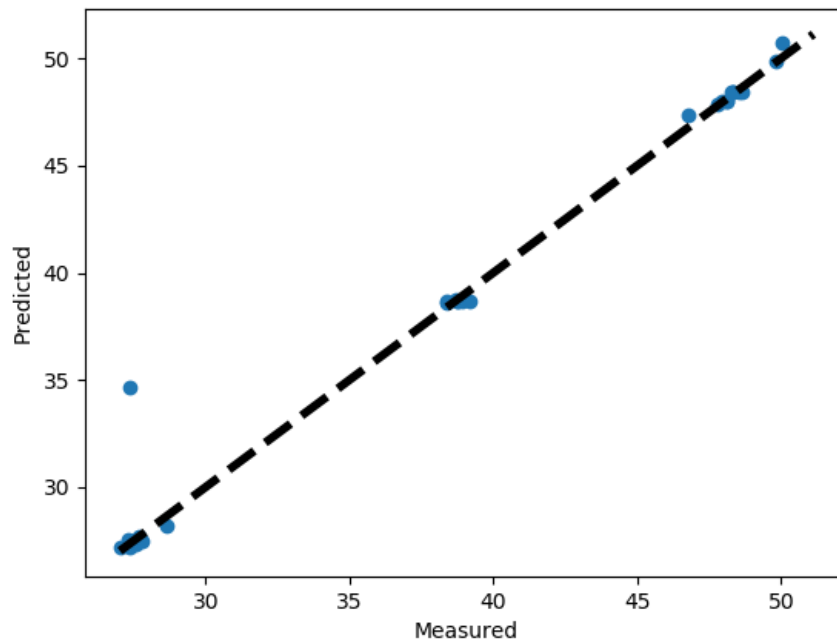
Τυχαία Δάση

Ο αλγόριθμος περιέχει μια επανάληψη (από 5 έως 100 με βήμα 5), όπου η μεταβλητή $n_estimators$ (αριθμός δέντρων) λαμβάνει την τιμή της επανάληψης, για καλύτερη διεξαγωγή συμπερασμάτων. Στο πρώτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στον επεξεργαστή. Χρησιμοποιώντας τρία χαρακτηριστικά ως εισόδους, όπως γίνεται αντιληπτό στο σχήμα 4.73, καθώς αυξάνεται η μεταβλητή $n_estimators$, υπάρχει μικρή αυξομείωση στο test score, το οποίο παίρνει τη μέγιστη τιμή του για αριθμό δέντρων 100. Για τη βέλτιστη τιμή της μεταβλητής $n_estimators$, το test score είναι 0.98.

Σχήμα 4.73: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (RF)



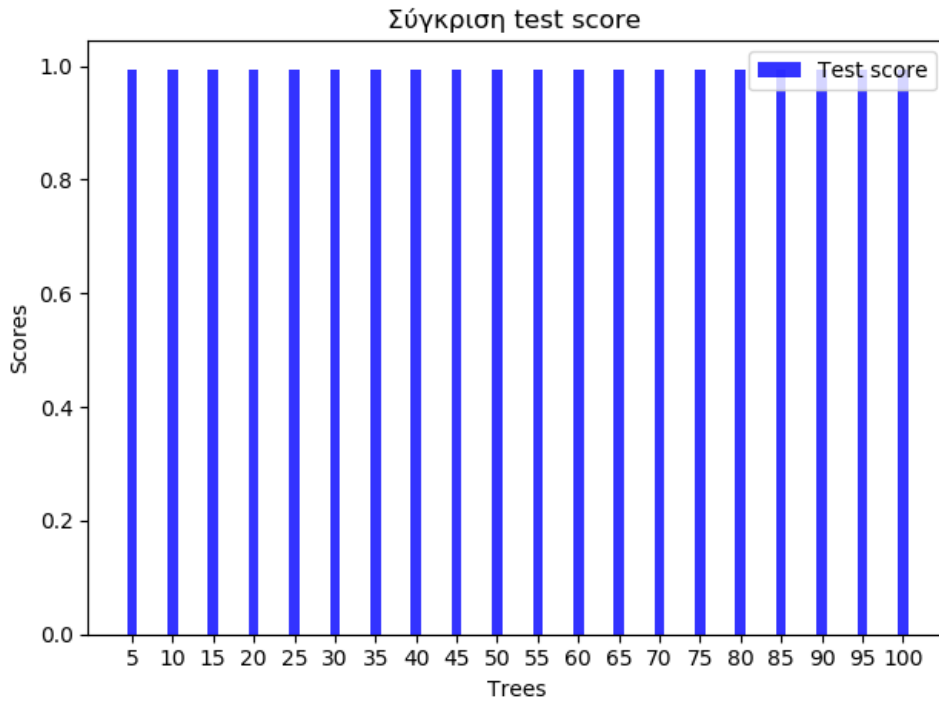
Σχήμα 4.74: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 3 εισόδους (RF)



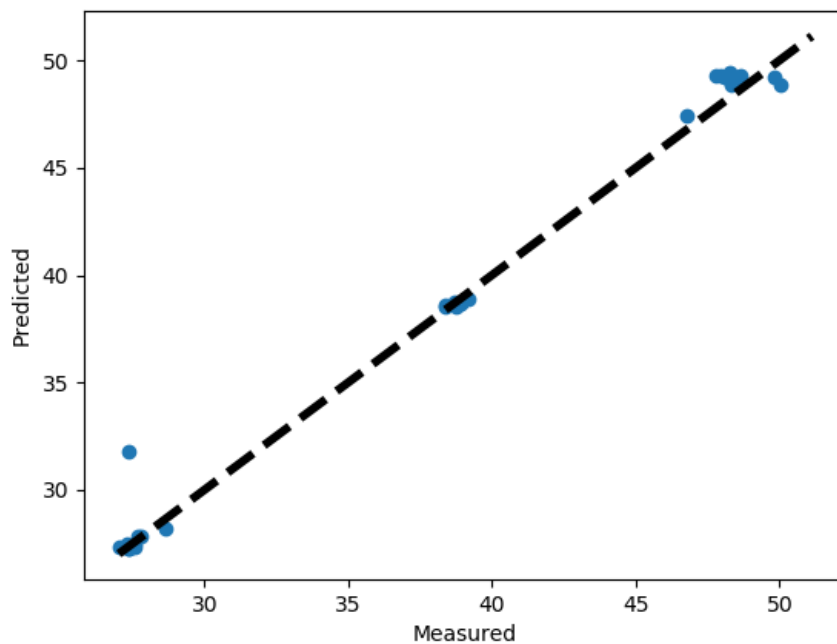
Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση του επεξεργαστή, όπως παρατηρείται στο σχήμα 4.75, καθώς αυξάνεται η μεταβλητή $n_estimators$ υπάρχει μικρή αυξομείωση στο test score, το οποίο παίρνει

τη μέγιστη τιμή του για αριθμό δέντρων 85. Για τη βέλτιστη τιμή της μεταβλητής $n_estimators$, το test score είναι 0.99.

Σχήμα 4.75: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (RF)



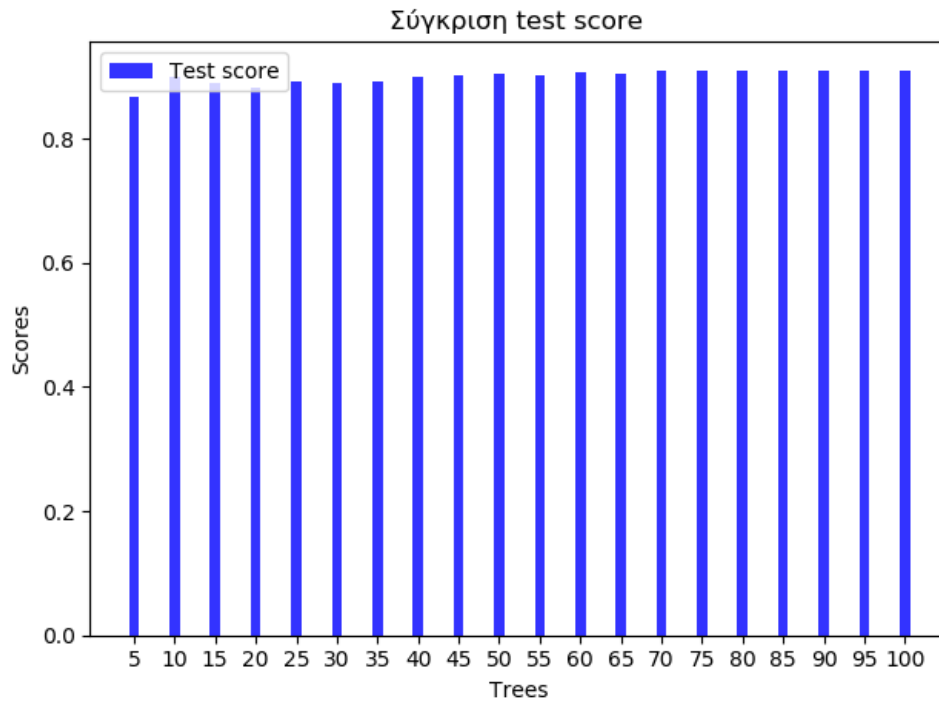
Σχήμα 4.76: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (RF)



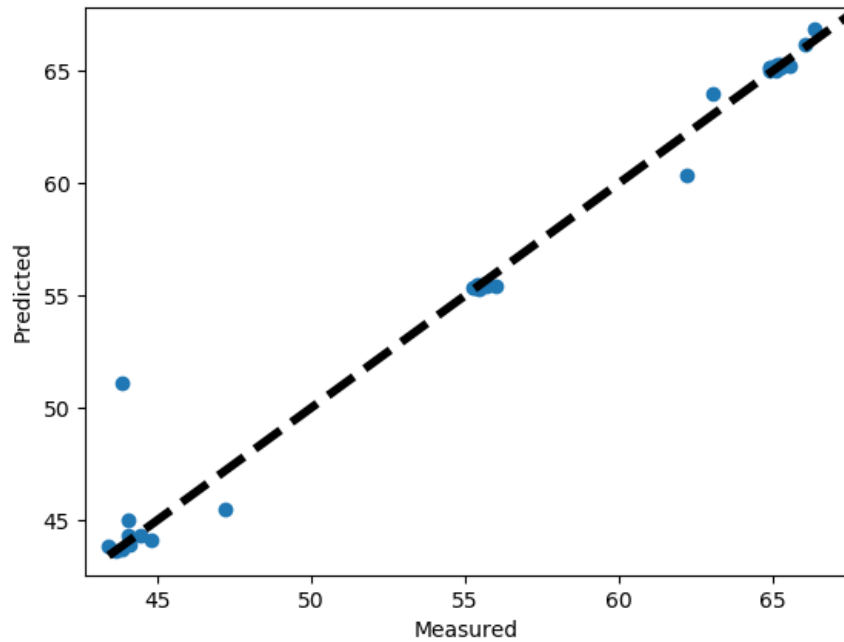
Στο δεύτερο πείραμα κάνουμε πρόβλεψη του αθροίσματος κατανάλωσης ενέρ-

γείας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 3 χαρακτηριστικά ως εισόδους. Όπως γίνεται αντιληπτό στο σχήμα 4.77 καθώς αυξάνεται η μεταβλητή $n_estimators$, υπάρχει αυξομείωση στο test score, το οποίο παίρνει τη μέγιστη τιμή του για αριθμό δέντρων 80. Για τη βέλτιστη τιμή της μεταβλητής $n_estimators$, το test score είναι 0.91.

Σχήμα 4.77: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (RF)

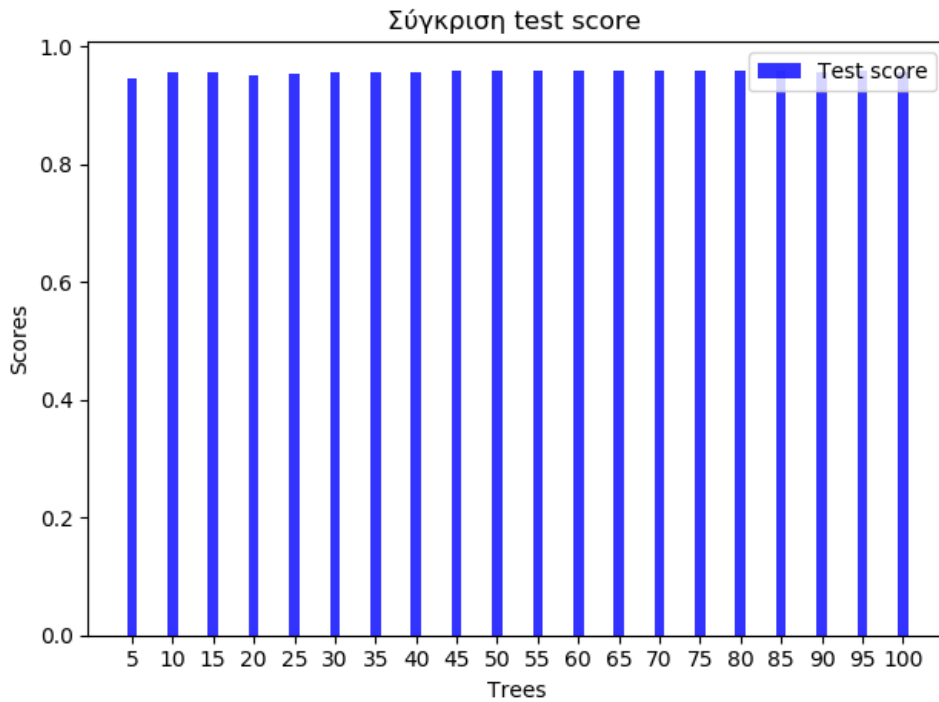


Σχήμα 4.78: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (RF)

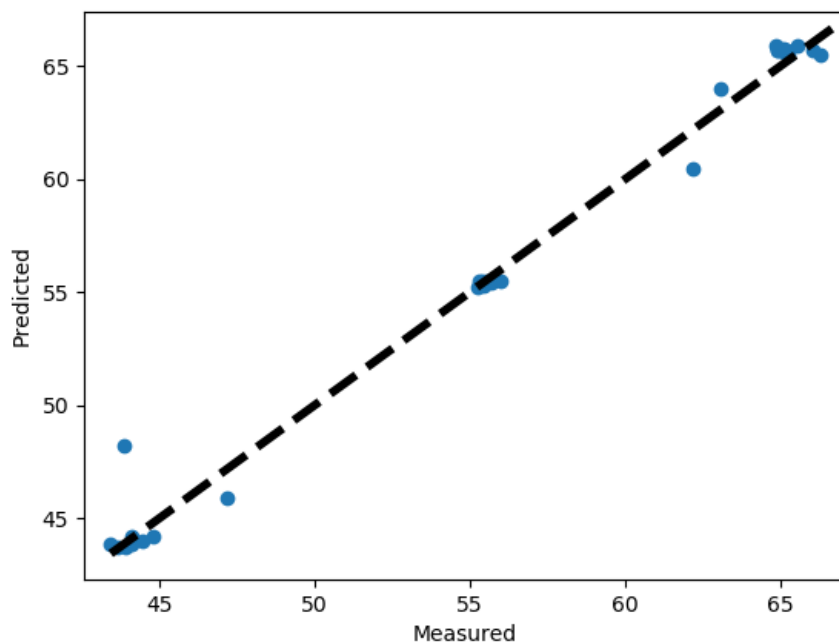


Για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 2 χαρακτηριστικά ως εισόδους, όπως παρατηρείται στο σχήμα 4.79, καθώς αυξάνεται η μεταβλητή $n_estimators$ υπάρχει μικρή αζομείωση στο test score, το οποίο παίρνει τη μέγιστη τιμή του για αριθμό δέντρων 65. Για τη βέλτιστη τιμή της μεταβλητής $n_estimators$, το test score είναι 0.95.

Σχήμα 4.79: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (RF)



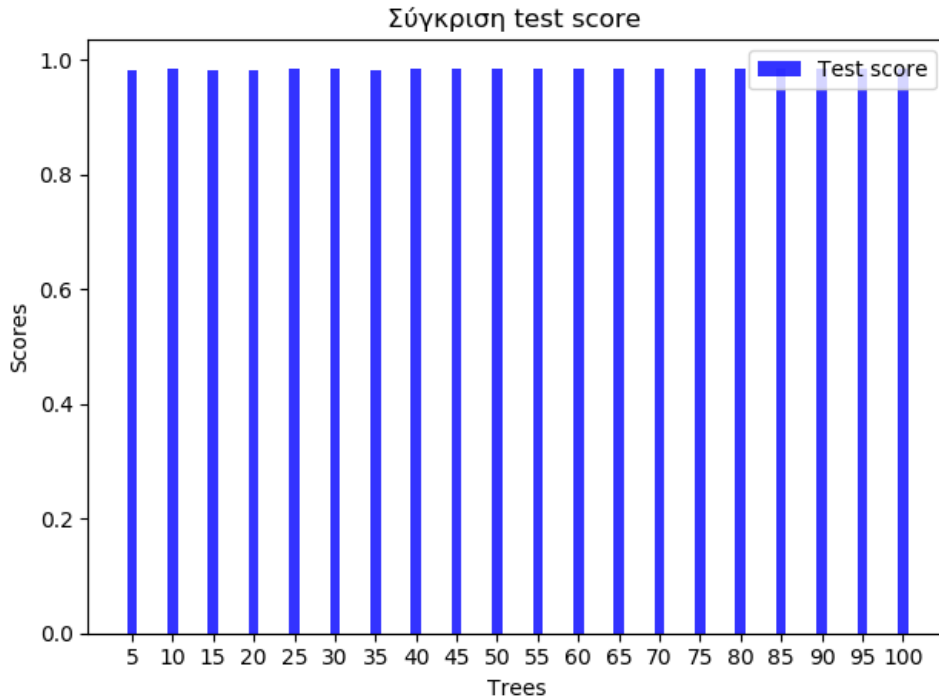
Σχήμα 4.80: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (RF)



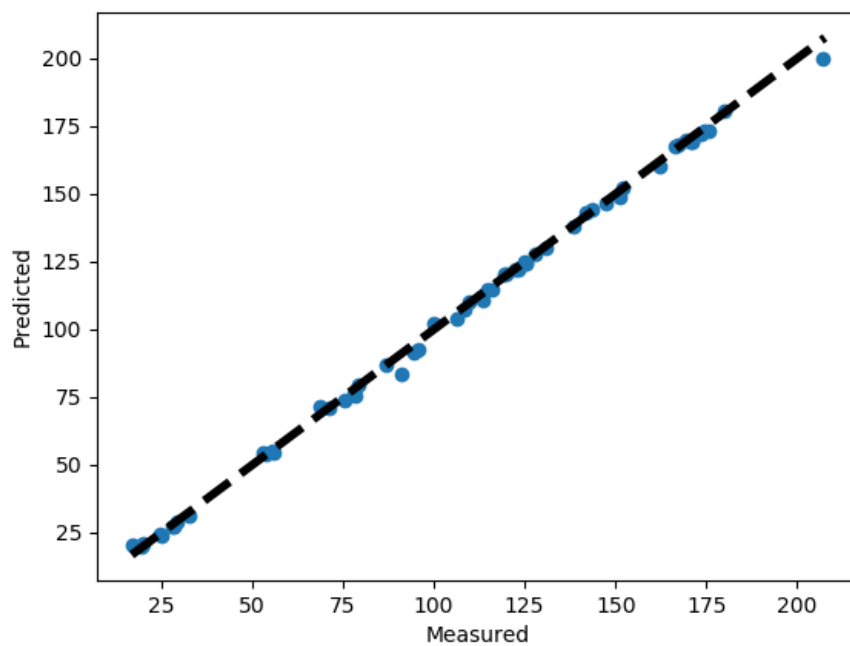
Στο τρίτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στην κάρτα γραφικών. Χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους, όπως γίνεται αντιληπτό στο σχήμα 4.81, καθώς αυξάνεται η μεταβλητή $n_estimators$ υπάρχει

μικρή αυξομείωση στο test score, το οποίο παίρνει τη μέγιστη τιμή του για αριθμό δέντρων 45. Για τη βέλτιστη τιμή της μεταβλητής $n_estimators$, το test score είναι 0.98.

Σχήμα 4.81: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (RF)

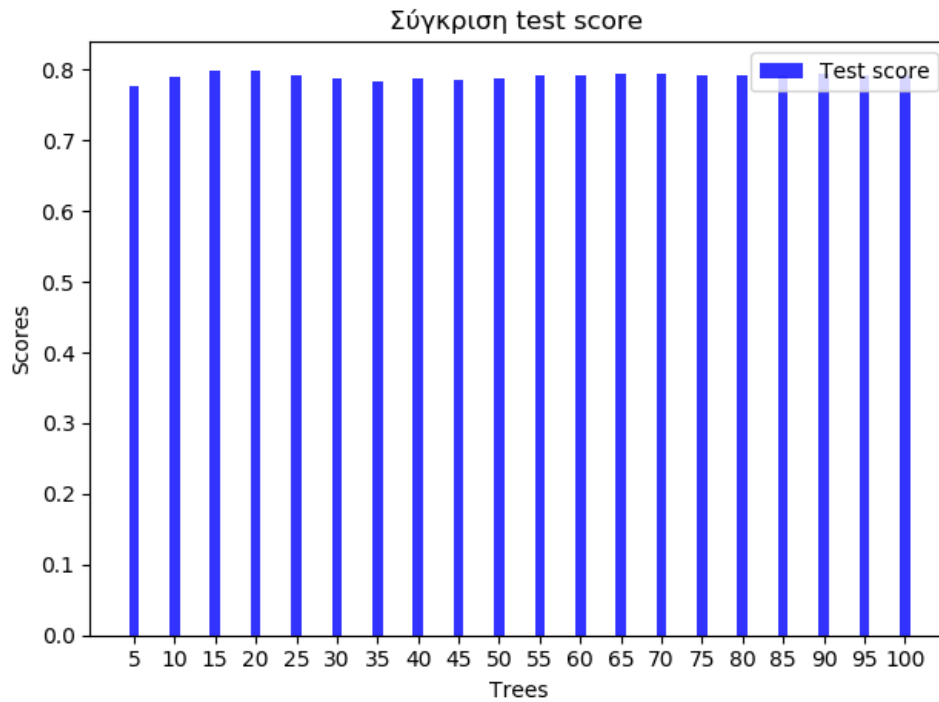


Σχήμα 4.82: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 4 εισόδους (RF)

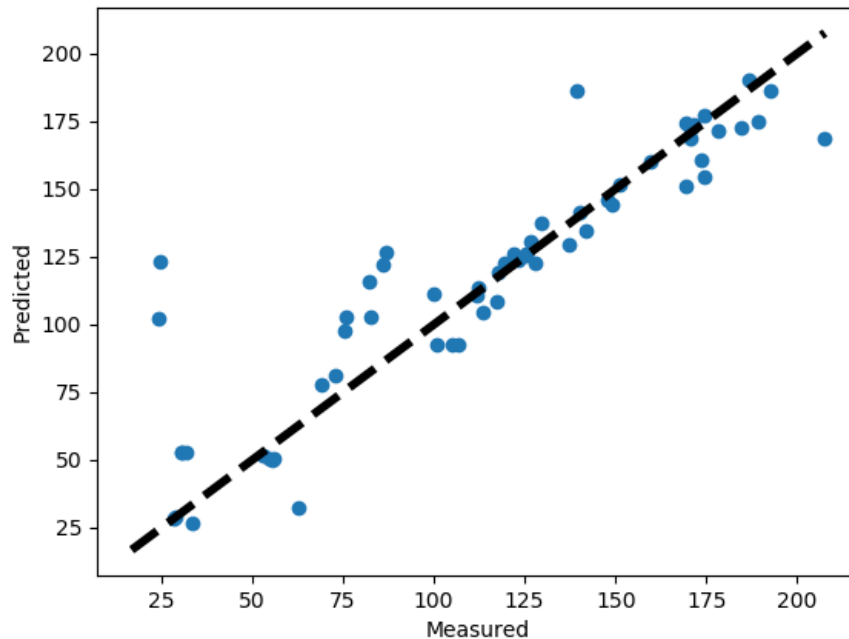


Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση της κάρτας γραφικών, παρατηρώντας το σχήμα 4.83, καθώς αυξάνεται η μεταβλητή $n_estimators$ υπάρχει μικρή αυξομείωση στο test score, το οποίο παίρνει τη μέγιστη τιμή του για αριθμό δέντρων 15. Για τη βέλτιστη τιμή της μεταβλητής $n_estimators$, το test score είναι 0.79.

Σχήμα 4.83: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (RF)

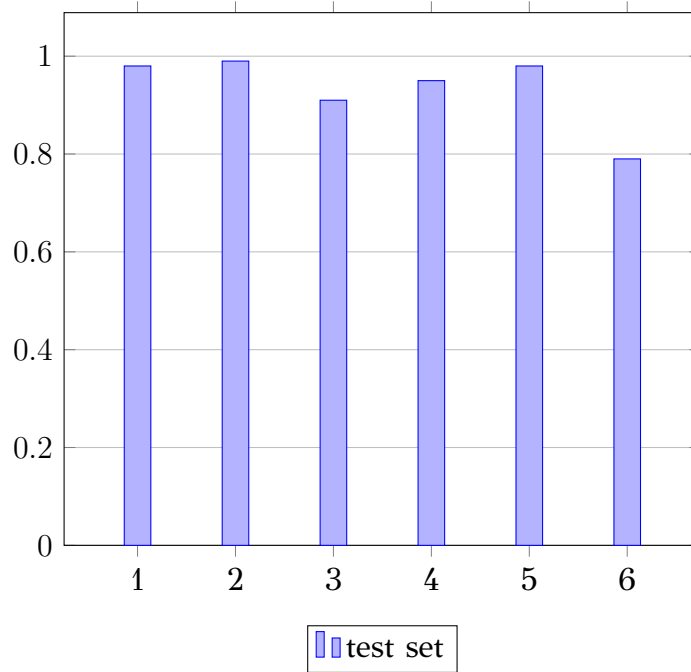


Σχήμα 4.84: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 2 εισόδους (RF)



Τα σχήματα 4.74, 4.76, 4.78, 4.80, 4.82 και 4.84 απεικονίζουν διαγράμματα πραγματικών με προβλεπόμενων τιμών για κάθε μία από τις παραπάνω περιπτώσεις. Για κάθε σημείο, στον άξονα X αναπαριστάται η πραγματική τιμή του, ενώ στον άξονα Y η προβλεπόμενη τιμή. Όσο πιο κοντά είναι κάποιο σημείο στη γραμμή (μοντέλο), τόσο πιο σωστά πραγματοποιήθηκε η πρόβλεψη.

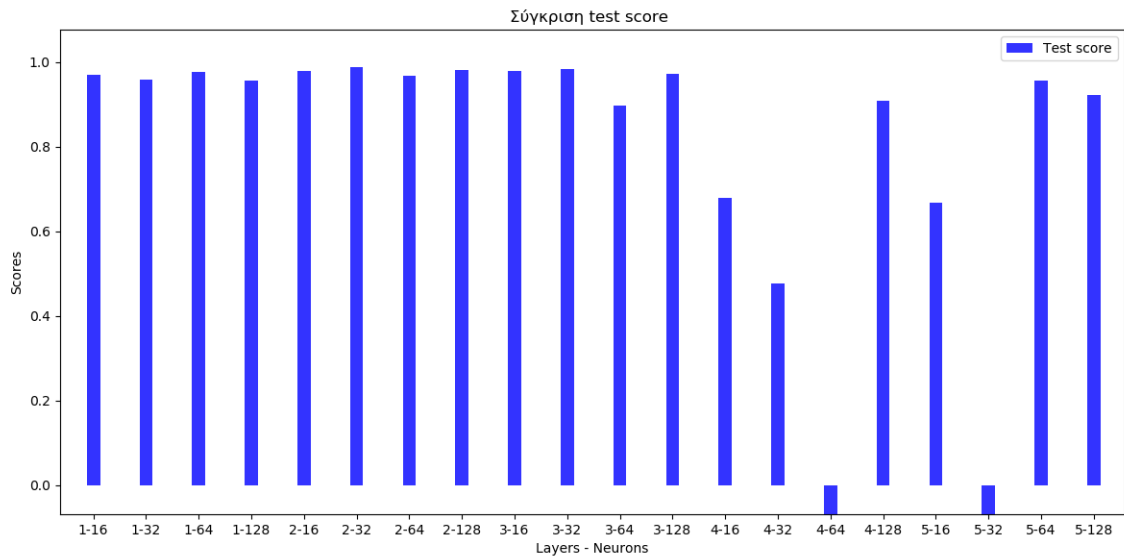
Στο παρακάτω σχήμα συνοφίζονται συγκεντρωτικά τα scores των train και test σετ για όλες τις παραπάνω περιπτώσεις. Παρατηρούμε ότι η πρόβλεψη κατανάλωσης ενέργειας στην κάρτα γραφικών είχε μεγαλύτερη επιτυχία χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους. Η πρόβλεψη κατανάλωσης ενέργειας του επεξεργαστή έχει ελάχιστα καλύτερο αποτέλεσμα χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους. Η πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών έχει μεγαλύτερη επιτυχία για δύο χαρακτηριστικά ως εισόδους.



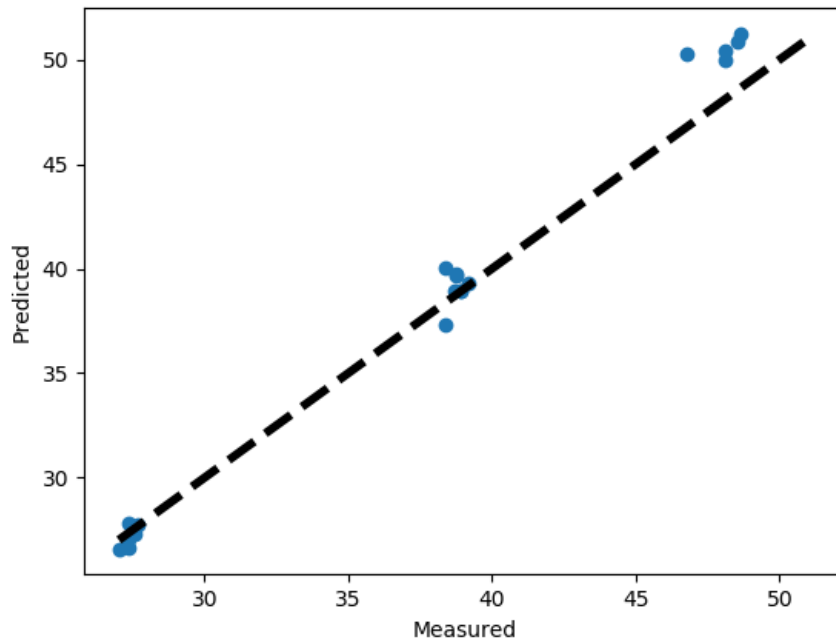
Βαθιά Νευρωνικά Δίκτυα

Ο αλγόριθμος χρησιμοποιεί δύο επαναλήψεις, από 1 μέχρι 5 με βήμα 1 για τον ορισμό των κρυφών επιπέδων και από 4 μέχρι 8 με βήμα 1 για τον ορισμό των νευρώνων (οι νευρώνες ορίζονται ως 2^j), στις οποίες εναλλάσσονται τα κρυφά επίπεδα (1-5) καθώς και οι νευρώνες που περιέχουν (16, 32, 64 και 128). Στο πρώτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στον επεξεργαστή. Χρησιμοποιώντας τρία χαρακτηριστικά ως εισόδους, όπως παρατηρείται και στο σχήμα 4.85, το test score αυξομειώνεται. Για 4 κρυφά επίπεδα με 64 νευρώνες και 5 κρυφά επίπεδα με 32 νευρώνες παρατηρείται αρνητικό test score, ενώ το test score λαμβάνει τη μέγιστη τιμή του (0.98) για 2 κρυφά επίπεδα με 32 νευρώνες.

Σχήμα 4.85: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (NN)

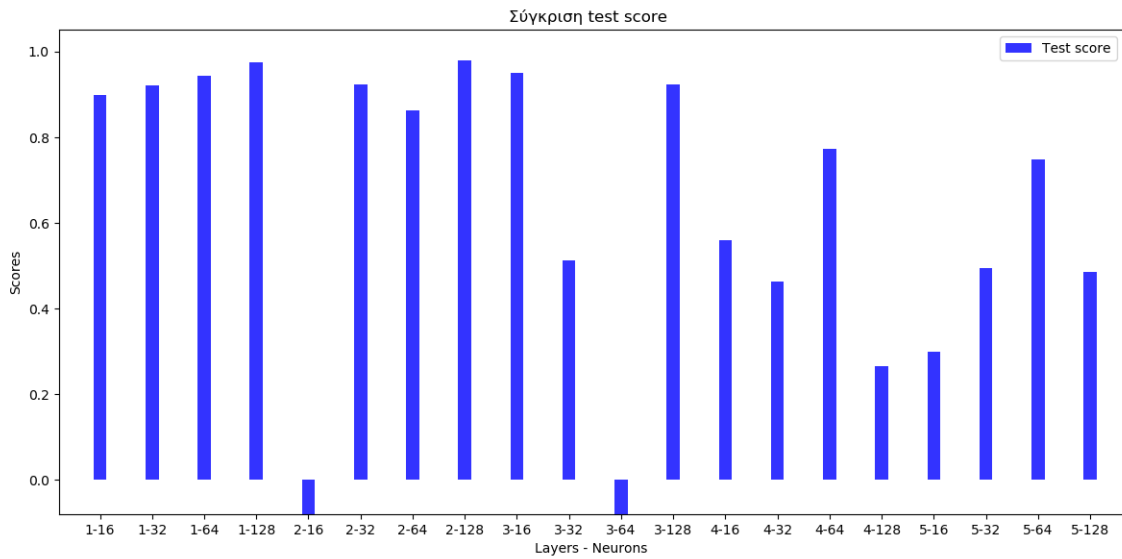


Σχήμα 4.86: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 3 εισόδους (NN)

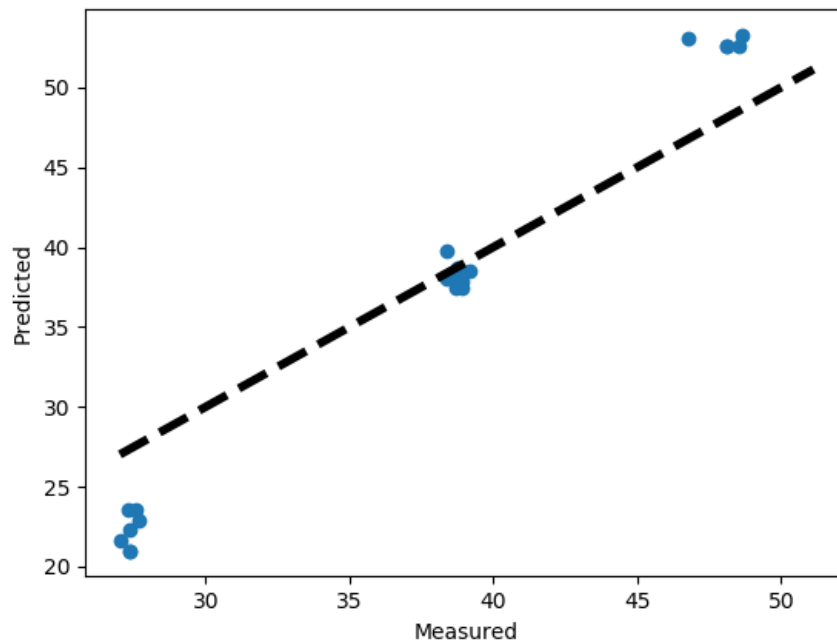


Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση του επεξεργαστή, όπως παρατηρείται και στο σχήμα 4.87, το test score αυξομειώνεται. Για 2 κρυφά επίπεδα με 16 νευρώνες καθώς και για 3 κρυφά επίπεδα με 64 νευρώνες παρατηρείται αρνητικό test score. Το test score λαμβάνει τη μέγιστη τιμή του (0.97) για 2 κρυφά επίπεδα με 128 νευρώνες.

Σχήμα 4.87: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (NN)

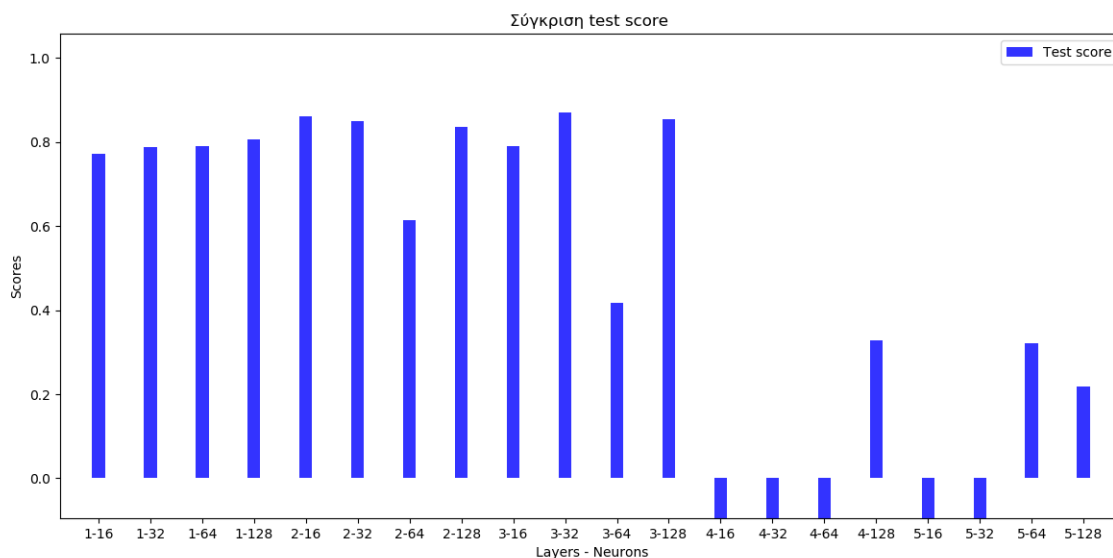


Σχήμα 4.88: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (NN)

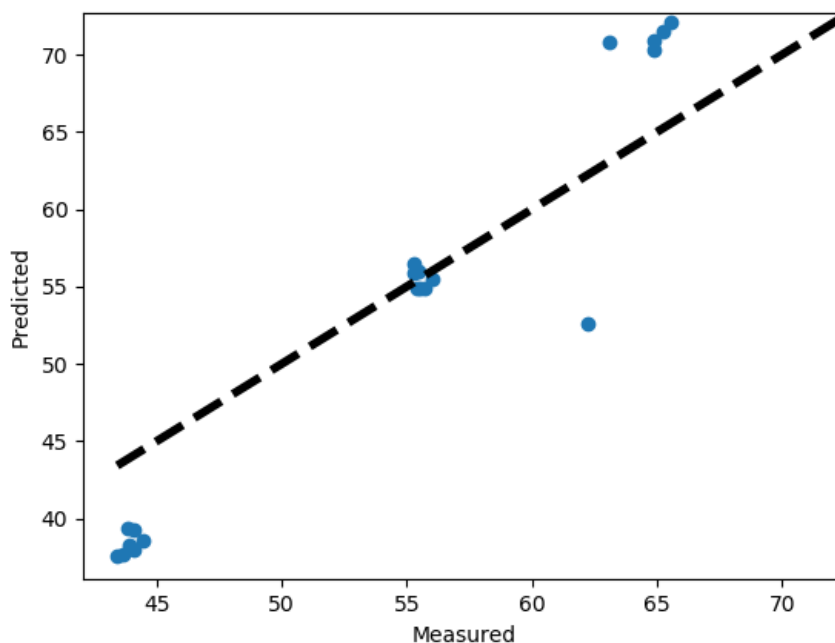


Στο δεύτερο πείραμα κάνουμε πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 3 χαρακτηριστικά ως εισόδους, όπως παρατηρείται και στο σχήμα 4.89, το test score αυξομειώνεται. Για 4 κρυφά επίπεδα με 16,32 και 64 νευρώνες καθώς και για 5 κρυφά επίπεδα με 16 και 32 νευρώνες παρατηρείται αρνητικό test score. Το test score λαμβάνει τη μέγιστη τιμή του (0.87) για 3 κρυφά επίπεδα με 32 νευρώνες.

Σχήμα 4.89: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (NN)



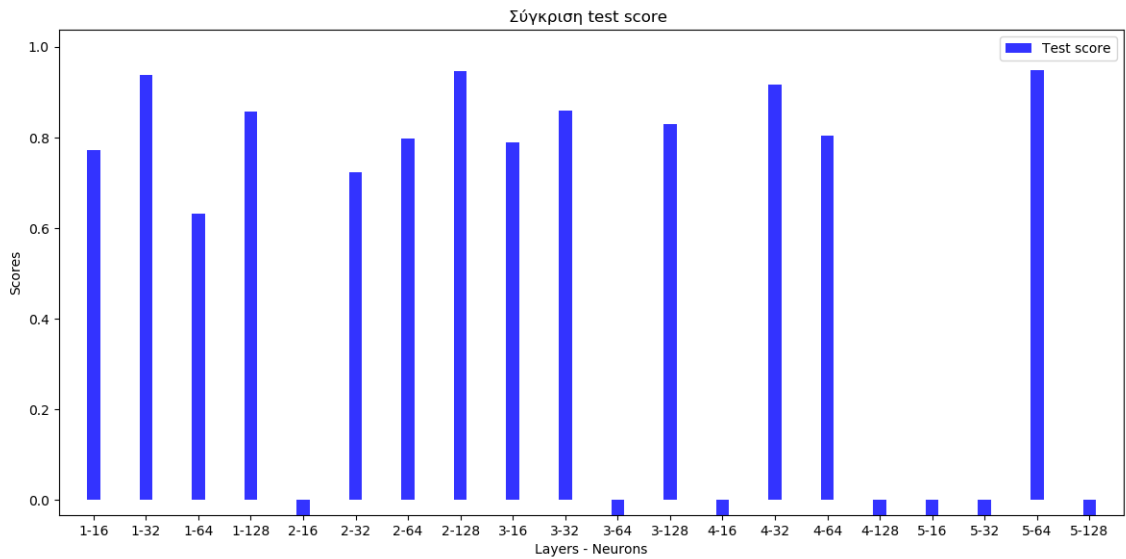
Σχήμα 4.90: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (NN)



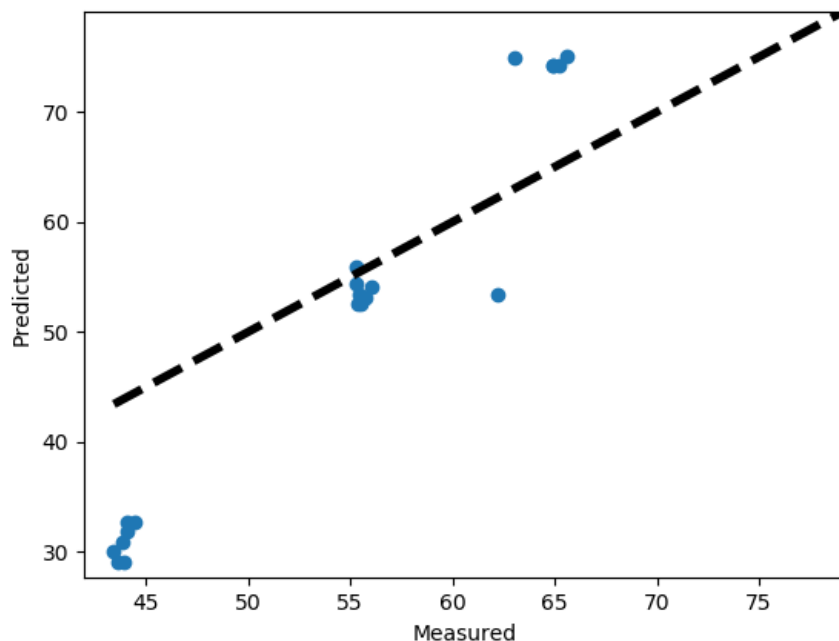
Για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 2 χαρακτηριστικά ως εισόδους, όπως γίνεται αντιληπτό και στο σχήμα 4.91, το test score αυξομειώνεται. Για 2 κρυφά επίπεδα με 16 νευρώνες, 3 κρυφά επίπεδα με 64 νευρώνες, 4 κρυφά επίπεδα με 16 και 128 νευρώνες και 5 κρυφά επίπεδα με 16, 32 και 128 νευρώνες παρατηρείται αρνητικό test score. Το test score λαμβάνει τη μέγιστη τιμή του (0.94) για 5 κρυφά

επίπεδα με 64 νευρώνες.

Σχήμα 4.91: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (NN)



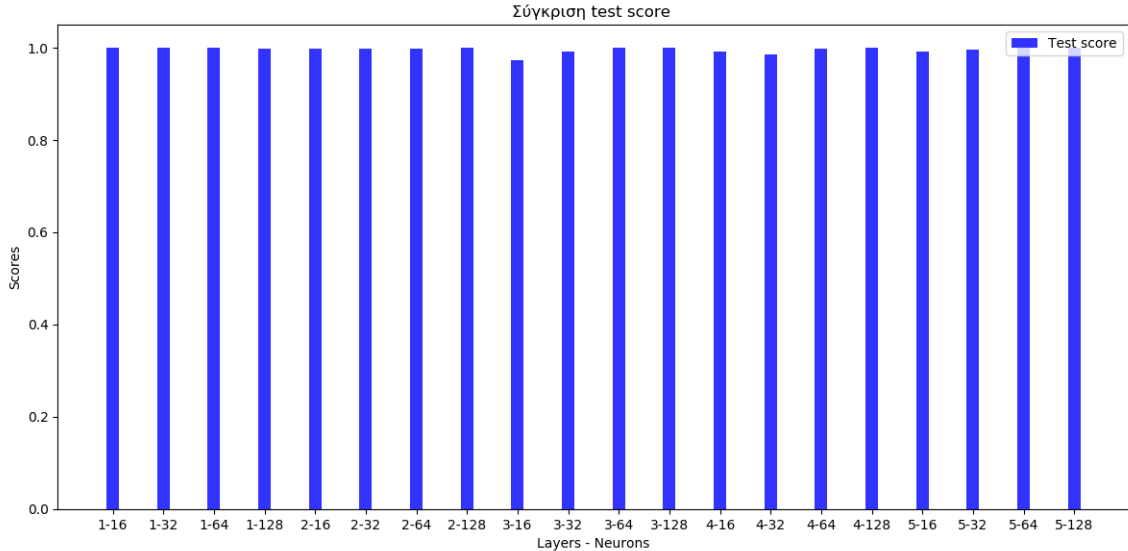
Σχήμα 4.92: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (NN)



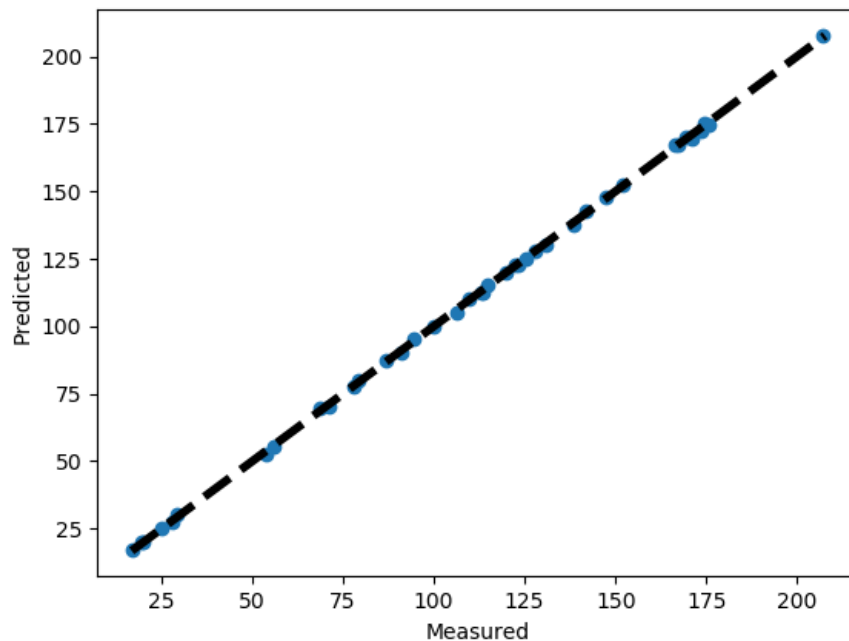
Στο τρίτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στην κάρτα γραφικών. Χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους, όπως παρατηρείται και στο σχήμα 4.93, υπάρχει μικρή αυξομείωση στο test score. Για 2 κρυφά επίπεδα με 16 νευρώνες τα train και test scores είναι αρνητικά, όπως και για 3

κρυφά επίπεδα με 64 νευρώνες το test score είναι αρνητικό. Το test score λαμβάνει τη μέγιστη τιμή του (0.97) για 1 κρυφό επίπεδο με 64 νευρώνες.

Σχήμα 4.93: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (NN)

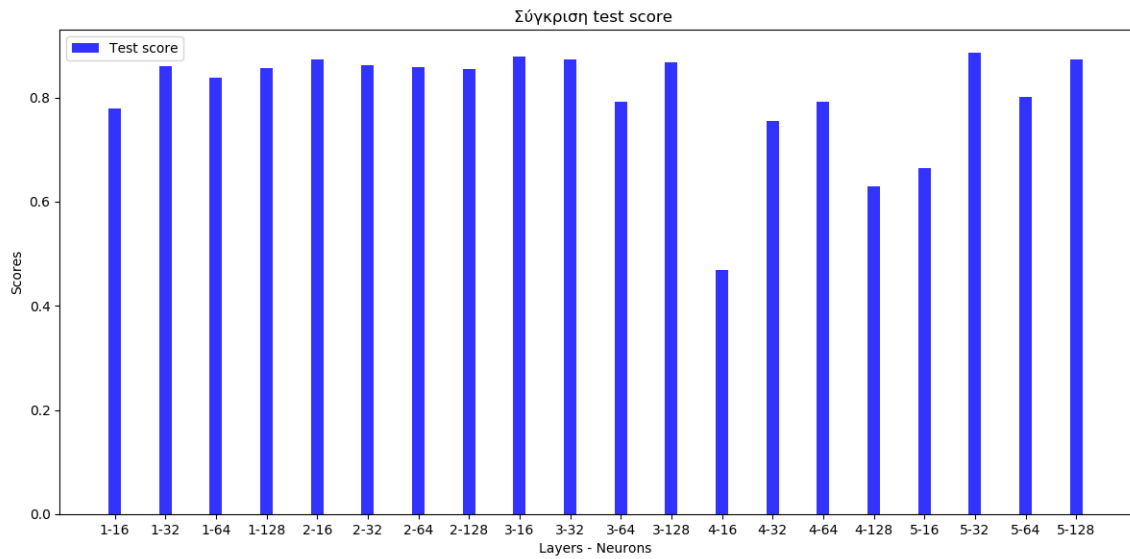


Σχήμα 4.94: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 4 εισόδους (NN)

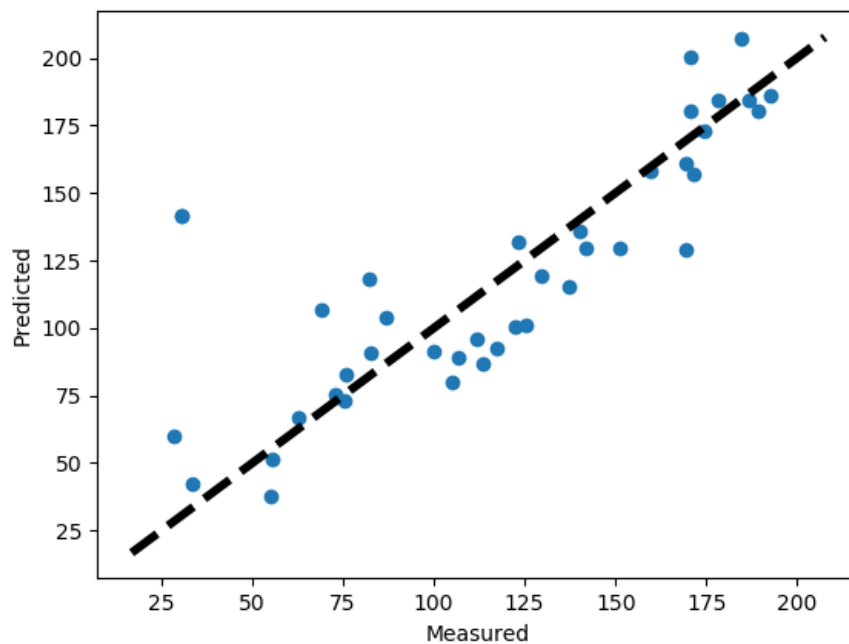


Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση της κάρτας γραφικών, όπως παρατηρείται στο σχήμα 4.95, υπάρχει αυξομείωση στο test score. Το test score λαμβάνει τη μέγιστη τιμή του (0.88) για 5 κρυφά επίπεδα με 32 νευρώνες.

Σχήμα 4.95: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (NN)



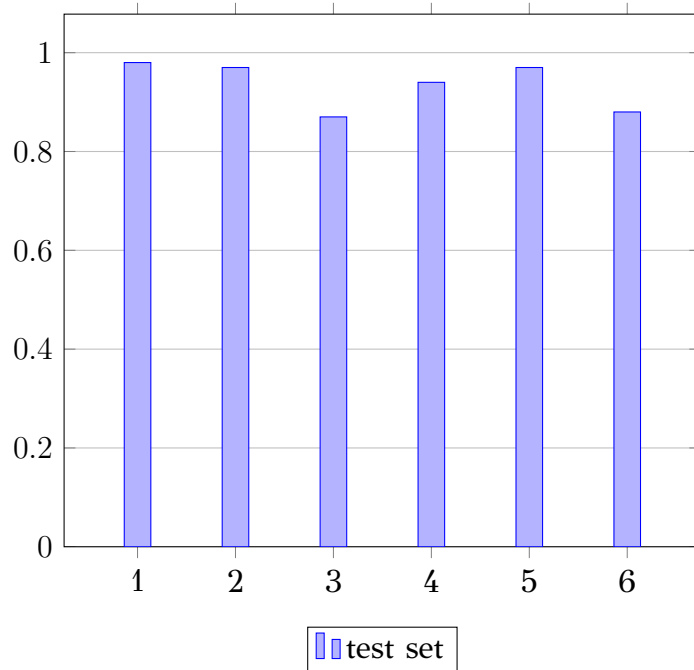
Σχήμα 4.96: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 2 εισόδους (NN)



Τα σχήματα 4.86, 4.88, 4.90, 4.92, 4.94 και 4.96 απεικονίζουν διαγράμματα πραγματικών με προβλεπόμενων τιμών για κάθε μία από τις παραπάνω περιπτώσεις. Για κάθε σημείο, στον άξονα X αναπαριστάται η πραγματική τιμή του, ενώ στον άξονα Y η προβλεπόμενη τιμή. Όσο πιο κοντά είναι κάποιο σημείο στη γραμμή (μοντέλο), τόσο πιο σωστά πραγματοποιήθηκε η πρόβλεψη.

Στο παρακάτω σχήμα συνοψίζονται συγκεντρωτικά τα scores των train και test

σεν για όλες τις παραπάνω περιπτώσεις. Παρατηρούμε ότι η πρόβλεψη κατανάλωσης ενέργειας στην κάρτα γραφικών είχε μεγαλύτερη επιτυχία χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους. Η πρόβλεψη κατανάλωσης ενέργειας του επεξεργαστή έχει ελάχιστα καλύτερο αποτέλεσμα χρησιμοποιώντας τρία χαρακτηριστικά ως εισόδους. Η πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών έχει μεγαλύτερη επιτυχία για δύο χαρακτηριστικά ως εισόδους.

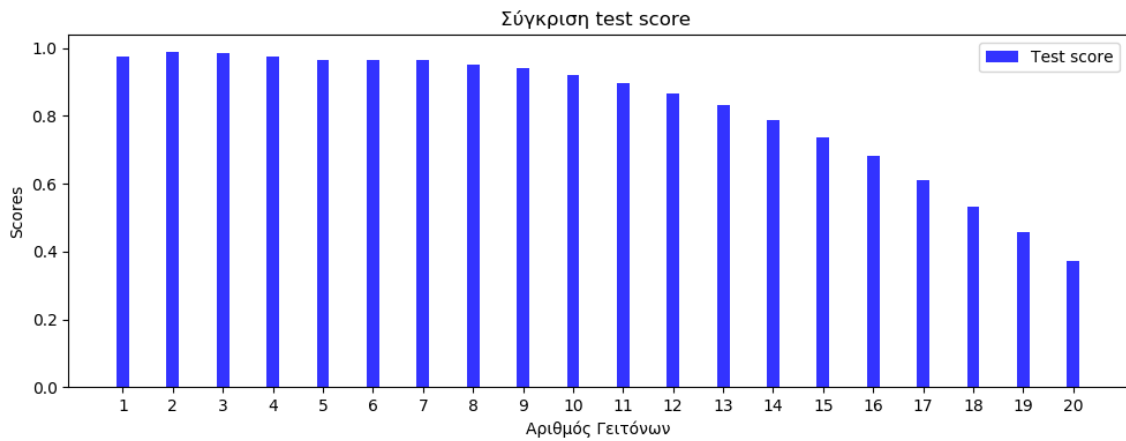


4.4 Πειράματα στον υπολογιστή υψηλών προδιαγραφών

Κ-Πλησιέστεροι Γείτονες

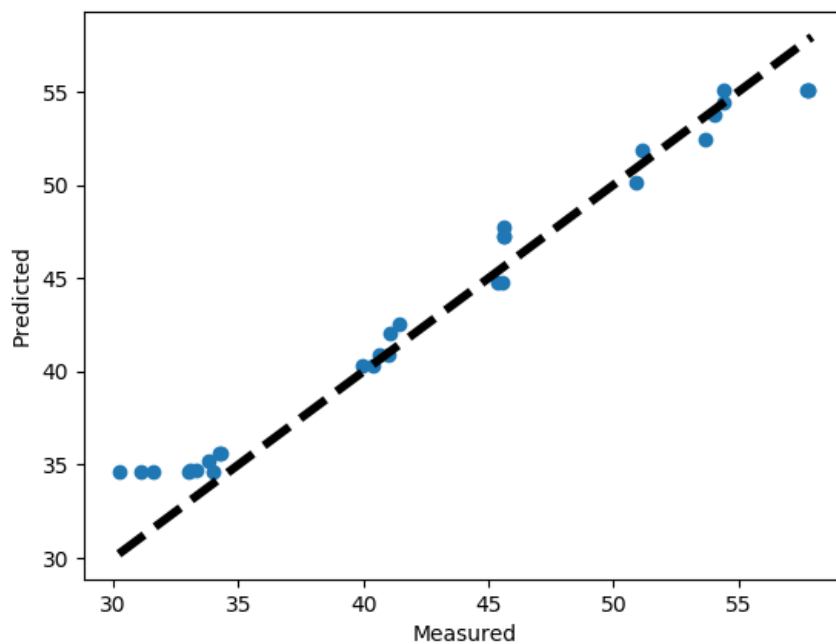
Ο αλγόριθμος περιέχει μια επανάληψη (από 1 μέχρι 20 με βήμα 1), όπου η μεταβλητή $n_neighbors$ (αριθμός γειτόνων) λαμβάνει την τιμή της επανάληψης, για καλύτερη διεξαγωγή συμπερασμάτων. Στο πρώτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στον επεξεργαστή. Χρησιμοποιώντας τρία χαρακτηριστικά ως εισόδους, η βέλτιστη τιμή της μεταβλητής $n_neighbors$ (αριθμός γειτόνων) είναι 2 (σύμφωνα με το σχήμα 4.97). Ορίζοντας τη μεταβλητή $n_neighbors$ στη τιμή 2, λαμβάνουμε το καλύτερο score στο test σετ, το οποίο είναι 0.98.

Σχήμα 4.97: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (KNN)



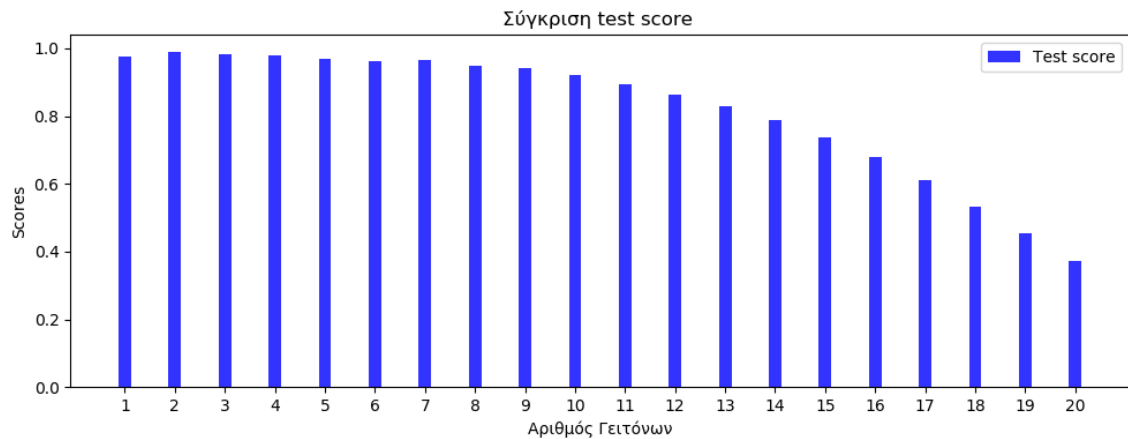
Όπως φαίνεται στο σχήμα 4.97, το test score αυξάνεται μέχρι η τιμή της μεταβλητής $n_neighbors$ να λάβει την τιμή 2. Έπειτα, καθώς η μεταβλητή $n_neighbors$ αυξάνεται, παρατηρείται αισθητή μείωση του test score.

Σχήμα 4.98: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 3 εισόδους (KNN)



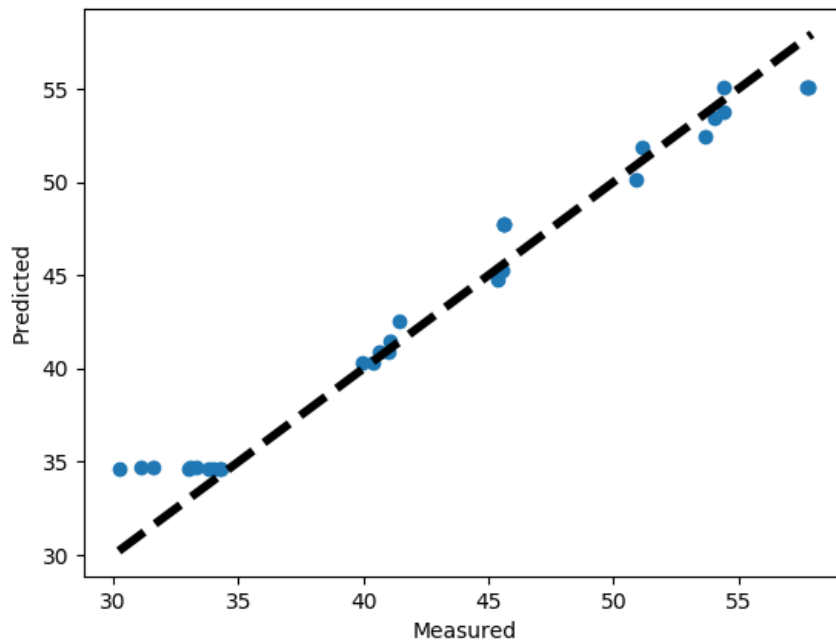
Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση του επεξεργαστή, η βέλτιστη τιμή της μεταβλητής $n_neighbors$ (αριθμός γειτόνων) είναι 2 (σύμφωνα με το σχήμα 4.99). Ορίζοντας τη μεταβλητή $n_neighbors$ στη τιμή 2, λαμβάνουμε το καλύτερο score στο test σετ, το οποίο είναι 0.98.

Σχήμα 4.99: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (KNN)



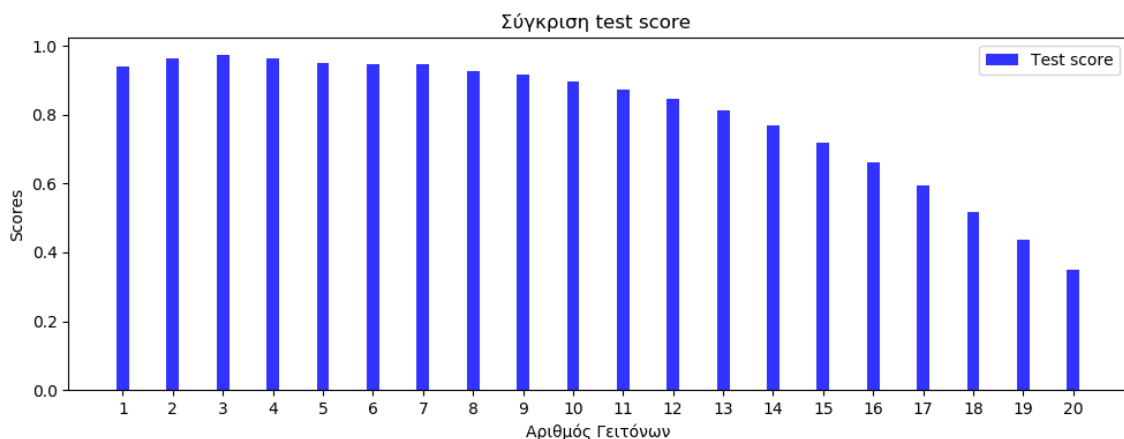
Όπως φαίνεται και στο σχήμα 4.99, το test score αυξάνεται μέχρι η τιμή της μεταβλητής $n_neighbors$ να λάβει την τιμή 2. Έπειτα, καθώς η μεταβλητή $n_neighbors$ αυξάνεται, παρατηρείται αισθητή μείωση του test score.

Σχήμα 4.100: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (KNN)



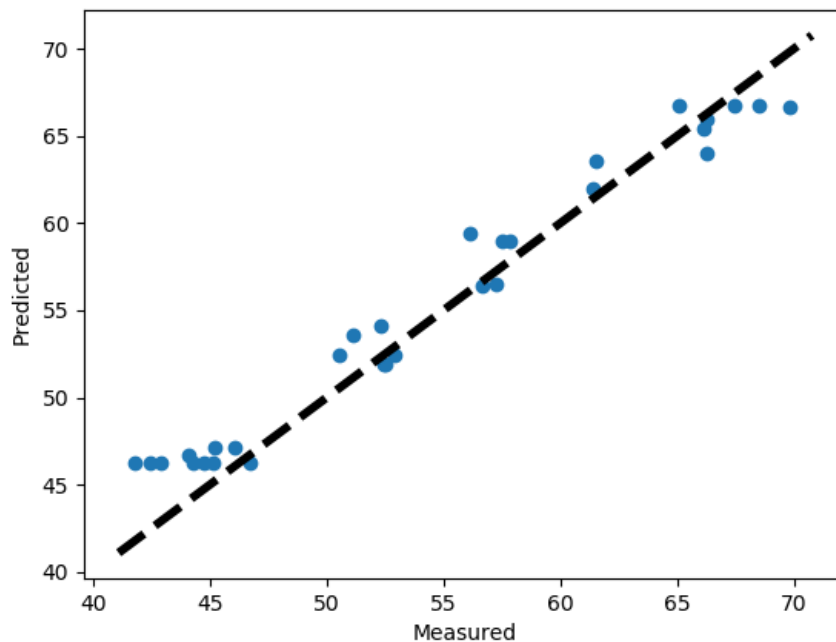
Στο δεύτερο πείραμα κάνουμε πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 3 χαρακτηριστικά ως εισόδους. Ορίζοντας τη μεταβλητή $n_neighbors$ (αριθμός γειτόνων) στη βέλτιστη τιμή 3, σύμφωνα με το σχήμα 4.101, λαμβάνουμε το καλύτερο score στο test σετ, το οποίο είναι 0.97.

Σχήμα 4.101: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (KNN)



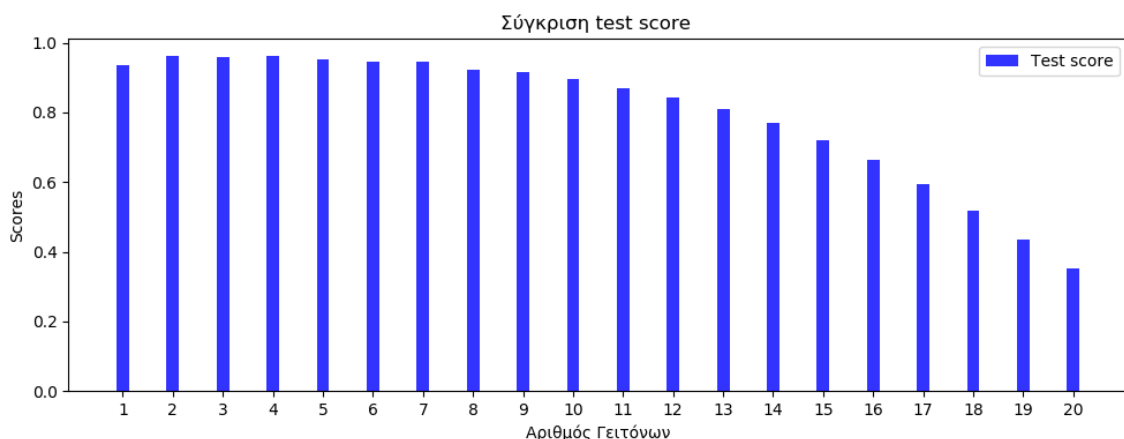
Όπως φαίνεται στο σχήμα 4.101, το test score αυξάνεται μέχρι η τιμή της μεταβλητής $n_neighbors$ να λάβει την τιμή 3. Έπειτα, καθώς η μεταβλητή $n_neighbors$ αυξάνεται, παρατηρείται αισθητή μείωση του test score.

Σχήμα 4.102: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (KNN)



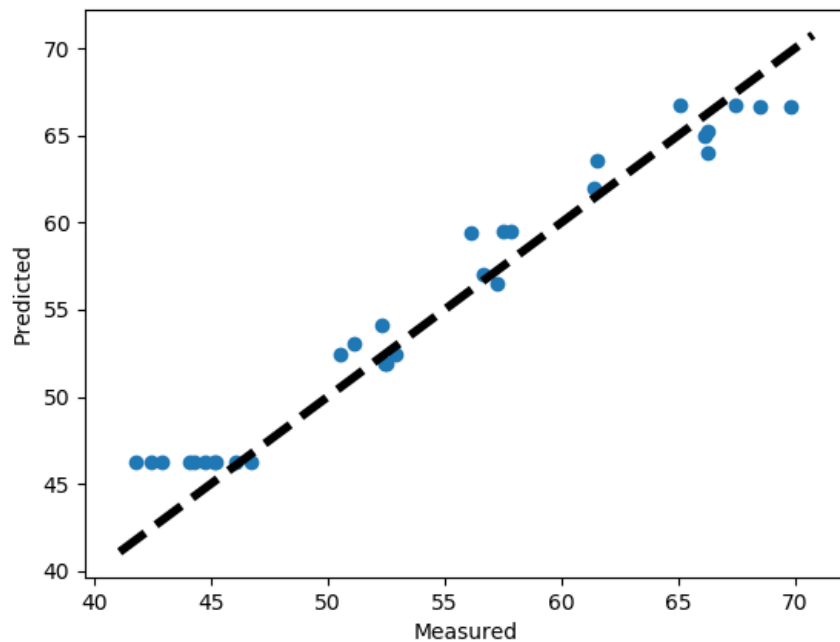
Για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 2 χαρακτηριστικά ως εισόδους, ορίζουμε τη μεταβλητή $n_neighbors$ (αριθμός γειτώνων) στη βέλτιστη τιμή 2, σύμφωνα με το σχήμα 4.103. Ορίζοντας τη μεταβλητή $n_neighbors$ στη τιμή 2, λαμβάνουμε το καλύτερο score στο test σετ, το οποίο είναι 0.96.

Σχήμα 4.103: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (KNN)



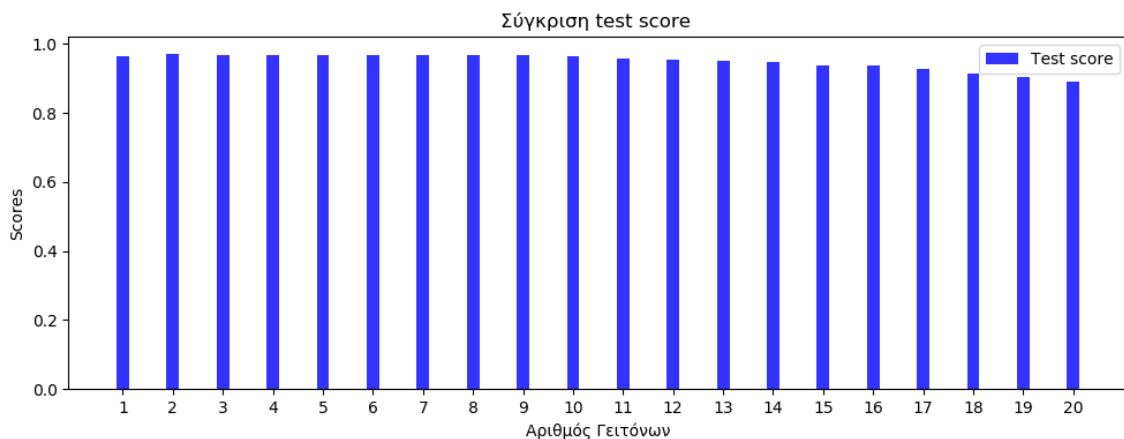
Όπως φαίνεται και στο σχήμα 4.103, το test score αυξάνεται μέχρι η τιμή της μεταβλητής $n_neighbors$ να λάβει την τιμή 2. Έπειτα, καθώς η μεταβλητή $n_neighbors$ αυξάνεται, παρατηρείται αισθητή μείωση του test score.

Σχήμα 4.104: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (KNN)



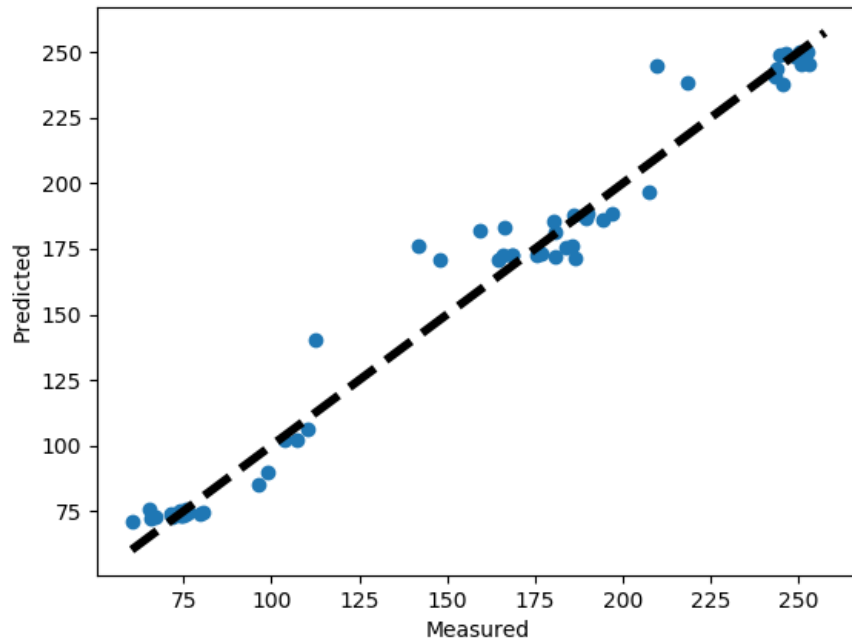
Στο τρίτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στην κάρτα γραφικών. Χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους, η βέλτιστη τιμή της μεταβλητής $n_neighbors$ (αριθμός γειτόνων) είναι 2 (σύμφωνα με το σχήμα 4.105) με test score 0.97.

Σχήμα 4.105: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (KNN)



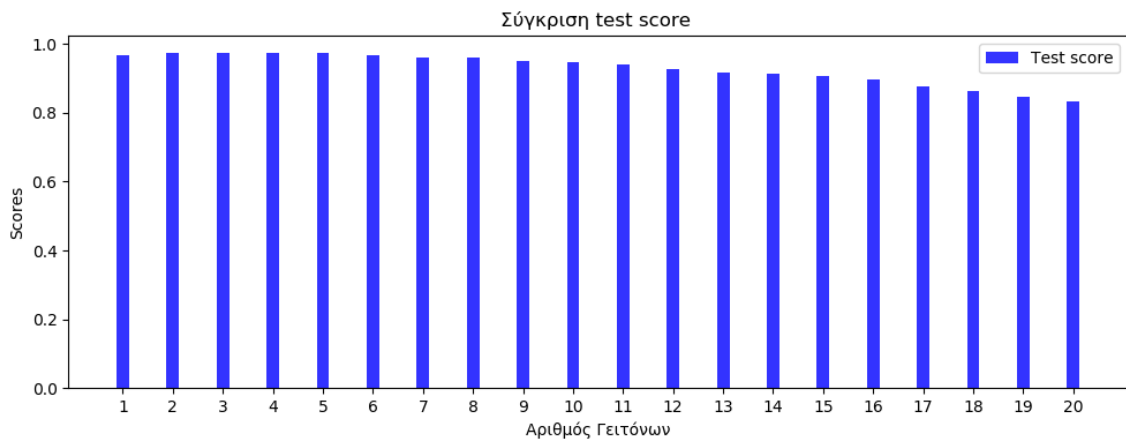
Όπως φαίνεται στο σχήμα 4.105, το test score παίρνει τη βέλτιστη τιμή του όταν η μεταβλητή $n_neighbors$ παίρνει την τιμή 2. Καθώς η μεταβλητή $n_neighbors$ αυξάνεται, το test score μειώνεται.

Σχήμα 4.106: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 4 εισόδους (KNN)



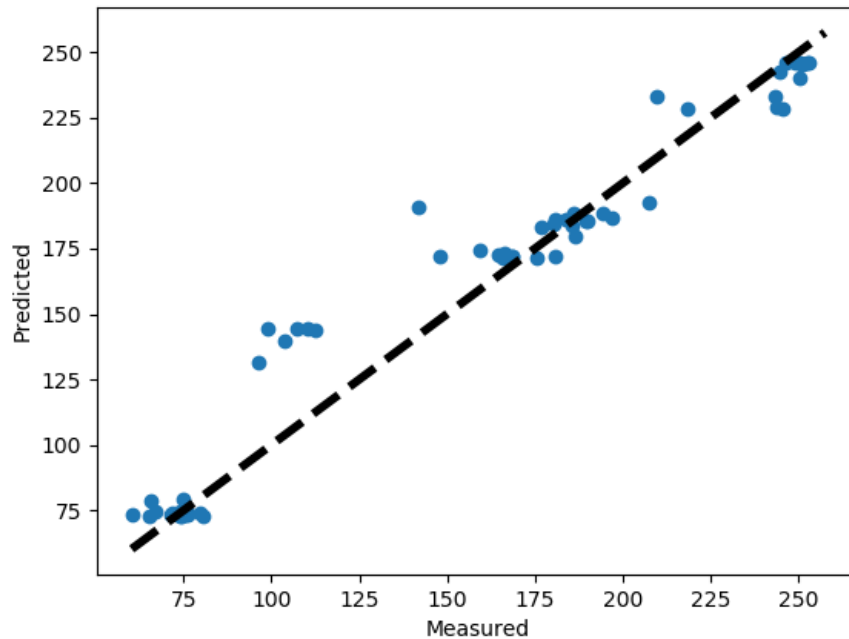
Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση της κάρτας γραφικών, η βέλτιστη τιμή της μεταβλητής $n_neighbors$ (αριθμός γειτόνων) είναι 4 (σύμφωνα με το σχήμα 4.107) με test score 0.97.

Σχήμα 4.107: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (KNN)

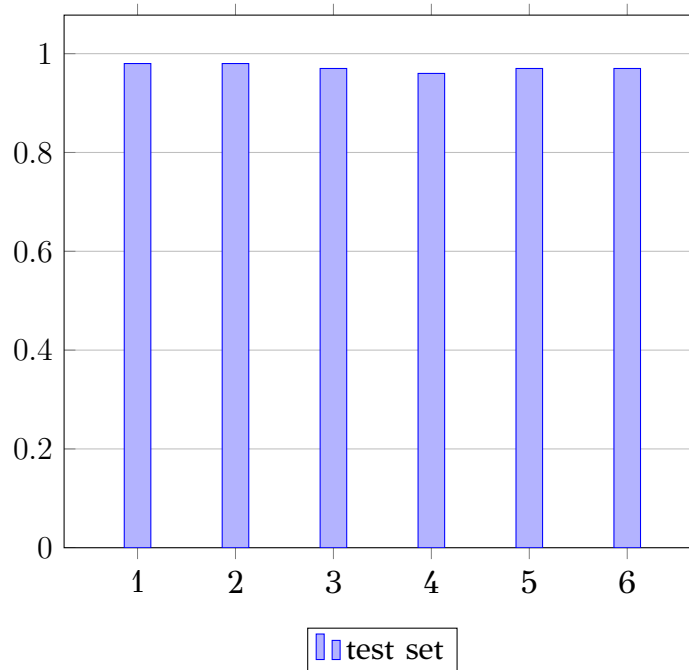


Όπως φαίνεται στο σχήμα 4.107, το test score παίρνει τη βέλτιστη τιμή του όταν η μεταβλητή $n_neighbors$ παίρνει την τιμή 4. Καθώς η μεταβλητή $n_neighbors$ αυξάνεται, το test score μειώνεται.

Σχήμα 4.108: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατά-
νάλωσης της κάρτας γραφικών με 2 εισόδους (KNN)



Τα σχήματα 4.98, 4.100, 4.102, 4.104, 4.106 και 4.108 απεικονίζουν διαγράμματα πραγματικών με προβλεπόμενων τιμών για κάθε μία από τις παραπάνω περιπτώσεις. Για κάθε σημείο, στον άξονα X αναπαριστάται η πραγματική τιμή του, ενώ στον άξονα Y η προβλεπόμενη τιμή. Όσο πιο κοντά είναι κάποιο σημείο στη γραμμή (μοντέλο), τόσο πιο σωστά πραγματοποιήθηκε η πρόβλεψη.



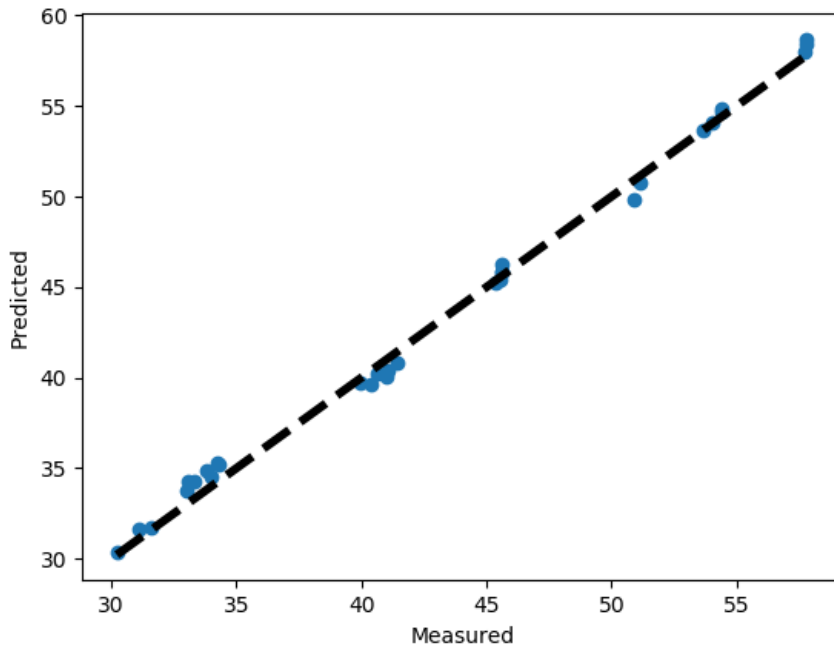
Στο παραπάνω σχήμα συνοψίζονται συγκεντρωτικά τα scores των test σετ για όλα τα παραπάνω σενάρια με τη σειρά που πραγματοποιήθηκαν. Παρατηρούμε ότι η πρόβλεψη κατανάλωσης ενέργειας της κάρτας γραφικών είχε τα ίδια αποτελέσματα χρησιμοποιώντας τέσσερα και δύο χαρακτηριστικά ως εισόδους. Ομοίως, η πρόβλεψη κατανάλωσης ενέργειας του επεξεργαστή είχε τα ίδια αποτελέσματα χρησιμοποιώντας τρία και δύο χαρακτηριστικά ως εισόδους. Η πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών ήταν ελάχιστα πιο αποτελεσματική χρησιμοποιώντας 3 χαρακτηριστικά ως εισόδους.

Γραμμική Παλινδρόμηση

Στο πρώτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στον επεξεργαστή. Χρησιμοποιώντας τρία χαρακτηριστικά ως εισόδους, το μοντέλο της γραμμικής παλινδρόμησης δίνεται από την παρακάτω εξίσωση:

$$y = 0.37cpu_{\%} + 0.07gpu_{\%} - 0.007cpu_c + 24.67$$

Σχήμα 4.109: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 3 εισόδους (LR)

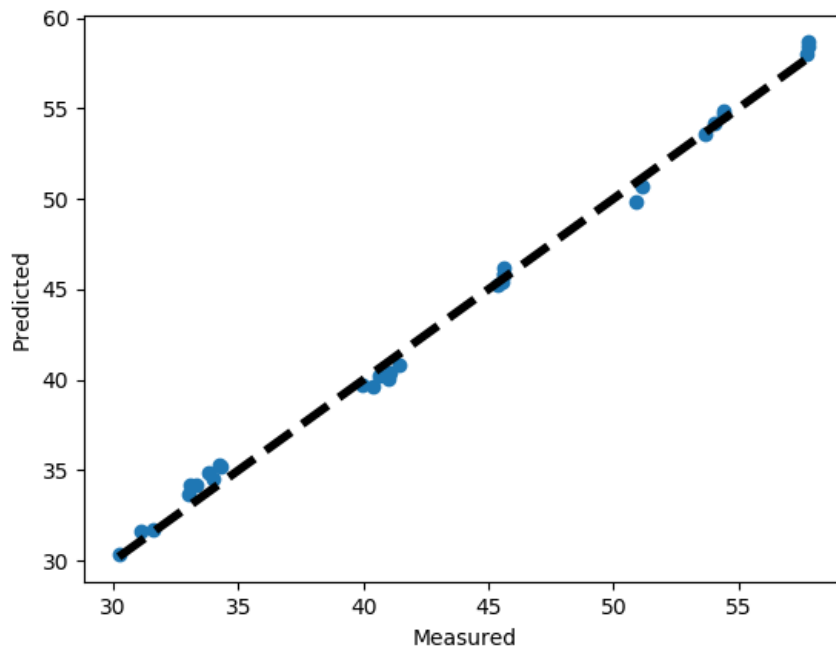


Όπως φαίνεται και στη παραπάνω εξίσωση, η χρήση του επεξεργαστή και της κάρτας γραφικών έχουν θετικούς συντελεστές, με τη χρήση της κάρτας γραφικών να

μην επηρεάζει πολύ το μοντέλο. Αντίθετα, η θερμοκρασία του επεξεργαστή έχει αρνητικό αμελητέο συντελεστή, μην επηρεάζοντας σημαντικά το μοντέλο. Η σταθερά του μοντέλου φέρει τη τιμή 24.67, ενώ το score στο test σετ είναι 0.99. Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση του επεξεργαστή, το μοντέλο της γραμμικής παλινδρόμησης δίνεται από την παρακάτω εξίσωση:

$$y = 0.37cpu_{\%} + 0.07gpu_{\%} + 24.37$$

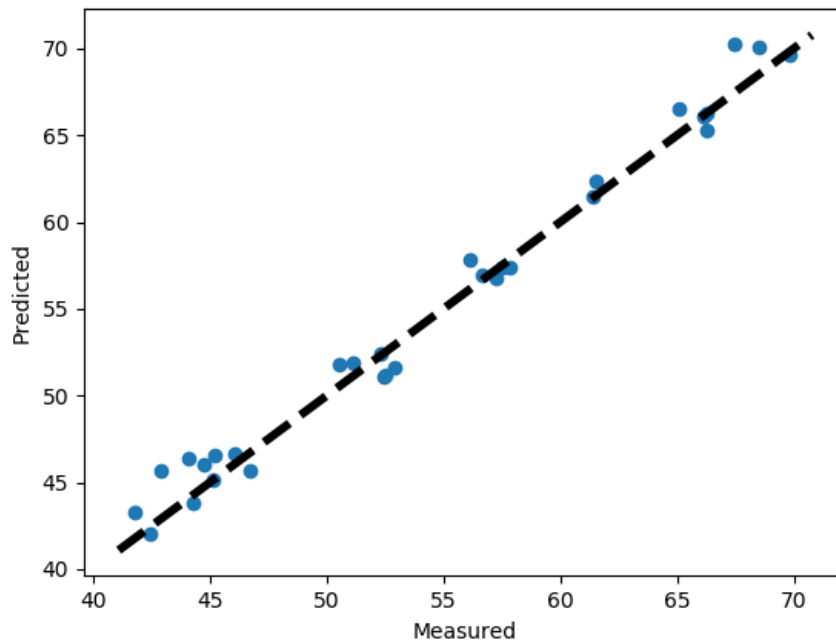
Σχήμα 4.110: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (LR)



Όπως φαίνεται στη παραπάνω εξίσωση, η χρήση του επεξεργαστή και της κάρτας γραφικών έχουν θετικούς συντελεστές, με τη χρήση της κάρτας γραφικών να μην επηρεάζει σημαντικά το μοντέλο. Η σταθερά του μοντέλου έχει τιμή 24.37, ενώ το score στο test σετ είναι 0.99. Στο δεύτερο πείραμα κάνουμε πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών. Χρησιμοποιώντας 3 χαρακτηριστικά ως εισόδους, το μοντέλο της γραμμικής παλινδρόμησης δίνεται από την παρακάτω εξίσωση:

$$y = 0.36cpu_{\%} + 0.08gpu_{\%} + 0.04cpu_c + 34.03$$

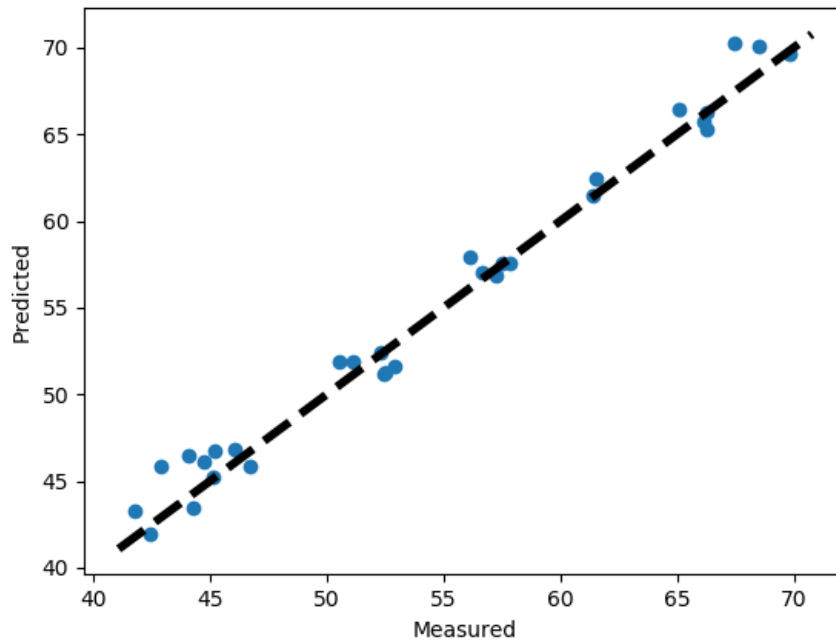
Σχήμα 4.111: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (LR)



Όπως γίνεται αντιληπτό στη παραπάνω εξίσωση, όλα τα χαρακτηριστικά έχουν θετικούς συντελεστές. Η χρήση του επεξεργαστή έχει κύριο ρόλο στο μοντέλο, ενώ η χρήση της κάρτας γραφικών και η θερμοκρασία του επεξεργαστή δε το επηρεάζουν σημαντικά. Η σταθερά του μοντέλου φέρει την τιμή 34.03, ενώ το score στο test σετ είναι 0.97. Για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 2 χαρακτηριστικά ως εισόδους, το μοντέλο της γραμμικής παλινδρόμησης δίνεται από την παρακάτω εξίσωση:

$$y = 0.37cpu_{\%} + 0.10gpu_{\%} + 35.89$$

Σχήμα 4.112: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (LR)

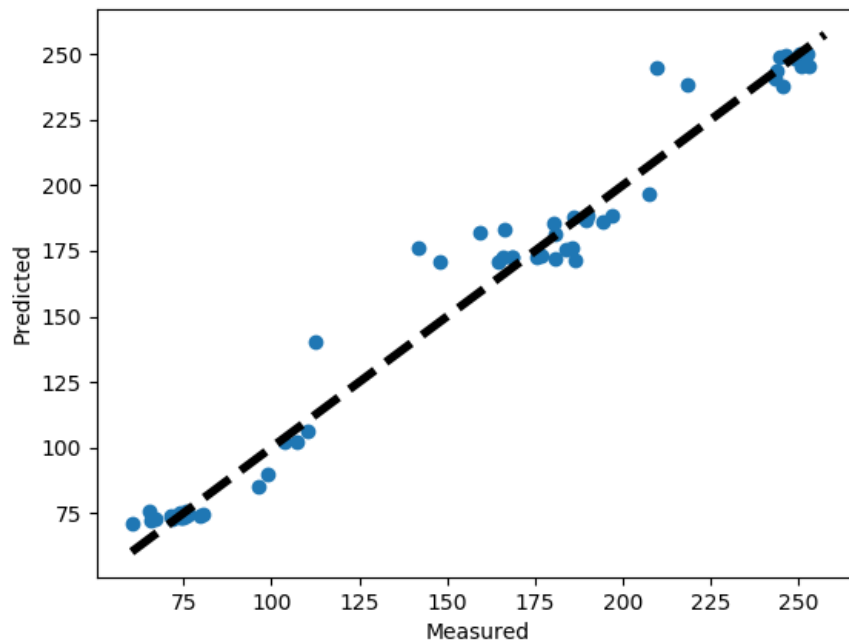


Παρατηρώντας την παραπάνω εξίσωση βλέπουμε πως, όλα τα χαρακτηριστικά έχουν θετικούς συντελεστές, με τη χρήση του επεξεργαστή να επηρεάζει περισσότερο το μοντέλο από ότι η χρήση της κάρτας γραφικών. Η σταθερά του μοντέλου έχει τιμή 35.89, ενώ το score στο test σετ είναι 0.97.

Στο τρίτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στην κάρτα γραφικών. Χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους, το μοντέλο της γραμμικής παλινδρόμησης δίνεται από την παρακάτω εξίσωση:

$$y = 0.01cpu_{\%} + 0.34gpu_{\%} + 0.61gpu_c + 1.94gpu_{tdp} - 22.43$$

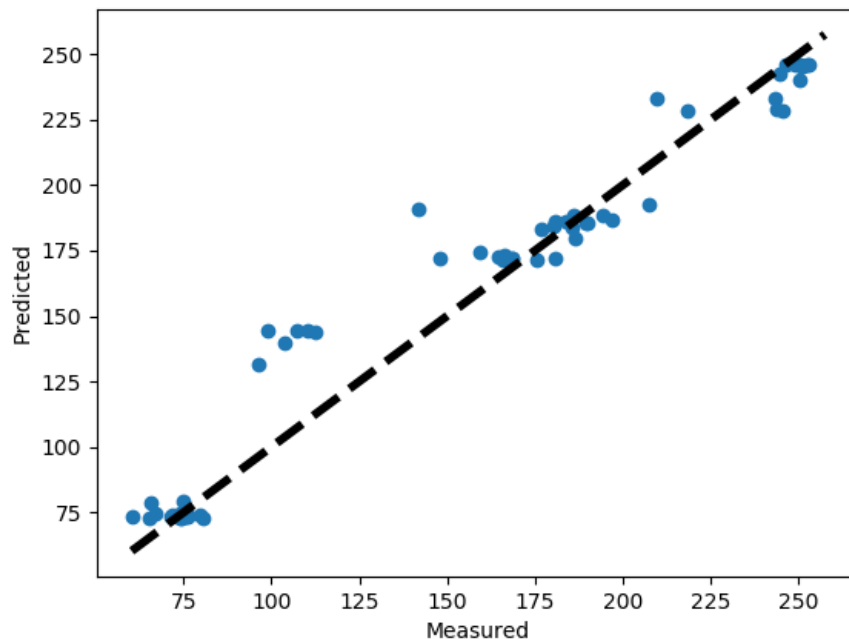
Σχήμα 4.113: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 4 εισόδους (LR)



Όπως φαίνεται και στην παραπάνω εξίσωση, όλα τα χαρακτηριστικά έχουν θετικούς συντελεστές, με τη θερμική σχεδιαστική ισχύ να έχει κύριο ρόλο στο μοντέλο, ενώ η χρήση του επεξεργαστή να μη το επηρεάζει σημαντικά. Η σταθερά του μοντέλου έχει τιμή -22.43 , ενώ το score στο test σετ είναι 0.96 . Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση της κάρτας γραφικών, το μοντέλο της γραμμικής παλινδρόμησης δίνεται από την παρακάτω εξίσωση:

$$y = 0.42cpu\% + 2.50gpu\% + 14.06$$

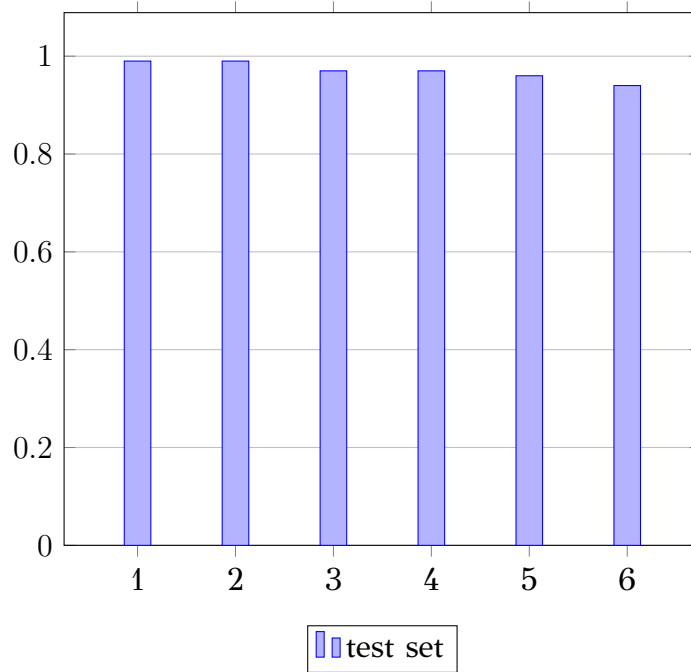
Σχήμα 4.114: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 εισόδους (LR)



Όπως γίνεται αντιληπτό στην παραπάνω εξίσωση, όλα τα χαρακτηριστικά έχουν θετικούς συντελεστές, με τη χρήση της κάρτας γραφικών να έχει κύριο ρόλο στο μοντέλο, ενώ η χρήση του επεξεργαστή να μη το επηρεάζει σημαντικά. Η σταθερά του μοντέλου έχει τιμή 14.06, ενώ το score στο test σετ είναι 0.94.

Τα σχήματα 4.109, 4.110, 4.111, 4.112, 4.113 και 4.114 απεικονίζουν διαγράμματα πραγματικών με προβλεπόμενων τιμών για κάθε μία από τις παραπάνω περιπτώσεις. Για κάθε σημείο, στον άξονα X αναπαριστάται η πραγματική τιμή του, ενώ στον άξονα Y η προβλεπόμενη τιμή. Όσο πιο κοντά είναι κάποιο σημείο στη γραμμή (μοντέλο), τόσο πιο σωστά πραγματοποιήθηκε η πρόβλεψη.

Στο παρακάτω σχήμα συνοψίζονται συγκεντρωτικά τα scores των train και test σετ για όλες τις παραπάνω περιπτώσεις με τη σειρά που πραγματοποιήθηκαν. Παρατηρούμε ότι η πρόβλεψη κατανάλωσης ενέργειας της κάρτας γραφικών είχε μεγαλύτερη επιτυχία χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους. Η πρόβλεψη κατανάλωσης ενέργειας του επεξεργαστή είχε τα ίδια αποτελέσματα χρησιμοποιώντας τρία και δύο χαρακτηριστικά ως εισόδους. Η πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών την ίδια αποτελεσματικότητα χρησιμοποιώντας τρία και δύο χαρακτηριστικά ως εισόδους.



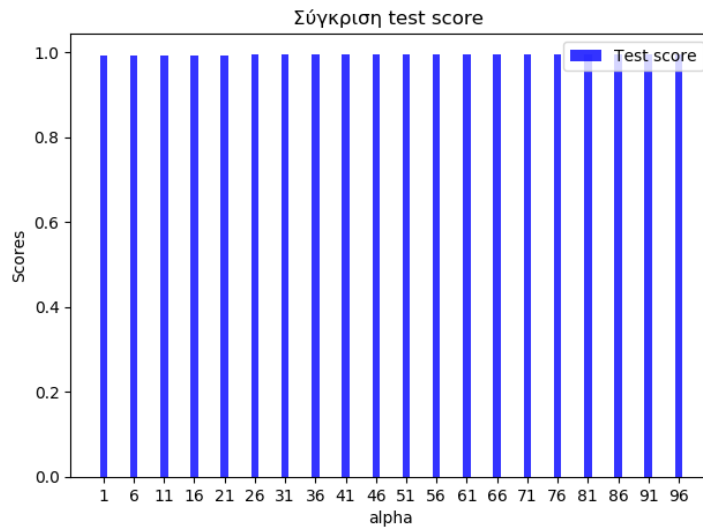
Παλινδρόμηση Ridge

Ο αλγόριθμος περιέχει μια επανάληψη (από 1 μέχρι 100 με βήμα 5), όπου η μεταβλητή $alpha$ λαμβάνει την τιμή της επανάληψης, για καλύτερη διεξαγωγή συμπερασμάτων. Στο πρώτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στον επεξεργαστή. Χρησιμοποιώντας τρία χαρακτηριστικά ως εισόδους, το μοντέλο της παλινδρόμησης ridge με τη μεταβλητή $alpha$ να έχει τη μέγιστη τιμή 96 δίνεται από την παρακάτω εξίσωση:

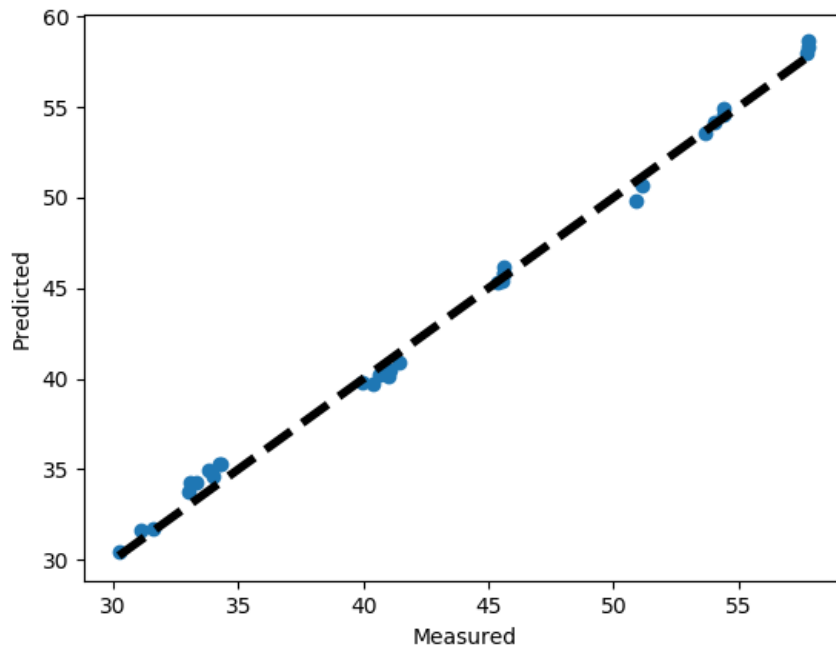
$$y = 0.36cpu_{\%} + 0.04gpu_{\%} + 0.0001cpu_c + 24.51$$

Όπως φαίνεται και στη παραπάνω εξίσωση, όλα τα χαρακτηριστικά έχουν θετικούς συντελεστές, με τη χρήση του επεξεργαστή να έχει κύριο ρόλο στο μοντέλο, ενώ η θερμοκρασία του επεξεργαστή δεν επηρεάζει σχεδόν καθόλου το μοντέλο, ως αμελητέα ποσότητα. Η σταθερά του μοντέλου φέρει τη τιμή 24.51, ενώ το score στο test σετ είναι 0.99. Καθώς αυξάνεται η μεταβλητή $alpha$ παρατηρείται μεταβολή στο 4ο δεκαδικό ψηφίο των test score (σχήμα 4.115).

Σχήμα 4.115: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (RR)



Σχήμα 4.116: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 3 εισόδους (RR)

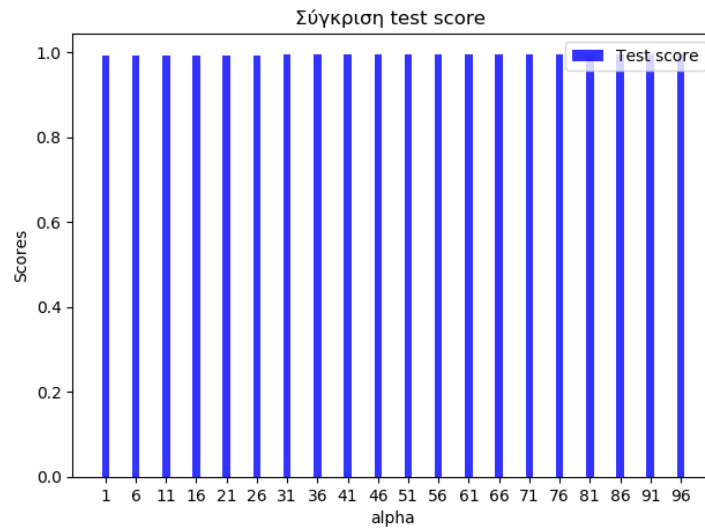


Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση του επεξεργαστή, το μοντέλο της παλινδρόμησης ridge με τη μεταβλητή $alpha$ να έχει τη μέγιστη τιμή 96 δίνεται από την παρακάτω εξίσωση:

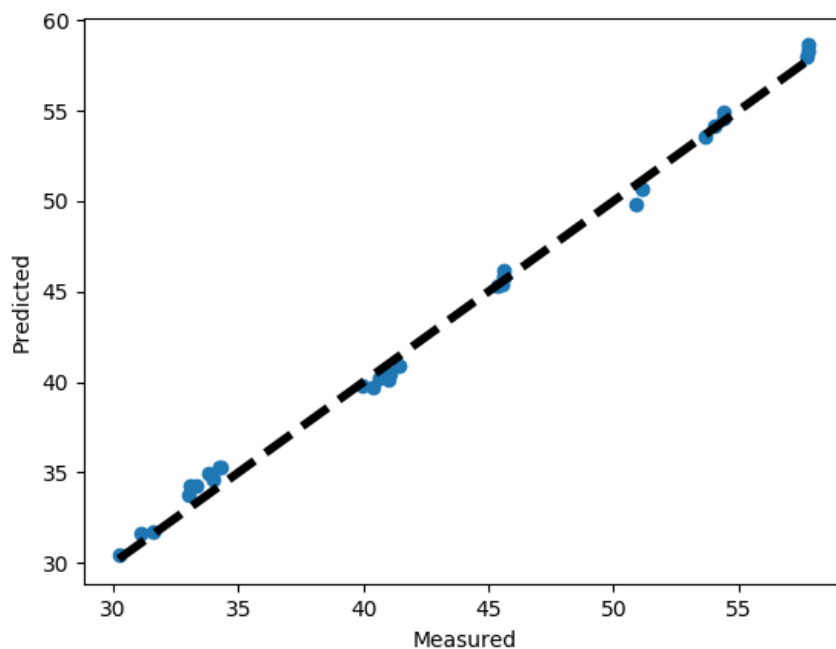
$$y = 0.36cpu\% + 0.04gpu\% + 24.52$$

Όπως φαίνεται στη παραπάνω εξίσωση, η χρήση του επεξεργαστή και της κάρτας γραφικών έχουν θετικούς συντελεστές, με τη χρήση της κάρτας γραφικών να μην επηρεάζει σημαντικά το μοντέλο. Η σταθερά του μοντέλου έχει τιμή 24.52, ενώ το score στο test σετ είναι 0.99. Καθώς αυξάνεται η μεταβλητή α παρατηρείται μεταβολή στο 4ο δεκαδικό ψηφίο των test score (σχήμα 4.117).

Σχήμα 4.117: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (RR)



Σχήμα 4.118: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (RR)



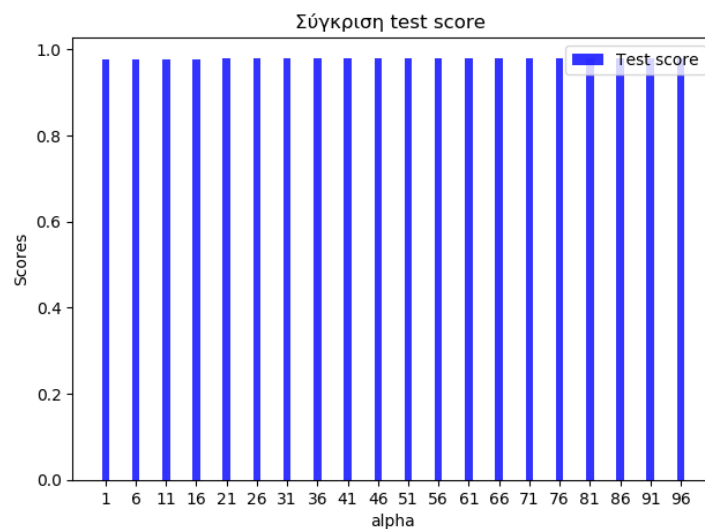
Στο δεύτερο πείραμα κάνουμε πρόβλεψη του αθροίσματος κατανάλωσης ενέρ-

γειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 3 χαρακτηριστικά ως εισόδους, το μοντέλο της παλινδρόμησης ridge με τη μεταβλητή $alpha$ να έχει τη μέγιστη τιμή 66 δίνεται από την παρακάτω εξίσωση:

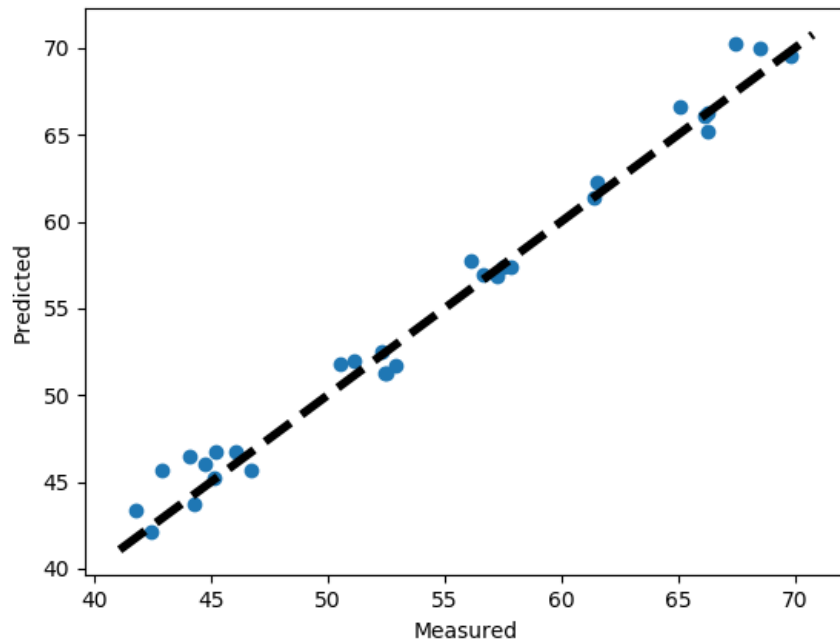
$$y = 0.36cpu_{\%} + 0.05gpu_{\%} + 0.04cpu_c + 34.26$$

Όπως γίνεται αντιληπτό στη παραπάνω εξίσωση, όλα τα χαρακτηριστικά έχουν θετικούς συντελεστές. η χρήση του επεξεργαστή έχει κύριο ρόλο στο μοντέλο, ενώ η χρήση της κάρτας γραφικών και η θερμοκρασία του επεξεργαστή δεν το επηρεάζουν σημαντικά. Η σταθερά του μοντέλου φέρει την τιμή 34.26, ενώ το score στο test σετ είναι 0.97. Καθώς αυξάνεται η μεταβλητή $alpha$ παρατηρείται μεταβολή στο 3ο δεκαδικό ψηφίο των test score (σχήμα 4.119).

Σχήμα 4.119: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (RR)



Σχήμα 4.120: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (RR)

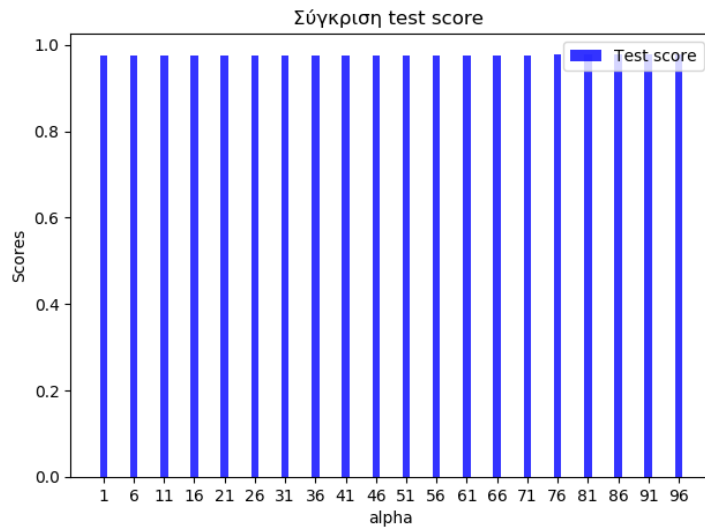


Για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 2 χαρακτηριστικά ως εισόδους, το μοντέλο της παλινδρόμησης ridge με τη μεταβλητή α να έχει τη μέγιστη τιμή 1 δίνεται από την παρακάτω εξίσωση:

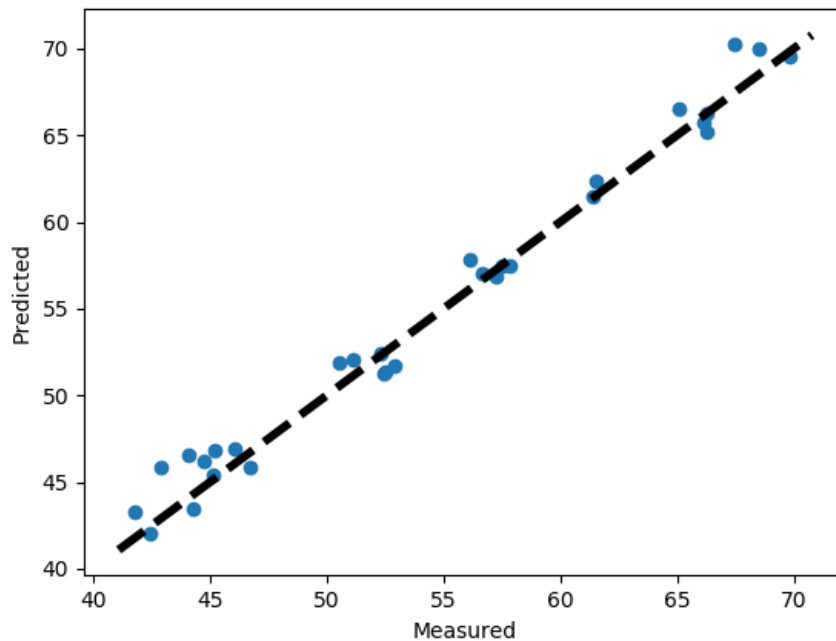
$$y = 0.36cpu\% + 0.06gpu\% + 36.08$$

Παρατηρώντας την παραπάνω εξίσωση βλέπουμε πως, όλα τα χαρακτηριστικά έχουν θετικούς συντελεστές, με τη χρήση του επεξεργαστή να έχει κύριο ρόλο στο μοντέλο και τη χρήση της κάρτας γραφικών να μη το επηρεάζει σημαντικά. Η σταθερά του μοντέλου έχει τιμή 36.08, ενώ το score στο test σετ είναι 0.97. Καθώς αυξάνεται η μεταβλητή α παρατηρείται μεταβολή στο 3ο δεκαδικό ψηφίο των test score (σχήμα 4.121).

Σχήμα 4.121: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (RR)



Σχήμα 4.122: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (RR)

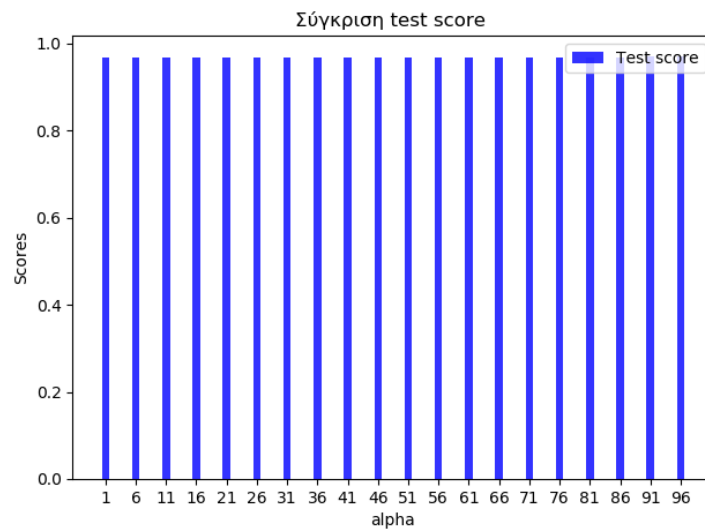


Στο τρίτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στην κάρτα γραφικών. Χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους, το μοντέλο της παλινδρόμησης ridge με τη μεταβλητή $alpha$ να έχει τη μέγιστη τιμή 96 δίνεται από την παρακάτω εξίσωση:

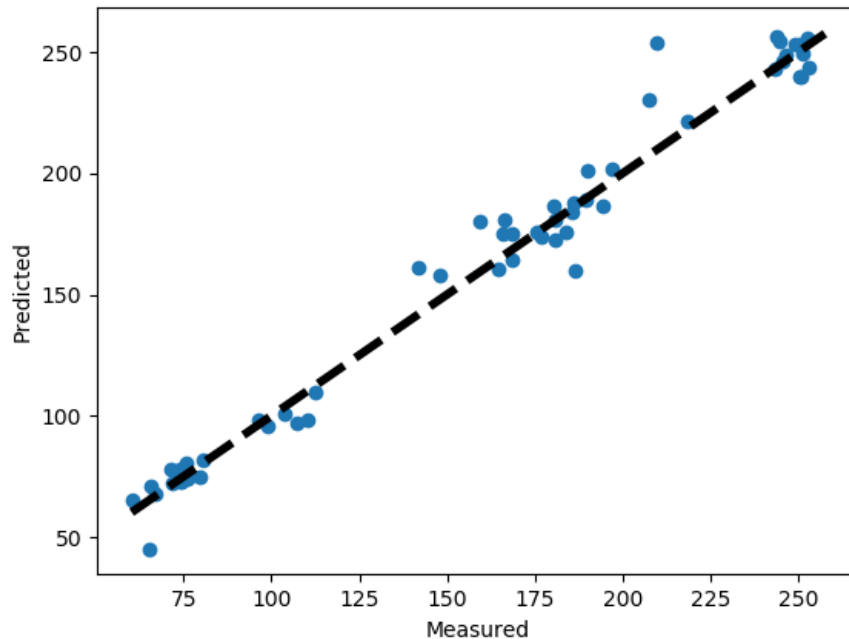
$$y = 0.01cpu_{\%} + 0.37gpu_{\%} + 0.61gpu_c + 1.91gpu_{tdp} - 22.31$$

Όπως φαίνεται και στην παραπάνω εξίσωση, όλα τα χαρακτηριστικά έχουν θετικούς συντελεστές. Η θερμική σχεδιαστική ισχύ της κάρτας γραφικών έχει κύριο ρόλο στο μοντέλο, ενώ η χρήση του επεξεργαστή δε το επηρεάζει σημαντικά. Η σταθερά του μοντέλου έχει τιμή -22.31, ενώ το score στο test σετ είναι 0.96. Καθώς αυξάνεται η μεταβλητή *alpha* παρατηρείται μεταβολή στο 4ο δεκαδικό ψηφίο των test score (σχήμα 4.123).

Σχήμα 4.123: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (RR)



Σχήμα 4.124: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 4 εισόδους (RR)

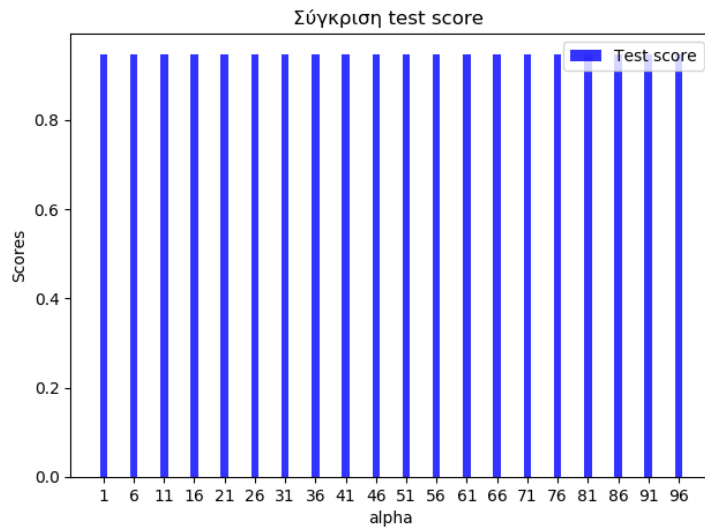


Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση της κάρτας γραφικών, το μοντέλο της παλινδρόμησης ridge με τη μεταβλητή $alpha$ να έχει τη μέγιστη τιμή 1 δίνεται από την παρακάτω εξίσωση:

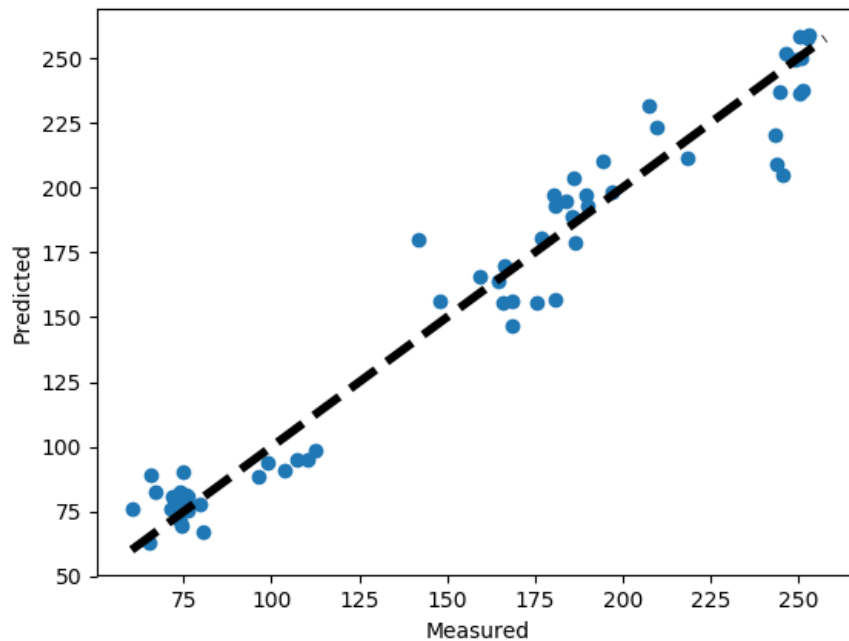
$$y = 0.42cpu_{\%} + 2.50gpu_{\%} + 14.06$$

Όπως γίνεται αντιληπτό στην παραπάνω εξίσωση, η χρήση του επεξεργαστή και της κάρτας γραφικών έχουν θετικούς συντελεστές, με τη χρήση της κάρτας γραφικών να έχει κύριο ρόλο στο μοντέλο ενώ η χρήση του επεξεργαστή να το επηρεάζει λιγότερο. Η σταθερά του μοντέλου έχει τιμή 14.06, ενώ το score στο test σετ είναι 0.94. Καθώς αυξάνεται η μεταβλητή $alpha$ παρατηρείται μεταβολή στο 5ο δεκαδικό ψηφίο των test score (σχήμα 4.125).

Σχήμα 4.125: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (RR)



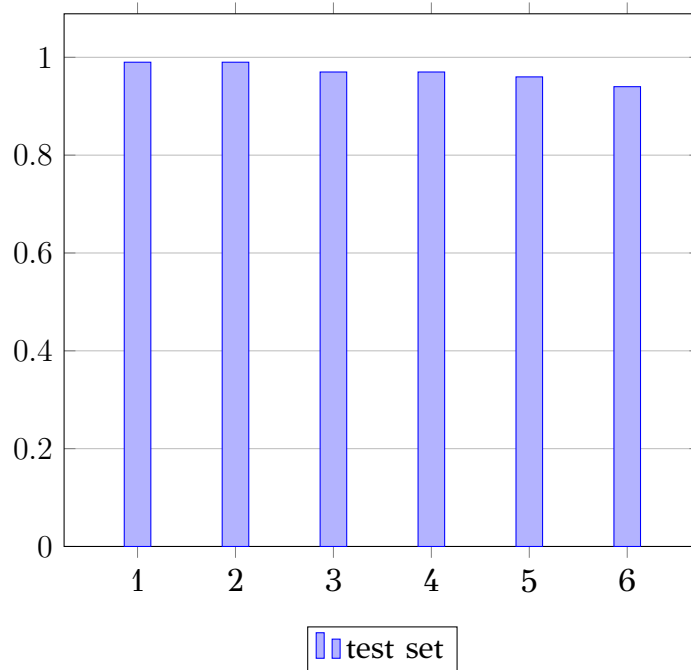
Σχήμα 4.126: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 εισόδους (RR)



Τα σχήματα 4.116, 4.118, 4.120, 4.122, 4.124 και 4.126 απεικονίζουν διαγράμματα πραγματικών με προβλεπόμενων τιμών για κάθε μία από τις παραπάνω περιπτώσεις. Για κάθε σημείο, στον άξονα X αναπαριστάται η πραγματική τιμή του, ενώ στον άξονα Y η προβλεπόμενη τιμή. Όσο πιο κοντά είναι κάποιο σημείο στη γραμμή (μοντέλο), τόσο πιο σωστά πραγματοποιήθηκε η πρόβλεψη.

Στο παρακάτω σχήμα συνοψίζονται συγκεντρωτικά τα scores των train και test

σετ για όλες τις παραπάνω περιπτώσεις με τη σειρά που πραγματοποιήθηκαν. Παρατηρούμε ότι η πρόβλεψη κατανάλωσης ενέργειας της κάρτας γραφικών είχε μεγαλύτερη επιτυχία χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους. Η πρόβλεψη κατανάλωσης ενέργειας του επεξεργαστή ήταν ήταν το ίδιο αποτελεσματική χρησιμοποιώντας τρία και δύο χαρακτηριστικά ως εισόδους. Η πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών είχε το ίδιο αποτέλεσμα χρησιμοποιώντας τρία και δύο χαρακτηριστικά ως εισόδους.



Παλινδρόμηση Lasso

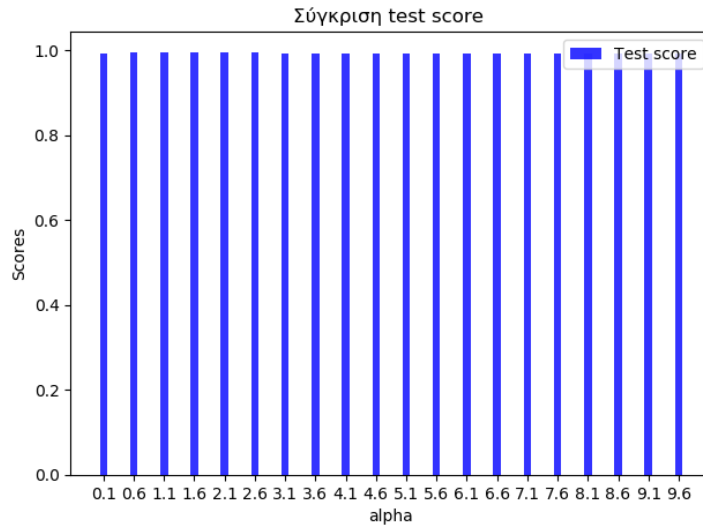
Ο αλγόριθμος περιέχει μια επανάληψη (από 0.1 μέχρι 10 με βήμα 0.5), όπου η μεταβλητή $alpha$ λαμβάνει την τιμή της επανάληψης, για καλύτερη διεξαγωγή συμπερασμάτων. Στο πρώτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στον επεξεργαστή. Χρησιμοποιώντας τρία χαρακτηριστικά ως εισόδους, το μοντέλο της παλινδρόμησης lasso με τη μεταβλητή $alpha$ να έχει τη βέλτιστη τιμή 0.6 δίνεται από την παρακάτω εξίσωση:

$$y = 0.36cpu_{\%} + 0gpru_{\%} + 0cpu_c + 24.66$$

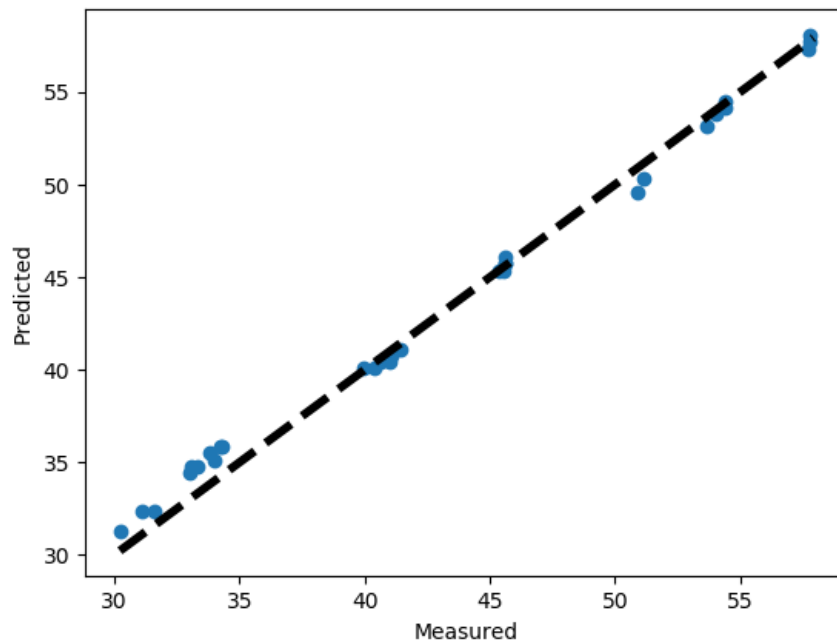
Όπως φαίνεται και στη παραπάνω εξίσωση, η χρήση του επεξεργαστή έχει θετικό συντελεστή. Η χρήση της κάρτας γραφικών και η θερμοκρασία του επεξεργαστή

έχουν μηδενικό συντελεστή, μην επηρεάζοντας το μοντέλο. Η σταθερά του μοντέλου φέρει τη τιμή 24.66, ενώ το score στο test σετ είναι 0.99. Καθώς αυξάνεται η μεταβλητή $alpha$ παρατηρείται μικρή μεταβολή στο test score (σχήμα 4.127).

Σχήμα 4.127: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (LR)



Σχήμα 4.128: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 3 εισόδους (LR)

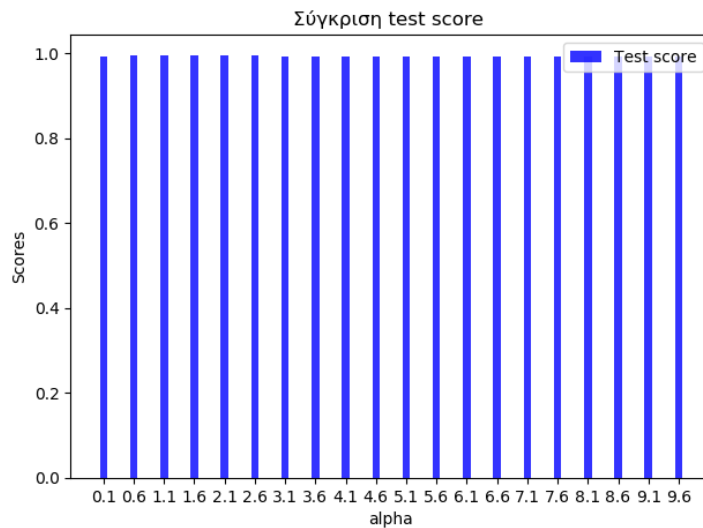


Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση του επεξεργαστή, το μοντέλο της παλινδρόμησης lasso με τη μεταβλητή $alpha$ να έχει τη βέλτιστη τιμή 2.6 δίνεται από την παρακάτω εξίσωση:

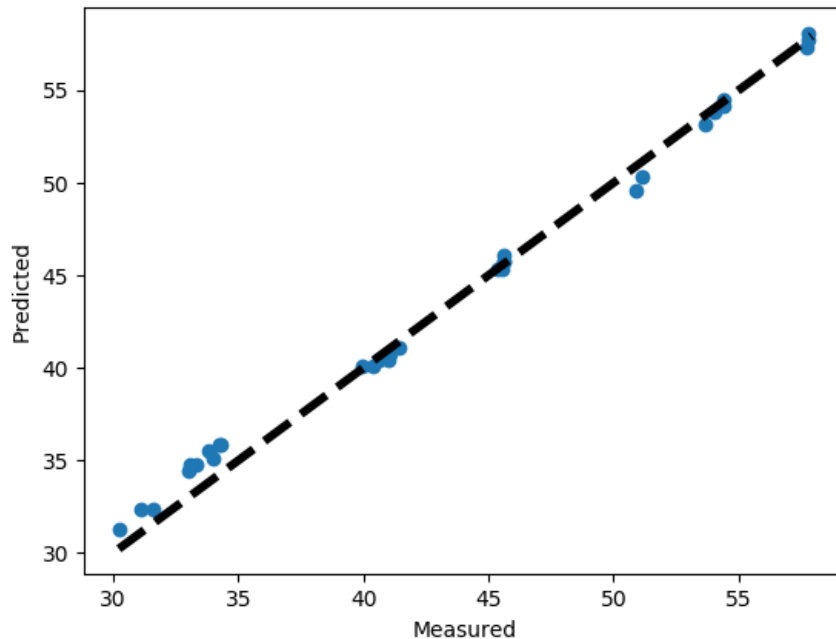
$$y = 0.36cpu\% + 0gru\% + 24.66$$

Όπως φαίνεται στη παραπάνω εξίσωση, η χρήση του επεξεργαστή έχει θετικό συντελεστή ενώ η χρήση της κάρτας γραφικών έχει μηδενικό συντελεστή, με τη χρήση της κάρτας γραφικών να μην επηρεάζει καθόλου το μοντέλο. Η σταθερά του μοντέλου έχει τιμή 24.66, ενώ το score στο test σετ είναι 0.99. Καθώς αυξάνεται η μεταβλητή *alpha* παρατηρείται μεταβολή στο 3ο δεκαδικό ψηφίο των test score (σχήμα 4.129).

Σχήμα 4.129: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (LR)



Σχήμα 4.130: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (LR)

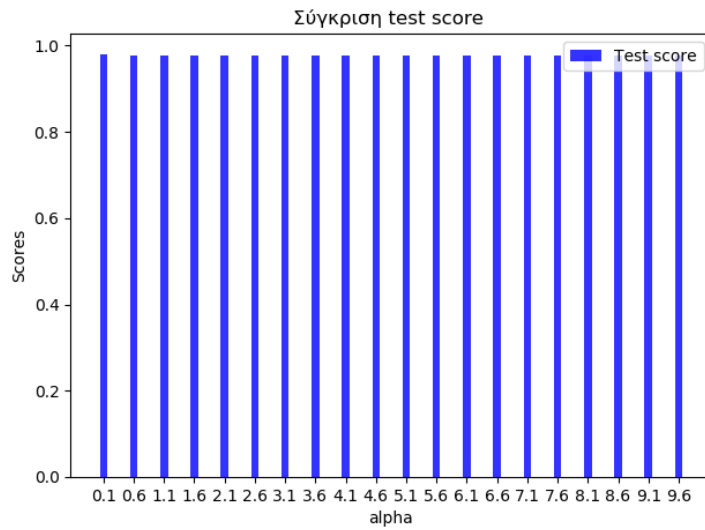


Στο δεύτερο πείραμα κάνουμε πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 3 χαρακτηριστικά ως εισόδους. Το μοντέλο της παλινδρόμησης lasso με τη μεταβλητή $alpha$ να έχει τη βέλτιστη τιμή 0.1 δίνεται από την παρακάτω εξίσωση:

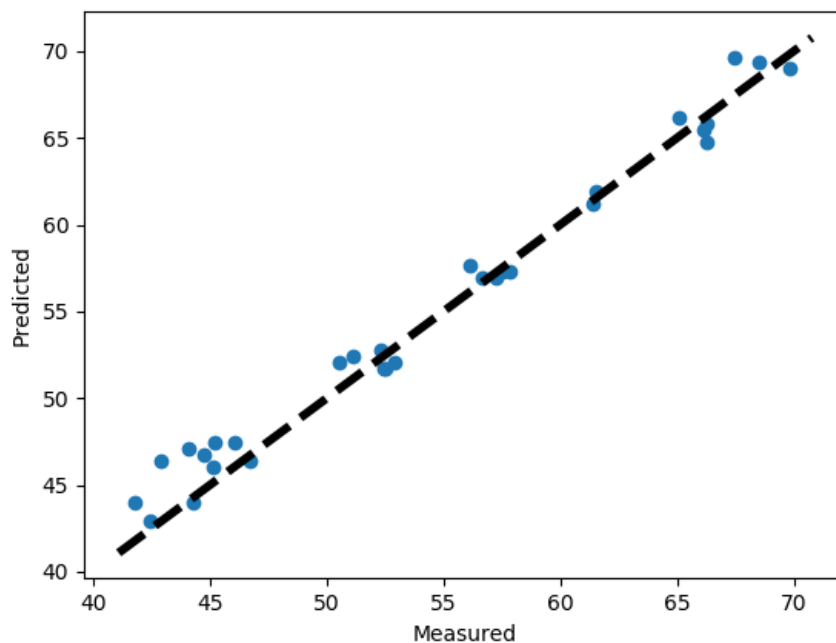
$$y = 0.36cpu_{\%} + 0.03gpu_{\%} + 0.03cpu_c + 34.65$$

Όπως γίνεται αντιληπτό στη παραπάνω εξίσωση, όλα τα χαρακτηριστικά έχουν θετικούς συντελεστές, με τη χρήση του επεξεργαστή να έχει κύριο ρόλο στο μοντέλο, ενώ η χρήση της κάρτας γραφικών και η θερμοκρασία του επεξεργαστή να μη το επηρεάζουν σημαντικά. Η σταθερά του μοντέλου φέρει την τιμή 34.65, ενώ το score στο test σετ είναι 0.97. Καθώς αυξάνεται η μεταβλητή $alpha$ παρατηρείται μεταβολή στο 3ο δεκαδικό ψηφίο των test score (σχήμα 4.131).

Σχήμα 4.131: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (LR)



Σχήμα 4.132: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (LR)

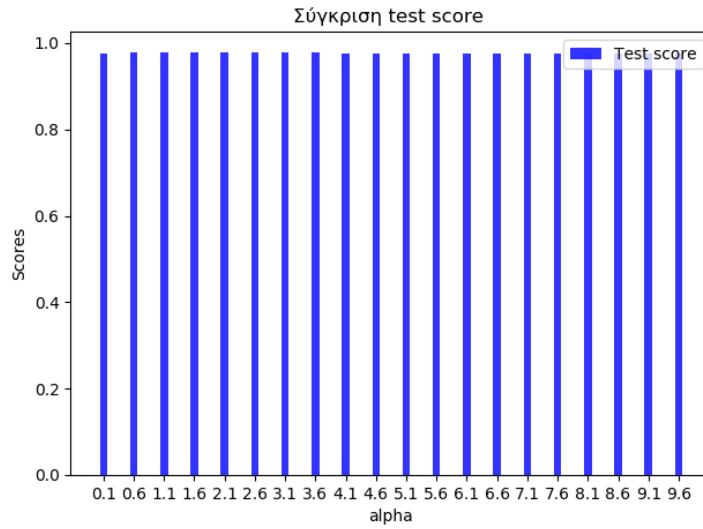


Για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 2 χαρακτηριστικά ως εισόδους, το μοντέλο της παλινδρόμησης lasso με τη μεταβλητή α να έχει τη βέλτιστη τιμή 1.6 δίνεται από την παρακάτω εξίσωση:

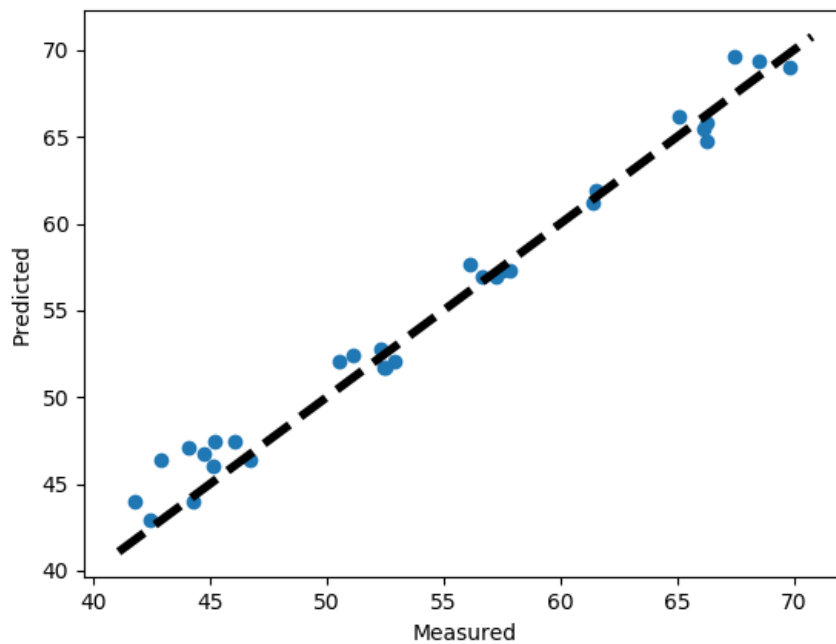
$$y = 0.36cpu_{\%} + 0.9gpu_{\%} + 36.40$$

Παρατηρώντας την παραπάνω εξίσωση βλέπουμε πως, η χρήση του επεξεργαστή έχει θετικό συντελεστή ενώ η χρήση της κάρτας γραφικών έχει μηδενικό συντελεστή. Η σταθερά του μοντέλου έχει τιμή 36.40, ενώ το score στο test σετ είναι 0.97. Καθώς αυξάνεται η μεταβλητή α παρατηρείται μεταβολή στο 3ο δεκαδικό ψηφίο των test score (σχήμα 4.133).

Σχήμα 4.133: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (LR)



Σχήμα 4.134: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (LR)



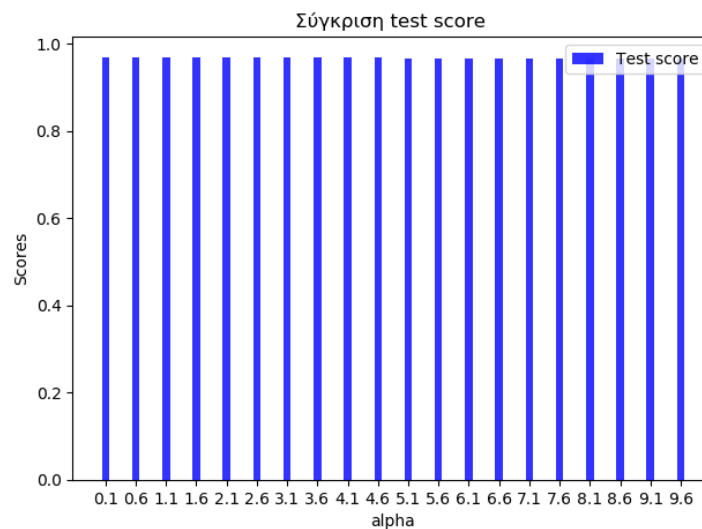
Στο τρίτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στην κάρτα

γραφικών. Χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους, το μοντέλο της παλινδρόμησης lasso με τη μεταβλητή $alpha$ να έχει τη βέλτιστη τιμή 0.1 δίνεται από την παρακάτω εξίσωση:

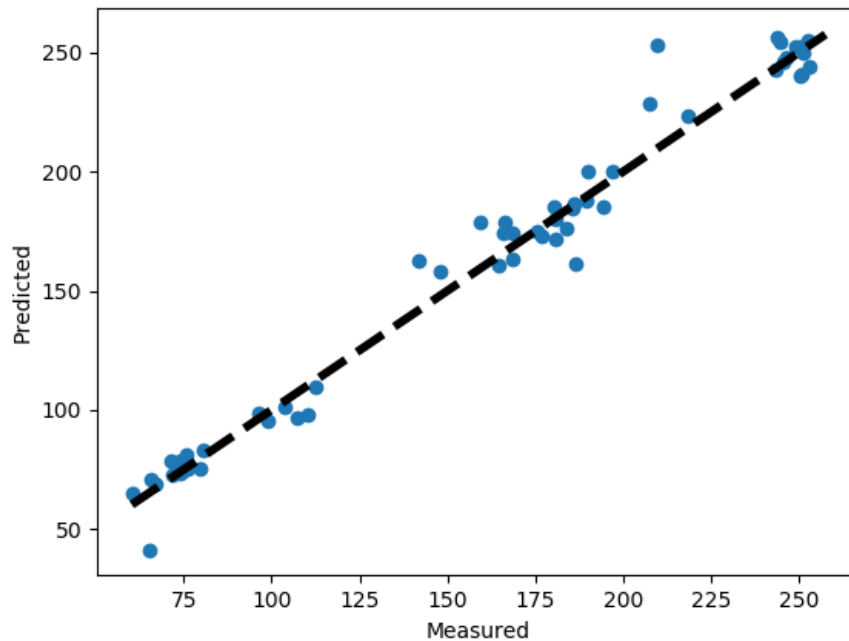
$$y = 0.01cpu_{\%} + 0.34gpu_{\%} + 0.61gpu_c + 1.94gpu_{tdp} - 22.40$$

Όπως φαίνεται και στην παραπάνω εξίσωση, όλα τα χαρακτηριστικά έχουν θετικούς συντελεστές. Η θερμική σχεδιαστική ισχύ της κάρτας γραφικών έχει κύριο ρόλο στο μοντέλο, ενώ η χρήση του επεξεργαστή έχει τη μικρότερη επιρροή. Η σταθερά του μοντέλου έχει τιμή -22.40, ενώ το score στο test σετ είναι 0.96. Καθώς αυξάνεται η μεταβλητή $alpha$ παρατηρείται μεταβολή στο 4ο δεκαδικό ψηφίο των test score (σχήμα 4.135).

Σχήμα 4.135: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (LR)



Σχήμα 4.136: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 4 εισόδους (LR)

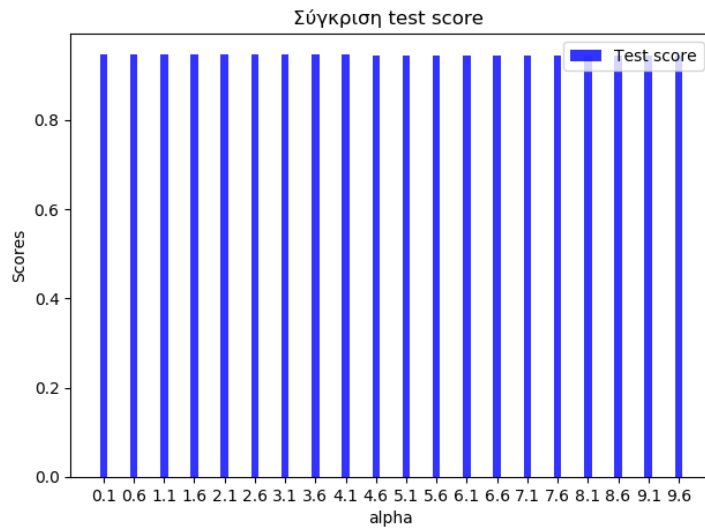


Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση της κάρτας γραφικών, το μοντέλο της παλινδρόμησης lasso με τη μεταβλητή α να έχει τη βέλτιστη τιμή 0,1 δίνεται από την παρακάτω εξίσωση:

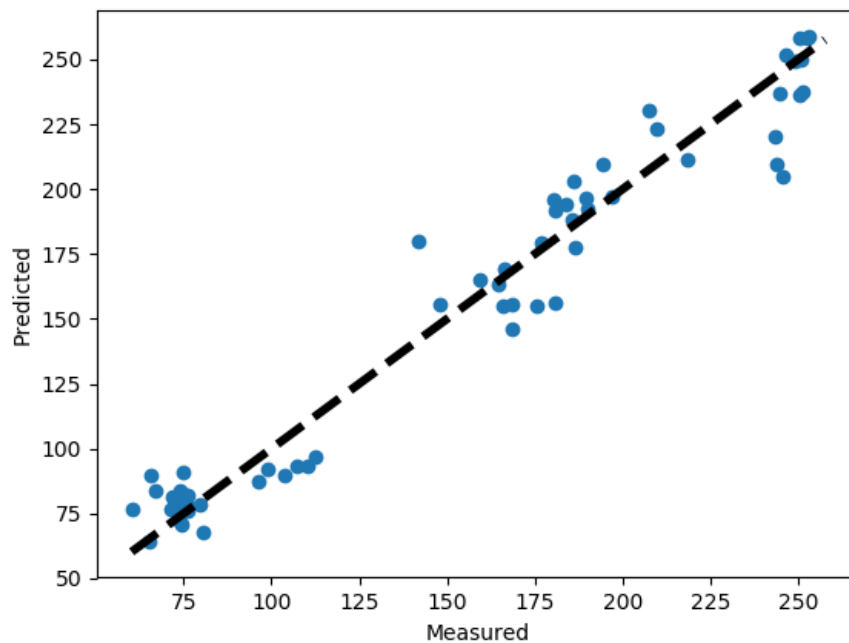
$$y = 0.42cpu\% + 2.50gpu\% + 14.08$$

Όπως γίνεται αντιληπτό στην παραπάνω εξίσωση, η χρήση του επεξεργαστή και της κάρτας γραφικών έχουν θετικούς συντελεστές, με τη χρήση της κάρτας γραφικών να έχει κύριο ρόλο στο μοντέλο και τη χρήση του επεξεργαστή να το επηρεάζει λιγότερο. Η σταθερά του μοντέλου έχει τιμή 14.08, ενώ το score στο test σετ είναι 0.94. Καθώς αυξάνεται η μεταβλητή α παρατηρείται μεταβολή στο 3ο δεκαδικό ψηφίο των test score (σχήμα 4.137).

Σχήμα 4.137: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (LR)



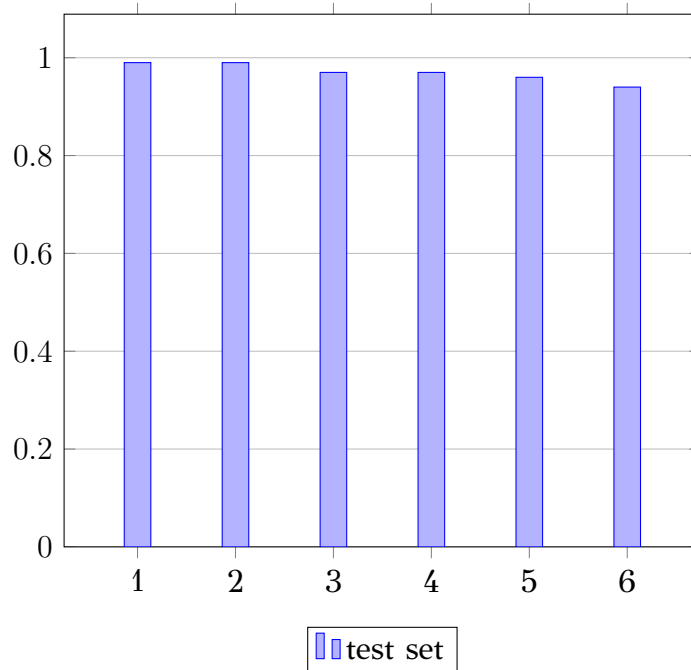
Σχήμα 4.138: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 εισόδους (LR)



Τα σχήματα 4.128, 4.130, 4.132, 4.134, 4.136 και 4.138 απεικονίζουν διαγράμματα πραγματικών με προβλεπόμενων τιμών για κάθε μία από τις παραπάνω περιπτώσεις. Για κάθε σημείο, στον άξονα X αναπαριστάται η πραγματική τιμή του, ενώ στον άξονα Y η προβλεπόμενη τιμή. Όσο πιο κοντά είναι κάποιο σημείο στη γραμμή (μοντέλο), τόσο πιο σωστά πραγματοποιήθηκε η πρόβλεψη.

Στο παρακάτω σχήμα συνοψίζονται συγκεντρωτικά τα scores των train και test

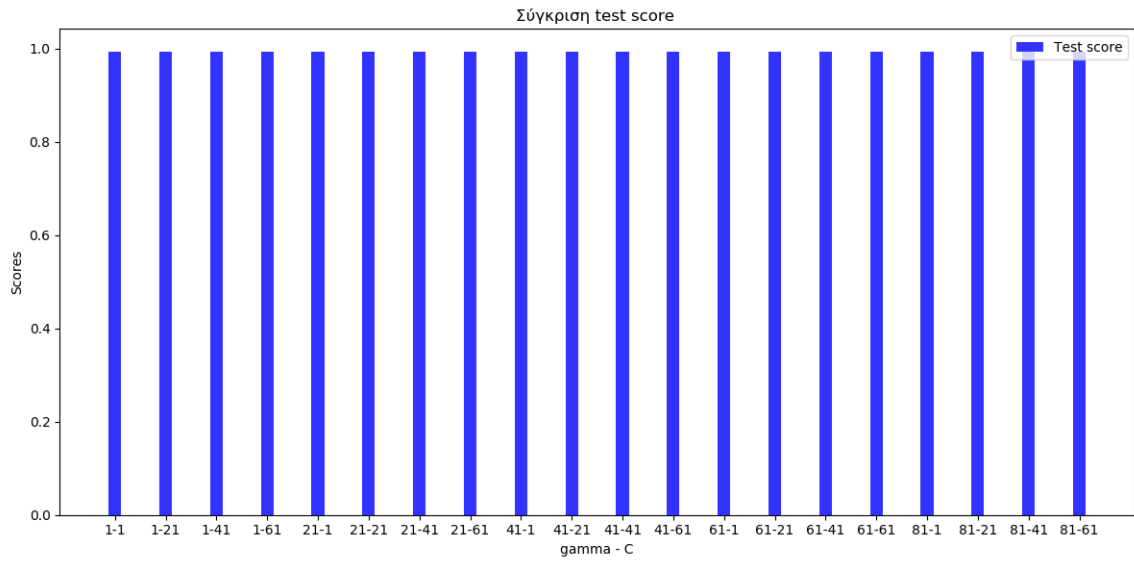
σετ για όλες τις παραπάνω περιπτώσεις με τη σειρά που πραγματοποιήθηκαν. Παρατηρούμε ότι η πρόβλεψη κατανάλωσης ενέργειας της κάρτας γραφικών είχε μεγαλύτερη επιτυχία χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους. Η πρόβλεψη κατανάλωσης ενέργειας του επεξεργαστή ήταν ήταν το ίδιο αποτελεσματική χρησιμοποιώντας τρία και δύο χαρακτηριστικά ως εισόδους. Η πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών είχε το ίδιο αποτέλεσμα χρησιμοποιώντας τρία και δύο χαρακτηριστικά ως εισόδους.



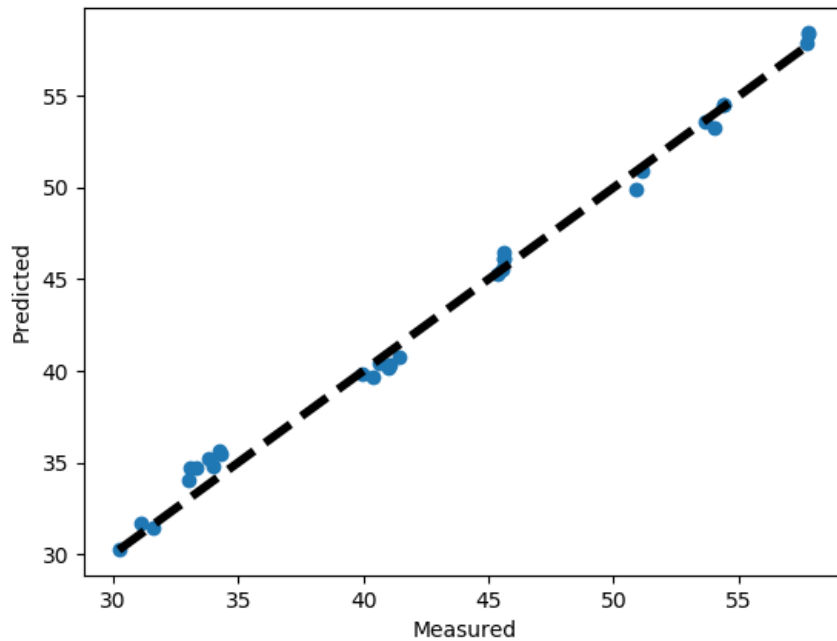
Μηχανές Διανυσμάτων Υποστήριξης

Ο αλγόριθμος περιέχει δύο επαναλήψεις (από 1 μέχρι 100 με βήμα 20 για τη μεταβλητή γ και από 1 μέχρι 80 με βήμα 20 για τη μεταβλητή C), όπου οι μεταβλητές γ και C λαμβάνουν την τιμή των επαναλήψεων, για καλύτερη διεξαγωγή συμπερασμάτων. Στο πρώτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στον επεξεργαστή. Χρησιμοποιώντας τρία χαρακτηριστικά ως εισόδους, ορίζουμε τις μεταβλητές γ και C στη τιμή 1 και λαμβάνουμε test score 0.99 (σχήμα 4.139).

Σχήμα 4.139: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (SVM)

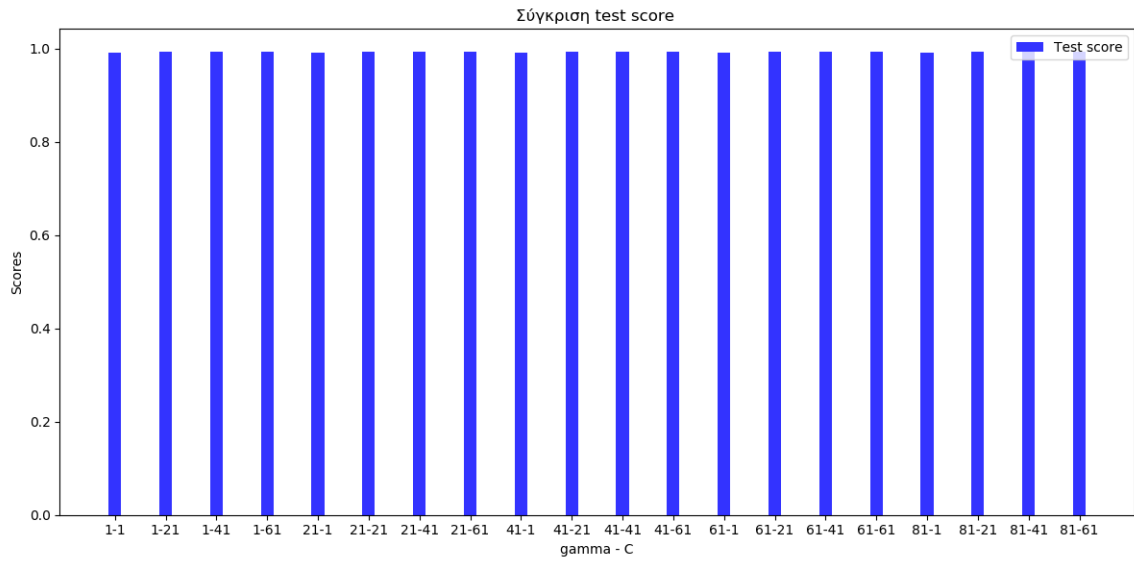


Σχήμα 4.140: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 3 εισόδους (SVM)

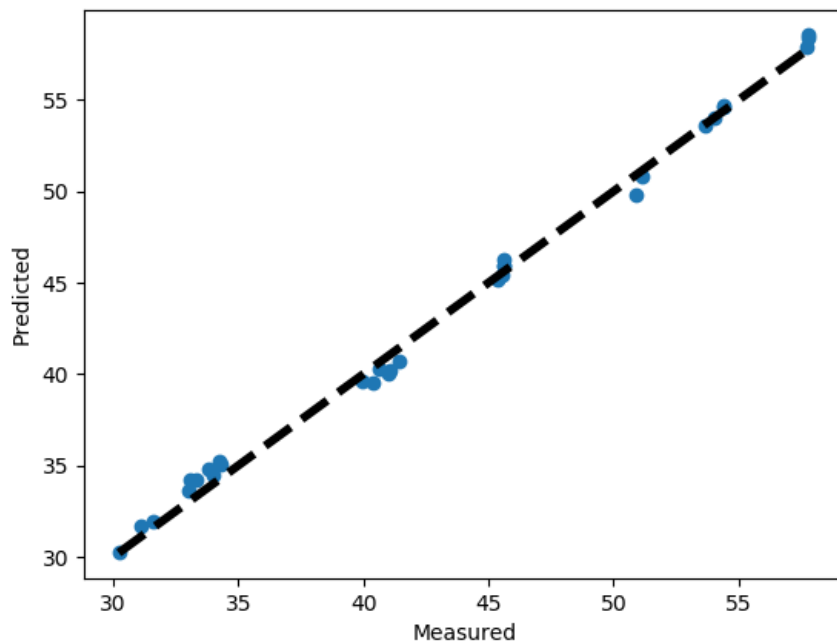


Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση του επεξεργαστή, ορίζουμε τις μεταβλητές $gamma$ και C στη τιμή 1 και λαμβάνουμε test score 0.99 (σχήμα 4.141).

Σχήμα 4.141: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (SVM)

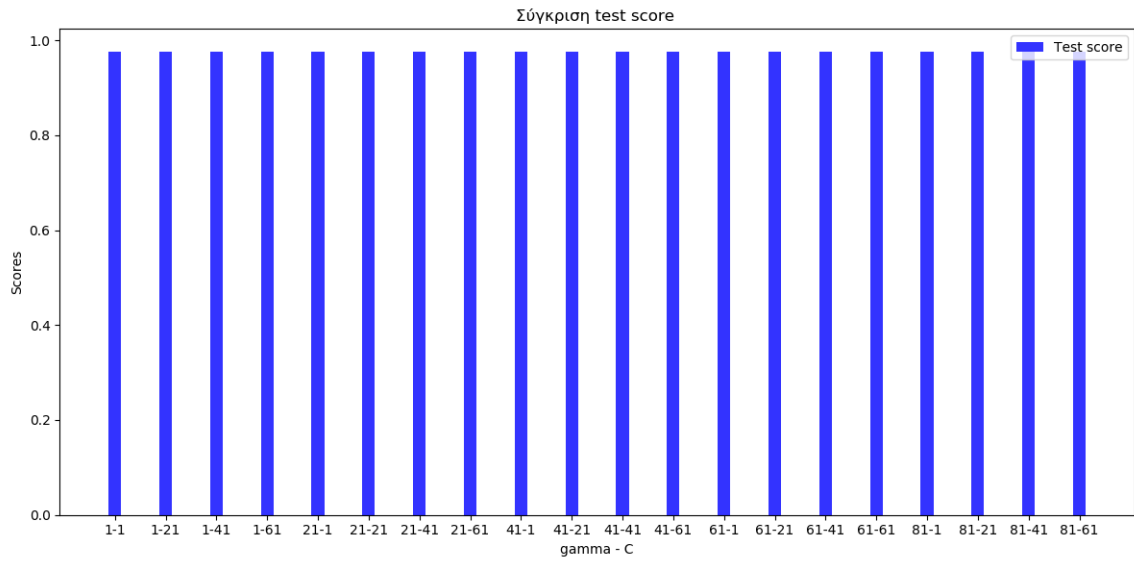


Σχήμα 4.142: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (SVM)

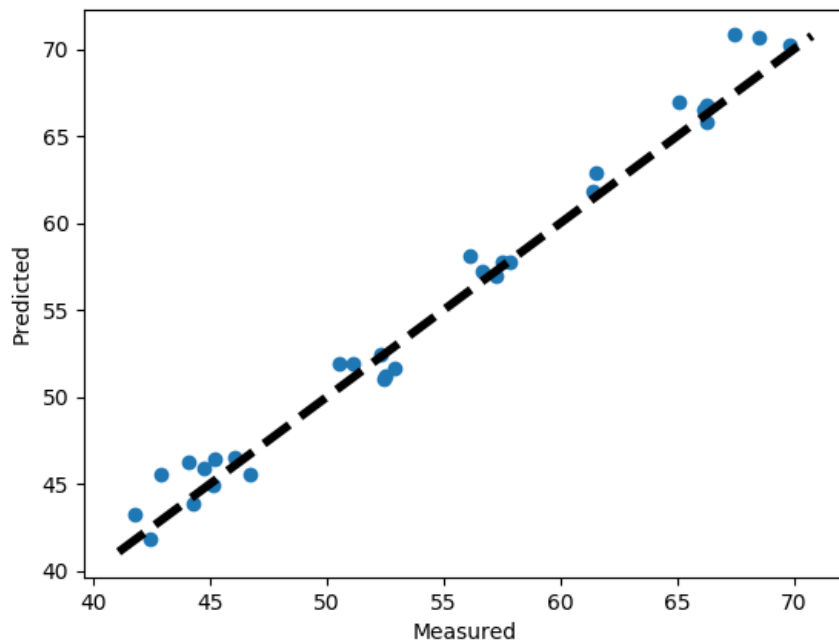


Στο δεύτερο πείραμα κάνουμε πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιούμε 3 χαρακτηριστικά ως εισόδους. Ορίζουμε τις μεταβλητές $gamma$ και C στη τιμή 1 και λαμβάνουμε test score 0.97 (σχήμα 4.143).

Σχήμα 4.143: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (SVM)

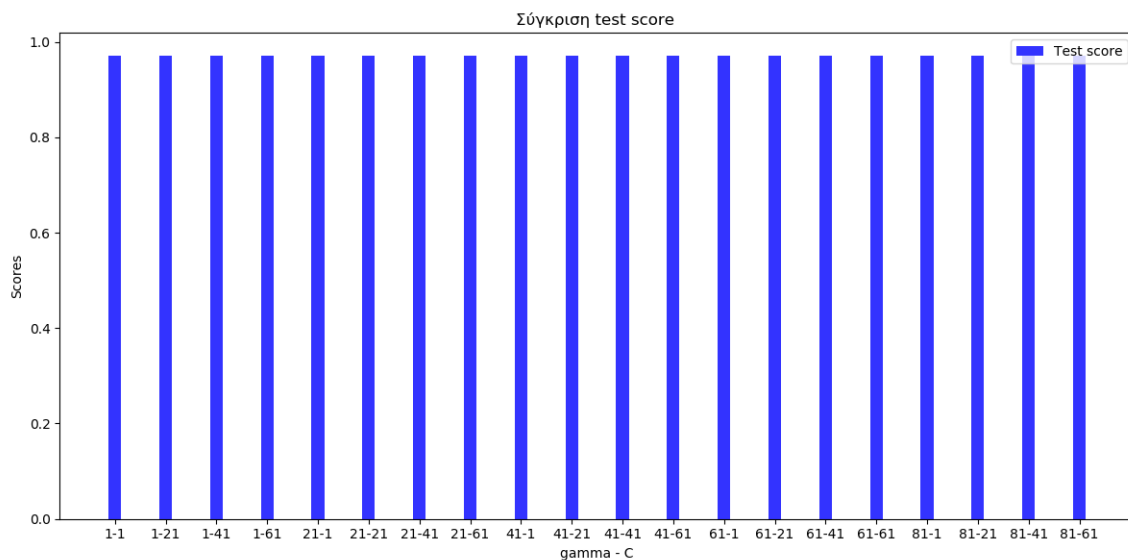


Σχήμα 4.144: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (SVM)

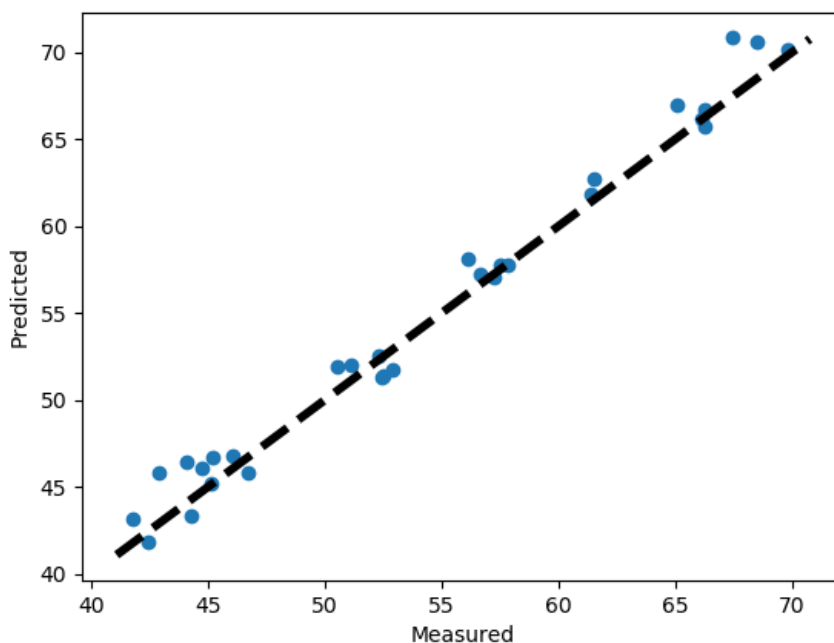


Για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιούμε 2 χαρακτηριστικά ως εισόδους. Ορίζουμε τις μεταβλητές $gamma$ και C στη τιμή 1 και λαμβάνουμε test score 0.97 (σχήμα 4.145).

Σχήμα 4.145: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (SVM)

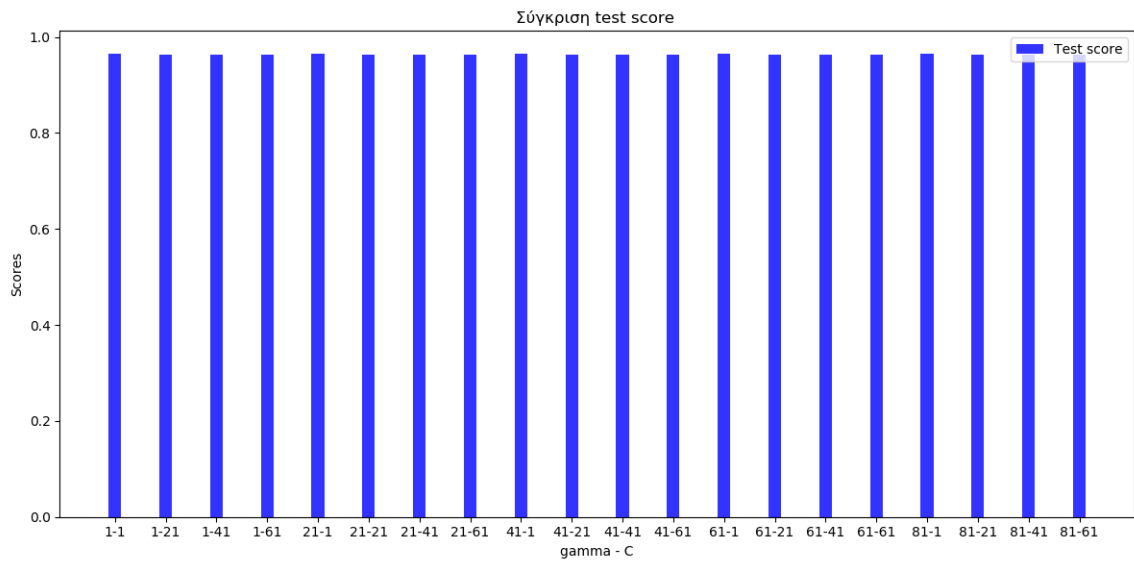


Σχήμα 4.146: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (SVM)

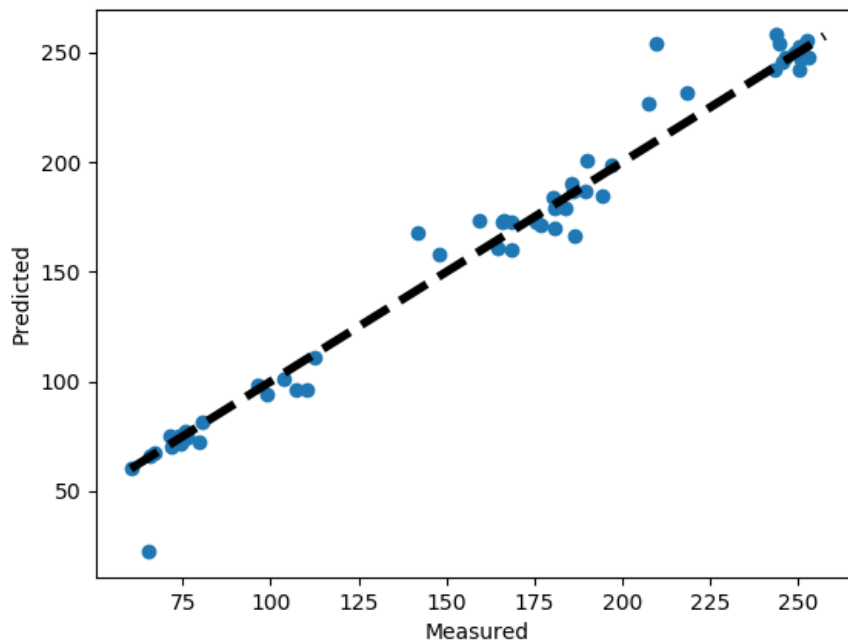


Στο τρίτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στην κάρτα γραφικών. Χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους, ορίζουμε τις μεταβλητές $gamma$ και C στη τιμή 1 και λαμβάνουμε test score 0.96 (σχήμα 4.147).

Σχήμα 4.147: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (SVM)

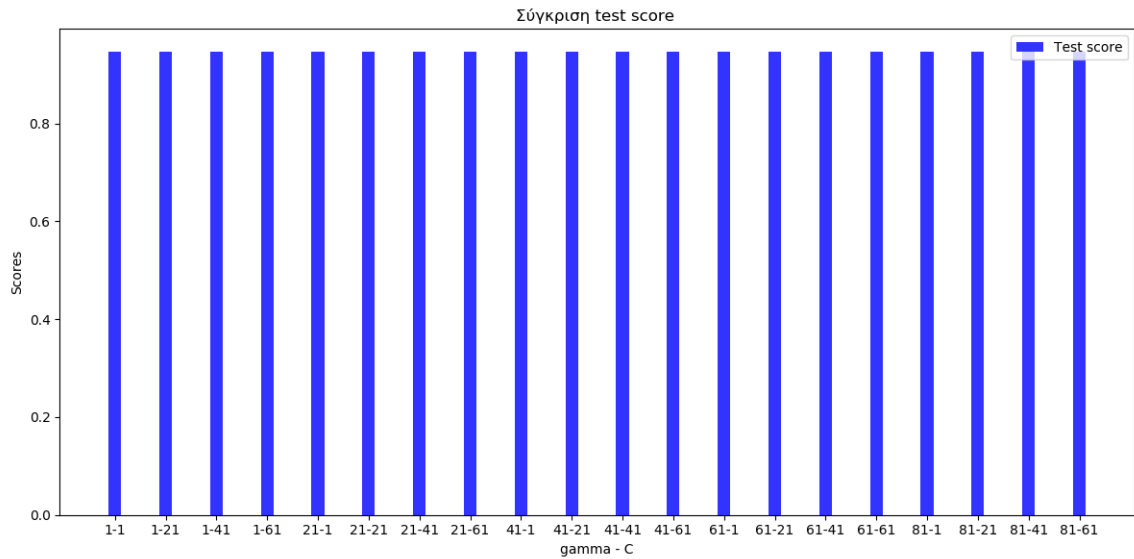


Σχήμα 4.148: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης του επεξεργαστή με 4 εισόδους (SVM)

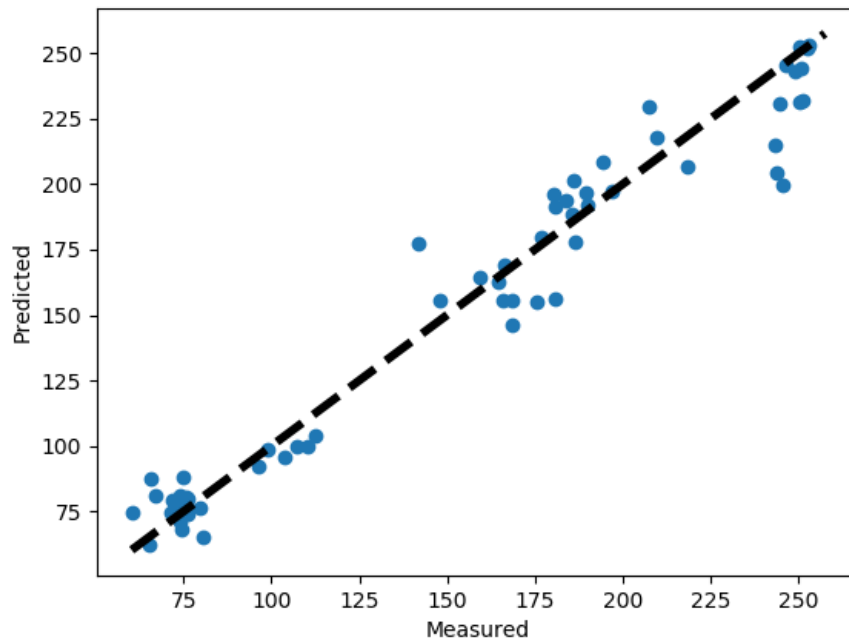


Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση της κάρτας γραφικών, ορίζουμε τις μεταβλητές γ και C στη τιμή 1 και λαμβάνουμε test score 0.94 (σχήμα 4.149).

Σχήμα 4.149: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (SVM)



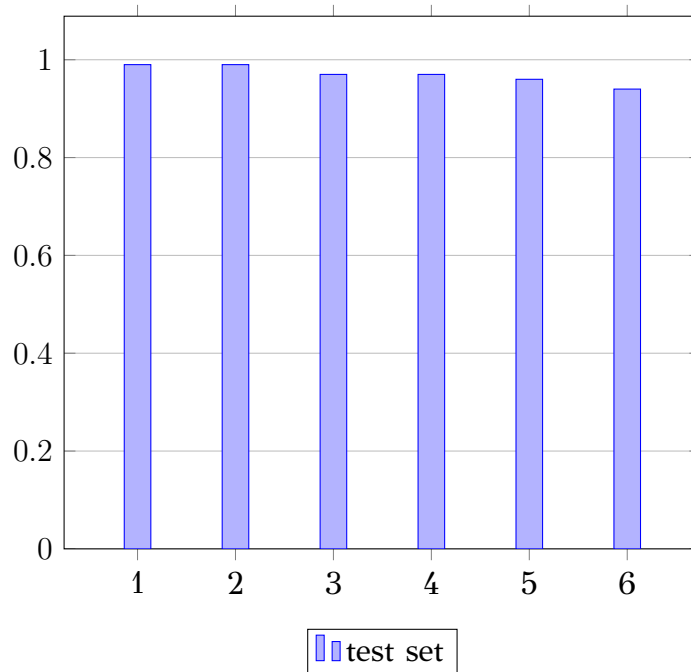
Σχήμα 4.150: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 εισόδους (SVM)



Τα σχήματα 4.140, 4.142, 4.144, 4.146, 4.148 και 4.150 απεικονίζουν διαγράμματα πραγματικών με προβλεπόμενων τιμών για κάθε μία από τις παραπάνω περιπτώσεις. Για κάθε σημείο, στον άξονα X αναπαριστάται η πραγματική τιμή του, ενώ στον άξονα Y η προβλεπόμενη τιμή. Όσο πιο κοντά είναι κάποιο σημείο στη γραμμή (μοντέλο), τόσο πιο σωστά πραγματοποιήθηκε η πρόβλεψη.

Στο παρακάτω σχήμα συνοψίζονται συγκεντρωτικά τα scores των train και test

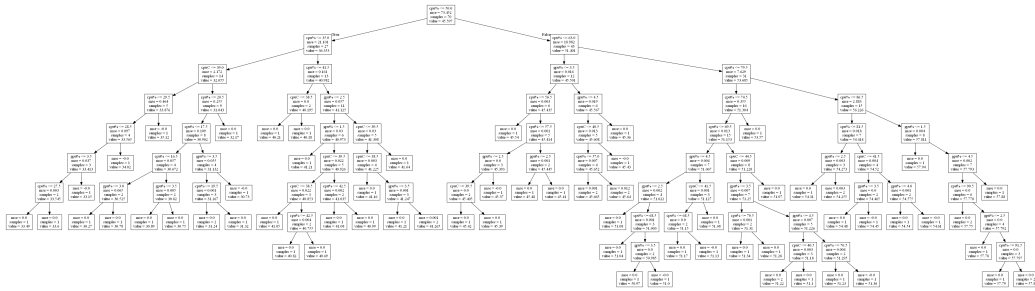
σετ για όλες τις παραπάνω περιπτώσεις. Παρατηρούμε ότι η πρόβλεψη κατανάλωσης ενέργειας της κάρτας γραφικών είχε μεγαλύτερη επιτυχία χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους. Η πρόβλεψη κατανάλωσης ενέργειας του επεξεργαστή ήταν ήταν το ίδιο αποτελεσματική χρησιμοποιώντας τρία και δύο χαρακτηριστικά ως εισόδους. Η πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών είχε το ίδιο αποτέλεσμα χρησιμοποιώντας τρία και δύο χαρακτηριστικά ως εισόδους.



Δέντρα Απόφασης

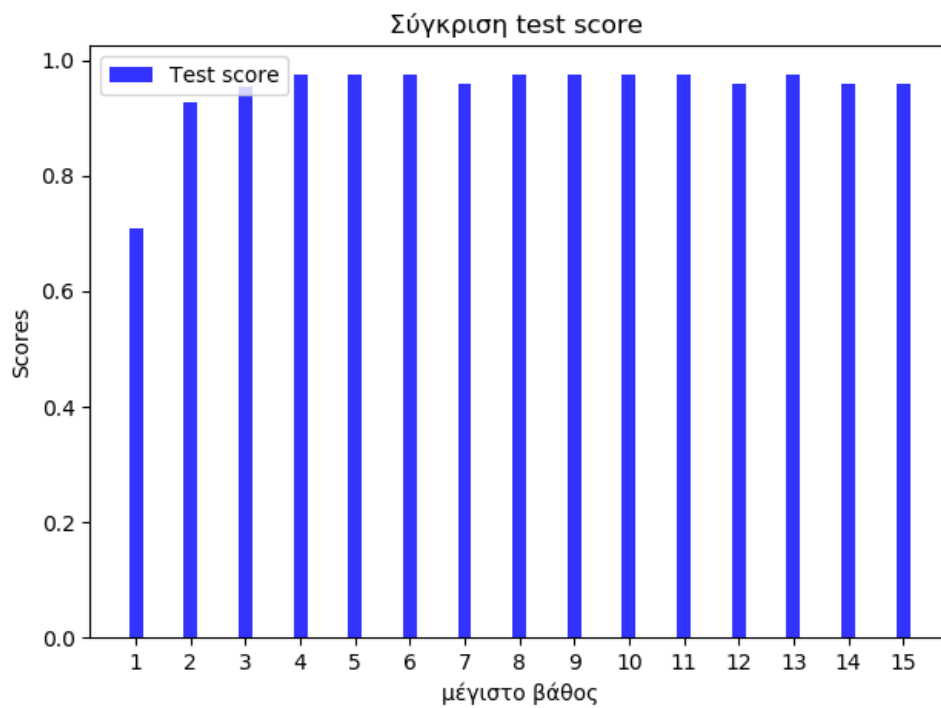
Ο αλγόριθμος περιέχει μια επανάληψη (από 1 μέχρι 15 με βήμα 1), η οποία προσδιορίζει το βάθος το οποίο θα φτάσει το δέντρο απόφασης (η μεταβλητή *max_depth* λαμβάνει την τιμή της επανάληψης), για την καλύτερη διεξαγωγή συμπερασμάτων. Στο πρώτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στον επεξεργαστή. Χρησιμοποιώντας τρία χαρακτηριστικά ως εισόδους, το δέντρο απόφασης φαίνεται στο σχήμα 4.151.

Σχήμα 4.151: Δέντρο απόφασης για τη πρόβλεψη κατανάλωσης ενέργειας στον επεξεργαστή με 3 χαρακτηριστικά (DT)

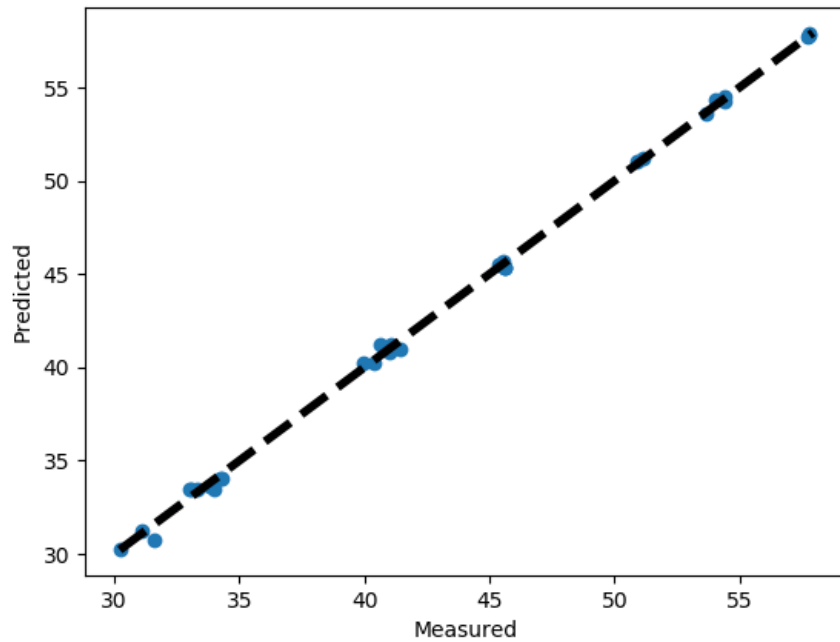


Όπως γίνεται αντιληπτό στο σχήμα 4.152, καθώς αυξάνεται το βάθος του δέντρου απόφασης το test score αυξομειώνεται. Για βάθος δέντρου 13 το test score έχει τη βέλτιστη τιμή του (0.97).

Σχήμα 4.152: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (DT)

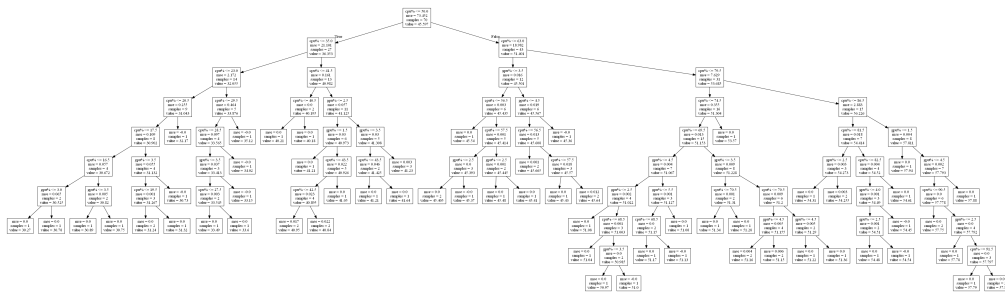


Σχήμα 4.153: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 3 εισόδους (DT)



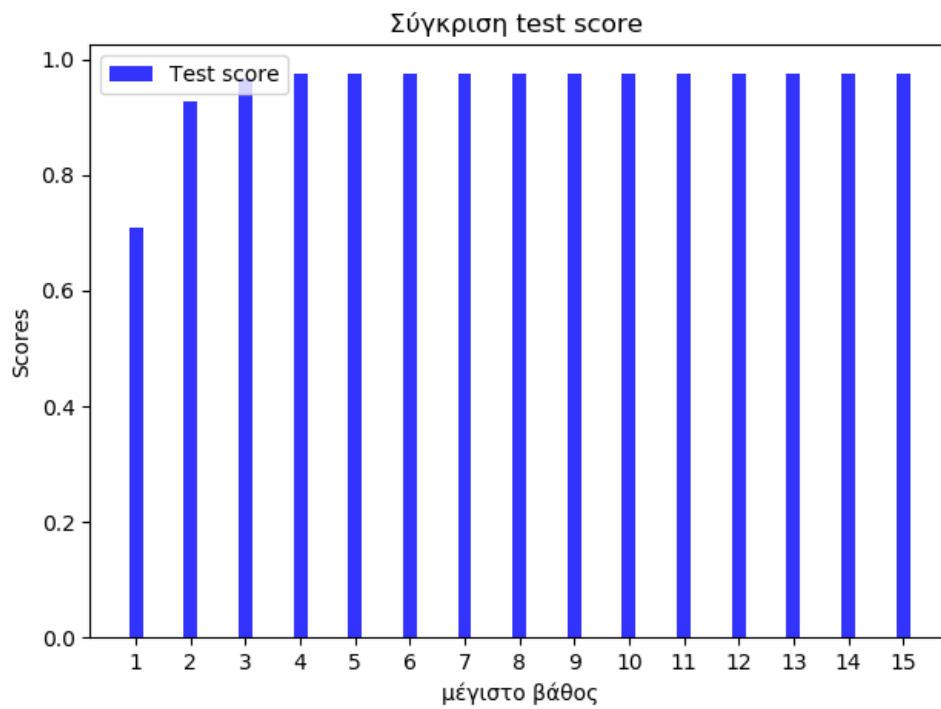
Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση του επεξεργαστή, το δέντρο απόφασης για τη πρόβλεψη κατανάλωσης ενέργειας στον επεξεργαστή φαίνεται στο σχήμα 4.154.

Σχήμα 4.154: Δέντρο απόφασης για τη πρόβλεψη κατανάλωσης ενέργειας στον επεξεργαστή με 2 χαρακτηριστικά (DT)

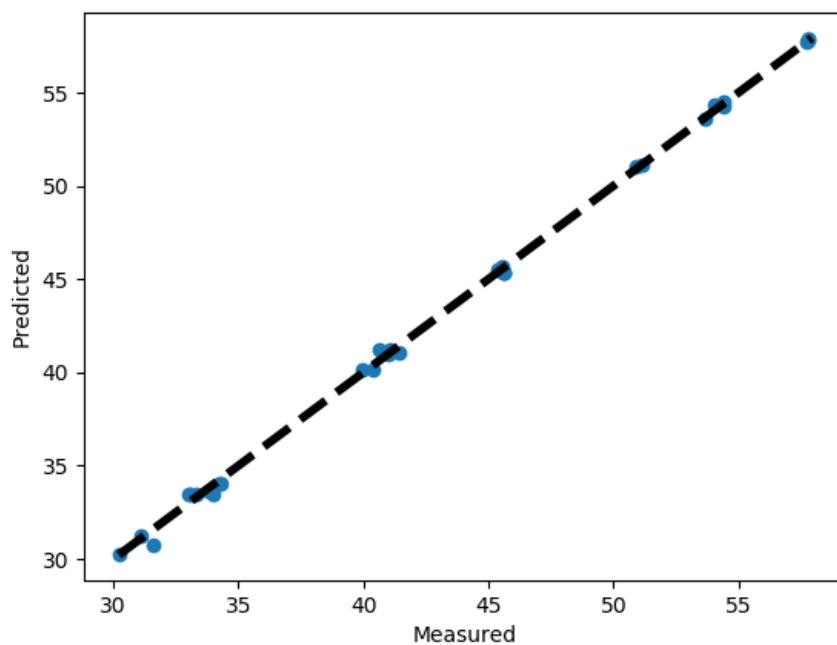


Όπως φαίνεται και στο σχήμα 4.155, καθώς αυξάνεται το βάθος του δέντρου απόφασης το test score αυξομειώνεται. Για βάθος δέντρου 11 το test score έχει τη βέλτιστη τιμή του (0.97).

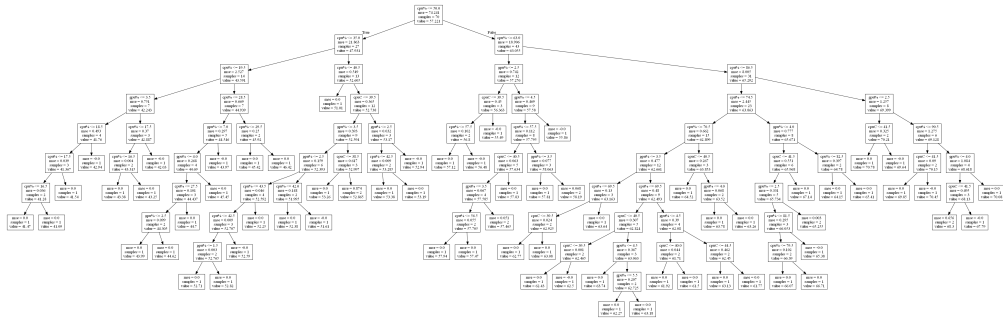
Σχήμα 4.155: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (DT)



Σχήμα 4.156: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (DT)

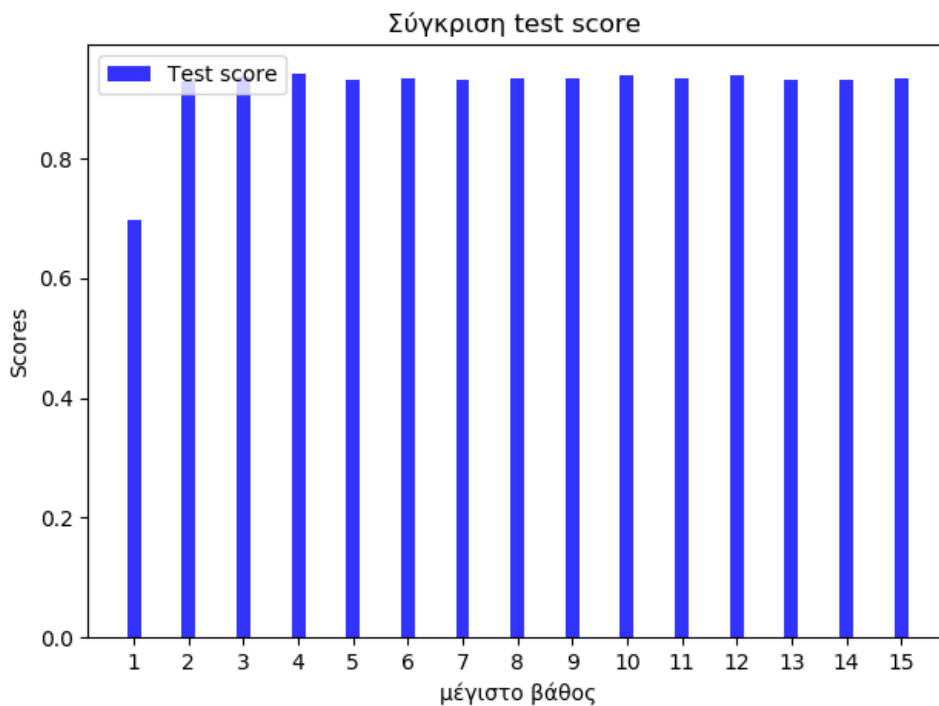


Σχήμα 4.157: Δέντρο απόφασης για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (DT)

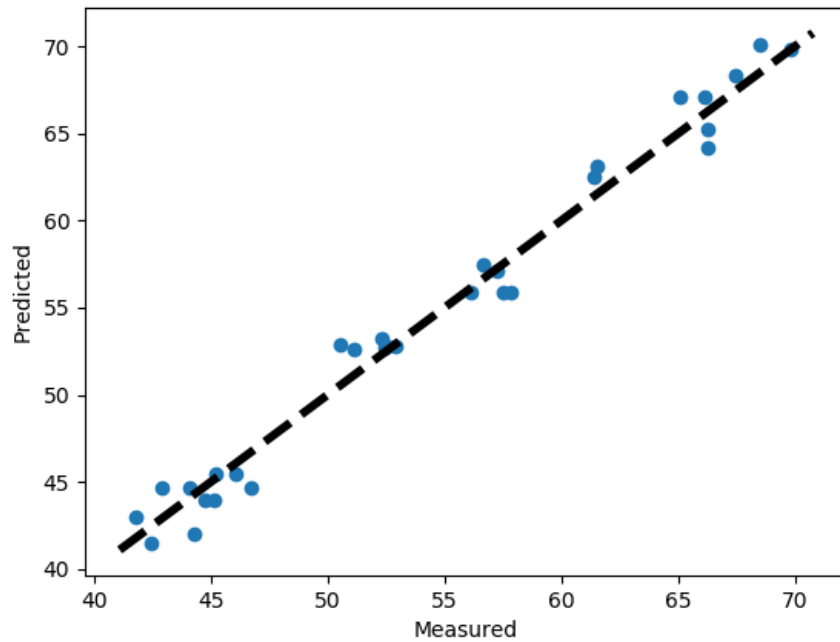


Στο δεύτερο πείραμα κάνουμε πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 3 χαρακτηριστικά ως εισόδους, το δέντρο απόφασης δίνεται από το σχήμα 4.157. Όπως είναι εμφανές στο σχήμα 4.158, καθώς αυξάνεται το βάθος του δέντρου απόφασης, το test score αυξομειώνεται. Για βάθος δέντρου 4 το test score έχει τη βέλτιστη τιμή του (0.94).

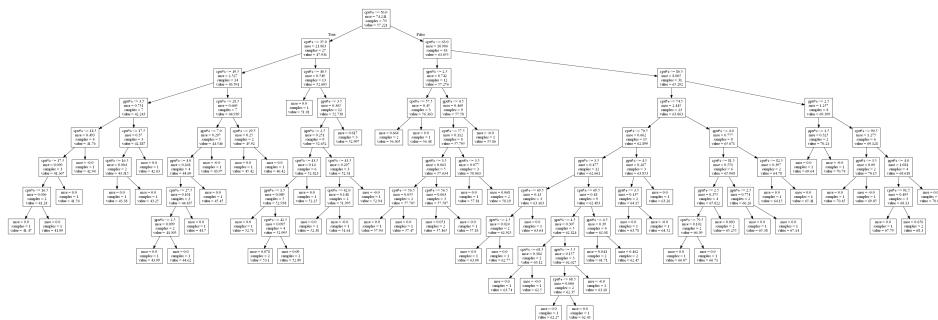
Σχήμα 4.158: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (DT)



Σχήμα 4.159: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (DT)

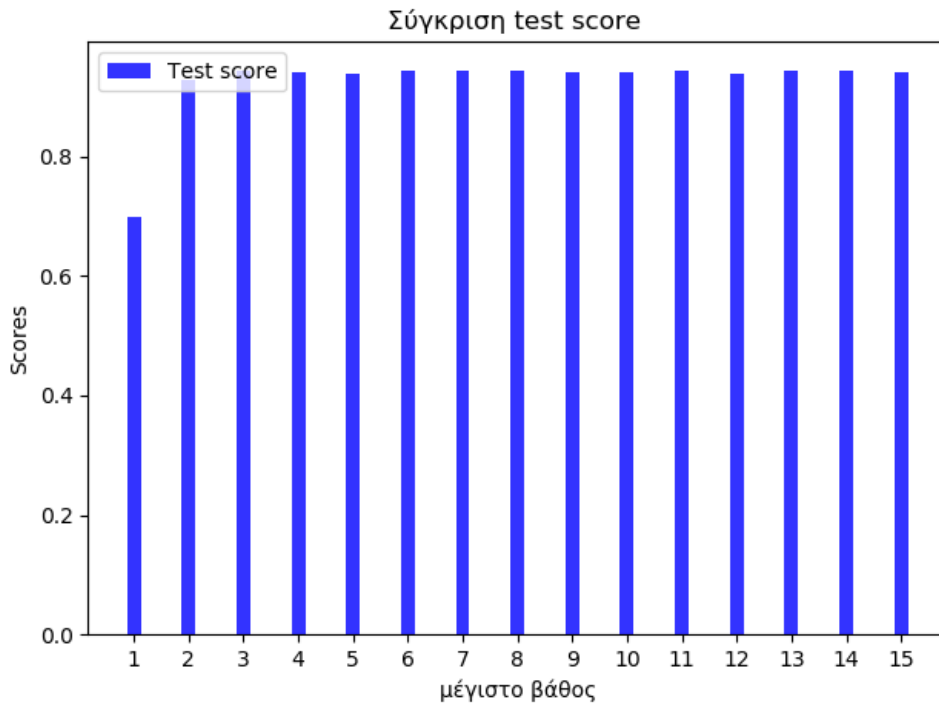


Σχήμα 4.160: Δέντρο απόφασης για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (DT)

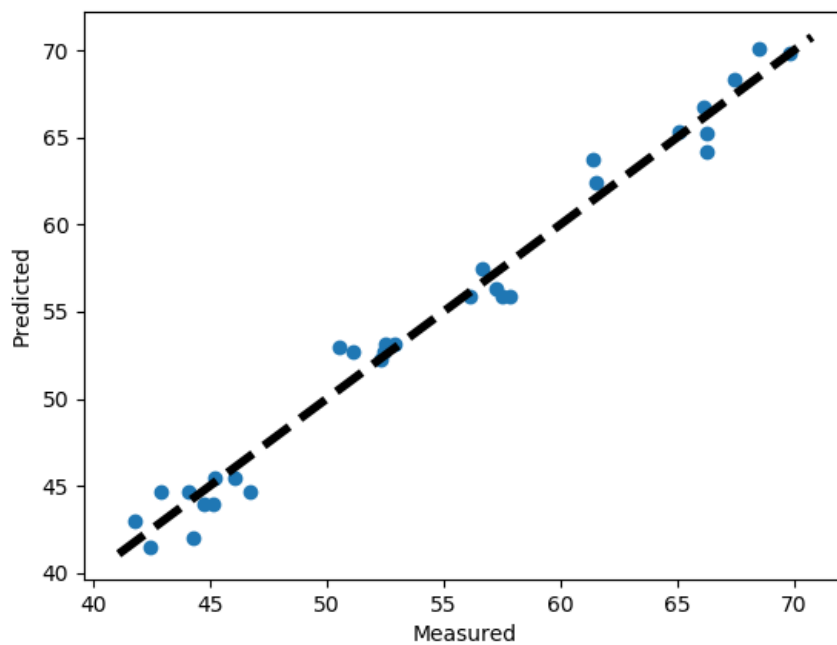


Για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 2 χαρακτηριστικά ως εισόδους, το δέντρο απόφασης δίνεται από το σχήμα 4.160. Όπως παρατηρείται και στο σχήμα 4.161, καθώς αυξάνεται το βάθος του δέντρου απόφασης, το test score αυξομειώνονται. Η βέλτιστη τιμή του test score βρίσκεται σε βάθος 3 με τιμή 0.94.

Σχήμα 4.161: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (DT)



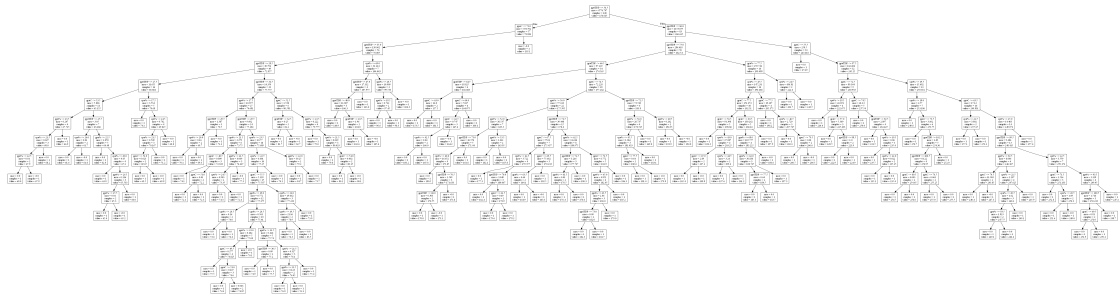
Σχήμα 4.162: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (DT)



Στο τρίτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στην κάρτα γραφικών. Χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους, το δέντρο απόφασης για τη πρόβλεψη κατανάλωσης ενέργειας στον επεξεργαστή φαίνεται στο

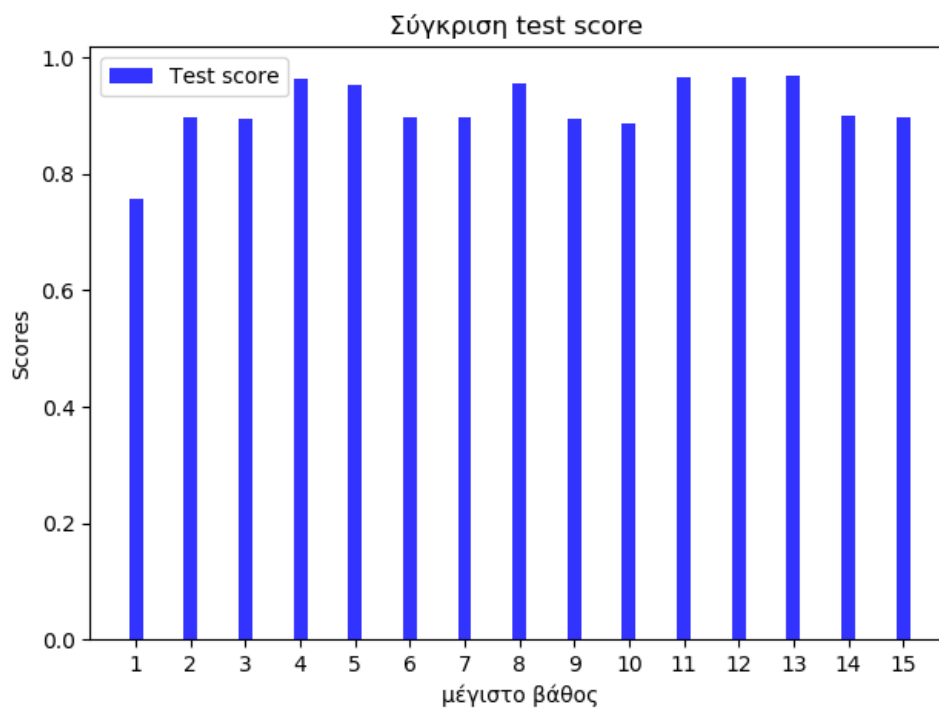
σχήμα 4.163.

Σχήμα 4.163: Δέντρο απόφασης για τη πρόβλεψη κατανάλωσης ενέργειας στην κάρτα γραφικών με 4 χαρακτηριστικά (DT)

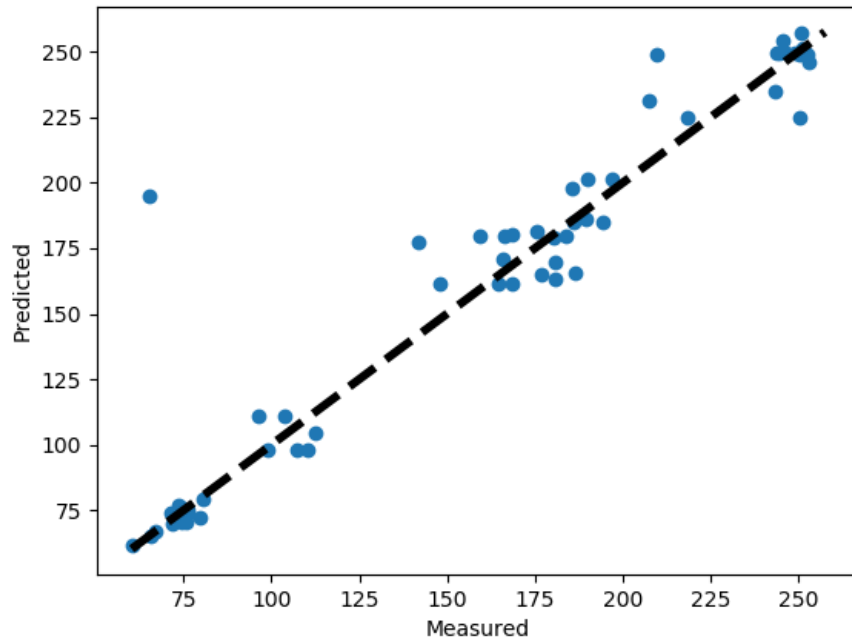


Όπως φαίνεται και στο σχήμα 4.164, καθώς αυξάνεται το βάθος του δέντρου απόφασης το test score αυξομειώνεται. Για βάθος δέντρου 13 το test score έχει τη βέλτιστη τιμή του (0.97).

Σχήμα 4.164: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (DT)

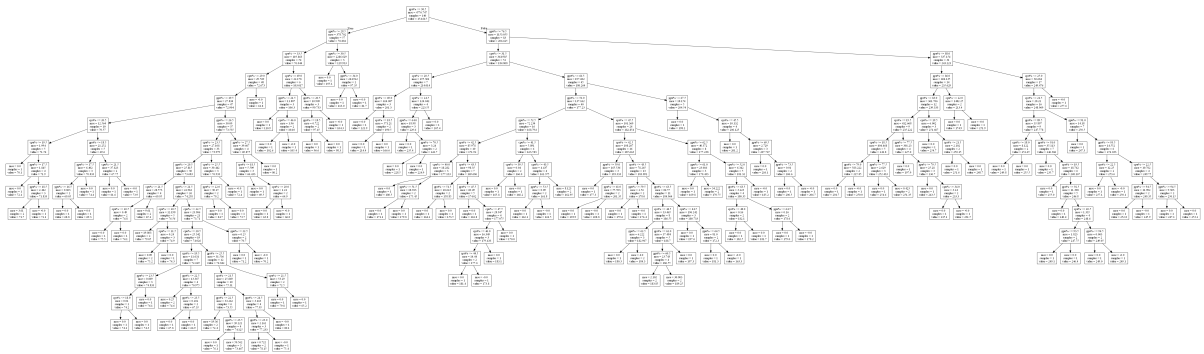


Σχήμα 4.165: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 4 εισόδους (DT)



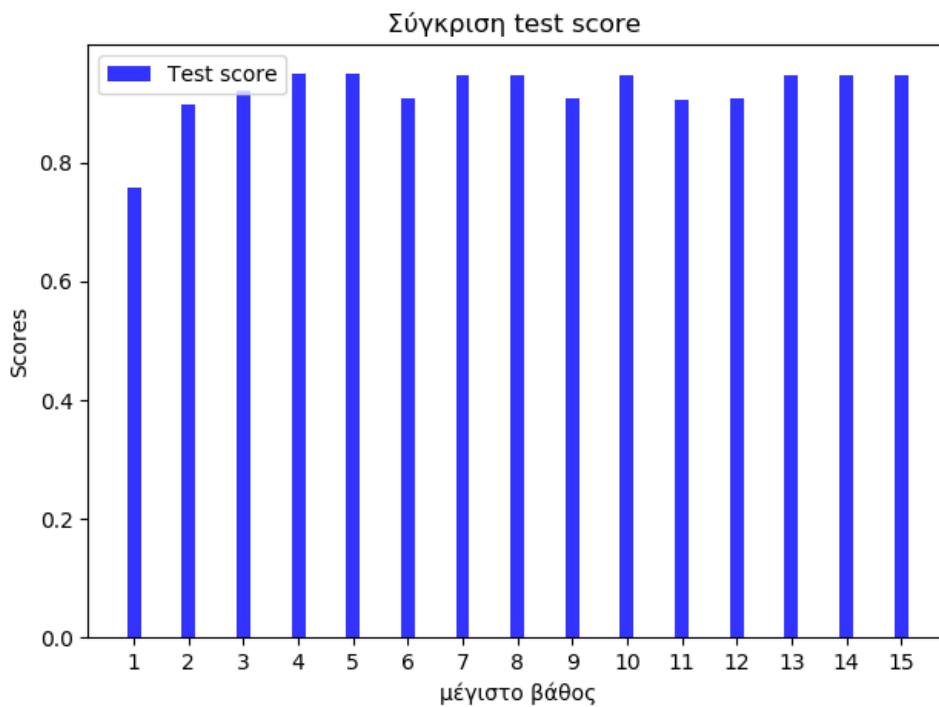
Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση της κάρτας γραφικών, το δέντρο απόφασης φαίνεται στο σχήμα 4.166.

Σχήμα 4.166: Δέντρο απόφασης για τη πρόβλεψη κατανάλωσης ενέργειας στην κάρτα γραφικών με 2 χαρακτηριστικά (DT)

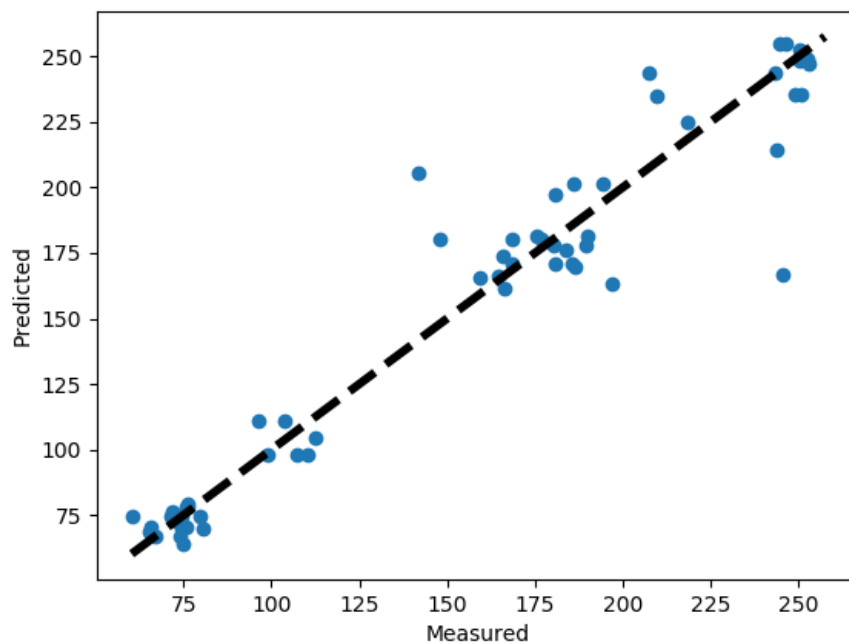


Όπως φαίνεται και στο σχήμα 4.167, καθώς αυξάνεται το βάθος του δέντρου απόφασης το test score αυξομειώνεται. Για βάθος δέντρου 4 το test score έχει τη βέλτιστη τιμή του (0.95).

Σχήμα 4.167: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (DT)



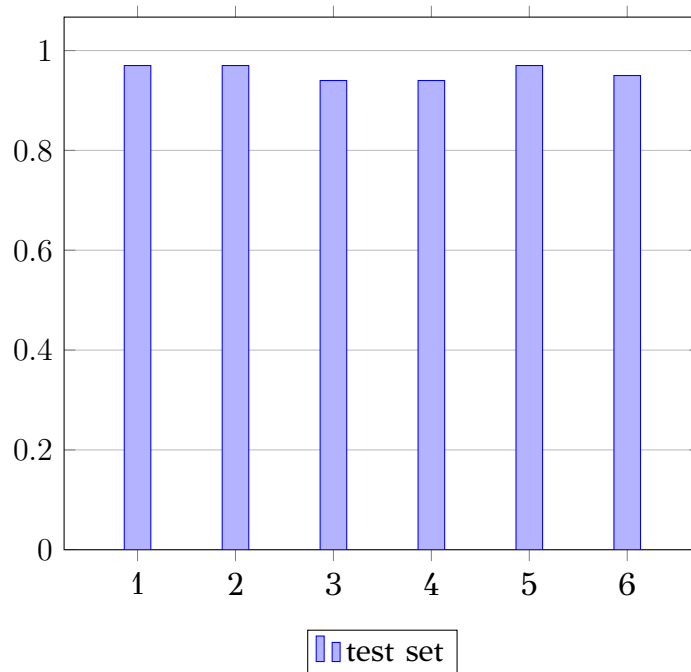
Σχήμα 4.168: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 2 εισόδους (DT)



Τα σχήματα 4.153, 4.156, 4.159, 4.162, 4.165 και 4.168 απεικονίζουν διαγράμματα πραγματικών με προβλεπόμενων τιμών για κάθε μία από τις παραπάνω περιπτώσεις. Για κάθε σημείο, στον άξονα X αναπαριστάται η πραγματική τιμή του, ενώ

στον άξονα Y η προβλεπόμενη τιμή. Όσο πιο κοντά είναι κάποιο σημείο στη γραμμή (μοντέλο), τόσο πιο σωστά πραγματοποιήθηκε η πρόβλεψη.

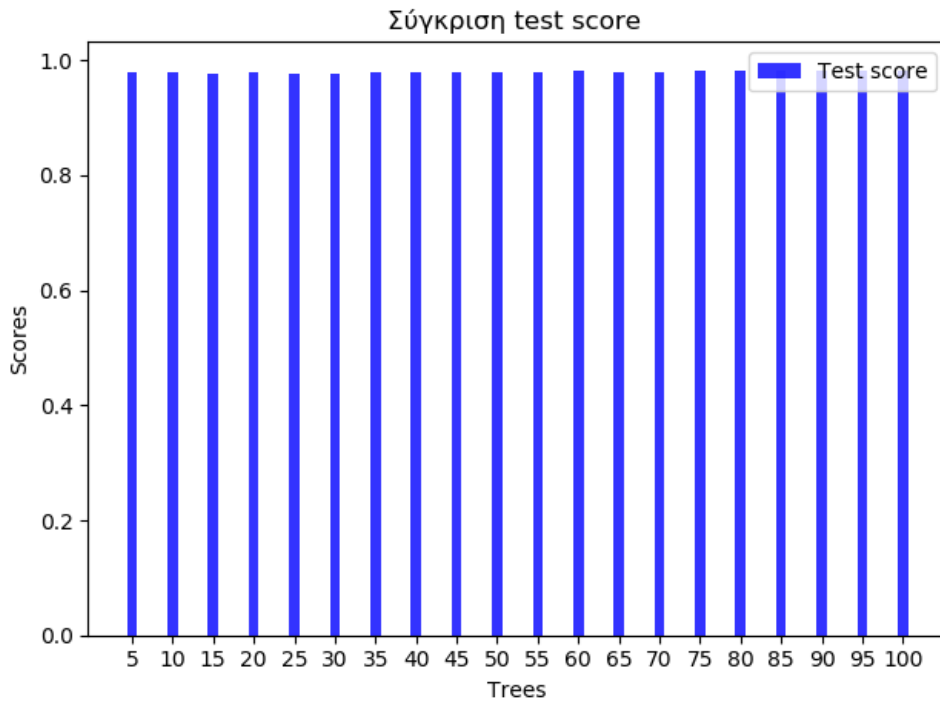
Στο παρακάτω σχήμα συνοψίζονται συγκεντρωτικά τα scores των train και test σετ για όλες τις παραπάνω περιπτώσεις. Παρατηρούμε ότι η πρόβλεψη κατανάλωσης ενέργειας στην κάρτα γραφικών είχε μεγαλύτερη επιτυχία χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους. Η πρόβλεψη κατανάλωσης ενέργειας του επεξεργαστή έχει το ίδιο αποτέλεσμα χρησιμοποιώντας τρία και δύο χαρακτηριστικά ως εισόδους. Η πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών είχε το ίδιο αποτέλεσμα τόσο για δύο όσο και για τρία χαρακτηριστικά ως εισόδους.



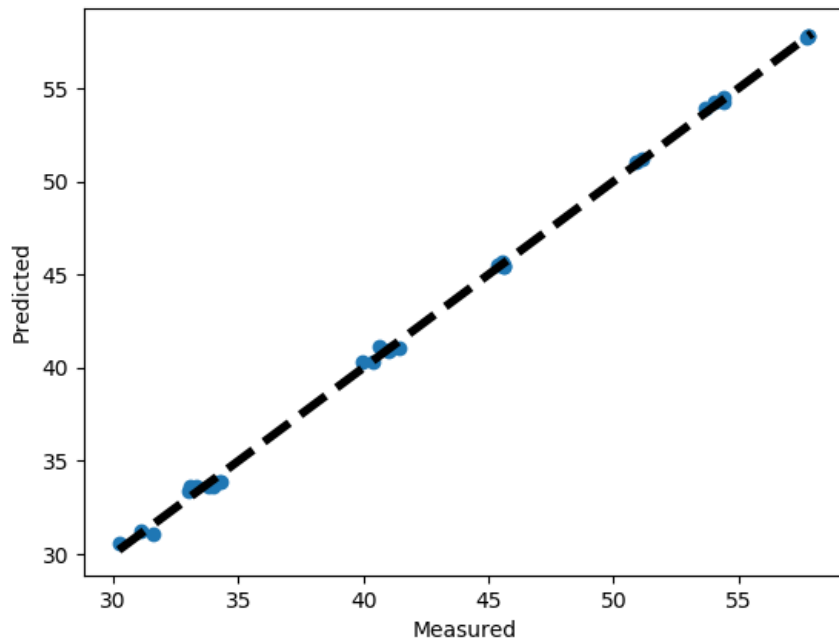
Τυχαία Δάση

Ο αλγόριθμος περιέχει μια επανάληψη (από 5 έως 100 με βήμα 5), όπου η μεταβλητή $n_estimators$ (αριθμός δέντρων) λαμβάνει την τιμή της επανάληψης, για καλύτερη διεξαγωγή συμπερασμάτων. Στο πρώτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στον επεξεργαστή. Χρησιμοποιώντας τρία χαρακτηριστικά ως εισόδους, όπως γίνεται αντιληπτό στο σχήμα 4.169, καθώς αυξάνεται η μεταβλητή $n_estimators$, υπάρχει μικρή αυξομείωση στο test score, το οποίο παίρνει τη μέγιστη τιμή του για αριθμό δέντρων 95. Για τη βέλτιστη τιμή της μεταβλητής $n_estimators$, το test score είναι 0.98.

Σχήμα 4.169: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (RF)



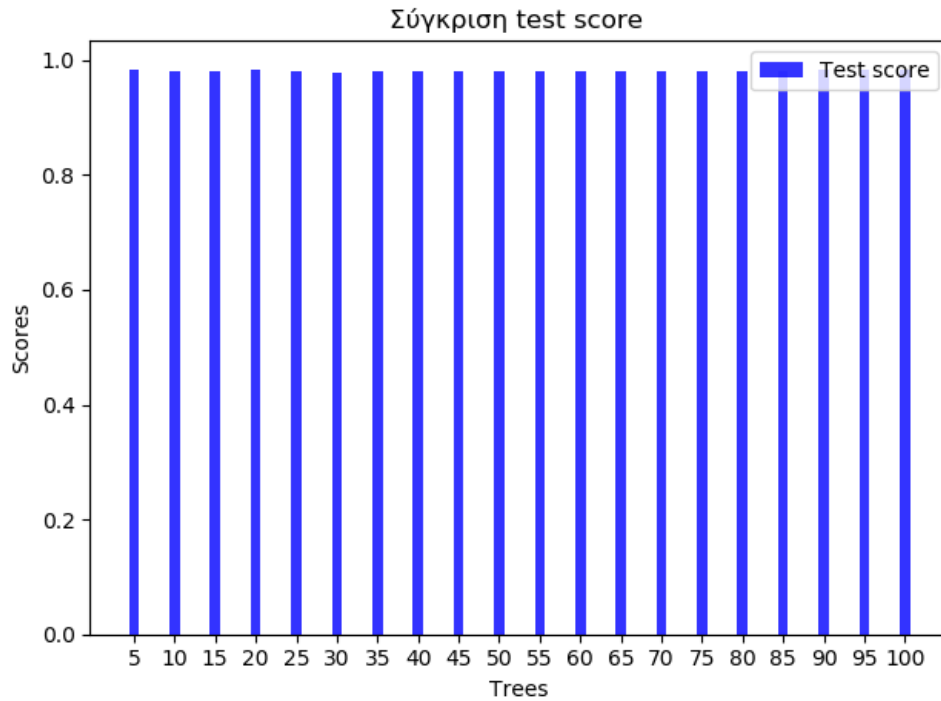
Σχήμα 4.170: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 3 εισόδους (RF)



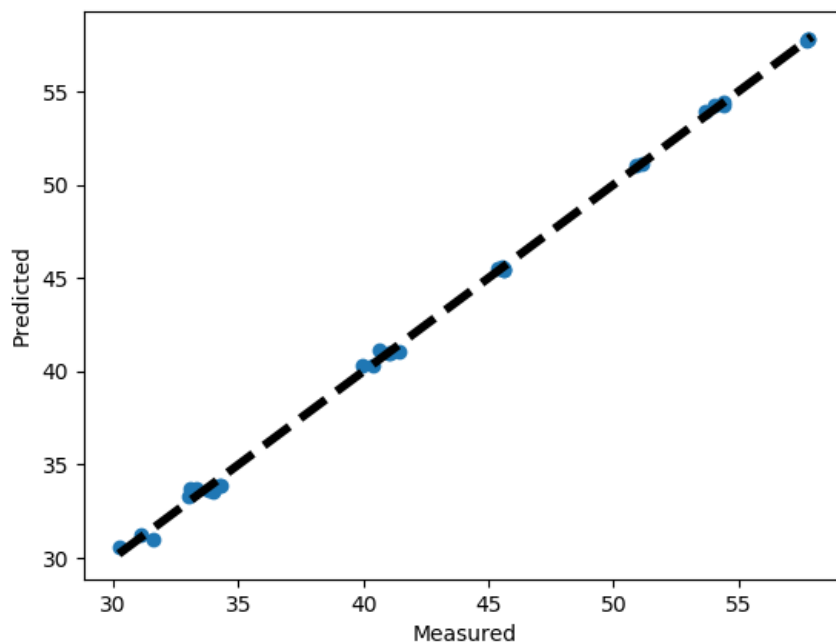
Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση του επεξεργαστή, όπως φαίνεται στο σχήμα 4.171, καθώς αυξάνεται η μεταβλητή $n_estimators$, υπάρχει μικρή αυξομείωση στο test score, το οποίο παίρνει

τη μέγιστη τιμή του για αριθμό δέντρων 5. Για τη βέλτιστη τιμή της μεταβλητής $n_estimators$, το test score είναι 0.98.

Σχήμα 4.171: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (RF)



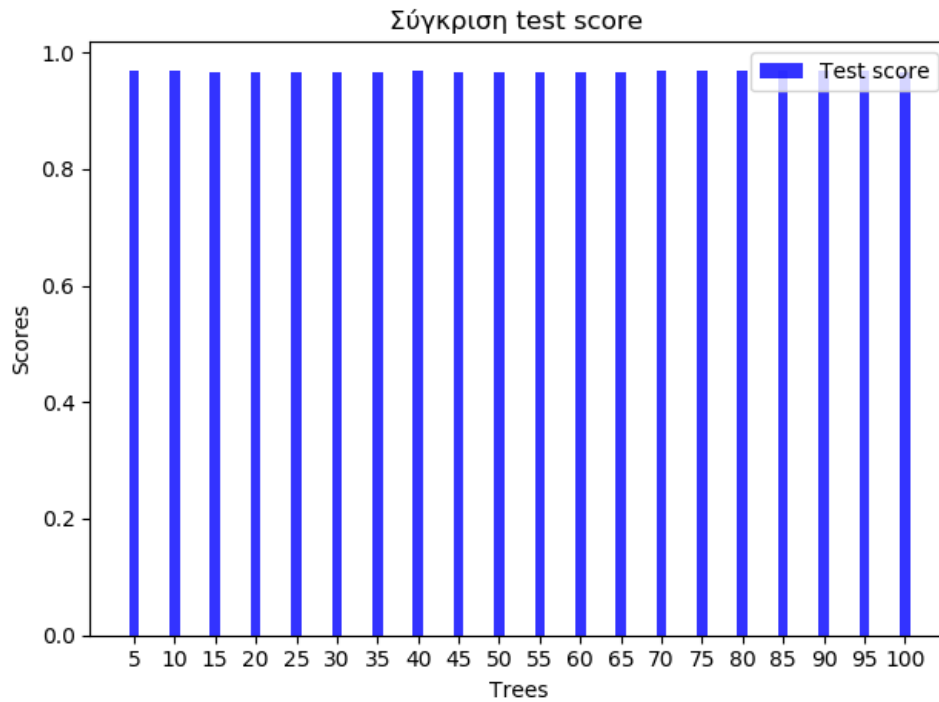
Σχήμα 4.172: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (RF)



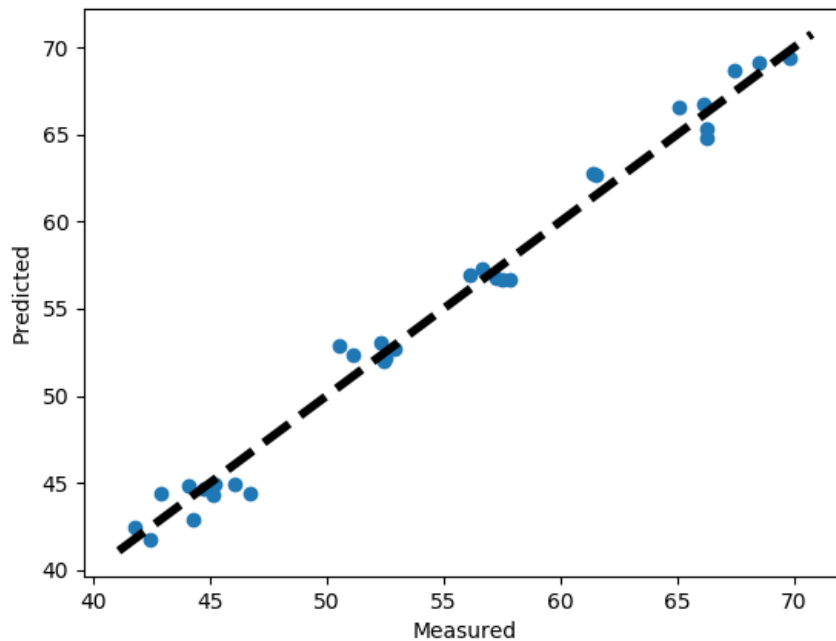
Στο δεύτερο πείραμα κάνουμε πρόβλεψη του αθροίσματος κατανάλωσης ενέρ-

γειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 3 χαρακτηριστικά ως εισόδους. Όπως παρατηρείται στο σχήμα 4.173, καθώς αυξάνεται η μεταβλητή $n_estimators$ υπάρχει αυξομείωση στο test score, το οποίο παίρνει τη μέγιστη τιμή του για αριθμό δέντρων 5. Για τη βέλτιστη τιμή της μεταβλητής $n_estimators$, το test score είναι 0.96.

Σχήμα 4.173: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (RF)

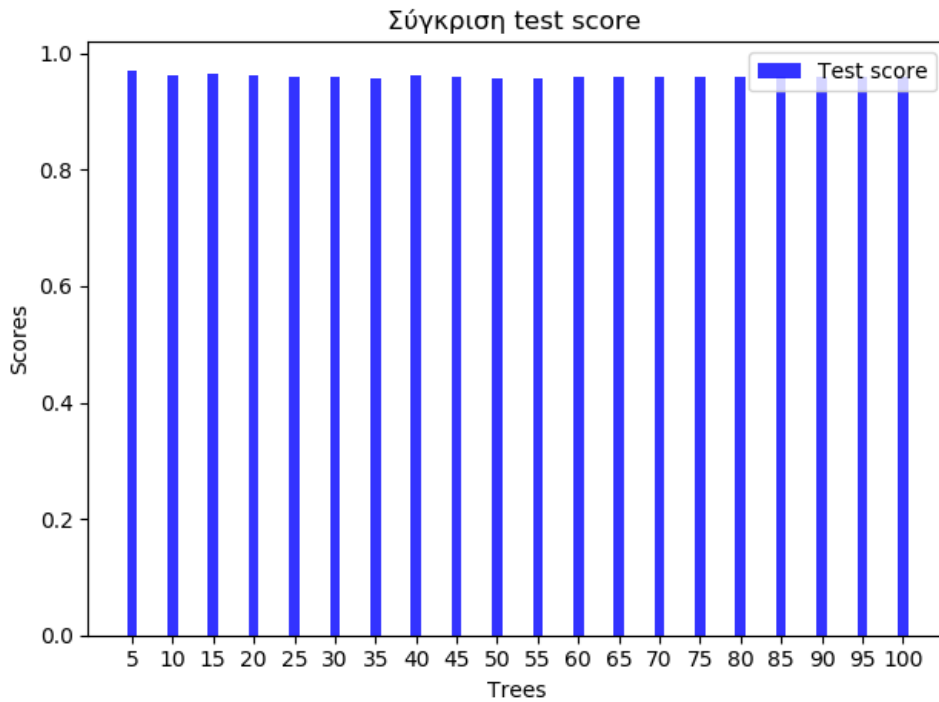


Σχήμα 4.174: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (RF)

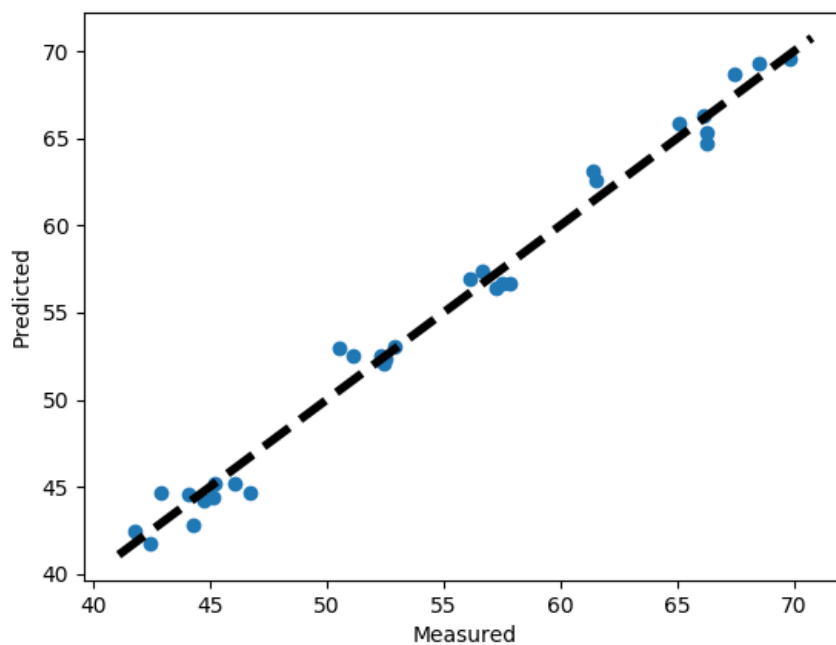


Για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 2 χαρακτηριστικά ως εισόδους, όπως γίνεται αντιληπτό στο σχήμα 4.175, καθώς αυξάνεται η μεταβλητή $n_estimators$ υπάρχει μικρή αυξομείωση στο test score, το οποίο παίρνει τη μέγιστη τιμή του για αριθμό δέντρων 15. Για τη βέλτιστη τιμή της μεταβλητής $n_estimators$, το test score είναι 0.96.

Σχήμα 4.175: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (RF)



Σχήμα 4.176: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (RF)

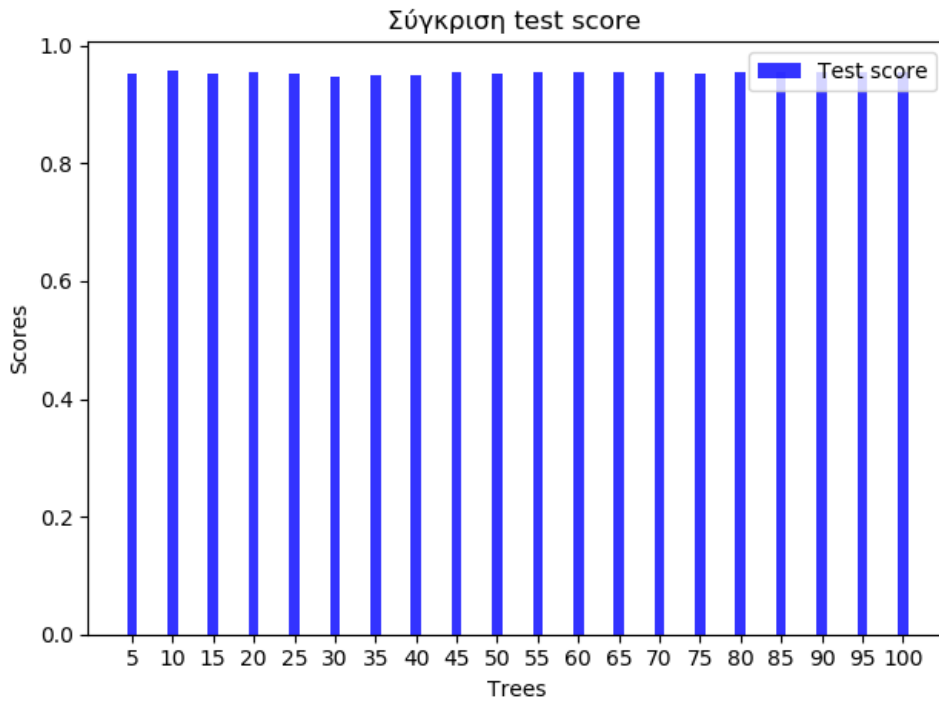


Στο τρίτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στην κάρτα γραφικών. Χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους, όπως φαίνεται στο σχήμα 4.177, καθώς αυξάνεται η μεταβλητή $n_estimators$ υπάρχει μικρή αυξο-

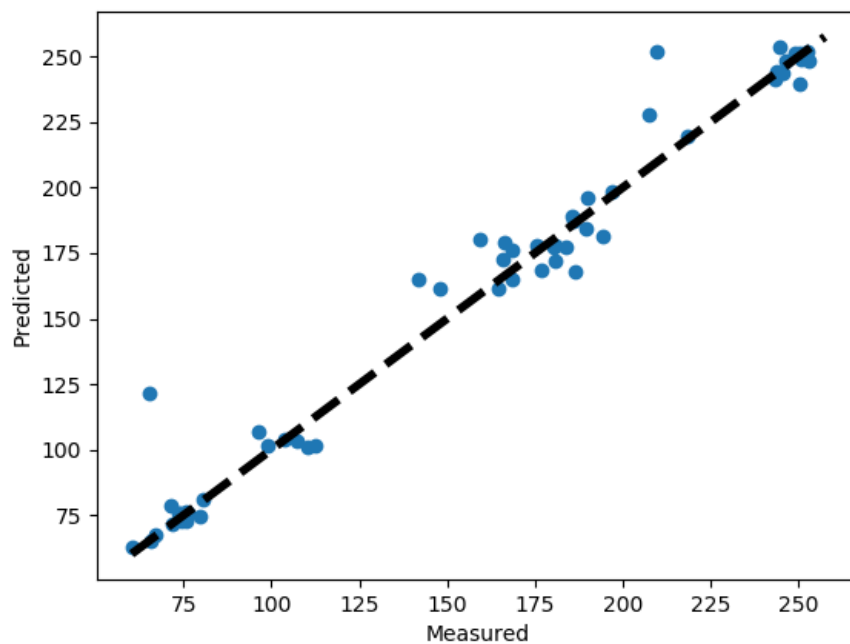
μείωση στο test score, το οποίο παίρνει τη μέγιστη τιμή του για αριθμό δέντρων 10.

Για τη βέλτιστη τιμή της μεταβλητής $n_estimators$, το test score είναι 0.95.

Σχήμα 4.177: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (RF)



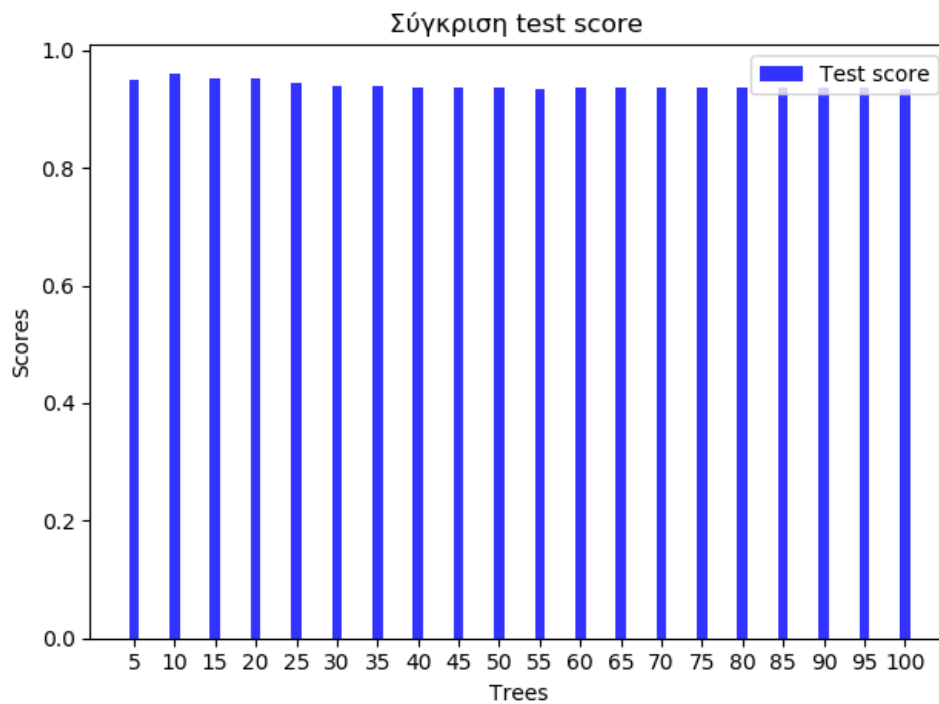
Σχήμα 4.178: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 4 εισόδους (RF)



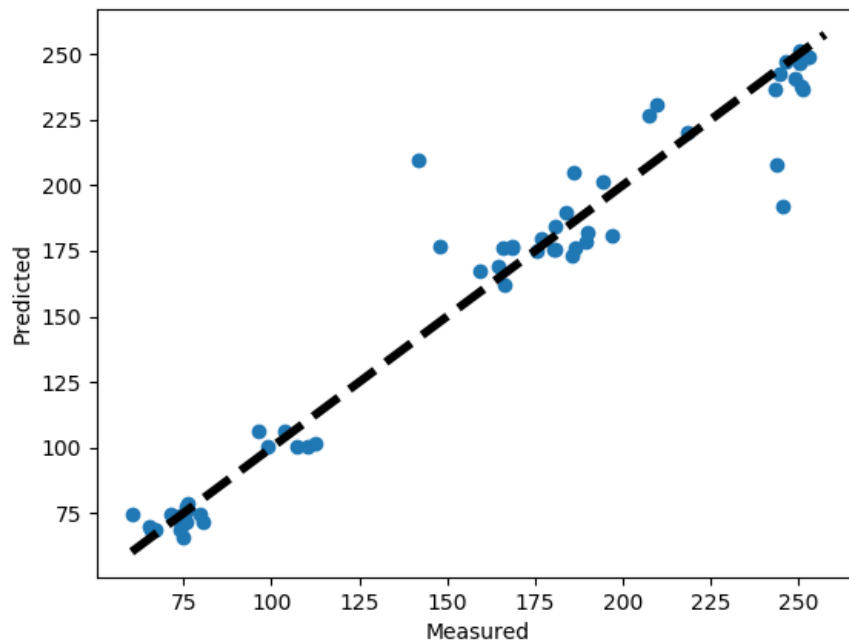
Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κα-

τανάλωση της κάρτας γραφικών, όπως γίνεται αντιληπτό στο σχήμα 4.179, καθώς αυξάνεται η μεταβλητή $n_estimators$ υπάρχει μικρή αυξομείωση στο test score, το οποίο παίρνει τη μέγιστη τιμή του για αριθμό δέντρων 10. Για τη βέλτιστη τιμή της μεταβλητής $n_estimators$, το test score είναι 0.96.

Σχήμα 4.179: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (RF)

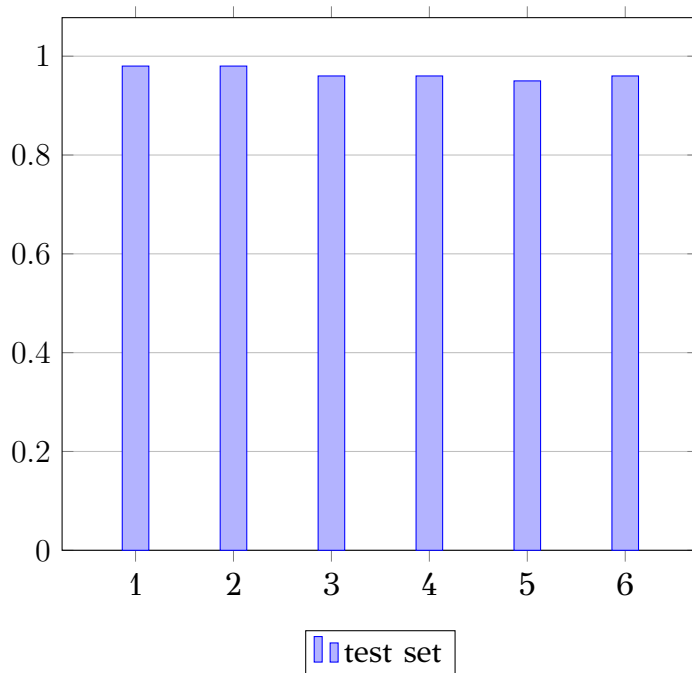


Σχήμα 4.180: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 2 εισόδους (RF)



Τα σχήματα 4.170, 4.172, 4.174, 4.176, 4.178 και 4.180 απεικονίζουν διαγράμματα πραγματικών με προβλεπόμενων τιμών για κάθε μία από τις παραπάνω περιπτώσεις. Για κάθε σημείο, στον άξονα X αναπαριστάται η πραγματική τιμή του, ενώ στον άξονα Y η προβλεπόμενη τιμή. Όσο πιο κοντά είναι κάποιο σημείο στη γραμμή (μοντέλο), τόσο πιο σωστά πραγματοποιήθηκε η πρόβλεψη.

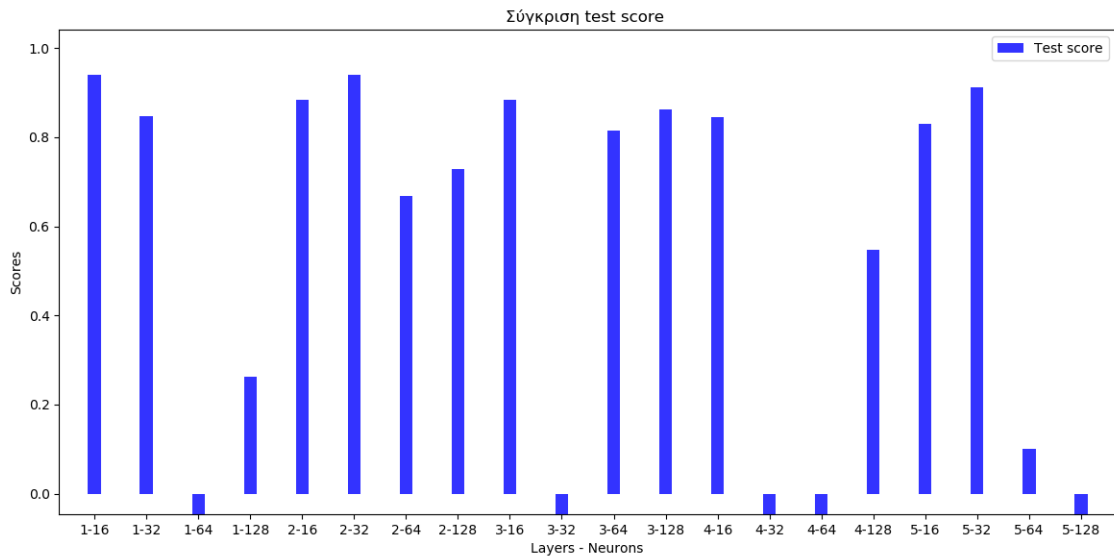
Στο παρακάτω σχήμα συνοφίζονται συγκεντρωτικά τα scores των train και test σετ για όλες τις παραπάνω περιπτώσεις. Παρατηρούμε ότι η πρόβλεψη κατανάλωσης ενέργειας στην κάρτα γραφικών είχε ελάχιστα μεγαλύτερη επιτυχία χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους. Η πρόβλεψη κατανάλωσης ενέργειας του επεξεργαστή έχει το ίδιο αποτέλεσμα χρησιμοποιώντας τρία και δύο χαρακτηριστικά ως εισόδους. Η πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών είχε το ίδιο αποτέλεσμα τόσο για δύο όσο και για τρία χαρακτηριστικά ως εισόδους.



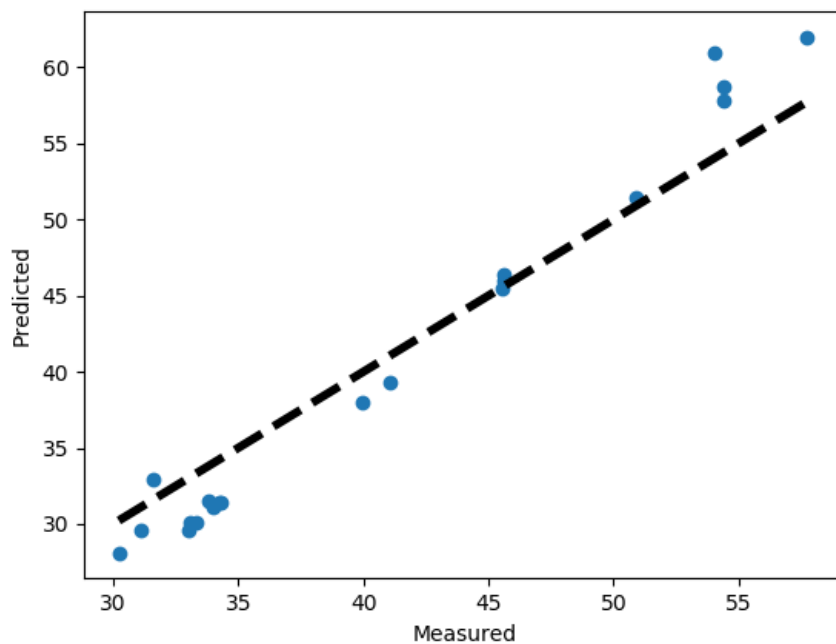
Βαθιά Νευρωνικά Δίκτυα

Ο αλγόριθμος χρησιμοποιεί δύο επαναλήψεις, από 1 μέχρι 5 με βήμα 1 για τον ορισμό των κρυφών επιπέδων και από 4 μέχρι 8 με βήμα 1 για τον ορισμό των νευρώνων (οι νευρώνες ορίζονται ως 2^j), στις οποίες εναλλάσσονται τα κρυφά επίπεδα (1-5) καθώς και οι νευρώνες που περιέχουν (16, 32, 64 και 128). Στο πρώτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στον επεξεργαστή. Χρησιμοποιώντας τρία χαρακτηριστικά ως εισόδους, όπως παρατηρείται και στο σχήμα 4.181, το test score αυξομειώνεται. Για 1 κρυφό επίπεδο με 64 νευρώνες, 3 κρυφά επίπεδα με 32 νευρώνες, 4 κρυφά επίπεδα με 32 και 64 νευρώνες και 5 κρυφά επίπεδα με 128 νευρώνες παρατηρείται αρνητικό test score, ενώ το test score λαμβάνει τη μέγιστη τιμή του (0.94) για 1 κρυφό επίπεδο με 16 νευρώνες.

Σχήμα 4.181: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 3 χαρακτηριστικά (NN)



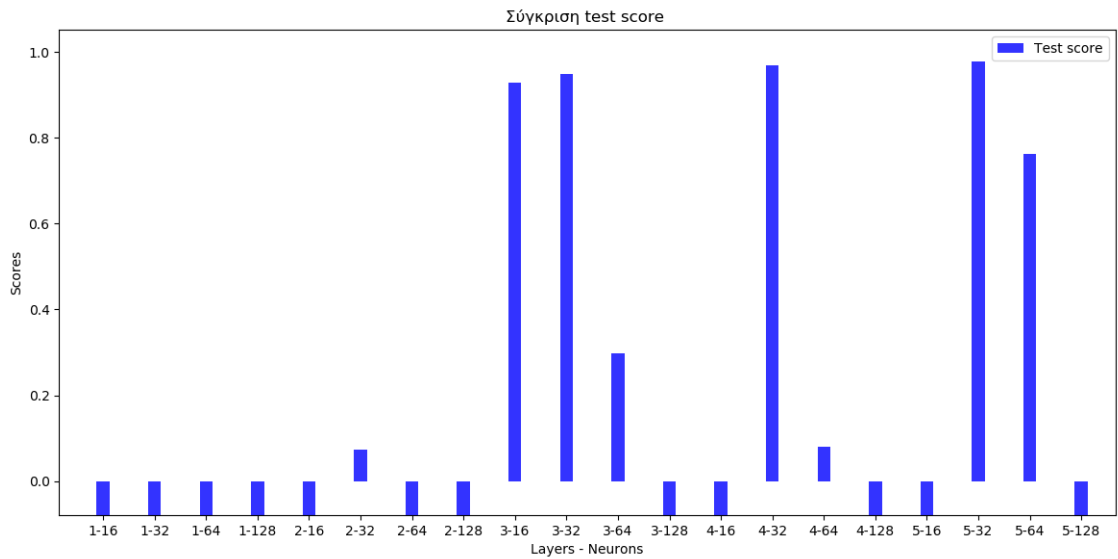
Σχήμα 4.182: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 3 εισόδους (NN)



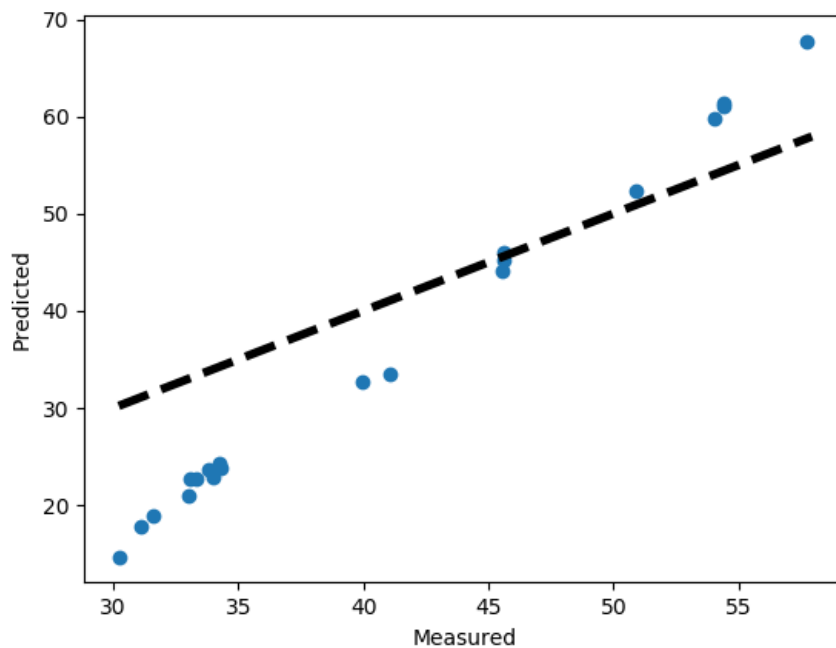
Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση του επεξεργαστή, όπως παρατηρείται και στο σχήμα 4.183, το test score αυξομειώνεται. Για όλο το πρώτο κρυφό επίπεδο, 2 κρυφά επίπεδα με 64 και 128 νευρώνες, 3 κρυφά επίπεδα με 128 νευρώνες, 4 κρυφά επίπεδα με 16, 64 και 128 νευρώνες και με 5 κρυφά επίπεδα με 16 και 128 νευρώνες παρατηρείται αρνητικό test score. Το test score λαμβάνει τη μέγιστη τιμή του (0.97) για 5 κρυφά επίπεδα

με 32 νευρώνες.

Σχήμα 4.183: Σύγκριση test score για την πρόβλεψη κατανάλωσης του επεξεργαστή με 2 χαρακτηριστικά (NN)



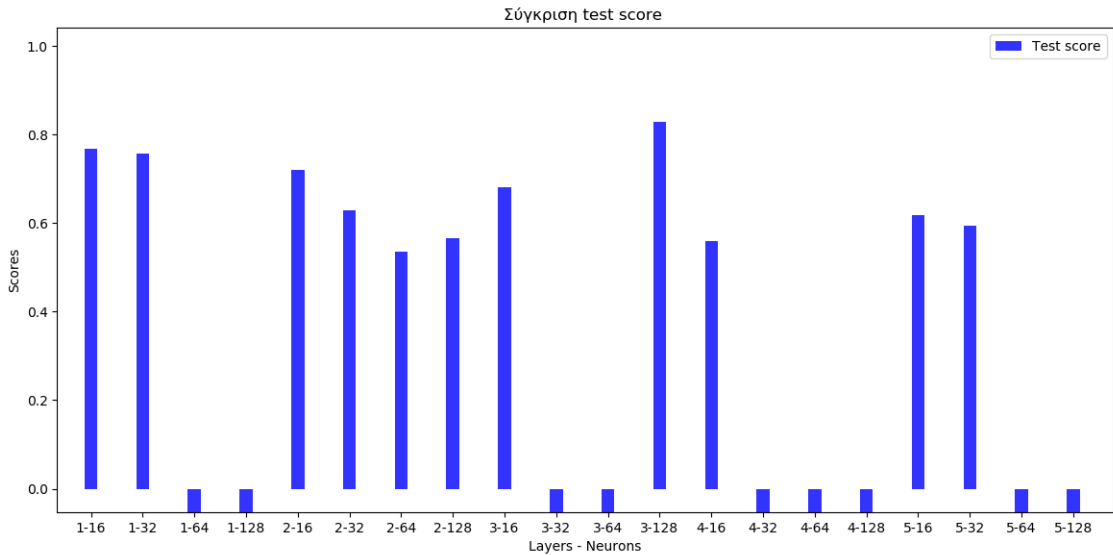
Σχήμα 4.184: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στον επεξεργαστή με 2 εισόδους (NN)



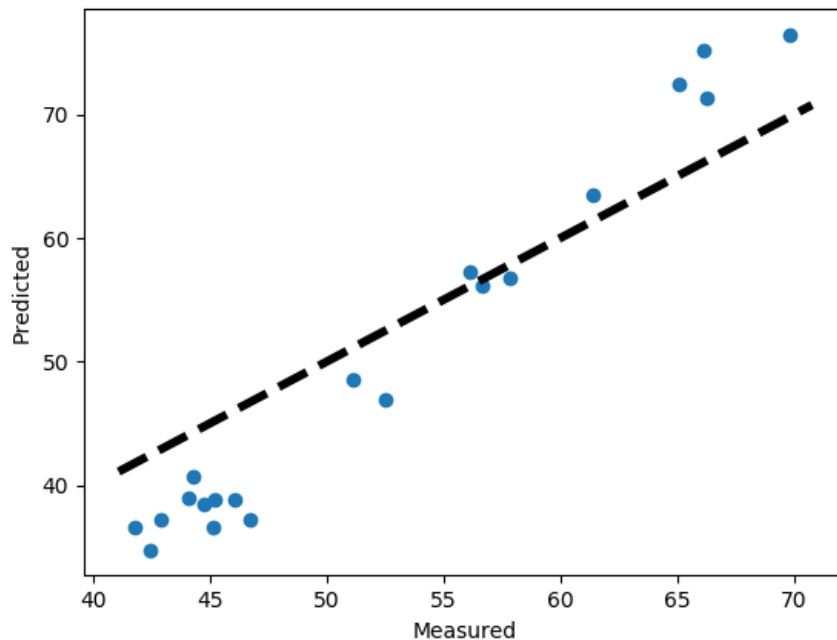
Στο δεύτερο πείραμα κάνουμε πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 3 χαρακτηριστικά ως εισόδους, όπως παρατηρείται και στο σχήμα 4.185, το test score αυξομειώνεται. Για 1 κρυφό επίπεδο με 64 και 128 νευρώνες, 3 κρυφά επίπεδα με 32 και 64

νευρώνες, 4 κρυφά επίπεδα με 32, 64 και 128 νευρώνες και για 5 κρυφά επίπεδα με 64 και 128 νευρώνες παρατηρείται αρνητικό test score. Το test score λαμβάνει τη μέγιστη τιμή του (0.82) για 3 κρυφά επίπεδα με 128 νευρώνες.

Σχήμα 4.185: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 χαρακτηριστικά (NN)



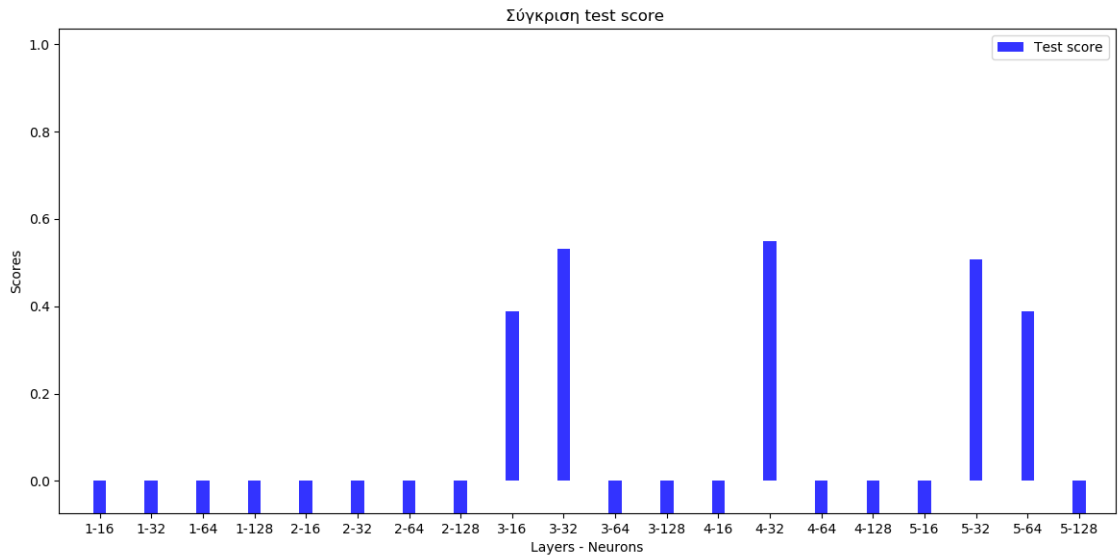
Σχήμα 4.186: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 3 εισόδους (NN)



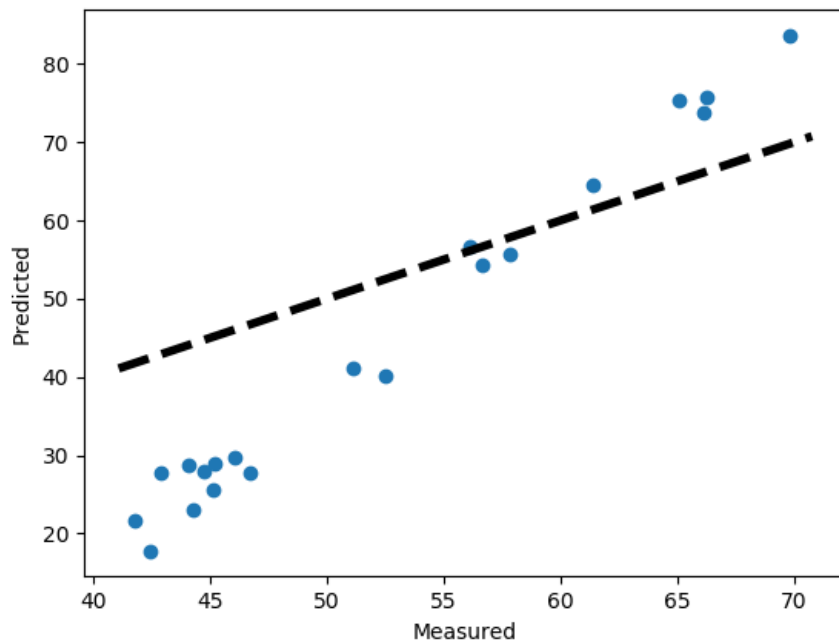
Για τη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας του επεξεργαστή και της κάρτας γραφικών χρησιμοποιώντας 2 χαρακτηριστικά ως εισόδους, όπως

γίνεται αντιληπτό και στο σχήμα 4.187, το test score αυξομειώνεται. Για 1 και 2 κρυφά επίπεδα, 3 κρυφά επίπεδα με 64 και 128 νευρώνες, 4 κρυφά επίπεδα με 16, 64 και 128 νευρώνες και 5 κρυφά επίπεδα με 16, και 128 νευρώνες παρατηρείται αρνητικό test score. Το test score λαμβάνει τη μέγιστη τιμή του (0.54) για 4 κρυφά επίπεδα με 32 νευρώνες.

Σχήμα 4.187: Σύγκριση test score για την πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 χαρακτηριστικά (NN)



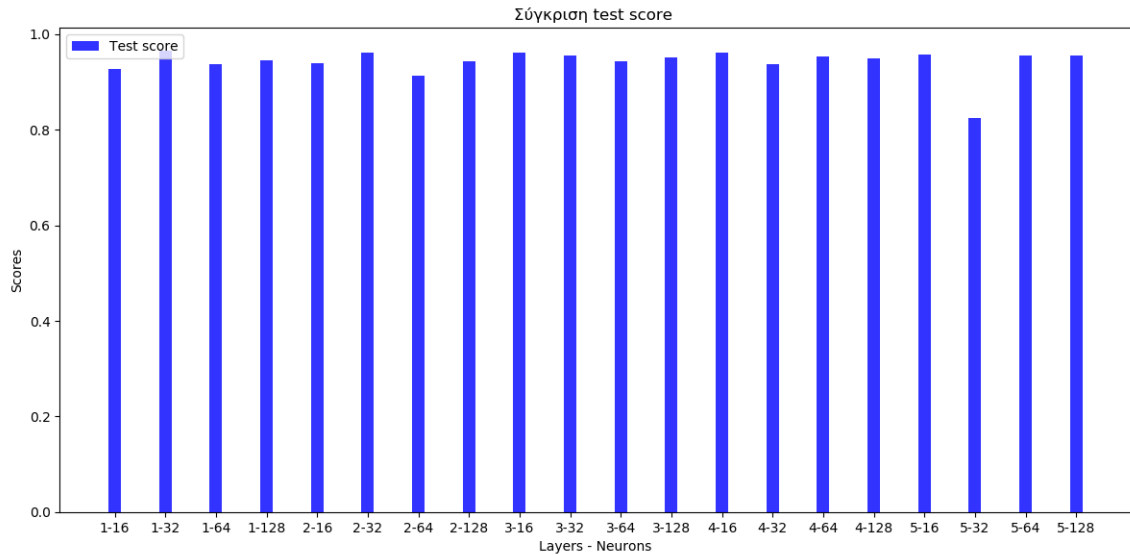
Σχήμα 4.188: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών με 2 εισόδους (NN)



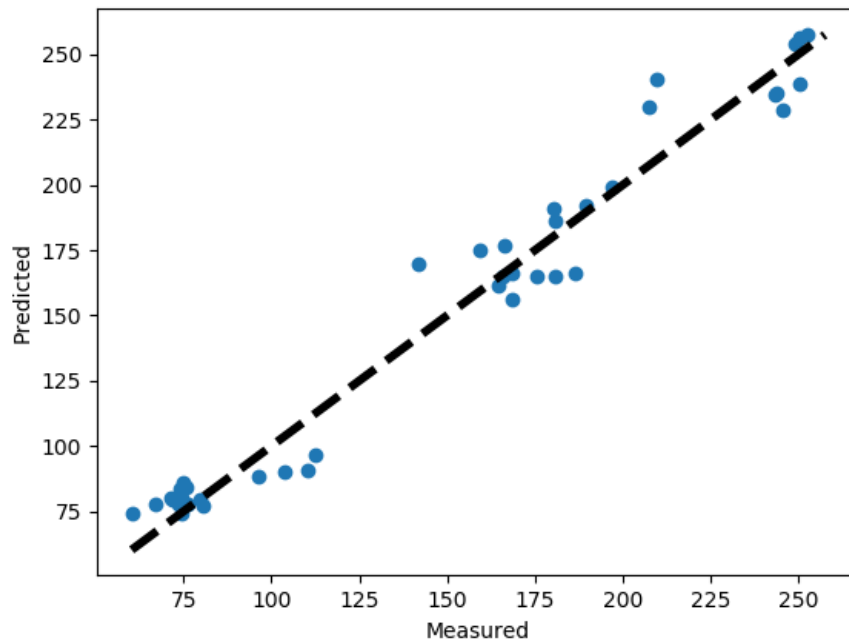
Στο τρίτο πείραμα κάνουμε πρόβλεψη για την κατανάλωση ενέργειας στην κάρτα

γραφικών. Χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους, όπως παρατηρείται και στο σχήμα 4.189, υπάρχει μικρή αυξομείωση στο test score. Το test score λαμβάνει τη μέγιστη τιμή του (0.96) για 1 κρυφό επίπεδο με 32 νευρώνες.

Σχήμα 4.189: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 4 χαρακτηριστικά (NN)



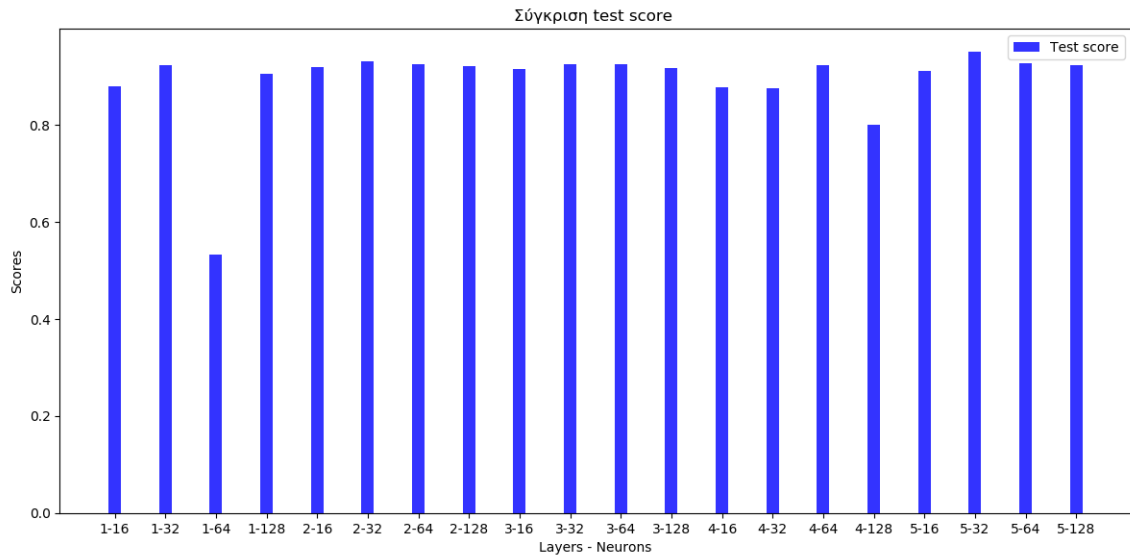
Σχήμα 4.190: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 4 εισόδους (NN)



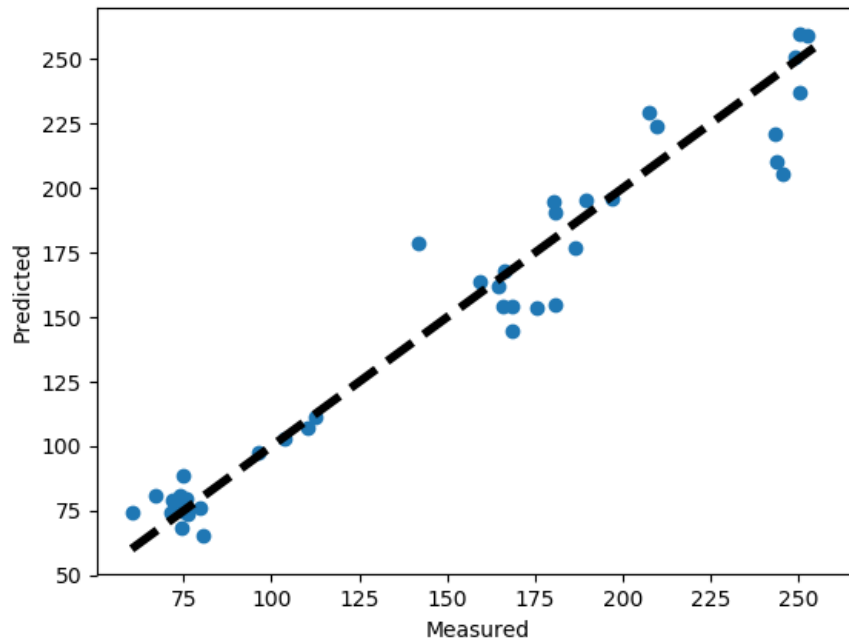
Χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους και προβλέποντας την κατανάλωση της κάρτας γραφικών, όπως γίνεται αντιληπτό στο σχήμα 4.191, υπάρχει

αυξομείωση στο test score. Το test score λαμβάνει τη μέγιστη τιμή του (0.95) για 5 κρυφά επίπεδα με 32 νευρώνες.

Σχήμα 4.191: Σύγκριση test score για την πρόβλεψη κατανάλωσης της κάρτας γραφικών με 2 χαρακτηριστικά (NN)

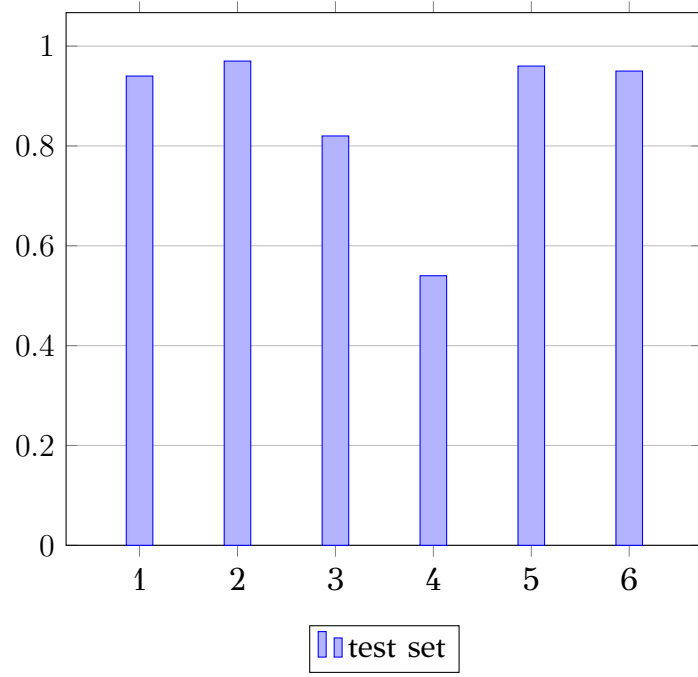


Σχήμα 4.192: Διάγραμμα προβλεπόμενων vs πραγματικών τιμών για την καλύτερη πρόβλεψη στην κάρτα γραφικών με 2 εισόδους (NN)



Τα σχήματα 4.182, 4.184, 4.186, 4.188, 4.190 και 4.192 απεικονίζουν διαγράμματα πραγματικών με προβλεπόμενων τιμών για κάθε μία από τις παραπάνω περιπτώσεις. Για κάθε σημείο, στον άξονα X αναπαριστάται η πραγματική τιμή του, ενώ στον άξονα Y η προβλεπόμενη τιμή. Όσο πιο κοντά είναι κάποιο σημείο στη γραμμή (μοντέλο), τόσο πιο σωστά πραγματοποιήθηκε η πρόβλεψη.

Στο παρακάτω σχήμα συνοψίζονται συγκεντρωτικά τα scores των train και test σετ για όλες τις παραπάνω περιπτώσεις. Παρατηρούμε ότι η πρόβλεψη κατανάλωσης ενέργειας στην κάρτα γραφικών είχε ελάχιστα μεγαλύτερη επιτυχία χρησιμοποιώντας τέσσερα χαρακτηριστικά ως εισόδους. Η πρόβλεψη κατανάλωσης ενέργειας του επεξεργαστή έχει καλύτερο αποτέλεσμα χρησιμοποιώντας δύο χαρακτηριστικά ως εισόδους. Η πρόβλεψη του αθροίσματος κατανάλωσης ενέργειας επεξεργαστή και κάρτας γραφικών έχει μεγαλύτερη επιτυχία για τρία χαρακτηριστικά ως εισόδους.



Κεφάλαιο 5

Συμπεράσματα

Στη συγκεκριμένη εργασία μελετήσαμε την πρόβλεψη κατανάλωσης ενέργειας σε κέντρα δεδομένων, χρησιμοποιώντας αλγορίθμους μηχανικής μάθησης. Μέσω της πειραματικής διαδικασίας, εφαρμόζοντας τρία πειράματα σε δύο υπολογιστές διαφορετικών προδιαγραφών, προσπαθήσαμε να προβλέψουμε την κατανάλωση ενέργειας σε επεξεργαστή (χρησιμοποιώντας τρία και δύο χαρακτηριστικά ως εισόδους), κάρτα γραφικών (χρησιμοποιώντας τέσσερα και δύο χαρακτηριστικά ως εισόδους), καθώς και στο άθροισμα των δύο εξαρτημάτων (χρησιμοποιώντας τρία και δύο χαρακτηριστικά ως εισόδους).

Για τον υπολογιστή μεσαίων προδιαγραφών, στο πρώτο πείραμα, χρησιμοποιώντας τρία χαρακτηριστικά, παρατηρούμε πως το test score κυμαίνεται από 0.98 (βαθιά νευρωνικά δίκτυα, τυχαία δάση) έως 0.99 (K-πλησιέστεροι γείτονες, παλινδρομικά μοντέλα, μηχανές διανυσμάτων υποστήριξης και δέντρα απόφασης). Από τις εξισώσεις των παλινδρομικών μοντέλων συμπεραίνουμε πως η χρήση και η θερμοκρασία του επεξεργαστή έχουν κύριο ρόλο στη πρόβλεψη, ενώ η χρήση της κάρτας γραφικών δεν επηρεάζει σημαντικά. Χρησιμοποιώντας δύο χαρακτηριστικά, παρατηρούμε πως το test score κυμαίνεται από 0.97 (K-πλησιέστεροι γείτονες και βαθιά νευρωνικά δίκτυα) έως 0.99 (δέντρα απόφασης και τυχαία δάση). Από τις εξισώσεις των παλινδρομικών μοντέλων συμπεραίνουμε πως η χρήση του επεξεργαστή έχει κύριο ρόλο στη πρόβλεψη, ενώ η χρήση της κάρτας γραφικών δεν επηρεάζει σημαντικά. Στο δεύτερο πείραμα, χρησιμοποιώντας τρία χαρακτηριστικά, παρατηρούμε πως το test score κυμαίνεται από 0.87 (βαθιά νευρωνικά δίκτυα) έως 0.99 (K-πλησιέστεροι γείτονες). Από τις εξισώσεις των παλινδρομικών μοντέλων συμπεραίνουμε πως η χρήση της κάρτας γραφικών έχει κύριο ρόλο στη πρόβλεψη, ακολουθεί η χρήση του

επεξεργαστή, ενώ η θερμοκρασία του επεξεργαστή δεν επηρεάζει σημαντικά. Χρησιμοποιώντας δύο χαρακτηριστικά παρατηρούμε πως το test score κυμαίνεται από 0.94 (δέντρα απόφασης και βαθιά νευρωνικά δίκτυα) έως 0.97 (K-πλησιέστεροι γείτονες, παλινδρομικά μοντέλα και μηχανές διανυσμάτων υποστήριξης). Από τις εξισώσεις των παλινδρομικών μοντέλων συμπεραίνουμε πως η χρήση της κάρτας γραφικών έχει κύριο ρόλο στη πρόβλεψη, με τη χρήση του επεξεργαστή να έχει μικρότερη επιρροή. Στο τρίτο πείραμα, χρησιμοποιώντας τέσσερα χαρακτηριστικά, παρατηρούμε πως το test score κυμαίνεται από 0.92 (K-πλησιέστεροι γείτονες) έως 0.99 (παλινδρομικά μοντέλα και μηχανές διανυσμάτων υποστήριξης). Από τις εξισώσεις των παλινδρομικών μοντέλων συμπεραίνουμε πως η θερμική σχεδιαστική ισχύ έχει κύριο ρόλο στη πρόβλεψη, ενώ η χρήση και η θερμοκρασία της κάρτας γραφικών καθώς και η χρήση του επεξεργαστή δεν επηρεάζουν σχεδόν καθόλου. Χρησιμοποιώντας 2 χαρακτηριστικά, παρατηρούμε πως το test score κυμαίνεται από 0.68 (μηχανές διανυσμάτων υποστήριξης) έως 0.88 (βαθιά νευρωνικά δίκτυα). Από τις εξισώσεις των παλινδρομικών μοντέλων συμπεραίνουμε πως η χρήση της κάρτας γραφικών έχει κύριο ρόλο στη πρόβλεψη, ενώ η χρήση του επεξεργαστή επηρεάζει λιγότερο.

Για τον υπολογιστή υψηλών προδιαγραφών, στο πρώτο πείραμα, χρησιμοποιώντας τρία χαρακτηριστικά, παρατηρούμε πως το test score κυμαίνεται από 0.94 (βαθιά νευρωνικά δίκτυα) έως 0.99 (παλινδρομικά μοντέλα και μηχανές διανυσμάτων υποστήριξης). Από τις εξισώσεις των παλινδρομικών μοντέλων συμπεραίνουμε πως η χρήση του επεξεργαστή έχει κύριο ρόλο στη πρόβλεψη, ενώ η θερμοκρασία του επεξεργαστή και η χρήση της κάρτας γραφικών δεν επηρεάζουν σημαντικά. Χρησιμοποιώντας δύο χαρακτηριστικά, παρατηρούμε πως το test score κυμαίνεται από 0.97 (K-πλησιέστεροι γείτονες και βαθιά νευρωνικά δίκτυα) έως 0.99 (παλινδρομικά μοντέλα και μηχανές διανυσμάτων υποστήριξης). Από τις εξισώσεις των παλινδρομικών μοντέλων συμπεραίνουμε πως η χρήση του επεξεργαστή έχει κύριο ρόλο στη πρόβλεψη, ενώ η χρήση της κάρτας γραφικών δεν επηρεάζει σημαντικά. Στο δεύτερο πείραμα, χρησιμοποιώντας τρία χαρακτηριστικά, παρατηρούμε πως το test score κυμαίνεται από 0.82 (βαθιά νευρωνικά δίκτυα) έως 0.97 (K-πλησιέστεροι γείτονες, παλινδρομικά μοντέλα και μηχανές διανυσμάτων υποστήριξης). Από τις εξισώσεις των παλινδρομικών μοντέλων συμπεραίνουμε πως η χρήση του επεξερ-

γαστή έχει κύριο ρόλο στη πρόβλεψη, ενώ η θερμοκρασία του επεξεργαστή και η χρήση της κάρτας γραφικών δεν επηρεάζουν σημαντικά. Χρησιμοποιώντας δύο χαρακτηριστικά παρατηρούμε πως το test score κυμαίνεται από 0.54 (βαθιά νευρωνικά δίκτυα) έως 0.97 (παλινδρομικά μοντέλα και μηχανές διανυσμάτων υποστήριξης). Από τις εξισώσεις των παλινδρομικών μοντέλων συμπεραίνουμε πως η χρήση του επεξεργαστή έχει κύριο ρόλο στη πρόβλεψη, ενώ η χρήση της κάρτας γραφικών δεν το επηρεάζει σημαντικά. Στο τρίτο πείραμα, χρησιμοποιώντας τέσσερα χαρακτηριστικά, παρατηρούμε πως το test score κυμαίνεται από 0.95 (τυχαία δάση) έως 0.97 (K-πλησιέστεροι γείτονες και δέντρα απόφασης). Από τις εξισώσεις των παλινδρομικών μοντέλων συμπεραίνουμε πως η θερμική σχεδιαστική ισχύ έχει κύριο ρόλο στη πρόβλεψη, ενώ η χρήση και η θερμοκρασία της κάρτας γραφικών καθώς και η χρήση του επεξεργαστή δεν επηρεάζουν σχεδόν καθόλου. Χρησιμοποιώντας 2 χαρακτηριστικά, παρατηρούμε πως το test score κυμαίνεται από 0.94 (παλινδρομικά μοντέλα και μηχανές διανυσμάτων υποστήριξης) έως 0.97 (K-πλησιέστεροι γείτονες). Από τις εξισώσεις των παλινδρομικών μοντέλων συμπεραίνουμε πως η χρήση της κάρτας γραφικών έχει κύριο ρόλο στη πρόβλεψη, ενώ η χρήση του επεξεργαστή επηρεάζει λιγότερο.

Από όλα τα παραπάνω, συμπεραίνουμε πως η θερμοκρασία του επεξεργαστή αλλά και της κάρτας γραφικών δεν βοήθησαν σημαντικά στις προβλέψεις. Αντιθέτως, η χρήση του επεξεργαστή είχε σημαντική συνεισφορά τόσο στη πρόβλεψη κατανάλωσης ενέργειας του επεξεργαστή, όσο και στη πρόβλεψη κατανάλωσης ενέργειας του αθροίσματος των δύο εξαρτημάτων, και στους δύο υπολογιστές. Ομοίως, η χρήση της κάρτας γραφικών είχε σημαντική συνεισφορά στη πρόβλεψη κατανάλωσης ενέργειας του επεξεργαστή και του αθροίσματος των δύο εξαρτημάτων, μόνο στον υπολογιστή μεσαίων προδιαγραφών, ενώ στον υπολογιστή υψηλών προδιαγραφών επηρέαζε λιγότερο τις προβλέψεις. Ακόμη, είχε σημαντική συνεισφορά στη πρόβλεψη κατανάλωσης της κάρτας γραφικών. Τέλος, η θερμική σχεδιαστική ισχύ είχε τη μεγαλύτερη συνεισφορά στη πρόβλεψη κατανάλωσης ενέργειας της κάρτας γραφικών.

Βιβλιογραφία

- [3DM, 2019] (2019). 3DMark. <https://benchmarks.ul.com/>. Last Accessed September 25, 2019.
- [AID, 2019] (2019). AIDA64 Extreme. <https://www.aida64.com/products/aida64-extreme>. Last Accessed September 25, 2019.
- [sho, 2019] (2019). Data Center History. <http://opticalcloudinfra.com/index.php/what-why-and-how/short-data-center-history/>. Last Accessed September 24, 2019.
- [GTX, 2019] (2019). GIGABYTE GTX 970 G1 Gaming. <https://www.techpowerup.com/gpu-specs/gigabyte-gtx-970-g1-gaming.b3050>. Last Accessed September 25, 2019.
- [Hea, 2019] (2019). HeavyLoad. <https://www.jam-software.com/heavyload/>. Last Accessed September 25, 2019.
- [Mac, 2019] (2019). History of Machine Learning. <https://www.doc.ic.ac.uk/~jce317/history-machine-learning.html>. Last Accessed September 24, 2019.
- [Int, 2019a] (2019a). Intel Core i5-4690k Processor. <https://ark.intel.com/content/www/us/en/ark/products/80811/intel-core-i5-4690k-processor-6m-cache-up-to-3-90-ghz.html>. Last Accessed September 25, 2019.
- [Int, 2019b] (2019b). Intel Core i7-3770 Processor. <https://ark.intel.com/content/www/us/en/ark/products/65719/intel-core-i7-3770-processor-8m-cache-up-to-3-90-ghz.html>. Last Accessed September 25, 2019.
- [Tit, 2019] (2019). NVIDIA TITAN X Pascal. <https://www.techpowerup.com/gpu-specs/titan-x-pascal.c2863>. Last Accessed September 25, 2019.
- [gpu, 2019] (2019). TechPowerUp GPU-Z. <https://www.techpowerup.com/gpuz/>. Last Accessed September 25, 2019.
- [BBC, 2018] BBC (2018). The history of machine learning. <https://www.bbc.com/timelines/zypd97h>. Last Accessed September 25, 2019.
- [Bircher and John, 2011] Bircher, W. L. and John, L. K. (2011). Complete system power estimation using processor performance events. *IEEE Transactions on Computers*, 61(4):563–577.
- [Contreras and Martonosi, 2005] Contreras, G. and Martonosi, M. (2005). Power prediction for Intel XScale/spl reg/processors using performance monitoring unit events. In *ISLPED'05. Proceedings of the 2005 International Symposium on Low Power Electronics and Design, 2005.*, pages 221–226. IEEE.

-
- [Dayarathna et al., 2015] Dayarathna, M., Wen, Y., and Fan, R. (2015). Data center energy consumption modeling: A survey. *IEEE Communications Surveys & Tutorials*, 18(1):732–794.
- [Economou et al., 2006] Economou, D., Rivoire, S., Kozyrakis, C., and Ranganathan, P. (2006). Full-system power analysis and modeling for server environments. *International Symposium on Computer Architecture-IEEE*.
- [Garris, 2018] Garris, L. (2018). The Current State of Data Center Energy Efficiency in Europe. <https://www.opencompute.org/documents/the-current-state-of-data-center-energy-efficiency-in-europe-ocp-white-paper>. Last Accessed September 24, 2019.
- [Géron, 2017] Géron, A. (2017). Hands-on machine learning with Scikit-Learn and TensorFlow: concepts, tools, and techniques to build intelligent systems. ” O’Reilly Media, Inc.”.
- [Harvey, 2017] Harvey, C. (2017). Data Centers. <https://www.datamation.com/data-center/what-is-data-center.html>. Last Accessed September 25, 2019.
- [Joseph and Martonosi, 2001] Joseph, R. and Martonosi, M. (2001). Run-time power estimation in high performance microprocessors. In *ISLPED’01: Proceedings of the 2001 International Symposium on Low Power Electronics and Design (IEEE Cat. No. 01TH8581)*, pages 135–140. IEEE.
- [Li and John, 2003] Li, T. and John, L. K. (2003). Run-time modeling and estimation of operating system power consumption. In *ACM SIGMETRICS Performance Evaluation Review*, volume 31, pages 160–171. ACM.
- [Marr, 2016] Marr, B. (2016). A Short History of Machine Learning. <https://www.forbes.com/sites/bernardmarr/2016/02/19/a-short-history-of-machine-learning-every-manager-should-read/#2f570ae15e78>. Last Accessed September 25, 2019.
- [Martin, 2017] Martin, N. (2017). GPU-accelerated computing makes its way into the data center. <https://searchdatacenter.techtarget.com/feature/GPU-accelerated-computing-makes-its-way-into-the-data-center>. Last Accessed September 24, 2019.
- [McNerney, 2019] McNerney, M. (2019). The Data Center Dilemma: Is Our Data Destroying the Environment? <https://www.datacenterknowledge.com/industry-perspectives/data-center-dilemma-our-data-destroying-environment>. Last Accessed September 25, 2019.
- [Mitchell, 2018] Mitchell, D. (2018). Evolution of Data Centers [A Look at how Data Center’s have evolved over the past 50 years]. <https://racksimply.com/evolution-of-data-centers-a-look-at-how-data-centers-have-evolved-over-the-past-50-years/>. Last Accessed September 24, 2019.

-
- [Müller et al., 2016] Müller, A. C., Guido, S., et al. (2016). Introduction to machine learning with Python: a guide for data scientists. ” O’Reilly Media, Inc.”.
- [Pramatarov, 2018] Pramatarov, M. (2018). The History of Data Centers. <https://blog.cloudflare.com/en/the-history-of-data-centers/>. Last Accessed September 25, 2019.
- [Pupale, 2018] Pupale, R. (2018). Support Vector Machines(SVM) — An Overview. <https://towardsdatascience.com/https-medium-com-pupalerushikesh-svm-f4b42800e989>. Last Accessed September 24, 2019.
- [Rodrigues et al., 2013] Rodrigues, R., Annamalai, A., Koren, I., and Kundu, S. (2013). A study on the use of performance counters to estimate power in microprocessors. *IEEE Transactions on Circuits and Systems II: Express Briefs*, 60(12):882–886.
- [Rodriguez-Martinez et al., 2011] Rodriguez-Martinez, M., Valdivia, H., Seguel, J., and Greer, M. (2011). Estimating power/energy consumption in database servers. *Procedia Computer Science*, 6:112–117.
- [Ross, 2018] Ross, A. (2018). A perfect storm: the environmental impact of data centres. <https://www.information-age.com/a-perfect-storm-the-environmental-impact-of-data-centres-123474834/>. Last Accessed September 25, 2019.
- [Shalev-Shwartz and Ben-David, 2014] Shalev-Shwartz, S. and Ben-David, S. (2014). Understanding machine learning: From theory to algorithms. Cambridge university press.
- [Singh et al., 2009] Singh, K., Bhadauria, M., and McKee, S. A. (2009). Real time power estimation and thread scheduling via performance counters. *ACM SIGARCH Computer Architecture News*, 37(2):46–55.
- [Soni, 2018] Soni, D. (2018). Supervised vs. Unsupervised Learning. <https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d>. Last Accessed September 24, 2019.
- [Supermicro, 2018] Supermicro (2018). Data Centers and The Environment. https://www.supermicro.com/wekeepitgreen/Data_Centers_and_the_Environment_Dec2018_Final.pdf. Last Accessed September 25, 2019.