ΕΡΓΑΣΙΑ ΣΤΑ

ΑΣΑΦΗ ΣΥΣΤΗΜΑΤΑ

ΟΝΟΜΑ: ΚΩΝΣΤΑΝΤΙΝΟΣ

ΕΠΙΘΕΤΟ: ΛΕΤΡΟΣ

ΣΧΟΛΗ: ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΘΕΣΣΑΛΟΝΙΚΗΣ

ΤΜΗΜΑ: ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧ. ΚΑΙ ΜΗΧ. ΥΠΟΛΟΓΙΣΤΩΝ

ΑΕΜ: 8851

ΕΞΑΜΗΝΟ: 8ο

ΕΤΟΣ: 2019

**Επίλυση προβλημάτων παλινδρόμησης με χρήση μοντέλων TSK**

**Ομάδα 4 – S08**

Περιεχόμενα

[Περιγραφή του Προβλήματος 3](#_Toc14721084)

[Εφαρμογή στο Σετ Δεδομένων Avila 3](#_Toc14721085)

[Προετοιμασία του Σετ Δεδομένων 3](#_Toc14721086)

[Περιγραφή της Διαδικασίας Εκπαίδευσης 4](#_Toc14721087)

[Αποτελέσματα TSK Μοντέλων και Δείκτες Απόδοσης 5](#_Toc14721088)

[TSK Μοντέλο 1 5](#_Toc14721089)

[TSK Μοντέλο 2 6](#_Toc14721090)

[TSK Μοντέλο 3 6](#_Toc14721091)

[TSK Μοντέλο 4 7](#_Toc14721092)

[TSK Μοντέλο 5 8](#_Toc14721093)

[Δείκτες Απόδοσης και Χρόνοι Εκτέλεσης 8](#_Toc14721094)

# Περιγραφή του Προβλήματος

Στόχος αυτής της εργασίας είναι η διερεύνηση της ικανότητας των TSK μοντέλων στην επίλυση προβλημάτων ταξινόμησης (classification) με χρήση ασαφών νευρωνικών μοντέλων. Η εργασία διακρίνεται σε δύο τμήματα στα οποία θα χρησιμοποιηθούν δύο διαφορετικά σετ δεδομένων. Σκοπός του πρώτου τμήματος είναι η εκπαίδευση και αξιολόγηση ενός πλήθους μοντέλων (με διαφορετικό πλήθος IF THEN κανόνων) τα οποία να μπορούν να ταξινομούν με επιτυχία το κάθε δείγμα στην κλάση που ανήκει. Αντίθετα στο δεύτερο μέρος, το μεγάλο πλήθος χαρακτηριστικών του δεύτερου σετ δεδομένων της εργασίας καθιστά τη διαδικασία εκμάθησης πολύ δυσκολότερη. Για την επίλυση του παραπάνω προβλήματος γίνεται μια επιλογή των πιο σημαντικών χαρακτηριστικών του σετ δεδομένων με τη βοήθεια του αλγορίθμου Relief και τέλος αναζητείται από μια σειρά μοντέλων με διαφορετικές παραμέτρους, το μοντέλο με το ελάχιστο σφάλμα.

# Εφαρμογή στο Σετ Δεδομένων Avila

## Προετοιμασία του Σετ Δεδομένων

Το Avila Dataset της UCI περιλαμβάνει 10 χαρακτηριστικά και 20876 δείγματα ενώ έχει εξαχθεί από 800 εικόνες της «Βίβλου Avila», ενός γιγάντιου λατινικού αντιγράφου της Βίβλου του 12ου αιώνα. Η στόχος ταξινόμησης συνίσταται στη συσχέτιση κάθε σχεδίου με έναν αντιγραφέα.

Αρχικά ταξινομούμε κατά αύξουσα σειρά το σετ δεδομένων με βάση τη στήλη που περιέχει τις διάφορες τιμές εξόδων (κλάσεις) 1–12 και προχωρούμε στην καταμέτρηση της συχνότητας εμφάνισης κάθε διαφορετικής τιμής εξόδου.

Στη συνέχεια, πραγματοποιούμε διαχωρισμό του σετ δεδομένων σε τρία μη επικαλυπτόμενα υποσύνολα ως εξής:

1. 60% : Σετ Εκπαίδευσης – training set
2. 20% : Σετ Επικύρωσης – validation set
3. 20% : Σετ Ελέγχου – check set

με τρόπο τέτοιο ώστε οι παραπάνω συχνότητες εμφάνισης να διατηρούνται περίπου σταθερές και ανακατεύουμε το κάθε σετ ξεχωριστά.

Επίσης, κανονικοποιούμε τις τιμές **κάθε στήλης** του Dataset με βάση τον παρακάτω τύπο:



**όπου οι min και max τιμές αφορούν κάθε στήλη του training set ξεχωριστά και με βάση την οποία κανονικοποιούμε και τα validation και check set.**

Ο λόγος που γίνεται αυτό είναι ότι κάθε χαρακτηριστικό περιλαμβάνει τιμές σε διαφορετικό εύρος, με αποτέλεσμα αν αγνοήσουμε το παραπάνω βήμα να μην υλοποιείται ορθά η εκπαίδευση του δικτύου. Μετά το πέρας της παραπάνω διαδικασίας όλα τα χαρακτηριστικά, αλλά και η έξοδος παίρνουν τιμές εύρους από 0 έως 1.

## Περιγραφή της Διαδικασίας Εκπαίδευσης

Η εκπαίδευση γίνεται με την υβριδική μέθοδο, δηλαδή οι παράμετροι των συναρτήσεων συμμετοχής βελτιστοποιούνται με τη μέθοδο Backpropagation.

Δημιουργούμε, με τη συνάρτηση genfis() του MATLAB, το μοντέλο προς εκπαίδευση με βάση τα χαρακτηριστικά του πίνακα ανάλογα με τον αριθμό του μοντέλου και τη μέθοδο Subtractive Clustering, δίνοντας ως είσοδο τα δεδομένα εκπαίδευσης. Επίσης, οι συναρτήσεις εξόδου μεταβάλλονται από την προκαθορισμένη μορφή Polynomial (Linear) σε Singleton (Constant) .

Στη συνέχεια εκπαιδεύουμε το μοντέλο με χρήση της συνάρτησης anfis() του MATLAB για 250 εποχές, προχωρούμε στην αξιολόγησή του και τέλος υπολογίζουμε τις ζητούμενα μεγέθη - δείκτες απόδοσης Error Matrix, Overall Accuracy, Producer’s Accuracy, User’s Accuracy, .

Σημειώνουμε επίσης ότι πριν την αξιολόγηση του μοντέλου με το σετ δεδομένων ελέγχου, στρογγυλοποιούμε τις τιμές εξόδου στον κοντινότερο ακέραιο, καθώς το παρών πρόβλημα

είναι πρόβλημα ταξινόμησης οπότε ο εκτιμώμενος αριθμός της κλάσης θα πρέπει να ανήκει στο σύνολο τιμών της εξόδου (δηλαδή ακέραιος από 1 έως 12).

Συνολικά εξετάζονται πέντε μοντέλα με διαφορετικές τιμές ακτίνας, δηλαδή της παραμέτρου της μεθόδου Subtractive Clustering που προσδιορίζει το εύρος επιρροής του κέντρου κάθε κλάσης, άρα και το πλήθος των IF THEN κανόνων. Τα πέντε μοντέλα λαμβάνουν τιμές ακτίνας από το σύνολο τιμών όπως φαίνεται στο Σχήμα 1.



*Σχήμα 1: TSK Μοντέλα για διάφορες τιμές ακτίνας*

## Αποτελέσματα TSK Μοντέλων και Δείκτες Απόδοσης

### TSK Μοντέλο 1

Οι συναρτήσεις συμμετοχής για το πρώτο μοντέλο πριν από τη διαδικασία εκπαίδευσης φαίνονται στο Σχήμα 2.

**(Σχήμα 2)**

*Σχήμα 2: Αρχικές Συναρτήσεις Συμμετοχής - TSK Μοντέλο 1*

Τα αποτελέσματα της παραπάνω διαδικασίας φαίνονται στη συνέχεια. Αρχικά βλέπουμε τη μορφή των συναρτήσεων συμμετοχής του μοντέλου μετά την εκπαίδευση.

**(Σχήμα 3)**

*Σχήμα 3: Τελικές Συναρτήσεις Συμμετοχής - TSK Μοντέλο 1*

Ακολουθούν οι καμπύλες εκμάθησης στο πέρας των εποχών.

**(Σχήμα 4)**

*Σχήμα 4: Καμπύλες Εκμάθησης - TSK Μοντέλο 1*

Τέλος, βλέπουμε τα σφάλματα πρόβλεψης και τις τιμές πραγματικής και εκτιμήτριας εξόδου για το σύνολο των δεδομένων ελέγχου.

**(Σχήμα 5)**

*Σχήμα 5: Σφάλματα Πρόβλεψης - TSK Μοντέλο 1*

**(Σχήμα 6)**

*Σχήμα 6: Πραγματική και Εκτιμήτρια Έξοδος - TSK Μοντέλο 1*

### TSK Μοντέλο 2

Οι συναρτήσεις συμμετοχής για το δεύτερο μοντέλο πριν από τη διαδικασία εκπαίδευσης φαίνονται στο Σχήμα 7.

**(Σχήμα 7)**

*Σχήμα 7: Αρχικές Συναρτήσεις Συμμετοχής - TSK Μοντέλο 2*

Τα αποτελέσματα της παραπάνω διαδικασίας φαίνονται στη συνέχεια. Αρχικά βλέπουμε τη μορφή των συναρτήσεων συμμετοχής του μοντέλου μετά την εκπαίδευση.

**(Σχήμα 8)**

*Σχήμα 8: Τελικές Συναρτήσεις Συμμετοχής - TSK Μοντέλο 2*

Ακολουθούν οι καμπύλες εκμάθησης στο πέρας των εποχών.

**(Σχήμα 9)**

*Σχήμα 9: Καμπύλες Εκμάθησης - TSK Μοντέλο 2*

Τέλος, βλέπουμε τα σφάλματα πρόβλεψης και τις τιμές πραγματικής και εκτιμήτριας εξόδου για το σύνολο των δεδομένων ελέγχου.

**(Σχήμα 10)**

*Σχήμα 10: Σφάλματα Πρόβλεψης - TSK Μοντέλο 2*

**(Σχήμα 11)**

*Σχήμα 11: Πραγματική και Εκτιμήτρια Έξοδος - TSK Μοντέλο 2*

### TSK Μοντέλο 3

Οι συναρτήσεις συμμετοχής για το τρίτο μοντέλο πριν από τη διαδικασία εκπαίδευσης φαίνονται στο Σχήμα 12.

**(Σχήμα 12)**

*Σχήμα 12: Αρχικές Συναρτήσεις Συμμετοχής - TSK Μοντέλο 3*

Τα αποτελέσματα της παραπάνω διαδικασίας φαίνονται στη συνέχεια. Αρχικά βλέπουμε τη μορφή των συναρτήσεων συμμετοχής του μοντέλου μετά την εκπαίδευση.

**(Σχήμα 13)**

*Σχήμα 13: Τελικές Συναρτήσεις Συμμετοχής - TSK Μοντέλο 3*

Ακολουθούν οι καμπύλες εκμάθησης στο πέρας των εποχών.

**(Σχήμα 14)**

*Σχήμα 14: Καμπύλες Εκμάθησης - TSK Μοντέλο 3*

Τέλος, βλέπουμε τα σφάλματα πρόβλεψης και τις τιμές πραγματικής και εκτιμήτριας εξόδου για το σύνολο των δεδομένων ελέγχου.

**(Σχήμα 15)**

*Σχήμα 15: Σφάλματα Πρόβλεψης - TSK Μοντέλο 3*

**(Σχήμα 16)**

*Σχήμα 16: Πραγματική και Εκτιμήτρια Έξοδος - TSK Μοντέλο 3*

### TSK Μοντέλο 4

Οι συναρτήσεις συμμετοχής για το τέταρτο μοντέλο πριν από τη διαδικασία εκπαίδευσης φαίνονται στο Σχήμα 17.

**(Σχήμα 17)**

*Σχήμα 17: Αρχικές Συναρτήσεις Συμμετοχής - TSK Μοντέλο 4*

Τα αποτελέσματα της παραπάνω διαδικασίας φαίνονται στη συνέχεια. Αρχικά βλέπουμε τη μορφή των συναρτήσεων συμμετοχής του μοντέλου μετά την εκπαίδευση.

**(Σχήμα 18)**

*Σχήμα 18: Τελικές Συναρτήσεις Συμμετοχής - TSK Μοντέλο 4*

Ακολουθούν οι καμπύλες εκμάθησης στο πέρας των εποχών.

**(Σχήμα 19)**

*Σχήμα 19: Καμπύλες Εκμάθησης - TSK Μοντέλο 4*

Τέλος, βλέπουμε τα σφάλματα πρόβλεψης και τις τιμές πραγματικής και εκτιμήτριας εξόδου για το σύνολο των δεδομένων ελέγχου.

**(Σχήμα 20)**

*Σχήμα 20: Σφάλματα Πρόβλεψης - TSK Μοντέλο 4*

**(Σχήμα 21)**

*Σχήμα 21: Πραγματική και Εκτιμήτρια Έξοδος - TSK Μοντέλο 4*

### TSK Μοντέλο 5

Οι συναρτήσεις συμμετοχής για το πέμπτο μοντέλο πριν από τη διαδικασία εκπαίδευσης φαίνονται στο Σχήμα 22.

**(Σχήμα 22)**

*Σχήμα 22: Αρχικές Συναρτήσεις Συμμετοχής - TSK Μοντέλο 5*

Τα αποτελέσματα της παραπάνω διαδικασίας φαίνονται στη συνέχεια. Αρχικά βλέπουμε τη μορφή των συναρτήσεων συμμετοχής του μοντέλου μετά την εκπαίδευση.

**(Σχήμα 23)**

*Σχήμα 23: Τελικές Συναρτήσεις Συμμετοχής - TSK Μοντέλο 5*

Ακολουθούν οι καμπύλες εκμάθησης στο πέρας των εποχών.

**(Σχήμα 24)**

*Σχήμα 24: Καμπύλες Εκμάθησης - TSK Μοντέλο 5*

Τέλος, βλέπουμε τα σφάλματα πρόβλεψης και τις τιμές πραγματικής και εκτιμήτριας εξόδου για το σύνολο των δεδομένων ελέγχου.

**(Σχήμα 25)**

*Σχήμα 25: Σφάλματα Πρόβλεψης - TSK Μοντέλο 5*

**(Σχήμα 26)**

*Σχήμα 26: Πραγματική και Εκτιμήτρια Έξοδος - TSK Μοντέλο 5*

### Δείκτες Απόδοσης και Χρόνοι Εκτέλεσης

Στον παρακάτω πίνακα βλέπουμε τους διαφορετικούς δείκτες απόδοσης και το χρόνο εκτέλεσης για τα πέντε μοντέλα.

**(Σχήμα 27)**

*Σχήμα 27: Πίνακας Δεικτών Απόδοσης – Χρόνου Εκτέλεσης*

**ΣΧΟΛΙΑ!!!!!!!!!**

# Εφαρμογή στο Σετ Δεδομένων Isolet

**Αντιμετώπιση σετ δεδομένων υψηλής διαστασιμότητας**

Το Isolet Dataset πρόκειται για ένα πολύ μεγαλύτερο σετ δεδομένων σε σχέση με το Avila, καθώς περιέχει 617 διαφορετικά χαρακτηριστικά και 7797 δεδομένα. Στόχος του τμήματος της εργασίας αυτού είναι η ορθή ταξινόμηση των δειγμάτων με βάση τα χαρακτηριστικά αυτά. Ο μεγάλος όγκος των χαρακτηριστικών και δεδομένων καθιστά την εκπαίδευση του ζητούμενου μοντέλου πρακτικά ανέφικτη, καθώς ο χρόνος που απαιτείται για αυτήν είναι υπερβολικά μεγάλος. Για το λόγο αυτό, θα χρειαστεί να επιλέξουμε ένα αρκετά πιο περιορισμένο πλήθος χαρακτηριστικών, και συγκεκριμένα τα πιο αντιπροσωπευτικά του δείγματος, η επιλογή των οποίων γίνεται με χρήση του αλγορίθμου Relief.

## Εύρεση Πλήθους Χαρακτηριστικών και Κανόνων για βέλτιστη Μοντελοποίηση

Αρχικά ταξινομούμε κατά αύξουσα σειρά το σετ δεδομένων με βάση τη στήλη που περιέχει τις διάφορες τιμές εξόδων (κλάσεις) και προχωρούμε στην καταμέτρηση της συχνότητας εμφάνισης κάθε διαφορετικής τιμής εξόδου.

Στη συνέχεια, πραγματοποιούμε διαχωρισμό του σετ δεδομένων σε τρία μη επικαλυπτόμενα υποσύνολα ως εξής:

1. 60% : Σετ Εκπαίδευσης – training set
2. 20% : Σετ Επικύρωσης – validation set
3. 20% : Σετ Ελέγχου – check set

με τρόπο τέτοιο ώστε οι παραπάνω συχνότητες εμφάνισης να διατηρούνται περίπου σταθερές και ανακατεύουμε το κάθε σετ ξεχωριστά.

Στο σημείο αυτό είναι καλό να εφαρμόσουμε μια προεπεξεργασία στα δεδομένα μας και συγκεκριμένα να ελέγξουμε ότι δεν υπάρχουν κενές τιμές και διπλότυπα δείγματα. Με τον τρόπο αυτό, θα είναι αποτελεσματικότερη, αλλά και ταχύτερη, η εκπαίδευση του δικτύου. Για αυτό το λόγο, αφού διαπιστώσουμε ότι δεν υπάρχουν NaN τιμές εφαρμόζουμε κανονικοποίηση **σε κάθε στήλη** του Dataset με βάση τον παρακάτω τύπο:



**όπου οι min και max τιμές αφορούν κάθε στήλη του training set ξεχωριστά και με βάση την οποία κανονικοποιούμε και τα validation και check set.**

Ο λόγος που γίνεται αυτό είναι ότι κάθε χαρακτηριστικό περιλαμβάνει τιμές σε διαφορετικό εύρος, με αποτέλεσμα αν αγνοήσουμε το παραπάνω βήμα να μην υλοποιείται ορθά η εκπαίδευση του δικτύου. Μετά το πέρας της παραπάνω διαδικασίας όλα τα χαρακτηριστικά, αλλά και η έξοδος παίρνουν τιμές εύρους από 0 έως 1.

Στη συνέχεια εφαρμόζουμε τον αλγόριθμο Relief επιλέγοντας ως αριθμό γειτόνων το 100 ώστε να γίνει εκτίμηση των σημαντικότερων χαρακτηριστικών με τη σειρά που εμφανίζονται στον πίνακα ranks.

Έπειτα χρησιμοποιούμε το συνδυασμό των μεθόδων Grid Search και 5-Fold Cross Validation ώστε να βρούμε το μοντέλο που εκτιμάει καλύτερα την επιθυμητή έξοδο. Συγκεκριμένα η μέθοδος k-Fold Cross Validation, με τιμή k=5, αποτελείται από τα εξής βήματα:

1. Αρχικά, διαχωρίζουμε το set δεδομένων εκπαίδευσης σε πέντε νέα τμήματα και κάθε νέο τμήμα σε ένα νέο set δεδομένων εκπαίδευσης (80% του αρχικού set εκπαίδευσης) και ένα νέο set δεδομένων επικύρωσης (20% του αρχικού set εκπαίδευσης), δημιουργώντας πέντε νέα δευτερεύοντα μοντέλα.
2. Εκπαιδεύουμε καθένα από αυτά τα δευτερεύοντα μοντέλα και στη συνέχεια υπολογίζουμε το σφάλμα του καθενός ως την ευκλείδεια νόρμα της διαφοράς της πραγματικής εξόδου από την εκτιμήτρια. **Ως σετ ελέγχου χρησιμοποιείται το σετ επικύρωσης του κύριου μοντέλου. (Το check set είναι άχρηστο ???)**
3. Τέλος, υπολογίζουμε τη μέση τιμή των προηγουμένως υπολογισμένων σφαλμάτων, η οποία αποτελεί αντιπροσωπευτικό δείγμα του πραγματικού σφάλματος για το συνολικό κύριο μοντέλο.

Η παραπάνω διαδικασία συνδυάζεται με τη μέθοδο Grid Search, δηλαδή εκτελείται μια επαναληπτική διαδικασία στην οποία εφαρμόζεται συνεχώς η μέθοδος 5-Fold Cross Validation για διάφορα κύρια μοντέλα μεταβάλλοντας κάθε φορά τόσο το πλήθος των IF THEN κανόνων όσο και το πλήθος χαρακτηριστικών που λαμβάνονται υπόψιν. Έπειτα συγκεντρώνονται όλα τα μέσα σφάλματα, που υπολογίζονται όπως αναφέρθηκε προηγουμένως για κάθε κύριο μοντέλο, και επιλέγεται το βέλτιστο μοντέλο ως αυτό που παρουσιάζει το ελάχιστο μέσο σφάλμα.

Για την ομαδοποίηση και τη δημιουργία των IF THEN κανόνων χρησιμοποιείται η μέθοδος Fuzzy C-Means (FCM) ενώ οι διάφορες περιπτώσεις των μοντέλων που διερευνώνται αποτελούνται από τους συνδυασμούς πλήθους χαρακτηριστικών και IF THEN κανόνων όπως προκύπτουν από το καρτεσιανό γινόμενο των συνόλων αντίστοιχα,



Δηλαδή εξετάζονται 20 διαφορετικά κύρια μοντέλα του αρχικού σετ εκπαίδευσης, καθένα από τα οποία αξιολογείται με βάση τα 5 δευτερεύοντα μοντέλα (Μέθοδος 5-Fold Validation) στα οποία χρησιμοποιούνται τα μικρότερα υποσύνολα-σετ εκπαίδευσης όπως αναφέρθηκε προηγουμένως. Συνεπώς, τελικά πραγματοποιείται εκπαίδευση, συνολικά, 100 μοντέλων για 150 εποχές, το καθένα ώστε να αποφασιστεί ποιο από τα κύρια μοντέλα είναι το βέλτιστο.

Στον παρακάτω πίνακα παρουσιάζεται το μέσο σφάλμα για τα 20 διαφορετικά μοντέλα.

**(Σχήμα 28)**

*Σχήμα 28: Πίνακας Μέσου Σφάλματος για τα διάφορα μοντέλα*

Στα παρακάτω διαγράμματα φαίνονται γραφικά οι τιμές του μέσου σφάλματος για τις διάφορες τιμές χαρακτηριστικών και κανόνων.

**(Σχήμα 29)**

*Σχήμα 29: Μέσο σφάλμα μοντέλων για τις διάφορες τιμές πλήθους χαρακτηριστικών και κανόνων*

Τέλος, τα παραπάνω σφάλματα παρουσιάζονται και σε ένα κοινό διάγραμμα τριών διαστάσεων.

**(Σχήμα 30)**

*Σχήμα 30: Κοινό 3D Διάγραμμα Μέσου Σφάλματος των διάφορων μοντέλων*

Από τα παραπάνω είναι εμφανές ότι το βέλτιστο, από τα εξεταστέα μοντέλα, είναι αυτό με τα **xxx**  χαρακτηριστικά και τους **xxx** κανόνες. Παρατηρούμε επίσης ότι όσο αυξάνεται η πολυπλοκότητα του μοντέλου (πλήθος χαρακτηριστικών και κανόνων) τόσο αυξάνεται και ο χρόνος εκτέλεσης του αλγορίθμου, ωστόσο δεν βελτιώνεται απαραίτητα η ικανότητα εκτίμησης του μοντέλου.

## Εκπαίδευση βέλτιστου TSK μοντέλου

Αρχικά παρουσιάζουμε ορισμένες από τις συναρτήσεις συμμετοχής του βέλτιστου μοντέλου πριν την εκπαίδευσή του.

**(Σχήμα 31)**

*Σχήμα 31: Συναρτήσεις Συμμετοχής πριν την εκπαίδευση*

Μετά από εκπαίδευση σε **xxx** εποχές οι παραπάνω συναρτήσεις συμμετοχής λαμβάνουν την παρακάτω μορφή.

**(Σχήμα 32)**

*Σχήμα 32: Συναρτήσεις Συμμετοχής μετά την εκπαίδευση*

Ακολουθούν οι καμπύλες εκμάθησης στο πέρας των εποχών.

**(Σχήμα 33)**

*Σχήμα 33: Καμπύλες Εκμάθησης - TSK Μοντέλο*

Τέλος, βλέπουμε τα σφάλματα πρόβλεψης και τις τιμές πραγματικής και εκτιμήτριας εξόδου για το σύνολο των δεδομένων ελέγχου.

**(Σχήμα 34)**

*Σχήμα 34: Σφάλματα Πρόβλεψης - TSK Μοντέλο*

**(Σχήμα 35)**

*Σχήμα 35: Πραγματική και Εκτιμήτρια Έξοδος - TSK Μοντέλο*

## Δείκτες Απόδοσης και Χρόνος Εκτέλεσης

Στον παρακάτω πίνακα βλέπουμε τους δείκτες απόδοσης και το χρόνο εκτέλεσης για την εκπαίδευση και αξιολόγηση του βέλτιστου μοντέλου.

**(Σχήμα 36)**

*Σχήμα 36: Πίνακας Μετρικών Σφάλματος – Χρόνου Εκτέλεσης*

**ΣΧΟΛΙΑ ΜΕ ΒΑΣΗ ΤΑ ΑΠΟΤΕΛΕΣΜΑΤΑ**

# Αρχεία MATLAB

1. avilaModel.m : MATLAB Script – Υλοποίηση πρώτου τμήματος της εργασίας (Avila Dataset).
2. gridSearch.m : MATLAB Script – Υλοποίηση δεύτερου τμήματος της εργασίας (Isolet Dataset). Ο χρήστης ενημερώνεται σε ζωντανό χρόνο για την πρόοδο της διαδικασίας εκπαίδευσης των 100 μοντέλων και τις παραμέτρους (πλήθος χαρακτηριστικών, πλήθος κανόνων, αριθμός πτυχής) του μοντέλου που εκπαιδεύεται κάθε φορά. Τέλος, δημιουργείται και ένα αρχείο με όνομα optimum\_model.mat, το οποίο περιλαμβάνει τον αριθμό των χαρακτηριστικών και κανόνων του βέλτιστου μοντέλου καθώς και το απαραίτητο τμήμα του του πίνακα ranks, που καθορίζει με φθίνουσα σειρά σημασίας ποιες από τις στήλες των χαρακτηριστικών χρησιμοποιήθηκαν.
3. optimumModel.m : MATLAB Script – Εκπαίδευση του βέλτιστου TSK μοντέλου και υπολογισμός των απαραίτητων δεικτών απόδοσης.