

Αναφορά για την Εξαμηνιαία Εργασία στο μάθημα Θεωρία Αποφάσεων

Μέλη Ομάδας

Δημήτριος Κωστορρίζος: AM: 1054419, Έτος Σπουδών: 5ο, Υπεύθυνος ομάδας

Λάμπρος Παπαδόπουλος: AM: 1054433, Έτος Σπουδών: 5ο

Σπυρίδων Τσιμπούκης: AM: 1058123, Έτος Σπουδών: 5ο

Github Repository Link: <https://github.com/DimitrisKostorrizos/TheoriaApofaseon>

Επιλεγμένο Θέμα: Θέμα 5

Αριθμός Ομάδας: 38

Για την εκπόνηση της άσκησης, χρησιμοποιήσαμε το IDE VsCode και την Python version 3.8.5.

Στα αρχεία έχουμε comments και documentation strings πάνω από κάθε μέθοδο που περιγράφουν, σε λίγες γραμμές, την λειτουργία της εκάστοτε συνάρτησης.

Κατά την υλοποίηση, όπως θα δείτε και στο Github repository, αναζητήσαμε τους τρόπους με τους οποίους μπορούμε να αναζητήσουμε το αποτέλεσμα. Καταλήξαμε στους εξής τρόπους:

- Χρήση ενός SVC(support vector classifier)
- Logistic Regression
- Kmeans Clustering

Στην αρχή του προγράμματος, φορτώνουμε τα περιεχόμενα του train και του test αρχείου στην μνήμη. Έπειτα, εκπαιδεύουμε το SVC μοντέλο, χρησιμοποιώντας το train set και χρησιμοποιούμε το μοντέλο αυτό για να δημιουργήσουμε μία αρχική πρόβλεψη των τιμών για την ποιότητα του κρασιού.

Έπειτα, αποφασίσαμε να πειραματιστούμε με τις τιμές ορισμένων πεδίων, ώστε να δούμε την συμβολή τους στην αρχική πρόβλεψη. Δοκιμάσαμε δύο εκδοχές αυτής της τεχνικής. Στην πρώτη εκδοχή, αφαιρέσαμε πλήρως την στήλη του pH από ολόκληρο το train dataset, αναπροσαρμόσαμε το SVC και δημιουργήσαμε μία νέα πρόβλεψη για το αποτέλεσμα.

Στην δεύτερη εκδοχή, αφαιρέσαμε τις τιμές του pH από το ένα τρίτο του train dataset. Για τις εγγραφές, από τις οποίες αφαιρέσαμε την τιμή του pH, τους εισήγαμε ως τιμή pH, την μέση του pH των υπολοίπων εγγραφών του train dataset. Έπειτα, αναπροσαρμόσαμε το SVC και δημιουργήσαμε μία νέα πρόβλεψη για το αποτέλεσμα.

Εφαρμόσαμε μία παρόμοια τεχνική, αλλά αυτήν την φορά, αφαιρέσαμε πλήρως την τιμή του pH από όλες τις εγγραφές του train dataset. Χρησιμοποιώντας το Logistic Regression, δημιουργήσαμε μία πρόβλεψη για την στήλη του pH. Την πρόβλεψη αυτή, την εισήγαμε στις εγγραφές του train dataset, αναπροσαρμόσαμε το SVC και δημιουργήσαμε μία νέα πρόβλεψη για το αποτέλεσμα.

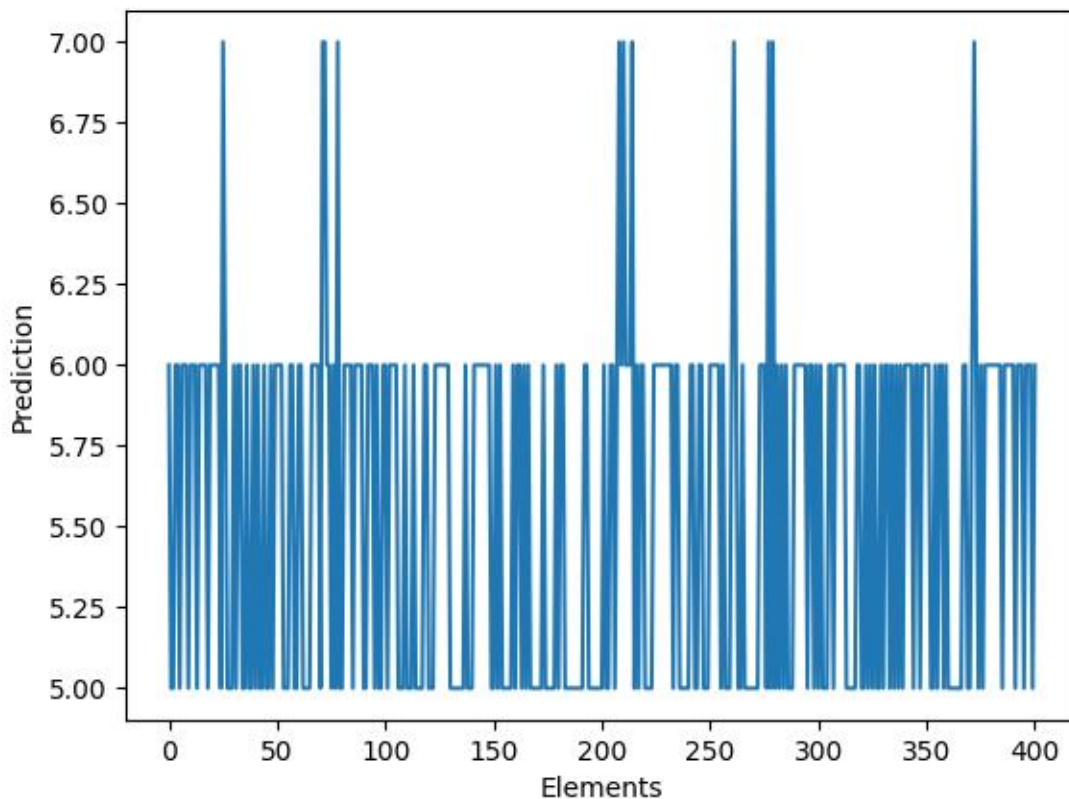
Τέλος, εφαρμόσαμε Kmeans Clustering ώστε να ομαδοποιήσουμε τις εγγραφές του train dataset. Χρησιμοποιώντας την μέση τιμή του cluster, αντικαταστήσαμε την τιμή του pH των εγγραφών που ανήκουν στο cluster, με την μέση τιμή του pH των εγγραφών αυτών. Έπειτα, αναπροσαρμόσαμε το SVC και δημιουργήσαμε μία νέα πρόβλεψη για το αποτέλεσμα.

Ελέγξαμε την πρόβλεψη, για κάθε μία από τις προαναφερθείσες τεχνικές και καταλήξαμε στο συμπέρασμα ότι η καλύτερη τεχνική από θέμα ακρίβειας ήταν η τεχνική με το Kmeans Clustering. Ωστόσο, η αποτελεσματική τεχνική από άποψη χρόνου ήταν το classification χρησιμοποιώντας το SVC.

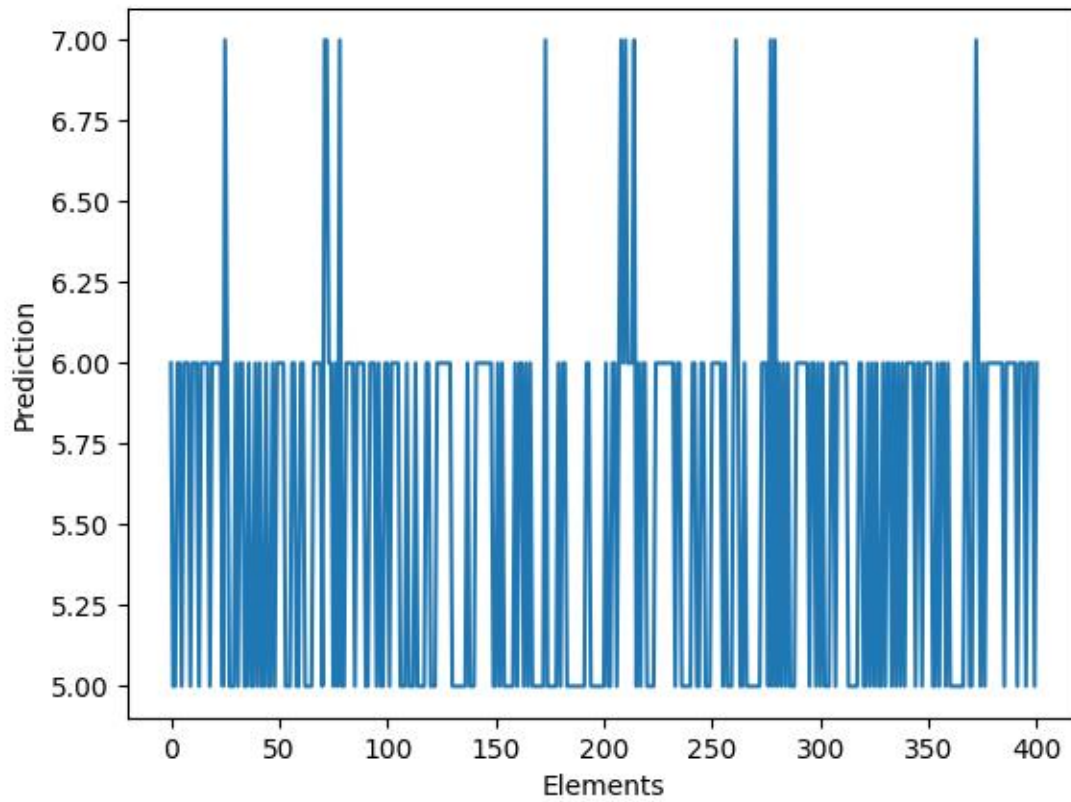
Κατά την εκτέλεση του κώδικα, αποθηκεύονται τα γραφήματα ως png αρχεία, στον τοπικό φάκελο του αρχείου.

Γραφήματα Μετρήσεων

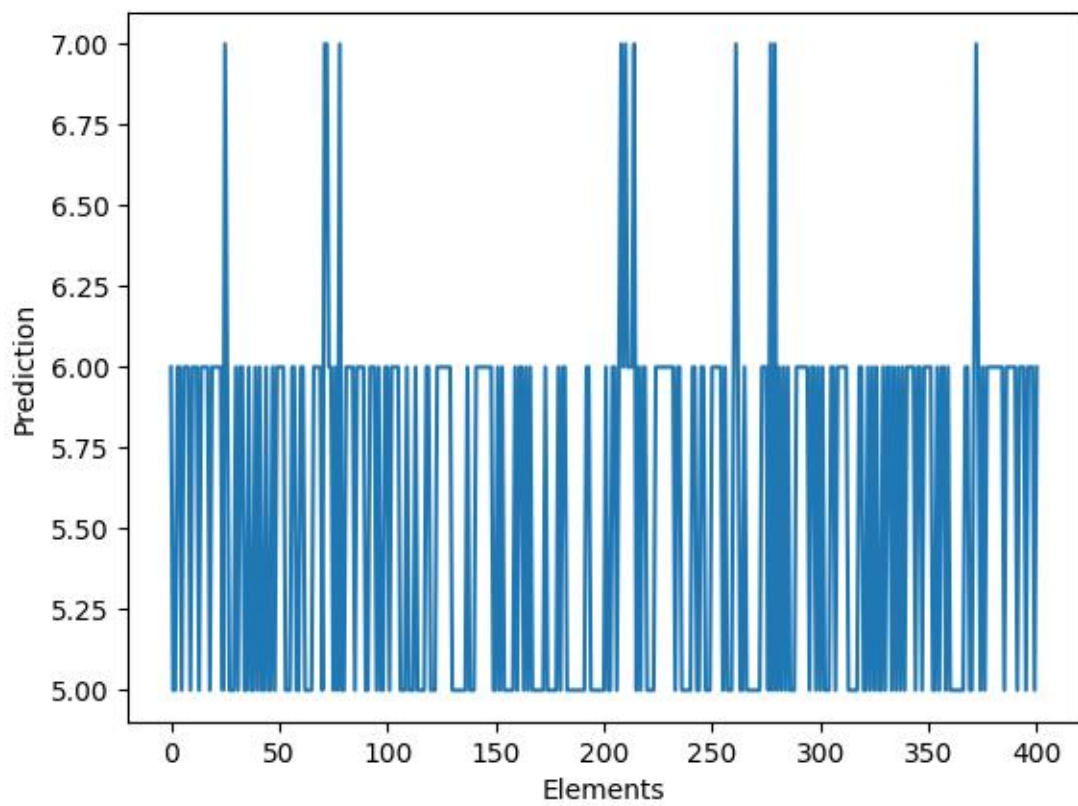
Αποτελέσματα για το SVC



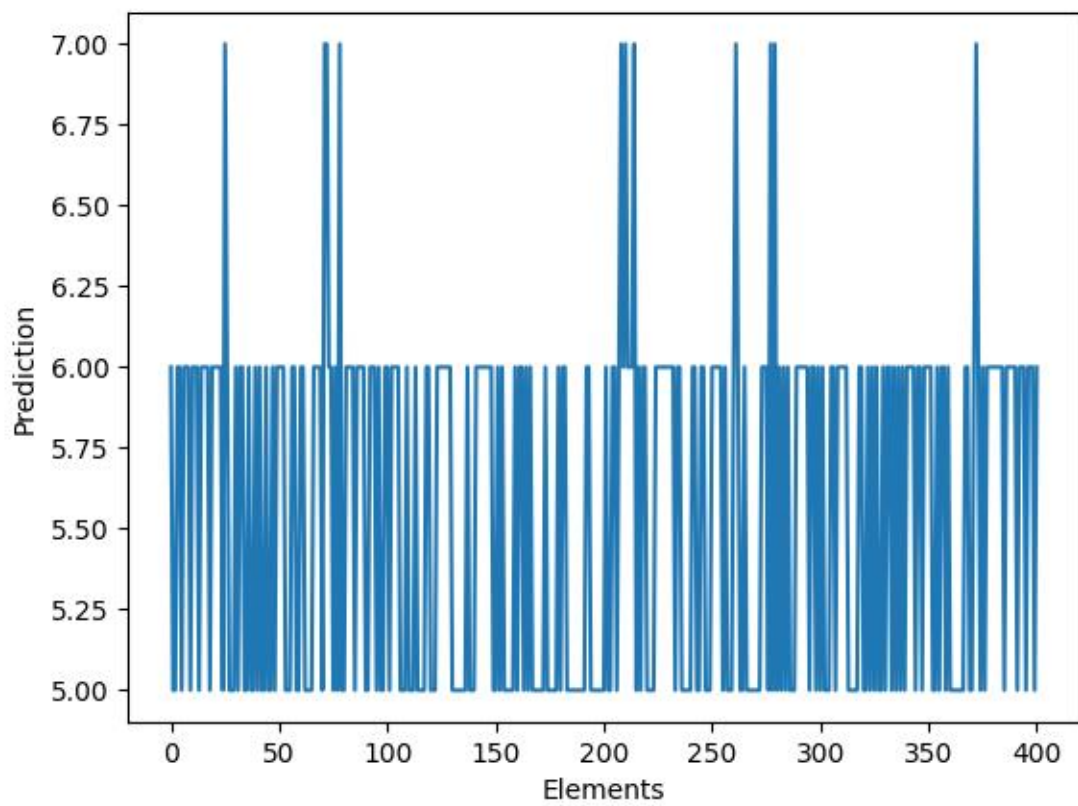
Αποτελέσματα για το Remove PH και SVC



Αποτελέσματα για το Average PH και SVC



Αποτελέσματα για το Logistic Regression PH και SVC



Αποτελέσματα για το Kmeans Classification PH και SVC

