A black text on a white background

Description automatically generated

A blue and white logo

Description automatically generated

Movies in Box Office and their Analysis

Εισαγωγή

Στα πλαίσια του τελευταίου ακαδημαϊκού έτους όλοι οι φοιτητές ήρθαν αντιμέτωποι με την πτυχιακή εργασία από την αρχή έως το τέλος του έτους. Ο στόχος της πτυχιακής εργασίας με τίτλο Movies in Box Office and their Analysis ήταν με την βοήθεια της Μηχανικής Μάθησης, η πρόβλεψη της επιτυχίας από ένα συγκεκριμένο αριθμό ταινιών που ανήκουν στο Imported Dataset στο MATLAB, χρησιμοποιώντας κάποιους συγκεκριμένους παράγοντες. Μερικοί από τους παράγοντες αυτούς ήταν το Runtime (διάρκεια ταινίας) της ταινίας ή το Budget (Προϋπολογισμός ταινίας) της ταινίας που συνεισφέρουν στην ακρίβεια της πρόβλεψης. Συνήθως η πρόβλεψη αποσκοπούσε με την χρήση δύο προβλεπόμενων παραγόντων, τους Gross και του Gross-Class (έσοδα και κατάταξη των ταινιών με βάση τα έσοδα αντίστοιχα), που χρησιμοποιήθηκαν για να εγκριθεί η εμπιστευτικότητα ως προς τον δείκτη ακρίβειας σε Regression και Classification Learner είτε σε Confusion Matrix, είτε σε Response. Σημασία έχει ότι κατά την διάρκεια της υλοποίησης του project πολλά Datasets με εφαρμοσμένες διεργασίες χρησιμοποιήθηκαν ώστε να γίνει η διαδικασία του testing, ώσπου να δημιουργηθεί το τελικό Dataset.

Κύριο Μέρος

Η πτυχιακή εργασία αποτελείται από δύο κομμάτια το θεωρητικό και το πρακτικό. Στο πρακτικό κομμάτι υλοποιήθηκε η έρευνα και η διεργασίες των μοντέλων και των Datasets. Μερικές από αυτές είναι ο αλγόριθμος One-Hot Encode στο MATLAB ή διαδικασία remove outliers κατά την οποία αφαιρούνται τα outliers του Dataset μέσα από τον Live Editor του MATLAB. Το MATLAB είναι ένα προγραμματιστικό περιβάλλον στο οποίο εφαρμόζονται αλγόριθμοι Μηχανικής Μάθησης και διαθέτει ήδη έτοιμα και εγκατεστημένα εργαλεία Μηχανικής Μάθησης, συγκεκριμένα αυτά που χρησιμοποιήθηκαν ήταν το Regression και το Classification. Παρακάτω φαίνονται δύο εικόνες που αντικατοπτρίζουν δύο μοντέλα με την επιθυμητή ακρίβεια της επιτυχίας στο MATLAB.

A screenshot of a computer

Description automatically generated A screenshot of a computer

Description automatically generated

Αποτελέσματα

Τα αποτελέσματα είναι πως το accuracy του Classification Learner κυμαίνεται μεταξύ 40 με 45 της εκατό ενώ το RMSE είναι ένας δείκτης ακρίβειας μέτριας έως κακής απόδοσης.

Συμπέρασμα

Γενικά η Μηχανική Μάθηση είναι μια επιστήμη που εξελίσσεται συνεχόμενα συσχετίσει της επιστήμης Data Mining και Data Analysis. Είναι μια αλληλένδετη αλυσίδα όπου αν σπάσει ένας κρίκος, σπάει ολόκληρη η αλυσίδα.