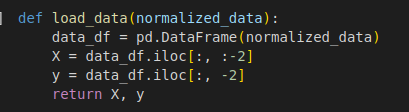
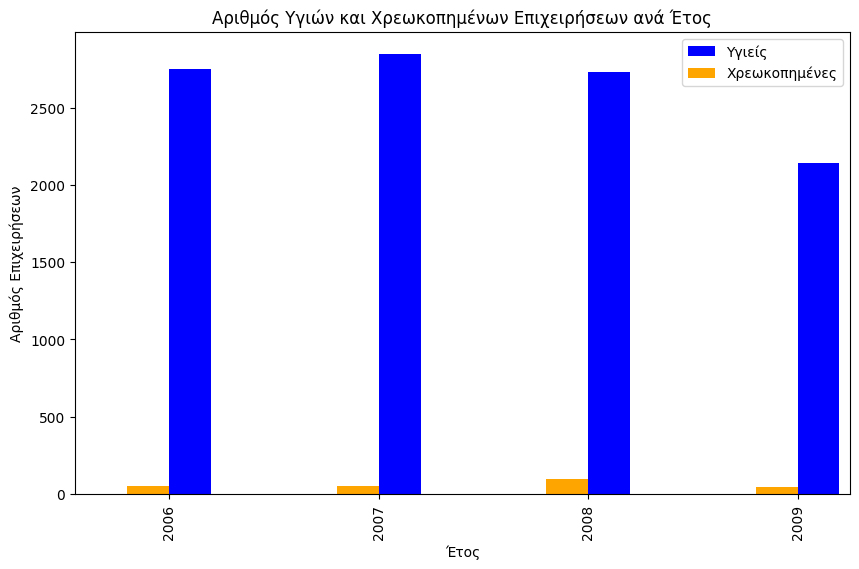
Η παρούσα αναφορά αναφέρεται στην εργασία που διενεργήθηκε στα πλαίσια του προγράμματος μεταπτυχιακών σπουδών 'Εφαρμοσμένη Πληροφορική' στο μάθημα Μέθοδοι και Εργαλεία της Τεχνητής Νοημοσύνης. Το πρόβλημα που αντιμετωπίσαμε εστιάζεται στον εντοπισμό εταιρειών που ενδέχεται να κηρύξουν χρεωκοπία, χρησιμοποιώντας δεδομένα που παρέχονται έτοιμα από κάποιον οργανισμό. Σκοπός της εργασίας μας ήταν να αναπτύξουμε ένα μοντέλο ταξινόμησης που θα μπορεί να προβλέπει με υψηλή ακρίβεια τις εταιρείες που ενδέχεται να κηρύξουν χρεωκοπία, λαμβάνοντας υπόψη διάφορους δείκτες απόδοσης και δραστηριοτήτων των εταιρειών.

Η μεθοδολογία που ακολουθήθηκε για την ανάπτυξη του μοντέλου περιλάμβανε τα εξής βήματα:

1. **Φόρτωση Δεδομένων**: Τα δεδομένα διαβάστηκαν από ένα αρχείο Excel, το οποίο περιείχε πληροφορίες σχετικά με την οικονομική κατάσταση εταιριών στις στήλες από Α μέχρι και Κ. Η τελευταία στήλη με το ΈΤΟΣ δεν τη χρειαζόμαστε, ενώ θα διαλέξουμε τη στήλη ΕΝΔΕΙΞΗ ΑΣΥΝΕΠΕΙΑΣ (=2) (ν+1) για τη μεταβλητή y.



1. **Εξερεύνηση Δεδομένων**: Πραγματοποιήθηκε εξερεύνηση των δεδομένων με την εμφάνιση των πρώτων δειγμάτων και τη δημιουργία γραφημάτων για να αναλυθεί η κατάσταση των υγιών και των χρεωκοπημένων εταιριών ανά έτος όπου παρατηρείται ότι το dataset είναι unbalanced.



1. **Ανάλυση και Επεξεργασία Δεδομένων**: Εκτελέστηκε ανάλυση στατιστικών μεγεθών για κάθε δείκτη, συμπεριλαμβανομένων των μέγιστων, ελαχίστων και μέσων τιμών, ώστε να κατανοηθεί η κατανομή των δεδομένων.
2. **Επιλογή Μοντέλων**: Επιλέχθηκαν πολλά διαφορετικά μοντέλα μηχανικής μάθησης, συμπεριλαμβανομένων του LDA, Logistic Regression, Decision Tree, Random Forest, k-Nearest Neighbors, Naïve Bayes, Support Vector Machines και Gradient Boosting.
3. **Εκπαίδευση και Αξιολόγηση Μοντέλων**: Τα μοντέλα εκπαιδεύτηκαν στα δεδομένα εκπαίδευσης και αξιολογήθηκαν χρησιμοποιώντας διάφορες μετρικές απόδοσης, όπως ο F1 Score, ο AUC ROC και οι επιτυχίες στις προβλέψεις των υγιών και των χρεωκοπημένων εταιριών.
4. **Εξαγωγή Αποτελεσμάτων**: Τα αποτελέσματα της ανάλυσης και της αξιολόγησης αποθηκεύτηκαν σε ένα αρχείο CSV για μελλοντική αναφορά.

Τα προβλήματα που αντιμετωπίστηκαν περιλάμβαναν την ανάγκη για καθαρισμό και προεπεξεργασία των δεδομένων, την επιλογή του κατάλληλου μοντέλου μηχανικής μάθησης και τη βελτιστοποίηση των υπερπαραμέτρων για καλύτερη απόδοση.