



Π.Μ.Σ ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΑΝΑΠΤΥΞΗ ΛΟΓΙΣΜΙΚΟΥ & ΝΕΦΟΣ

Μέθοδοι και Εργαλεία Τεχνητής Νοημοσύνης (SD0203)

Εργασία 2 – Επιβλεπόμενη μάθηση - Ταξινόμηση

Ονοματεπώνυμο : Χρυσοχοΐδης Αναστάσιος
Αριθμός Μητρώου : mai25067
email : mai25067@uom.edu.gr
Ακαδημαϊκό έτος 2024-2025
Μάιος 2025

Περιεχόμενα

1. Εισαγωγή.....	5
2. Μεθοδολογία	5
3. Πειραματικά Αποτελέσματα	7
3.1 Περιγραφή Dataset	7
3.2 Εκπαίδευση Μοντέλων.....	8
3.3 Σύγκριση Μοντέλων.....	8
3.4 Έλεγχος περιορισμών	10
4. Συμπεράσματα.....	12
5. Βιβλιογραφία.....	12
Παράρτημα 1 – Boxplots.....	13
Παράρτημα 2 – Confusion Matrices	17
Παράρτημα 3 – Αποτελέσματα επίδοσης μοντέλων	49

Πίνακας Διαγραμμάτων

Διάγραμμα 1 - Υγιείς και Χρεωκοπημένες εταιρείες ανά έτος	7
Διάγραμμα 2 - Επίδοση μοντέλων στο Train set	9
Διάγραμμα 3 - Επίδοση μοντέλων στο Test set	9
Διάγραμμα 4 – Positive Recall (χρεωκοπημένων)	11
Διάγραμμα 5 – Negative Recall (Υγιών)	11
Διάγραμμα 6 - Κατανομή δείκτη απόδοσης 365 * (BT/Κόστος Πωλήσεων)	13
Διάγραμμα 7 - Κατανομή δείκτη Λειτουργικά Αποτελέσματα / Συν. Ενεργητικό	13
Διάγραμμα 8 - Κατανομή δείκτη Χρηματοοικονομικές Δαπάνες / Πωλήσεις	14
Διάγραμμα 9 - Κατανομή δείκτη Πραγματική Ρευστότητα	14
Διάγραμμα 10 - Κατανομή δείκτη (Εμπορικές Απαιτήσεις * 365) / Πωλήσεις	15
Διάγραμμα 11 - Κατανομή δείκτη Σύνολο Υποχρεώσεων /Σύνολο Ενεργητικού	15
Διάγραμμα 12 - Κατανομή δείκτη Διάρκεια Παραμονής Αποθεμάτων	16
Διάγραμμα 13 - Κατανομή δείκτη Λογάριθμος Προσωπικού	16
Διάγραμμα 14 - LDA - Train set-Fold 1	17
Διάγραμμα 15 - LDA - Test set - Fold 1	17
Διάγραμμα 16 - LDA - Train set - Fold 2	18
Διάγραμμα 17 - LDA - Test set - Fold 2	18
Διάγραμμα 18- LDA - Train set - Fold 3	19
Διάγραμμα 19 - LDA - Test set - Fold 3	19
Διάγραμμα 20 - LDA - Train set - Fold 4	20
Διάγραμμα 21 - LDA - Test set - Fold 4	20
Διάγραμμα 22 - LR - Train set - Fold 1	21
Διάγραμμα 23 - LR - Test set - Fold 1	21
Διάγραμμα 24 - LR - Train set - Fold 2	22
Διάγραμμα 25 - LR - Test set - Fold 2	22
Διάγραμμα 26 - LR - Train set - Fold 3	23
Διάγραμμα 27 - LR - Test set - Fold 3	23
Διάγραμμα 28 - LR - Train set - Fold 4	24
Διάγραμμα 29 - LR - Test set - Fold 4	24
Διάγραμμα 30 - CTree - Train set - Fold 1	25
Διάγραμμα 31 - CTree - Test set - Fold 1	25
Διάγραμμα 32 - CTree - Train set - Fold 2	26
Διάγραμμα 33 - CTree - Test set - Fold 2	26
Διάγραμμα 34 - CTree - Train set - Fold 3	27
Διάγραμμα 35 - CTree - Test set - Fold 3	27
Διάγραμμα 36 - CTree - Train set - Fold 4	28
Διάγραμμα 37 - CTree - Test set - Fold 4	28
Διάγραμμα 38 - RF - Train set - Fold 1	29
Διάγραμμα 39 - RF - Test set - Fold 1	29
Διάγραμμα 40 - RF - Train set - Fold 2	30
Διάγραμμα 41 - RF - Test set - Fold 2	30
Διάγραμμα 42 - RF - Train set - Fold 3	31
Διάγραμμα 43 - RF - Test set - Fold 3	31

Διάγραμμα 44 - RF - Train set - Fold 4	32
Διάγραμμα 45 - RF - Test set - Fold 4	32
Διάγραμμα 46 - kNN – Train set – Fold 1	33
Διάγραμμα 47 - kNN – Test set – Fold 1	33
Διάγραμμα 48 - kNN – Train set – Fold 2	34
Διάγραμμα 49 - kNN – Test set – Fold 2	34
Διάγραμμα 50 - kNN – Train set – Fold 3	35
Διάγραμμα 51 - kNN – Test set – Fold 3	35
Διάγραμμα 52 - kNN – Train set – Fold 4	36
Διάγραμμα 53 - kNN – Test set – Fold 4	36
Διάγραμμα 54 - NV – Train set – Fold 1	37
Διάγραμμα 55 - NV – Test set – Fold 1	37
Διάγραμμα 56 - NV – Train set – Fold 2	38
Διάγραμμα 57 - NV – Test set – Fold 2	38
Διάγραμμα 58 - NV – Train set – Fold 3	39
Διάγραμμα 59 - NV – Test set – Fold 3	39
Διάγραμμα 60 - NV – Train set – Fold 4	40
Διάγραμμα 61 - NV – Test set – Fold 4	40
Διάγραμμα 62 - SVM – Train set – Fold 1	41
Διάγραμμα 63 - SVM – Test set – Fold 1	41
Διάγραμμα 64 - SVM – Train set – Fold 2	42
Διάγραμμα 65 - SVM – Test set – Fold 2	42
Διάγραμμα 66 - SVM – Train set – Fold 3	43
Διάγραμμα 67 - SVM – Test set – Fold 3	43
Διάγραμμα 68 - SVM – Train set – Fold 4	44
Διάγραμμα 69 - SVM – Test set – Fold 4	44
Διάγραμμα 70 - GB – Train set – Fold 1	45
Διάγραμμα 71 - GB – Test set – Fold 1	45
Διάγραμμα 72 - GB – Train set – Fold 2	46
Διάγραμμα 73 - GB – Test set – Fold 2	46
Διάγραμμα 74 - GB – Train set – Fold 3	47
Διάγραμμα 75 - GB – Test set – Fold 3	47
Διάγραμμα 76 - GB – Train set – Fold 4	48
Διάγραμμα 77 - GB – Test set – Fold 4	48
Διάγραμμα 78 - Accuracy Test set - 4 fold Average	49
Διάγραμμα 79 – AUC-ROC - Test set - 4 fold Average	49
Διάγραμμα 80 - Precision Test set - 4 fold Average	50
Διάγραμμα 81 - Prevalence Test set - 4 fold Average	50
Διάγραμμα 82 - Recall Test set - 4 fold Average	51
Διάγραμμα 83 - FDR Test set - 4 fold Average	51

1. Εισαγωγή

Το παρόν έγγραφο αποτελεί αναφορά για την εργασία πάνω στην Επιβλεπόμενη μάθηση (Supervised learning) και την Ταξινόμηση (Classification) στα πλαίσια του μαθήματος Μέθοδοι και Εργαλεία Τεχνητής Νοημοσύνης. Σκοπός της εργασίας είναι να βρεθεί ποιο είναι το καλύτερο δυνατό μοντέλο ταξινόμησης στο πρόβλημα του εντοπισμού εταιρειών που θα κηρύξουν χρεωκοπία, παραθέτοντας συγκριτικά αποτελέσματα. Προκειμένου να μπορεί να απαντηθεί αυτό με τεκμηρίωση, ακολουθήθηκε συγκεκριμένη μεθοδολογία που αναλύεται παρακάτω.

2. Μεθοδολογία

Για την εξαγωγή του αποτελέσματος χρειάστηκε να ακολουθηθούν συγκεκριμένα βήματα. Ως δεδομένο υπήρχε διαθέσιμο ένα dataset σε μορφή .xlsx που περιέχει δεδομένα με δείκτες απόδοσης εταιρειών, δυαδικούς δείκτες δραστηριοτήτων, την κατάσταση της εταιρείας (υγιής ή σε χρεωκοπία) και το έτος αναφοράς. Για την χρήση των δεδομένων γράφτηκε κώδικας σε python στην πλατφόρμα Google Colab, και χρησιμοποιήθηκαν διάφορες βιβλιοθήκες για τον χειρισμό των δεδομένων, την οπτικοποίησή τους σε γραφήματα και την εκπαίδευση των μοντέλων μηχανικής μάθησης, όπως pandas, numpy, sklearn, matplotlib.

Αρχικά χωρίστηκαν τα δεδομένα σε 2 κλάσεις, 1 για τις υγιείς εταιρείες και 1 για τις χρεωκοπημένες. Στη συνέχεια, και αφού ελέγχθηκε το dataset για τυχόν ελλείψεις εγγραφές (NaN values), δημιουργήθηκαν θηκογράμματα (boxplots) για κάθε δείκτη απόδοσης, ανάλογα με την κατάσταση της εταιρείας, για να υπάρχει μια πρώτη εικόνα της κατανομής των τιμών, προς καλύτερη κατανόηση του dataset.

Τα μοντέλα μηχανικής μάθησης που θα αξιολογηθούν είναι:

- Linear Discriminant Analysis
- Logistic Regression
- Decision Trees
- Random Forests
- K Nearest Neighbors
- Naïve Bayes
- Support Vector Machines
- Gradient Boosting Classifier (προσωπική επιλογή)

Για την αξιολόγηση των μοντέλων χρειάζεται τα μοντέλα αυτά να εκπαιδευτούν. Για την εκπαίδευση τους χρησιμοποιήθηκε η τεχνική Stratified kfold, με την αξιοποίηση της αντίστοιχης βιβλιοθήκης (sklearn). Με το Stratified kfold διατηρείται η αναλογία ανάμεσα στις δύο κλάσεις σε κάθε fold. Προκειμένου να είναι αποτελεσματική η αξιολόγηση όλα τα δεδομένα των δεικτών απόδοσης του dataset κανονικοποιήθηκαν χρησιμοποιώντας Min Max Scaler. Η κανονικοποίηση είναι απαραίτητη προκειμένου τα δεδομένα των δεικτών να έρθουν σε μια ενιαία κλίμακα για να γίνει σωστά η αξιολόγηση των μοντέλων.

Η εκπαίδευση των μοντέλων έγινε με 4 folds. Για κάθε fold, επιχειρήθηκε να γίνει εξισορρόπηση των δεδομένων τόσο στο training όσο και στο test set, στην περίπτωση που η κατανομή ξεπερνούσε τις 3 υγιείς εταιρείες για κάθε 1 χρεωκοπημένη. Έπειτα εκπαιδεύτηκαν τα μοντέλα, σχεδιάστηκαν τα αντίστοιχα confusion matrices και με βάσει αυτά, υπολογίζεται η επίδοσή τους τόσο στο training όσο και στο test set, βάσει των μετρικών Accuracy, Precision, Recall, F1 score και AUC-ROC.

Για κάθε train/test set, για κάθε μοντέλο, για κάθε fold set, εξάγουμε πληροφορίες για τον υπολογισμό των ακόλουθων μετρικών :

- Accuracy
- Precision
- Recall
- F1 score
- AUC-ROC

Τέλος, αφού υπολογιστούν οι παραπάνω μετρικές, μπορούμε να δούμε ποιο μοντέλο αποδίδει καλύτερα βάση του F1 score.

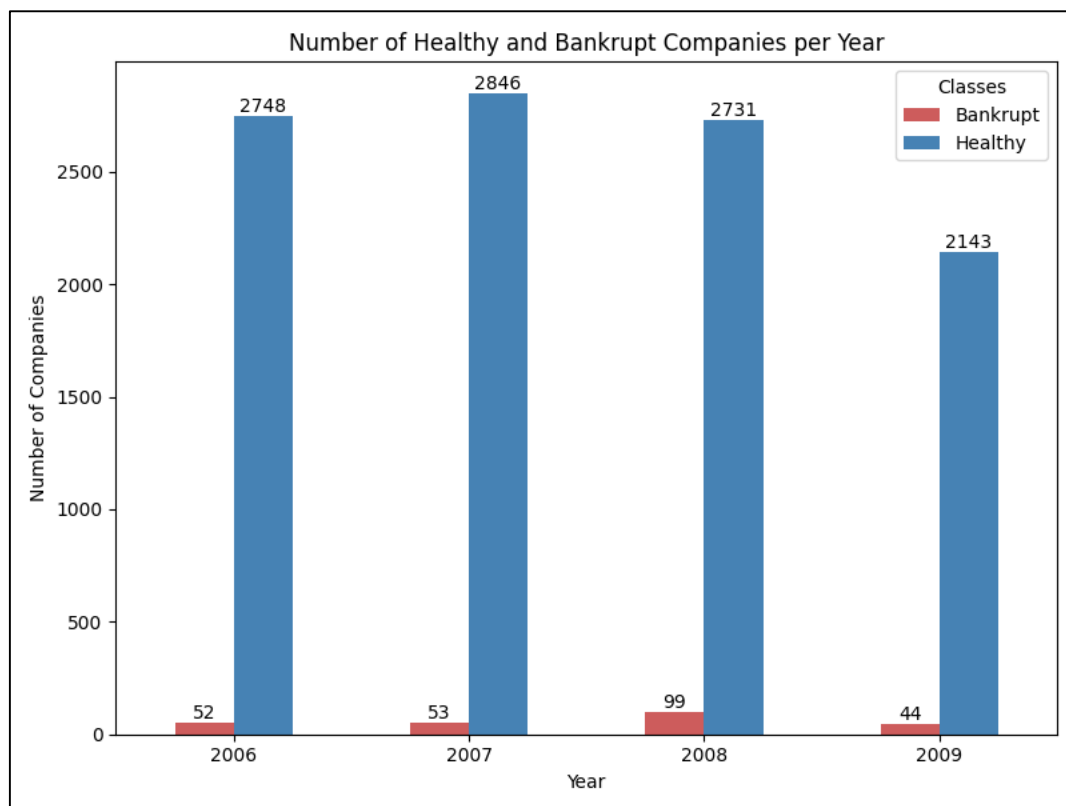
Στην επόμενη ενότητα, αναλύεται το dataset και παραθέτονται τα αποτελέσματα που προκύπτουν σύμφωνα με την παραπάνω μεθοδολογία.

3. Πειραματικά Αποτελέσματα

3.1 Περιγραφή Dataset

Το dataset που χρησιμοποιήθηκε είναι μορφής .xlsx και περιέχει δείκτες απόδοσης για εταιρείες, δυαδικούς δείκτες δραστηριοτήτων, την κατάσταση της εταιρείας (υγιής/χρεωκοπημένη) και το έτος αναφοράς.

Ως πρώτο βήμα μεθοδολογίας είναι ο διαχωρισμός του dataset σε 2 κλάσεις, 1. Υγιής εταιρεία (Healthy) 2. Χρεωκοπημένη (Bankrupt). Αυτό φαίνεται από την στήλη L του αρχείου .xlsx, όπου αν η εταιρεία έχει τιμή 1 είναι υγιής και αν έχει την τιμή 2 είναι χρεωκοπημένη. Ακολουθεί το ραβδόγραμμα που δείχνει τον διαχωρισμό των κλάσεων ανά έτος :



Διάγραμμα 1 - Υγιείς και Χρεωκοπημένες εταιρείες ανά έτος

Παρατηρούμε ότι το dataset δεν είναι ισορροπημένο, καθώς για κάθε έτος αναφοράς υπάρχουν πολύ περισσότερες υγιείς εταιρείες έναντι των χρεωκοπημένων. Αυτό δεν μας εξυπηρετεί και στην συνέχεια θα επιχειρήσουμε να κάνουμε undersampling προκειμένου να έρθει σε μια ισορροπία για την εκπαίδευση των μοντέλων.

Για κάθε δείκτη απόδοσης που υπάρχει στο dataset, με διαχωρισμό των υγιών και των χρεωκοπημένων εταιρειών σχεδιάστηκαν boxplots που βρίσκονται στο [Παράρτημα 1](#). Στα boxplots επισημαίνονται οι μέγιστες (max), οι ελάχιστες (min) και οι MO (avg) των τιμών τους.

3.2 Εκπαίδευση Μοντέλων

Για την εκπαίδευση των μοντέλων τα δεδομένα κανονικοποιούνται όπως ειπώθηκε παραπάνω. Για τον διαχωρισμό σε folds χρησιμοποιήθηκε η τεχνική k Fold Cross Validation με 4 folds, τεχνική που διατηρεί την αναλογία των κλάσεων σε κάθε fold.

Σε κάθε fold set τα 3 folds αποτελούν το training set και το 1 fold το test set. Το σύνολο των δειγμάτων μας είναι 10.716, οπότε σε κάθε fold set υπάρχουν 8.037 δείγματα και σε κάθε test set 2.679 δείγματα. Βάσει του Διαγράμματος 1, παρατηρούμε ότι το δείγμα μας δεν είναι καθόλου ισορροπημένο, έτσι για κάθε fold set έγινε εξισορρόπηση στο training set έτσι ώστε η αναλογία να είναι 3 Υγιείς προς 1 Χρεωκοπημένη. Προτιμήθηκε να γίνει undersampling της κλάσης των Υγιών εταιρειών αφαιρώντας τυχαία τιμές μέχρι να ισχύει η αναλογία των 3:1. Η εξισορρόπηση είναι απαραίτητη για να αποφύγουμε την προκατάληψη των μοντέλων υπέρ της μεγάλης κλάσης, στην περίπτωση μας των Υγιών εταιρειών. Επισημαίνεται ότι δεν

Έτσι, μετά την εξισορρόπηση προσχωρήσαμε στην εκπαίδευση των παρακάτω μοντέλων :

- Linear Discriminant Analysis
- Logistic Regression
- Decision Trees
- Random Forests
- K Nearest Neighbors
- Naïve Bayes
- Support Vector Machines
- Gradient Boosting Classifier (προσωπική επιλογή)

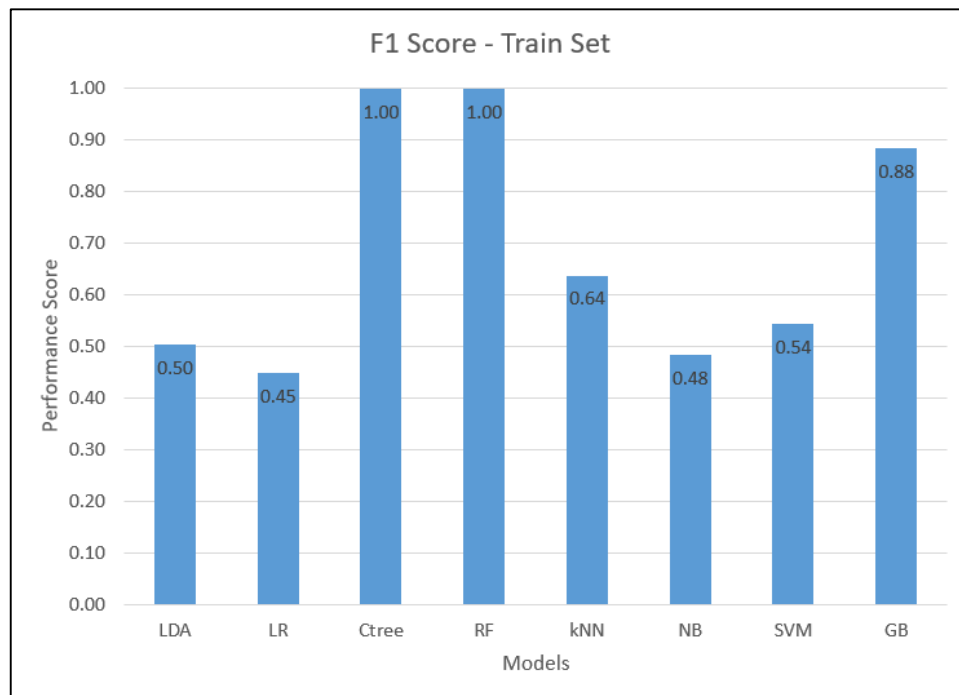
Όλα τα μοντέλα εκπαιδεύτηκαν 4 φορές, όσες και τα folds, και σε κάθε πέρασμα εκτυπώθηκαν τα Confusion matrices τόσο για το train όσο και για το test set. Σημειώνεται ότι το Positive στην εξεταζόμενη περίπτωση είναι μια εταιρεία που έχει κηρύξει χρεοκοπία. Στο [Παράρτημα 2](#) βρίσκονται τα Confusion Matrices για κάθε πέρασμα όλων των μοντέλων.

3.3 Σύγκριση Μοντέλων

Με την δημιουργία των confusion matrices υπολογίστηκαν επιπλέον και οι μετρικές :

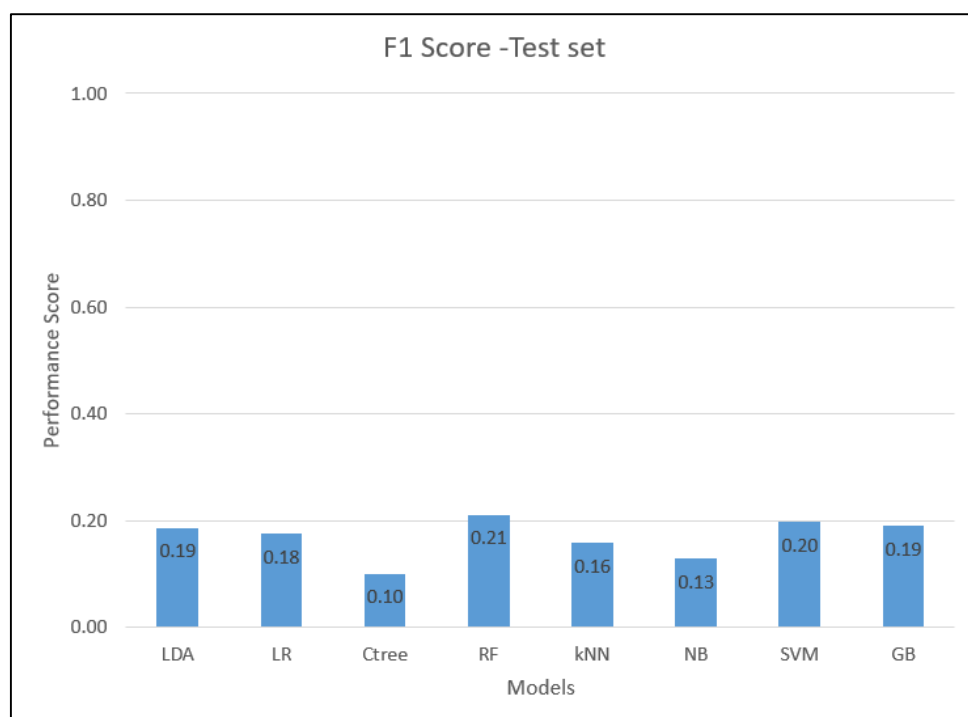
- Accuracy
- Precision
- Recall
- F1-Score
- Prevalence
- False Discovery Rate

Βάσει της εκφώνησης τα μοντέλα συγκρίνονται με βάση το F1 Score. Ακολουθεί γράφημα που δείχνει την επίδοση των μοντέλων για το train και το test set. Επειδή είχαμε 4 folds, το F1 Score κάθε μοντέλου καθορίζεται από τον Μ.Ο. .



Διάγραμμα 2 - Επίδοση μοντέλων στο Train set

Παρατηρούμε ότι στο train set τα μοντέλα Decision Trees και Random Forest σημείωσαν την καλύτερη επίδοση. Αυτό οφείλεται και στο γεγονός ότι τα train set εξισορροπήθηκαν έτσι ώστε να διατηρούν την αναλογία 3:1. **Ωστόσο για την σωστή αξιολόγηση των μοντέλων μας ενδιαφέρει η επίδοσή τους στο train set, όπως φαίνεται στο ακόλουθο γράφημα.**



Διάγραμμα 3 - Επίδοση μοντέλων στο Test set

Σύμφωνα με το Διάγραμμα 3, κανένα μοντέλο δεν κατάφερε να έχει επίδοση παραπάνω από 0,21, με το **Random Forests** να πετυχαίνει 0,21 σημειώνοντας την καλύτερη επίδοση. Το F1 Score αποτελεί τον αρμονικό ΜΟ μεταξύ του Precision, το οποίο δηλώνει το πόσες από τις positive προβλέψεις είναι όντως positive, και του Recall, το οποίο δηλώνει πόσες από τις πραγματικά positive περιπτώσεις εντοπίστηκαν σωστά. Συνεπώς, ο χαμηλό F1 Score οφείλεται σε χαμηλές τιμές του precision ή του recall ή και των δύο, κάτι που επιβεβαιώνεται και από τα αποτελέσματα τα οποία φαίνονται στα διαγράμματα του [Παραρτήματος 3](#). Αυτό οφείλεται στο γεγονός ότι το test set **δεν είναι ισορροπημένο** και τα μοντέλα έχουν την τάση να ευνοούν την «ισχυρή» κλάση.

3.4 Έλεγχος περιορισμών

Για την αξιολόγηση των μοντέλων μας τέθηκε το ερώτημα στο αν πρέπει να πληρούνται οι εξής περιορισμοί :

(α) Το μοντέλο πρέπει να βρίσκει με ποσοστό επιτυχίας τουλάχιστον 60% τις εταιρείες που θα χρεωκοπήσουν.

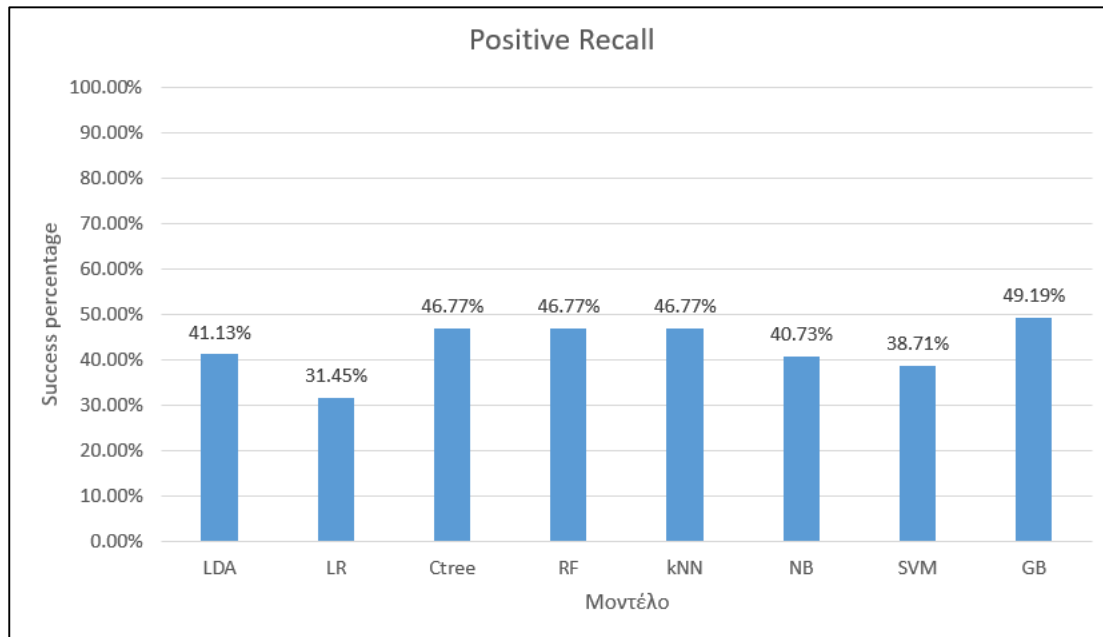
Πρακτικά αυτό σημαίνει πως το μοντέλο πρέπει να έχει $Recall > 0,60$, δηλαδή

$Recall = \frac{TP}{TP+FN}$, μιας και το positive στην περίπτωση μας είναι οι χρεωκοπημένες εταιρείες.

(β) Το μοντέλο πρέπει να βρίσκει με ποσοστό επιτυχίας τουλάχιστον 70% τις εταιρείες που δεν θα χρεωκοπήσουν.

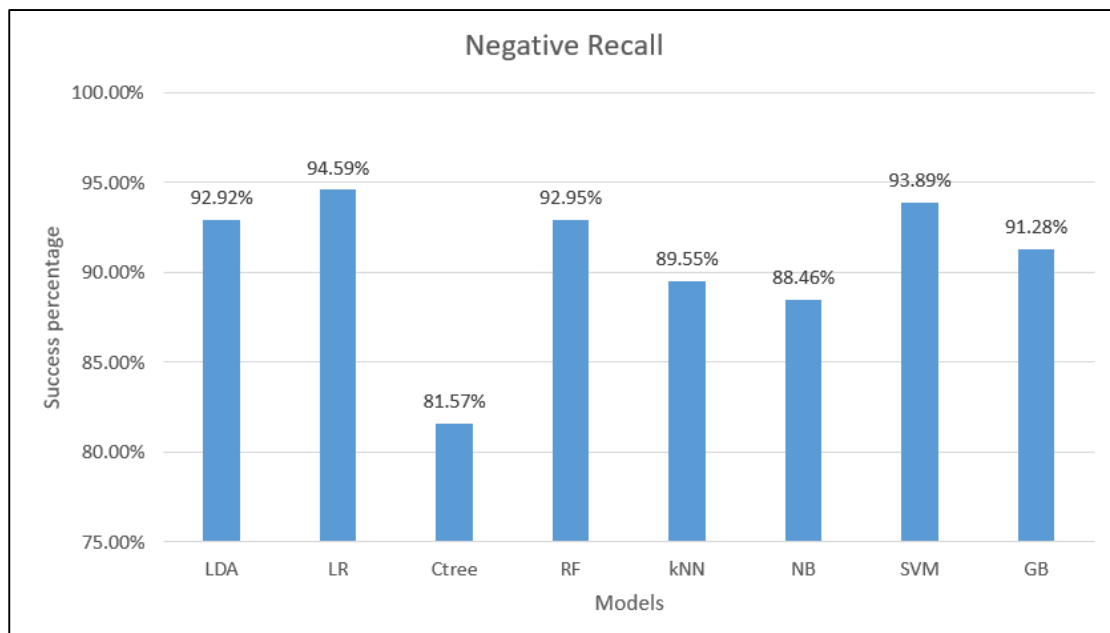
Πρακτικά αυτό σημαίνει πως το μοντέλο πρέπει να έχει $Recall$ (των υγιών) $> 0,70$ δηλαδή $Recall = \frac{TN}{TN+FP}$, μιας και το negative στην περίπτωση μας είναι οι υγιείς εταιρείες.

Οι περιορισμοί πρέπει να ισχύουν στην απόδοση ενός μοντέλου καθώς σε περίπτωση που δεν πληρούνται, δηλώνει πως το μοντέλο μας αδυνατεί να προβλέψει με μεγάλη αξιοπιστία αν μια εταιρεία δείχνει να είναι προς χρεωκοπία ή όχι. Ο έλεγχος περιορισμών θα γίνει, όπως και πριν, μόνο στο test set. Ακολουθούν τα γραφήματα που δείχνουν την επίδοση των εξεταζόμενων μοντέλων στους παραπάνω περιορισμούς.



Διάγραμμα 4 – Positive Recall (χρεωκοπημένων)

Σύμφωνα με το παραπάνω διάγραμμα, παρατηρούμε ότι ο (α) περιορισμός αποτυγχάνει. Κανένα από τα μοντέλα δεν καταφέρνει να εντοπίσει τουλάχιστον το 60% των εταιρειών που χρεωκοπούν, κάτι που μας δείχνει ότι τα μοντέλα δεν είναι επαρκώς ευαίσθητα στην κρίσιμη κλάση, που στην περίπτωση μας είναι οι χρεωκοπημένες εταιρείες.



Διάγραμμα 5 – Negative Recall (Υγιών)

Σύμφωνα με το παραπάνω διάγραμμα, παρατηρούμε ότι όλα τα μοντέλα πληρούν τον (β) περιορισμό, καθώς με επιτυχία βρίσκουν τουλάχιστον το 70% των εταιρειών που

δεν θα χρεοκοπήσουν. Αυτό μας δείχνει ότι η πλειοψηφία των δειγμάτων βρίσκονται σε αυτήν την κλάση και τα μοντέλα ευνοούν την πλειοψηφία.

4. Συμπεράσματα

Η παρούσα αναφορά αξιολόγησε συγκεκριμένα μοντέλα πάνω στην δυνατότητα πρόβλεψης χρεωκοπίας ή όχι εταιρειών, χρησιμοποιώντας ένας dataset που είχε οικονομικούς δείκτες απόδοσης.

Οι δοκιμές που πραγματοποιήθηκαν μας δείχνουν ότι το δείγμα μας είναι εξαιρετικά ανισόρροπο, κάτι που προκαλεί τα μοντέλα να έχουν προκατάληψη προς την «ισχυρή» κλάση. Αφού έγινε εξισορρόπηση των train sets, παρατηρήθηκε ότι με βάση το F1 score τα μοντέλα Decision Trees, Random Forests και Gradient Boosting Classifier σημείωσαν την καλύτερη επίδοση. Ωστόσο, η αξιολόγηση των μοντέλων γίνεται στο test set, το οποίο δεν είναι ισορροπημένο, και φάνηκε ότι κανένα μοντέλο δεν κατάφερε να πετύχει υψηλό F1 score. **Την καλύτερη επίδοση με βάση το F1 Score σημείωσε το μοντέλο Random Forests με F1 score 0.21 .**

Επιπλέον, **κανένα από τα μοντέλα δεν πληροί τον περιορισμό (α)**, δηλαδή την ικανότητα να βρίσκει τουλάχιστον το 60% των χρεωκοπημένων εταιρειών. Αυτό υποδηλώνει ότι τα μοντέλα αποτυγχάνουν να εντοπίσουν επαρκώς τις προβληματικές εταιρείες. Τα μοντέλα που σημείωσαν το μεγαλύτερο score είναι Gradient Boosting Classifier, Decision Trees, Random Forests και k Nearest Neighbours.

Αντιθέτως, **όλα τα μοντέλα πληρούν τον περιορισμό (β)**, καθώς επιτυγχάνουν να βρουν τουλάχιστον το 70% υγιών εταιρειών. Τα μοντέλα που σημείωσαν το μεγαλύτερο score είναι Linear Discriminant Analysis, Random Forests και Support Vector Machines.

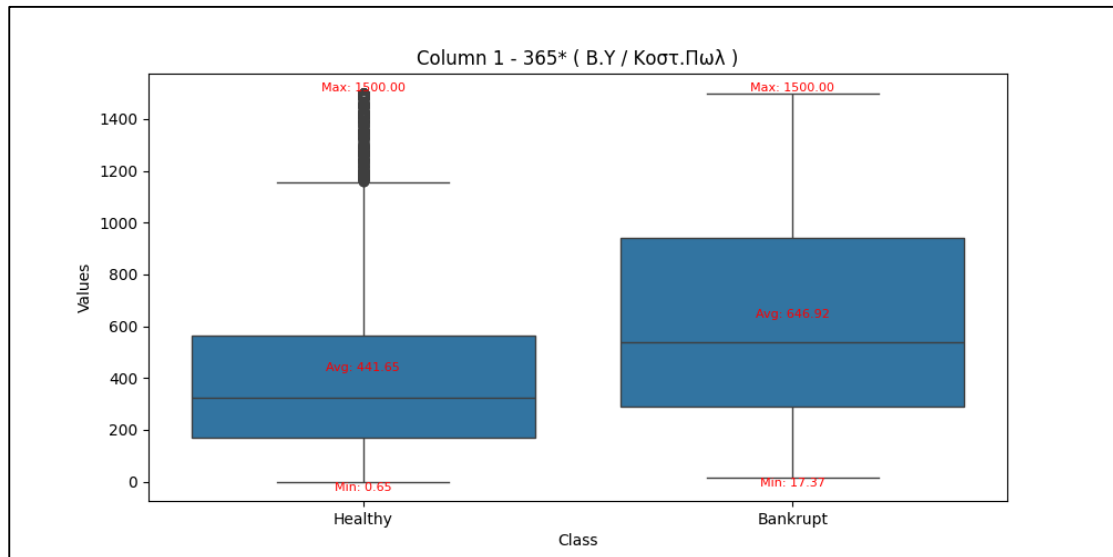
Και οι δύο περιορισμοί οφείλονται στην ανισορροπία των 2 κλάσεων, με αποτέλεσμα τα μοντέλα να έχουν την τάση προς την «ισχυρή» κλάση

Συνολικά, το καλύτερο δυνατό μοντέλο αποτελεί το **Random Forests** υπό τις υπάρχουσες συνθήκες.

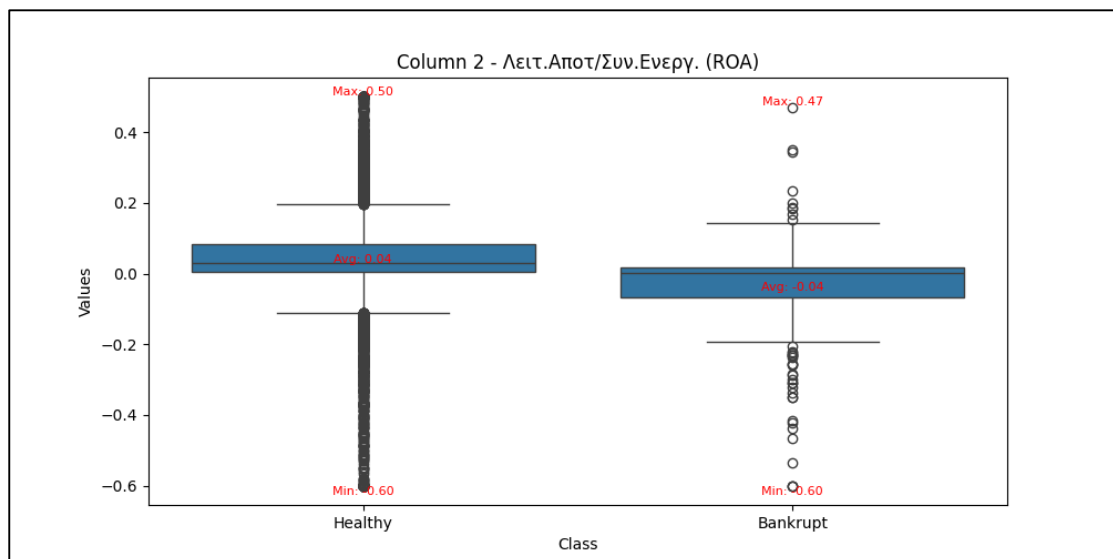
5. Βιβλιογραφία

- [1]. https://en.wikipedia.org/wiki/Confusion_matrix
- [2]. Διαφάνειες διαλέξεων μαθήματος «Μέθοδοι και Εργαλεία Τεχνητής Νοημοσύνης»
- [3]. Παραδείγματα κώδικα πάνω σε προβλήματα Classification του μαθήματος «Μέθοδοι και Εργαλεία Τεχνητής Νοημοσύνης»

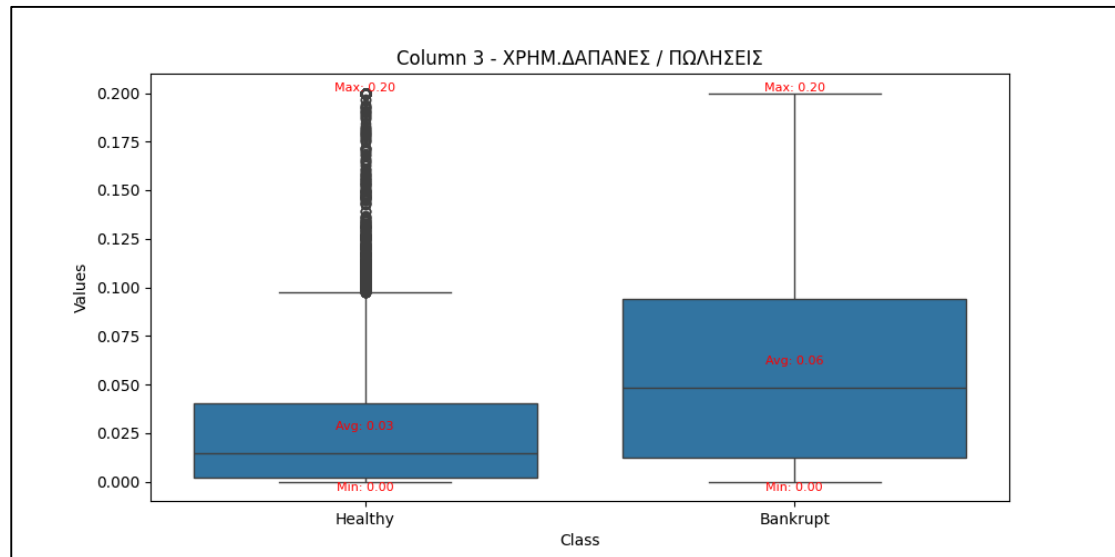
Παράρτημα 1 – Boxplots



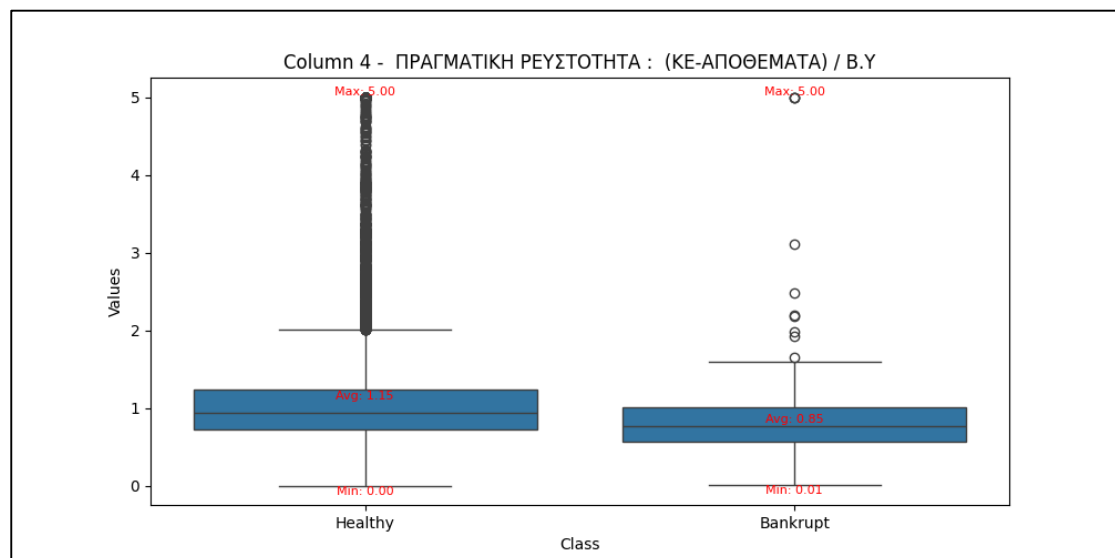
Διάγραμμα 6 - Κατανομή δείκτη απόδοσης 365 * (BT/Κόστος Πωλήσεων)



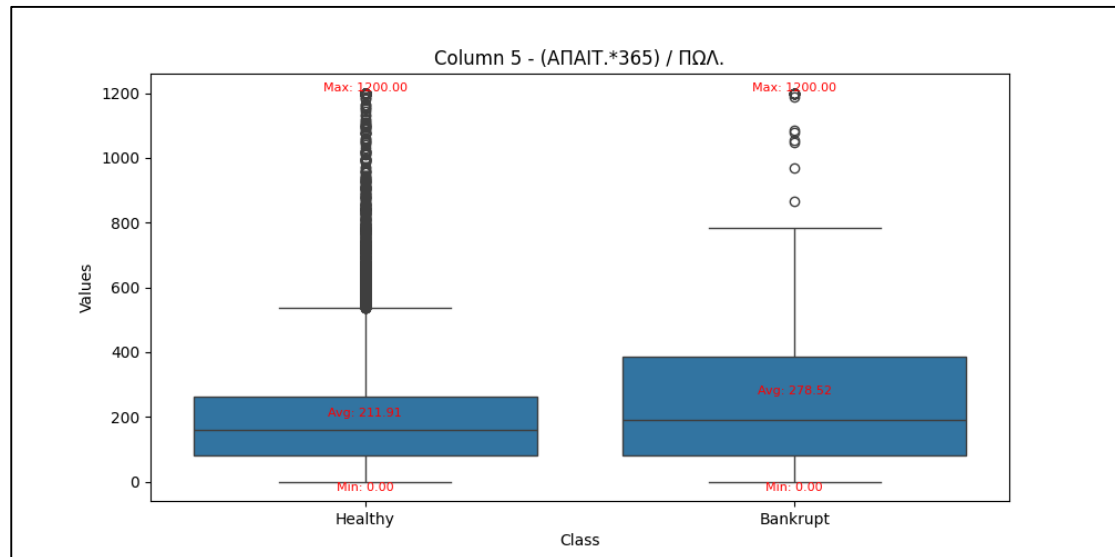
Διάγραμμα 7 - Κατανομή δείκτη Λειτουργικά Αποτελέσματα / Συν. Ενεργητικό



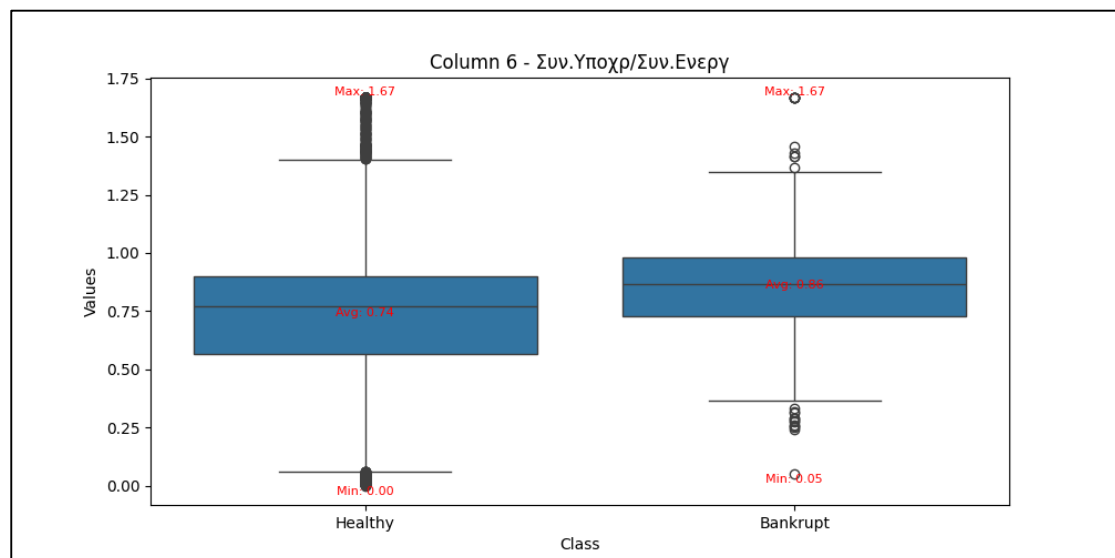
Διάγραμμα 8 - Κατανομή δείκτη Χρηματοοικονομικές Δαπάνες / Πωλήσεις



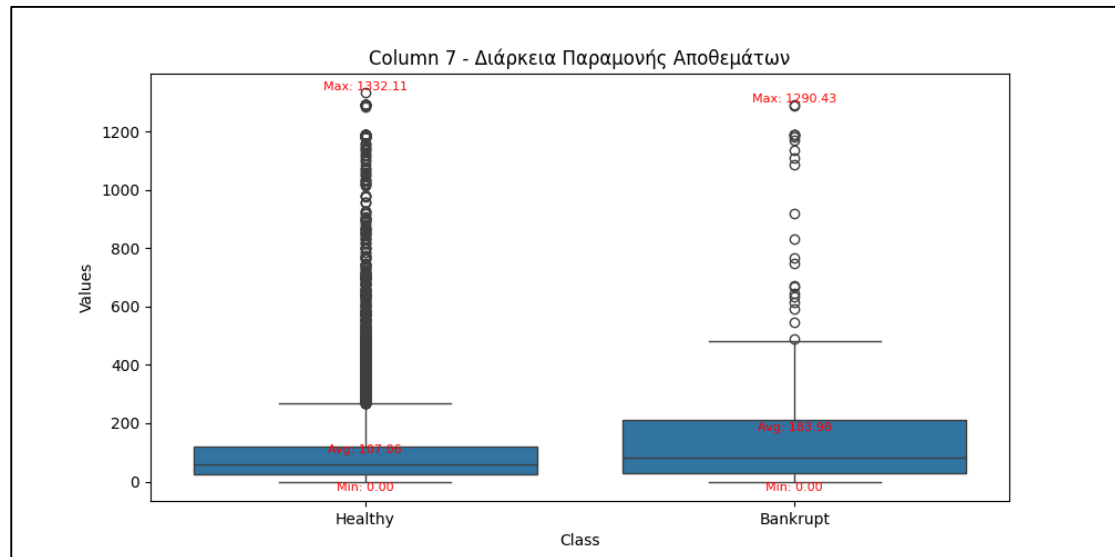
Διάγραμμα 9 - Κατανομή δείκτη Πραγματική Ρευστότητα



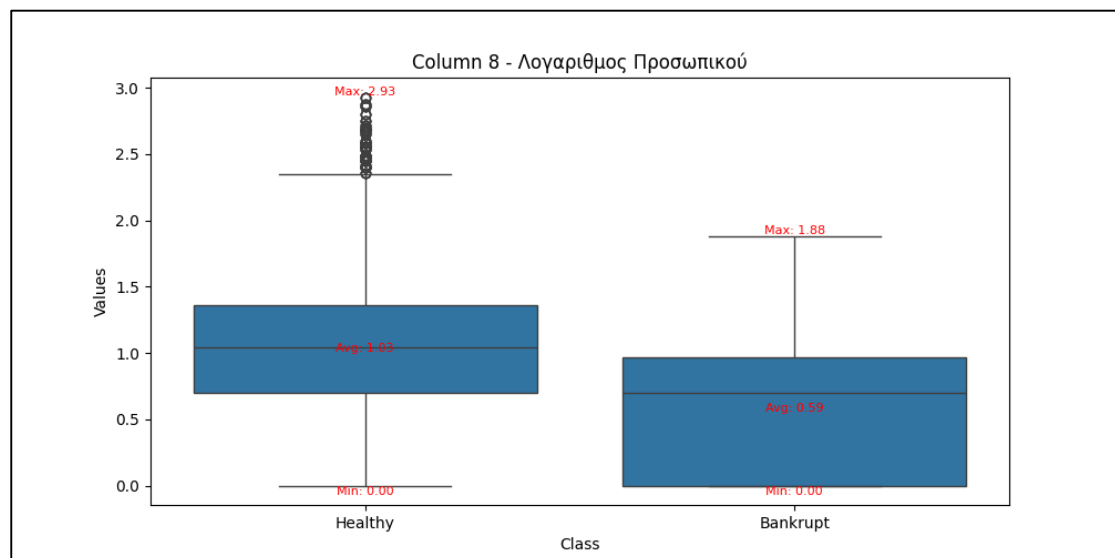
Διάγραμμα 10 - Κατανομή δείκτη (Εμπορικές Απαιτήσεις * 365) / Πωλήσεις



Διάγραμμα 11 - Κατανομή δείκτη Σύνολο Υποχρεώσεων / Σύνολο Ενεργητικού

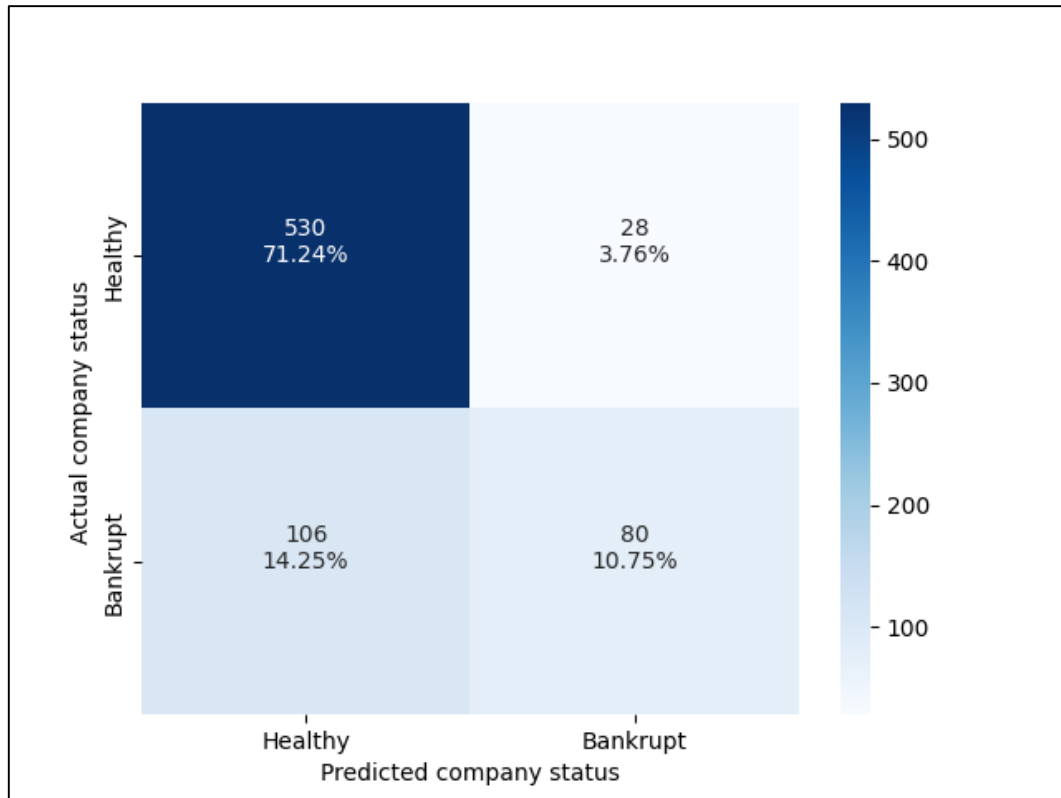


Διάγραμμα 12 - Κατανομή δείκτη Διάρκεια Παραμονής Αποθεμάτων

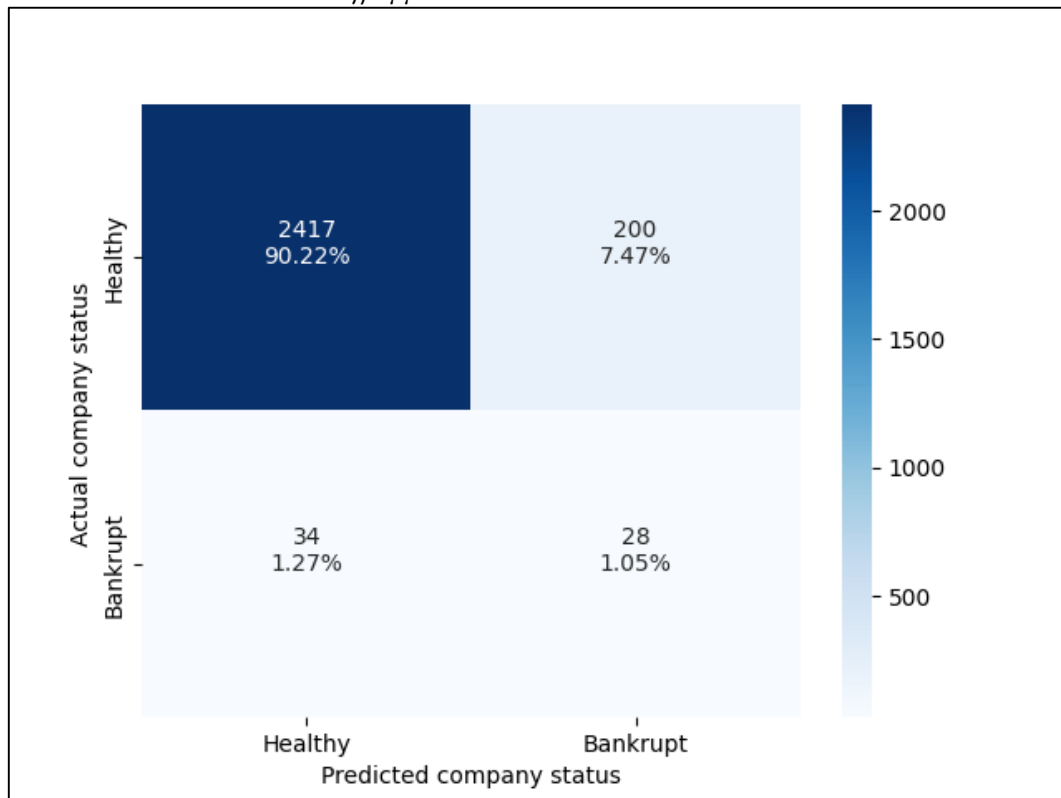


Διάγραμμα 13 - Κατανομή δείκτη Λογάριθμος Προσωπικού

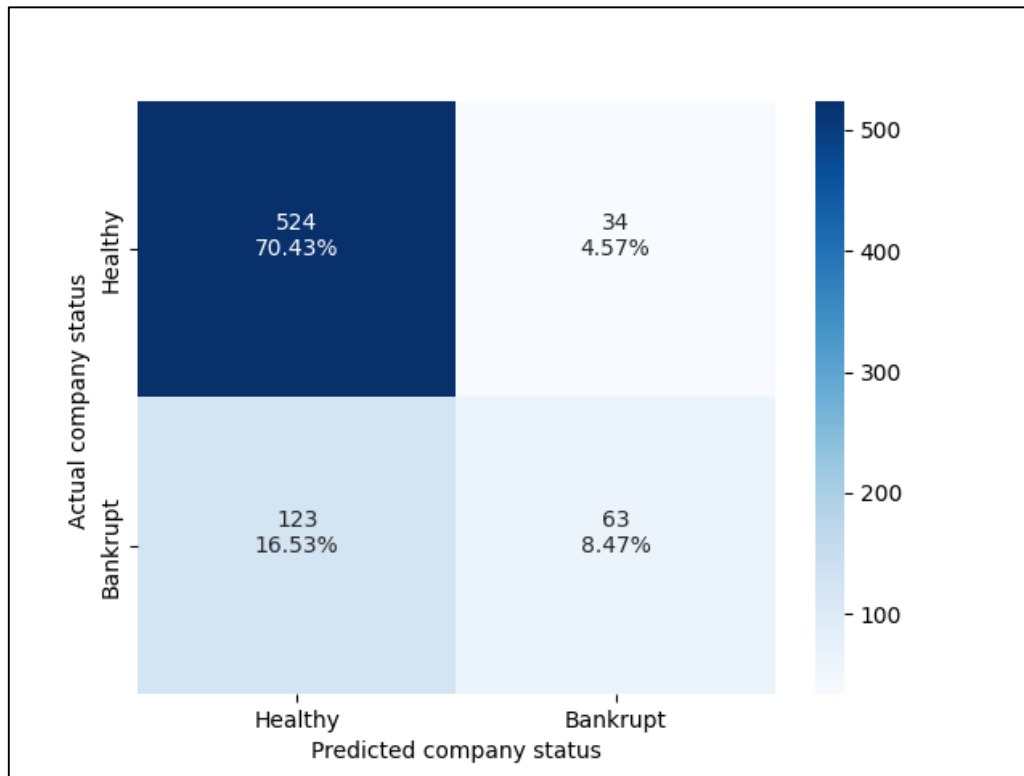
Παράρτημα 2 – Confusion Matrices



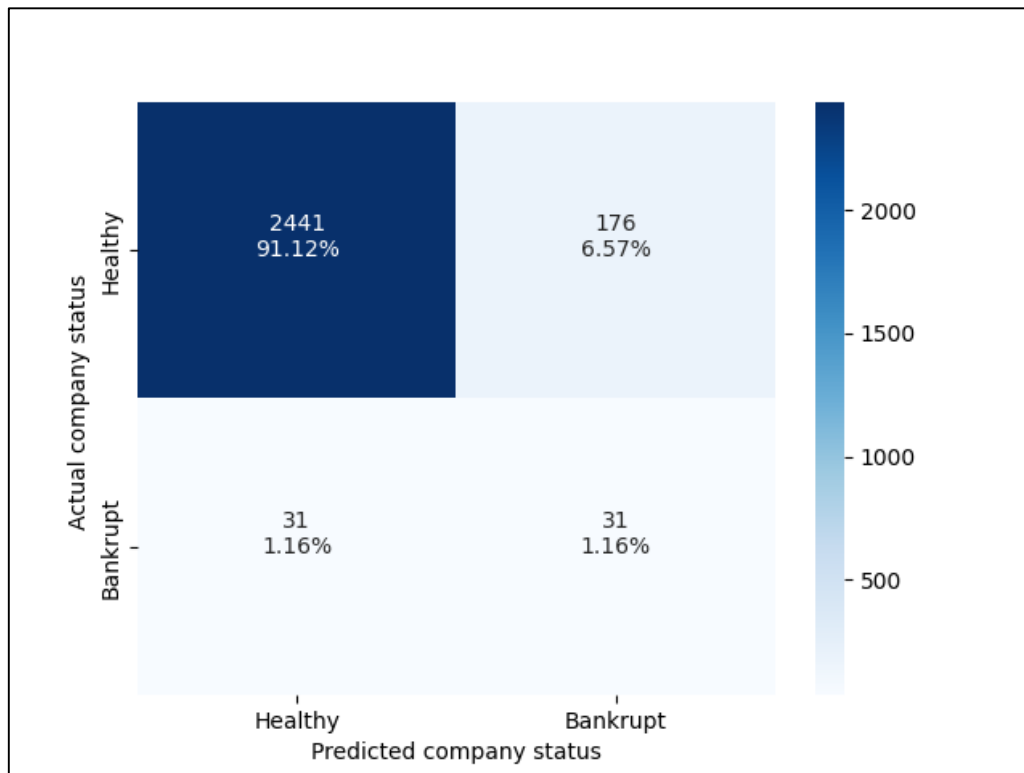
Διάγραμμα 14 - LDA - Train set-Fold 1



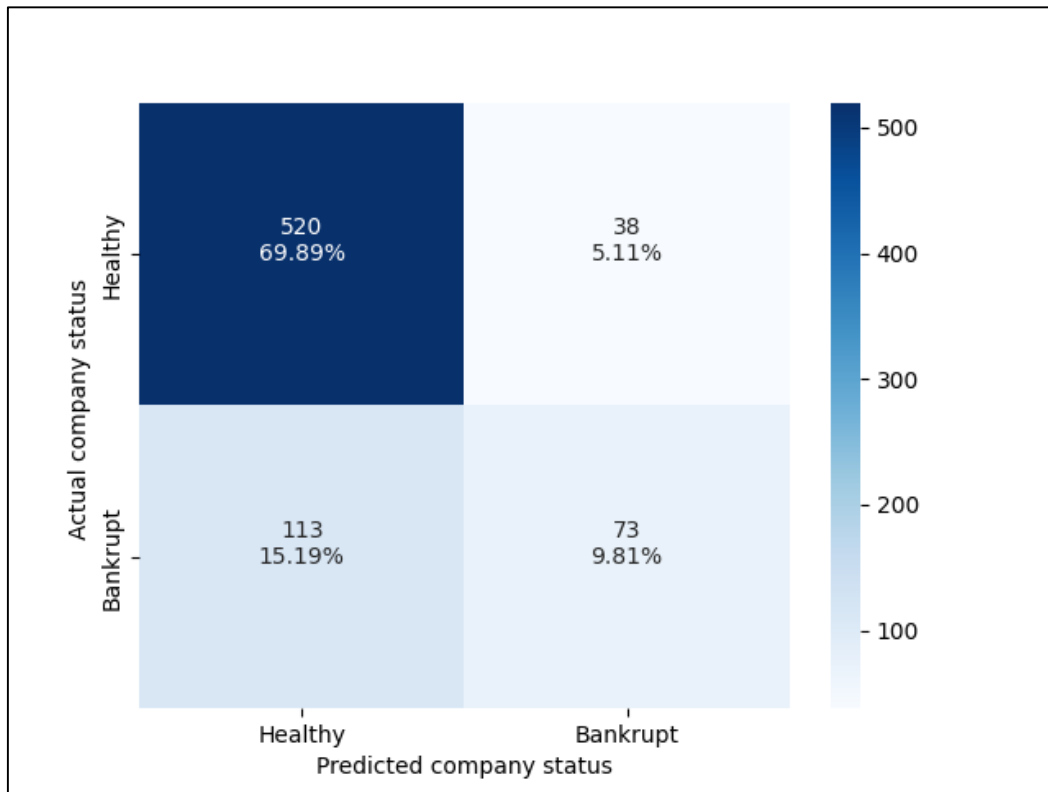
Διάγραμμα 15 - LDA - Test set - Fold 1



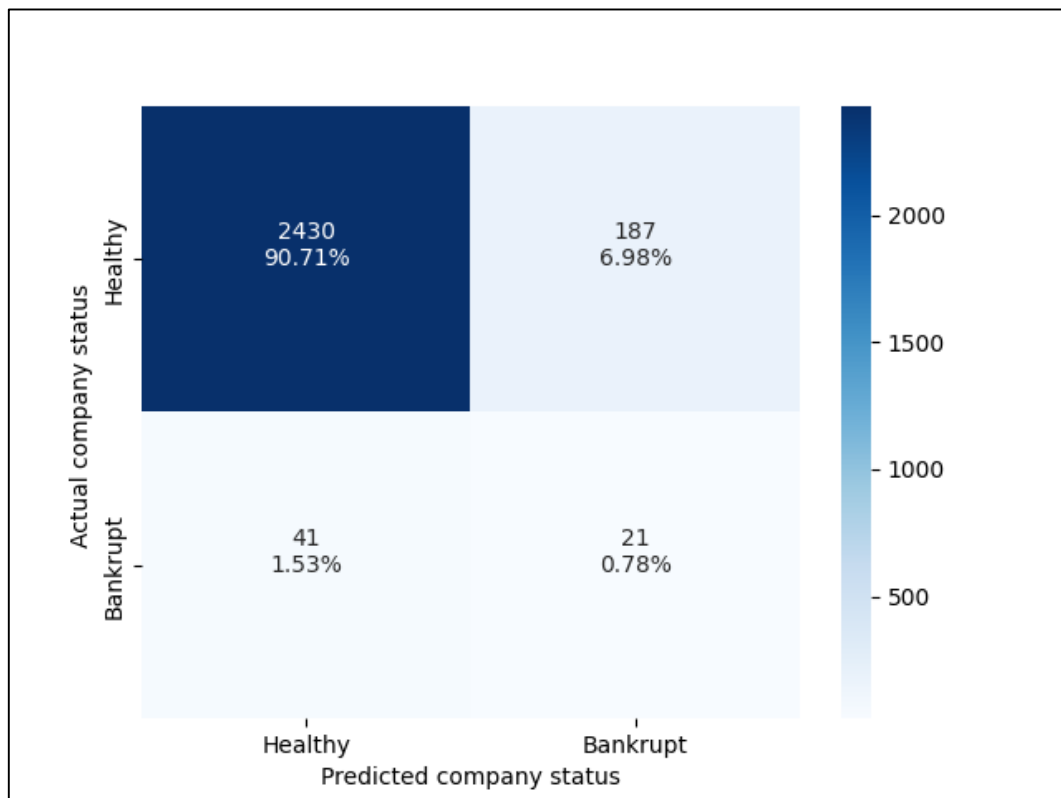
Διάγραμμα 16 - LDA - Train set - Fold 2



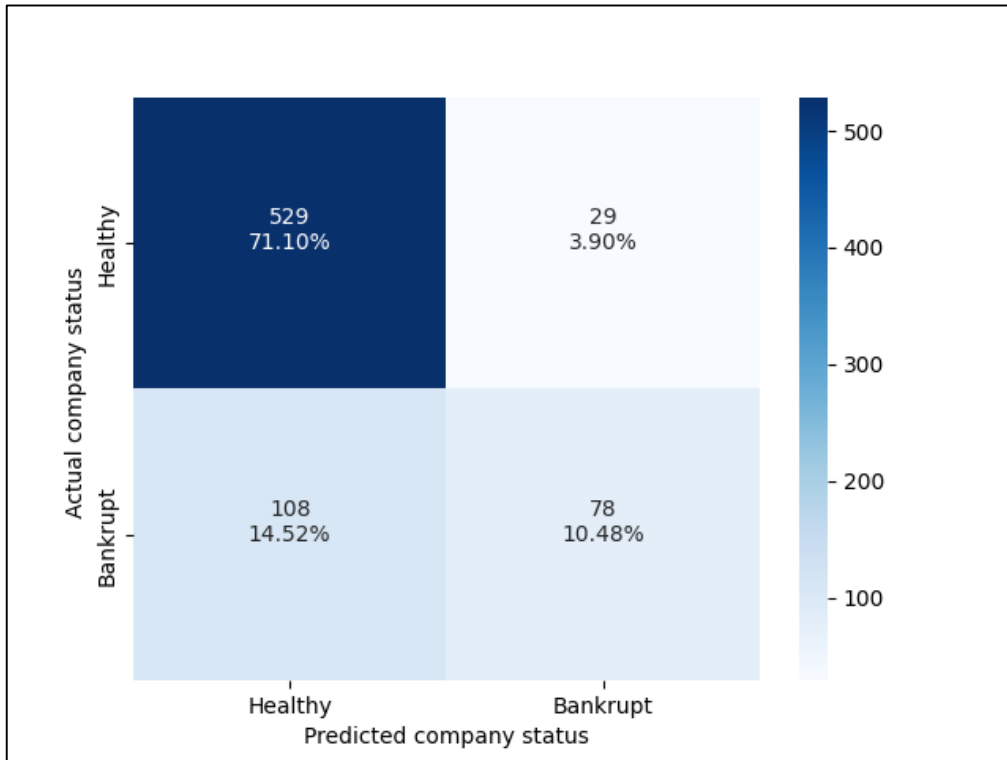
Διάγραμμα 17 - LDA - Test set - Fold 2



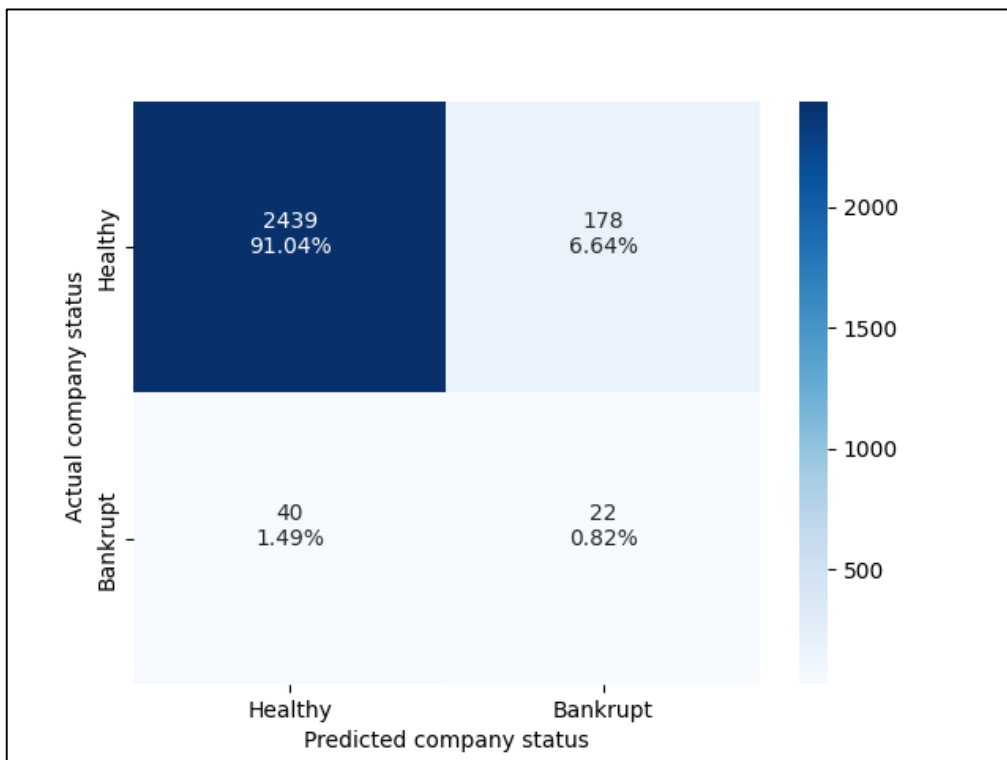
Διάγραμμα 18- LDA - Train set - Fold 3



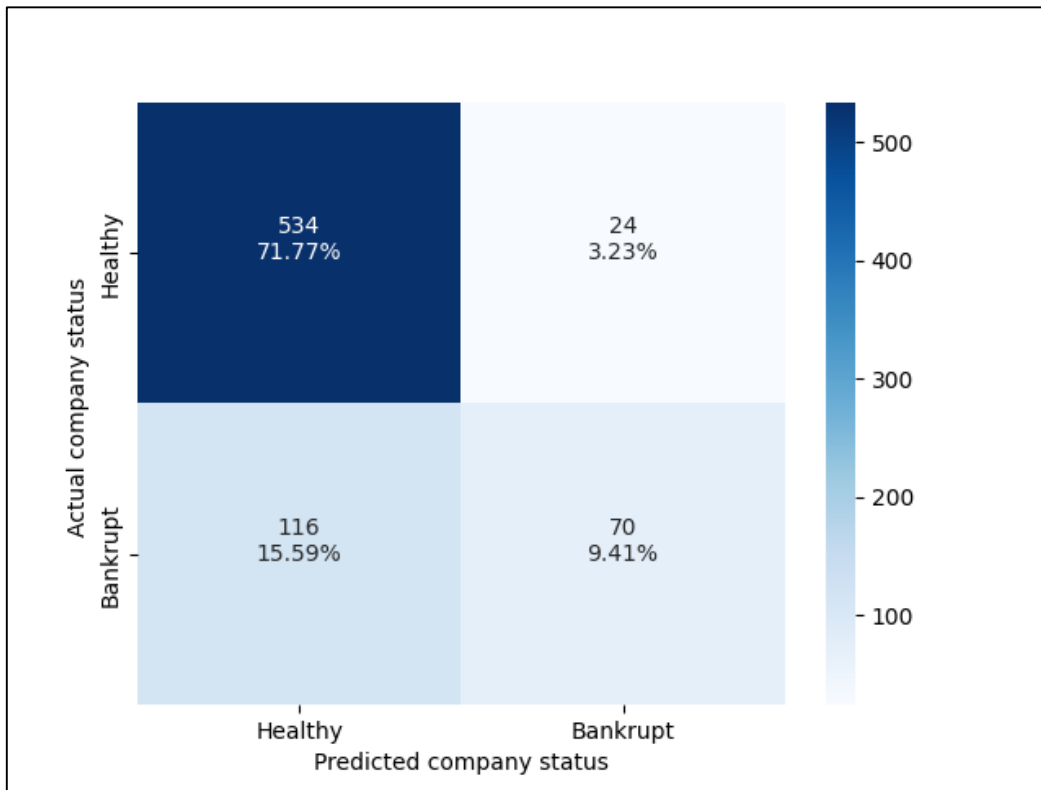
Διάγραμμα 19 - LDA - Test set - Fold 3



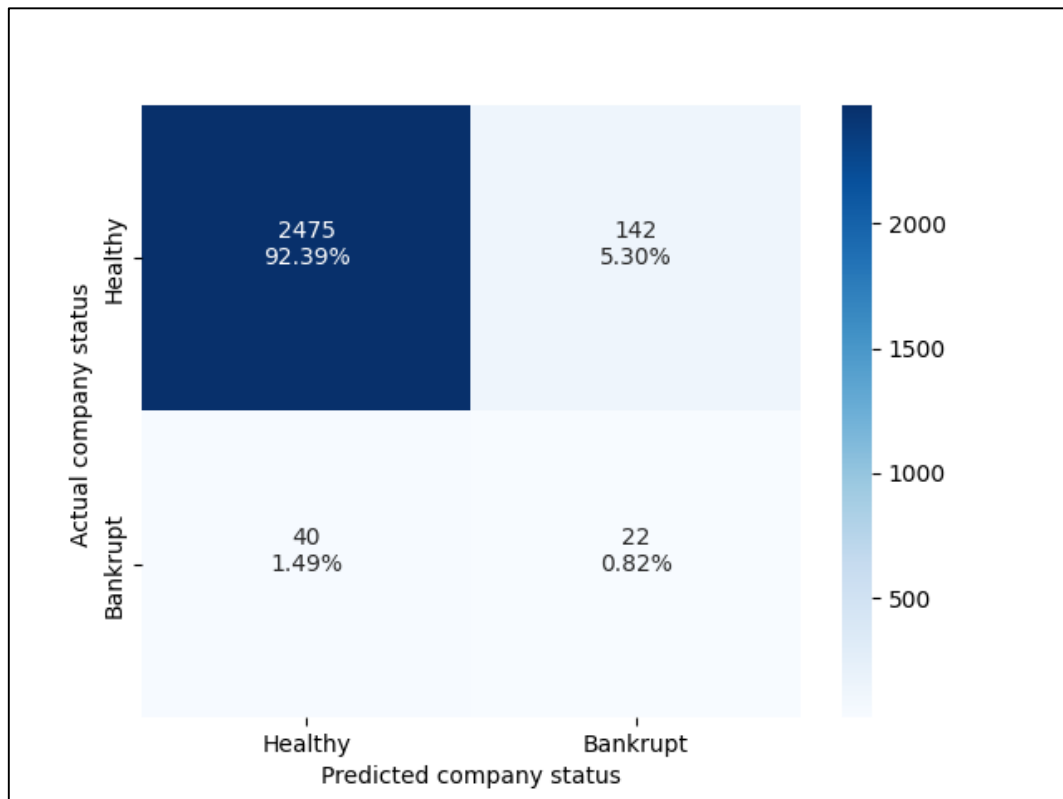
Διάγραμμα 20 - LDA - Train set - Fold 4



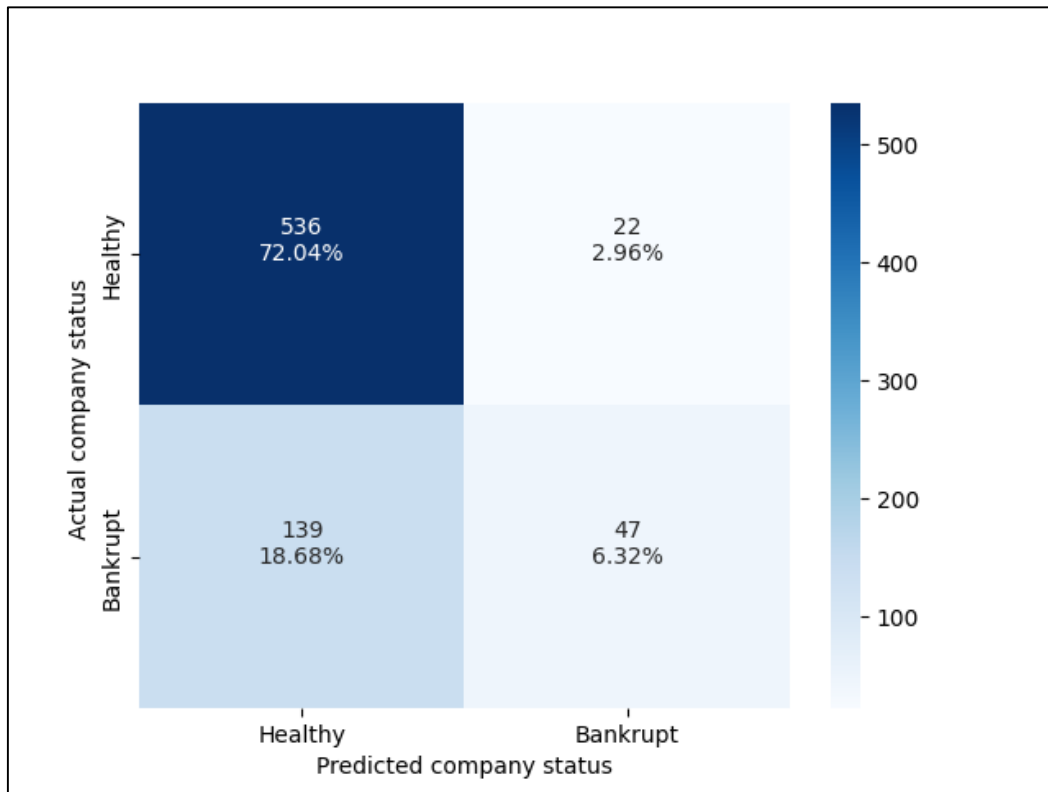
Διάγραμμα 21 - LDA - Test set - Fold 4



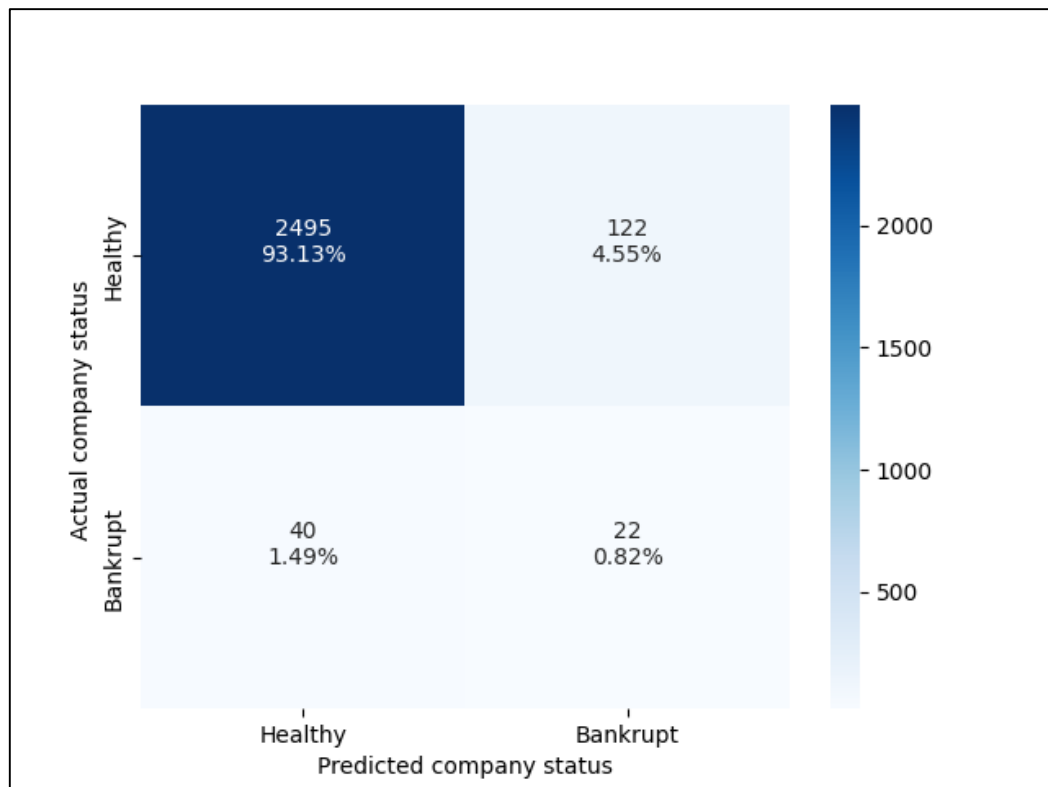
Διάγραμμα 22 - LR - Train set - Fold 1



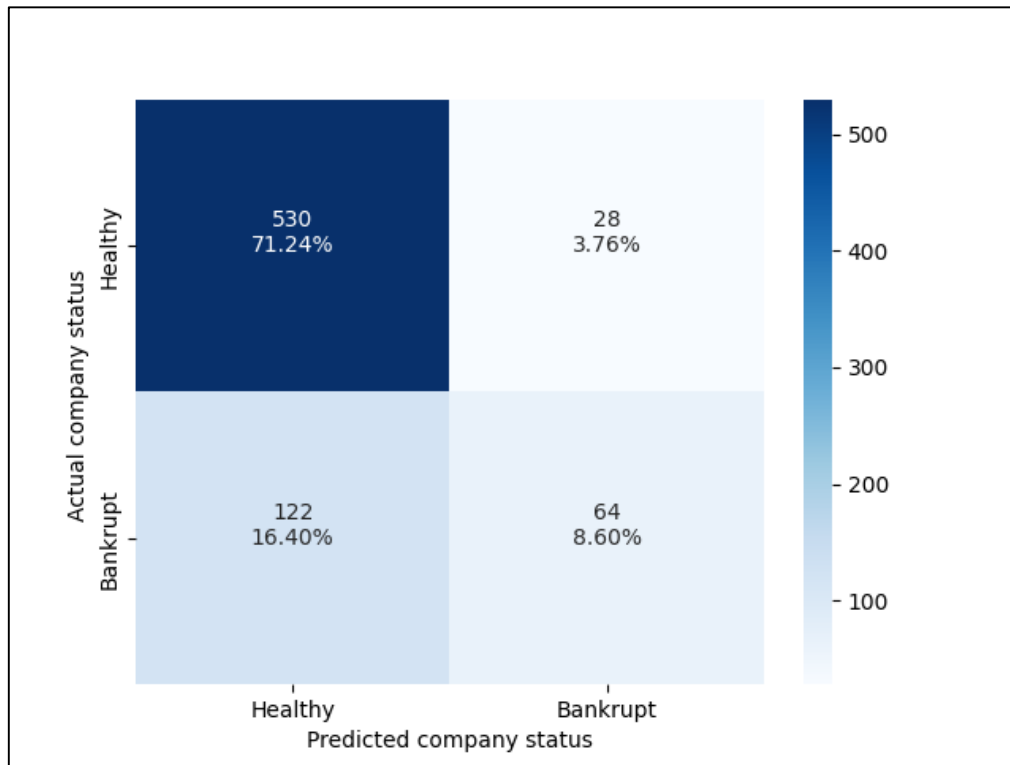
Διάγραμμα 23 - LR - Test set - Fold 1



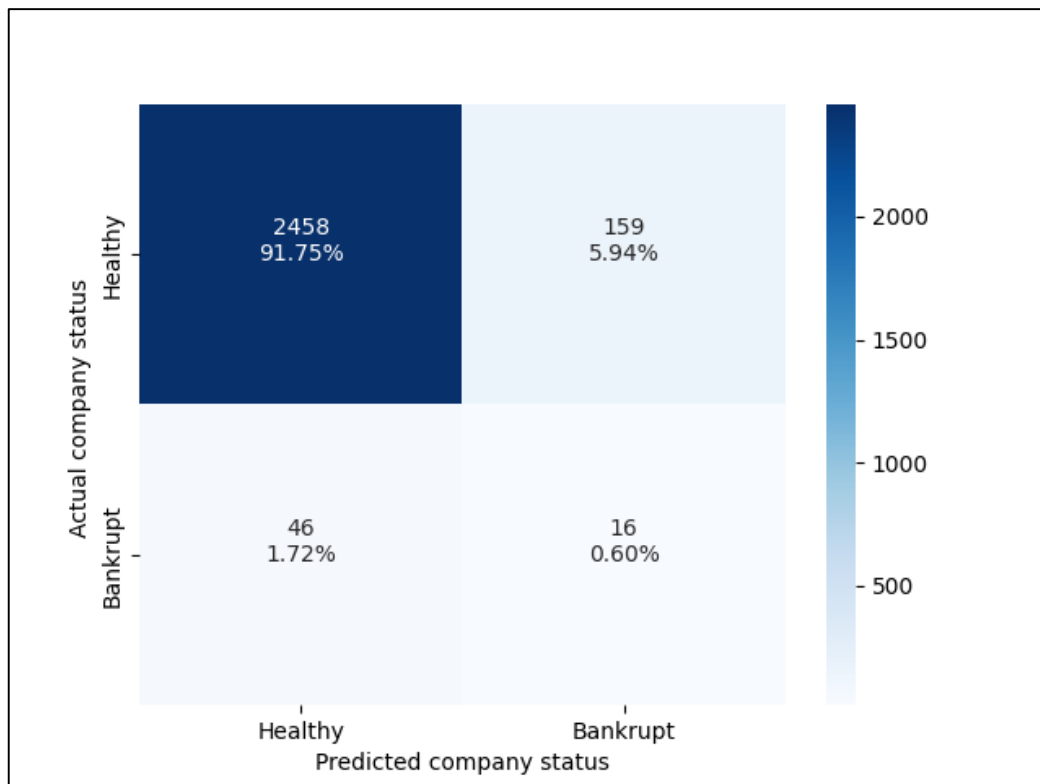
Διάγραμμα 24 - LR - Train set - Fold 2



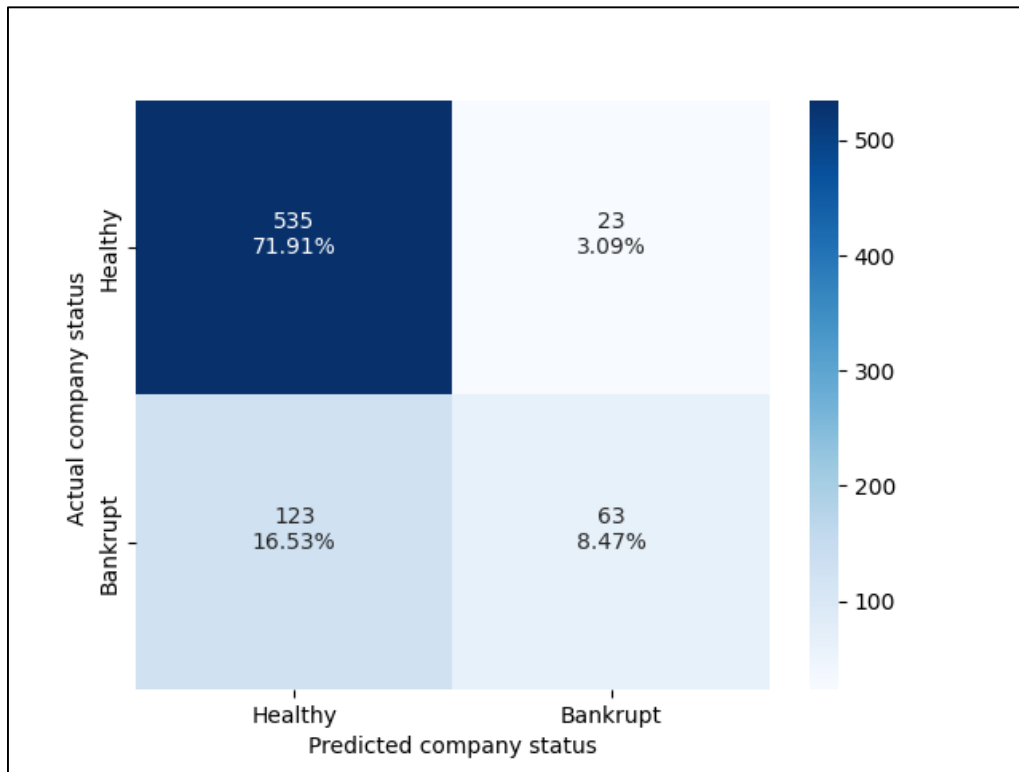
Διάγραμμα 25 - LR - Test set - Fold 2



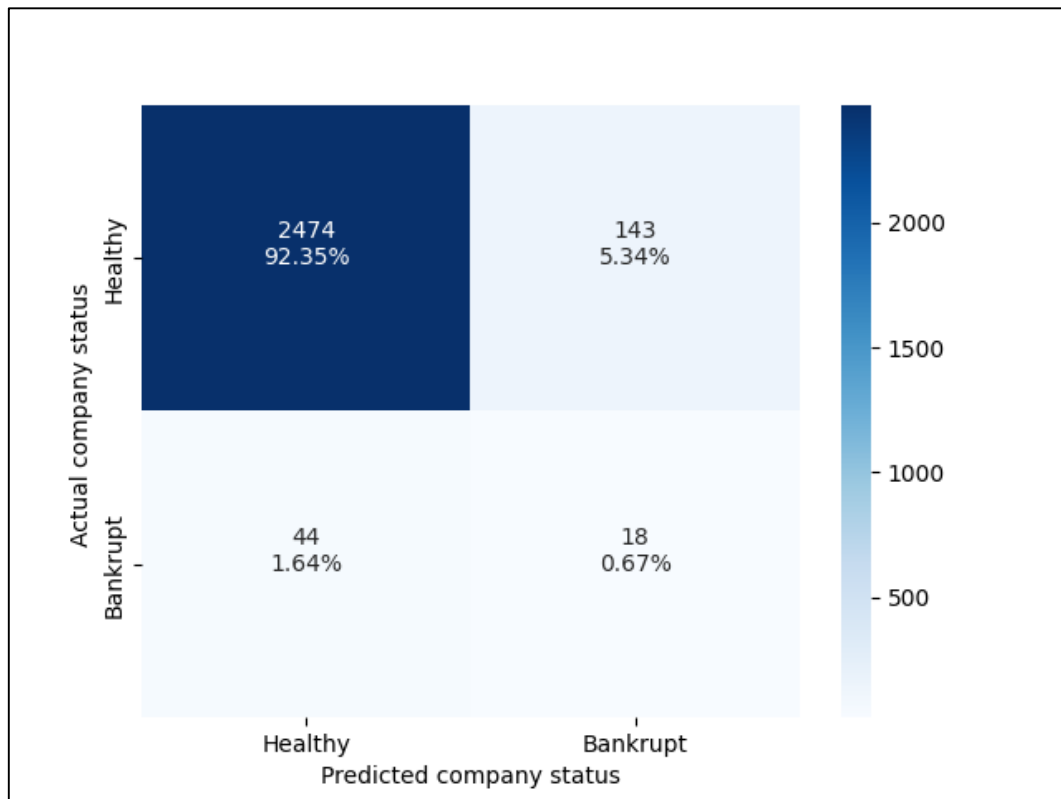
Διάγραμμα 26 - LR - Train set - Fold 3



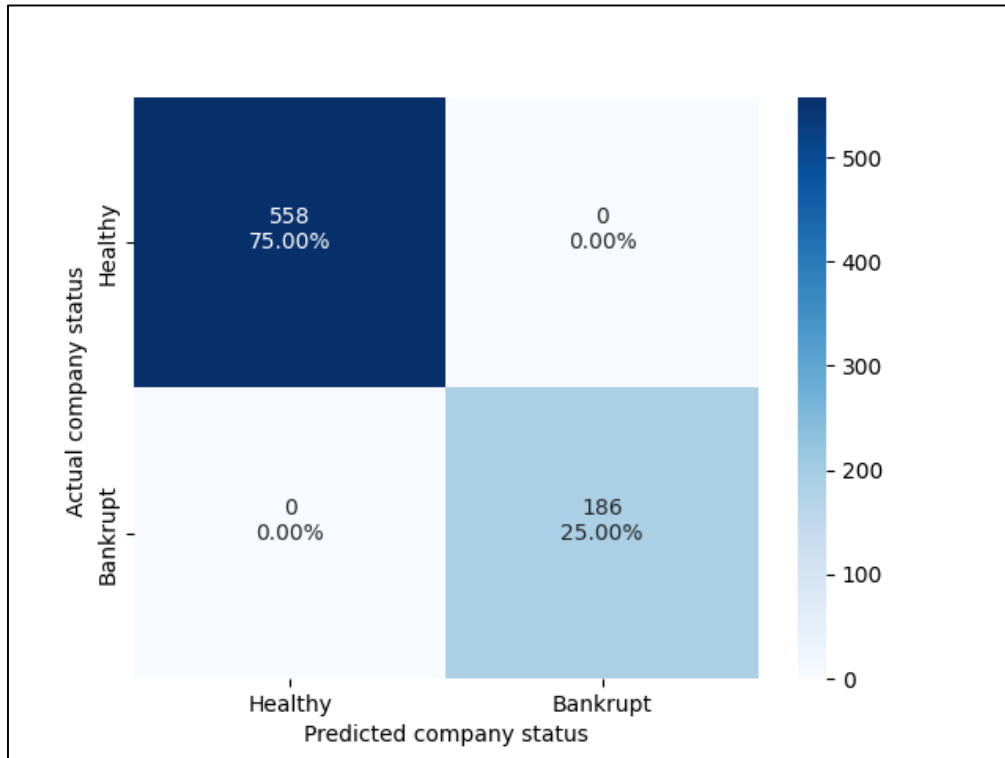
Διάγραμμα 27 - LR - Test set - Fold 3



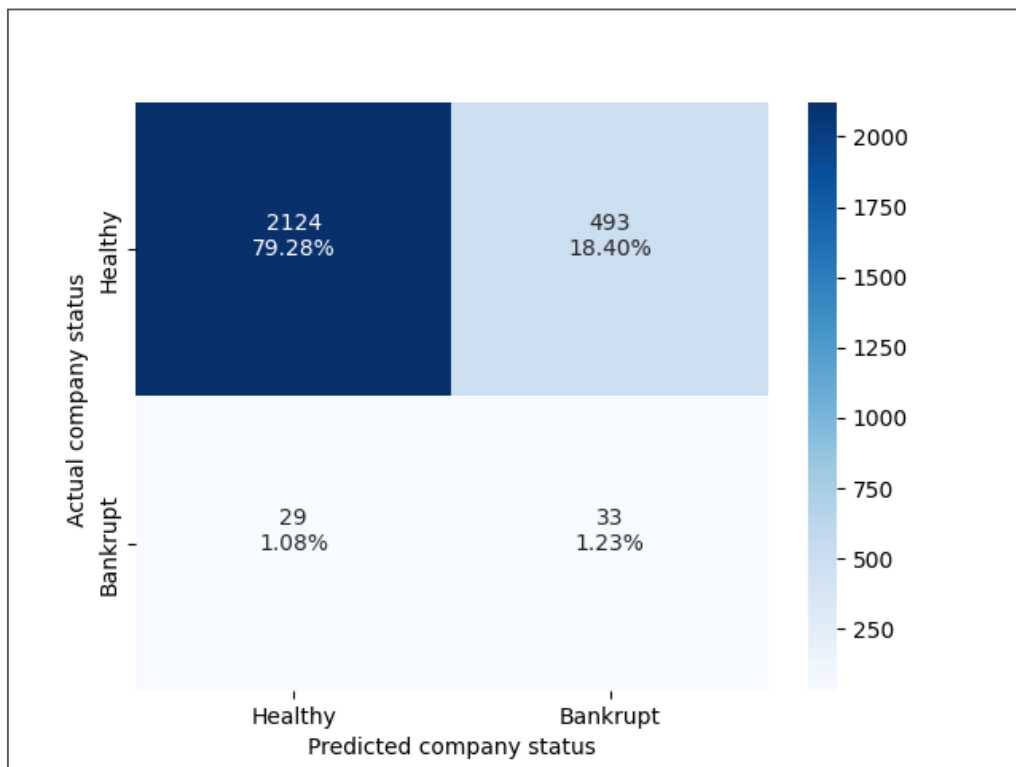
Διάγραμμα 28 - LR - Train set - Fold 4



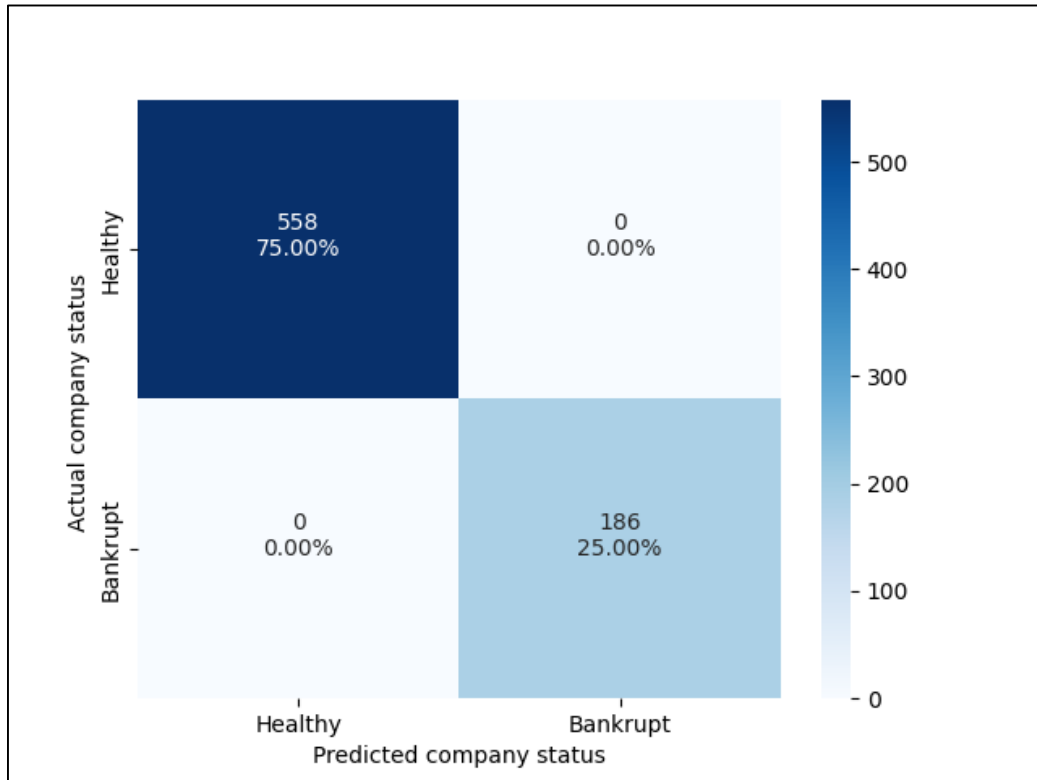
Διάγραμμα 29 - LR - Test set - Fold 4



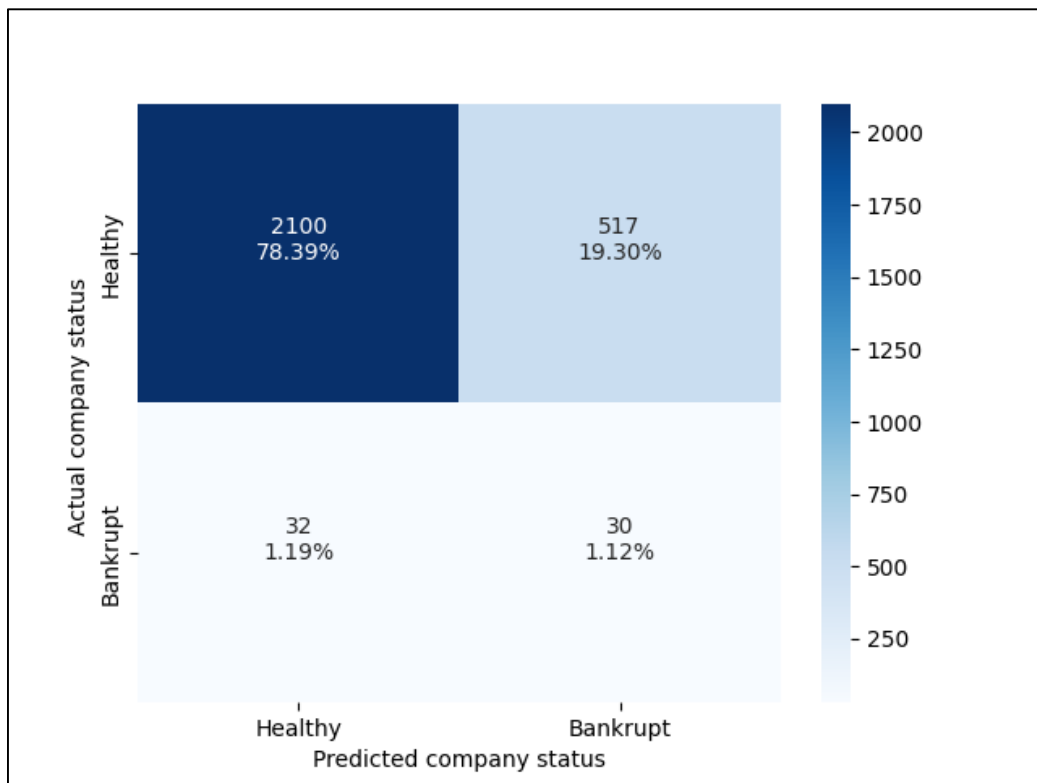
Διάγραμμα 30 - CTree - Train set - Fold 1



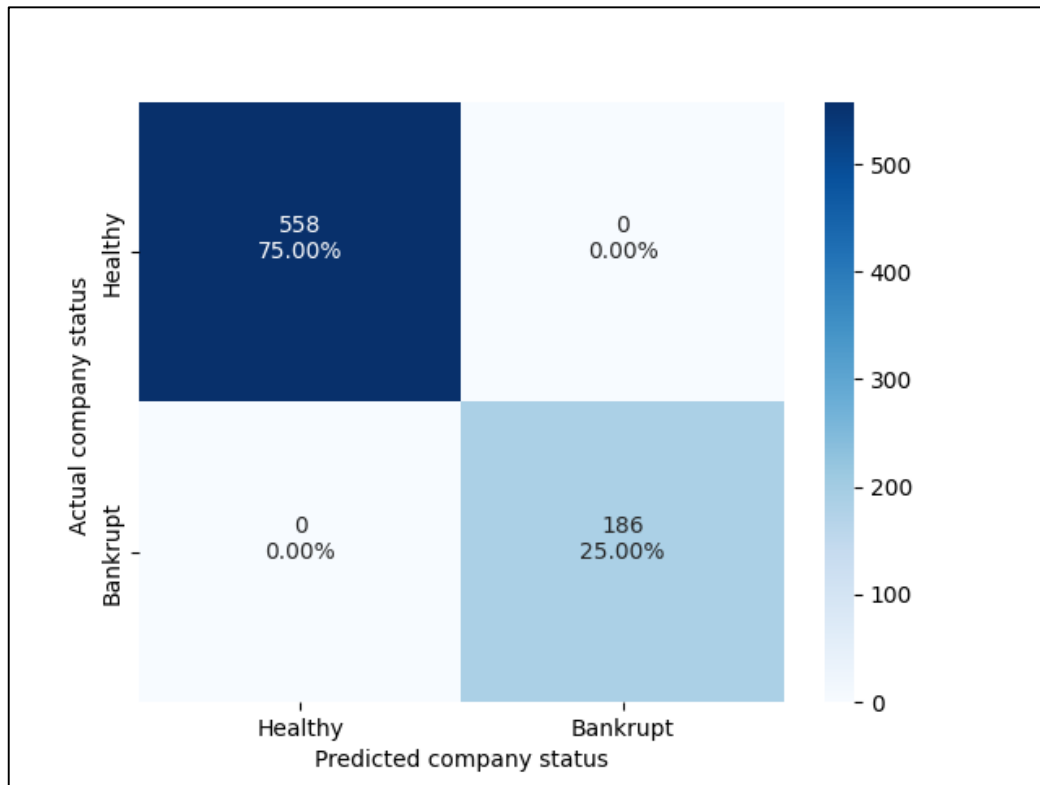
Διάγραμμα 31 - CTree - Test set - Fold 1



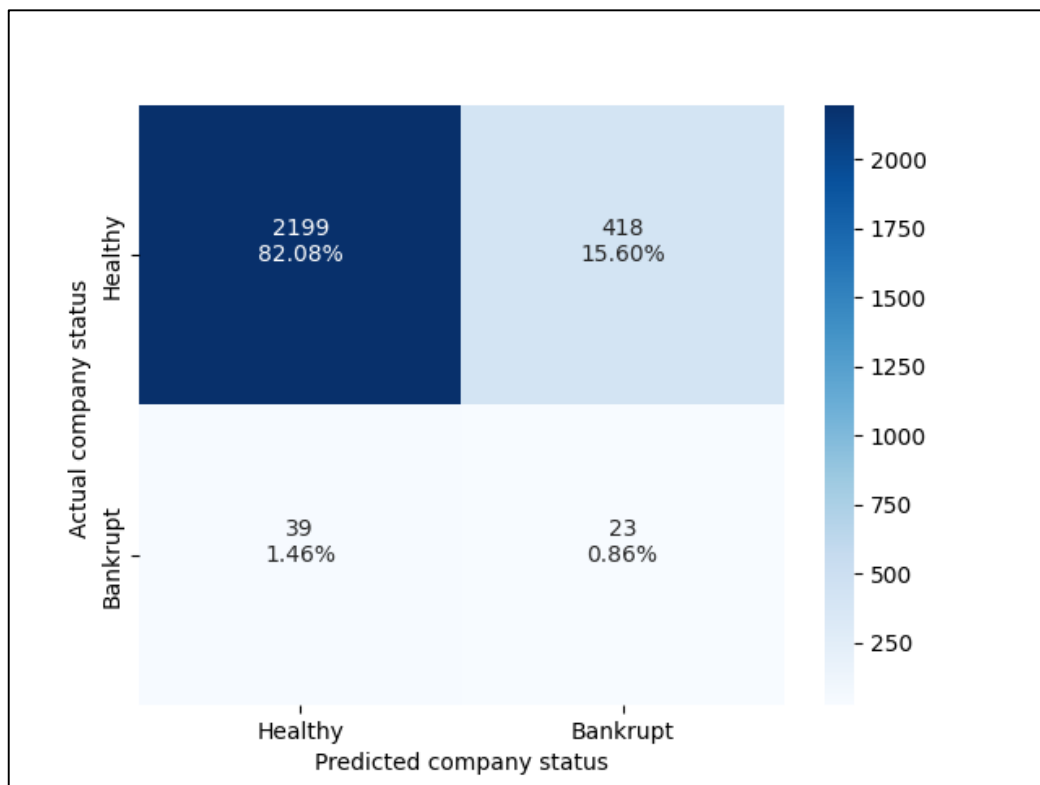
Διάγραμμα 32 - CTree - Train set - Fold 2



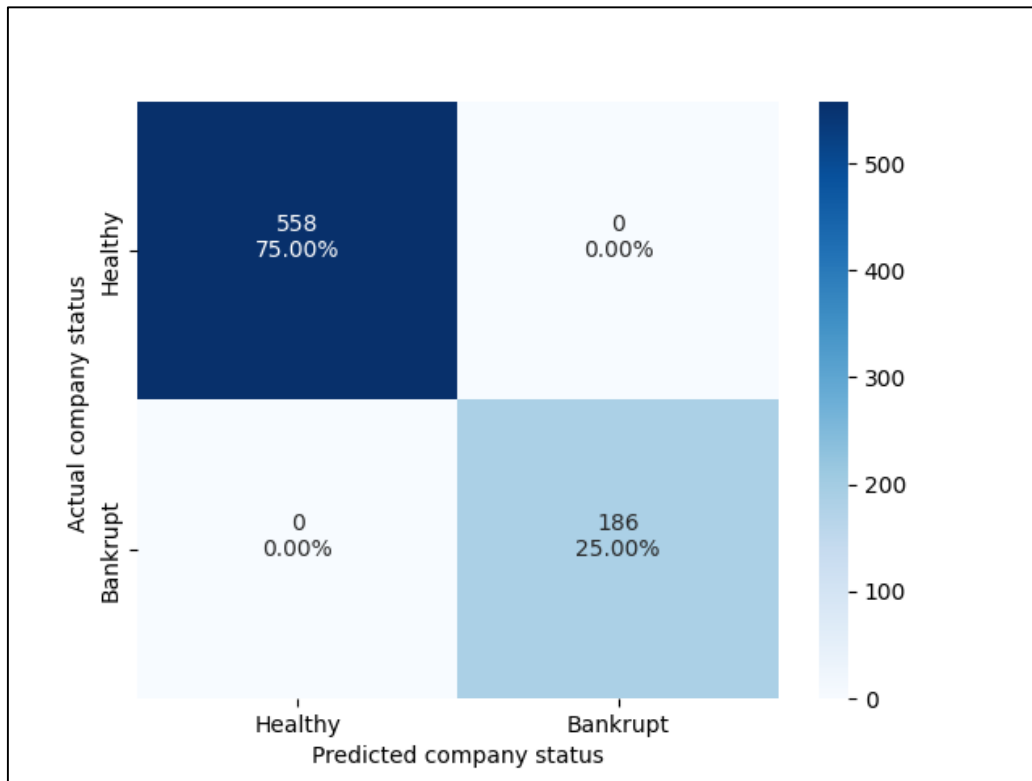
Διάγραμμα 33 - CTree - Test set - Fold 2



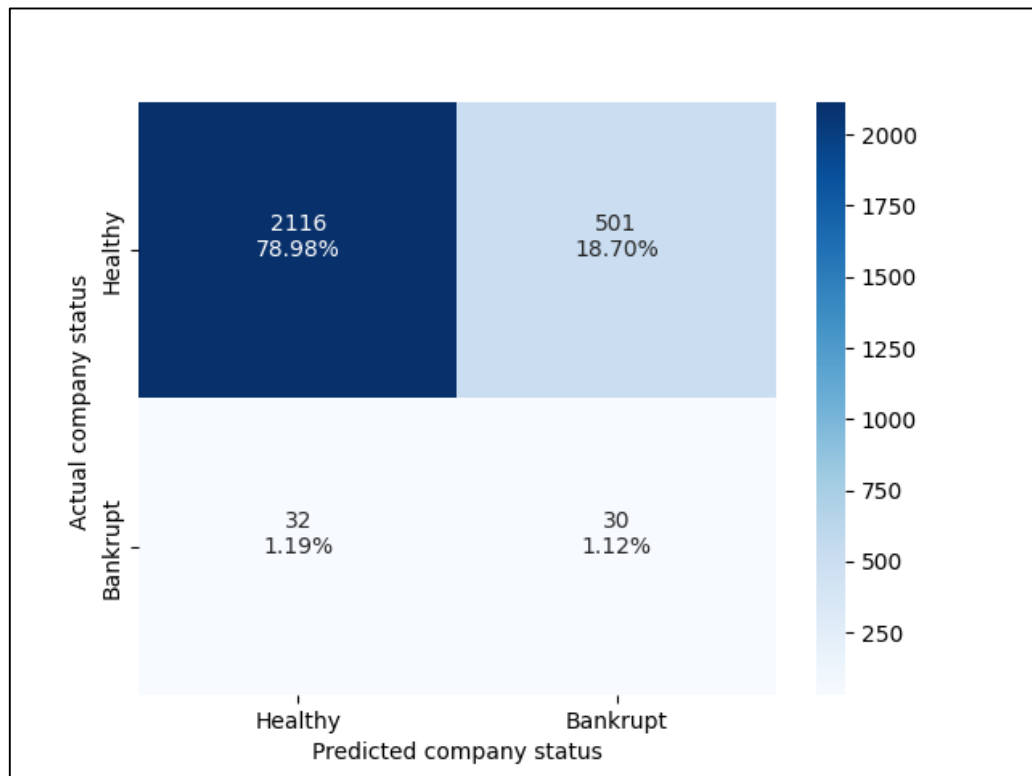
Διάγραμμα 34 - CTREE - Train set - Fold 3



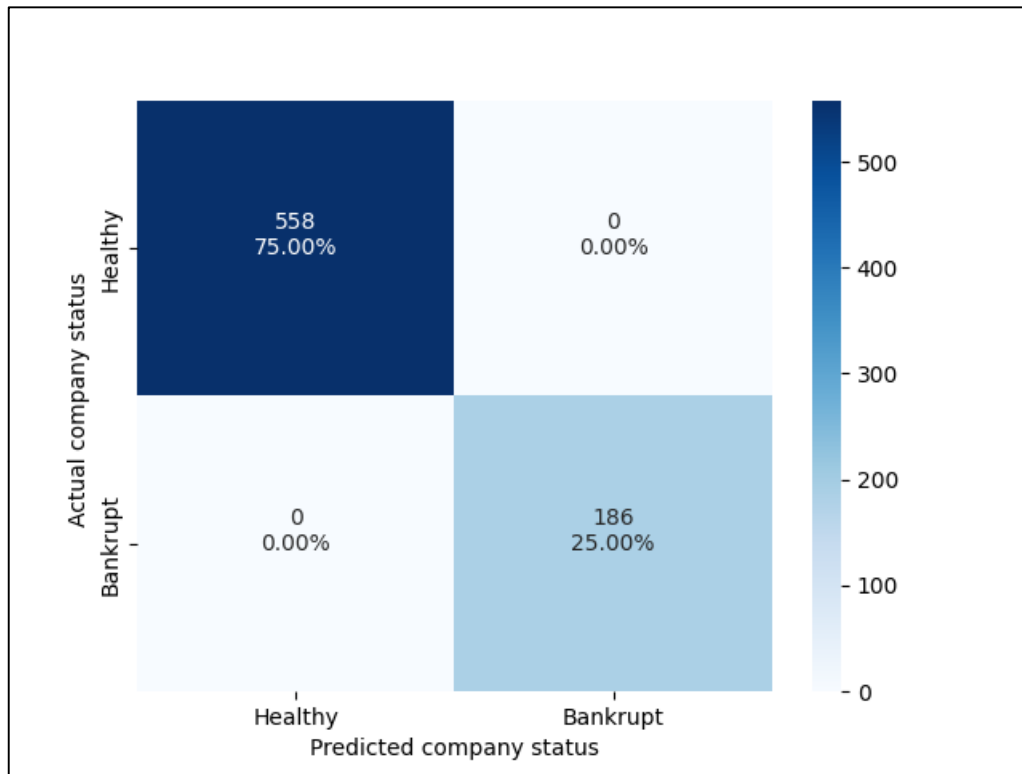
Διάγραμμα 35 - CTREE - Test set - Fold 3



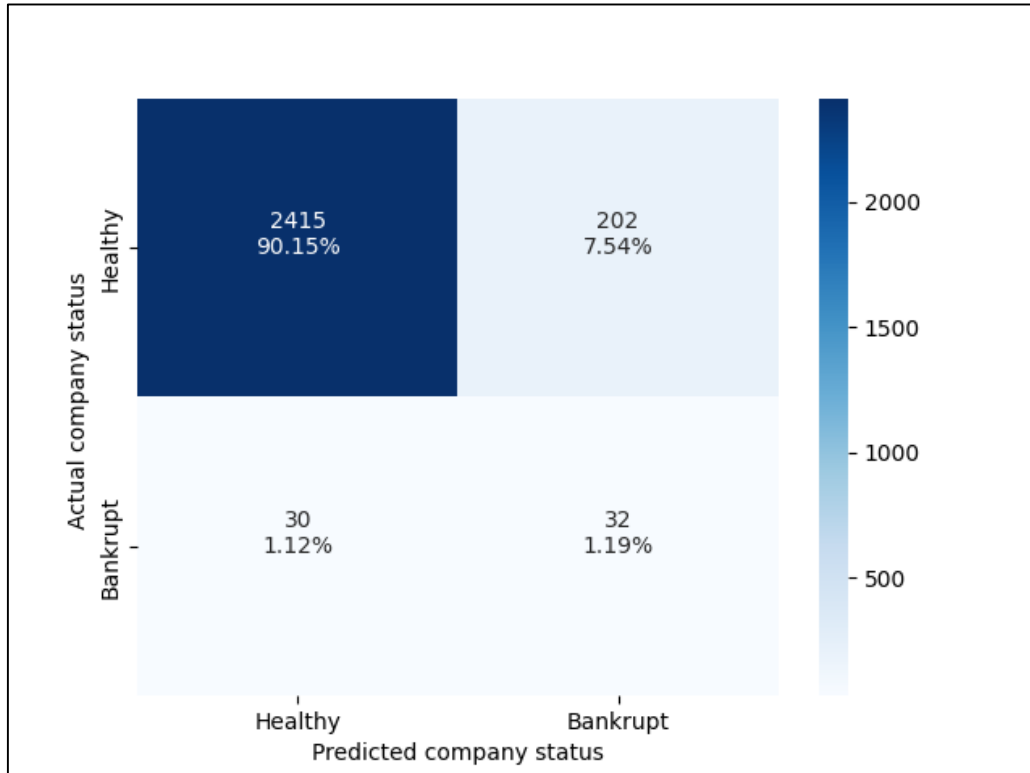
Διάγραμμα 36 - CTree - Train set - Fold 4



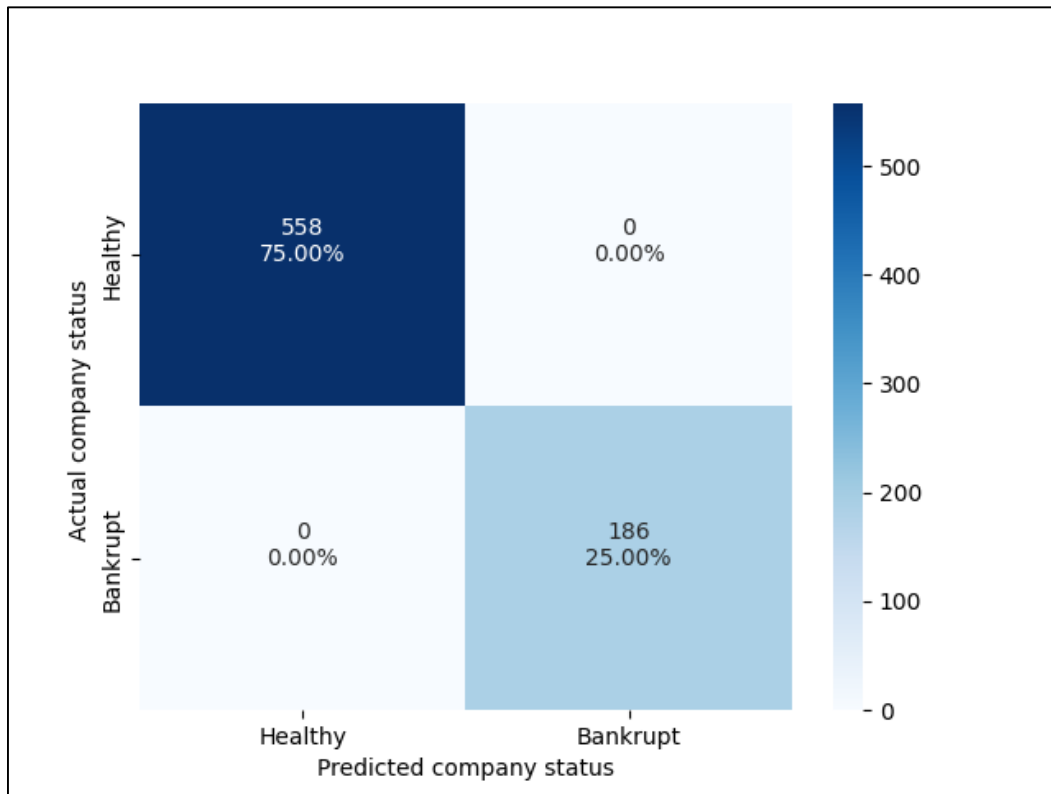
Διάγραμμα 37 - CTree - Test set - Fold 4



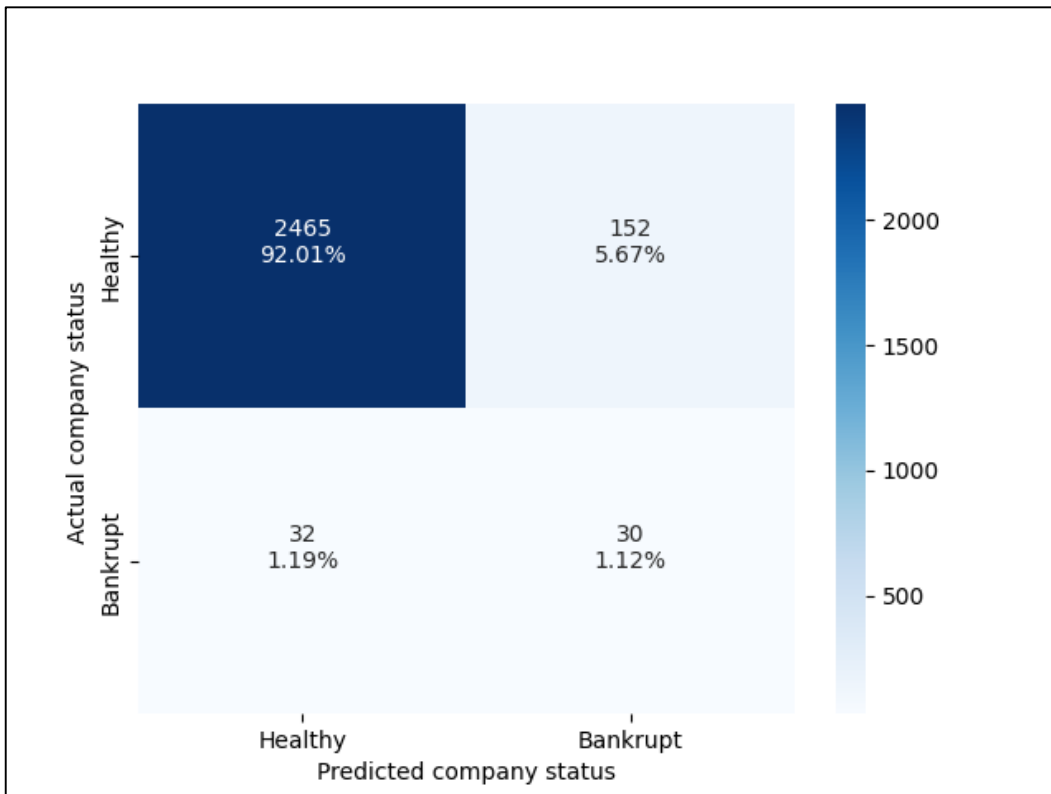
Διάγραμμα 38 - RF - Train set - Fold 1



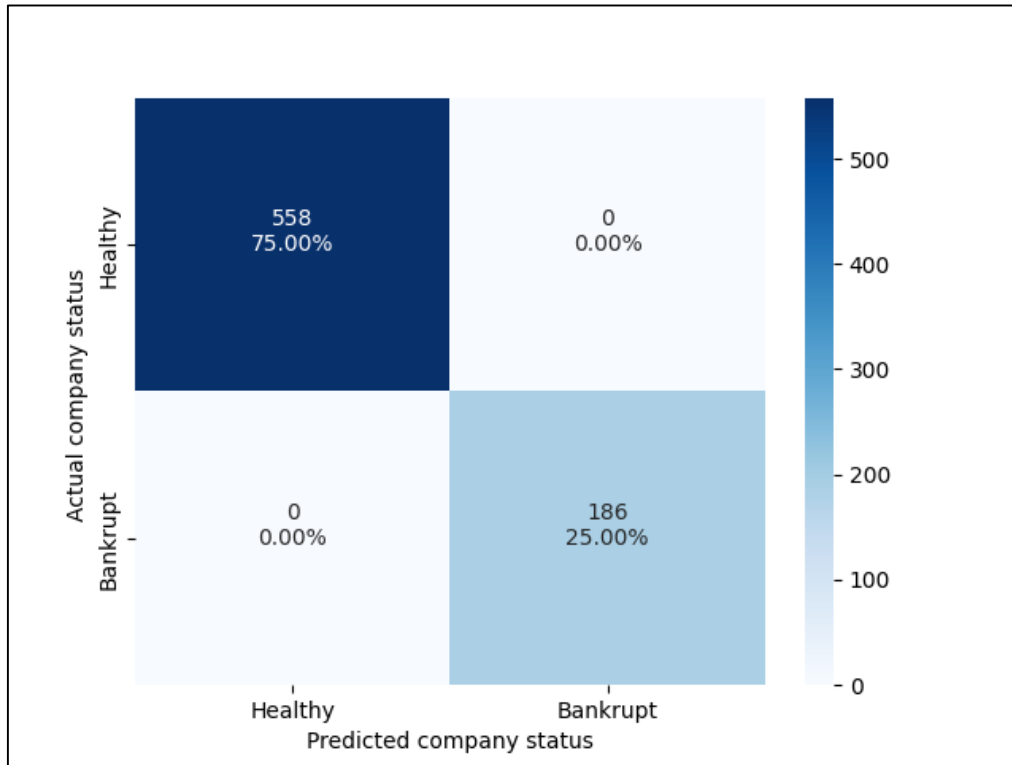
Διάγραμμα 39 - RF - Test set - Fold 1



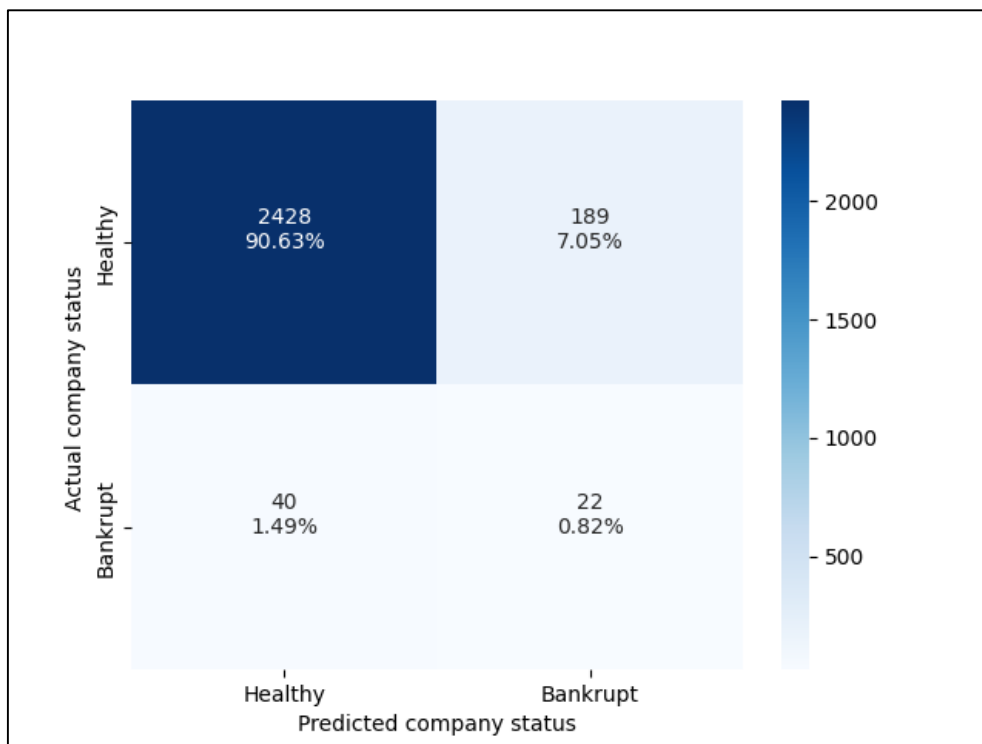
Διάγραμμα 40 - RF - Train set - Fold 2



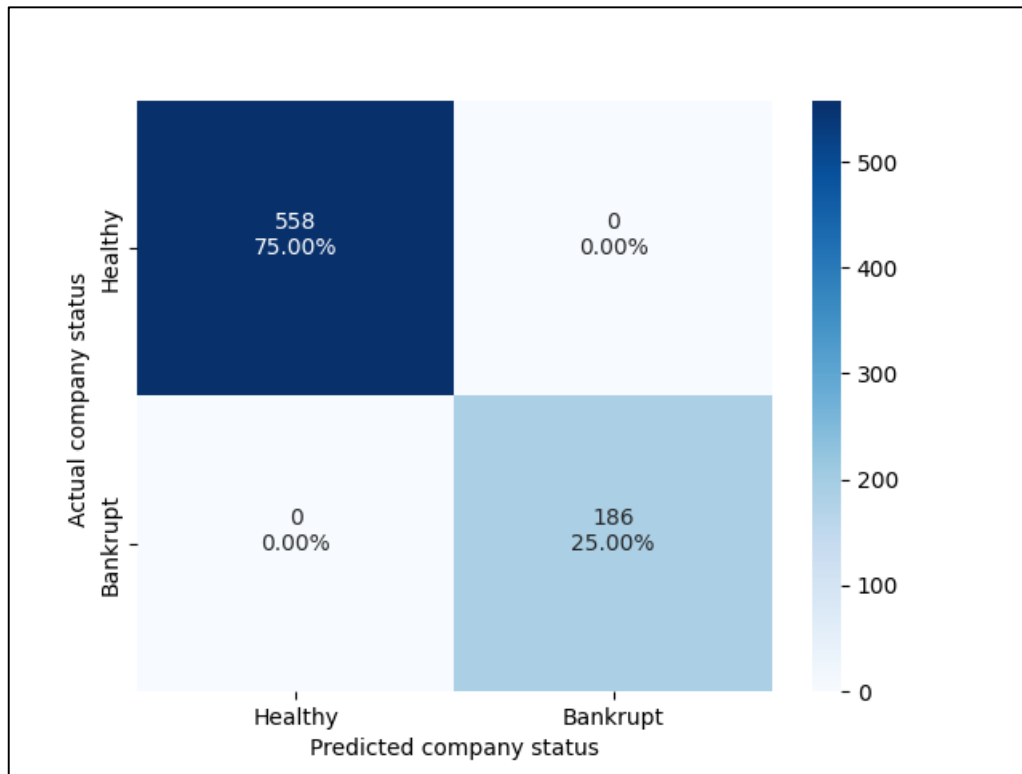
Διάγραμμα 41 - RF - Test set - Fold 2



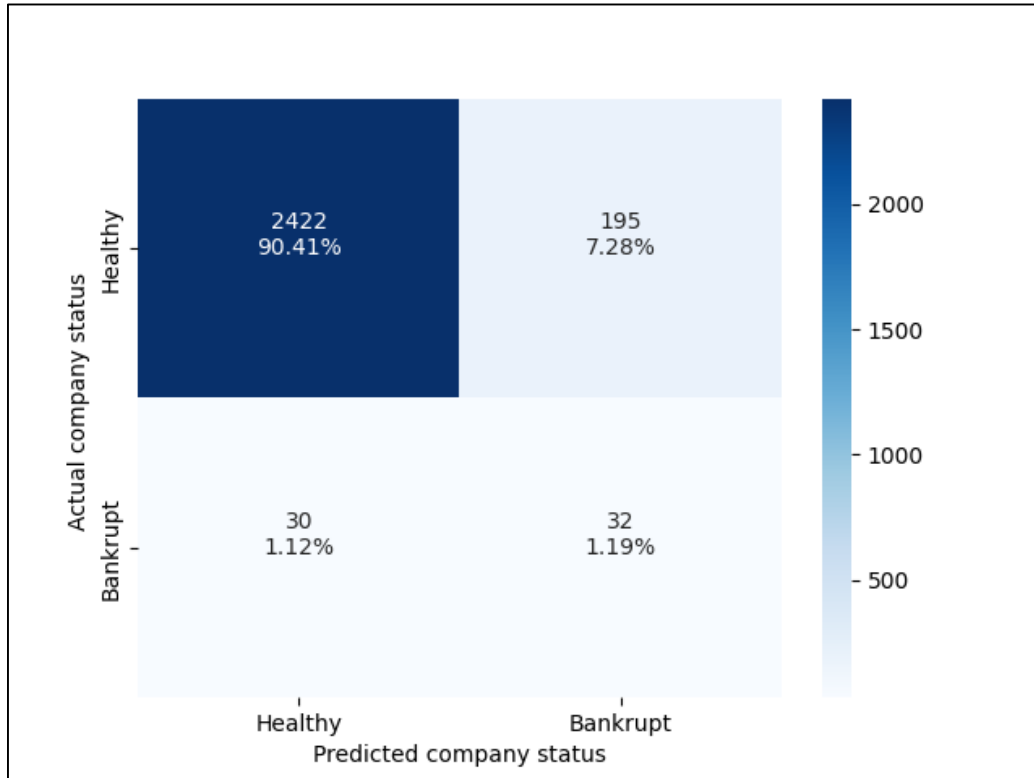
Διάγραμμα 42 - RF - Train set - Fold 3



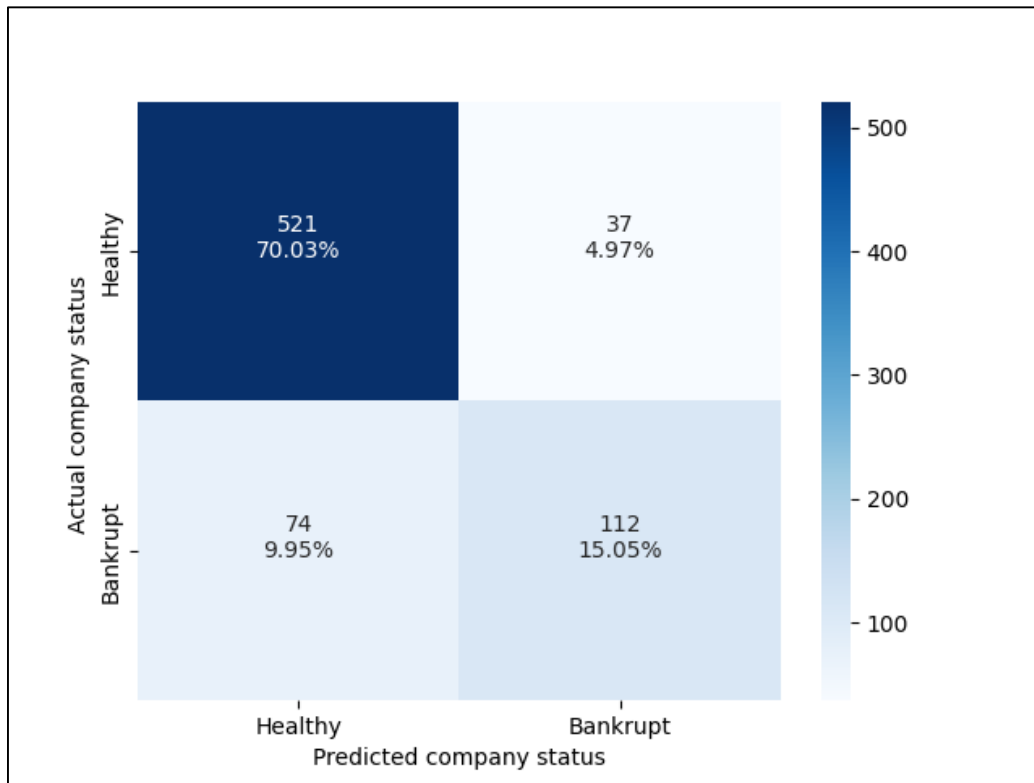
Διάγραμμα 43 - RF - Test set - Fold 3



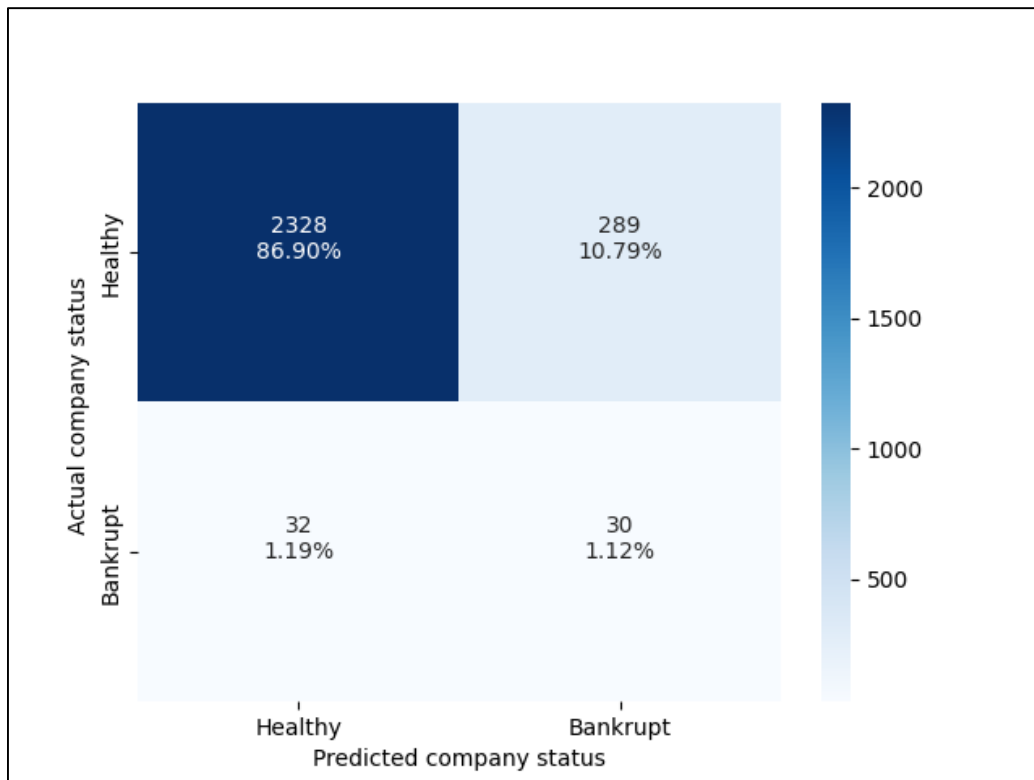
Διάγραμμα 44 - RF - Train set - Fold 4



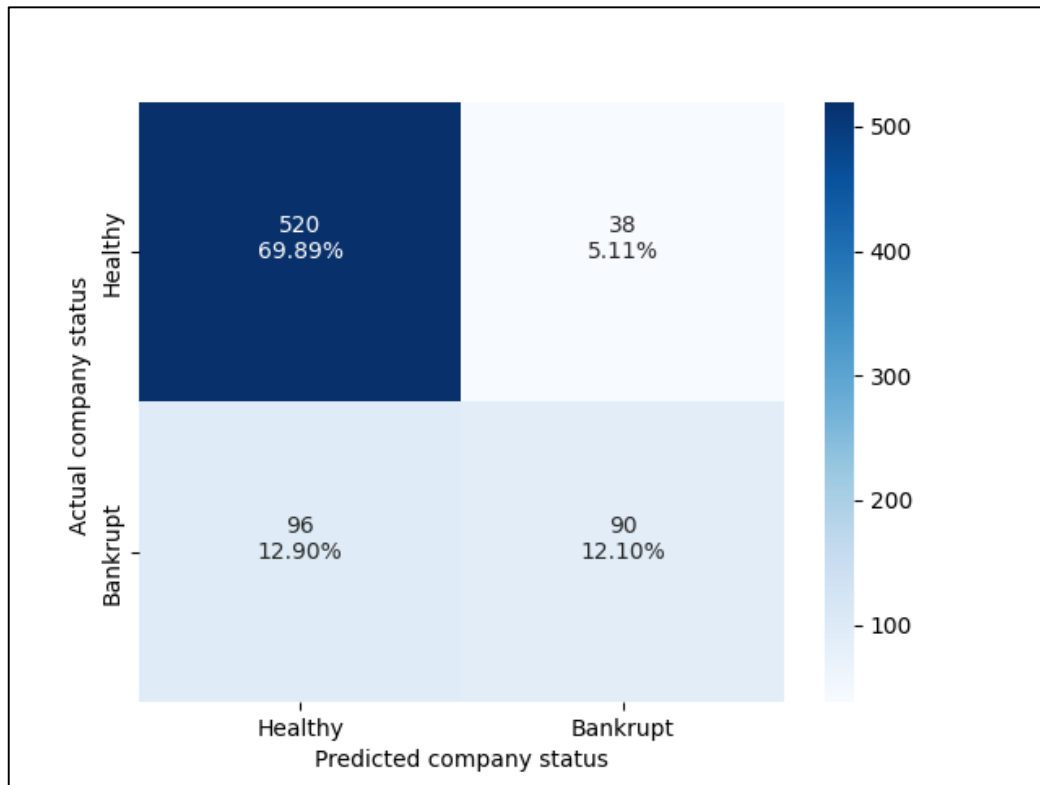
Διάγραμμα 45 - RF - Test set - Fold 4



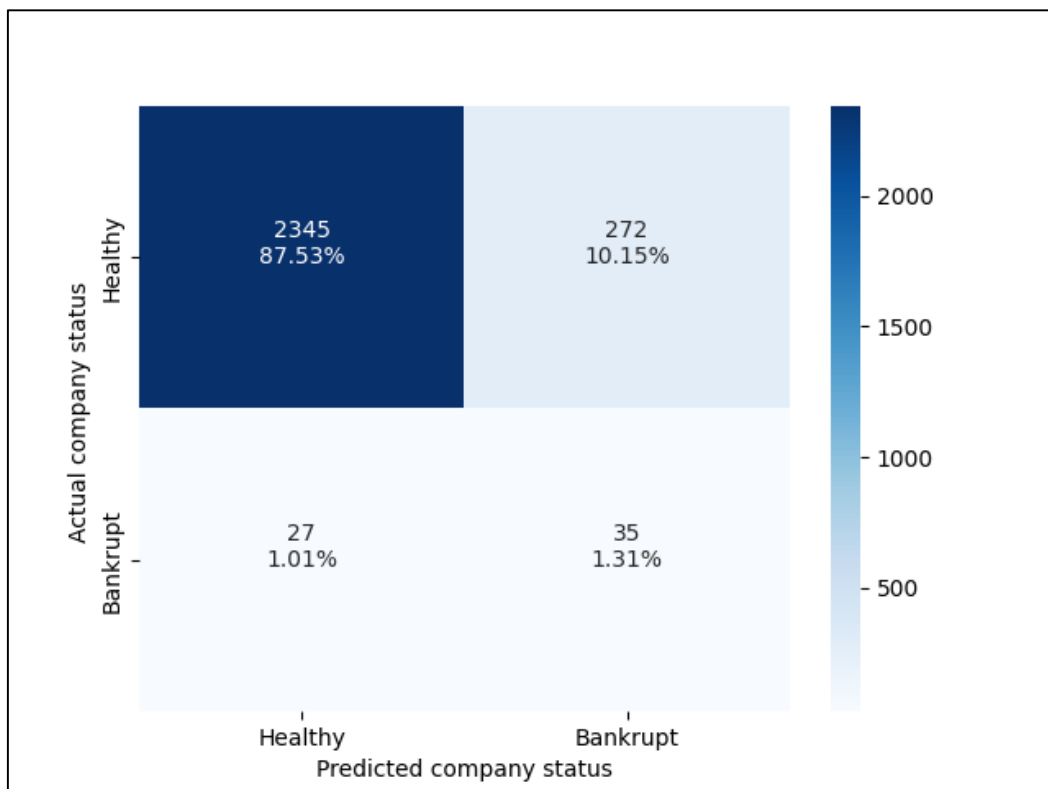
Διάγραμμα 46 - kNN – Train set – Fold 1



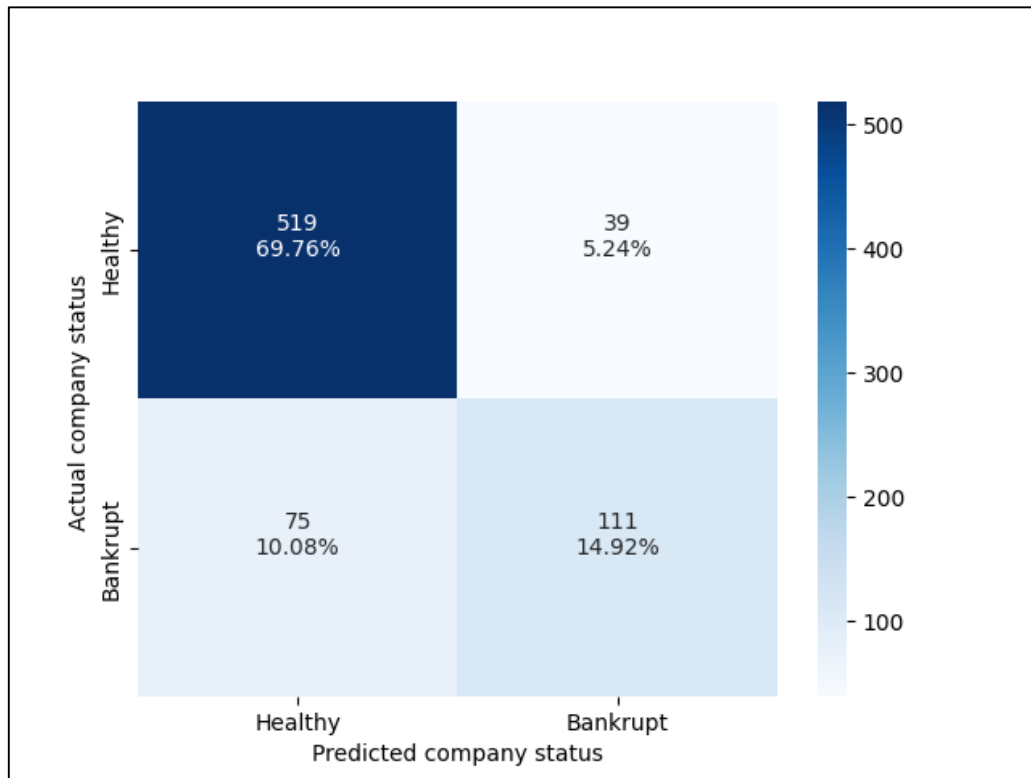
Διάγραμμα 47 - kNN – Test set – Fold 1



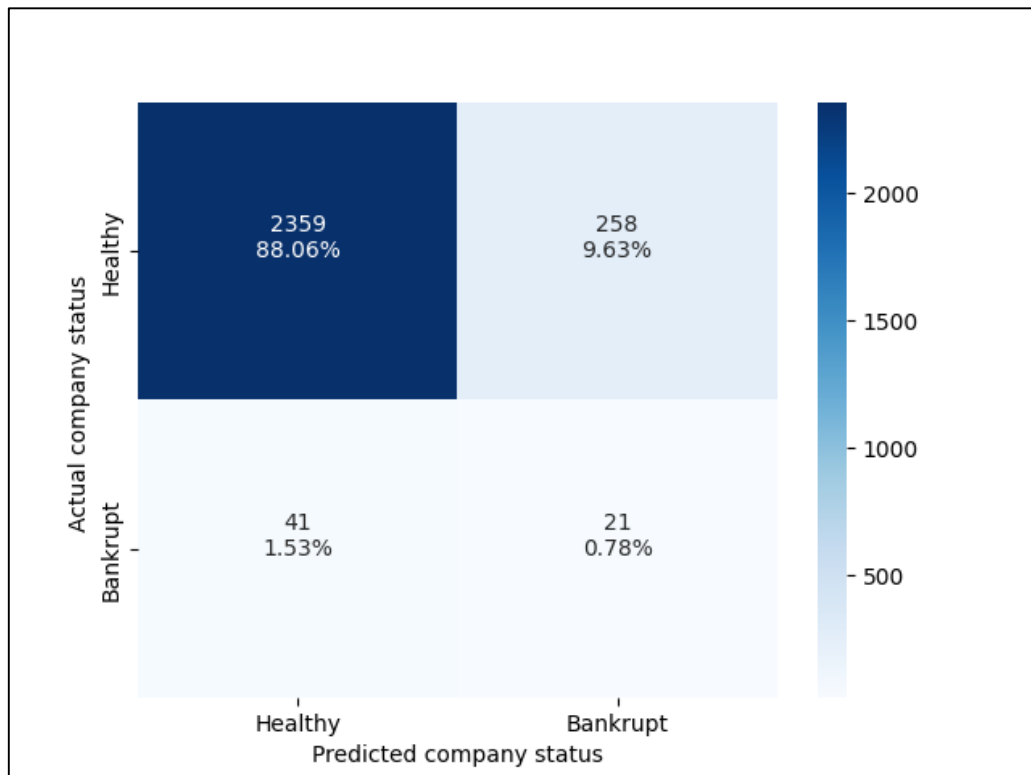
Διάγραμμα 48 - kNN – Train set – Fold 2



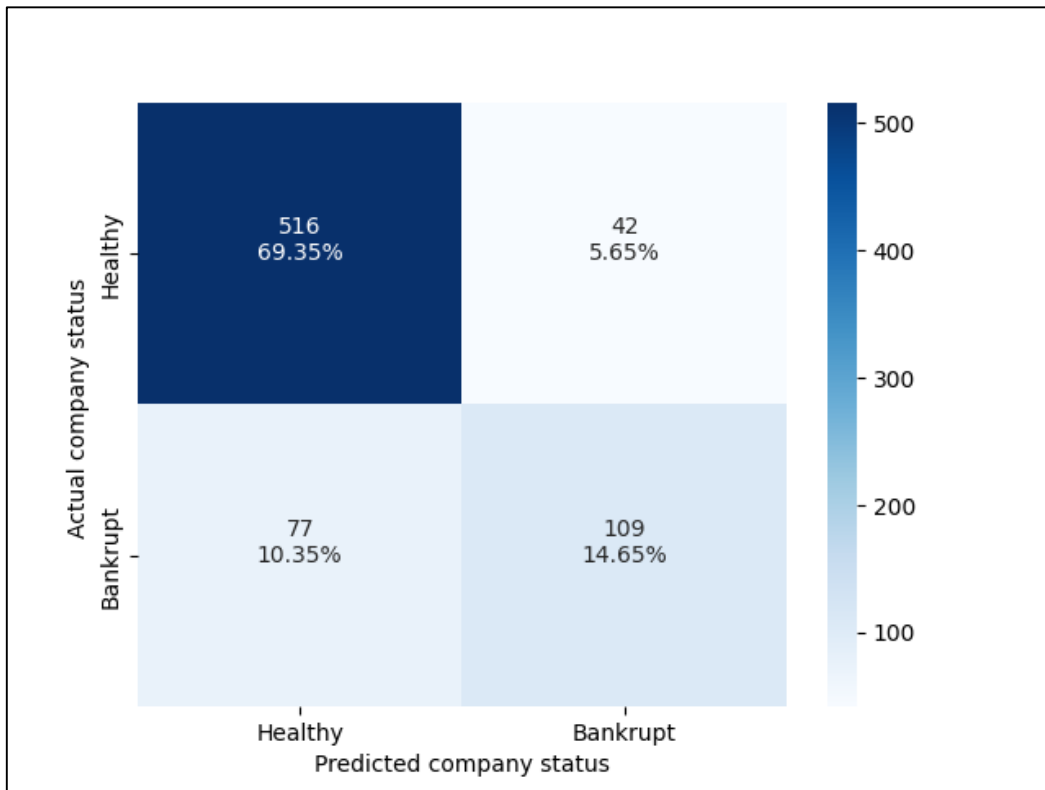
Διάγραμμα 49 - kNN – Test set – Fold 2



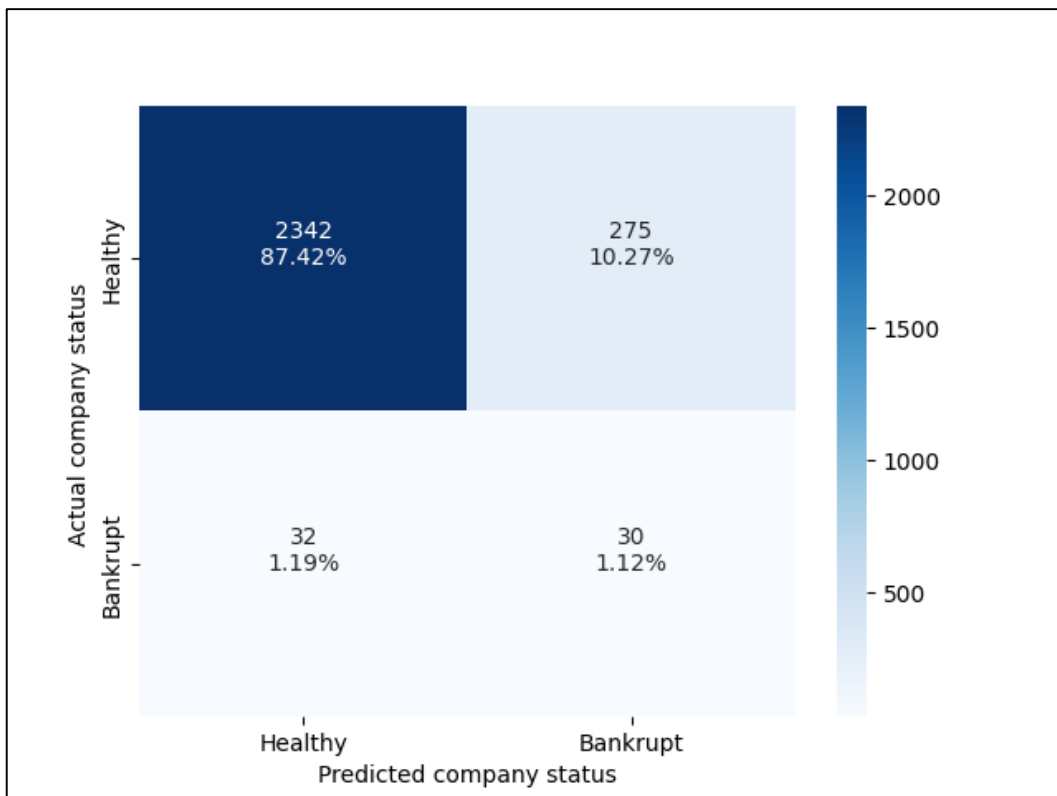
Διάγραμμα 50 - kNN – Train set – Fold 3



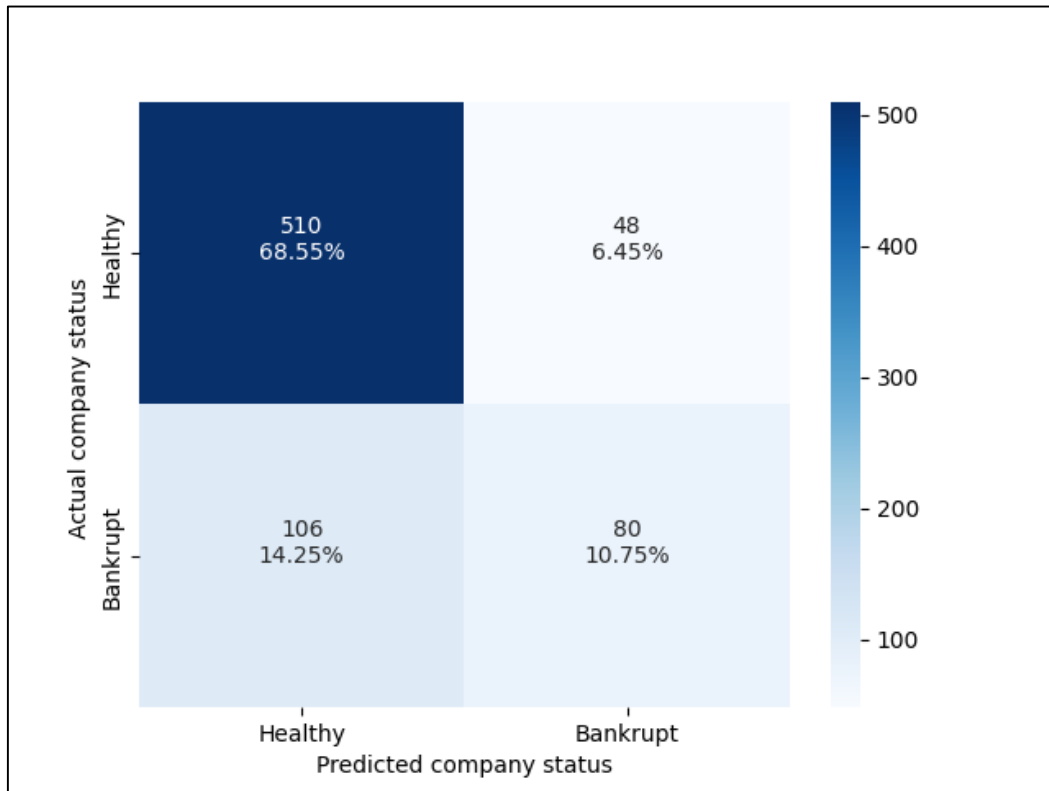
Διάγραμμα 51 - kNN – Test set – Fold 3



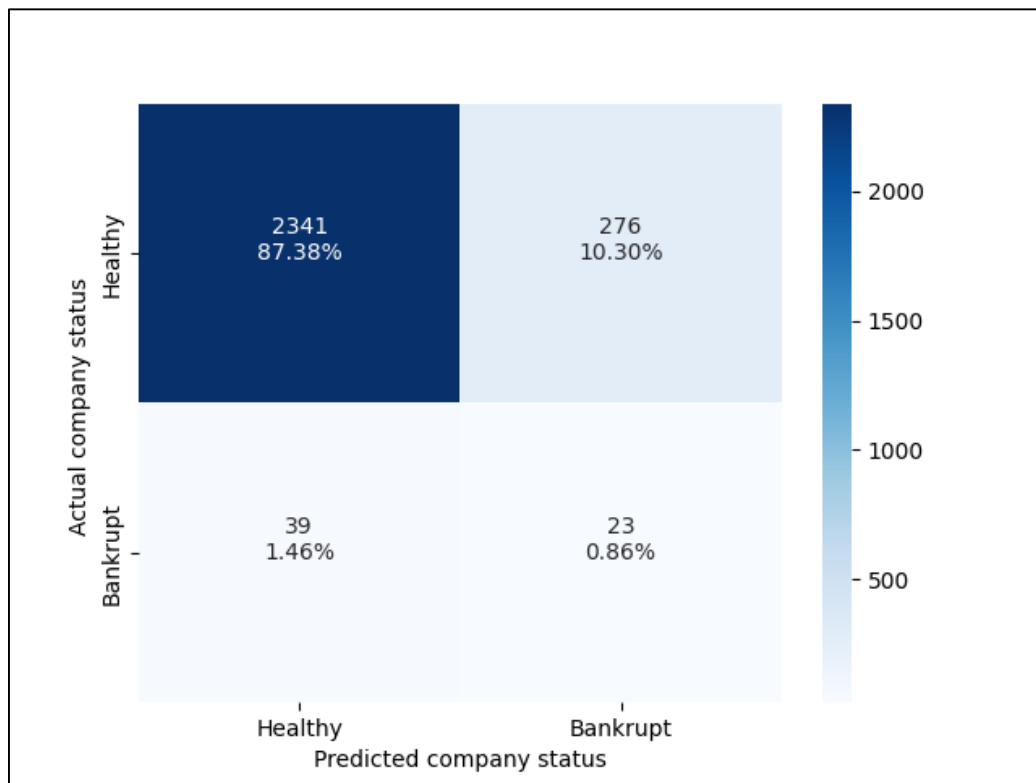
Διάγραμμα 52 - kNN – Train set – Fold 4



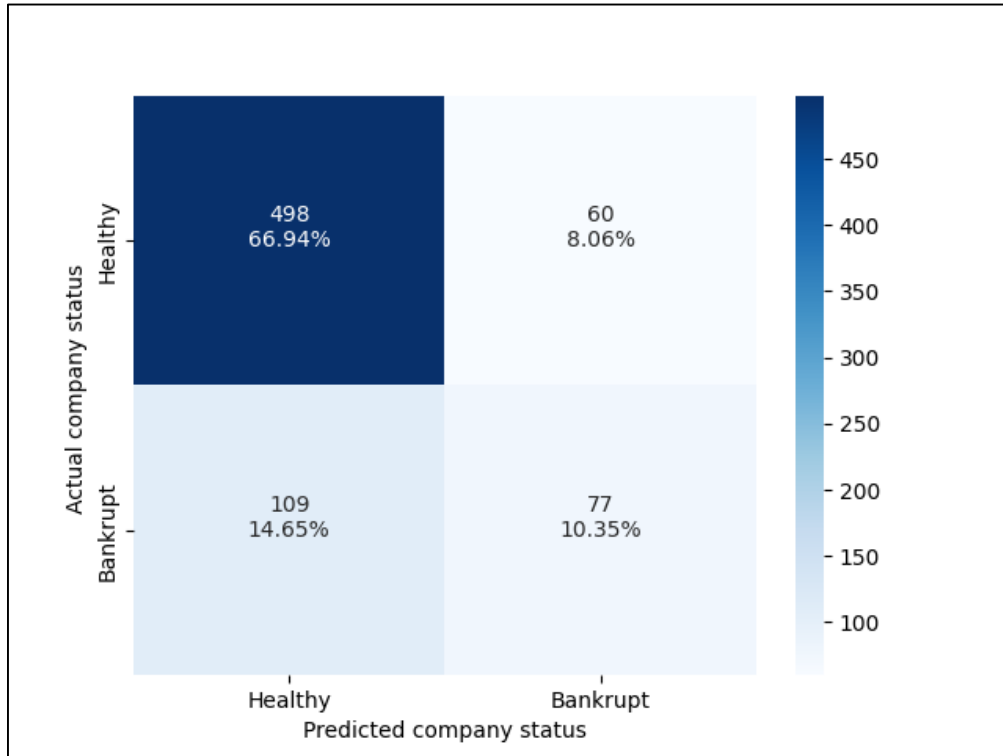
Διάγραμμα 53 - kNN – Test set – Fold 4



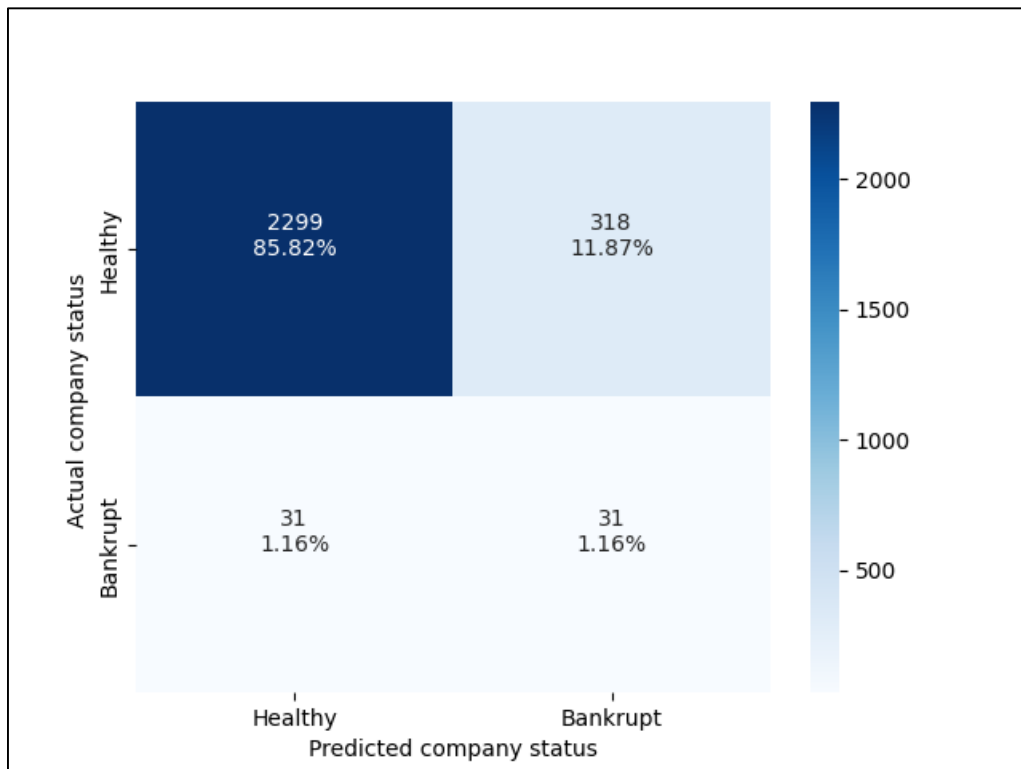
Διάγραμμα 54 - NV – Train set – Fold 1



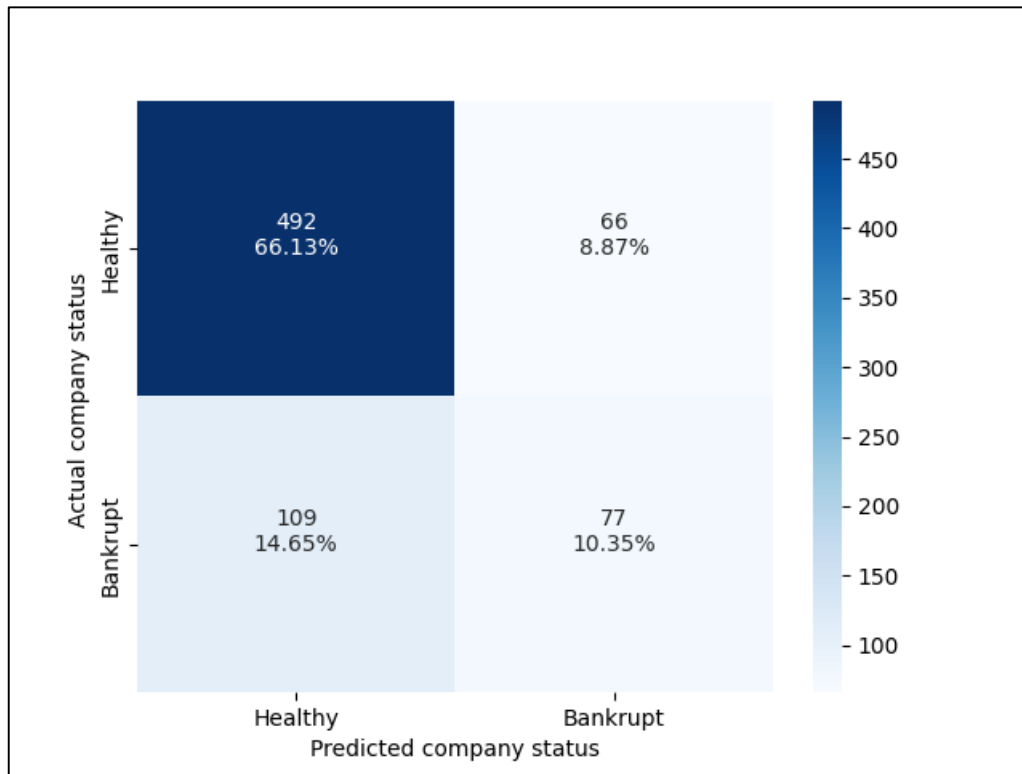
Διάγραμμα 55 - NV – Test set – Fold 1



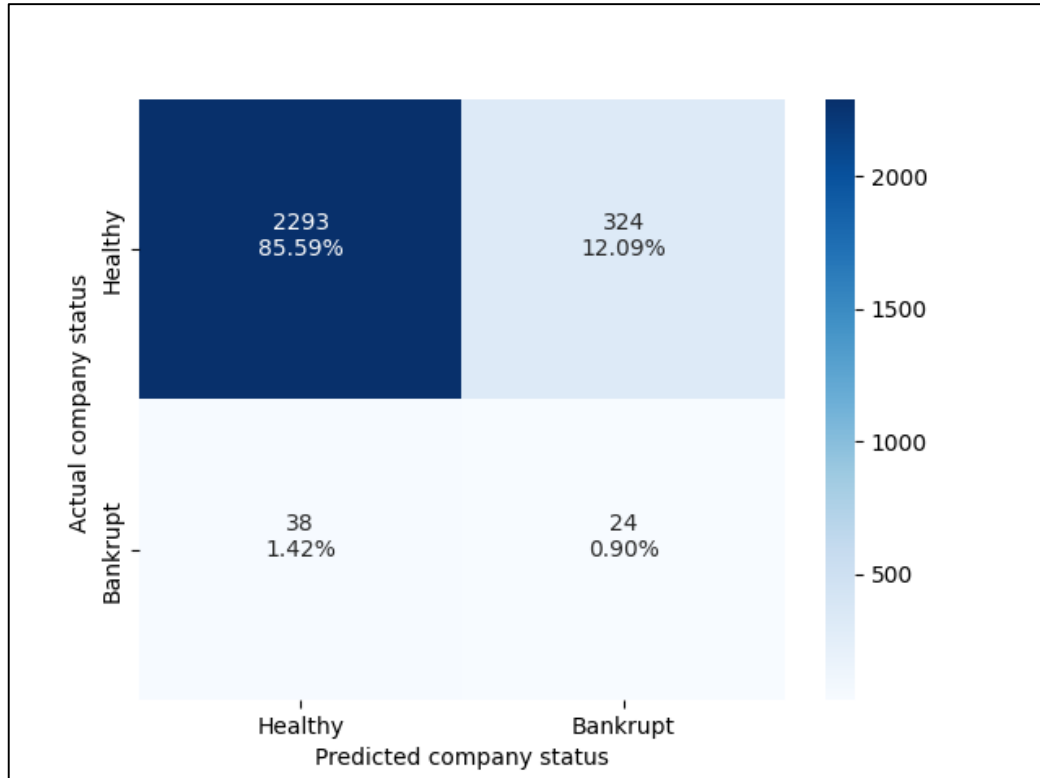
Διάγραμμα 56 - NV – Train set – Fold 2



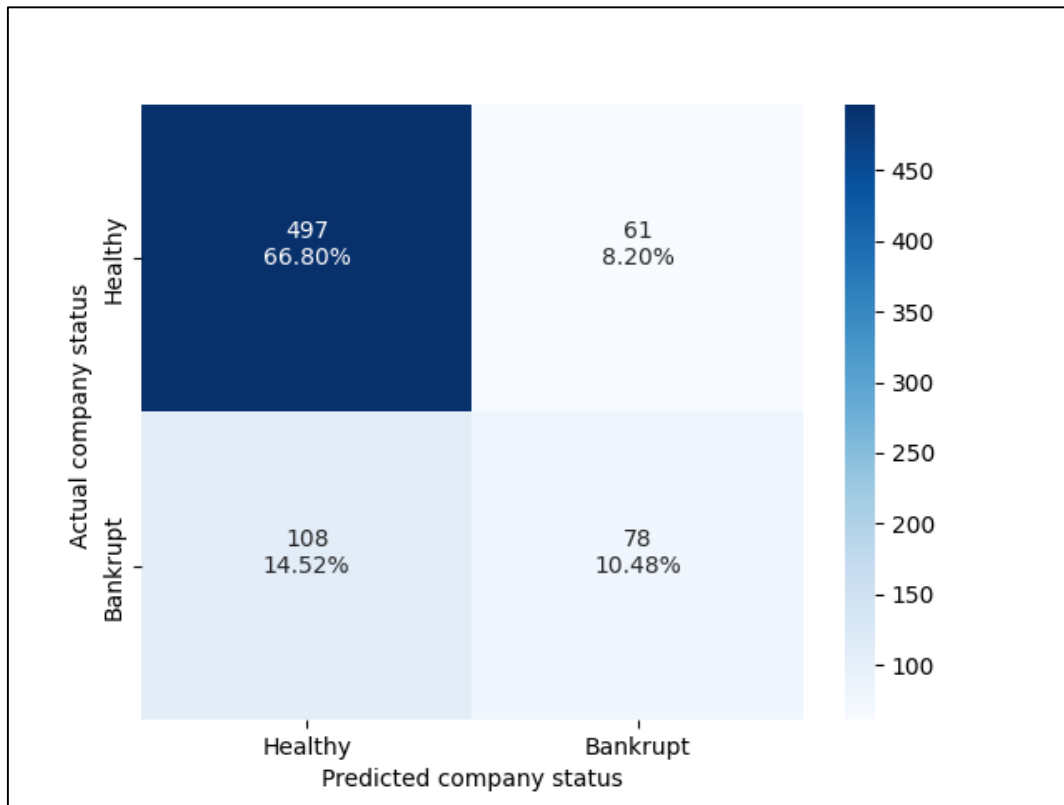
Διάγραμμα 57 - NV – Test set – Fold 2



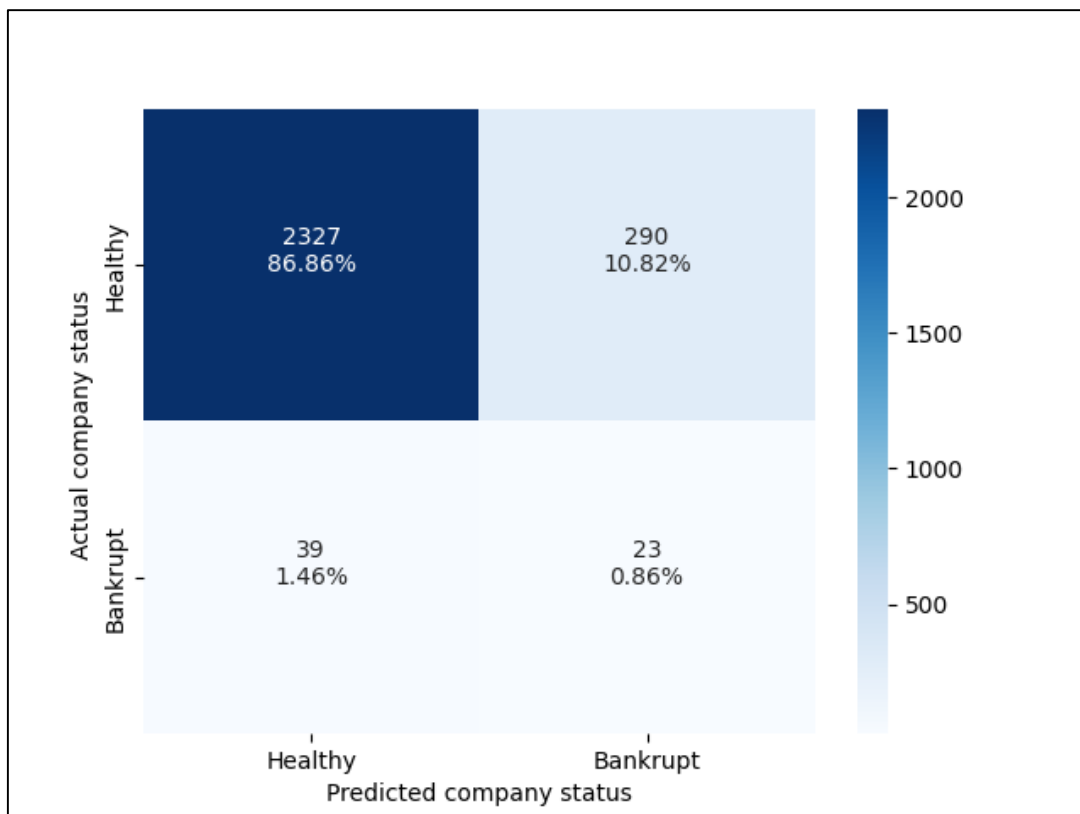
Διάγραμμα 58 - NV – Train set – Fold 3



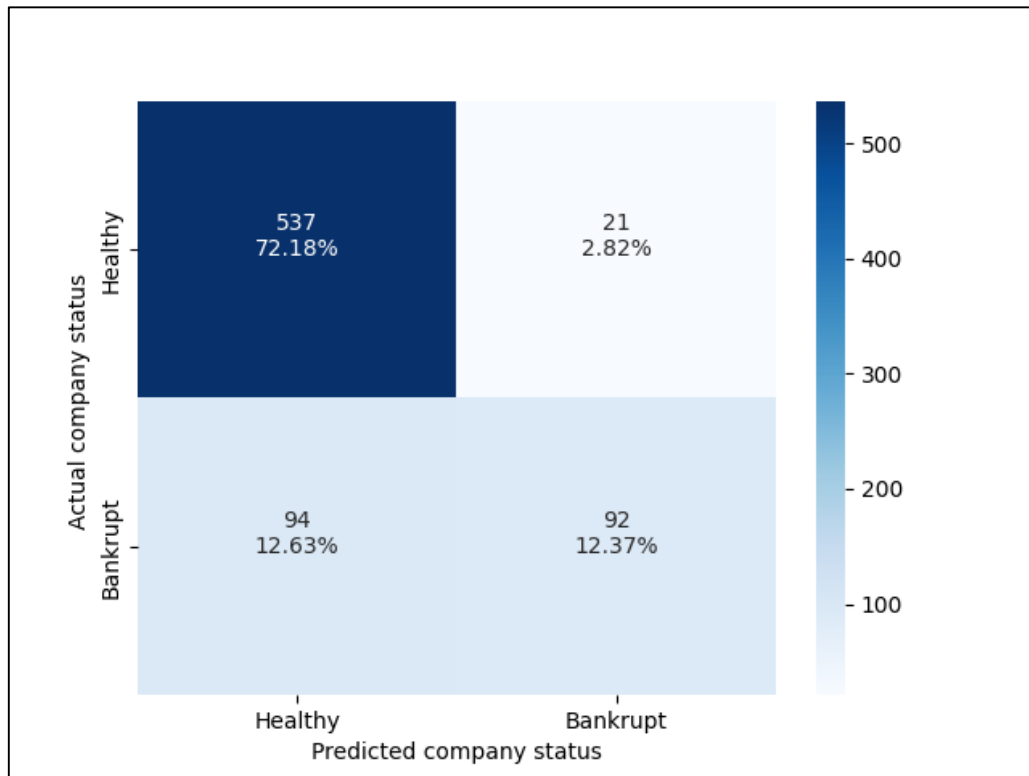
Διάγραμμα 59 - NV – Test set – Fold 3



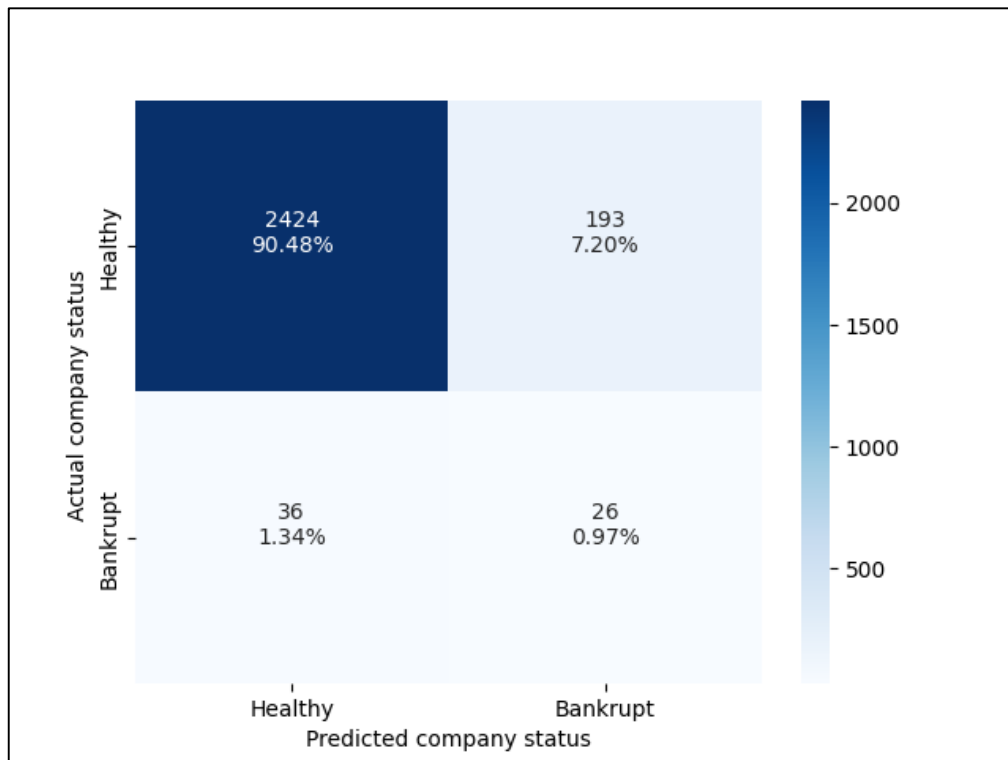
Διάγραμμα 60 - NV – Train set – Fold 4



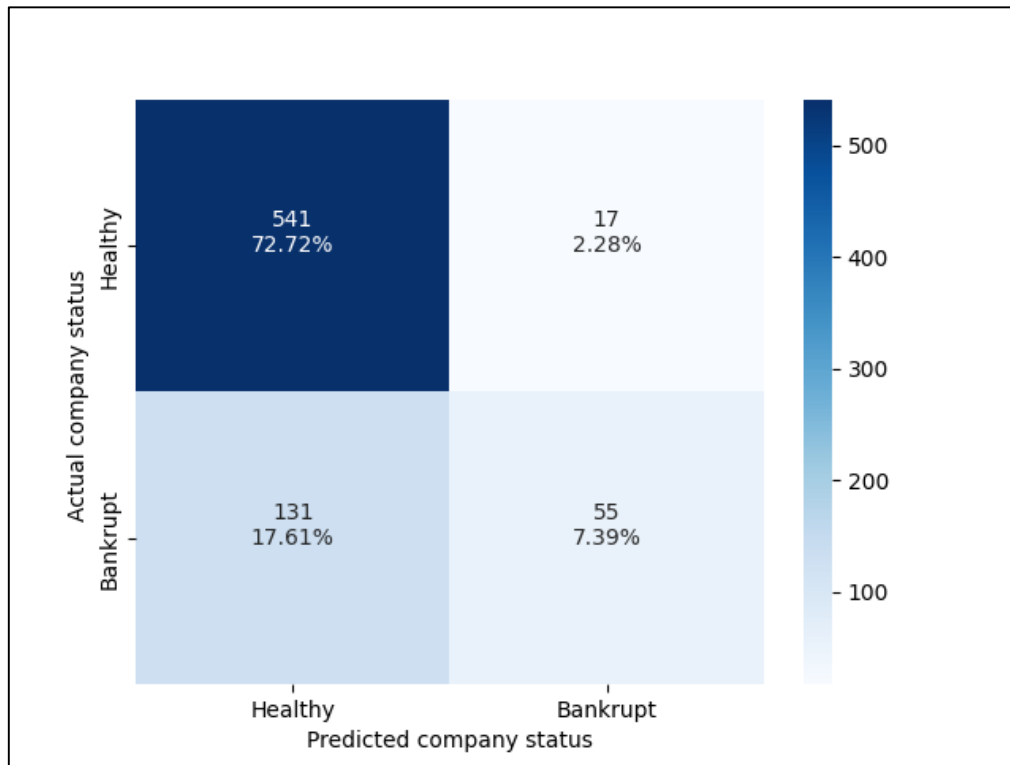
Διάγραμμα 61 - NV – Test set – Fold 4



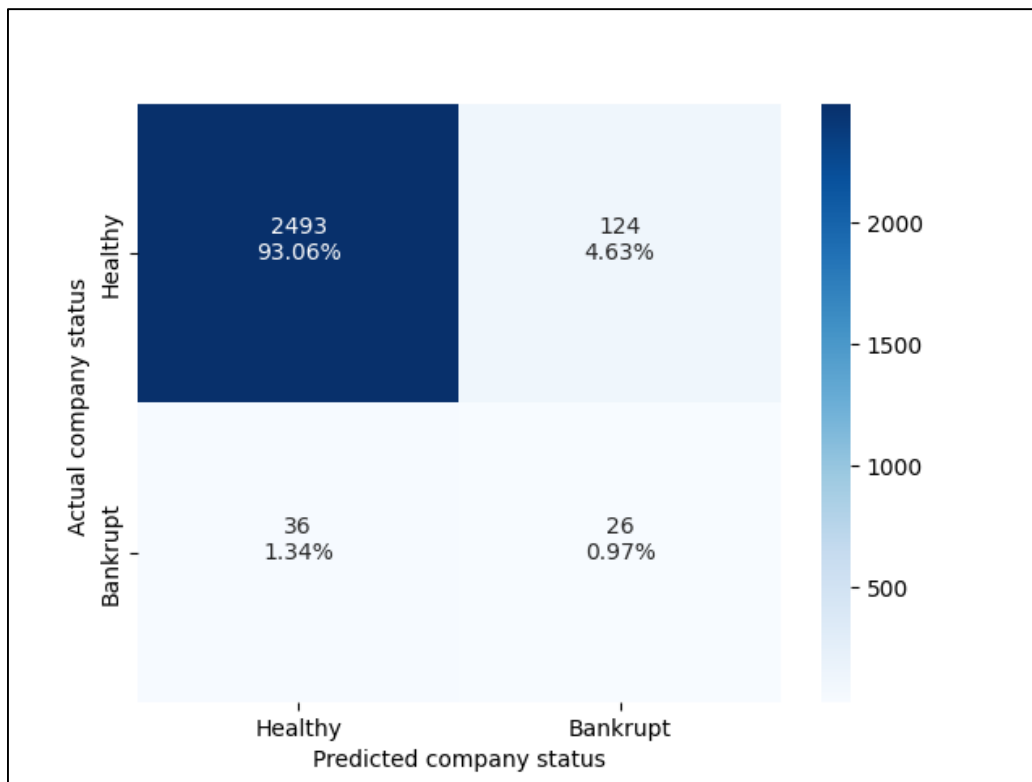
Διάγραμμα 62 - SVM – Train set – Fold 1



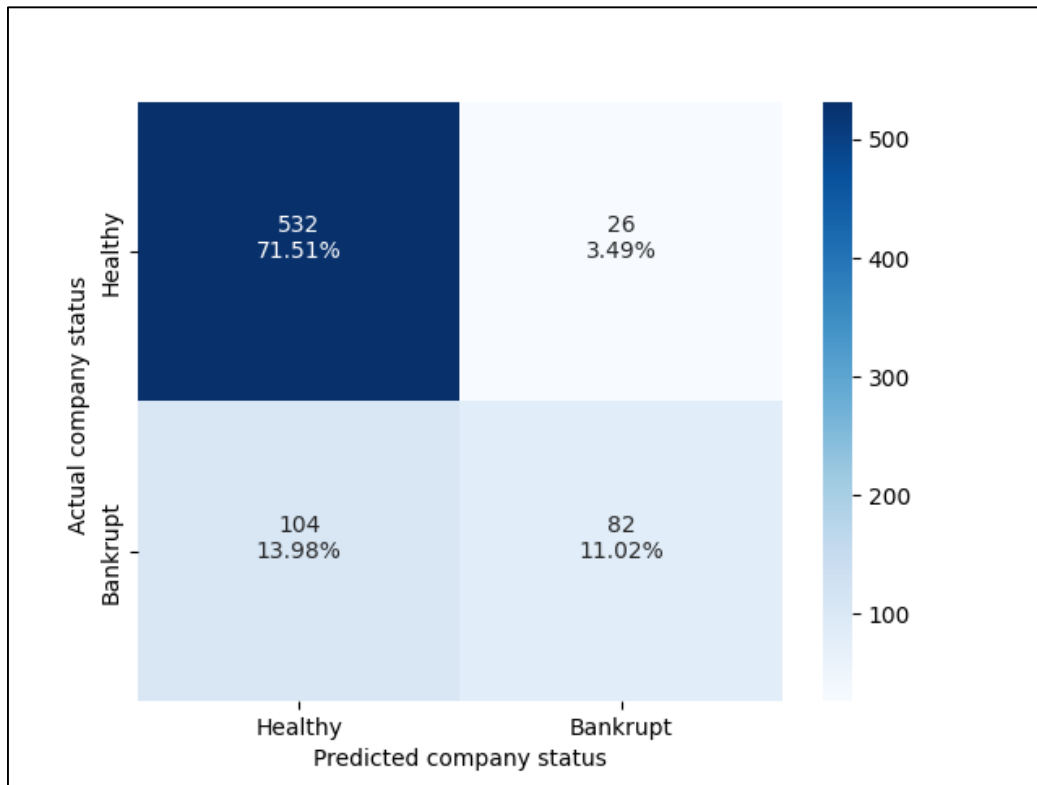
Διάγραμμα 63 - SVM – Test set – Fold 1



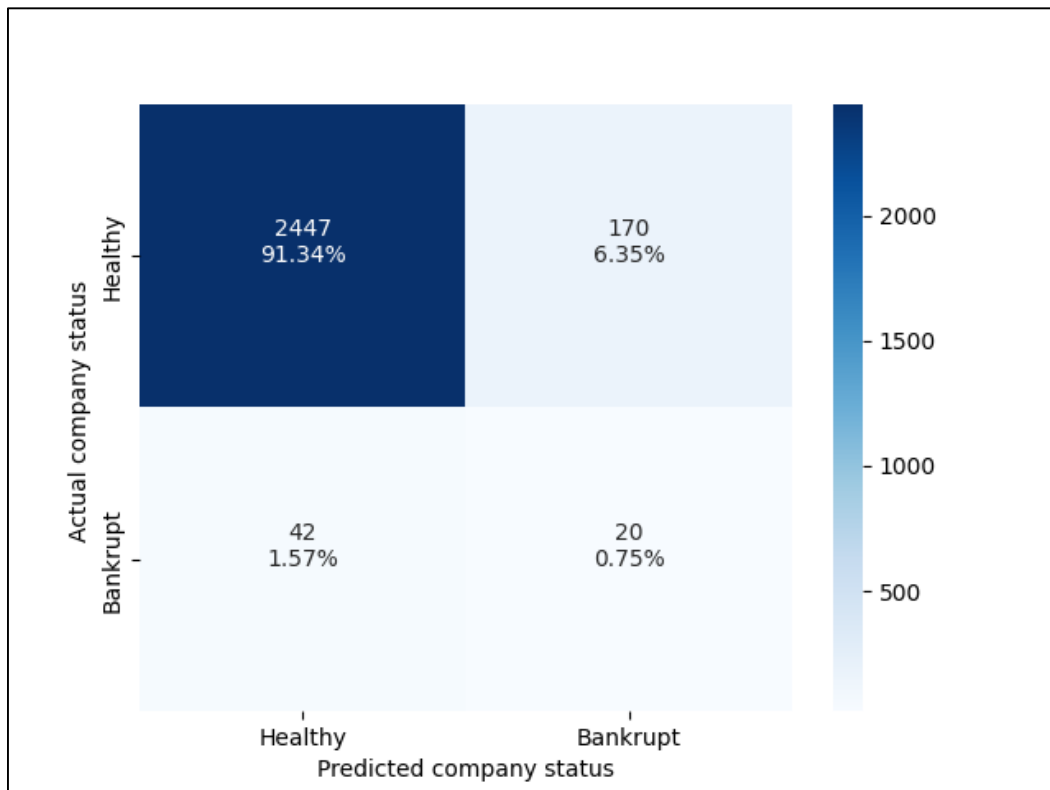
Διάγραμμα 64 - SVM – Train set – Fold 2



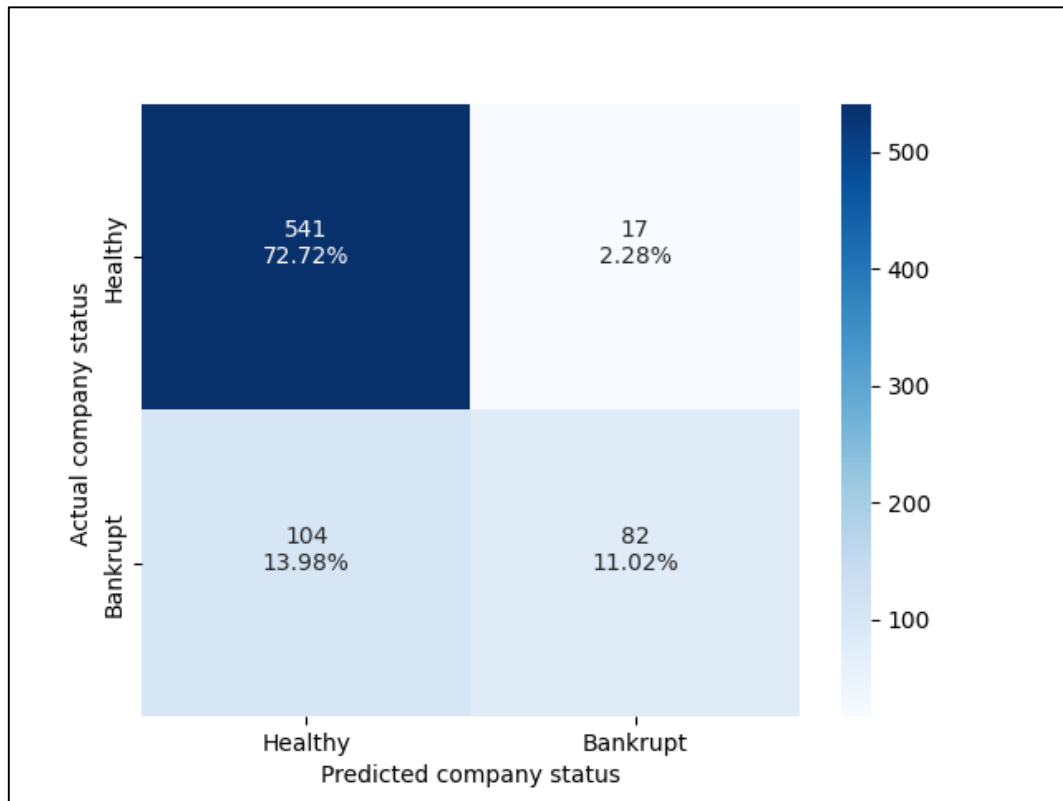
Διάγραμμα 65 - SVM – Test set – Fold 2



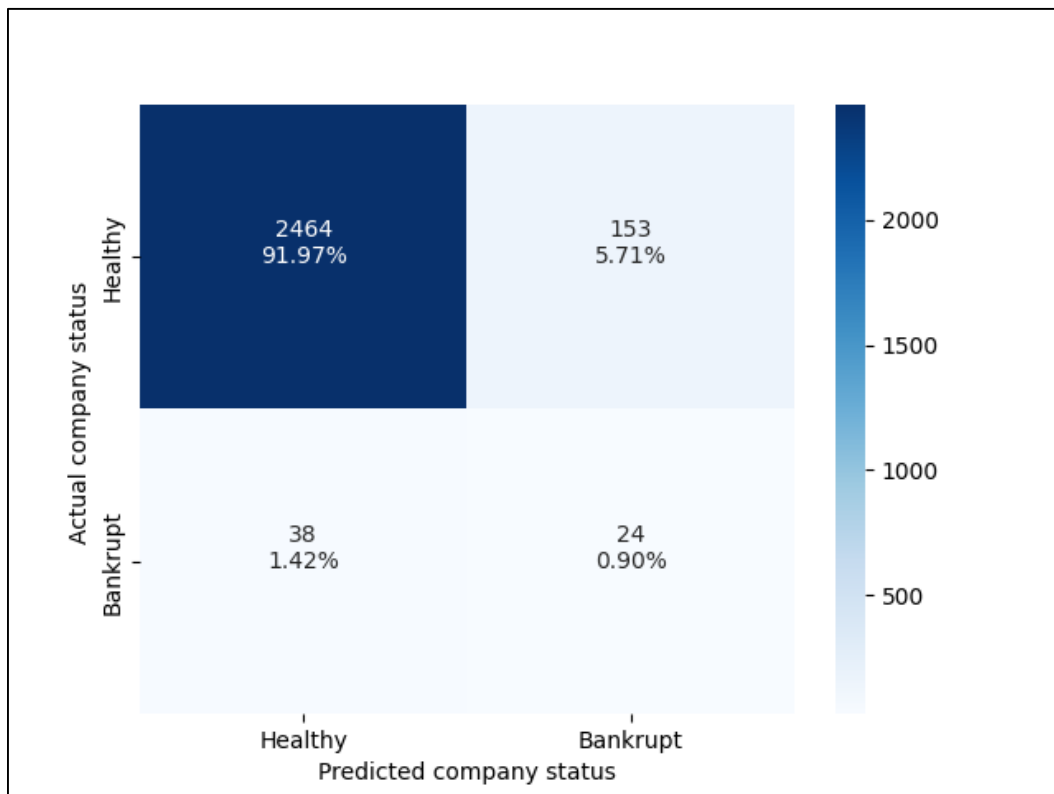
Διάγραμμα 66 - SVM – Train set – Fold 3



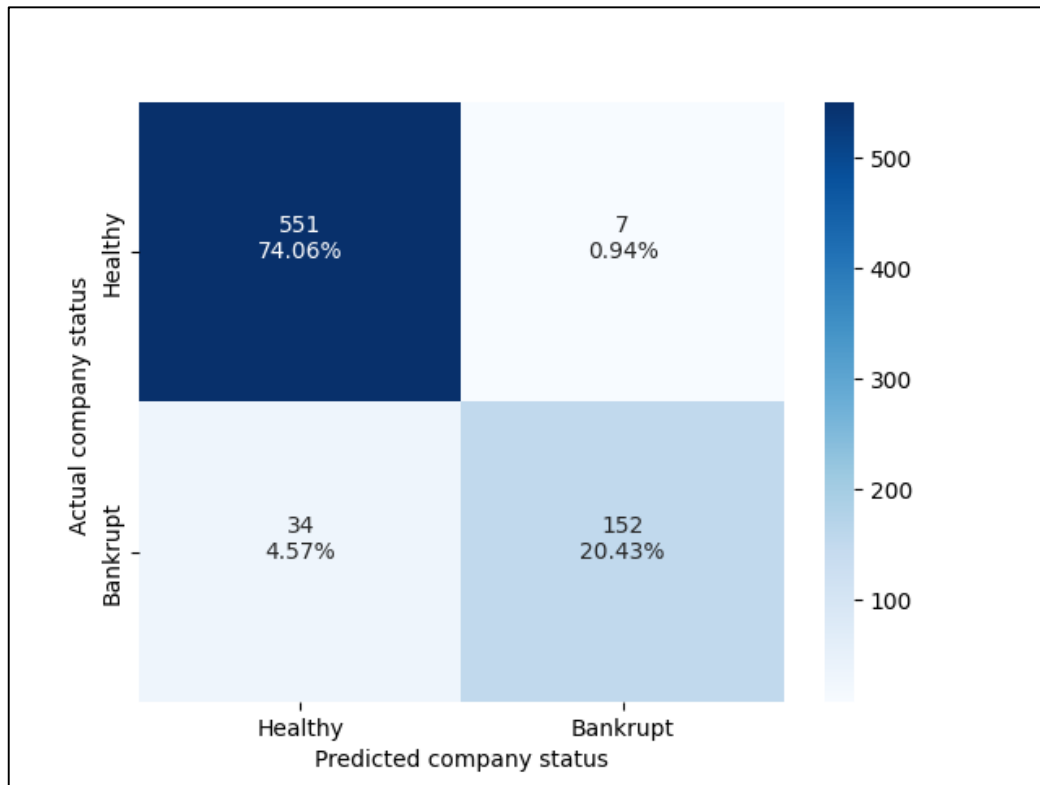
Διάγραμμα 67 - SVM – Test set – Fold 3



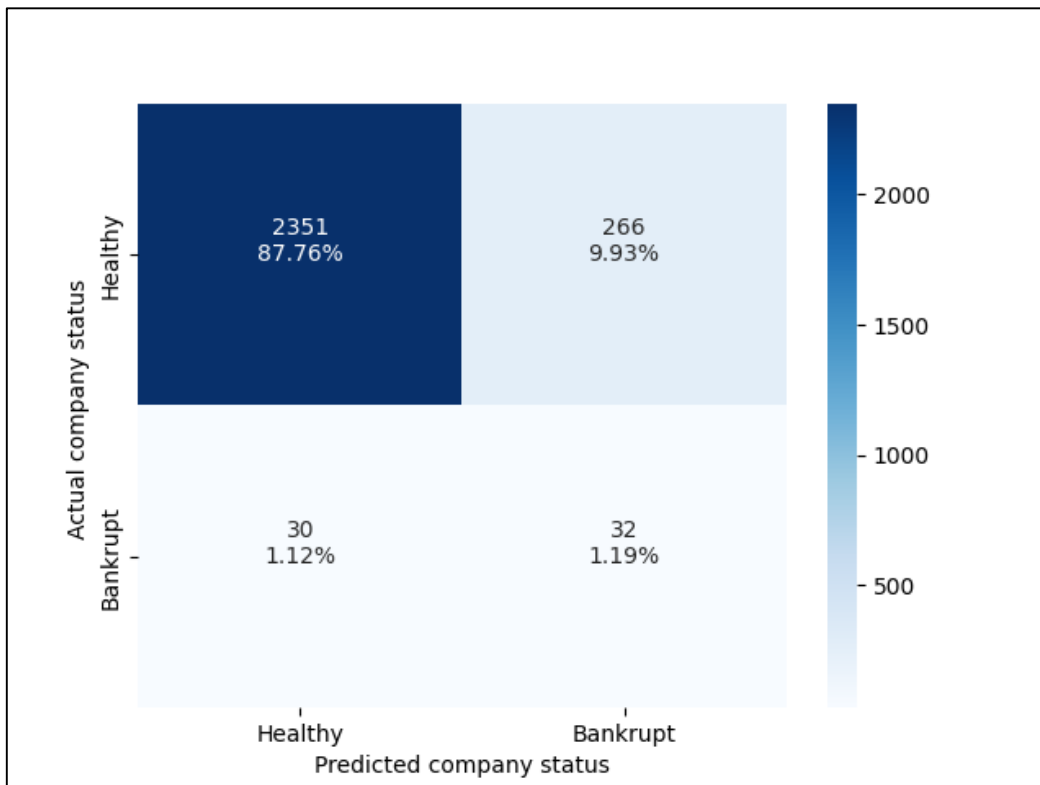
Διάγραμμα 68 - SVM – Train set – Fold 4



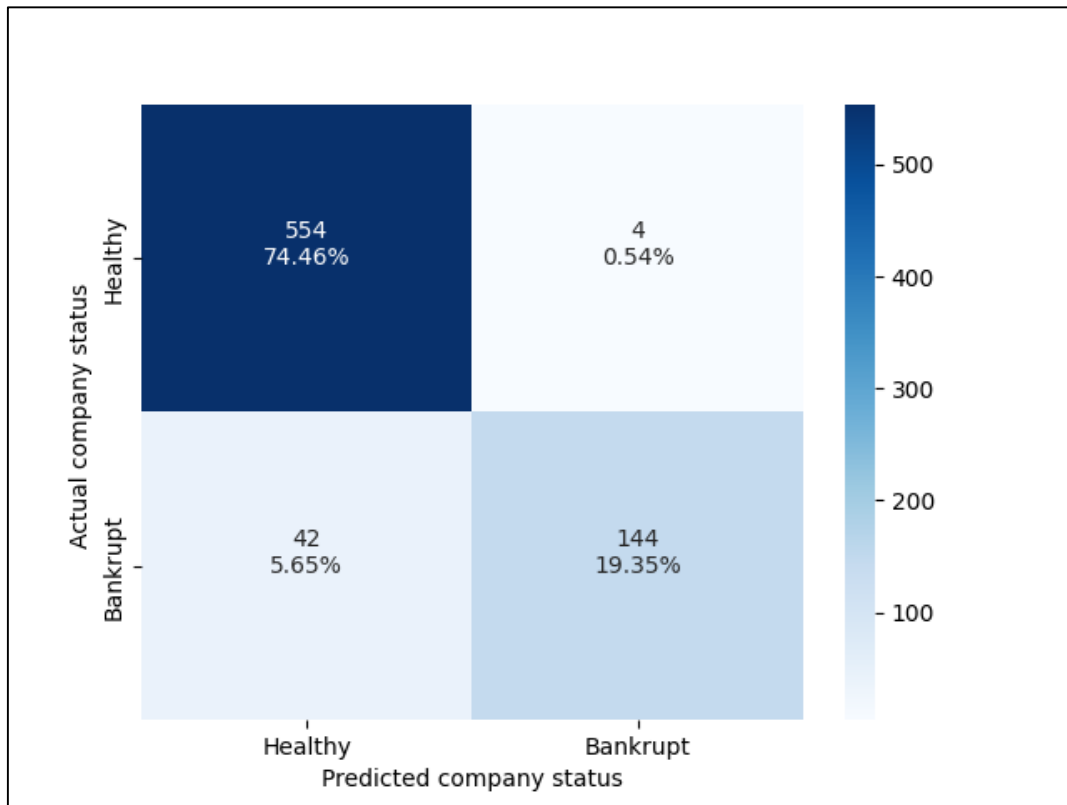
Διάγραμμα 69 - SVM – Test set – Fold 4



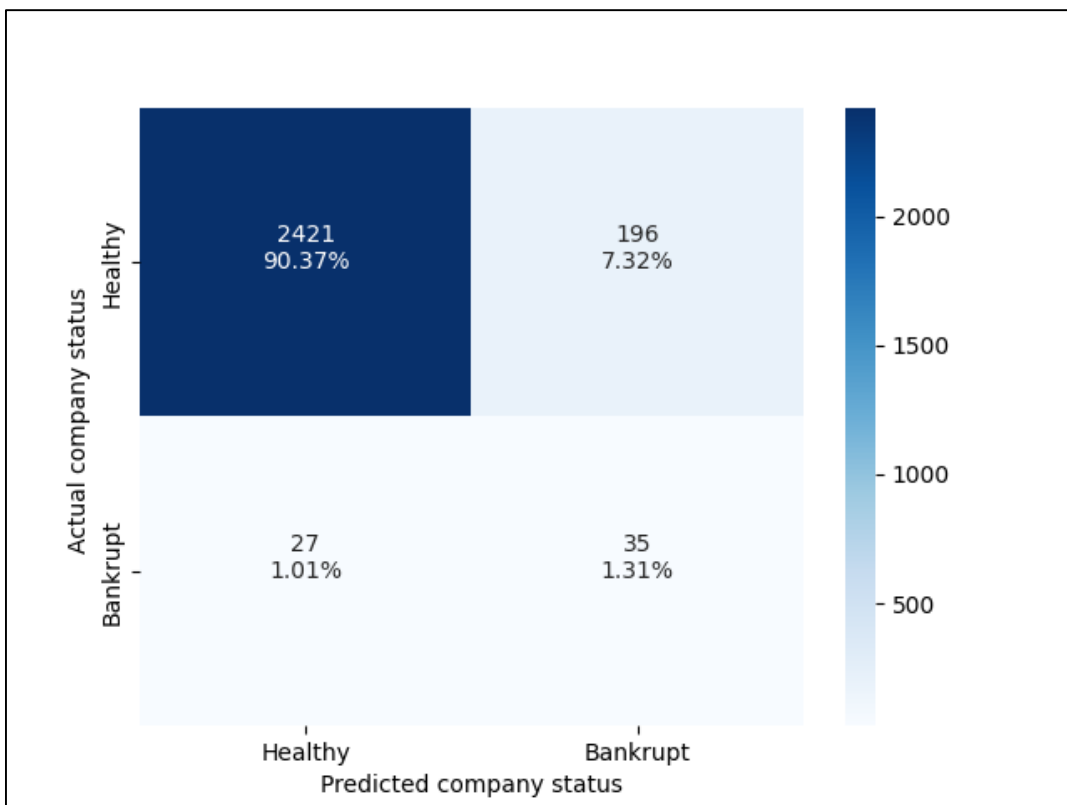
Διάγραμμα 70 - GB – Train set – Fold 1



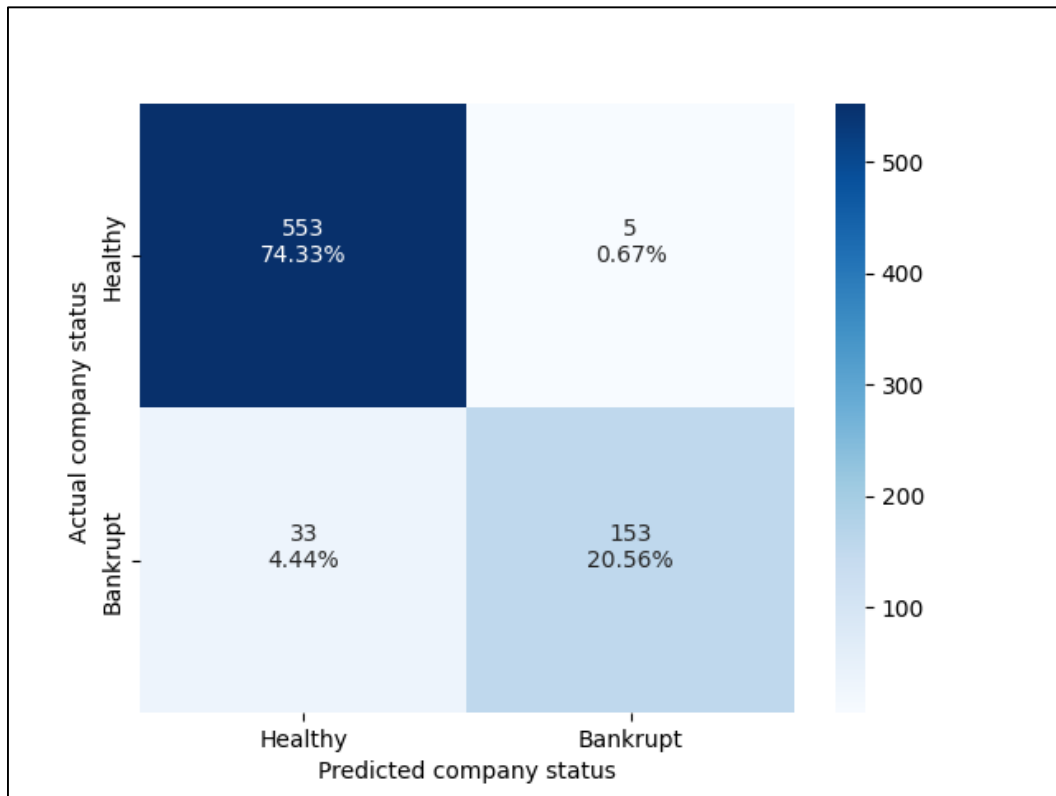
Διάγραμμα 71 - GB – Test set – Fold 1



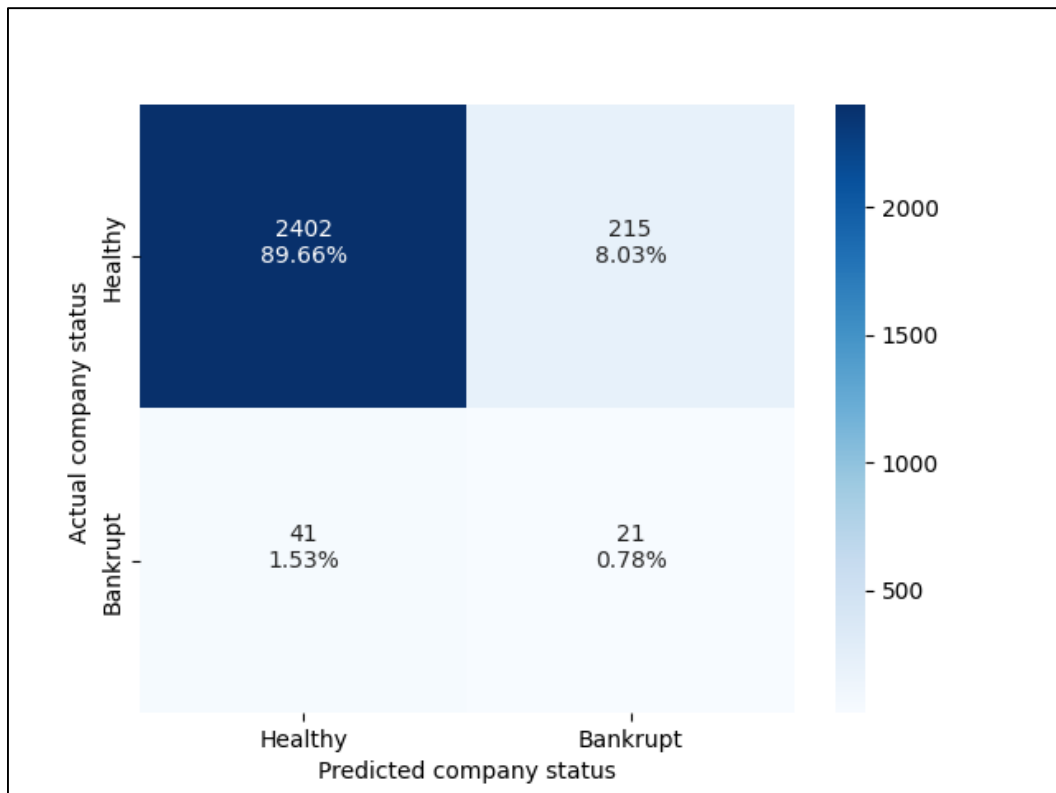
Διάγραμμα 72 - GB – Train set – Fold 2



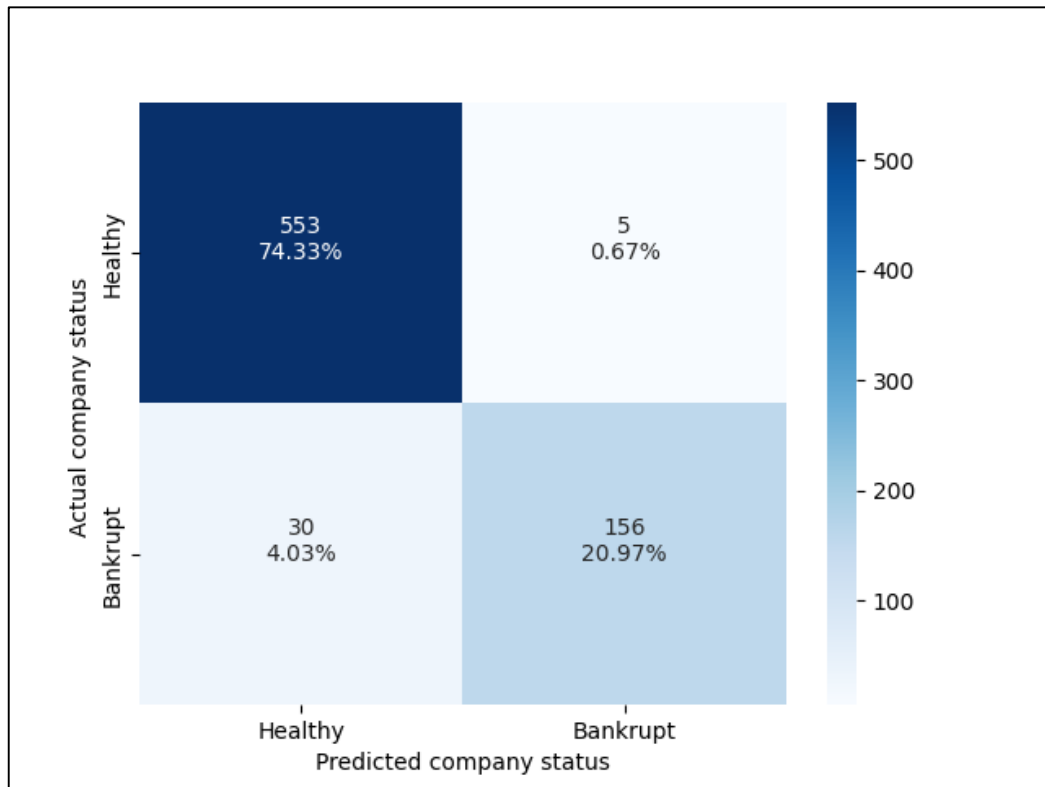
Διάγραμμα 73 - GB – Test set – Fold 2



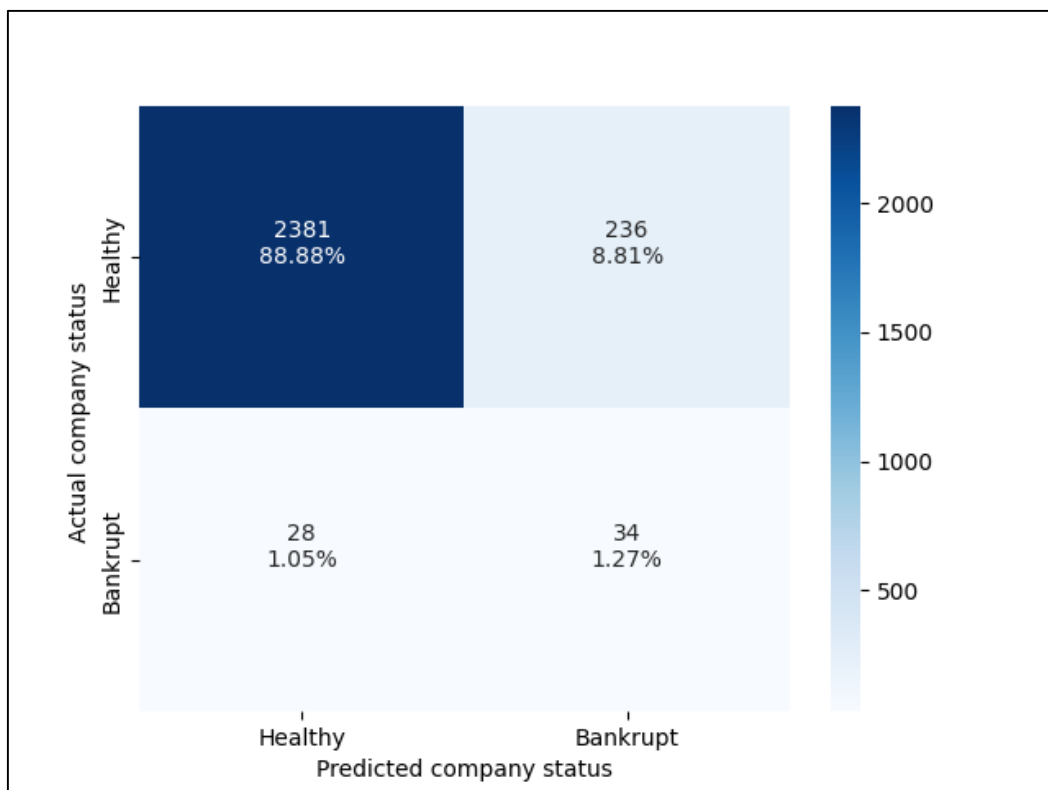
Διάγραμμα 74 - GB – Train set – Fold 3



Διάγραμμα 75 - GB – Test set – Fold 3

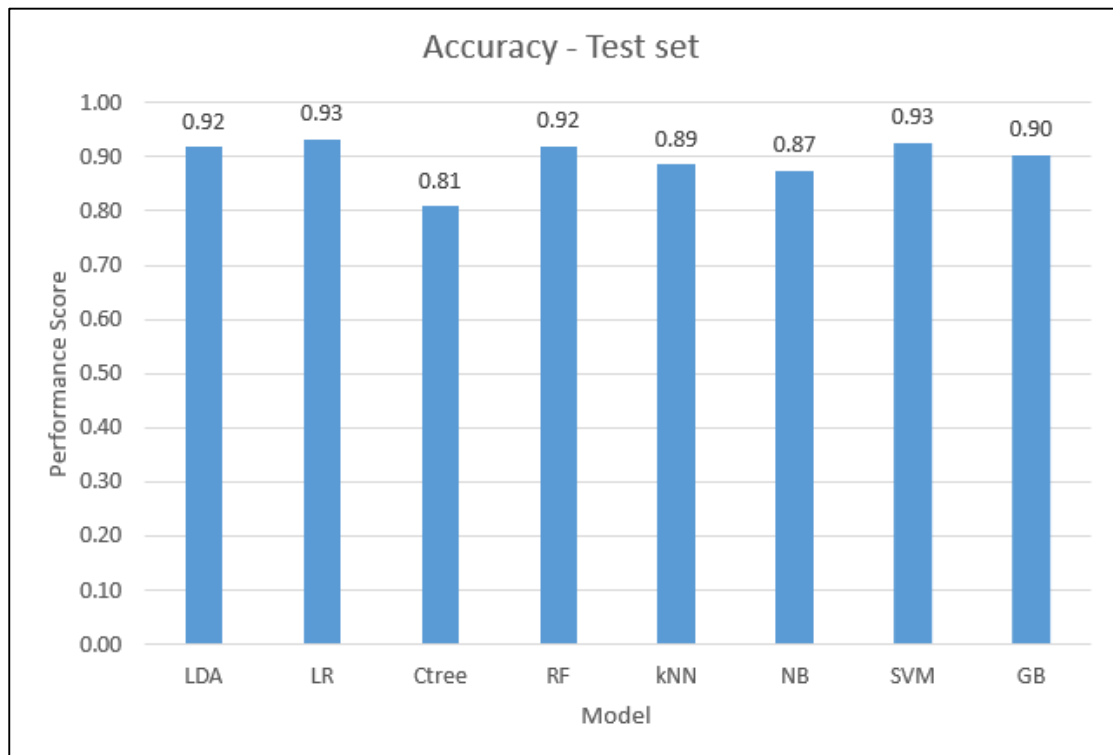


Διάγραμμα 76 - GB – Train set – Fold 4

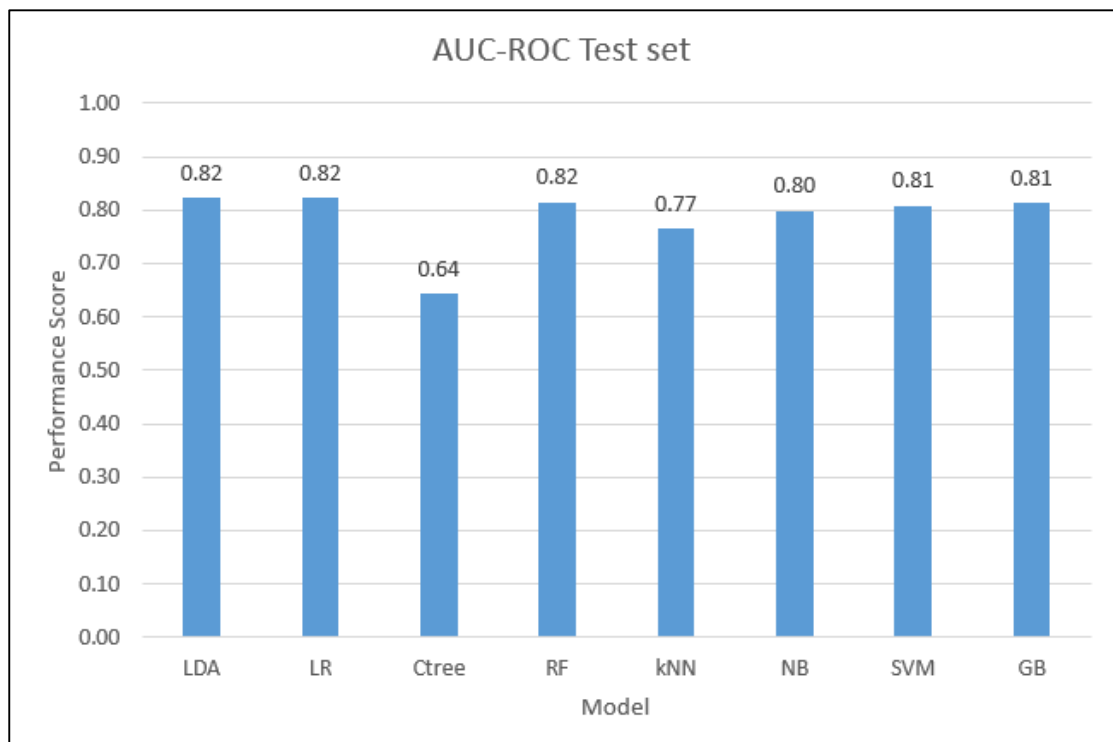


Διάγραμμα 77 - GB – Test set – Fold 4

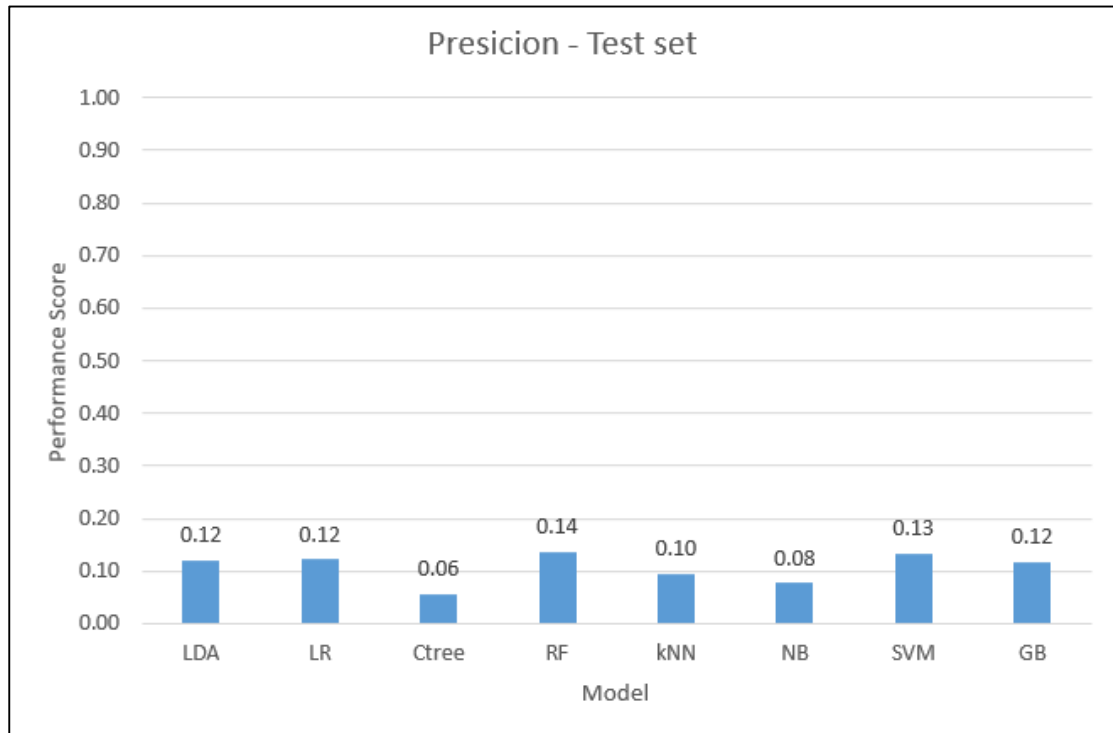
Παράρτημα 3 – Αποτελέσματα επίδοσης μοντέλων



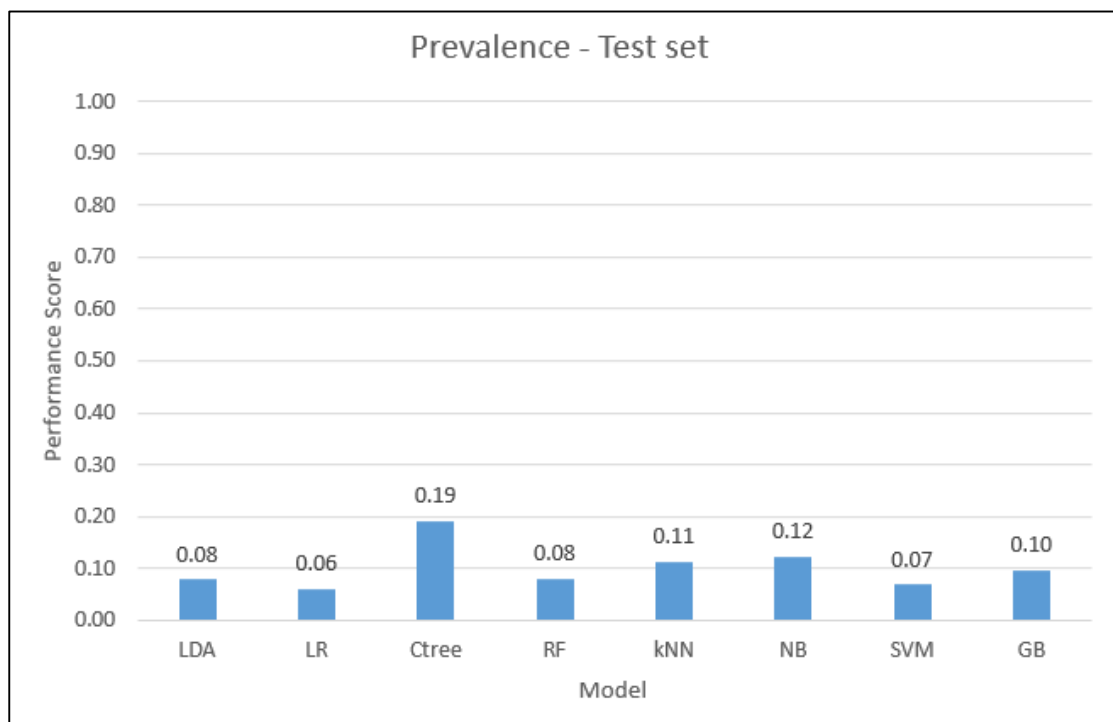
Διάγραμμα 78 - Accuracy Test set - 4 fold Average



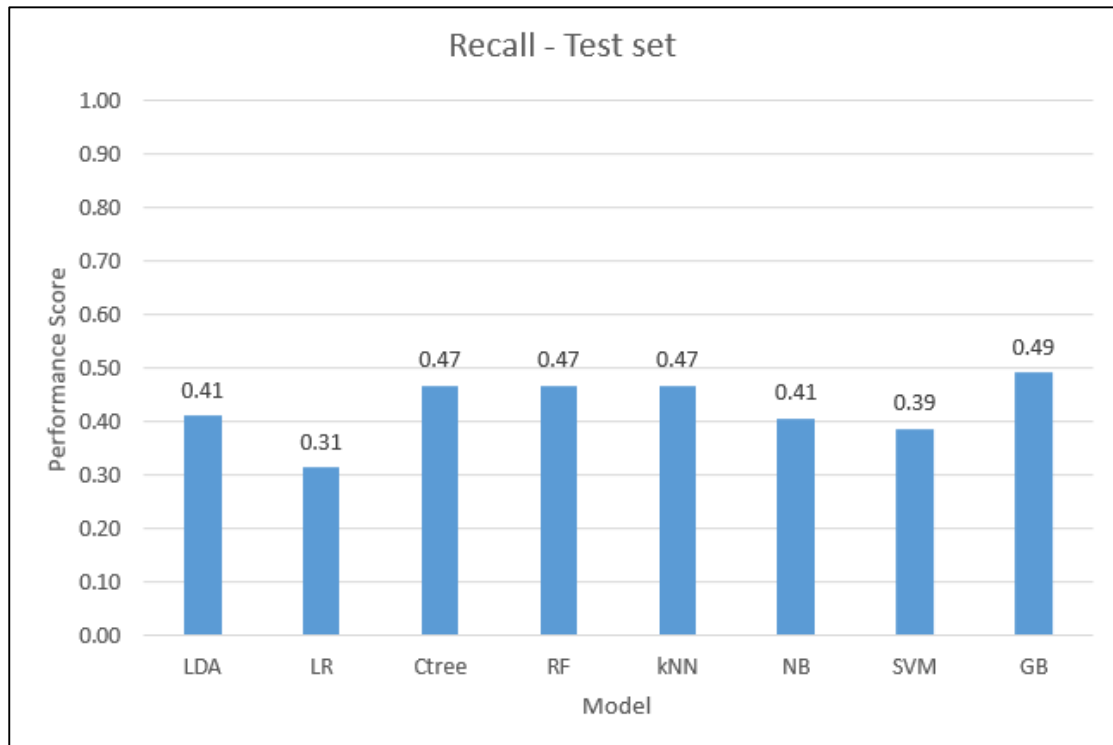
Διάγραμμα 79 – AUC-ROC - Test set - 4 fold Average



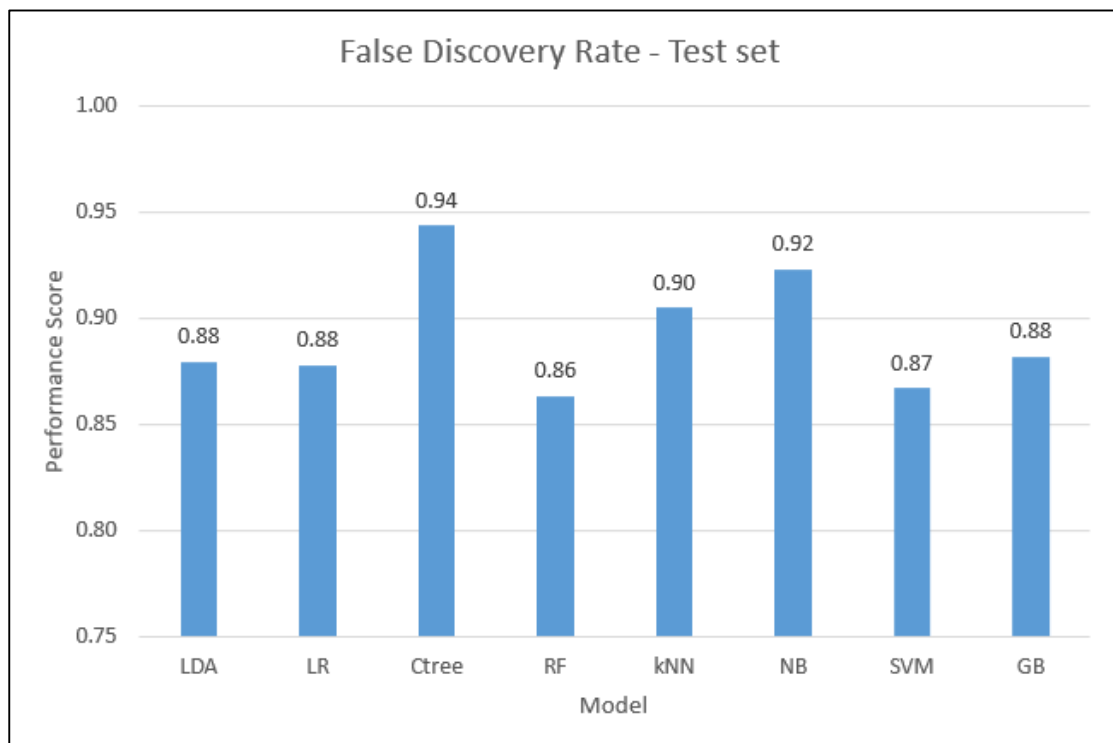
Διάγραμμα 80 - Precision Test set - 4 fold Average



Διάγραμμα 81 - Prevalence Test set - 4 fold Average



Διάγραμμα 82 - Recall Test set - 4 fold Average



Διάγραμμα 83 - FDR Test set - 4 fold Average