

**ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ**

**ΜΕΘΟΔΟΙ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ ΓΙΑ ΤΗΝ  
ΠΡΟΒΛΕΨΗ ΧΡΕΟΚΟΠΙΑΣ**

**Ελευθεριάδου Ίννα  
Α.Μ. 100/05**

**Επιβλέπων Καθηγητής: Ευστάθιος Κύρκος**

**ΑΤΕΙ ΘΕΣΣΑΛΟΝΙΚΗΣ 2010**

**ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΚΑΙ ΟΙΚΟΝΟΜΙΑΣ**

**ΤΜΗΜΑ ΛΟΓΙΣΤΙΚΗΣ**

# ΠΕΡΙΕΧΟΜΕΝΑ

Περιεχόμενα.....	2
Περίληψη.....	4
1.Εισαγωγή.....	4
1.1 Διάρθρωση της εργασίας.....	4
2. Τι είναι το Data mining;.....	5
3. Παρουσίαση των άρθρων.....	6
<b>Εργασία 1. Threshold accepting trained principal component neural network and feature subset selection: Application to bankruptcy prediction in banks.....</b>	<b>6</b>
<b>Εργασία 2. An integrative model with subject weight (IMSW) based on neural network learning for bankruptcy prediction.....</b>	<b>10</b>
<b>Εργασίας 3. Using neural network ensembles for bankruptcy prediction and credit scoring.....</b>	<b>11</b>
<b>Εργασία 4. Bankruptcy prediction with neural logic networks by means of grammar-guided genetic programming.....</b>	<b>14</b>
<b>Εργασία 5. Early bankruptcy prediction using ENPC.....</b>	<b>15</b>
<b>Εργασία 6. An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring.....</b>	<b>18</b>
<b>Εργασία 7. Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods:A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey.....</b>	<b>20</b>
<b>Εργασία 8. Financial failure prediction using efficiency as a predictor.....</b>	<b>23.</b>
<b>Εργασία 9. Differential evolution trained wavelet neural networks: Application to bankruptcy prediction in banks.....</b>	<b>25</b>
<b>Εργασία 10. A selective ensemble based on expected probabilities for bankruptcy prediction.....</b>	<b>27</b>

4. Αξιολόγηση – Σύγκριση των άρθρων.....	32
5. Συμπεράσματα.....	35
Αναφορές.....	

## ΠΕΡΙΛΗΨΗ

Αυτό το έγγραφο παρουσιάζει διάφορες μελέτες που έχουν γίνει πάνω στο θέμα της πρόβλεψης χρεοκοπίας. Αναλύοντας τις εργασίες βγαίνει το συμπέρασμα ότι αν και η κάθε μελέτη χρησιμοποιεί διαφορετικές μεθόδους εξόρυξης δεδομένων και διαφορετικά σύνολα δεδομένων, όλες έχουν κοινό στόχο, την επίτευξη καλύτερης πρόβλεψης πτώχευσης. Είναι σημαντικό να σημειωθεί ότι σε κάθε μελέτη γίνεται αναφορά στην πηγή από την οποία προέρχονται τα δεδομένα, στη χώρα στην οποία αναφέρεται η μελέτη καθώς και στην χρονική περίοδο στην οποία αναφέρεται. Τέλος γίνεται σύγκριση όλων αυτών των άρθρων για να προκύψει ένα τελικό αποτέλεσμα.

### 1. ΕΙΣΑΓΩΓΗ

Λόγω της ριζικής αλλαγής στη παγκόσμια οικονομία, η οικονομική πρόβλεψη αποτυχίας μιας εταιρίας παίζει όλο και περισσότερο σημαντικό ρόλο. Συχνά η οικονομική αποτυχία εμφανίζεται όταν έχει μια εταιρία τις χρόνιες και σοβαρές απώλειες ή και όταν η εταιρία γίνεται αφερέγγυη με τα στοιχεία του παθητικού. Οι ευρέως διαδεδομένες αιτίες και τα συμπτώματα της οικονομικής αποτυχίας περιλαμβάνουν τη φτωχή διαχείριση, την αυταρχική ηγεσία και τις δυσκολίες να λειτουργήσουν επιτυχώς στην αγορά. Η εταιρική πτώχευση δεν προκαλεί μόνο σημαντικές απώλειες στην επιχειρησιακή κοινότητα αλλά και στην κοινωνία συνολικά. Επομένως, τα ακριβή μοντέλα πρόβλεψης πτώχευσης είναι κρίσιμης σπουδαιότητας στους διάφορους συμμετάσχοντες όπως είναι οι επενδυτές, οι υπάλληλοι, οι μέτοχοι και άλλα ενδιαφερόμενα μέλη.

Η ανάπτυξη των οικονομικών μοντέλων για την πρόβλεψη των επιχειρησιακών αποτυχιών μπορούν να θεωρηθούν “συστήματα έγκαιρης προειδοποίησης”, που αποδεικνύονται πολύ χρήσιμα για τους διευθυντές και τις αρμόδιες αρχές που μπορούν να αποτρέψουν το περιστατικό των αποτυχιών. Επιπλέον, αυτά τα μοντέλα είναι σε θέση να βοηθήσουν τους ιθύνοντες των οικονομικών οργάνων να αξιολογήσουν και να επιλέξουν τις εταιρίες για να συνεργαστούν ή να επενδύσουν

#### 1.1 Διάρθρωση της εργασίας

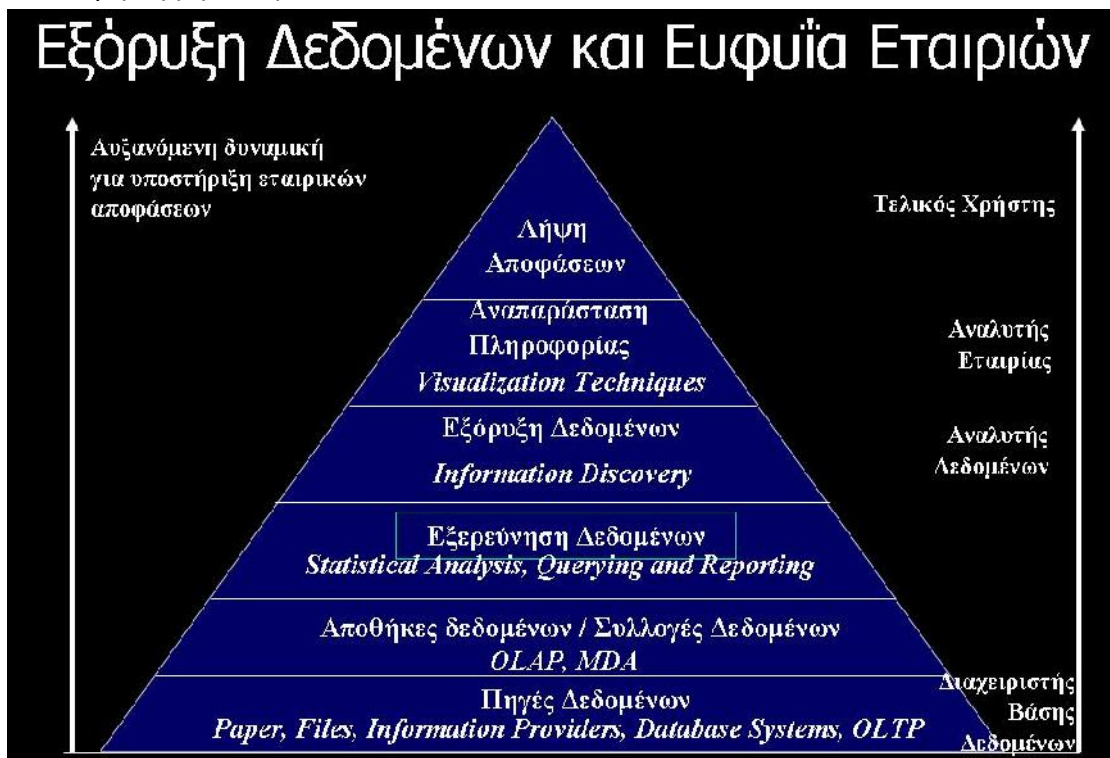
Η υπόλοιπη της εργασίας ξεκινάει με την παρουσίαση του Data Mining στο Κεφάλαιο 2. Ακολουθεί στο Κεφάλαιο 3 η αναλυτική παρουσίαση των εργασιών που μελετήθηκαν. Στο Κεφάλαιο 4 γίνεται σύγκριση και αξιολόγηση των

αποτελεσμάτων των ερευνών. Τέλος στο Κεφάλαιο 5 αναφέρονται τα συμπεράσματα.

## 2. Data mining;

Η εξόρυξη δεδομένων παίρνει το όνομά της από τις ομοιότητες που έχει η αναζήτηση πολύτιμων επιχειρηματικών πληροφοριών σε μια μεγάλη βάση δεδομένων με την εξόρυξη πολύτιμων ορυκτών από μια ορεινή μάζα. Η εξόρυξη δεδομένων δηλαδή η εξαγωγή των κρυμμένων πληροφοριών από μεγάλες βάσεις δεδομένων, αποτελεί ισχυρή τεχνολογία με πολύ μεγάλες δυνατότητες να βοηθήσει τις εταιρείες να επικεντρωθούν στις πιο σημαντικές πληροφορίες που περιέχονται στις βάσεις δεδομένων τους. Με τη χρήση περίπλοκων αλγορίθμων οι χρήστες έχουν τη δυνατότητα να αναγνωρίζουν τα χαρακτηριστικά των επιχειρηματικών διαδικασιών. Οι αναλύσεις που προσφέρονται από την εξόρυξη δεδομένων πάνε πέρα από τις στατιστικές αναλύσεις. Τα εργαλεία εξόρυξης δεδομένων είναι δυνατόν να απαντήσουν σε επιχειρηματικές ερωτήσεις, πράγμα που, παραδοσιακά, ήταν πολύ χρονοβόρο. Η επιστήμη της εξόρυξης δεδομένων έχει ευρύ φάσμα και με πολλαπλές εφαρμογές.

Τα συστήματα Εξόρυξης Δεδομένων είναι φτιαγμένα για να διαχειρίζονται τεράστιες ποσότητες πληροφορίας, να μπορούν να έχουν και να ανατρέχουν σε ιστορικά δεδομένα, να χειρίζονται από υψηλόβαθμα στελέχη εταιριών έτσι ώστε σε μικρό χρονικό διάστημα να έχουν οπτική αναπαράσταση πληροφοριών για την καλύτερη λήψη αποφάσεων.



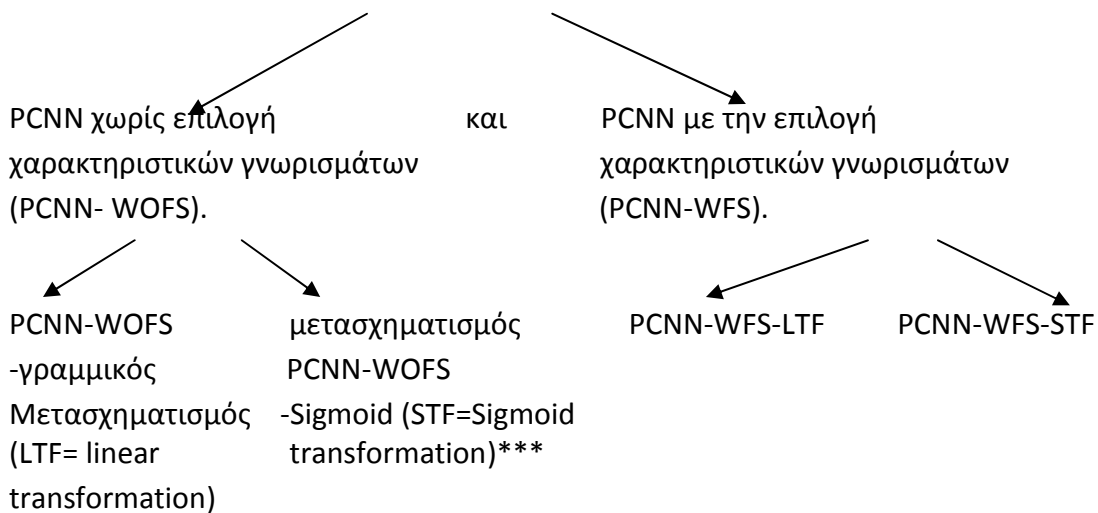
### 3. ΠΑΡΟΥΣΙΑΣΗ ΤΩΝ ΑΡΘΡΩΝ

#### Εργασία 1. Threshold accepting trained principal component neural network and feature subset selection: Application to bankruptcy prediction in banks.

Αυτό το άρθρο προτείνει μια εφαρμογή της νέας αρχιτεκτονικής κύριων συστατικών νευρωνικών δικτύων (PCNN principal component neural network) στο πρόβλημα πρόβλεψης πτώχευσης στις εμπορικές τράπεζες. Παράλληλα παρουσιάζει έναν αλγόριθμο βασισμένο στην threshold accepting (TA) για να εκπαιδεύσει το PCNN. Η αποτελεσματικότητα του αλγορίθμου εξετάζεται στο σύνολο δεδομένων Ισπανικών τραπεζών και στο σύνολο δεδομένων Τούρκικων τραπεζών.

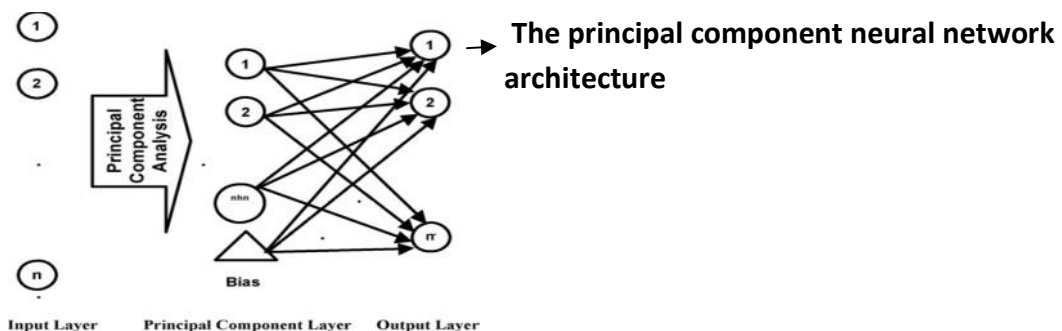
Παρακάτω παρουσιάζονται οι τέσσερις παραλλαγές του PCNN που εκπαιδεύονται από έναν TA-based αλγόριθμο.

Αρχικά, αναπτύχθηκαν δύο παραλλαγές του ταξινομητή PCNN:



\* **WOFS** = without feature selection

\*\*\* Η sigmoid λειτουργία μεταφοράς χρησιμοποιείται μεταξύ της εισόδου και του κρυμμένου στρώματος και μεταξύ του κρυμμένου και στρώματος εξόδου.



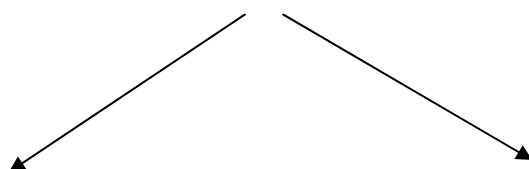
Το προτεινόμενο hybrid neural network και FSS feature subset selection algorithm εξετάστηκαν για το πρόβλημα πρόβλεψης πτώχευσης σε 2 σύνολα:

Ισπανικό σύνολο δεδομένων τραπεζών που —> αποτελείται από 66 τράπεζες με 9 οικονομικές αναλογίες (ratios) η κάθε μια.

Τουρκικό σύνολο δεδομένων τραπεζών που —> αποτελείται από 40 τράπεζες με 12 οικονομικές αναλογίες (ratios) η κάθε μία.

Και στις περιπτώσεις που αφορούν τις Ισπανικές τράπεζες αλλά και στις περιπτώσεις που αφορούν τις Τούρκικες τράπεζες χρησιμοποιείται η 10-fold cross-validation.

Προκειμένου να ερευνηθεί η δύναμη των παραλλαγών PCNN, όπου το κρυμμένο στρώμα (hidden layer) είναι εμφανώς απόν, συγκρίναμε την απόδοσή τους με:



την ανάλυση-βασισμένη στην αποδοχή κατώτατων ορίων του νευρωνικού δικτύου (PCA- TANN principal component analysis-based threshold accepting neural network)

και την ανάλυση-βασισμένη στο πίσω νευρωνικό δίκτυο διάδοσης κύριων συστατικών (PCA- BPNN principal component analysis-based back propagation neural network)

Σε αντίθεση με το PCNN στα (PCA-TANN) και (PCA-BPNN) ένα κρυμμένο στρώμα είναι **ΠΑΡΟΝ**

### ROC καμπύλες

Ισπανικές τράπεζες

-Οι παραλλαγές: PCANN-WOFS-LTF, PCANN-WFS-LTF ξεπέρασαν όλες τις άλλες τεχνικές όσον αφορά το ποσοστό και την περιοχή ταξινόμησης κάτω από την καμπύλη ROC (AUC)

Τουρκικές τράπεζες

-PCNN-WFS-LTF ξεπέρασε όλες τις τεχνικές με το ποσοστό ταξινόμησης του 100% και AUC 10.000

- PCNN-WOFS-LTF ξεπέρασαν όλες τις άλλες τεχνικές με το ποσοστό ταξινόμησης 96,6% και το AUC 9660.

-Το διάφορο(The variant) PCA- TANN είχε καλή απόδοση με το ποσοστό ταξινόμησης 97,5% και AUC 9840

- PCNN-WFSLTF είναι δίπλα στο PCNN WOFS-LTF με το ποσοστό ταξινόμησης 92,5 και AUC 9435.

- TANN δίπλα σε PCA-TANN με το ποσοστό ταξινόμησης του 91.6% και AUC 9000, PCNN-WOFS-LTF εκτελείται καλά όσον αφορά το ποσοστό ταξινόμησης 97,5%

- TANN ξεπέρασε όλες τις άλλες τεχνικές όσον αφορά το ποσοστό ταξινόμησης 97,5% αλλά έχει την πολύ χαμηλή ιδιομορφία.

- Από τις τέσσερις τεχνικές TANN, PCA- TANN, και PCABPNN οι παραλλαγές TANN έχουν καλύτερο αποτέλεσμα.

- Οι παραλλαγές PCNN WFS-STF, PCNN-WOFS-STF δεν εκτέλεσαν καλά και κατέλαβε τις τελευταίες θέσεις

-Από TANN και PCA- TANN, το TANN είχε καλό αποτέλεσμα.

Το συμπέρασμα που προκύπτει από τα παραπάνω είναι ότι οι PCNN-WOFS/WFS-LTF μπορούν να χρησιμοποιηθούν ως πρότυποι ταξινομητές. Οι παραλλαγές με την sigmoid , PCNN-WOFS - STF και PCNN-WFS - STF δεν έχουν αποδώσει καλά και στις δυο περιπτώσεις.

PCNN= principal component neural network(νευρωνικό δίκτυο κύριων συστατικών)

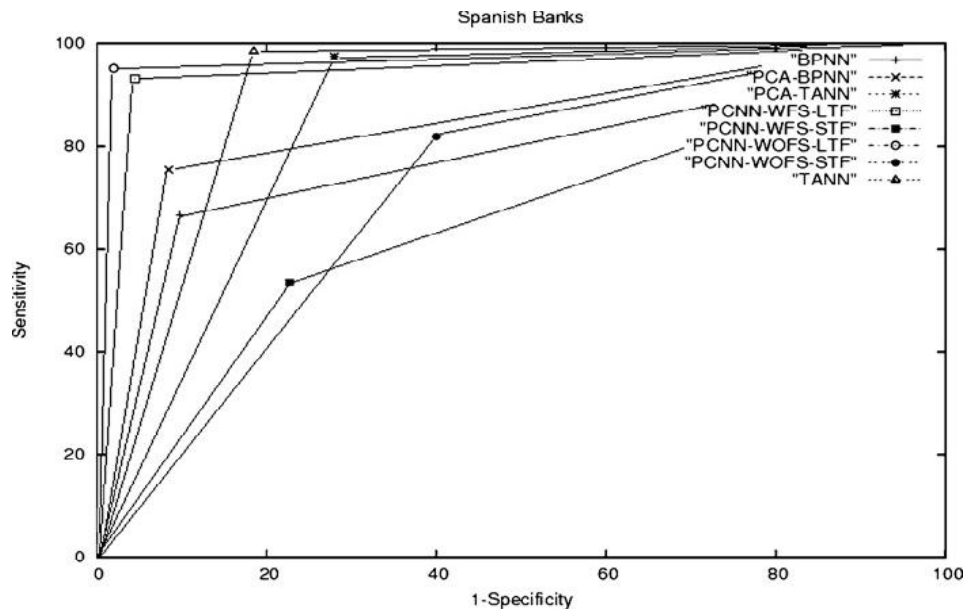
FSS =feature subset selection (επιλογή υποσυνόλων χαρακτηριστικών γνωρισμάτων)

TA = threshold accepting (αποδοχή κατώτατων ορίων)

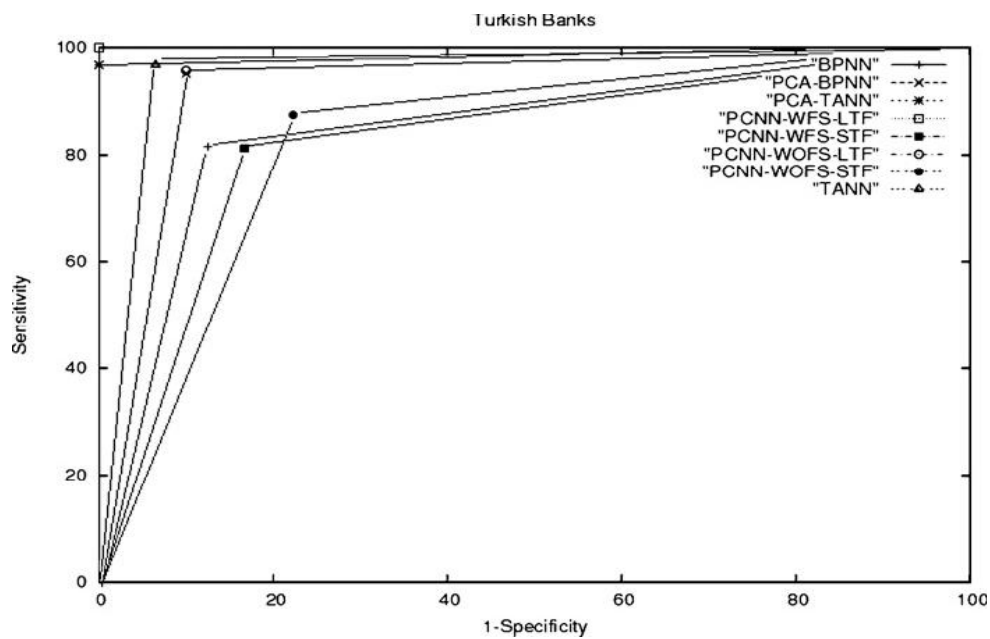
TANN= threshold accepting neural network(κατώτατο όριο που δέχεται το νευρωνικό δίκτυο)

BPNN= back propagation neural network(πίσω νευρωνικό δίκτυο)





**ROC καμπύλες για όλους τους ταξινομητές όσον αφορά τις ισπανικές τράπεζες.**



**ROC καμπύλες για όλους τους ταξινομητές όσον αφορά τις τούρκικες τράπεζες.**

\* Να σημειωθεί ότι όσο υψηλότερη είναι η αξία AUC τόσο καλύτερος είναι ο ταξινομητής

## **Εργασία 2. An integrative model with subject weight (IMSW) based on neural network learning for bankruptcy prediction.**

Προτείνεται μια στρατηγική ολοκλήρωσης σχετικά με το πώς να συνδυαστούν αποτελεσματικά οι τεχνικές στατιστικής και της τεχνητής νοημοσύνης. Ο συνδυασμός της multiple discriminant analysis MDA, της logistic regression, των neural networks NN και της decision trees induction εισάγει ένα περιεκτικό μοντέλο με το υπαγόμενο βάρος βασισμένο στο νευρωνικό δίκτυο ερευνώντας την πρόβλεψη πτώχευσης (an integrative model with subject weight based on neural network learning for bankruptcy prediction) IMSW.

Συγκριτικά με τη Multiple discriminant analysis (MDA), η Logistic regression προτιμάται διότι η Logistic regression μπορεί να ενσωματώσει τα μη γραμμικά αποτελέσματα. Τα Neural networks NN παρέχουν καλή πρόβλεψη για τα σύνθετα προβλήματα όπου οι μεταβλητές συσχετίζονται η μια με την άλλη με έναν μη γραμμικό τρόπο. Ένα ευδιάκριτο χαρακτηριστικό γνώρισμα των νευρωνικών δικτύων έναντι στις παραδοσιακές στατιστικές μεθόδους είναι η ύπαρξη των κρυμμένων στρωμάτων. Οι κρυμμένοι κόμβοι είναι η λειτουργία των ανεξάρτητων μεταβλητών ενώ οι εξαρτημένες μεταβλητές είναι η λειτουργία των κρυμμένων κόμβων. Η rule induction αναφέρεται στους κανόνες που παράγονται από τον αλγόριθμο δέντρων απόφασης στον τομέα της εξόρυξης δεδομένων (Data mining).

### **Source methods (Μέθοδοι πηγής)**

Για να αναπτυχθεί το integrative prediction model, χρησιμοποιήθηκαν οι ακόλουθες τέσσερις μέθοδοι ως πηγές, η MDA multiple discriminant analysis, η logistic regression, τα neural networks και η rule induction.

Το integrative model with subject weight (IMSW) περιλαμβάνει κάποιες διαδικασίες διαμόρφωσης και οι οποίες είναι οι ακόλουθες:

ΒΗΜΑ 1: Αν  $S < 0$  τότε bankrupt  
αλλιώς healthy

ΒΗΜΑ 2:  $I_i = 1$  αν η πρόβλεψη είναι σωστή  
αλλιώς 0

ΒΗΜΑ 3: Αν  $Pr > 0.50$  τότε bankrupt  
αλλιώς healthy

Η εργασία βαρών μπορεί να δοθεί με τη σύγκριση των ακόλουθων τεσσάρων περιπτώσεων:

α) Αν  $I_i = 1$  και  $Pr(Y=1)_i = 0,95 \rightarrow$  bankrupt (υψηλό επίπεδο εμπιστοσύνης)

β) Αν  $I_i = 1$  και  $Pr(Y=1)_i = 0,05 \rightarrow$  healthy (με υψηλό επίπεδο εμπιστοσύνης)

γ) Αν  $I_i = 0$  και  $Pr(Y=1)_i = 0,95 \rightarrow$  ανακριβώς bankrupt

δ) Αν  $I_i = 0$  και  $Pr(Y=1)_i = 0,55 \rightarrow$  ανακριβώς bankrupt με χαμηλό επίπεδο εμπιστοσύνης

\*S= αποτέλεσμα

Pr= πιθανότητα της ύπαρξης πτώχευσης

i= MDA, LOGIC, ANN, RULE  
Xs= dependent variables (ανεξάρτητες μεταβλητές)  
Y= independent variables (εξαρτημένες μεταβλητές)  
I= μεταβλητή δεικτών

### **Πείραμα και αποτελέσματα**

Για το πείραμα χρησιμοποιήθηκαν τα ετήσια οικονομικά στοιχεία που συλλέχθηκαν από το Ταμείο πιστωτικών Εγγυήσεων της Κορέας (Korea Credit Guarantee Fund). Τα στοιχεία αποτελούνται από 900 χρεοκοπημένες εταιρείες και 900 υγιείς την περίοδο : 1999-2005. Υιοθετήθηκε μια πενταπλάσια διαγώνια επικύρωση (five-fold cross validation) για να ενισχύσει τη δυνατότητα γενίκευσης των test results. Η ακρίβεια πρόβλεψης μετρείται με 3 τρόπους:

- correct classification ratio (RC)
- false positive ratio (FP) that refers to incorrectly predicting “bankrupt”
- false negative ratio (FN) for implying incorrectly predicting “healthy”

Το συμπέρασμα με βάση τα πειράματα που έχουν γίνει είναι ότι το IMSW οδηγεί στην καλύτερη απόδοση. Ο μέσος όρος των σωστών αναλογιών ταξινόμησης (RC) του IMSW είναι 78,92%, το οποίο είναι μεγαλύτερο από άλλα μοντέλα. Όσον αφορά τη λάθος ταξινόμηση, το LOGIC δίνει τη χαμηλότερη ψεύτικη θετική αναλογία (FP), 11,79%, ενώ η ANN δίνει τη χαμηλότερη ψεύτικη αρνητική αναλογία (FN), 7,61%. Όσον αφορά τον απλό αλγόριθμο ψηφοφορίας (SVM) ΔΕΝ είναι επιτυχής.

### **Financial variables (οικονομικές μεταβλητές εισόδου)**

Οι μεταβλητές που επιλέγονται περιλαμβάνουν: κέρδος στις πωλήσεις (Profit to sales), λειτουργούν κέρδος στις πωλήσεις (Operating profit to sales), (Ordinary profit to total capital), τρέχοντα στοιχεία του παθητικού (Current liabilities to total capital), το καθαρό κόστος χρηματοδότησης (Net financing cost), το καθαρό κεφάλαιο κίνησης (Net working capital to total capital), το ποσοστό αύξησης των παρόντων προτερημάτων (Growth rate of current assets) και το συνηθισμένο εισόδημα στην καθαρή αξία (Ordinary income to net worth).

### **Εργασία 3. Using neural network ensembles for bankruptcy prediction and credit scoring**

Σε αυτό το έγγραφο γίνεται η σύγκριση της απόδοσης του single neural network classifier με τους multiple classifiers και τους diversified multiple neural network classifiers άνω των τριών συνόλων δεδομένων για τα προβλήματα πρόβλεψης πτώχευσης και πιστωτικής σημείωσης.

Για την λύση των οικονομικών προβλημάτων λήψεων αποφάσεων έχουν χρησιμοποιηθεί: - Artificial intelligence (η τεχνητή νοημοσύνη)

- Machine learning techniques

Σε αυτές τις 2 τεχνικές συμπεριλαμβάνονται:

α) τεχνητά νευρωνικά δίκτυα ANN

β) δέντρα αποφάσεων DT και

γ) διανυσματικές μηχανές υποστήριξης SVM

Η περισσότερο όμως χρησιμοποιημένη τεχνική για τα οικονομικά προβλήματα λήψης αποφάσεων είναι το multilayer perceptron (MLP) network

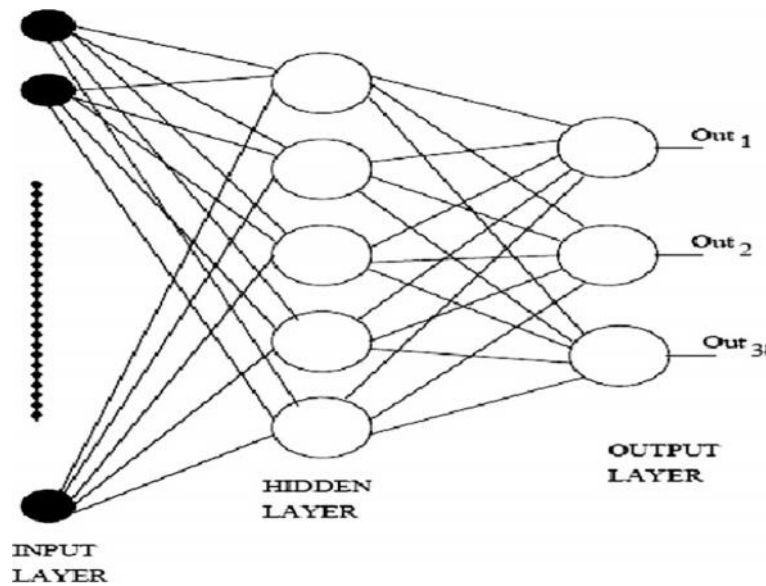
Κατά μέσον όρο οι single classifiers είναι καταλληλότεροι για την πρόβλεψη πτώχευσης και την πιστωτική σημείωση. Εντούτοις, όταν εξετάζουμε τα ποσοστά σφάλματος τύπων I και II, δεν υπάρχει κανένας ακριβής νικητής. Οι περισσότεροι από τους diversified multiple classifiers αποδίδουν χειρότερα από τους single classifiers και τους multiple classifiers. Αν και οι single classifiers εκτελούν καλύτερα στην περίπτωση της μέσης ακρίβειας πρόβλεψης, multiple classifiers ή οι diversified multiple classifiers δεν πρέπει να αγνοηθούν στην πρόβλεψη πτώχευσης και την πιστωτική σημείωση.

Τύπος I σφάλματος: Αυτό εμφανίζει το ποσοστό σφαλμάτων πρόβλεψης ενός μοντέλου, στο οποίο η ταξινόμηση δεν είναι σωστή, δηλαδή ταξινομείται η κακή πιστωτική ομάδα στην καλή πιστωτική ομάδα.

- Τύπος II σφάλματος: Αντίθετα από το σφάλμα τύπων I, αυτό παρουσιάζει το ποσοστό σφαλμάτων πρόβλεψης ενός μοντέλου στο οποίο η ταξινόμηση επίσης δεν είναι σωστή δηλαδή ταξινομείται η καλή πιστωτική ομάδα στην κακή πιστωτική ομάδα

### **Artificial neural networks ANN**

Ο πιο κοινός τύπος νευρωνικών δικτύων αποτελείται από τρία στρώματα: στρώματα εισόδου (input layers), κρυμμένα στρώματα (hidden layers) και στρώματα εξόδου (output layers) και ο οποίος καλείται multilayer perceptron MLP (πολυστρωματικό perceptron). Ένα στρώμα των μονάδων εισόδου συνδέεται με ένα στρώμα των κρυμμένων μονάδων, το οποίο συνδέεται με ένα στρώμα των μονάδων εξόδου. Η δραστηριότητα των στρωμάτων εισόδου αντιπροσωπεύει τις ακατέργαστες πληροφορίες που τροφοδοτούν το δίκτυο. Η δραστηριότητα κάθε κρυμμένης μονάδας καθορίζεται από τις δραστηριότητες των μονάδων εισόδου και τα βάρη στις συνδέσεις μεταξύ της εισόδου και των κρυμμένων μονάδων. Η συμπεριφορά των μονάδων εξόδου εξαρτάται από τη δραστηριότητα των κρυμμένων μονάδων και των βαρών μεταξύ των κρυμμένων και μονάδων εξόδου.



**Σχήμα:** *The three-layer neural network.*

### Πολλαπλάσιοι ταξινομητές (Multiple classifiers)

Ως εναλλακτική λύση μιας ενιαίας προσέγγισης ταξινομητών, εξετάζεται ο συνδυασμός πολλαπλάσιων ταξινομητών. Η βασική ιδέα στα πολλαπλάσια συστήματα ταξινομητών είναι να συνδυαστούν διάφοροι ταξινομητές έτσι ώστε το προκύπτον συνδυασμένο σύστημα να επιτυγχάνει την υψηλότερη ακρίβεια και αποδοτικότητα ταξινόμησης από τους αρχικούς ενιαίους ταξινομητές.

Στο πείραμα αυτού του άρθρου συμπεριλαμβάνονται τρία οικονομικά σύνολα, τα οποία είναι Αυστραλιανή πίστωση (Australian credit), Γερμανική πίστωση (German credit) και Ιαπωνική πίστωση (Japanese credit)

### ΣΥΜΠΕΡΑΣΜΑΤΑ

- Έγινε η σύγκριση της απόδοσης του single neural network classifier με τους diversified multiple neural network classifiers άνω των τριών συνόλων δεδομένων για τα προβλήματα πρόβλεψης πτώχευσης και πιστωτικής σημείωσης.

- ο ενιαίος νευρωνικός ταξινομητής δικτύων είναι καταλληλότερος από τους πολλαπλάσιους

- Τα αποτελέσματα δείχνουν ότι εξετάζοντας τα προβλήματα πιστωτικής σημείωσης της Αυστραλίας της Γερμανίας και της Ιαπωνίας, ο ιθύνων πρέπει να εξετάσει όχι μόνο τους single classifiers αλλά και τους multiple classifiers και τους diversified multiple classifiers διότι όταν εξετάζουμε τα ποσοστά σφάλματος τύπων I και II, οι ενιαίοι ταξινομητές δεν ξεπερνούν συνολικά τους πολλαπλάσιους ή διαφοροποιημένους πολλαπλάσιους ταξινομητές, ειδικά για τον τύπο II σφάλμα.

#### **Εργασία 4. Bankruptcy prediction with neural logic networks by means of grammar-guided genetic programming**

Στο έγγραφο αυτό γίνεται χρήση υβριδικών ευφυών συστημάτων (hybrid intelligent systems) για το πρόβλημα ταξινόμησης της πτώχευσης. Στόχος της μελέτης είναι η ικανότητα πρόβλεψης της επιχειρησιακής αποτυχίας. Προτείνεται η εφαρμογή των neural logic networks με τη βοήθεια του genetic programming.

Για την μελέτη αυτή χρησιμοποιήθηκαν στοιχεία από 118 Ελληνικές επιχειρήσεις. Το συνολικό σύνολο 118 περιπτώσεων, αποτελείται από δύο άνισα υποσύνολα, ένα που αντιπροσωπεύει το training set (σύνολο κατάρτισης) (80 περιπτώσεις) και άλλο το testing data set (εξεταστικό σύνολο στοιχείων) (38 περιπτώσεις).

Το πρώτο, training σύνολο στοιχείων αποτελείται από 80 ελληνικές εταιρίες, οι οποίες ανήκουν σε διαφορετικές βιομηχανίες. Το ολόκληρο σύνολο στοιχείων μπορεί να διαιρεθεί σε δύο ίσα υποσύνολα, συμπεριλαμβανομένων 40 εταιριών που χρεοκόπησαν κατά τη διάρκεια περιόδου 1986-1990 και 40 μη-χρεοκοπημένων εταιριών για το ίδιο χρονικό διάστημα. Η επιλογή γίνεται κατά τέτοιο τρόπο ώστε να υπάρχουν παρόμοια χαρακτηριστικά, όπως ο αριθμός υπαλλήλων και συνολικών προτερημάτων.

Οι περισσότερες από τις πτωχεύσασες εταιρίες (42%) άνηκαν στον κλωστοϋφαντουργικό τομέα και το υπόλοιπο ποσοστό ισορροπήθηκε μεταξύ των διάφορων τομέων όπως τα τρόφιμα, η ένδυση και τα υποδήματα, το ξύλο, τα πλαστικά, οι χημικές ουσίες, η βιομηχανία μετάλλων, τα οχήματα μεταφορών, κ.λπ. Ομοίως, το δεύτερο testing (εξεταστικό) δείγμα που αποτελείται από 38 περιπτώσεις χωρίζεται σε δύο ίσους υποτομείς, δηλαδή 19 χρεοκοπημένες εταιρίες και 19 μη χρεοκοπημένες το χρονικό διάστημα 1991-1993. Οι περισσότερες από τις πτωχεύσασες εταιρίες (37%) του δείγματος δοκιμής ανήκουν επίσης στον κλωστοϋφαντουργικό τομέα ενώ το υπόλοιπο ποσοστό ισορροπήθηκε πάλι μεταξύ των διάφορων τομέων όπως τα τρόφιμα, η ένδυση και τα υποδήματα, το ξύλο, τα πλαστικά, οι χημικές ουσίες, η βιομηχανία μετάλλων και τα οχήματα μεταφορών. Οι οικονομικές δηλώσεις για τις πτωχεύσασες επιχειρήσεις συλλέχθηκαν για μια περίοδο 1 έτους πριν από την αποτυχία.

Για την πρόβλεψη πτώχευσης εξετάστηκαν 12 financial ratios και οι οποίες είναι οι εξής:

1. Net income/Gross profit (Εισόδημα δικτύου/ακαθάριστο κέρδος)
2. Gross profit/Total assets
3. Net income/Total assets
4. Net income/Net worth (εισόδημα δικτύου/δίκτυο αξίας)

5. Current assets/Current liabilities (παρόντων προτερημάτων/των παρόντων στοιχείων του παθητικού)
6. Quick assets/Current liabilities (γρήγορα προτερήματα/των παρόντων στοιχείων του παθητικού)
7. Net worth/Net fixed assets (αξίας/ πάγια ενεργητικά δικτύου)
8. Inventories/Working capital (κατάλογοι/κεφάλαιο κίνησης)
9. Current liabilities/Total assets (παρόντα στοιχεία του παθητικού/συνολικά προτερήματα)
10. Working capital/Net worth (κεφάλαιο κίνησης/δίκτυο αξίας)

### **Genetic programming**

GP genetic programming είναι μια μεθοδολογία αναζήτησης που ανήκει στην οικογένεια του evolutionary computation (EC). Σήμερα αυτοί οι αλγόριθμοι έχουν εφαρμοστεί σε ένα ευρύ φάσμα των πραγματικών προβλημάτων. Το genetic programming με την κανονική μορφή του επιτρέπει την αυτόματη παραγωγή των μαθηματικών εκφράσεων. Ένα πλεονέκτημά του πέρα από τους αλγόριθμους, είναι η δυνατότητα να κατασκευάσει λειτουργικά δέντρα του ποικίλου μήκους. Το βασικότερο όμως πλεονέκτημα αυτής της διαδικασίας είναι ότι η επιλογή χαρακτηριστικών γνωρισμάτων και η διαμόρφωση συστημάτων, εμφανίζονται κατά τη διάρκεια του κανονικού τρεξίματος και δεν απαιτούν οποιαδήποτε ανθρώπινη προεπεξεργασία.

### **Εργασία 5. Early bankruptcy prediction using ENPC**

\* ENPC = - Evolutionary Nearest Neighbor Classifier

Αυτή η εργασία ασχολείται με την αξιολόγηση της απόδοσης του ENPC συγκρίνοντας την με έξι (6) εναλλακτικές λύσεις. Υπάρχουν 2 τρόποι να εξεταστεί η πρόβλεψη πτώχευσης: ο 1<sup>ος</sup> είναι δομική προσέγγιση (structural approach), περιλαμβάνει τις λεπτομερείς προβολές των οικονομικών δηλώσεων και ο 2<sup>ος</sup> είναι μια στατιστική προσέγγιση και η οποία χρησιμοποιείται σε αυτήν την εργασία.

Λόγω της δικτυακής γειτονιάς οι προσπάθειες πρόβλεψης γίνονται με πολύ διαφορετικές τεχνικές όπως:

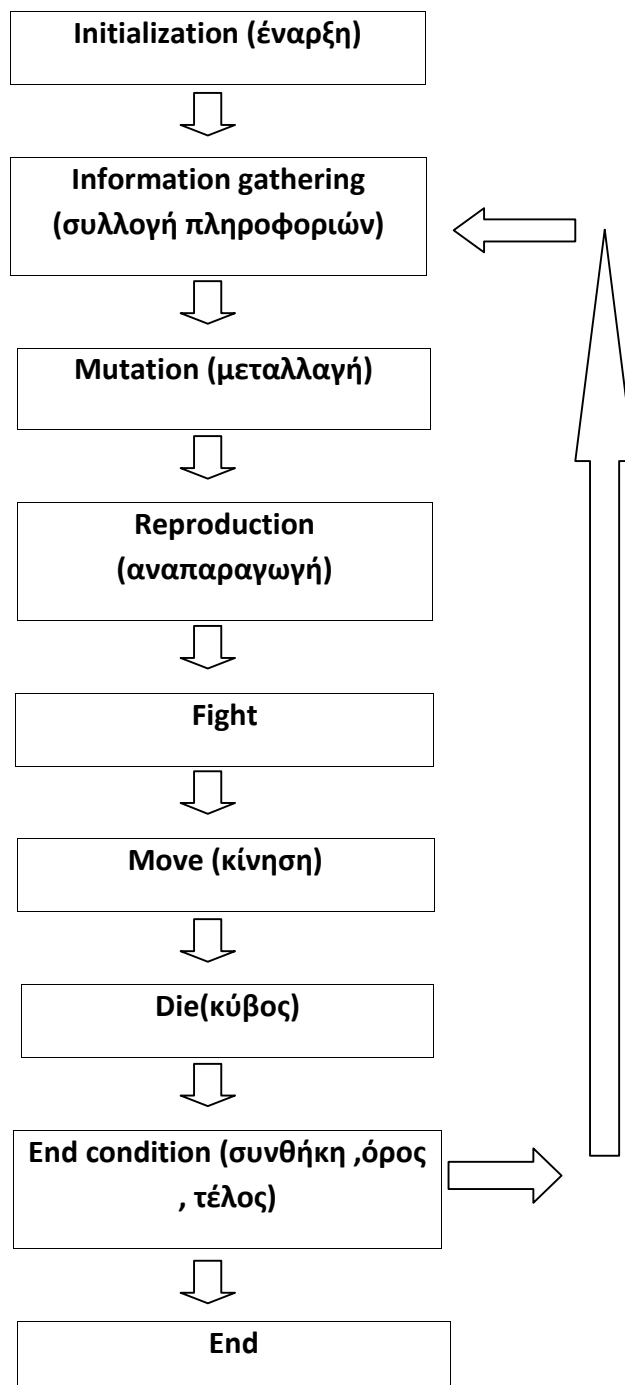
- k-nearest neighbor
- multiple discriminant analysis (η πολλαπλάσια διακρίνουσα ανάλυση MDA)
- logit models (τα μοντέλα λογικού)

- artificial neural networks ANN (τα τεχνητά νευρικά δίκτυα)
- classification trees (τα δέντρα ταξινόμησης)

ENPC : Ένα από τα βασικά πλεονεκτήματα αυτού του αλγορίθμου είναι η έλλειψη για τις αρχικές παραμέτρους που καθορίζονται από το χρήστη για να καθοδηγήσει τη διαδικασία

**Η διαδικασία κατάρτισης (training process) αποτελείται από 8 στάδια.**

### Ροή του αλγορίθμου ENPC





- Initialization(Έναρξη) : Η αρχική θέση του αλγορίθμου είναι άσχετη και ρυθμίζεται αυτόματα.
- Information gathering (συλλογή πληροφοριών): Στην αρχή κάθε επανάληψης ο αλγόριθμος συγκεντρώνει τις πληροφορίες σχετικά με τα πρωτότυπα, τις κλάσεις και τα σύνολα προτύπων που απαιτούνται για τα ακόλουθα πέντε στάδια.
- Reproduction (αναπαραγωγή): Αυτό το στάδιο εισάγει νέα πρωτότυπα στον ταξινομητή.
- Όρος τέλους. Η διαδικασία συνεχίζεται έως ότου αποφασίζει να σταματήσει ο χρήστης.

Ο αλγόριθμος ENPC έχει συγκριθεί με 6 ταξινομητές οι οποίοι φαίνονται στον επόμενο πίνακα ο οποίος αναφέρει τα συνοπτικά αποτελέσματα για το δείγμα δοκιμής.

	NB	LR	C4.5	PART	SVM	MLP	ENPC
Accuracy	73.30%	76.47%	71.95%	74.66%	74.66%	79.19%	80.09%
T. I Error	76.79%	85.71%	96.43%	99.99%	99.99%	76.79%	71.43%
T. II Error	9.70%	2.42%	4.85%	0.00%	0.00%	1.81%	2.42%

\* NB= Naive Bayes

LR= logistic regression with a ridge estimator (λογιστική οπισθοδρόμηση με έναν εκτιμητή κορυφογραμμών)

SVM= support vector machine trained with sequential minimal optimization

MLP= multilayer perceptron

Καθορίζεται ο Type I error ως λανθασμένη καταχώρηση μιας αποτυχημένης εταιρίας ως υγιούς απ' την άλλη ο Type II error θα εμφανίζεται όποτε οι μη-χρεοκοπημένες επιχειρήσεις ταξινομούνται ως πτωχεύσασες. Με βάση το πίνακα το MLP εκτελεί καλύτερα από τα υπόλοιπα. Τα C4.5, NB δεν είναι χρήσιμα αφού εμφανίζουν ένα βαθμό για την ακρίβεια κάτω των 77.46%. Τα SVM, PART έχουν τα χειρότερα αποτελέσματα. Τέλος το ENPC προσφέρει τα καλύτερα αποτελέσματα αν και δεν διαφέρει και πολύ από το MLP

### Μεταβλητές- αναλογίες (Variables)

Στην ανάλυση χρησιμοποιούνται 5 επεξηγηματικές μεταβλητές

Μεταβλητές πρόβλεψης	
NAME	Definition (καθορισμός)
WC/TA	Κεφάλαιο κίνησης /Συνολικά

	προτερήματα
RE/TA	Διατηρημένες αποδοχές/ Συνολικά προτερήματα
EBIT/TA	Αποδοχές πριν από το φόρο/ Συνολικά προτερήματα
MVE/TA	Αγοραστική αξία / Συνολικά προτερήματα
S/TA	Πωλήσεις / Συνολικά προτερήματα
CR	Παρόντα προτερήματα/ Παρόντα στοιχεία του παθητικού

Το δείγμα στην μελέτη αυτού του άρθρου αποτελείται από 552 Αμερικάνικες επιχειρήσεις, 138 εκ των οποίων χρεοκόπησαν μεταξύ των ετών 1995-2004.

Ως συμπέρασμα της μελέτης που αναλύθηκε πριν μπορεί να αναφερθεί η καινοτομία του ENPC, η οποία αναλύεται ως εξής:

- α) ευκολία στην χρήση
- β) αντίθετα με MLP βρίσκει τις απαραίτητες παραμέτρους αυτόνομα στο χρόνο εκτέλεσης
- γ) είναι άξιος ανταγωνιστικός αλγόριθμος

### **Εργασία 6. An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring.**

Είναι το πρώτο έγγραφο που συγκρίνει διάφορες μεθόδους για την πρόβλεψη πτώχευσης και την πιστωτική σημείωση, δείχνοντας έτσι ότι το σύνολο ταξινομητών μπορεί να χρησιμοποιηθεί για την ώθηση της απόδοσης του “αυτόνομου ταξινομητή”( “stand-alone”)

Το άρθρο αυτό αναφέρεται σε διάφορες οικονομικές μεθόδους λήψης αποφάσεων (decision-making methods) όπως : multilayer perceptron (MLP), Decision Tree (δέντρο απόφασης) και η Support Vector Machine SVM (διανυσματική μηχανή υποστήριξης).

Το σύνολο μεθόδων ταξινομητών που εξετάζονται σε αυτό το έγγραφο είναι:

- α) Bagging (BA)
- β) Random Subspace (RS) Κάθε αυτόνομος ταξινομητής (stand-alone classifier) χρησιμοποιεί μόνο ένα υποσύνολο όλων των χαρακτηριστικών γνωρισμάτων για την κατάρτιση και τη δοκιμή (training and testing).

- γ) Class Switching (CW) Κάθε νέο σύνολο κατάρτισης (training set) λαμβάνεται τυχαία μεταστρέφοντας τις κλάσεις του παραδείγματος κατάρτισης
- δ) Rotation Forest (RF)

Στη συνέχεια εξετάστηκαν οι ακόλουθοι ταξινομητές:

- α) Levenberg–Marquardt neural net με πέντε κρυμμένες μονάδες (LM)
- β) Multi-layer perceptron neural net με πέντε κρυμμένες μονάδες (MLP).
- γ) Radial Basis function Support Vector Machine (RV)<sup>2</sup> (Ακτινωτή διανυσματική μηχανή (RV)<sup>2</sup> υποστήριξης)
- δ) The old 5-nearest neighbour. (5nn)

Για τα πειράματα αυτού του εγγράφου επιλέγονται τρία οικονομικά σύνολα δεδομένων(financial datasets): Αυστραλιανή πίστωση (Australian credit), Γερμανική πίστωση(German credit), και Ιαπωνική πίστωση ( Japanese credit)

### Πείραμα

Συγκρίνεται ένας ενιαίος ταξινομητής MLP ( multilayer perceptron ) με τους πολλαπλάσιους ταξινομητές και τους διαφοροποιημένους πολλαπλάσιους ταξινομητές σε τρία σύνολα δεδομένων.

Στον πίνακα φαίνονται τα χαρακτηριστικά των συνόλων δεδομένων που χρησιμοποιούνται στο πείραμα.

Σύνολο στοιχείων	#A	#E	#C
Ιαπωνία	15	690	2
Αυστραλία	14	690	2
Γερμανία	20	1000	2

#A= αριθμός ιδιοτήτων

#E= αριθμός παραδειγμάτων

#C=αριθμός κλάσεων

### Παράμετροι

Οι παράμετροι που χρησιμοποιούνται για να συγκρίνουν τις δοκιμασμένες μεθόδους είναι:

- α) Accuracy (Ακρίβεια)
- β) The area under the Receiver Operating Characteristic curve (AUC)  
Η περιοχή κάτω από τη λειτουργούσα χαρακτηριστική καμπύλη δεικτών σχεδιάζει το αληθινό θετικό ποσοστό ενάντια στο ψεύτικο θετικό ποσοστό. Η AUC είναι μια από τις καλύτερες μεθόδους για τους ταξινομητές στα προβλήματα δύο-κλάσης (two-class problems).
- γ) Type I error αριθμός προτύπων που ανήκουν στην κακή πιστωτική ομάδα και ταξινομείται ανακριβώς στην καλή πιστωτική ομάδα.
- δ) Type II error αριθμός προτύπων που ανήκουν στην καλή πιστωτική ομάδα που ταξινομείται ανακριβώς στην κακή πιστωτική ομάδα.

**Συμπέρασμα:** Η Accuracy και τα Type I error και Type II error ΔΕΝ είναι αξιόπιστοι παράμετροι(αυτοί οι δείκτες ΔΕΝ εξετάζουν τα αποτελέσματα των ταξινομητών). Αντιθέτως το AUC είναι ο καλύτερος δείκτης για να συγκρίνουμε τις μεθόδους και είναι βασισμένο στην πλοκή του αληθινού θετικού ποσοστού ενάντια στο ψεύτικο θετικό ποσοστό.

**Συμπέρασμα:** Ο καλύτερος αυτόνομος ταξινομητής(stand-alone classifier) είναι MLP και το καλύτερο σύστημα είναι το RS LMNC. Η καινοτομία αυτού του εγγράφου είναι ότι κάνει σύγκριση ενός συνόλου ταξινομητών για την πρόβλεψη πτώχευσης και την πιστωτική σημείωση.

### **Εργασία 7. Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods:A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey.**

Αυτή η μελέτη στοχεύει να εφαρμόσει τις διάφορες neural network techniques NN, τις support vector machines SVM και τις multivariate statistical methods στο πρόβλημα πρόβλεψης αποτυχίας τραπεζών σε μια τουρκική περίπτωση και να παρουσιάσει μια περιεκτική σύγκριση των αποδόσεων ταξινόμησης των δοκιμασμένων τεχνικών.

#### **Μέθοδοι**

Artificial neural networks (ANN) = Τεχνητά Νευρωνικά Δίκτυα

- ⇒ Multi-layer Perceptron(MLP)
- ⇒ Competitive Learning (CL)
- ⇒ Selforganizing Map (SOM)
- ⇒ Learning Vector Quantization (LVQ)=  
διανυσματική κβαντοποίηση εκμάθησης

Support vector machines (SVM) = Διανυσματικές μηχανές υποστήριξης

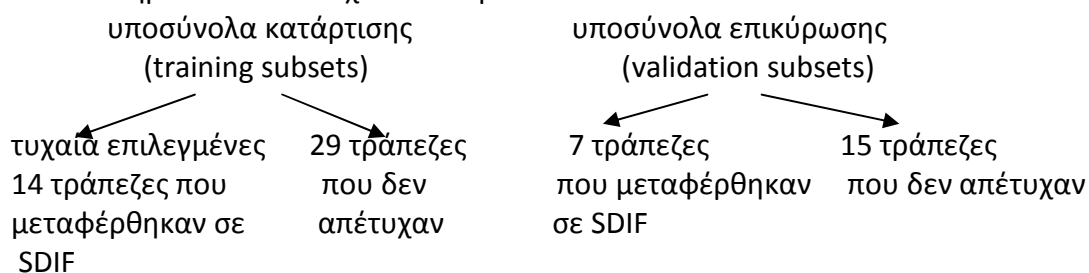
Multivariate statistical methods = Πολλών μεταβλητών στατιστικές μέθοδοι

- ✓ Multivariate Discriminant Analysis (MDA)
- ✓ K-Means Cluster Analysis (CA)
- ✓ Logistic Regression Analysis (LRA)

**Δείγμα:** Συνολικά 21 Τράπεζες μεταφέρθηκαν στο SDIF από τις οποίες οι 21 τράπεζες ήταν αποτυχημένες και οι 44 τράπεζες ήταν μη-αποτυχημένες. Ως εκ τούτου, το ολόκληρο σύνολο στοιχείων αποτελούνταν από 65 τράπεζες και αναφέρεται στην περίοδο 1997-2003. Οι μη-αποτυχημένες τράπεζες επιλέχτηκαν τυχαία με το ποσό διπλάσιο του αριθμού αποτυχημένων τραπεζών. Επιπλέον, δύο πιο μη-

αποτυχημένες τράπεζες επιλέχθηκαν και περιλήφθηκαν επίσης τυχαία στα ερευνητικά στοιχεία (δηλ:  $21 \times 2 = 42 + 2$ )

Το ολόκληρο σύνολο στοιχείων διαιρείται έπειτα σε:



Ο αριθμός τραπεζών στα ερευνητικά στοιχεία

ΕΤΟΣ	Ολόκληρο σύνολο		Σύνολο κατάρτισης Training set		Σύνολο επικύρωσης Validation set	
	Αποτυχημένος	μη-αποτυχημένος	Αποτυχημένος	μη-αποτυχημένος	Αποτυχημένος	μη-αποτυχημένος
1997	1	2	0	1	1	1
1998	1	2	1	1	0	1
1999	6	12	4	8	2	4
2000	3	6	2	4	1	2
2001	8	16	5	11	3	5
2002	1	2	1	1	0	1
2003	1	2	1	2	0	0
2004	0	2	0	1	0	1
Σύνολο	21	44	14	29	7	15

Αποτελέσματα απόδοσης για την κατάρτιση και την επικύρωση των τεχνικών πρόβλεψης

Μοντέλο πρόβλεψης	Απόδοση κατάρτισης (%) Training performance	Απόδοση επικύρωσης (%)
Multi-layer perceptron (MLP)	100.000	95.50
Competitive learning (CL)	58.14	68.18

Self-organizing map (SOM)	58.14	63.63
Learning vector quantization (LVQ)	83.72	100.00
Support vector machines (SVMs)	95.34	90.90
Multivariate discriminant analysis (MDA)	88.37	68.18
K-means cluster analysis (CA)	86.04	81.81
Logistic regression analysis (LRA)	86.04	81.81

Τα αποτελέσματα για το MLP έδειξαν ότι ολόκληρο το σύνολο στοιχείων με 20 καθαρές μεταβλητές εισόδου δίνουν τα ακριβέστερα αποτελέσματα ταξινόμησης. Το MLP ταξινόμησε σωστά 100% των τραπεζών στο σύνολο στοιχείων κατάρτισης (Training performance %), και 95,5% των τραπεζών στο σύνολο επικύρωσης (Validation performance %). Αυτά τα αποτελέσματα σημαίνουν ότι MLP ταξινόμησε ακριβώς όλα τα πρότυπα εισόδου του συνόλου στοιχείων κατάρτισης και το μεγαλύτερο μέρος του συνόλου στοιχείων επικύρωσης. Η απόδοση του MLP μπορεί να θεωρηθεί ως ικανοποιητική.

⇒ Artificial neural networks.

Σε αυτό το άρθρο για την πρόβλεψη αποτυχίας των τραπεζών χρησιμοποιήθηκαν τα εξής μοντέλα

- Multi-layer Perceptron (MLP)
- Competitive Learning (CL),
- Selforganizing Map (SOM) and
- Learning Vector Quantization (LVQ)

⇒ SVMs Support vector machines.

Είναι μια τεχνική:

- Ταξινόμησης (classification)
- Αναγνώρισης (recognition)
- Regression και
- time series technique

⇒ Multivariate statistical methods.

Για την πρόβλεψη οικονομικών αποτυχιών των τραπεζών χρησιμοποιούνται:

- Multivariate Discriminant Analysis (MDA) (πολλών μεταβλητών διακρίνουσα ανάλυση)
- K-Means Cluster Analysis (CA) και
- Logistic Regression Analysis (LRA)

#### **Οικονομικές αναλογίες-μεταβλητές:**

Capital adequacy: (CA3).

Asset quality: (AQ5).

Management: Personnel expenses/average assets (M1).

Earnings: Net profit/average assets.

Liquidity: Liquid assets/total assets

Sensitivity to market risk: Trading securities/total assets (SMR1).

#### **Εργασία 8. Financial failure prediction using efficiency as a predictor.**

Σε αυτό το έγγραφο, προτείνεται μια προσέγγιση πρόβλεψης που χρησιμοποιεί την αποδοτικότητα της εταιρίας ως προάγγελο (predictor). Στην προτεινόμενη μέθοδο, η data envelopment analysis (DEA) υιοθετείται ως εργαλείο για να αξιολογήσει την αποδοτικότητα εισόδου/εξόδου κάθε εταιρίας. Δίνεται έμφαση στο γεγονός ότι μια βασική αιτία της οικονομικής αποτυχίας είναι η φτωχή διαχείριση και ότι η αποδοτικότητα επιχειρησιακής λειτουργίας είναι μια καλή αντανάκλαση της διαχείρισης της εταιρείας.

Για να ελέγξουμε την αποτελεσματικότητα της αποδοτικότητας ως προάγγελο, χρησιμοποιούμε τα στοιχεία των εταιριών που εμφανίζονται στην λίστα της ανταλλαγής αποθεμάτων της Σαγγάης (SSE) (Shanghai stock exchange (SSE)), και συγκρίνουμε την ακρίβεια της ίδιας μεθόδου πρόβλεψης με και χωρίς τη μεταβλητή αποδοτικότητας (efficiency).

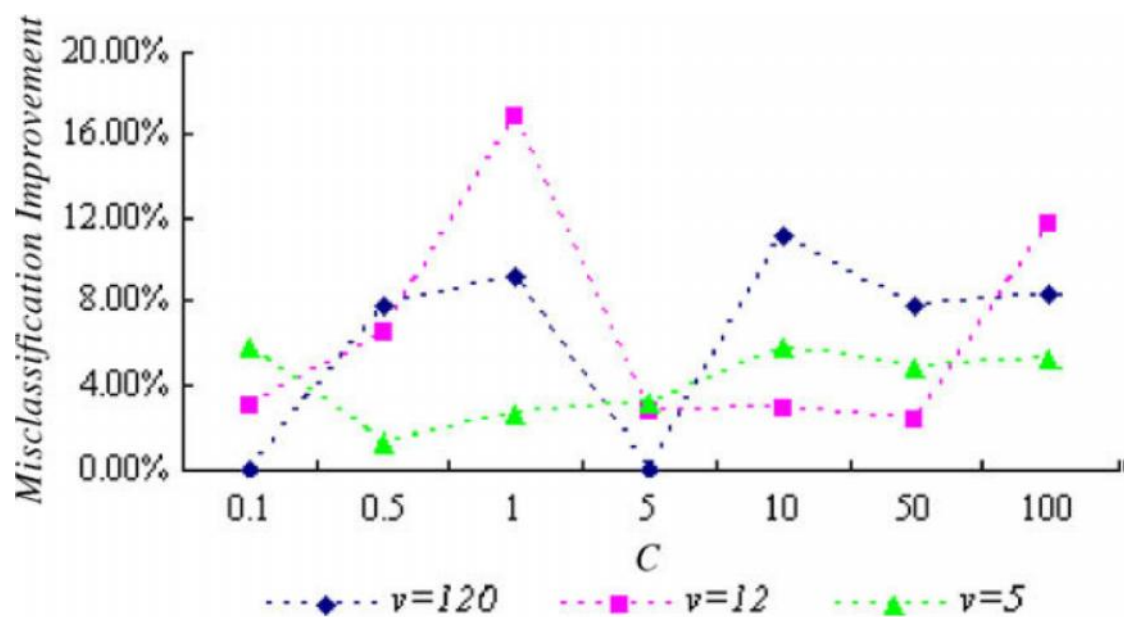
Τα μοντέλα πρόβλεψης αποτυχίας που χρησιμοποιεί η έρευνα αυτή είναι τα εξής: multiple discriminant approach (MDA), logistic regression και support vector machines (SVMs). Και τα τρία μοντέλα προτείνουν ότι η αποδοτικότητα λειτουργεί πολύ καλά ως προάγγελος.

Οι τρεις προαναφερθείσες μέθοδοι εφαρμόζονται στην οικονομική πρόβλεψη αποτυχίας των εταιριών στην Κίνα. Το σύνολο δεδομένων που υιοθετείται σε αυτήν την έρευνα παρέχεται από τις οικονομικές δηλώσεις των εταιριών που παρατίθενται στην ανταλλαγή αποθεμάτων της Σαγγάης (SSE) (Shanghai stock exchange). Το σύνολο δεδομένων περιέχει 120 ελεγχόμενες εταιρίες, από τις οποίες οι 60 περιπτώσεις έχουν τον κίνδυνο της πτώχευσης και οι 60 περιπτώσεις δεν κινδυνεύουν και αναφέρεται για την περίοδο 1999-2005. Μια επιχείρηση βρίσκεται

στην λίστα για τον οικονομικό κίνδυνο εάν εμφανίζεται ως αφερέγγυα με τα στοιχεία του παθητικού. Επιλέγονται οι πτωχεύσασες και μη-πτωχεύσασες εταιρίες τυχαία, έτσι η επιλογή καλύπτει ολόκληρο το φάσμα από στις κατασκευαστικές εταιρίες.

**Misclassification C** (εσφαλμένη κατηγοριοποίηση): Για να ερευνηθεί η αποτελεσματικότητα της προτεινόμενης προσέγγισης, πραγματοποιούνται δύο πειράματα για τη σύγκριση μέσω των διαφορετικών Cross-validation. Στο πρώτο πείραμα εφαρμόζεται η SVMs, η logistic regression και η πολλαπλάσια διακρίνουσα ανάλυση (MDA) στην οικονομική πρόβλεψη αποτυχίας που χρησιμοποιεί επτά (7) οικονομικές αναλογίες ως μεταβλητές ενώ το δεύτερο πείραμα χρησιμοποιεί τις επτά μεταβλητές και την αποδοτικότητα ως μεταβλητή. Με βάση τα αποτελέσματα που προτείνονται, η κατάλληλη σειρά για το C τίθεται μεταξύ 0,1 και 100. Ποσοστό της βελτίωσης λάθος ταξινόμησης (misclassification) του SVMs-E ανω των SVMs με διάφορο το C είναι διευκρινισμένο στο Σχ 1. Η βελτίωση λάθος ταξινόμησης (misclassification) του SVMs-ε πέρα από SVMs ορίζεται ως:  

$$\frac{(\text{misclassification of SVMs} - \text{misclassification of SVMs-E})}{(\text{misclassification of SVMs})} \cdot 100\%$$



Σχήμα. 1. Misclassification improvement of SVMs-E over SVMs. (Βελτίωση λάθος ταξινόμησης του SVMs-E ανω των SVMs.)

\* Parameter v represents the number of subsets in v-fold cross-validation.

\*SVMs-E= SVMs–efficiency=αποδοτικότητα

Μπορεί να παρατηρηθεί από το σχέδιο ότι η απόδοση SVMs που χρησιμοποιεί την αποδοτικότητα (E) είναι με συνέπεια καλύτερη από το SVMs χωρίς αποδοτικότητα. Κάνουμε το ίδιο πείραμα αλλά αντί να χρησιμοποιήσουμε το SVMs χρησιμοποιούμε τη logistic regression και το MDA στο σύνολο δεδομένων χρησιμοποιώντας επτά και οκτώ μεταβλητές, αντίστοιχα. Το συμπέρασμα είναι ότι οι γενικές αποδόσεις πρόβλεψης του SVMs, της logistic regression και της MDA που χρησιμοποιούν την



αποδοτικότητα ως προάγγελο είναι καλύτερες από εκείνες χωρίς αποδοτικότητα. Επομένως, η χρησιμοποίηση της αποδοτικότητας που λήφθηκε μέσω DEA ως προάγγελος στην εταιρική οικονομική πρόβλεψη αποτυχίας θα μπορούσε να βελτιώσει την ακρίβεια πρόβλεψης.

Σε αυτό το έγγραφο, ως μεταβλητές εισόδου για DEA επιλέχθηκαν οι συνολικές ευθύνες και οι δαπάνες των πωλήσεων για DEA, και ως μεταβλητή εξόδου επιλέχθηκε το εισόδημα των πωλήσεων.

### **Εργασία 9. Differential evolution trained wavelet neural networks: Application to bankruptcy prediction in banks.**

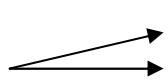
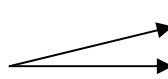
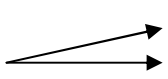
Σε αυτήν την μελέτη, αναπτύσσεται το DEWNN (differential evolution trained wavelet neural network) και συγκρίνεται με TAWNN threshold accepting trained wavelet neural network και με το αρχικό WNN a wavelet neural network όσον αφορά τα σύνολα δεδομένων συγκριτικής μέτρησης επιδόσεων (benchmark datasets), δηλαδή σύνολο δεδομένων IRIS, σύνολο δεδομένων κρασιού και σύνολο δεδομένων καρκίνου του μαστού του Wisconsin καθώς επίσης και σύνολα στοιχείων πτώχευσης δηλαδή σύνολο δεδομένων αμερικανικών τραπεζών, σύνολο δεδομένων τουρκικών τραπεζών και σύνολο δεδομένων ισπανικών τραπεζών.

\*\*DEWNN = διαφορικό εκπαιδευμένο νευρωνικό δίκτυο κυματοακτιών.

TAWNN = το εκπαιδευμένο νευρωνικό δίκτυο κυματοακτιών.

WNN = νευρωνικό δίκτυο κυματοακτιών.

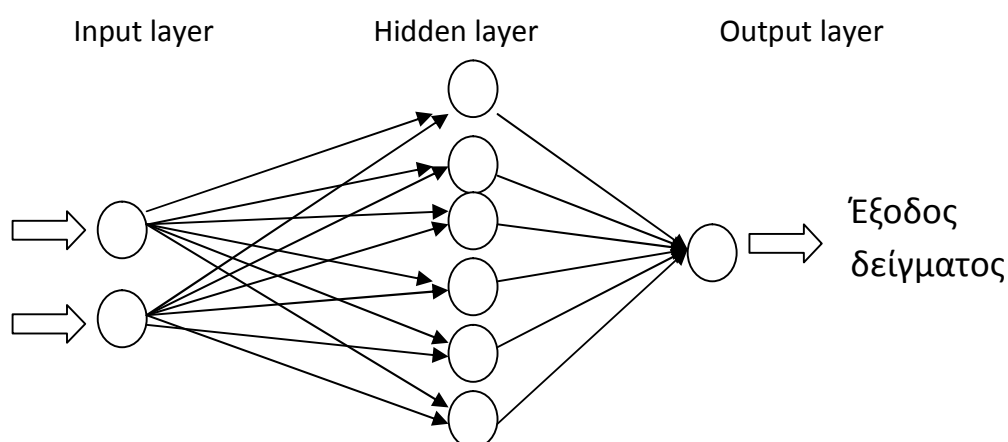
Όσον αφορά τα σύνολα στοιχείων πτώχευσης έχουμε τα εξής:

- Το σύνολο δεδομένων που αφορά την Τουρκία περιέχει:  
40 τράπεζες  22 που έχουν χρεοκοπήσει  
18 που είναι υγιείς
- Το σύνολο δεδομένων που αφορά την Ισπανία περιέχει:  
66 τράπεζες  37 που έχουν χρεοκοπήσει  
29 που είναι υγιείς (περίοδος 1982)
- Το σύνολο δεδομένων που αφορά την Αμερική περιέχει:  
129 τράπεζες  65 που έχουν χρεοκοπήσει  
64 που είναι υγιείς (περίοδος 1975-1982)

\*Στο πείραμα γίνεται η χρήση της 10-fold cross validation method

Differential evolution algorithm (DE) (διαφορικός αλγόριθμος εξέλιξης) προτείνεται για να εκπαιδεύσει ένα wavelet neural network (WNN) (νευρωνικό δίκτυο κυματακιών). Το προκύπτον δίκτυο ονομάζεται ως differential evolution trained wavelet neural network (DEWNN) (διαφορικό εκπαιδευμένο νευρωνικό δίκτυο κυματακιών). Η απόδοση DEWNN συγκρίνεται με αυτήν της ευαισθησίας που αποδέχεται το trained wavelet neural network (TAWNN) (εκπαιδευμένο νευρωνικό δίκτυο κυματακιών).

Το WNN (a wavelet neural network) αποτελείται από τρία στρώματα: στρώμα εισόδου, κρυμμένο στρώμα και στρώμα εξόδου (input layer, hidden layer and output layer). Η δομή του WNN στην συγκεκριμένη έρευνα αποτελείται από δύο εισαγμένους και έξι κρυμμένους κόμβους



## Financial ratios of the datasets and the selected features

### Στοιχεία των τουρκικών τραπεζών

- 1 Δαπάνες ενδιαφέροντος/μέσος όρος κερδοφόρων προτερημάτων (Interest expenses/average profitable assets)
- 2 Δαπάνες ενδιαφέροντος/μέσος όρος μη-κερδοφόρων προτερημάτων (Interest expenses/average non-profitable assets)
- 3 (Share holders' equity + total income)/(deposits + non-deposit funds)
- 4 (Interest income/interest expenses)
- 5 (Share holders' equity + total income)/total assets
- 6 (Share holders' equity + total income)/(total assets + contingencies & commitments)
- 7 Κεφάλαιο δικτύωσης /συνολικά προτερήματα (Networking capital / total assets)

### Στοιχεία των ισπανικών τραπεζών

- 1 Ενεργητικό/συνολικά προτερήματα (Current assets/total assets)
- 2 (Current assets-cash/total assets)

- 3 Ενεργητικό/δάνεια (Current assets/loans)
- 4 Επιφυλάξεις/δάνεια (Reserves/loans)
- 5 Καθαρό εισόδημα/συνολικά προτερήματα (Net income/total assets)
- 6 Καθαρό εισόδημα/συνολικό κεφάλαιο δικαιοσύνης (Net income/total equity capital)
- 7 Καθαρό εισόδημα/δάνεια (Net income/loans)
- 8 Κόστος των πωλήσεων/πωλήσεις (Cost of sales/sales)
- 9 Ταμειακή ροή/δάνεια (Cash flow/loans)

### Στοιχεία των αμερικανικών τραπεζών

- 1 Κεφάλαιο κίνησης/συνολικά προτερήματα (Working capital/total assets)
- 2 Διατηρημένες αποδοχές/συνολικά προτερήματα (Retained earnings/total assets)
- 3 Αποδοχές πριν από τους φόρους/τα συνολικά προτερήματα (Earnings before interest and taxes/total assets)
- 4 (Market value of equity/total assets)
- 5 Πωλήσεις/συνολικά προτερήματα (Sales/total assets)

Συμπέρασμα: -Τα αποτελέσματα δείχνουν πως το DEWNN ξεπερνά τους άλλους αλγόριθμους.

- Το DEWNN είναι ένας αποτελεσματικός αλγόριθμος για τα προβλήματα ταξινόμησης που εμφανίζονται στην χρηματοδότηση.

- Η παρούσα έρευνα καταλήγει στο συμπέρασμα ότι η κατάρτιση WNN με την διαφορική εξέλιξη (DE) λύνει τα προβλήματα ταξινόμησης με την αυξανόμενη ακρίβεια.

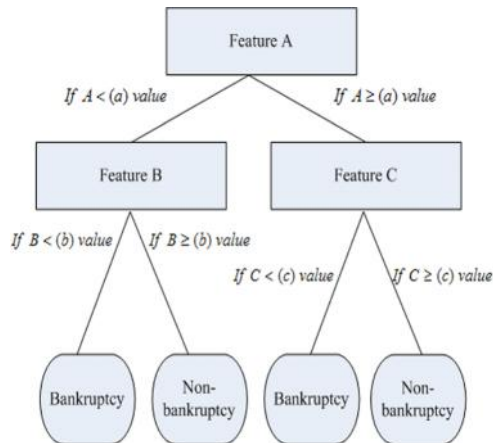
### **Εργασία 10. A selective ensemble based on expected probabilities for bankruptcy prediction.**

Σε αυτό το έγγραφο, προτείνεται ένα σύνολο τριών ταξινομητών, the Decision Tree, the Back propagation neural network, **BPN** και the support vector machine, **SVM**. Με βάση τις αναμενόμενες πιθανότητες της πτώχευσης και της μη-πτώχευσης, αυτό το σύνολο παρέχει μια προσέγγιση που κληρονομεί τα πλεονεκτήματα και αποφεύγει τα μειονεκτήματα των διαφορετικών τεχνικών ταξινόμησης.

#### **Μέθοδοι:**

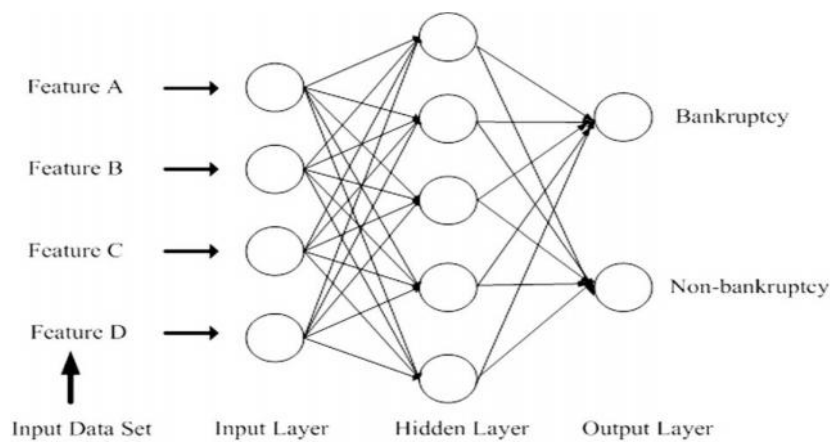
**α) Decision trees** Ένα δέντρο απόφασης καταχωρεί μερικούς κανόνες ταξινόμησης στους κόμβους κλάδων προκειμένου να ομαδοποιηθούν τα παρόμοια δείγματα στους ίδιους κόμβους φύλλων. Επομένως, ένα δέντρο απόφασης είναι μια τεχνική ταξινόμηση ή πρόβλεψη. Υπάρχουν πολλοί αλγόριθμοι των δέντρων αποφάσεων. Η βασική διαφορά τους είναι ο τρόπος να αποφασιστεί η ακολουθία ιδιοτήτων που πρέπει να χρησιμοποιηθεί για κάθε κόμβο κλάδων. ID3 (Quinlan, 1986) είναι ένας διάσημος αλγόριθμος δέντρων απόφασης, ο οποίος χρησιμοποιεί το κέρδος πληροφοριών για την επιλογή της ακολουθίας ιδιοτήτων. C4.5 (Quinlan,

1993) είναι μια εκτεταμένη έκδοση ID3, η οποία χρησιμοποιεί αναλογία κέρδους αντί του κέρδους πληροφοριών για να αποφύγει αυτό το πρόβλημα.



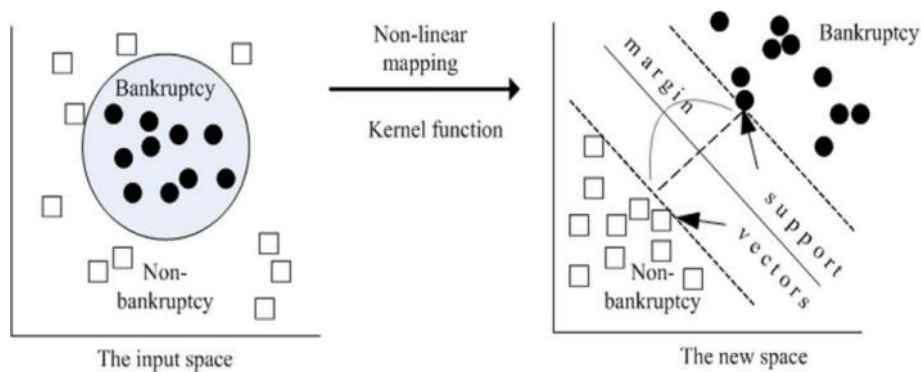
An example of decision tree.

**β) Artificial neural networks(ANN).** ANNs έχουν χρησιμοποιηθεί ευρέως για πολλές στοιχειώδεις εργασίες ταξινόμησης και πρόβλεψης όπως η αξιολόγηση επιχειρησιακού κινδύνου, η πρόβλεψη πτώχευσης, η πρόβλεψη τιμών αποθεμάτων κ.λπ...)



An example of ANN.

**γ) Support vector machine (SVM).** Χρησιμοποιείται κυρίως για τις δυαδικές στοιχειώδεις εργασίες ταξινόμησης. Το SVM έχει γίνει ένα από τα δημοφιλέστερα μοντέλα ταξινόμησης λόγω της καλής εκτέλεσής του στα μεγάλα διαστήματα χαρακτηριστικών γνωρισμάτων.



An example of SVM.

### Η έννοια της αναμενόμενης πιθανότητας.

Υπάρχουν δύο κατευθύνσεις πρόβλεψης για κάθε ταξινομητή πτώχευσης, δηλ. ένας για την πτώχευση και άλλη για την μη-πτώχευση. Η ακρίβεια (CA) ταξινόμησης χρησιμοποιείται συνήθως ως κριτήριο απόδοσης.

Περιγράψουμε το CA (ακρίβεια ταξινόμησης) ως εξής:

**A** δείχνουμε τον αριθμό μη-χρεοκοπημένων επιχειρήσεων που προβλέπονται επίσης ως μη-χρεοκοπημένες επιχειρήσεις, **B** δείχνει τον αριθμό μη-χρεοκοπημένων επιχειρήσεων που προβλέπονται ως χρεοκοπημένες επιχειρήσεις, **C** δείχνει τον αριθμό χρεοκοπημένων επιχειρήσεων που προβλέπονται ως μη-χρεοκοπημένες επιχειρήσεις. **D** δείχνει τον αριθμό χρεοκοπημένων επιχειρήσεων που προβλέπονται επίσης ως χρεοκοπημένες επιχειρήσεις. Καθορίζουμε την ακρίβεια ταξινόμησης ενός ταξινομητή, CA model, ως (1), της ακρίβειας ταξινόμησης για τις χρεοκοπημένες επιχειρήσεις, CAbankruptcy, ως (2) και ακρίβεια ταξινόμησης για τις μη-χρεοκοπημένες επιχειρήσεις, CA μη-πτώχευση, ως (3)

$$(1) \text{ CA model} = \frac{A + D}{A+B+C+D} \quad (1)$$

$$(2) \text{ CAbankruptcy} = \frac{D}{C+D} \quad (2)$$

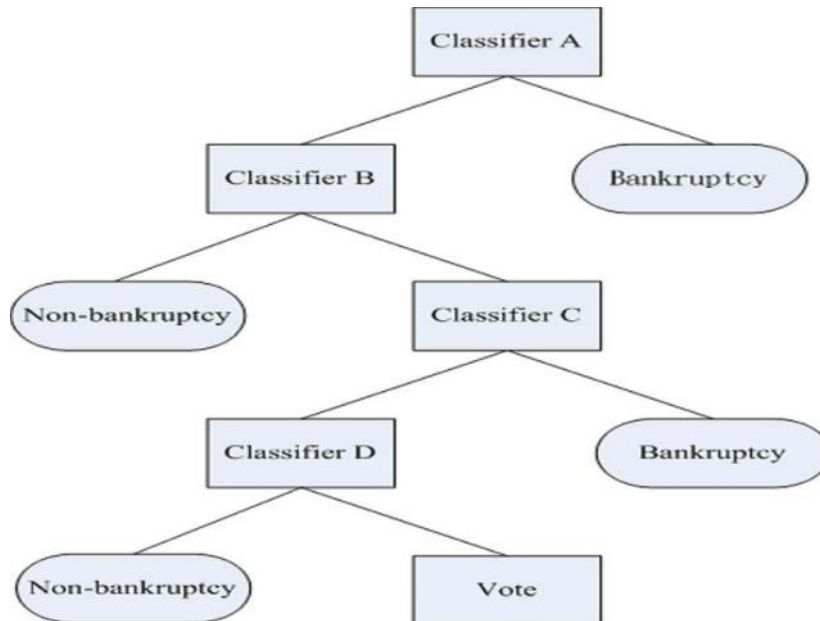
$$(3) \text{ CA μη-πτώχευση} = \frac{A}{A+B} \quad (3)$$

$$P_{\text{bankruptcy}} = \frac{D}{B+A}$$

$$P_{\text{non-bankruptcy}} = \frac{A}{A+C}$$

### Το προτεινόμενο εκλεκτικό σύνολο. (The proposed selective ensemble)

Αυτό το έγγραφο ενσωματώνει την έννοια του εκλεκτικού συνόλου με την αναμενόμενη πιθανότητα προκειμένου να κατακτηθεί το παραδοσιακό δίλημμα για το συνδυασμό των διαφορετικών ταξινομητών. Γίνεται τακτοποίηση της ακολουθίας ταξινομητών σε ένα δέντρο απόφασης βασισμένο στην αναμενόμενη πιθανότητα κάθε ταξινομητή. Ο ταξινομητής A έχει τη μέγιστη αναμενόμενη πιθανότητα για την μη-πτώχευση και αυτός ο ταξινομητής είναι ο κορυφαίος κόμβος του δέντρου συνόλων.



An example of selective ensemble tree

#### Πειραματικό σχέδιο.

#### Περιγραφή του συνόλου στοιχείων

Σε αυτό το έγγραφο χρησιμοποιείται το σύνολο στοιχείων Wieslaw (Wieslaw, 2004) Περιέχει 30 οικονομικές αναλογίες από 56 πτωχεύσασες επιχειρήσεις και 64 μη-πτωχεύσασες επιχειρήσεις μεταξύ 1997 και 2001. Κάθε επιχείρηση παρέχει δύο συνεχή ετήσια στοιχεία έτσι υπάρχουν 240 δείγματα στο σύνολο.

#### Wieslaw dataset

X01 – cash/current liabilities

X02 – cash/total assets

X03 – current assets/current liabilities

X04 – current assets/total assets

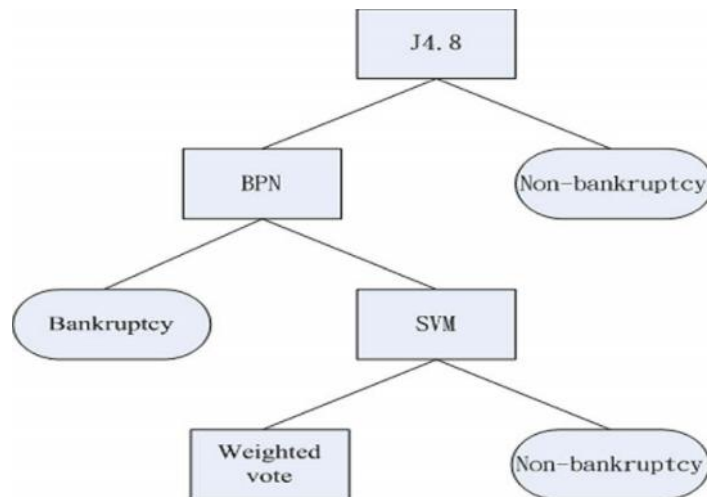
X05 – working capital/total assets

X06 – working capital/sales  
 X07 – sales/inventory  
 X08 – sales/receivables  
 X09 – net profit/total assets  
 X10 – net profit/current assets  
 X11 – net profit/sales  
 X12 – gross profit/sales  
 X13 – net profit/liabilities  
 X14 – net profit/equity  
 X15 – net profit/(equity + long term liabilities)  
 X16 – sales/receivables  
 X17 – sales/total assets  
 X18 – sales/current assets  
 X19 – (365 \* receivables)/sales  
 X20 – sales/total assets  
 X21 – liabilities/total income  
 X22 – current liabilities/total income  
 X23 – receivables/liabilities  
 X24 – net profit/sales  
 X25 – liabilities/total assets  
 X26 – liabilities/equity  
 X27 – long term liabilities/equity  
 X28 – current liabilities/equity  
 X29 – EBIT (earnings before interests and taxes)/total assets  
 X30 – current assets/sales

Υπάρχουν πολλοί αλγόριθμοι δέντρων απόφασης. Χρησιμοποιούμε το J4.8, το οποίο είναι μια τροποποίηση της C4.5. Για την τεχνική του SVM, χρησιμοποιούμε το διαδοχικό ελάχιστο αλγόριθμο βελτιστοποίησης (SMO= sequential minimal optimization). Για ANN, χρησιμοποιούμε το BPN= back propagation neural networks και γίνεται χρήση του 10-fold cross-validation.

Σύμφωνα με τον αλγόριθμό μας, υπολογίζουμε αρχικά όλες τις αναμενόμενες πιθανότητες για J4.8, SVM και BPN όπως φαίνεται παρακάτω. Η μέγιστη αναμενόμενη πιθανότητα είναι P non-πτώχευση για J4.8 και έτσι είναι ο πρώτος κόμβος κλάδων του δέντρου συνόλων.

	J4.8 (%)	BPN (%)	SVM (%)
Pbankruptcy	64.29	70.21	70.11
Pnon-bankruptcy	76.14	74.53	71.68



The selective ensemble tree using the Wieslaw data set.

#### 4. ΑΞΙΟΛΟΓΗΣΗ - ΣΥΓΚΡΙΣΗ

Η εταιρική οικονομική πρόβλεψη αποτυχίας είναι κρίσιμης σπουδαιότητας τόσο για τους διευθυντές όσο και για τα άτομα που σχετίζονται ακόμα και έμμεσα με την συγκεκριμένη τράπεζα ή επιχείρηση. Στη λογοτεχνία, πολλές έρευνες έχουν εστιάσει στις οικονομικές αναλογίες των εταιριών για την πρόβλεψη αποτυχίας. Είναι αξιοσημείωτο ότι από κάθε εργασία προκύπτει και μια μέθοδος που θεωρείται η καλύτερη για την πρόβλεψη πτώχευσης της επιχείρησης ή της τράπεζας. Παρακάτω γίνεται σύγκριση και αξιολόγηση όλων των παραπάνω άρθρων.

##### Τι επιχειρήσεις αφορούν;

PAPERS	Τράπεζες	Επιχειρήσεις ή Εταιρείες	Πίστωση μιας χώρας
3	√		
5		√	
2			√

Από την ανάλυση των παραπάνω papers προκύπτουν χρήσιμες πληροφορίες οι οποίες αφορούν την χρονική περίοδο στην οποία αναφέρεται το κάθε paper καθώς και για την χώρα από την οποία προέρχονται τα δεδομένα που χρησιμοποιήθηκαν στις έρευνες.

Συγκεκριμένα, όσον αφορά τις Τράπεζες: Ένα paper εξετάζει την πορεία των Ισπανικών και των Τούρκικων Τραπεζών. Ένα άλλο paper ασχολείται με το πρόβλημα πρόβλεψης αποτυχίας Τραπεζών σε μια Τούρκικη περίπτωση το χρονικό



διάστημα 1995-2004. Τέλος, ένα άλλο paper ασχολείται με τα σύνολα στοιχείων πτώχευσης Αμερικάνικων, Τούρκικων καθώς και Ισπανικών Τραπεζών. Όσον αφορά τα papers που ασχολούνται με τα προβλήματα πτώχευσης των τραπεζών μπορούμε να βγάλουμε το συμπέρασμα ότι και στις 3 περιπτώσεις εκτός των άλλων γίνεται αναφορά στις Τράπεζες της Τουρκίας.

Αναλύοντας τα papers που ασχολήθηκαν με επιχειρήσεις και εταιρείες έχουμε τα εξής: Ένα paper αναλύει ένα πείραμα για το οποίο χρησιμοποιήθηκαν τα ετήσια οικονομικά στοιχεία που συλλέχθηκαν από το Ταμείο πιστωτικών Εγγυήσεων της Κορέας (Korea Credit Guarantee Fund). Τα στοιχεία αποτελούνται από 900 χρεοκοπημένες εταιρείες και 900 μη-χρεοκοπημένες την περίοδο 1995-2005. Ένα άλλο paper μελετά την ικανότητα πρόβλεψης της επιχειρησιακής αποτυχίας παίρνοντας στοιχεία από 118 Ελληνικές επιχειρήσεις την περίοδο 1999-2000. Το συνολικό σύνολο 118 περιπτώσεων, αποτελείται από 2 άνισα υποσύνολα, ένα που αντιπροσωπεύει το σύνολο κατάρτισης (training set) (80 περιπτώσεις) και άλλο το εξεταστικό σύνολο (testing data set) (38 περιπτώσεις). Τέλος έχουμε ένα paper που αναλύει μια περίπτωση η οποία αναφέρεται σε 56 πτωχεύσασες επιχειρήσεις και 64 μη πτωχεύσασες επιχειρήσεις μεταξύ 1997-2001. Η ιδιαιτερότητα της μελέτης που αναφέρει το άρθρο αυτό είναι ότι κάθε επιχείρηση παρέχει 2 συνεχή ετήσια στοιχεία, έτσι υπάρχουν 240 δείγματα στο σύνολο.

Ένα πολύ σημαντικό συμπέρασμα όσον αφορά τις επιχειρήσεις είναι ότι μόνο σε μια περίπτωση τα σύνολα των παρατηρήσεων έχουν ίσο αριθμό χρεοκοπημένων και μη χρεοκοπημένων εταιρειών. Στις άλλες δυο περιπτώσεις τα σύνολα είναι άνισα.

Τέλος στα papers που γίνεται αναφορά στην πιστωτική κατάσταση μιας χώρας έχουμε τα εξής: και τα 2 papers ασχολούνται με την Αυστραλιανή, την Γερμανική και την Ιαπωνική πίστωση.

### **Μοντέλα, μέθοδοι που χρησιμοποιούνται.**

MODELS	PAPERS
Neural Networks	5
Decision Trees	3
Hybrid	1
Support Vector Machines (SVM)	5
MDA (πολλών μεταβλητών διανυσματικές μέθοδοι)	4
Logistic Regression	2
MLP(Multilayer perceptron)	2
PCNN (Principal component neural network)	1
DE (Differential evolution algorithm)	

WNN (Wavelet neural network) DEWNN (Differential evolution trained wavelet neural network)	1
--	---

Αξίζει να σημειωθεί ότι τα δημοφιλέστερα μοντέλα είναι τα Neural Networks και τα Decision Trees. Αυτό όμως δεν σημαίνει ότι τα υπόλοιπα είναι λιγότερο καλύτερα ως προς το αποτέλεσμα καλύτερης επίτευξης πρόβλεψης πτώχευσης.

#### Ποιο είναι το ερευνητικό ζητούμενο;

#### PAPERS

Σύγκριση διαφόρων μεθόδων για την πρόβλεψη πτώχευσης.	4
Προτείνεται σύνολο ταξινομητών ή γίνεται συνδυασμός ταξινομητών	4
Προτείνεται μια μέθοδος για την πρόβλεψη πτώχευσης εταιρίας ή τράπεζας.	2

Ο παραπάνω πίνακας δείχνει ότι για να βρεθεί ο καλύτερος τρόπος πρόβλεψης πτώχευσης οι περισσότερες έρευνες είτε κάνουν σύγκριση μιας συγκεκριμένης μεθόδου με πολλές είτε γίνεται συνδυασμός πολλών ταξινομητών.

Τέλος μπορούμε να αναφερθούμε στην 10 fold cross validation η οποία χρησιμοποιείται σε πολλές μελέτες. Στην μέθοδο επικύρωσης 10 τμημάτων (10 fold cross validation) το δείγμα διαιρείται σε 10 υποσύνολα. Κάθε υποσύνολο περιέχει διαφορετικές παρατηρήσεις. Η επιλογή των υποσυνόλων είναι τυχαία. Ένα από τα υποσύνολα χρησιμοποιείται ως σύνολο επικύρωσης και τα υποσύνολα 9 συνενώνονται και δημιουργούν το δείγμα εκπαίδευσης. Το μοντέλο εκπαιδεύεται χρησιμοποιώντας το δείγμα εκπαίδευσης και δοκιμάζεται έναντι του δείγματος επικύρωσης. Η διαδικασία επαναλαμβάνεται 10 φορές, κάθε φορά χρησιμοποιώντας ένα διαφορετικό σύνολο ως σύνολο επικύρωσης και τα υπόλοιπα εννέα ως σύνολο εκπαίδευσης. Στο τέλος υπολογίζεται η μέση επίδοση του μοντέλου. Η μέθοδος μπορεί να διαφοροποιηθεί ως προς το πλήθος των τμημάτων.

## 5. ΣΥΜΠΕΡΑΣΜΑΤΑ

Ο κύριος σκοπός της παρούσας εργασίας είναι να παρουσιάσει μεθόδους εξόρυξης δεδομένων για την καλύτερη πρόβλεψη πτώχευσης τραπεζών ή εταιριών.

Ένα σημαντικό συμπέρασμα που μπορεί να προέλθει από την ανάλυση των πιο πάνω εργασιών είναι ότι οι ερευνητές υιοθέτησαν σχεδόν όλες τις ευφυείς τεχνικές για να λύσουν αυτό το σημαντικό πρόβλημα. Τα Νευρωνικά δίκτυα και τα Δέντρα αποφάσεων είναι οι δύο μέθοδοι που χρησιμοποιούνται πιο συχνά από κάθε άλλη μέθοδο. Είναι σημαντικό να σημειωθεί ότι περίπου 80% των πιο πάνω εργασιών χρησιμοποιούν την 10 fold cross validation. Αν και η κάθε εργασία αντλεί τα δεδομένα από διαφορετικές πηγές, διαφορετικές χώρες και περιόδους, όλες στο τέλος βρίσκουν μία μέθοδο ή έναν συνδυασμό μεθόδων ώστε να γίνει όσο το δυνατόν καλύτερη πρόβλεψη πτώχευσης.

## ΑΝΑΦΟΡΕΣ

1. V. Ravi C. Pramodh (2008), "Threshold accepting trained principal component neural network and feature subset selection: Application to bankruptcy prediction in banks", Applied Soft Computing Vol. 8, pp. 1539–1548.
2. Sungbin Cho, Jinhwa Kim, Jae Kwon Bae (2009), "An integrative model with subject weight based on neural network learning for bankruptcy prediction", Expert Systems with Applications Vol.36, pp. 403–410.
3. Chih-Fong Tsai \*, Jhen-Wei Wu (2008), Using neural network ensembles for bankruptcy prediction and credit scoring, Expert Systems with Applications Vol.34, pp 2639–2649.
4. Athanasios Tsakonas, George Dounias, Michael Doumpos, Constantin Zopounidis (2006), "Bankruptcy prediction with neural logic networks by means of grammar-guided genetic programming", Expert Systems with Applications Vol. 30, pp. 449–461.
5. David Quintana, Yago Saez, Asuncion Mochon, Pedro Isasi (2008), "Early bankruptcy prediction using ENPC" Appl Intell Vol.29, pp. 157–161.
6. Loris Nanni , Alessandra Lumini (2009), "An experimental comparison of ensemble of classifiers for bankruptcy prediction and credit scoring" Expert Systems with Applications Vol.36, pp. 3028–3033
7. Melek Acar Boyacioglu , Yakup Kara, Omer Kaan Baykan (2009), Predicting bank financial failures using neural networks, support vector machines and multivariate statistical methods:

A comparative analysis in the sample of savings deposit insurance fund (SDIF) transferred banks in Turkey, *Expert Systems with Applications* Vol.36, pp. 3355–3366.

8. Xiaoyan Xu, Yu Wang (2009), “Financial failure prediction using efficiency as a predictor”, *Expert Systems with Applications* Vol.36, pp. 366–373.

9. Nikunj Chauhan, V. Ravi, D. Karthik Chandra (2009), “Differential evolution trained wavelet neural networks: Application to bankruptcy prediction in banks” *Expert Systems with Applications* Vol. 36, pp. 7659–7665.

10. Chihli Hung, Jing-Hong Chen (2009), “A selective ensemble based on expected probabilities for bankruptcy prediction”, *Expert Systems with Applications* 36, pp. 5297–5303

