Prediction of Banks Financial Distress

Mohammed A. SirElkhatim[1] and Naomie Salim[2]

[1]Dept.Computer Science, Sudan University of Science and Technology (SUST)
[2]Faculty of Computing, Universiti Teknologi Malaysia, Malaysia
elnasrabi@gmail.com, anomie@utm.my

Received:27.03.2013
Accepted:12.05.2013

ABSTRACT–In this research we conduct a comprehensive review on the existing literature of prediction techniques that have been used to assist on prediction of the bank distress.
We categorized the review results on the groups depending on the prediction techniques method, our categorization started by firstly using time factors of the founded literature, so we mark the literature founded in the period (1990-2010) as history of prediction techniques, and after this period until 2013 as recent prediction techniques and then presented the strengths and weaknesses of both. We came out by the fact that there was no specific type fit with all bank distress issue although we found that intelligent hybrid techniques considered the most candidates methods in term of accuracy and reputation.

Keywords: *Bank Distress, Banks Factors, Prediction techniques, Text Mining, Data Mining.*

المستخلص – *تتناول هذه الورقة أهم ما سبق نشره عن تقنيات التنبؤ والتي تساعد بصورة مباشرة وأساسية في عمليات التنبؤ بمشاكل القطاع المصرفي، تم تصنيف مخرجات البحث لعدة مجموعات علي حسب طريقة التنبؤ المتبعة إبتداءا من التاريخ الذي نشرت به البحوث الخاصة بالطريقة المعينة وبهذا تم تقسيم العمل لفترتين، الفترة الأولي تاريخية تشمل الأعوام (1990-2010) ومن ثم تم عرض اخر ما نشر حتي العام 2013 مع التركيز علي مميزات ومساوئ كليهما.من أهم النتائج عدم التوصل لطريقة تنبؤ مثالية تفي بجميع الأغراض إلا أنه قد خلصنا إلي ان طرق التنبؤ الهجينة – الذكية تعتبر الأكثر إستخداما وموثوقية من قبل الباحثين نسبة لدقة النتائج التي تنتجها.*

INTRODUCTION

The study of bank distress is an important issue. First, it enhances regulators' ability to predict potential crisis, and enables them to manage, coordinate and supervise more efficiently. Second, the early distinction between troubled and sound banks allows for appropriate actions to prevent failure and to protect healthy institutions. Third, the direct fiscal cost of recapitalizing and restructuring a troubled sector is high, and may amount to as large as half of the country's GDP[1]. Fourth, the crisis in the financial sector may create other crisis, such as currency crisis, which may further weaken the economy, and aggravates the cost of distress. Finally, bank distress is accompanied with a credit crunch that leads to underutilization and misallocation of funds, which may further hamper growth in the economy.

Several studies have aimed to identify the main factors that cause banks to fail; we will summarize the main factors that agreed to have significant effect on the banks-firm-performance. As part of their audit process, regulators use a six-part rating system to indicate the safety and soundness of the institution. This rating, referred to as the CAMELS rating, evaluates banks in their basic functional areas: Capital adequacy, Asset quality, Management expertise, Earnings' strength, Liquidity, and Sensitivity to market risk. While CAMELS ratings clearly provide regulators with important information about the health of a bank, many researchers found that these CAMELS ratings decay rapidly. Moreover, financial expert's team is scarce and is an expensive resource. Hence, there was a call for developing new, off-line and computer-aided monitoring mechanisms but as beginning start with those CAMELS indicators to evaluate the bank performance will be a wise choice.

This paper will start by discussing the factors that affecting banks, we will cover the literature of our problem organizing by the type of techniques that found to manage the research problems in the existing literature which are broadly: statistical techniques and intelligent techniques so our research methodology is similar to kumar et al. [25] review process. This paper will cover them all with dedicated focusing on their application on banks prediction process which our research problem. Then the historical background of financial prediction techniques will be covered, then the recent studies and present their adopted techniques is highlighted, then we give a quick review of selected techniques followed by the conclusion of the research.

Factors Affecting Banks

There are many factors which we can depend on to classify banks like a distressed ones, following we list some of them :

i. The bank's operation was temporarily suspended
ii. The bank was recapitalized or has received any liquidity support from the monetary authority
iii. The bank eventually merged with another bank due to financial distress (i.e. distressed mergers).
iv. The bank was closed by the government.
v. The ratio of non-performing loans to total loans during two subsequent years belongs to the fourth Quartile of the sample empirical distribution of this ratio.

But ultimately we can group the factors into following groups as we noticed that in the literature review process:

CAMELS Type Factors

CAMELS rating evaluate banks in their basic functional areas, which give a good indicator of the health and financial solvency of rated institution.

Almost all CAMEL type variables are significant leading indicators at conventional levels. For instance, the coefficient of capital adequacy is might negatively associated with the probability of distress. This concept is robust across the three capital adequacy measures.

Out of these dentitions, we found that Tier 1 capital is the most significant as a leading indicator. The other two dentitions were also significant, but to a lower extent according to bank's experts. Thus, banks that are better capitalized are less likely to experience distress in the forthcoming year. Similarly, the asset quality was found to be an important. Loan loss provision is considered also from the important measures in CAMELS group [3, 4, 5].

Wheelock and Wilson [6] found that managerial inefficiency contributes significantly to bank distress. Similar results were found by Pantalone and Platt [7]. Another important factor that received attention is capital adequacy.

Oshinsky [8] summarized that well capitalized banks are less likely to suffer. Moreover, Berger [9] indicated that low capital may create an optimum environment for frauds and Moral hazard practices and riskier loan portfolios. The role of capital is reviewed by many researchers and the consensus is that low capital increases the likelihood of bank troubles.

The banks temptation to allocate more assets into risky marketable securities and to avoid liquidity stresses was studied by Wagner [10-15].

Non- CAMELS Type Factors

The non-CAMEL Type includes three factors: banks size, diversification and market power. The size is simply measured as the natural log of total assets. Generally, banks' revenues are generated from either interest on loans or from fees, commissions, foreign exchange dealing, brokerage and market trading income that is unrelated to the lending activities of banks. Therefore, a higher proportion of non-interest income to total operating income implies that the bank is more diversified and more engaged in other activities besides traditional lending.

A group of other studies have investigated non-CAMEL type indicators such as size, income diversification and market power. The probability of distress due to excessive risks assumed by large banks was studied by Louzis et al. [16]. They found that big banks have a high risk and more likely to fail. All these indicated that size and diversification increase agency cost, encourage value destruction, enhance risk taking activities, and reduce the incentives for monitoring.

Along the same lines, the studies of Stiroh and Rumble [17] and Goddard et al. [18] found that size and diversification may lead to volatile income and higher risk of failure.

Calomiris and Mason [19] gave a totally different and opposite point of view that indicated size reduces the chance of distress due to the benefits of diversification, and to the ease of raising additional equity capital in large banks compared to small banks. The effect of diversification of activities on banks' performance was also investigated by many others. In particular, the gains from economies of scale and economies of scope were studied by many other researchers which reflect the importance of such indicator in a bank's performance measurements.

Macroeconomic Factors

The macroeconomic environment is summarized using real GDP growth and inflation and calculated depending on the annual percentage change of real GDP (at constant certain year prices) Inflation rates measured by the annual percentage change of the consumer price index.

Another trend of research investigates the effect of macroeconomic aggregates such as GDP growth, inflation, money growth and profit margins . The studies of Mannasoo and Mayes[19], and Aubuchon and Wheelock [18] are along these lines. All these found predictive information in macroeconomic aggregates, and that conditioning on macroeconomic indicators give very promising results in banks failure prediction process.

Other Sectors

There could be factors in other sectors which can affect the banks performance such like factors on the agricultural sector like annual rain rate which indirectly affect the banks by introducing troubles which faced the farmers in terms of productivity. Other sectors or domain might have the similar affects for instance industrial sector, political issues, social issues, Table 1 summarize these factors.

Table 1: other sectors which can affect the bank's performance [20]

Sector	Factors	Affect
Agricultural sector	- Annual rain rate - Crops enemies.	If the rain rate is quite small banks might face troubles with farmer's obligations satisfaction, same thing for the ruined cropped.
Political sector	- Wrong decisions. - Public strikes.	One wrong decision like changing the exchange rate policies in same area the public strikes might affect the banks situation.
Social Sector	- Consumer behavior - Customers awareness	Some customers have a common believe that finance shouldn't pay back specially in the micro finance operation; also customer's behavior might affect the project's success.
Money Market	- Financial instruments	If for instance stock prices has declined suddenly that will affect banks

History of Financial Prediction Techniques

We have covered the literature of techniques that have been used to perform financial prediction process grouped by similar technology, we first started our review by the history of those techniques which are found in the literature (1990-2010) and in the next section we are covering the most recent prediction techniques that we were covered.

Statistical Techniques

Altman et al. [1] developed a new bankruptcy classification model called Zeta analysis, which incorporated comprehensive inputs. The data sample consists of 111 firms with seven variables each. Their results were proved that ZETA outperformed its counterparts in term of expected cost. Classification accuracy of this model ranged from 96% for one period prior to bankruptcy to 70% for five periods prior to bankruptcy. Martin [21] presented logistic regression to predict probability of failure of banks based on the data obtained from the Federal Reserve System, data sample. Ohlson [23] has developed the logit model to predict firm failure. The classification accuracy reported by him was 96.12%, 95.55% and 92.84% for prediction within one year, two years and one or two years respectively. Dietrich and Kaplan [25] developed a simple three-variable linear model to classify loan risks. He compared his model with Altman model and many other researchers' models and got better performance over them. They used six independent variables and one dependent variable suggested by individual experts. The model gave better accuracy than previous bankruptcy models in non-performing loan risk prediction accuracy. The data used in the study was obtained from the American and New York Stock Exchanges. The data set consisted of estimation sample of 40 bankrupt and 800 non-bankrupt firms and a prediction sample of 41 bankrupt and 800 non-bankrupt firms.

Canbas et al. [24] proposed an integrated early warning system (IEWS) by combining DA, logistic regression, pro-bit and principal component analysis (PCA), that provide a significant assistance to determining the banks suffering from serious financial problem.

PCA explored three financial components, which significantly explained the changes in the financial condition of banks. With these three factors they employed DA, logistic regression and probit models. Then, by combining all these together they constructed an IEWS. They pointed out their model has more powerful features in term of prediction accuracy than other proposed models. However from the review we have noticed that statistical techniques in stand-alone mode are no longer employed that most recent studies depend on intelligent techniques which prove high performance and more accurate results. So next we are going to cover non-statistic techniques that have been known to solve the problem of bank distress prediction.

ARTIFICIAL INTELLIGENCE METHODS

Neural networks

This section reviews the literature where different topologies of neural networks (NN) are proposed and compared with other techniques. We divided this section into three sub sections covering the applications of (i) back propagation trained NN (BPNN), (ii) self-organizing feature map (SOM) and (iii) other NN topologies such as probabilistic NN, auto associative NN and cascade correlation NN.

Back Propagation Trained Neural Network (BPNN)

Atiya [28] reviewed the applications of NN in bankruptcy prediction and developed an NN. He developed novel indicators for the NN, taking initial ideas from Merton [76]. For the study he collected data from defaulted and solvent US firms. He reported a prediction accuracy of 84.52% for the in-sample set and 81.46% for the out-of-sample set, In case of multiple prediction instants for one firm, the firm's data are either all in the in-sample set or all in the out-of-sample set in order to avoid bias. He proved that the use of the indicators in addition to financial ratios provided significant improvement.

BP neural networks are a nonlinear self-adaptive dynamic system, which simulates human's neural system structure. It is composed of a lot of collateral neural elements which have the ability of learning, memorizing, computing and intellectual handling. Generally include one input layer, connotative layers and one output layer. Each node between two close layers joins each other in single direction.

It used by [57] to set sets up early warning indicators for commercial bank credit risk, and carries out the warning for the credit risk in advance. The experiment has proved that this method is objective and effective. So it can provide theoretical basis, which is more scientific and credible for detection and early warning about commercial bank credit risk.

The factors used by [57] were shown in Table 2.

Table 2: Factors affect bank's performance [1]

Criterion	Factor name
The ability to refund	Asset-liability ratio
	Liquidity ratio
	Rapidly ratio
	Cash ratio
	Debt ratio of cash in business activities
	Current cash ratio
	Rate of main business profit and financial expenses
The ability to profit	Main business profit ratio
	Net assets profit ratio
	Asset profit ratio
Management and administration ability	Turnover rate of the account receivable
	Turnover rate of the stock
Development and potential ability	Increased income ratio of main business
	Rate of increase of the profit

The model is depend on the three layer design of the BP neural network, input layer deal only with numerical data so we need to standardize the quantitative indices (factors) , middle layer use 3 nodes to maintain minimum system error, output layer has one of three values, low risk, medium risk, high risk.

[58] Used BP artificial neural network model for the establishment of early warning and training and testing the sample data with it, and then prediction the state of the financial risk in 2008 with evaluation results show that as a result of the global financial crisis in 2007, in 2008 the financial situation of china is poor. In China's macro-control policy, the economic situation of china is a little better.

Swicegood and Clark [29] compared Discriminant Analysis (DA), BPNN and human judgment in predicting bank failures. The variables were taken from the bank call reports. Overall, Multivariate Discriminant Analysis (MDA) model correctly classified 86.4% and 79.5% of regional and community banks respectively. However, BPNN model correctly classified 81.4% and 78.25% of regional and community banks. They concluded that BPNN outperformed other two models (DA and human judgment) in identifying under performance banks. Lam [30] integrated fundamental and technical analysis in BPNN for financial performance prediction. The predictors included 16 financial and 11 macroeconomic variables. She concluded that BPNN significantly outperformed the minimum benchmark based on highly diversified investment strategy. Also, incorporation of previous years' financial data in the input vector for BPNN could significantly increase the return level, thereby, demonstrating the benefits of integrating fundamental analysis with technical analysis via BPNN. She also extracted rules from trained neural network and found that they outperformed the neural networks per se. Further, they performed as accurately as the maximum benchmark.

Lee et al. [31] compared BPNN with self-organizing feature map (SOM), DA and logistic regressions. The data sample consisted of 168 Korean firms taken from the Security and Exchange Commission (SEC)). The fourfold cross-validation testing was used for all the models. They concluded that the BPNN give better results than other models as well.

Self-Organizing Maps (SOM)

Lee et al. [31] proposed three hybrid BPNN which are : (i) MDA-assisted BPNN (ii) ID3-assisted BPNN and (iii) SOM-assisted BPNN for predicting bankruptcy in firms. The data of 166 firms is taken from the Korea Stock exchange. They selected 57 financial variables. They concluded that hybrid neural network models performed better than MDA.

Kaski et al. [32] worked on Fisher information matrix based metric and implemented SOM with it. They used novel method to understand the non-linear dependencies between bankruptcies and financial indicators. The dependencies were converted into a metric of the input space and the SOM was used to visualize the dependencies in a concise form. They obtained 23 financial indicators from Finnish small and medium-sized enterprises. They computed the accuracy of SOMs in the Euclidean metric (SOM-E) and in the Fisher metric (SOM-F) in representing the probability of bankruptcy, measured by the likelihood of data at the location of the best matching SOM units. He concluded that the SOM-F performed better than the SOM-E.

Other Neural Network Topologies

Lacher et al. [62] proposed a cascade-correlation neural network (Cascor) for classifying firm's financial performance. Altman's five financial ratios were used. Data was collected from Standard and Poor's COMPUSTAT financial database. He compared the performance of the Cascor model with that of Altman Z-score model. They concluded that the Cascor model consistently yielded higher overall classification rates.

Baek and Cho [33] proposed the auto-associative neural network (AANN) for Korean firm bankruptcy prediction. They trained the AANN with only solvent firms data. Then they applied the test data containing both solvent and insolvent firms. So, any solvent firm data that shared common characteristics with the training data resulted in small error at the output layer while the bankrupt firms data resulted in a large error at the output layer. AANN yielded classification rates of 80.45% for solvent and 50.6% for defaulted firms.

However, the 2-class BPNN produced classification rates of 79.26% for solvent and 24.1% for defaulted firms. Therefore, they concluded that AANN outperformed 2-class BPNN.

Case-based reasoning (CBR) Techniques

Park and Han [34] has introduced analytic reasoning model called the K-NN with analytic hierarchy process (AHP) feature weight approach for bankruptcy prediction. They proposed CBR for indexing and retrieving similar cases. The AHP-weighted K-NN was compared with pure K-NN algorithm.

For this study, they manipulate the data fetched from Industrial Bank of Korea. They used both financial and non-financial variables. The classification accuracy of pure KNN approach was 68.3% whereas the hybrids Logit-CBR and AHP-K-NN-CBR produced 79.2% and 83.0%, respectively. They concluded that weighted K-NN hybrid model outperformed other models.

Yip [35] used CBR with K-NN to predict Australian firm business failure. She used the statistical evaluations for assigning the relevancy of attributes in the retrieval phase of algorithm. She compared the performance of CBR + K-NN with that of DA. The overall accuracy of CBR with weighted K-NN, CBR with pure K-NN and DA were 90.9%, 79.5% and 86.4%, respectively. She concluded that CBR with weighted K-NN was better than DA.

DECISION TREES

In this section, we are reviewing the studies that have been published regarding decision trees application on financial health problem prediction. Decision trees use recursive partitioning algorithm to define rules on a given data set.

Nanda and Pendharkar [36] incorporated misclassification cost matrix into an evolutionary classification system. Using simulated and real-life bankruptcy data, they compared the proposed method with LDA, a goal programming and a GA-based classification without the asymmetric misclassification costs. For bankruptcy training set the classification accuracy of integrated cost preference based minimized sum of deviations (ICPBMSD) and integrated cost preference based GA (ICPB-GA) outperformed LDA, MSD and GA. For simulated holdout set the ICPB-GA outperformed others. They concluded that the ICPBMSD or ICPB-GA might be promising when compared to traditional MSD or GA.

Shin and Lee [37] also proposed a GA-based approach for bankruptcy prediction. The rules generated by GA were easily understood and could be used as expert systems. Data used contained 528 firms. The five rules generated by GA got 80.8% accuracy. They concluded that GA could successfully learn linear relationship among input variables.

Operational Research

Lam and Moy [38] compared different DA methods and proposed their combination to predict the classification of new observations. For testing this hybrid technique, they used simulation experiment. DA taken for this study was FLDF (Fisher linear discriminant analysis), cluster-based LP (CBLP) and MSD (minimized sum of deviations). Their simulation generated data from contaminated multivariate normal distribution for the first 24 cases and for remaining 10 cases contained non-contaminated data to test the robustness of LPC, which combined classification results of different DA. Their combined method outperformed other DA methods. LPC was reformulated as a mixed-integer programming model, which minimized weighted number of mis-classifications

Rough Sets

McKee [41] developed a rough set based bankruptcy prediction model. Rough set analysis produced better results when Here, the variables identified by recursive partitioning method were used to develop rough set based model which yielded 88% accuracy in predicting bankruptcy on a 100-company holdout sample. This was superior to the original recursive partitioning model, which was only 65% accurate on the same data set. Bioch and Popova [42] proposed a modification of the rough set approach for bankruptcy prediction. They used monotone extensions, decision lists and dualizations to compute classification rules that cover the whole space. They computed the monotone decision trees for the dataset. They concluded that there was close relationship between the decision rules obtained using the rough set approach and the prime implicants of the maximal extension.

Soft Computing Techniques

In this section, studies employing various hybrid intelligent techniques to solve bankruptcy prediction problem are reviewed. These studies developed new hybrid systems combining the intelligent techniques and sometimes some proven statistical models. Three varieties of soft computing architectures have been employed so far: (i) ensemble classifier [47], where techniques such as BPNN, DA, logistic regression, MARS, C4.5, fuzzy logic, rough set based approach, GA, GP were employed to solve the problem in a stand-alone mode and then their results are combined through an arbitrator which performs simple majority voting or weighted majority voting schemes or a linear combination of predictions; (ii) an intelligent technique is used for feature selection task and another intelligent technique performs classification by taking the selected features and (iii) tightly integrated hybrid systems such as GA trained NN, neuro-fuzzy, GA-neuro-fuzzy, etc.

McKee and Lensberg [41] presented a hybrid approach to bankruptcy prediction by integrating a rough set model with genetic programming (GP). They suggested a two-stage hybrid model: stage 1 used a rough set model to identify subsets of important explanatory variables and stage 2 comprised a GP algorithm to develop a structural model of bankruptcy based on those variables. Rough set model yielded 100% classification accuracy on training set and 67% on validation set. The GP model obtained 82.6% accuracy on the training set and 80.3% on the validation set. This was significantly higher than the accuracy achieved by the rough set model. The type I and type II errors were analyzed using second genetic model. New model was trained using the same GP algorithm, except that the validation sample of first GP model was used as the training data for second GP model. These two models yielded 81% and 83% accuracies respectively over the entire sample. They concluded that the GP approach coupled with rough sets could be efficient approach for bankruptcy prediction problem.

Bian et al. [49] proposed a new fuzzy-rough-nearest neighbour hybrid for bankruptcy prediction. They compared it with crisp K-NN and fuzzy K-NN. Fuzzy-rough K-NN provided better results in identifying the non-profitable companies. When using decision tree feature selection method, fuzzyrough K-NN had the lower type I error of 25%. Overall, they concluded that fuzzy-rough K-NN performed better in minimizing type I error, while having satisfactory type II error.

Pendharkar and Rodger [50] studied GA-based ANN on bankruptcy prediction problem. They chose arithmetic cross- over operator for the GA-based ANN. They used real-valued GA to learn the connection weights of the ANN.. They showed that GA-based ANN training showed resistance towards overfitting by keeping the ANN architecture constant. Further, they concluded that GA-based ANN with 1-point crossover, arithmetic crossover and uniform crossover performed comparably with BPNN on holdout sample.

Other Techniques

We will review the least commonly used techniques and group them into this section, the studies applying fuzzy techniques, SVM and isotonic separation are reviewed.

Alam et al. [44] proposed fuzzy clustering for identifying potentially failed banks and then compared it with two SOM networks which are, (i) competitive neural network and (ii) SOM. They concluded that both fuzzy clustering and SOM are good tools in identifying potentially failing banks.

Andres et al. [45] proposed fuzzy rule based classifiers for bankruptcy prediction problem. They compared the performance of LDA and logistic regression with multi-layer perceptron and fuzzy-rule based classifiers. They also used Monte Carlo simulation to measure the effects of sample size variations on the performance of classifiers. The distinctive features of this study were: (i) for each classifier and a wide range of sample sizes, average error rates were estimated from the results of a large number of Monte Carlo simulations. (ii) The focus was on business profitability analysis, which was not considered earlier. (iii) Their classification problem had a low separability degree. (iv) A slight variant of the class of additive fuzzy systems with Gaussian membership functions and consequent normalized to be probabilities was tested.

They used the database of commercial and industrial firms located in Spain. They concluded that MLP and fuzzy rule based classifier outperformed LDA and logistic regression.

Text Mining Techniques

Text mining, in general, refers to uncovering the hidden, meaningful and important information through unstructured data analysis and feature extraction process to extract hidden information,then processed and stored as reusable knowledge.

According to Sullivan [54], text mining is a kind of the process of editing, organization and analysis of a large number of documents. It mainly provides analysts or policy makers with distinguished knowledge resulting from processing those unstructured data.

In the literature there are rich information about how could we text mine unstructured sources to predict specific goals, most of the available studies concentrating on stock market prediction using text mining, however there are none researches on bank failure prediction using customer's complaint letters to the best of my knowledge.

As general the two problems (stock market and bank distress prediction) are similar and there are few discrepancies between them therefore we can address the possibility of applying text mining processes to predict bank's failure.

In categorizing the bank's distress prediction systems different dimensions can be considered as it's adopted also in stock market prediction systems [15]:

Input data: Some prediction methods are based on historical bank scope values and use technical analysis to predict the bank's distress. Some other methods are based on analyzing the news content, letters of complaints; however combination of both types can also be used.

Prediction goal: The possible bank failure prediction goal can be the possibility of closing the financial institution from government, financial insolvency, organizational collapse, etc.

Prediction horizon: prediction horizon is the time span in which the prediction would be valid. It can be short-term or long-term prediction. Short-term prediction starts from 1 month to one year after the information source is analyzed and long term starts from 2 year hours and can last longer.

Complaints Letters - based bank distress prediction can be considered as a text classification task. Generally the goal is to forecast some aspects of the banks such as insolvency or failure based on the news content. Based on prediction goal described in previous section, a set of final classes are defined, such as "distressed" (which means this news indicates that bank is going to distress), "Non-distressed" (which means this piece of news indicates that bank are not going to distress) and etc. the prediction system is supposed to classify the incoming news into one of these classes.

News based market prediction can be divided into two main phases. "Training phase" and "Operational phase". In operational phase, one of the predefined classes will be assigned to incoming news; however, to make the system ready for the operational phase a classifier should be trained in the training phase. Machine learning techniques are widely used to automate such processes.

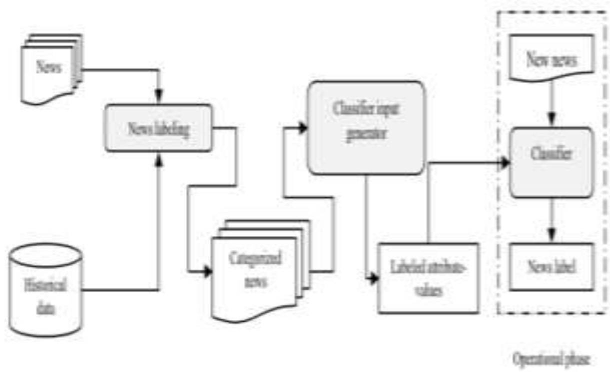

Figure 1: overall view of a news-based banks prediction system [15]

Previous studies

As we stated earlier most of the text mining prediction techniques were targeted the stock market prediction area, that this area is high potential to get instant benefit from especially with the degree of dependence on the internet and social media as one of the trust information sources to predict stock markets trend and prices.

Following in Table 3 we can have a glance to the summary of text mining prediction techniques that were adopted to predict specific goal using unstructured data along with other information such like which classification methods were applied in Table 3.

Zhai et al. [56] trained a classifier dedicated for 7 technical indicators in addition to another classifier for the news classification and finally combined the result of these classifiers. In the method proposed in [68], stock price at time of the news released included in the input vector which increased the prediction accuracy.

Table 3: Summary of text mining prediction techniques [15].

	Classification		Evaluation		
Developed Systems	Classifier	Number of categories	Period	Directional accuracy	Simulated trader
Fung/Lam/Yu [67]	SVM	3	7 month		Yes
Mittermayer/Knolmayer [69]	SVM	4 (3 for training)	9 month	45%	Yes
Soni/Eckt/Kaymakt[70]	SVM	2	11 years	56.2 %	NA
Zhai/Hsu/Halgamuge [66]	SVM	2	15 months	70.01%	Yes
Rachlin/Last/Alberg/Kandel [56]	Decession Trees	5	3 months	82.4 %	Yes
Schumaker and Chen[68]	SVM	3	1 month	58.2 %	Yes
Mahajan/Dey/Mirajul Haque [71]	stacked classifier(Decision tree+ SVM)	2	3 years	60%	No

Among proposed methods for feature selection and weighting, just [40] followed semantic approaches. The rest of the methods used Bag of words as their feature selection in which basically term frequency–inverse document frequency (tf-idf) technique is applied in term selection and weighting. To improve feature selection [60] and [66] included some predefined terms in the classifier input vector and reported that it increased the classification accuracy.

Recent Prediction Techniques for Financial Sector

Here we focused on the recent financial prediction techniques explain the technique then which factors and models has been used as well as evaluation results for all.

From our review process we conclude that two major categories of financial prediction systems are exist, stand -alone and non-stand-alone systems, so we are going to group our reviewed techniques regarding to those two groups depending on the most recent studies that was found in the literature (2010 – 2013).

Stand-Alone Techniques

It includes all techniques that are found in the literature applied on a stand-alone mode to predict any financial sector problems.

Logistic Regression

In logit model the failure indicator is a binary variable estimated using a set of explanatory variables such as financial ratios as Altman [1], by using the logit estimation, the predicted outcomes are restricted to lay in the unit interval, and are considered as the probability of failure. With This methodology, it is possible to evaluate the explanatory contribution of each independent variable, which can be seen as an advantage of the model.

The logit model has the statistical property of not assuming multivariate normality among the independent variables, contrary to the probit model that does assume a normal distribution of the data. This can be seen as an advantage when analyzing banking data, as it generally is not normally distributed so it has selected by [61] to predict the failure of large United States banks.

Also [51] worked to measure the failure risk of Turkish commercial banks. By using 29 financial ratios across 1996-2000 and apply principal component analysis to determine significant changes in the financial conditions of banks. Then employ these financial conditions, captured in factor scores, in the logit analysis to build an early warning model. Finally, they predicted the probabilities of failure for 25 commercial banks from 2002-to date.

Classifier Combination

These types of systems are varies in their design as example an intelligent technique can be used for feature selection task and another intelligent technique performs classification by taking the selected features and another design can be tightly integrated hybrid systems such as GA trained NN, neuro-fuzzy, GA-neuro fuzzy, etc. Following we are covering the studies that have been found on the recent literature:

Rough Sets and Neural Network

Most of the studies use rough sets to streamline, and then BP neural network was trained on the samples to determine risk of certain problem on hand, we know that Rough set is invented study of incomplete and uncertain knowledge and data representation which are very popular type of knowledge in financial systems early warning systems and as it clear neural network is came to achieve or imitate human brain information processing, learning, memory, knowledge storage and retrieval functions.

[16] Has used rough set and BP neural network to predict risk of default on Personal loan, so default index first constructed, and then use rough sets to streamline, and then BP neural network was trained on the samples to determine risk of default. Results showed that rough set and BP neural network test samples of prediction accuracy.

Their Data from a commercial bank committed to reducing the loan default rate data, including 850 former customers and prospective customers are the financial and demographic information. Observations have been made before the 700 has a lending value of the customer before and after observations of 150 banks need to do is assess the potential credit risk customers.

The results showed that the reduction of condition attributes after reduction, but the accuracy of early warning is higher. It should also be seen, limited to loans index system and the lack of sample data.

Modeling Logic and Neural Network

These techniques are aimed to analyze to what extent neural based modeling techniques can be framed within the long- established framework defined by the types of modeling techniques that exist in logic. The underlying philosophy of this study is that, potential synergies identified between the logic and neural basedapproaches, could set the basis for enhanced bankruptcy prediction models development via furtherexploitation and inclusion of human intelligence and reasoning capabilities on neural based systems.

[63] Prepared study synergies between logic and neural based approaches as the basis to enhance bankruptcy prediction models development.

The main takeaways are that it is indeed possible to apply the logic categorization to neural based systems and that deductive reasoning results are more appropriate for Neural Networks predictability models

Intelligent classification techniques

As we mentioned earlier that intelligent techniques can be used for feature selection task and another intelligent technique performs classification by taking the selected features, for instance [18] uses the same idea for bank failure prediction, one built using raw accounting variables and the other built using constructed financial ratios. Four popular data mining methods are used to learn the classifiers: logistic regression, decision tree, neural network, and k-nearest neighbor. He evaluated the classifiers on the basis of expected mis-classification cost under a wide range of possible settings. The results of the study strongly indicate that feature construction, guided by domain knowledge, significantly improves classifier performance and that the degree of improvement varies significantly across the methods.

With the four chosen classification techniques, the performance of each learned classifier was measured in terms of (normalized) expected mis-classification cost. Forecast of likely bank failures is subject to two types of errors. A false negative error is made when a classifier mis-classifies an actual failure as a survivor. A false positive error is made when a classifier mis-classifies an actual survivor as a failure.

The ANOVA results show that the use of the financial ratios, instead of raw accounting variables, significantly improves the performance (p<0.01), with respect to expected mis-classification cost, of classifiers learned using several widely used classification methods, including logistic regression, C4.5 decision tree, back-propagation neural network, and k-nearest neighbor. The results hold across all settings of prediction period, prior probability of failure, and cost ratio. The results imply that feature construction, guided by domain knowledge, significantly improves classification performance.

Another technique used Group Method of Data Handling neural network, counter propagation neural network and fuzzy ARTMAP proposed by [64]. They are present three hitherto unused neural network architectures such as CPNN, GMDH and fuzzy ARTMAP for bankruptcy prediction in banks. Ten-fold cross validation is performed throughout the study. Each of these techniques is tested in their stand-alone mode on Spanish, Turkish, US and UK banks' datasets. Out of these GMDH they have presented three hitherto unused neural network architectures for bankruptcy prediction in banks. These networks are Group Method of Data Handling (GMDH), Counter Propagation Neural Network (CPNN) and fuzzy Adaptive Resonance Theory Map (fuzzy ARTMAP). T-statistic, f-statistic and GMDH are used for feature selection purpose and the features so selected are fed as input to GMDH, CPNN and fuzzy ARTMAP for classification purpose. In each of these cases, top five features are selected in the case of Spanish dataset and top seven features are selected in the case of Turkish and UK datasets.

Their results were the features selected by t-statistic and f-statistic is identical in all datasets. Further, there is a good overlap in the features selected by t-statistic and GMDH. The performance of these hybrids is compared with that of GMDH, CPNN and fuzzy ARTMAP in their stand-alone mode without feature selection. Ten-fold cross validation is performed throughout the study. Results indicate that the GMDH outperformed all the techniques with or without feature selection. Furthermore, the results are much better than those reported in previous studies on the same datasets in terms of average accuracy, average sensitivity and average specificity.

Quick Review

We can summarize our review results in the following table containing the basic idea of each used technology its advantages and disadvantages (see Table 4).

Table 4: Quick Review.

Technology	Basic idea	Advantages	Disadvantages
FL	Models imprecision and ambiguity in the data using fuzzy sets and incorporates the human experiential knowledge into the model	Low computational requirements , more powerful to replace human knowledge	the excess of choices for membership function shapes, connectives for fuzzy sets and defuzzification operators are the disadvantage
NN	Learn from examples using several constructs and algorithms just like a human being learns new things	Good at function optimization tasks	Many neural network architectures need a lot of training data and training cycles (iterations)
GA	Copycats Darwinian principles of evolution to solve highly nonlinear optimization problems	Find solutions without getting trapped in local minima	May not yield global optimal solution always unless it is augmented by a suitable direct search method
CBR	Learns from examples using the Euclidean distance and k-nearest neighbor method	similar to the human like decision-making	Generalization problems
Rough Sets	model uncertainty in the data	Simplicity from using IF-Then rules to form classification	sensitive to changes in data and non-trusted accuracy
SVM	Exploit statistics to carry out classification and regression tasks	It can work well with few samples	It has high algorithmic complexity and requires extensive memory
Decision Trees	They use recursive partitioning technique to induce decision trees on a data set	They yield human comprehensible binary 'if–then' rules	they require a lot of data samples in order to get reliable predictions
TF	assumed that important terms occur in the document collection more Often than unimportant, used in text mining process.	have the highest explanatory power	High computational burden with large documents.

DISCUSSION

The following observations can be made from the current review; a vast majority of the studies reviewed here pertain to firms and not banks. Further, majority of the studies are conducted by taking the data sets relevant to the time period 1980–2012. Moreover, for a vast majority of the works, the country of origin of the data sets is USA followed by European countries. Further, it is observed that variety of statistical and intelligent techniques have already been applied to the bankruptcy prediction problem. The general observation is that the statistical techniques such as logistic regression, LDA, QDA, FA were all outperformed by the most popular NN architecture e.g. BPNN, wherever comparisons were made between these two families of techniques.

This is not surprising because the BPNN with logistic activation function can be thought of as a confluence of several logistic regressions fitted together in parallel. Thus, the non-linearity in the data can be modeled better by the BPNN.

Comparing decision trees with NN architectures, even in cases where they yield identical performance, we noticed the efficiency of using decision trees as they produce an important by-product i.e. ‘if–then’ rules. These rules can be used as an ‘early warning’ expert system later on. Depending on the data set, both techniques are capable of outperforming each other in terms of accuracy. However, in the case of bankruptcy prediction problems, BPNN outperformed the decision trees.

Further, BPNN, DA and logistic regression outperformed the CBR implemented without weights. Thus, CBR with the simple k-nearest neighbor method at its heart cannot solve classification problems with nonlinear boundaries. In addition, the CBR cannot generalize well. Then, rough set based approach outperformed DA, logistic regression and a decision tree. However, rough set based approaches can be, in general, inaccurate and sensitive to changes in data.

With regard to DEA, it yields us only the relative scores of the efficiency levels of the banks. On the other hand, SVM can be very accurate and does not have local minima problems unlike BPNN and it can also be trained with a small training set. It is observed that SVM outperformed DA, logistic regression and BPNN. Fuzzy logic based techniques are least exploited in bankruptcy prediction research. In the few studies where they were applied, fuzzy rule based classifier outperformed LDA and logistic regression, but BPNN outscored fuzzy rule based classifier. More importantly, fuzzy ‘if–then’ rules can be used as ‘early warning’ fuzzy expert systems. In general, in the soft computing architectures, ensemble classifiers outperformed the individual models, which is on the expected lines.

Further, it can be observed that the interest to design and employ a variety of soft computing architectures has increased significantly over the past decade. Statistical techniques are no longer preferred in view of their low accuracy.

It’s clear from the text mining literature review that SVM prediction techniques is most used by the previous studies and there is no other techniques have been used except decision trees and just one study try to get benefit from both by utilizing SVM and decision trees as on stacked classifier [17].

Bank’s failure factors have been study elaborately in literature, however we have presented new and not-studied before factors which come from other sectors, those factors have been identified after consulting subject matter experts with special attention to Sudan banking sector that is highly depend on agriculture as a backbone of its national economic..

So from the provided facts we can conclude that in order to have more accurate and depend-upon results the only trend that researchers should adopt in their new study is the hybrid soft computing methods and TF text mine indicator.

CONCLUSION

So far we have done provided a comprehensive literature review for the bank distress, we have discussed the techniques that have been known to deal with the problem of banks prediction, which are broadly divided to two sections: statistical and intelligent techniques.

We have discussed their sub types and applications as well as provide discussion of its feasibility of being implemented as banks distress prediction technique. Also we have presented a quick reference of those techniques explained their ideas, positive and negative aspects. As a future work, we will try to combine the results of text mining the complaints letters with the CAMEL ratio system to reach an optimal decision regarding the bank's distress condition.

REFERENCES

[1] E. Altman, (1968), "Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy", Journal of Finance Vol. 23, PP. 589–609.

[2] http://www.solvency.com/solvency.htm (Retrieved in July 2005).

[3] R.L. Wilson, R. Sharda, (1994),"Bankruptcy Prediction Using Neural Networks", Decision Support Systems Vol. 11, PP. 545–557.

[4] R. Cole J. Gunther, (1995), "A CAMEL Rating's Shelf Life", Federal Reserve Bank of Dallas Review, PP.13–20.

[5] G.V. Karels A. J. Prakash, (1987), "Multivariate Normality and Forecasting for Business Bankruptcy", Journal of Business Finance and Accounting Vol. 14, PP. 573–593.

[6] Wheelock D., Wilson, P., (2000), "Why Do Banks Disappear? The Determinants of US Bank Failures and Acquisitions". The Review of Economics and Statistics Vol. 82, No.1, PP. 127–138.

[7] Pantalone C., C. Platt, (1987) "Commercial Bank Failure since Deregulation". Economic Review, PP. 37–47.

[8] Oshinsky R. Olin, (2006), "Troubled Banks: Why Don't They All Fail?" FDIC Banking Review, Vol. 18, PP. 23–44.

[9] Berger A., Klapper L.F., Turk-Ariss R., (2009), "Bank Competition and financial Stability". Journal of Financial Services Research Vol. 35, PP. 99–118.

[10] Cole R.A., White L.J. Déjà Vu, (2012), "All Over Again: The Causes of Us Commercial Bank Failures This Time Around. Journal of Financial Services Research Vol. 42, PP. 5–29.

[11] Barrell et al., (2010), "Bank Regulation, Property Prices and Early Warning Systems for Banking Crises In OECD Countries". Journal of Banking and Finance Vol. 34, PP.2255–2264.

[12] Torna, G, (2010), "Understanding Commercial Bank Failures in the Modern Banking Era".

[13] Barth M. Landsman, (2010), "How Did financial Reporting Contribute To The financial Crisis?" European Accounting Review Vol. 19, PP. 399–423.

[14] Laeven, L, (2010), "Banking Crises: A Review", Annual Review of Financial Economics Vol. 3 No. 4, PP. 1–24.

[15] Azadeh Nikfarjam, M,(2010), "Text Mining approaches for stock market prediction" Computer and Automation Engineering, Vol. 4, PP. 256-260.

[16] Louzis, D. P., (2012), "Macroeconomic and Bank-Specific Determinants of Nonperforming Loans In Greece: A Comparative Study Of Mortgage, Business And Consumer Loan Portfolios". Journal of Banking and Finance Vol. 36, PP. 1012–1027.

[17] A. Mahajan, L. Dey, (2008), "Mining Financial News for Major Events and Their Impacts on the Market", IEEE /WIC /ACM International Conference on Web Intelligence and Intelligent Agent Technology, PP. 936 - 947.

[18] Malek A, (2009), "Study in bank failure prediction, Expert Systems with Applications" Vol. 36, PP. 3355-3366.

[19] Männasoo, K., (2009)," Explaining bank distress in Eastern European transition economies". Journal of Banking and Finance Vol. 33, PP. 244–253.

[20] Central Bank of Sudan Studies, (2011), Research Department, Annual Department Report.

[21] D. Martin, (1999), "Early Warning of Bank Failure: A Logit Regression Approach "Journal of Banking and Finance Vol. 1, PP. 249–276.

[22] J.A. Ohlson, (1980), Financial Rations and The Probabilistic Prediction of Bankruptcy, Journal of Accounting Research Vol. 18, PP. 109–131.

[23] S. Canbas, A. (2005),"Prediction of Commercial Bank Failure via Multivariate Statistical Analysis Of financial Structure: The Turkish Case, European" Journal of Operational Research Vol. 166, PP. 528–546.

[24] J.R. Dietrich, R.S. Kaplan, (1982),"Empirical Analysis of the Loan Classification Decision", the Accounting Review Vol. 57, PP. 18–38.

[25] R.C. West, (1985), "A Factor Analytic Approach to Bank Condition", Journal of Banking and Finance Vol. 9, PP. 253–266.

[26] G.V. Karels, A. J. Prakash, (1987),"Multivariate Normality and Forecasting of Business Bankruptcy", Journal of Business Finance and Accounting Vol. 14, No. 4.

[27] A.F. Atiya, (2001), "Bankruptcy Prediction For Credit Risk Using Neural Networks: A Survey and New Results", IEEE Transactions on Neural Networks Vol. 12, No. 4, PP. 929–935.

[28] P. Swicegood, (2001), "Off-Site Monitoring For Predicting Bank Under Performance: A Comparison of Neural Networks, Discriminant Analysis And Professional Human Judgment", International Journal of Intelligent Systems in Accounting, Finance and Management Vol. 10, PP. 169–186.

[29] M. Lam, (2004),"Neural Networks Techniques for Fnancial Performance Prediction: Integrating Fundamental and Technical Analysis", Decision Support Systems Vol. 37, PP.567–581.

[30] K. Lee, D. (2005), "A Comparison of Supervised and Unsupervised Neural Networks in Predicting Bankruptcy of Korean Frms", Expert Systems with Applications Vol. 29 .PP. 1–16.

[31] S. Kaski, J. (2001), "Bankruptcy Analysis with Self-Organizing Maps in Learning Metrics", IEEE Transactions on Neural Networks Vol. 12 No. 4.

[32] J. Baek and S. Cho, (2003),"Bankruptcy Prediction for Credit Risk using an Auto Associative Neural Networks in Korean firms", in: IEEE International Conference on Computational Intelligence for Financial Engineering, Hong-Kong.

[33] C.-S. Park, I, (2009), "A Case-Based Reasoning with the Feature Weights Derived By Analytic Hierarchy Process for Bank. Expert Systems with Applications, Vol. 23, PP. 255 – 264.

[34] A.Y.N. Yip, Predicting Business Failure with A Case-Based Reasoning Approach, lecture notes in computer science, in: M.G. Negoita, R.J. Howlett, L.C. Jain (Eds.), Knowledge Based Intelligent Information and Engineering Systems: 8th International Conference, KES, Wellington, New Zealand, September 3215/2004, Proceedings, Part III, PP. 20–25.

[35] S. Nanda, P. (2010), "Linear Models for Minimizing Mis-classification Costs in Bankruptcy Prediction, International Journal of Intelligent Systems in Accounting", Finance and Management Vol. 10, PP. 155–168.

[36] K. -S. Shin, (2012),"A Genetic Algorithm Application in Bankruptcy Prediction Modeling", Expert Systems with Applications Vol. 23 No. 3, PP. 321–328.

[37] K.F. Lam, (2012), "Combining Discriminant Methods In Solving Classification Problems In Two-Group Discriminant Analysis", European Journal of Operational Research Vol. 138, No. 294–301.

[38] A. Cielen, (2004), "Bankruptcy Prediction Using A Data Envelopment Analysis", European Journal of Operational Research Vol. 154, PP. 526–532.

[39] C. Kao, S.-T. Liu, (2004),"Prediction Bank Performance With financial Forecasts: A Case Of Taiwan Commercial Banks", Journal of Banking & Finance Vol. 28, PP. 2353–236.

[40] T. E. McKee, (2010), "Developing A Bankruptcy Prediction Model via Rough Sets Theory", International Journal of Intelligent Systems in Accounting", Finance and Management Vol. 9, PP.59–173.

[41] J. C. Bioch, (2011), "Bankruptcy Prediction with Rough Sets", ERIM Report Series Research in Management (ERS2011-11-LIS).

[42] T. E. McKee, (2013),"Rough Sets Bankruptcy Prediction Models versus Auditor Signaling Rates", Journal of Forecasting Vol. 22, PP. 569–589.

[43] P. Alam, D. Booth, (2000), "The Use of Fuzzy Clustering Algorithm and Self-Organizing Neural Network For Identifying Potentially Failing Banks: An Experiment Study", Expert Systems with Applications Vol. 18, PP. 185–199.

[44] J.D. Andres, (2005), "Forecasting Business Profitability by using Classification Techniques: A Comparative Analysis Based on A Spanish Case", European Journal of Operational Research Vol. 167, PP. 518–542.

[45] J. H. Min, Y.-C., Lee, (2005),"Bankruptcy Prediction Using Support Vector Machine (SVM) With Optimal Choice of Kernel Function Parameters", Expert Systems with Applications Vol. 28, PP. 603–614.

[46] Y. U. Ryu, (2005), "Firm Bankruptcy Prediction: Experimental Comparison of Isotonic Separation And Other Classification Approaches", IEEE Transactions on Systems, Management and Cybernetics-Part A: Systems and Humans Vol. 35 No. 5, PP. 727–737.

[47] F.Y. Lin, (2011), "A Data Mining Approach To The Prediction Of Corporate Failure", Knowledge Based Systems Vol. 14, PP. 189–195.

[48] H. Bian, (2003), Fuzzy-Rough Nearest Neighbor Classification Approach, in: 22nd International Conference of the North American Fuzzy Information Processing Society (NAFIPS 2003) Proceedings Chicago, PP. 500–505.

[49] P.C. Pendharkar, (2004), "An Empirical Study of Impact of Crossover Operators on the Performance of Non Binary Genetic Algorithm Based Neural Approaches For Classification", Computers and Operations Research Vol. 31, PP. 481–498.

[50] W.L. Tung, (2005),"A Novel Neural-Fuzzy Based Early Warning System For Predicting Bank Failures", Neural Networks, Vol. 17, PP. 567–587.

[51] F.-M. Tseng,(2005), "A quadratic interval logit model for forecasting bankruptcy", Omega Vol. 33, PP. 85–91.

[52] P. Ravi Kumar, (2007), "Bankruptcy Prediction in Banks and Frms via Statistical and Intelligent Techniques – A Review". European Journal of Operational Research Vol. 180, No. 1, PP. 1–28.

[53] Sullivan, D., (2001), "Document Warehousing and Text Mining", Canada: Wiley Computer Publishing

[54] G. Rachlin, M. Last, (2007), "ADMIRAL: A Data Mining Based Financial Trading System," IEEE Symposium on Computational Intelligence and Data Mining, PP. 720-725.

[55] Zhi-Yuan Yu1, (2011), "Bank Credit Risk Management Early Warning and Decision-Making Based On BP Neural Networks", International Symposium on IT in Medicine and Education (ITME), Vol. 2, PP. 528 – 532.

[56] Feng Kong, (2009),"The Early Warning of Financial Crisis in China", IEEE International Conference on Information Engineering and Computer Science, PP. 1 - 4.

[57] Wang Baosen, (2010),"Applied Research of entropy in comprehensive evaluation of the risk of commercial banks."

[58] Kenan Lopcu (2012), Effects of Structural Changes in the Turkish Banking Sector, Topics in Middle Eastern and African Economies Vol. 14.

[59] Zhang Zenglian, (2011), "Research of Default Risk of Commercial Bank's Personal Loan Based on Rough Sets and Neural Network", University of Science and Technology Beijing.

[60] Yakup Kara, (2009),"Study In Bank Failure Prediction, Expert Systems with Applications" Vol. 36, PP. 2633–2644.

[61] P. Ravisankar, V. Ravi, (2010), Financial Distress Prediction in Banks using Group Method Of Data Handling Neural Network, counter propagation neural network and fuzzy ARTMAP, LR.

[62] Y. Zhai, A. Hsu, (2007),"Combining News and Technical Indicators in Daily Stock Price Trends Prediction," Lecture Notes in Computer Science, PP. 1087-1096.

[63] G. Fung, (2002), "News Sensitive Stock Trend Prediction," Lecture Notes in Computer Science, vol. Volume 233, p. 481– 493.

[64] R. P. Schumaker, (2009), "Textual analysis of stock market prediction using breaking financial news: The AZFin Text system," ACM Transactions on Information Systems, Vol. 27, PP. 1-19.

[65] M. Mittermayer, (2006), "News CATS: A News Categorization And Trading System," Proceedings of the Sixth International Conference on Data Mining (ICDM'06), PP. 1-5.

[66] A. Soni, (2007), "Prediction of stock price movements based on concept map information," IEEE Symposium on Computational Intelligence in Multicriteria Decision Making, PP. 205-211.

[67] A. Mahajan, L. (2008), "Mining Financial News for Major Events and Their Impacts on the Market," IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Vol. 1, PP. 423 - 426,