International Journal of Ambient Computing and Intelligence
Volume 8 • Issue 2 • April-June 2017

A Comparative Study of Statistical and Rough Computing Models in Predictive Data Analysis

Debi Acharjya, School of Computing Science and Engineering, VIT University, Vellore, India

A. Anitha, School of Information Technology and Engineering, VIT University, Vellore, India

ABSTRACT

Information and technology revolution has brought a radical change in the way data are collected. The data collected is of no use unless some useful information is derived from it. Therefore, it is essential to think of some predictive analysis for analyzing data and to get meaningful information. Much research has been carried out in the direction of predictive data analysis starting from statistical techniques to intelligent computing techniques and further to hybridize computing techniques. The prime objective of this paper is to make a comparative analysis between statistical, rough computing, and hybridized techniques. The comparative analysis is carried out over financial bankruptcy data set of Greek industrial bank ETEVA. It is concluded that rough computing techniques provide better accuracy 88.2% as compared to statistical techniques whereas hybridized computing techniques provides still better accuracy 94.1% as compared to rough computing techniques.

KEYWORDS

Almost Indiscernibility, Correlation, Equivalence Class, Fuzzy Proximity Relation, Fuzzy Relation, Mean Percentile Error, Mean Square Error, Neural Network, Prediction, Regression Analysis, Rough Set

1. INTRODUCTION

Few decades before, computer was a simple device used for doing computations, and calculations in a limited area. But emergence of networking and communication technologies, has replaced the role of computer from stand alone system to distributed systems. Simultaneously, the processing speed is considerably increased. This helps in processing data at a greater speed. At present age, enormous amount of data are exchanged, generated, stored, and manipulated through the internet and through numerous sources. But, what is the need of such huge accumulated data unless we extract or predict some useful information from it. So data analysis, information retrieval, and prediction of decisions for unseen associations is of recent research. Additionally, the branch of data mining concerned about the prediction of future probabilities and trends are referred as predictive analysis. It deals with the variables that can be measured based on other single or multiple factors to predict the decision.

DOI: 10.4018/IJACI.2017040103

In traditional approach of predictive modeling, data are collected, a statistical model is formulated and predictions are made with validating the available data. But statistical methods have its own limitations and cannot produce better prediction when the data contains uncertainty. Further to handle uncertainty in predictive data analysis, many intelligent techniques such as rough set (Pawlak, 1982), rough set on fuzzy approximation space (De, 1999), rough set on intuitionistic fuzzy approximation space (Acharjya & Tripathy, 2009), and hybridization of these concepts with other techniques such as neural network, genetic algorithm, formal concept analysis (Tripathy, Acharjya & Ezhilarasi, 2011), etc., were developed. Moreover, predictive analysis is applied in numerous areas such as health sector, telecommunications, financial services, marketing, actuarial science, travel, pharmaceuticals etc. Our basic objective in writing this paper is to make a comparative study between statistical approach and some of the computational intelligent approach. To show the viability of comparison, financial bankruptcy dataset is used to measure the financial distress of the public firms.

Corporate bankruptcy plays a significant role in the field of finance, for the economic phenomena of a country. Policy makers, investors, managers, consumers, industry shares holders, are the prominent entities for the healthy and successful business world (Cielen, Peters, & Vanhoof, 2004). Business failure is a world-wide problem. To enhance the growth throughout the country, some mechanism should be available to predict the number of firms that may fail due to bankruptcy. Simultaneously, the failure serves as an index for the continuous development and robustness of a country's economy (Min & Jeong, 2009; Zhang & Wu, 2011). The consequences raised by the corporate bankruptcies urge the researchers to carry out research work in this direction. Bankruptcy prediction technique is a vast area of finance and accounting research. The research on developing such prediction models initializes its process by focusing on various classification models to distinguish failed and non-failed firms. Such models are of major importance for the budgetary decision makers, as they serve as early-warning system for the failure probability of a corporate entity. To this end, varied traditional statistical methods are employed for predicting financial distress. As stated earlier, in this paper, we compare the statistical approach with various rough computing techniques, using the data collected from the Greek industrial bank, ETEVA, which finances industrial and commercial firms in Greece (Slowinski & Zopounidis, 1995; Greco, Matarazzo, Pappalardo, & Slowinski, 2005). Furthermore, it will help on forming a economic distress prediction system to provide information to the investors, policy makers, and monitoring organizations.

The rest of the paper is organized as follows: Section 2 provides literature review on the bankruptcy prediction models. Section 3, discuss the foundations of the techniques used for predictive data analysis whereas Section 4 explains about the data organization and the proposed research model using various predictive data analysis techniques followed by Section 5, that depicts an experimental comparative analysis, and the paper is concluded by conclusion in Section 6.

2. LITERATURE REVIEW

Bankruptcy prediction models can be classified into two broad categories: statistical and computational intelligence models. Since 1932, bankruptcy predictions are subject to formal analysis by the world (Fitzpatrick, 1932). The analysis was done based on the financial ratios but not with any statistical methods. Later in 1967, William beaver applied t-test for the evaluation of the financial ratios with a linear variable (Beaver, 1966). Instead of using a single variable, Edward I. Altman used multiple discriminant analysis with the pair-matched samples, along with various other stochastic models such as conditional logit model and probit models (Altman, 1968). However, the practical application of these statistical models are limited by their inherent strict assumptions such as linearity, normality, independence among predictor variables and pre-existing functional forms relating to the significant variable with predictor variable (Hua, Wang, Xu, Zhang & Liang, 2007). But, it is observed that the bankruptcy data contains uncertainty, and it may lead to imperfect knowledge or prediction. Over the past decade, a number of studies have applied computational intelligence techniques to deal with uncertainty. One of the best tools to deal with uncertainty was introduced by Pawlak (1982).

The introduction of rough set theory took the attention of the various researchers to carry out their research towards the prediction process and has been applied to various financial decision analysis problems (Slowinski & Zopounidis, 1995; Siegel, Korvin & Omer, 1995). Since 1930, the researchers are interested in bankruptcy prediction models using various techniques. But, it is fair to say that the research was carried out using rough set theory but not with various abstract rough set models such as rough set on fuzzy approximation space, rough set on intuitionistic fuzzy approximation space etc. (Acharjya, 2015).

Slowinski and Zopounidis (1995) discussed more on financial decision analysis, and used the rough set approach for bankruptcy prediction, with 39 firms as its sample size subject to prediction process. Greeco, Matarazzo and Slowinski (1998) used dominance relation in addition with indiscernibility relation for predicting bankruptcy in Greece companies resulted with 84.90% accuracy. M C Kee (1999) used 150 companies of USA, with the validation sample size of 141 and derived 86 rules using rough set approach. The accuracy attained by this model was 67% for the considered dataset. An extensive international literature search by Dimitras (1999) considered Greece companies for bankruptcy prediction with the sample size of 80. On considering various financial ratios for analysis, using rough set technique, the model attained an accuracy of 76.30% for the valid sample data set. M Y Chen (2011, 2012) compared some traditional statistical methods with support vector machine techniques for the financial bankruptcy prediction. The paper shows that the accuracy increases by using support vector machines than using statistical methods. Cao (2011) predicted the financial distress of Chinese listed companies using rough set theory and support vector machine. It helps in early prediction to make precautions from bankruptcy. Also it stated that rough set model provides better accuracy than the support vector machine model. Xiao (2012) discussed about the financial distress using multiple prediction methods with rough set as preprocessing and Dempster-Shafer evidence theory to Chinese listed companies for business failure prediction. Since the rough set theory deals with vague and uncertainty, the consistent data produced higher accuracy. The advancements in computational intelligence led to the various integrated methods discussed by Cheng (2013). They integrated rough set, k-means clustering, support vector machines, for bankruptcy prediction for analyzing Taiwan database of 132 companies. This research resulted with 81.22% accuracy. It is observed that the prime objective of rough set model in the study of bankruptcy is feature reduction and rule generation. The feature reduction was carried out, such that the power of classification sustains same with the reduced number of attributes.

3. FOUNDATIONS OF PREDICTIVE ANALYSIS TECHNIQUES

Generally, predictive analysis is to forecast the values for the unknown associations. The approaches used to conduct predictive analysis can be broadly categorized into two ways as statistical techniques, and computational intelligent techniques. On statistical technique, regression analysis is commonly used for predictive analysis. Similarly when computational intelligence is considered, rough computing places a major role in predictive analysis. We discuss briefly about these techniques before comparative analysis.

3.1. Regression Analysis

Regression analysis is extensively used for prediction process, since it has a considerable amount of features that overlap in the field of predictive analysis. It is a statistical technique for finding the relationship between a dependent (decision) variable with one or more independent (conditional) variables as they are represented in any information system. It has two uses in scientific literature such as prediction, including classification, and interpretation. Also, regression analysis can be used to identify the casual relationship between the decision and conditional attribute values for analyzing the prediction process. In a causal analysis, the independent variables are regarded as causes of the dependent variable. However, this can lead to illusions or mendacious relationships. There are different kinds of regression analysis techniques used to make predictions. However, these techniques are mostly based on three metrics. These are the number of independent variables, type of dependent variables and the regression line (Lindley, 1987).

Predictive analysis is to be carried out in real life application, where the relationship between one dependent (decision) variable with '*n*' independent (conditional) variables is to be identified. Multi variable regression analysis is the powerful method to compute the relationship between the objects and attributes with explanatory attribute values. Formally, the multiple regression model is represented as equation of d on $a_1, a_2, a_3, \cdots \cdots, a_n$ is given as:

$$d = b_o + b_1 a_1 + b_2 a_2 + \cdots\cdots + b_n a_n \tag{1}$$

where b_o is the interception and $b_i; 1 \leq i \leq n$ are the regression coefficients that represents the coefficient for the independent variables (attributes) $a_1, a_2, a_3, \cdots \cdots, a_n$. The regression coefficients b_i, influence the increase or decrease of the predictive variable 'd'.

3.2. Fundamentals of Rough Set

Rough set developed by Z. Pawlak (1982) is a mathematical tool to deal with vagueness and uncertainty. It was constructed based on the assumption that every object is associated with some data and information. Classification of objects is carried out with the help of an indiscernible relation. The indiscernibility relation partitions the objects into disjoint subsets of equivalence classes. The equivalence class generated by the indiscernible relation is known as concepts.

Let $U \ (\neq \varphi)$ be a finite set of objects called the universe. Let $R \subseteq (U \times U)$ is an equivalence relation on *U*. The equivalence relation *R* partitions the set *U* into disjoint subsets of equivalence classes. Elements of same equivalence class are said to be indistinguishable. Thus (*U*, *R*) is called an approximation space. Given a target set X of objects, we can characterize X by a pair of lower and upper approximations. The lower and upper approximations are defined below where $\underline{R}X$ represents the lower approximation and $\overline{R}X$ represents the upper approximation:

$$\underline{R}X = \bigcup \{Y \in U / R : Y \subseteq X\} \tag{2}$$

$$\overline{R}X = \bigcup \{Y \in U / R : Y \cap X \neq \varphi\} \tag{3}$$

The R-boundary of X, $BN_R(X)$ is given by $BN_R(X) = \overline{R}X - \underline{R}X$. We say X is rough with respect to R if and only if $\overline{R}X \neq \underline{R}X$ equivalently $BN_R(X) \neq \varphi$. Similarly, X is said to be $R-$ definable if and only if $\overline{R}X = \underline{R}X$ or $BN_R(X) = \varphi$.

An information system is a quadruple $I = (U, A, V, f)$, where $U = \{x_1, x_2, \cdots, x_n\}$ is a non-empty finite set of objects called the universe and $A = \{a_1, a_2, \cdots, a_n\}$ is a non-empty finite set of attributes. For every $a \in A$; V_a is the set of values that attribute a may take. Also, $V = \bigcup_{a \in A} V_a$. In addition for every $a \in A$; $f_a : U \to V_a$ is the information function. Further, if $A = (C \cup D)$, where C is the set of conditional attributes and D is the decision, we call the information system as decision system.

Let us explain the lower and upper approximation using a sample information system as shown in Table 1. The information system provides the information about the type of glass based on various components. Here $U = \{x_1, x_2, x_3, x_4, x_5, x_6\}$ be the set of objects; $A = \{$ sodium (a_1), magnesium (a_2), aluminum (a_3), silicon (a_4)$\}$ represents the conditional attributes, and *a* is the decision attribute.

On employing equivalence relation on Table 1, the equivalence classes thus generated are given as follows:

$$U / (A = \{a_1\}) = \{\{x_1, x_4\}, \{x_2\}, \{x_3\}, \{x_5, x_6\}\}$$

$$U / (A = \{a_2\}) = \{\{x_1, x_4\}, \{x_2, x_5\}, \{x_3\}, \{x_6\}\}$$

$$U / (A = \{a_3\}) = \{\{x_1, x_4\}, \{x_2, x_5\}, \{x_3\}, \{x_6\}\}$$

$$U / (A = \{a_4\}) = \{\{x_1, x_4\}, \{x_2, x_6\}, \{x_3\}, \{x_5\}\}$$

$$U / (A = \{d\}) = \{\{x_1, x_5\}, \{x_2, x_3, x_4\}, \{x_6\}\}$$

Furthermore, the objects x_1 and x_4, are indiscernible according to the conditional attributes, but having different decisions. These types of objects are called conflicting objects, and produce uncertain rules. Therefore, it is essential to take care while analyzing the information system.

3.3. Rough Set on Fuzzy Approximation Space

The concept of rough set which is based on indiscernible relation is generally employed over an information system. But in many real-life applications, the attribute values are almost indiscernible rather exactly indiscernible. For example, in Table 1, the attribute values of the objects x_2 and x_3 for sodium, *i.e.*, $72 \cdot 99$ and $72 \cdot 98$ are almost indiscernible. To handle such almost indiscernible cases, the equivalence relation is generalized to fuzzy proximity relation and the concept of rough set on fuzzy approximation space (RSFAS) was introduced (De, 1999). It has thoroughly studied by Acharjya in his papers (Acharjya & Tripathy, 2008). The RSFAS thus generated have better generality and applicability than the basic ones. Let us discuss briefly the foundations of RSFAS.

Let U be a universe of discourse. A fuzzy relation on U is a fuzzy subset of ($U \times U$). A fuzzy relation R on U is said to be fuzzy proximity relation if $\mu_R(x, x) = 1, \forall x \in U$ and

Table 1. Sample information system

Objects	a_1	a_2	a_3	a_4	d
x_1	70.65	3.54	1.31	73.99	1
x_2	72.99	3.98	0.89	72.86	2
x_3	72.98	3.55	0.89	70.36	2
x_4	70.65	3.54	1.31	73.99	2
x_5	75.24	3.98	1.43	77.25	1
x_6	75.24	8.21	2.11	72.86	3

$\mu_R(x_i, x_j) = \mu_R(x_j, x_i)$, $\forall\ x_i, x_j \in U$ where μ represents the membership between two objects. Then for a given criterion of proximity $\alpha \in [0,1]$, two objects x_i and x_j are $\alpha -$ similar with respect to R if $\mu_R(x_i, x_j) \geq \alpha$ and can be written as $(x_i, x_j) \in R_\alpha$ or $x_i R_\alpha x_j$. Two objects x_i and x_j are said to be $\alpha -$ identical with respect to R, if either x_i and x_j are $\alpha -$ similar or transitively $\alpha -$ similar. The $\alpha -$ equivalence classes thus generated is denoted as R_α^*. The pair $K = (U, R(\alpha))$ is said to be fuzzy approximation space. Let $X \subseteq U$. The lower and upper approximation of X in the generated approximation space K is defined as follows:

$$\underline{X}_\alpha = \bigcup\{Y \in R_\alpha^* \mid Y \subseteq X\} \tag{4}$$

$$\overline{X}_\alpha = \bigcup\{Y \in R_\alpha^* \mid Y \cap X \neq \varphi\} \tag{5}$$

The boundary $BN_R^\alpha(X)$ is defined as $BN_R^\alpha(X) = \overline{X}_\alpha - \underline{X}_\alpha$. The target set X is $\alpha -$ rough if $\overline{X}_\alpha \neq \underline{X}_\alpha$ or $BN_R^\alpha(X) \neq \varphi$. Similarly, the target set X is $\alpha -$ crisp if $\overline{X}_\alpha = \underline{X}_\alpha$ or $BN_R^\alpha = \varphi$.

3.4. Rough Set on Intuitionistic Fuzzy Approximation Space

The concept of RSFAS was introduced to handle almost indiscernibility. The notion uses fuzzy proximity relation to check the $\alpha -$ indiscernibility between two objects. But many real-life applications contain hesitation. To deal with hesitation, the notion of rough set on intuitionistic fuzzy approximation space (RSIFAS) was introduced (Acharjya & Tripathy, 2009). The introduced concept replaces the fuzzy proximity relation with intuitionistic fuzzy proximity relation. Further, it is also established that RSIFAS is a better model as compared to RSFAS (Acharjya, 2009). In order to complete the article, we briefly state the notions and notations used in RSIFAS.

Let U be a universe of discourse. An intuitionistic fuzzy relation on U is an intuitionistic fuzzy subset of $(U \times U)$. An intuitionistic fuzzy relation R is said to be an intuitionistic fuzzy proximity relation if $\mu_R(x, x) = 1$ and $\nu_R(x, x) = 0$ $\mu\left(x, x\right) = 1$ and $v_R\left(x, x\right) = 0$ for all $x \in U$; $\mu_R(x_i, x_j) = \mu_R(x_j, x_i)$, $\nu_R(x_i, x_j) = \nu_R(x_j, x_i)$ for all $x_i, x_j \in U$. Here we use the standard symbol μ and ν for membership and non-membership between two objects. Let R be an intuitionistic fuzzy proximity relation on U, then for any $(\alpha, \beta) \in J$, the $(\alpha, \beta) - cut$, $R_{\alpha,\beta}$ is given by $R_{\alpha,\beta} = \{(x_i, x_j) \mid \mu_R(x_i, x_j) \geq \alpha$ and $\nu_R(x_i, x_j) \leq \beta\}$ where $J = \{(\alpha, \beta) \mid \alpha, \beta \in [0,1]$ and $0 \leq \alpha + \beta \leq 1\}$. We say that two objects x_i and x_j are $(\alpha, \beta) -$ similar with respect to R if $(x_i, x_j) \in R_{\alpha,\beta}$ $\left(x, y\right) \in R_{a,\beta}$ and we write $x_i R_{\alpha,\beta} x_j$. Two objects x_i and x_j are $(\alpha, \beta) -$ identical with respect to R for $(\alpha, \beta) \in J$ written as $x_i R(\alpha, \beta) x_j$ if and only if x_i and x_j are $(\alpha, \beta) -$ similar or transitively $(\alpha, \beta) -$ similar. The almost equivalence, $(\alpha, \beta) -$ equivalence, classes thus generated is denoted as $R_{\alpha,\beta}^*$. The pair $(U, R(\alpha, \beta))$ is called an intuitionistic fuzzy approximation space. The RSIFAS on the target set X in the generalized approximation space $(U, R(\alpha, \beta))$ is denoted by $(\underline{X}_{\alpha,\beta}, \overline{X}_{\alpha,\beta})$ where:

$$\underline{X}_{\alpha,\beta} = \bigcup\{Y \mid Y \in R_{\alpha,\beta}^*\ and\ Y \subseteq X\} \tag{6}$$

$$\overline{X}_{\alpha,\beta} = \bigcup\{Y \mid Y \in R_{\alpha,\beta}^*\ and\ Y \cap X \neq \varphi\} \tag{7}$$

The $(\alpha,\beta)-$boundary of X with respect to R denoted by $BN_R^{\alpha,\beta}(X)$ as $BN_R^{\alpha,\beta} = \overline{X}_{\alpha,\beta} - \underline{X}_{\alpha,\beta}$. The target set X is $(\alpha,\beta)-$discernible if and only if $\overline{X}_{\alpha,\beta} = \underline{X}_{\alpha,\beta}$ or $BN_R^{\alpha,\beta}(X) = \varphi$. Similarly, X is said to be $(\alpha,\beta)-$rough if and only if $\overline{X}_{\alpha,\beta} \neq \underline{X}_{\alpha,\beta}$ or $BN_R^{\alpha,\beta} \neq \varphi$.

3.5. Hybridization Techniques

The idea of hybridizing artificial neural network with rough computing model is to eradicate indeterminacy rules generated by rough set. These indeterminacy rules are generally due to conflict objects. To handle the conflict objects, the solution is to find the best strategy for the selected objects using specific selection methods, and to conduct experiments. This ground for the hybridization of rough computing models with neural network. In hybrid system, rules delivered by the rough computing model were used to oversee the learning process of the neural network and to correct the output errors. It is found that the hybridized of rough set and neural network produces better accuracy (Anitha & Acharjya, 2015). Furthermore, rough set on intuitionistic fuzzy approximation space is a better model as compared to rough set on fuzzy approximation space. At the same time, rough set on fuzzy approximation space is a better model than rough set (Acharjya, 2009). Therefore, for comparative analysis we have considered hybridization of rough set on intuitionistic fuzzy approximation space and neural network. The proposed model is presented in Figure 1.

Before we process data at initial stage, a sequence of cleaning task such as abstracting noise, consistency check and data cleaning are carried out to ascertain that the data is as precise as possible. The target data is processed using intuitionistic fuzzy proximity relation to obtain almost indiscernibility classes. This condenses the quantitative information system to qualitative information system. Further rules are generated and passed to neural network phase for further processing. Anitha and Acharjya (2015) illustrated and discussed in details its computational procedure.

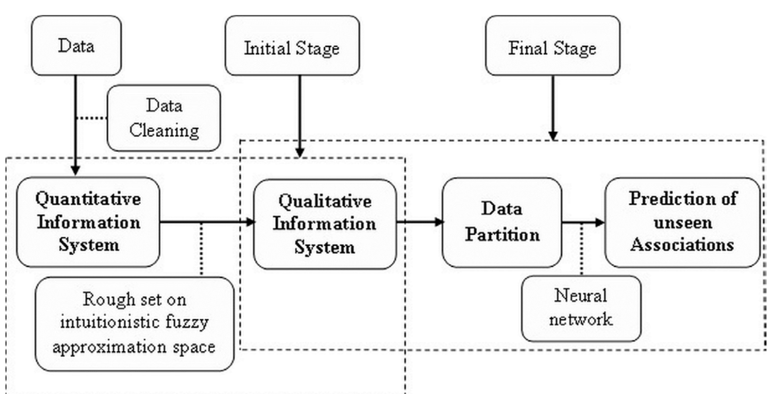
4. RESEARCH METHODOLOGY AND PREDICTIVE ANALYSIS

This section analyses the classes of bankruptcy risk using statistical, rough computing, and hybridized computing techniques. Before analyzing data, we briefly discuss about the data in the following section.

4.1. Data

To illustrate the comparative analysis, we use data provided by Greek industrial bank ETEVA (Slowinski & Zopounidis, 1995; Greco, Matarazzo, Pappalardo, & Slowinski, 2005). A data set composed of 39 firms has been chosen for the comparative study. The firms are classified into three

Figure 1. Abstract view of proposed model



classes of bankruptcy risk such as acceptable (1), uncertain (2) and non-acceptable (3). The firm was evaluated based on twelve financial ratios a_1 to a_{12}, represented as the conditional attributes. The first six conditional attributes are cardinal (financial ratios) and the last six are ordinal. On considering the length of the paper a sample data set of 15 firms is presented in Table 2:

$a_1 =$ Earnings before interests and taxes: Total assets
$a_2 =$ Net income: Net worth
$a_3 =$ Total liabilities: Total assets
$a_4 =$ Total liabilities: Cash flow
$a_5 =$ Interest expenses: Sales
$a_6 =$ General and administrative expense: Sales
$a_7 =$ Manager's work experience (very low = 1, low = 2, medium = 3, high = 4, very high = 5)
$a_8 =$ Firm's market niche/position (bad = 1, rather bad = 2, medium = 3, good = 4, very good = 5)
$a_9 =$ Technical structure facilities (bad = 1, rather bad = 2, medium = 3, good = 4, very good = 5)
$a_{10} =$ Organization personnel (bad = 1, rather bad = 2, medium = 3, good = 4, very good = 5)
$a_{11} =$ Special competitive advantage of firms (low = 1, medium = 2, high = 3, very high = 4)
$a_{12} =$ Market flexibility (very low = 1, low = 2, medium = 3, high = 4, very high = 5)

4.2. Statistical Descriptive Data Analysis

The original data collected is employed with fundamental descriptive statistical analysis tool and its results are depicted in Table 3. The positive skew ness results show many values at the low end, and few at the high end. The negative skew ness results show many values at the high end and few at the low end. It is seen from Table 3 that the skew ness values are positive and the kurtosis are significant.

4.3. Correlation Analysis

Correlation analysis measures the relationship between the conditional attributes and the decision attribute values. The correlation coefficient is a measure of linear association between two variables

Table 2. Sample data set

Firm	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	d
x_1	16.4	14.5	59.8	2.5	7.5	5.2	5	3	5	4	2	4	1
x_2	35.8	67	64.9	1.7	2.1	4.5	5	4	5	5	4	5	1
x_3	20.6	61.7	75.7	3.6	3.6	8	5	3	5	5	3	5	1
x_4	29.9	44	57.8	1.8	1.7	2.5	5	4	4	5	3	3	1
x_5	25.7	29.7	46.8	1.7	4.6	3.7	4	2	4	3	1	4	1
x_6	16.7	13.1	73.5	7.1	11.9	4.1	2	2	4	4	2	3	2
x_7	14.6	9.7	59.5	5.8	6.7	5.6	2	2	4	3	2	3	2
x_8	5.1	4.9	28.9	4.3	2.5	46	2	2	3	4	1	4	2
x_9	24.4	22.3	32.8	1.4	3.3	5	2	3	4	4	2	2	2
x_{10}	29.7	8.6	41.8	1.6	5.2	6.4	2	3	4	4	2	3	2
x_{11}	7.3	64.5	67.5	2.2	30.1	8.7	3	3	4	4	2	3	3
x_{12}	13.9	3.3	78.7	25.5	14.7	10.1	2	3	4	3	3	4	3
x_{13}	13.3	31.1	63	10	21.2	23.1	2	2	3	3	1	2	3
x_{14}	4.8	3.3	71.9	34.6	8.6	11.6	2	1	4	4	2	3	3
x_{15}	0.1	9.6	42.5	20	12.9	12.4	1	2	3	3	1	3	3

Table 3. Results of descriptive statistical analysis

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	d
Mean	15.7	19.1	56.7	7.0	7.9	7.9	3.4	2.5	3.9	3.7	2.0	3.5	1.7
Median	14.8	13.5	59.5	4.2	5.6	5.6	4.0	2.0	4.0	4.0	2.0	3.0	1.0
Std Dev.	7.7	16.9	15.8	8.0	6.1	7.4	1.3	0.8	0.7	0.8	0.8	0.8	0.8
Skew ness	0.4	1.5	0.5	2.4	1.7	3.9	0.3	0.4	1.8	1.3	0.3	0.1	0.6
Kurtosis	0.2	2.0	0.4	5.5	3.6	18.9	1.2	0.3	7.3	2.8	0.8	0.4	1.3
Std Error	2.6	2.3	9.5	0.4	1.2	0.8	0.8	0.4	0.8	0.6	0.3	0.6	0.1

and always lies between -1 and 1. A correlation coefficient of 1 indicates that two variables are perfectly related in a positive linear sense, whereas two variables are perfectly related in a negative linear sense if the correlation coefficient is -1. Correlation coefficient of 0 indicates that there is no linear relationship between the two variables. The correlation analysis is presented in Table 4.

From Table 4, it is clearly seen that the decision attribute values are correlated with the attributes $a_5, a_6, a_7, a_8, a_9, a_{11}$, and a_{12} where decision attribute value at its $P - value$ significance is 0.001.

4.4. Multi Variable Regression Data Analysis

Regression analysis is used to gain a better relationship between conditional and decision attribute. Regression analysis using R language is employed to find the coefficient of determination (accuracy) of the prediction model as listed in Table 5. The regression Equation (1) as discussed earlier in section 3.1 is given as:

$$d = 2 \cdot 357 + 0 \cdot 0021\, a_1 + 0 \cdot 0057\, a_2 - 0 \cdot 0064\, a_3 + 0 \cdot 016\, a_4 + 0 \cdot 08233\, a_5 + 0.018268\, a_6 - 0 \cdot 36547\, a_7 - 0 \cdot 10989\, a_8 - 0 \cdot 0145\, a_9 + 0 \cdot 07106\, a_{10} + 0 \cdot 26024\, a_{11} + 0.13029\, a_{12} \quad (8)$$

From Table 4 and Table 5, it is clear that the relationship between the decision variable and the attributes $a_5, a_6, a_7, a_8, a_9, a_{11}$ and a_{12} are statistically significant at $P - value < 0.001$.

Table 4. Correlation analysis

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	d
a_1	1.00												
a_2	0.764	1.00											
a_3	0.050	0.05	1.00										
a_4	0.022	0.12	0.13	1.00									
a_5	0.495*	0.53*	0.27*	0.27*	1.00								
a_6	0.473*	0.53*	0.11	0.05	0.18	1.00							
a_7	0.406*	0.48*	0.28*	0.27*	0.63*	0.42*	1.00						
a_8	0.485*	0.36	0.03	0.28*	0.15	0.17	0.4	1.00					
a_9	0.418*	0.38	0.01	0.13	0.37	0.39*	0.64	0.43	1.00				
a_{10}	0.370	0.33	0.00	0.15	0.28	0.12	0.52	0.14	0.65	1.00			
a_{11}	0.393*	0.59*	0.35	0.05	0.31	0.28	0.31	0.46	0.54*	0.38*	1.00		
a_{12}	0.350	0.50*	0.12	0.28*	0.64*	0.16	0.79*	0.31*	0.52*	0.50	0.19*	1.00	
d	0.433	0.54	0.28	0.27	0.64	0.42	0.83	0.31	0.54	0.38	0.19	0.49	1.00

* indicates the $P - value$ significance at 0.001.

Table 5. Results of multivariable regression analysis

Independent Variables	Coefficients	Std Error	t	$P-value$	Summary
Constant	2.357018	0.563511	4.183	0.00029	Coefficient of determination R^2 = 0.837 Adjusted R^2 = 0.761
a_1	0.002119	0.013672	0.155	0.878044	
a_2	0.005721	0.006389	0.895	0.378749	
a_3	- 0.006451	0.005853	-1.102	0.280516	
a_4	0.016015	0.008151	1.965	0.060227	
a_5	0.082332	0.020222	4.071	0.000388	
a_6	0.018268	0.011962	1.527	0.138793	
a_7	-0.36547	0.093252	-3.919	0.000577	
a_8	-0.109891	0.115971	-0.948	0.35207	
a_9	-0.014504	0.155364	-0.093	0.92634	
a_{10}	0.071062	0.114947	0.618	0.541808	
a_{11}	0.260248	0.121297	2.146	0.04142	
a_{12}	0.130297	0.124036	-1.05	0.303165	

Adjusted $R^2 = 0.761$ shows that the significant relationship among the attributes with accuracy of 83.7% of prediction.

4.5. Rough Set Approach

The information system as shown in Table 2 contains 39 objects with 12 conditional attribute and one decision attribute. The basic objective is to derive rules which could be useful in obtaining decisions. But, in general rough set cannot be applied directly to quantitative information system. Therefore, in order to apply rough set to bankruptcy data, we must normalize the quantitative data present in the information system. The normalization is coded based on domain expert for the cardinal conditional attributes present in dataset. The normalization of the attribute values is presented in Table 6.

On employing normalization, the bankruptcy data set is reduced to qualitative information system. The coded information system of 39 objects is considered as the decision table for predictive data analysis. A sample coded dataset obtained from Table 2 is given in Table 7. Additionally, the reduced information system is further divided into two parts such as training data of 22 objects (55%) and testing data of 17 objects (45%). Further, on employing rough set data analysis on training data set we get the decision rules. The decision rules obtained are presented in Table 8.

Table 6. Normalized information table

Attributes	Normalization	Classification	Attributes	Normalization	Classification
a_1	< 11	Low (1)	a_2	< 10	Low (1)
	11 – 19	Medium (2)		10 – 24	Medium (2)
	19 – 40	High (3)		24 – 70	High (3)
a_3	15 – 46	Low (1)	a_4	0 – 2	Low (1)
	46 – 63	Medium (2)		2 – 9	Medium (2)
	63 – 90	High (3)		9 – 35	High (3)
a_5	1 – 5	Low (1)	a_6	1 – 5	Low (1)
	5 – 11	Medium (2)		5 – 11	Medium (2)
	11 – 30	High (3)		11 – 23	High (3)
	> 30	Very high (4)		23 – 50	Very high (4)

Table 7. Reduced sample dataset

Firm	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	d
x_1	2	2	2	2	2	2	5	3	5	4	2	4	1
x_2	3	3	3	1	1	1	5	4	5	5	4	5	1
x_3	3	3	3	2	1	2	5	3	5	5	3	5	1
x_4	3	3	2	1	1	1	5	4	4	5	3	3	1
x_5	3	3	2	1	1	1	4	2	4	3	1	4	1
x_6	2	2	3	2	3	1	2	2	4	4	2	3	2
x_7	2	1	2	2	2	2	2	2	4	3	2	3	2
x_8	1	1	1	2	1	4	2	3	4	4	2	2	2
x_9	3	2	1	1	1	2	2	3	4	4	2	3	2
x_{10}	3	1	1	1	2	2	3	3	4	4	2	3	3
x_{11}	1	3	3	2	4	2	2	3	4	3	3	4	3
x_{12}	2	1	3	3	3	2	2	2	3	3	1	2	3
x_{13}	2	3	2	3	3	4	2	1	4	4	2	3	3
x_{14}	1	1	3	3	2	3	1	2	3	3	1	3	3
x_{15}	1	1	1	3	3	3	1	1	1	1	1	2	3

Table 8. Decision rules generated by RS approach

Rules	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	d	Supporting Objects
[1]	×	×	×	×	×	×	4	×	×	×	×	×	1	$x_5, x_6, x_7, x_8,$ x_9, x_{11}, x_{12}
[2]	×	×	×	×	×	×	5	×	×	×	×	×	1	$x_1, x_2, x_3, x_4, x_{10}$
[3]	×	×	2	×	×	×	2	×	×	×	×	×	2	x_{15}, x_{16}, x_{17}
[4]	×	3	×	×	×	×	×	3	×	×	3	×	2	x_{13}, x_{14}
[5]	×	×	×	4	×	×	×	×	×	×	×	×	3	$x_{19}, x_{20}, x_{21}, x_{22}$
[6]	×	×	×	5	×	×	×	×	×	×	×	×	3	x_{18}

4.6. RSIFAS Approach

Rough set on intuitionistic fuzzy approximation space (RSIFAS) reduces the quantitative information system into qualitative information system without employing any normalization on cardinal attributes. The almost equivalence classes or (α, β) equivalence classes are obtained on employing intuitionistic fuzzy proximity relation, where α is the measure of belongingness and β is the measure of non-belongingness. The membership and non-membership relation have been premeditated such that the sum of their values lies between [0, 1] and additionally these functions must be symmetric. The almost similarity between two objects x_i and x_j is defined as below where μ represents the membership and ν represents the non-membership between two objects x_i and x_j; and $V_{a_i}^{x_i}$ is the value of the object x_i for the attribute a_i (Acharjya, 2011):

$$\mu_R(x_i, x_j) = 1 - \frac{| V_{a_i}^{x_i} - V_{a_i}^{x_j} |}{Max\ Value\ (a_i)} \tag{9}$$

$$\nu_R(x_i, x_j) \;=\; \frac{|\, V_{a_i}^{x_i} - V_{a_i}^{x_j} \,|}{2 \times Max\ Value\ (a_i)} \tag{10}$$

On using intuitionistic fuzzy proximity relation defined in Equations (9) and (10) we get the $(\alpha, \beta)-$ equivalence classes for all cardinal attributes. Considering the length of the paper the intuitionistic fuzzy proximity relation for the attribute a_1 of 15 firms is presented in Figure 2.

On considering $\alpha \geq 0.95$ and $\beta \leq 0.03$, the $(\alpha,\ \beta)$ equivalence class obtained for attribute a_1 is given as follows:

$$U \,/\, R^{a_1}_{(\alpha,\beta)} = \{\{x_2\}, \{x_4, x_{10}\}, \{x_1, x_3, x_5, x_6, x_7, x_{8,} x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}\}, \{x_{15}\}\}$$

Similar computation is carried out for all 39 objects and the $(\alpha,\ \beta)$-equivalence class obtained for all attributes is furnished as follows:

$$U \,/\, R^{a_1}_{(\alpha,\beta)} = \{\{x_2\}, \{x_7, x_{30}\}, \{x_{38}\}, \{x_1, x_3, x_4, x_5, x_6, x_{8,} x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}, x_{19}, x_{20}, x_{21}, x_{22}, x_{23}, x_{24}, x_{25}, x_{26}, x_{27}, x_{28}, x_{29}, x_{31}, x_{32}, x_{33}, x_{34}, x_{35}, x_{36}, x_{37}, x_{39}\}\}$$

Figure 2. Intuitionistic fuzzy proximity for attribute (a_1)

$R^{a_1}_{(\alpha,\beta)}$	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}
x_1	1, 0	0.46, 0.27	0.88, 0.06	0.62, 0.19	0.74, 0.13	0.99, 0.00	0.95, 0.03	0.68, 0.16	0.78, 0.11	0.63, 0.19	0.75, 0.13	0.93, 0.03	0.91, 0.04	0.68, 0.16	0.54, 0.23
x_2	0.46, 0.27	1, 0	0.58, 0.21	0.84, 0.08	0.72, 0.14	0.47, 0.27	0.41, 0.30	0.14, 0.43	0.68, 0.16	0.83, 0.09	0.20, 0.40	0.39, 0.31	0.37, 0.31	0.13, 0.43	0.00, 0.50
x_3	0.88, 0.06	0.58, 0.21	1, 0	0.74, 0.13	0.86, 0.07	0.89, 0.05	0.83, 0.08	0.57, 0.22	0.89, 0.05	0.75, 0.13	0.63, 0.19	0.81, 0.09	0.80, 0.10	0.56, 0.22	0.43, 0.29
x_4	0.62, 0.19	0.84, 0.08	0.74, 0.13	1, 0	0.88, 0.06	0.63, 0.18	0.57, 0.21	0.31, 0.35	0.85, 0.08	0.99, 0.00	0.37, 0.32	0.55, 0.22	0.54, 0.23	0.30, 0.35	0.17, 0.42
x_5	0.74, 0.13	0.72, 0.14	0.86, 0.07	0.88, 0.06	1, 0	0.75, 0.13	0.69, 0.16	0.42, 0.29	0.96, 0.02	0.89, 0.06	0.49, 0.26	0.67, 0.16	0.65, 0.17	0.42, 0.29	0.28, 0.36
x_6	0.99, 0.00	0.47, 0.27	0.89, 0.05	0.63, 0.18	0.75, 0.13	1, 0	0.94, 0.03	0.68, 0.16	0.78, 0.11	0.64, 0.18	0.74, 0.13	0.92, 0.04	0.91, 0.05	0.67, 0.17	0.54, 0.23
x_7	0.95, 0.03	0.41, 0.30	0.83, 0.08	0.57, 0.21	0.69, 0.16	0.94, 0.03	1, 0	0.73, 0.13	0.73, 0.14	0.58, 0.21	0.80, 0.10	0.98, 0.01	0.96, 0.02	0.73, 0.14	0.59, 0.20
x_8	0.68, 0.16	0.14, 0.43	0.57, 0.22	0.31, 0.35	0.42, 0.29	0.68, 0.16	0.73, 0.13	1, 0	0.46, 0.27	0.31, 0.34	0.94, 0.03	0.75, 0.12	0.77, 0.11	0.99, 0.00	0.86, 0.07
x_9	0.78, 0.11	0.68, 0.16	0.89, 0.05	0.85, 0.08	0.96, 0.02	0.78, 0.11	0.73, 0.14	0.46, 0.27	1, 0	0.85, 0.07	0.52, 0.24	0.71, 0.15	0.69, 0.16	0.45, 0.27	0.32, 0.34
x_{10}	0.63, 0.19	0.83, 0.09	0.75, 0.13	0.99, 0.00	0.89, 0.06	0.64, 0.18	0.58, 0.21	0.31, 0.34	0.85, 0.07	1, 0	0.37, 0.31	0.56, 0.22	0.54, 0.23	0.30, 0.35	0.17, 0.41
x_{11}	0.75, 0.13	0.20, 0.40	0.63, 0.19	0.37, 0.32	0.49, 0.26	0.74, 0.13	0.80, 0.10	0.94, 0.03	0.52, 0.24	0.37, 0.31	1, 0	0.82, 0.09	0.83, 0.08	0.93, 0.03	0.80, 0.10
x_{12}	0.93, 0.03	0.39, 0.31	0.81, 0.09	0.55, 0.22	0.67, 0.16	0.92, 0.04	0.98, 0.01	0.75, 0.12	0.71, 0.15	0.56, 0.22	0.82, 0.09	1, 0	0.98, 0.01	0.75, 0.13	0.61, 0.19
x_{13}	0.91, 0.04	0.37, 0.31	0.80, 0.10	0.54, 0.23	0.65, 0.17	0.91, 0.05	0.96, 0.02	0.77, 0.11	0.69, 0.16	0.54, 0.23	0.83, 0.08	0.98, 0.01	1, 0	0.76, 0.12	0.63, 0.18
x_{14}	0.68, 0.16	0.13, 0.43	0.56, 0.22	0.30, 0.35	0.42, 0.29	0.67, 0.17	0.73, 0.14	0.99, 0.00	0.45, 0.27	0.30, 0.35	0.93, 0.03	0.75, 0.13	0.76, 0.12	1, 0	0.87, 0.07
x_{15}	0.54, 0.23	0.00, 0.50	0.43, 0.29	0.17, 0.42	0.28, 0.36	0.54, 0.23	0.59, 0.20	0.86, 0.07	0.32, 0.34	0.17, 0.41	0.80, 0.10	0.61, 0.19	0.63, 0.18	0.87, 0.07	1, 0

$$U / R^{a_2}_{(\alpha,\beta)} = \{\{x_2, x_3, x_{31}\}, \{x_7\}, \{x_1, x_4, x_5, x_6, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{19}, x_{20}, x_{21}, x_{22}, x_{23}, x_{24}, x_{25}, x_{26}, x_{27}, x_{28}, x_{29}, x_{30}, x_{31}, x_{32}, x_{33}, x_{34}, x_{35}, x_{36}, x_{37}, x_{38}, x_{39}\}\}$$

$$U / R^{a_3}_{(\alpha,\beta)} = \{\{x_{12}\}, \{x_8, x_{16}, x_{28}, x_{29}\}, \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_9, x_{10,} x_{11}, x_{13}, x_{14}, x_{15}, x_{17}, x_{18}, x_{19}, x_{20}, x_{21}, x_{22}, x_{23}, x_{24}, x_{25}, x_{26}, x_{27}, x_{30}, x_{31}, x_{32}, x_{33}, x_{34}, x_{35}, x_{36}, x_{37}, x_{38}, x_{39}\}\}$$

$$U / R^{a_4}_{(\alpha,\beta)} = \{\{x_{24}, x_{37}\}, \{x_{34}\}, \{x_{38}\}, \{x_{14}, x_{21}, x_{33}, x_{35}, x_{39}\}, \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{15}, x_{16}, x_{17}, x_{18}, x_{19}, x_{20}, x_{22}, x_{23}, x_{25}, x_{26}, x_{27}, x_{28}, x_{29}, x_{30}, x_{31}, x_{32}, x_{36}\}\}$$

$$U / R^{a_5}_{(\alpha,\beta)} = \{\{x_{31}\}, \{x_{35}\}, \{x_{18}, x_{23}, x_{24}, x_{26}, x_{32}, x_{33}, x_{34}, x_{38}\}, \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{19}, x_{20}, x_{21}, x_{22}, x_{25}, x_{27}, x_{28}, x_{29}, x_{30}, x_{36}, x_{37}, x_{39}\}\}$$

$$U / R^{a_6}_{(\alpha,\beta)} = \{\{x_{28}\}, \{x_{35}\}, \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}, x_{19}, x_{20}, x_{21}, x_{22}, x_{23}, x_{24}, x_{25}, x_{26}, x_{27}, x_{29}, x_{30}, x_{31}, x_{32}, x_{33}, x_{34}, x_{36}, x_{37}, x_{38}, x_{39}\}\}$$

From the above analysis, it is observed that the attribute values of a_1 has classified into four categories. Let the categories are High, Medium, Low, Very low. Similarly, the attributes a_2, a_3, and a_6 have classified into three categories, such as High, Medium, and Low. The attributes a_4 and a_5 have classified into five categories such as Very high, High, Medium, Low and Very low. To make our analysis simple, we assign 1, 2, 3, 4, and 5 to the categories Very low, Low, Medium, High, and Very high respectively. However, these values are optional and do not affect our analysis. This reduces quantitative information system to qualitative information system. The reduced qualitative information system is further divided arbitrarily into training (55%) of 22 firms and testing (45%) of 17 firms. The reduced sample qualitative information system of Table 2 is presented in Table 9.

Further, on employing rule generation algorithm () on training data set of 22 objects we get the decision rules. The decision rules obtained are presented in Table 10.

4.7. RSFAS Approach

Rough set on intuitionistic fuzzy approximation space reduces to rough set on fuzzy approximation space if there is no hesitation. Therefore, on neglecting non-membership values we will get fuzzy proximity relation. On considering $\alpha \geq 0.95$, the α- equivalence class obtained for attribute a_1 from Figure 2 is given as follows:

$$U / R^{a_1}_{\alpha} = \{\{x_2\}, \{x_4, x_{10}\}, \{x_1, x_3, x_5, x_6, x_7, x_8, x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}\}, \{x_{15}\}\}$$

Similar computation is carried out for all 39 objects and the -equivalence class obtained for all attributes is furnished below:

$$U / R^{a_1}_{\alpha} = \{\{x_2\}, \{x_7, x_{30}\}, \{x_{24,} x_{36}\}, \{x_{28}, x_{37}\}, \{x_{38}\}, \{x_1, x_3, x_4, x_5, x_6, x_{8,} x_9, x_{10}, x_{11}, x_{12}, x_{13}, x_{14}, x_{15}, x_{16}, x_{17}, x_{18}, x_{19}, x_{20}, x_{21}, x_{22}, x_{23}, x_{25}, x_{26}, x_{27}, x_{29}, x_{31}, x_{32}, x_{33}, x_{34}, x_{35}, x_{39}\}\}$$

Table 9. Qualitative sample information system after employing RSIFAS approach

Firms	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	d
x_1	3	1	3	1	1	2	5	3	5	4	2	4	1
x_2	4	3	3	1	1	1	5	4	5	5	4	5	1
x_3	3	3	3	1	1	1	5	3	5	5	3	5	1
x_4	2	2	3	1	1	1	5	4	4	5	3	3	1
x_5	3	1	3	1	1	1	4	2	4	3	1	4	1
x_6	3	1	3	1	2	1	2	2	4	4	2	3	2
x_7	3	1	3	1	1	1	2	2	4	3	2	3	2
x_8	3	1	2	1	1	3	2	2	3	4	1	4	2
x_9	3	1	2	1	1	1	2	3	4	4	2	2	2
x_{10}	2	1	3	1	1	1	2	3	4	4	2	3	2
x_{11}	3	3	3	1	5	1	3	3	4	4	2	3	3
x_{12}	3	1	3	4	3	1	2	3	4	3	3	4	3
x_{13}	3	1	3	2	4	2	2	2	3	3	1	2	3
x_{14}	3	1	3	5	1	1	2	1	4	4	2	3	3
x_{15}	1	1	3	3	3	1	1	2	3	3	1	3	3

Table 10. Decision rules generated by RSIFAS approach

Rules	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	d	Supporting Objects
[1]	×	×	×	×	×	×	4	×	×	×	×	×	1	$x_5, x_6, x_7, x_8,$ x_9, x_{11}, x_{12}
[2]	×	×	×	×	×	×	5	×	×	×	×	×	1	$x_1, x_2, x_3, x_4, x_{10}$
[3]	×	×	×	1	×	×	2	×	×	×	×	×	2	$x_{14}, x_{15}, x_{16}, x_{17}$
[4]	×	×	×	×	×	×	3	2	×	×	×	×	2	x_{13}
[5]	×	×	×	×	×	×	2	×	×	3	×	×	3	x_{19}, x_{20}
[6]	×	×	×	×	5	×	×	×	×	×	×	×	3	x_{21}
[7]	1	×	×	×	×	×	×	×	×	×	×	×	3	x_{22}
[8]	×	×	×	×	5	×	×	×	×	×	×	×	3	x_{18}

$$U / R_{\alpha}^{a_2} = \{\{x_2, x_3, x_{31}\}, \{x_7\}, \{x_5, x_6, x_9, x_{10}, x_{11}, x_{24}, x_{29}, x_{32}, x_{35}\}, \{x_1, x_4, x_{12}, x_{15}, x_{16}, x_{17}, x_{25}, x_{26}, x_{33}\}, \{x_{14}, x_{18}, x_{20}, x_{21}, x_{27}, x_{30}, x_{38}, x_{39}\}, \{x_8, x_{13}, x_{19}, x_{22}, x_{23}, x_{28}, x_{34}, x_{36}, x_{37}\}\}$$

$$U / R_{\alpha}^{a_3} = \{\{x_{12}, x_{29}\}, \{x_8, x_{16}, x_{28}\}, \{x_1, x_2, x_3, x_4, x_5, x_6, x_7, x_9, x_{10}, x_{11}, x_{13}, x_{14}, x_{15,} x_{17}, x_{18}, x_{19}, x_{20}, x_{21}, x_{22}, x_{23}, x_{24}, x_{25}, x_{26}, x_{27}, x_{30}, x_{31}, x_{32}, x_{33}, x_{34}, x_{35}, x_{36}, x_{37}, x_{38}, x_{39}\}\}$$

$$U / R_{\alpha}^{a_4} = \{\{x_{24}, x_{37}\}, \{x_{34}, x_{38}\}, \{x_{14}, x_{21}, x_{33}, x_{35}, x_{39}\}, \{x_{19}, x_{20}, x_{23}, x_{25}, x_{26}, x_{27}, x_{36}\}, \{x_3, x_4, x_8, x_{10}, x_{11}, x_{13}, x_{15}, x_{17}, x_{18}, x_{22}, x_{28}, x_{32}\}, \{x_1, x_2, x_5, x_6, x_7, x_9, x_{12}, x_{16}, x_{29}, x_{30}, x_{31}\}\}$$

$$U / R_{\alpha}^{a_5} = \{\{x_{31}\}, \{x_{35}\}, \{x_{18}, x_{23}, x_{24}, x_{26}, x_{32}, x_{33}, x_{34}, x_{38}, x_{39}\}, \{x_1, x_{13}, x_{15}, x_{17}, x_{19}, x_{21}, x_{27}, x_{37}\}, \{x_4, x_5, x_8, x_9, x_{16}, x_{20}, x_{25}, x_{30}, x_{36}\}, \{x_2, x_3, x_6, x_7, x_{10}, x_{11}, x_{12}, x_{14}, x_{22}, x_{28}, x_{29}\}\}$$

$$U / R_{\alpha}^{a_6} = \{\{x_{28}\}, \{x_{35}\}, \{x_{22}, x_{24}, x_{32}, x_{34}, x_{36}, x_{37}, x_{38}, x_{39}\}, \{x_3, x_5, x_6, x_{10}, x_{15}, x_{23}, x_{30}, x_{31}, x_{33}\}, \{x_1, x_2, x_4, x_8, x_9, x_{11}, x_{12}, x_{13}, x_{14}, x_{17}, x_{18}, x_{19}, x_{21}, x_{25}, x_{26}, x_{27}, x_{29}\}, \{x_7, x_{16}, x_{20}\}\}$$

From the above analysis, it is observed that the attribute values of a_1, a_2, a_4, a_5, a_6 has classified into six categories. Let the categories are Very high, High, Medium, Low, Very low and Poor. Similarly, the attribute a_3 have classified into three categories, such as High, Medium, and Low. To make our analysis simple, we assign 1, 2, 3, 4, 5 and 6 to the categories Poor, Very low, Low, Medium, High, and Very high respectively. However these values are optional and do not effect our analysis. This reduces quantitative information system to qualitative information system. The reduced qualitative information system is further divided arbitrarily into training (55%) of 22 firms and testing (45%) of 17 firms. The reduced sample qualitative information system of Table 2 is presented in Table 11.

Table 11. Qualitative sample information system after employing RSFAS approach

Firms	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	d
x_1	4	3	5	1	3	2	5	3	5	4	2	4	1
x_2	6	6	5	1	1	2	5	4	5	5	4	5	1
x_3	4	6	5	2	1	3	5	3	5	5	3	5	1
x_4	5	5	5	1	1	1	5	4	4	5	3	3	1
x_5	4	4	5	1	2	2	4	2	4	3	1	4	1
x_6	4	3	5	3	4	2	2	2	4	4	2	3	2
x_7	4	2	5	3	3	2	2	2	4	3	2	3	2
x_8	2	1	3	2	1	6	2	2	3	4	1	4	2
x_9	4	4	2	1	1	2	2	3	4	4	2	2	2
x_{10}	5	2	5	1	2	3	2	3	4	4	2	3	2
x_{11}	4	6	5	1	6	3	3	3	4	4	2	3	3
x_{12}	4	1	5	5	4	4	2	3	4	3	3	4	3
x_{13}	4	4	5	4	5	5	2	2	3	3	1	2	3
x_{14}	2	1	5	6	3	4	2	1	4	4	2	3	3
x_{15}	1	2	5	5	4	4	1	2	3	3	1	3	3

Further, on employing rule generation algorithm on training data set of 22 objects we get the decision rules. The decision rules obtained are presented in Table 12.

4.8. Hybridization of RSIFAS and Neural Network

This section analyzes bankruptcy data set by hybridizing rough set on intuitionistic fuzzy approximation space and neural network (RSIFASNN). Generally, RSIFAS generates two types of rules such as deterministic and non-deterministic rules. Due to non-deterministic rules, RSIFAS approach unable to predict the decision for newly entered object. The hybridization of RSIFAS with neural network involves two stages such as initial stage, and final stage. At initial stage, the RSIFAS data analysis is used to reduce the dataset by removing the conflicting and repeated objects, thus making the data to be consistent. Further, the reduced data set is arbitrarily divided in to training and testing data. The training data set fed into the neural network in the final stage. The trained data set is validated by applying the testing data into the trained network model to check its validity.

The qualitative data set obtained after employing RSIFAS is considered as the target set and checked for its consistency. Thus, the target data set is arbitrarily divided into training data 22 (55%) and testing data 17 (45%). The experiment is conducted with the help of MATLAB 2008, by the back propagation algorithm. The number of inputs to the back propagation neural network is considered as the number of attributes 12. The sigmoid function is considered as the activation function, and the learning rate ranges from 0 to 1. The iteration is carried out to get minimum mean square error (MSE) as 0.05. The number of hidden nodes increased from 1 to 20, and it is observed that the minimum MSE falls at the 12th hidden node. The number of hidden layers increased sequentially from one to five to obtain more accuracy. It is observed that the there is a negligible deviation in the accuracy obtained with the five hidden layers. Therefore, the analysis is carried out with single hidden layer and 12 hidden nodes. The learning rate is iterated from 0 to 1, and also observed that at 0.78, the training network model obtains minimum MSE, facilitates to stop the training model. The number of hidden nodes identified using MSE and minimum percentile error (MPE) is depicted in Figure 3 and 4 respectively.

5. EXPERIMENTAL COMPARATIVE ANALYSIS

In this section, we demonstrate how hybridization technique (RSIFASNN) provides better accuracy as compared to other rough computing and statistical techniques. The testing data set is used to obtain the accuracy of the model. We consider the example of finding the bankrupted firms by applying statistical technique, rough set, RSFAS, RSIFAS and RSIFASNN. The accuracy produced

Table 12. Decision rules generated by RSFAS approach

Rules	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}	d	Supporting Objects
[1]	×	×	×	×	×	×	4	×	×	×	×	×	1	$x_5, x_6, x_7, x_8, x_9, x_{11}, x_{12}$
[2]	×	×	×	×	×	×	5	×	×	×	×	×	1	$x_1, x_2, x_3, x_4, x_{10}$
[3]	×	×	×	×	×	×	2	3	×	×	2	×	2	x_{16}, x_{17}
[4]	×	3	×	3	×	×	×	×	×	×	×	×	2	x_{13}, x_{14}
[5]	×	×	×	×	×	4	×	×	4	×	×	×	2	x_{15}
[6]	×	×	×	×	4	4	×	×	×	×	×	×	3	x_{19}, x_{21}, x_{22}
[7]	×	×	×	4	×	×	×	×	×	×	×	×	3	x_{20}
[8]	×	×	×	×	×	4	×	×	×	×	1	×	3	x_{18}

Figure 3. Mean Square Errors (MSE)

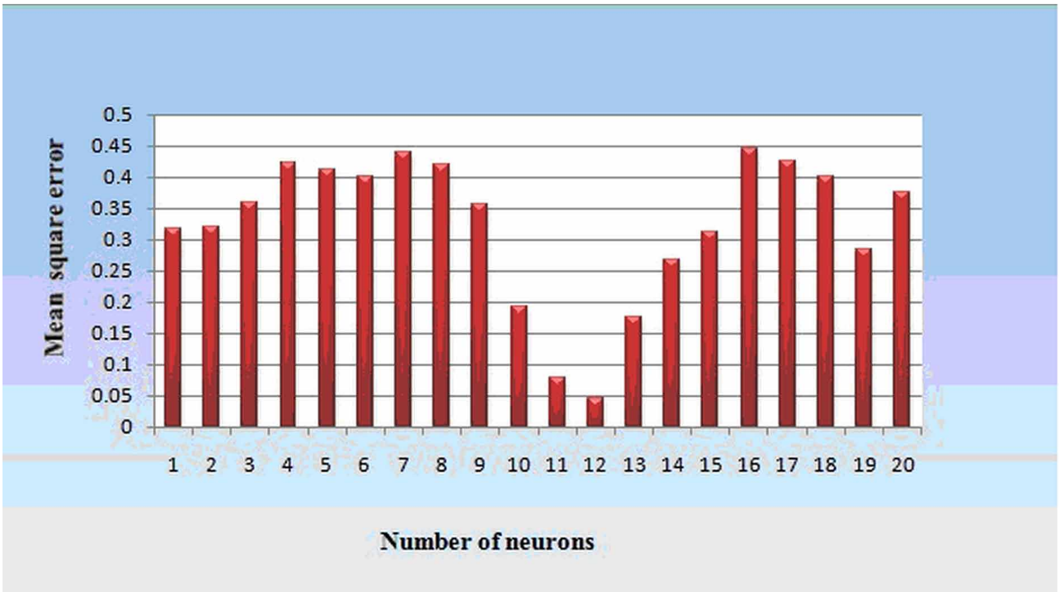
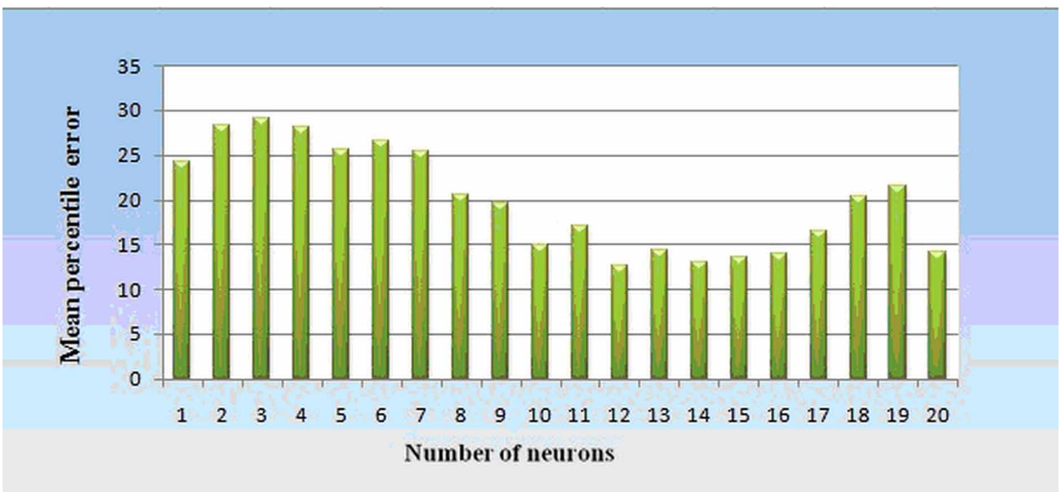


Figure 4. Mean Percentile Errors (MPE)



by above said techniques are listed in Table 13. From Table 13, it is observed that the RSIFASNN provides better accuracy than other techniques mentioned above. Also, we have analyzed the above models with agriculture dataset obtained from Krishi Vigyan Kendra, Vellore district of Tamilnadu to check its performance. Results pertaining to the analysis are presented in Table 13. It is seen that as the number of objects increases, the accuracy of the model RSIFASNN increases as compared to regression analysis, RS, RSFAS, and RSIFAS.

Table 13. Comparative study

Data sets	Bankruptcy	Agriculture
Sample size	39	2193
RSIFASNN		
Rules generated	13	681
Supporting objects	16 out of 17	924 out of 987
Accuracy obtained	**94.1 %**	**93.7 %**
RSIFAS		
Rules generated	8	678
Supporting objects	15 out of 17	901 out of 987
Accuracy obtained	**88.2 %**	**91.3 %**
RSFAS		
Rules generated	8	623
Supporting objects	15 out of 17	857 out of 987
Accuracy obtained	**88.2 %**	**86.9 %**
Rough Set (RS)		
Rules generated	6	515
Supporting objects	14 out of 17	815 out of 987
Accuracy obtained	**82.3 %**	**82.6 %**
Regression Analysis		
Multiple R^2	**81.1%**	**79.4%**
Adjusted R^2	76.1%	76.2%
p-value	0.00002	0.00005
Residual standard error	0.4029	0.4789
F-statistics	12.33	9.033

6. CONCLUSION

Much research has been carried out in the direction of predictive data analysis starting from statistical analysis to soft computing analysis and further to hybridized computing analysis. This paper is to make a comparative analysis between the existing techniques with Hybridization with neural network. To show the comparative analysis we have taken financial bankruptcy dataset and analyzed by using statistical techniques, rough computing techniques, and hybridized computing techniques. It is concluded that rough computing techniques provide better result as compared to statistical techniques whereas hybridized computing techniques provides slight notable increase in the accuracy.

REFERENCES

Acharjya, D. P. (2009). Comparative Study of Rough Sets on Fuzzy Approximation Spaces and Intuitionistic Fuzzy Approximation Spaces. *International Journal of Computational and Applied Mathematics*, *4*(2), 95–106.

Acharjya, D. P. (2015). Knowledge Extraction from Information System Using Rough Computing. In *Improving Knowledge Discovery through the Integration of Data Mining Techniques* (p. 161).

Acharjya, D. P., & Ezhilarasi, L. (2011). A Knowledge Mining Model for Ranking Institutions using Rough Computing with ordering rules and formal concept analysis. *International Journal of Computer Science Issues*, *8*(2), 417–425.

Acharjya, D. P., & Tripathy, B. K. (2008). Rough sets on Fuzzy Approximation Spaces and Applications to Distributed Knowledge Systems. *International Journal of Artificial Intelligence and Soft Computing*, *1*(1), 1–14. doi:10.1504/IJAISC.2008.021260

Acharjya, D. P., & Tripathy, B. K. (2009). Rough Sets on Intuitionistic Fuzzy Approximation Spaces and Knowledge Representation. *International Journal of Artificial Intelligence and Computational Research*, *1*(1), 29–36.

Altman, E. L. (1968). Financial ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy. *The Journal of Finance*, *23*(3), 589–609. doi:10.1111/j.1540-6261.1968.tb00843.x

Anitha, A., & Acharjya, D. P. (2015). Neural network and rough set hybrid scheme prediction of missing associations. *International Journal of Bioinformatics Research and Applications*, *11*(6), 503–524. doi:10.1504/IJBRA.2015.073237 PMID:26642360

Anitha, A., & Acharjya, D. P. (in press) Crop Suitability Prediction in Vellore district using Rough Set on Fuzzy Approximation Space and Neural Network.

Anitha, A., & Acharjya, D. P. (in press). Crop Suitability Prediction using Rough Set on Intuitionistic Fuzzy Approximation Space and Neural Network.

Beaver, W. (1966). Financial ratios as Predictors of Failure, Empirical Research in Accounting: Selected Studied. *Journal of Accounting Research*, *4*, 71–111. doi:10.2307/2490171

Cao,Y., Wan, G., & Wang, F. (2011). Predicting Financial Distress of Chinese listed companies using Rough Set Theory and Support Vector Machine. *Asia-Pacific Journal of Operational research,* 28(1), 95-109.

Chen, M. Y. (2011). Bankruptcy Prediction in Firms with Statistical and Intelligent Techniques and a Comparison of Evolutionary Computation Approaches. *Computers & Mathematics with Applications (Oxford, England)*, *62*(12), 4514–4524. doi:10.1016/j.camwa.2011.10.030

Chen, M. Y. (2012). Comparing Traditional Statistics, Decision tree Classification and Support Vector Machine Techniques for Financial Bankruptcy Prediction, *Intelligent Automation &. Soft Computing*, *18*(1), 65–73.

Cheng, J.-H., Lee, C.-M., Lee, T.-E., & Lee, L.-C. (2013). An Integrated Business Bankruptcy Prediction Model based on K-mean Clustering, Rough Sets and Support Vector Machines. *Advances in Information Sciences and Service Sciences*, *5*(17), 15.

Cielen, A., Peters, L., & Vanhoof, K. (2004). Bankruptcy Prediction using a Data Envelopment Analysis. *European Journal of Operational Research*, *154*(2), 526–532. doi:10.1016/S0377-2217(03)00186-3

De, S. K. (1999). Some Aspects of Fuzzy sets, Rough sets and Intuitionistic Fuzzy sets [Ph. D. Thesis]. IIT, Kharagpur, India.

Dimitras, A. I., Slowinski, R., Susmaga, R., & Zopounidis, C. (1999). Business Failure Prediction using Rough Sets. *European Journal of Operational Research*, *114*(2), 263–280. doi:10.1016/S0377-2217(98)00255-0

Fitzpatrick, P. J. (1932). *A Comparison of the ratios of Successful Industrial Enterprises with those of Failed Companies*. Accountants Publishing Company.

Greco, S., Matarazzo, B., Pappalardo, N., & Slowinski, R. (2005). Measuring Expected Effects of Iterventions based on Decision Rules. *Journal of Experimental & Theoretical Artificial Intelligence*, *17*(1-2), 103–118. do i:10.1080/09528130512331315864

Greco, S., Matarazzo, B., & Slowinski, R. (1998), A New Rough Set Approach to Evaluation of Bankruptcy Risk. In Operational tools in the management of financial risks (pp. 121-136). US: Springer.

Hua, Z., Wang, Y., Xu, X., Zhang, B., & Liang, L. (2007). Predicting Corporate Financial Distress based on Integration of Support Vector Machine and Logistic Regression. *Expert Systems with Applications*, *33*(2), 434–440. doi:10.1016/j.eswa.2006.05.006

Lindley, D.V. (1990). Regression and correlation analysis. In *Time Series and Statistics* (pp. 237-243). UK: Palgrave Macmillan.

Pawlak, Z. (1982). Rough Sets. *International Journal of Computer Information Science*, *11*(5), 341–356. doi:10.1007/BF01001956

Siegel, P. H., Andre, D. K., & Khursheed, O. (1995). Applications of Fuzzy Sets and the Theory of Evidence to Accounting. Jai Press.

Slowinski, R., & Zopounidis, C. (1995). Application of the Rough Set approach to Evaluation of Bankruptcy Risk. *Intelligent Systems in Accounting, Finance & Management*, *4*(1), 27–41. doi:10.1002/j.1099-1174.1995.tb00078.x

Xiao, Z., Yang, X., Pang, Y., & Dang, X. (2012). The Prediction for listed Companies Financial Distress by using Multiple Prediction Methods with Rough Set and Dempster–Shafer Evidence theory. *Knowledge-Based Systems*, *26*, 196–206. doi:10.1016/j.knosys.2011.08.001

Debi Prasanna Acharjya received his PhD in computer science from Berhampur University, India. He has been awarded with Gold Medal in M. Sc. from NIT, Rourkela. Currently he is working as a Professor in the School of Computing Science and Engineering, VIT University, Vellore, India. He has authored many national and international journal papers, book chapters, and five books to his credit. Additionally, he has edited four books to his credit. He is reviewer of many international journals such as Fuzzy Sets and Systems, Knowledge Based Systems, and Applied Soft Computing. He has been awarded with Eminent Academician Award from Khallikote Sanskrutika Parisad, Berhampur, Odisha; and Outstanding Educator and Scholar Award from National Foundation for Entrepreneurship Development, Coimbatore and Best Citizens of India Award from International Publishing House, New Delhi, India. Dr. Acharjya is actively associated with many professional bodies like CSI, ISTE, IMS, AMTI, ISIAM, OITS, IACSIT, CSTA, IEEE and IAENG. He was founder secretary of OITS Rourkela chapter. His current research interests include rough sets, formal concept analysis, knowledge representation, data mining, granular computing, bio-inspired computing, and business intelligence.

A. Anitha is an Associate Professor in the School of Information Technology at VIT University, Vellore, India. She received MCA degree from Adhi Parasakthi College of Science, Kalavai, Tamil Nadu, India, in 2000 and awarded PhD from School of Information Technology and Engineering, VIT University, Vellore. Her research interest includes data mining, fuzzy logic, neural network and rough sets. She has published many papers and is associated with the professional body CSI.