Adaptive Fuzzy Apriori - Tree Search Algorithm for Hybrid Information System

S. Bhuvaneswari, Dr.K.Meena

Abstract— **This paper illustrates a conceptual frame work for evaluating the impact of contemporary information systems mainly on business events. In the process of developing new data mining techniques and software to aid in better business solutions, we propose an 'efficient information system' that hybrids a tool of data mining incorporating rule induction on crisp database which is then transformed into a fuzzy database on which all samples Apriori tree search algorithm generates all possible prominent cases which are finally calibrated using a single layer feed forward rule based artificial neural network to make suggestive predictions and hence decisions. This model is hybrid to serve the aim of designing system architecture highly intelligent, both in system side aspects as well in the applicatory aspects. The system design is worked on an example case to deal with the customer risk management in the area of investment in the stock market.**

Index Terms—- **Adaptive Fuzzy Apriori(AFA)-Tree search Algorithm, Artificial Neural Network, Business application domain, Fuzzy database.**

I. Introduction

To operate in this world shaped by globalization and information revolution one should gather, analyze, understand, and act upon the information which successfully leads to the goal to become the winner in this information age [1]. To be more perfect and fittest the highest level of data accuracy and consistency should be the two essentialities needed in the real time applications.

Approaching in this way, eventually the partner of business or management should predict on making right moves to stay competitive. The vital component of the success is the prediction and execution of an actionable plan that supports the strategic goals and drives to all targeted decisions. Recent research publications in the area of Computer Science and Information Technology witness how the business events are performed by mining the hidden valuable information to improve the vital business operations like process, flow, data organization and social structures[2]. Especially in the domain of managerial decision making, the event logs of the system, popularly known as 'transaction log' [3] is oriented to some business activities and then leading to a process.

The event logs are to be incorporated effectively in a planned automation process so that an efficient system of business is set up [4]. Pursuing on this line, many system tools are developed in the market and the commercial interest of these tools towards the business domain is mounting due to the increasing awareness of the companies to incorporate effectively the same in the cross functional entities[5]. Mapping the various business functions on to the system the apt and the all purpose design architecture has to be borne in mind which will serve the activity. It is equally needed that the system oriented features [6] also has to be taken to their core importance, about which our design tries to elaborate in managing the time complexity and usage memory utility [7].

The corresponding author S. Bhuvaneswari is a Research Scholar in the Research Department of Computer Science in Srimathi Indira Gandhi College, Tiruchirapalli, TamilNadu, India. She acquired her post graduate degrees M.Sc in Physics, M.S in Information Technology and pre research degree M.Phil in Computer Science from Alagappa University and Bharathidasan University, TamilNadu, India. She has published adout five papers in the International Journals and four papers in National level seminars.

Prof.Dr.K.Meena , the Principal and HOD of Computer Science and Computer Applications, Srimathi Indira Gandhi College, Tiruchirappalli, Tamilnadu, South India, is the Ph.D. guide of the corresponding author and an eminent personality of inherent academic and administrative excellence. She has published many number of research papers and an appreciable number of persons were awarded Ph.D. in Computer Science under the guidance of her.

II. Design Architecture of the Proposed Hybrid System

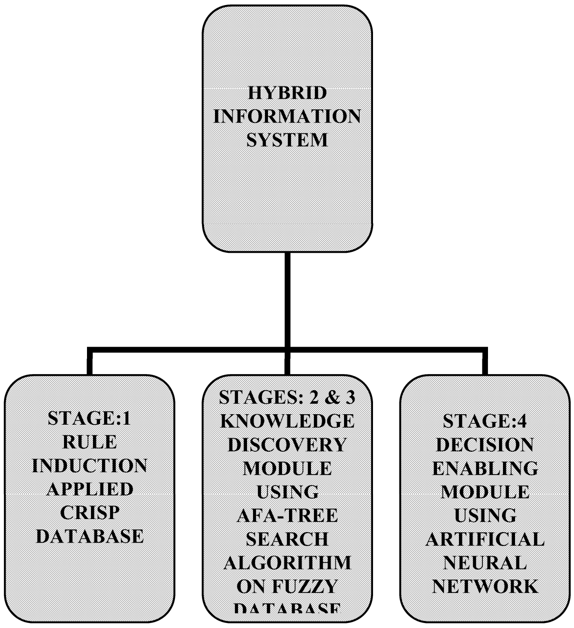

Figure: 1 Stages of the Hybrid Information System

The proposed hybrid Information System has different phases of knowledge discovery and decision making compartments. The various stages of the suggested hybrid model can be explained as follows.

STAGE: 1

The input to the model is a raw transactional data base which may be even operational in its type. Based on this master data base, a cluster of sub-data bases are extracted by applying rule induction (RI) technique, i.e., a technique focused on a specific (some) condition(s). The extracted cluster of sub-data bases may resemble as homogenous set due to the rule condition specified by RI and hence, efficient in providing specific, valuable information.

These sub data bases contain crisp form of classified dataset. As in this high time information scenario we are highly pertained with linguistic values on real time applications and as these values are in general, quantitative in nature and are abundant in random cluster we implement rule based conditions on such datasets to group them up to be homogeneous. It is well proven that operations on homogeneity means reduces time complexity and this idea led to apply RI in this fusion model.

STAGE: 2

The next step involves the fuzzy transformation of data [8] on the crisp sub-data bases by using fuzzy membership functions. Why that we use fuzzy transform at this stage is a valued question. The answer to this question is theories of fuzzy are highly able in predicting knowledge when worked on real world linguistic datasets. Most of the data mining works handle values in binary converted format while the proposed methodology handles values in a fuzzy database. This paper deals the ideology of training the real time quantitative values and for such values we suggest the fuzzy approach, which can trail a smooth transition of continuum of values with appreciable boundaries [9]. Thus for the supplied quantitative training data the logic of fuzzy generates the membership values which are collected as fuzzy sets and stored in a database very popularly called as the fuzzy database. The membership function used in this paper is the triangular type membership function for deriving the fuzzy values for the input quantitative data as it has been quite proven that triangular membership functions are highly versatile on real world applications.

STAGE: 3

This is the third stage in which tree search - Apriori algorithm is applied to distinguish frequent and dominant itemsets of transactions on the fuzzy database. Let us conjunct the general Apriori algorithm with that of the spanning tree search data structure on the fuzzy database which is hence patent as the Adaptive Fuzzy Apriori-Tree search algorithm which is briefed as below.

A. General Apriori Algorithm

General Apriori algorithm [10] can be for example, if D = $\{d_1, d_2, \ldots\ldots, d_n\}$ denotes the set of items of the methodology which are nothing but the sample dataset of the problem application in the database. (The proposed algorithm frames all possible meaningful clusters of the itemsets from the transformed fuzzy database.) If the association rule proposes both X and Y as decision variables which are the subsets of D and are independent, then the proposed association rule [11] implemented on the variables X and Y generates the support which is the percentage of dominant itemset transactions in the database. During the core phase of the algorithm, all dominant itemsets are generated. Now the rule over confidence standardizes the dominant itemsets generated by imposing the necessary probabilistic conditions to segregate the needed mined data to obtain the confidence. To make the rules more interesting and precise the appropriate threshold values are set for support and confidence

.

B. Adaptive Fuzzy Apriori (AFA) - Tree Search Algorithm on Samples

In the proposed model we assign fuzzy transformations to quantitative values to extract all possible versus impossible classifications and combinations of all degrees between the assigned intervals say ($\mathbf{i}_1$, $\mathbf{i}_2$). Here we fix up the maximum yield point say i_{max} and naturally by fuzzy logic i_{max} intends to the value 1. Thus all values of the fuzzy set are to be organized in between the points of minimum and maximum and this happens by assigning the membership functions. The type of function depends on the nature of the problem treatment but it is said to all and any type of problem assignment the triangular membership function suits describing the set values very well.Hence in our paper the triangular membership functions are used for training the quantitative data. Thus the quantitative data is being converted into fuzzy data and stored in the fuzzy database. Now the application of the data mining tool is proposed on the fuzzy database.

The proposed data mining algorithm classifies and mines all possible frequent and dominant itemsets, which enables knowledge discovery. The classification taxonomy is performed adapting the Apriori-tree search algorithm on the fuzzy data.

The proposed Adaptive Fuzzy Apriori (AFA) - Tree Search Algorithm is illustrated as below.

The algorithm first calculates the scalar cardinality of each linguistic term on all the transaction data. The design fabrication is so modeled in making essentially all needed clusters of decision variable samples, hereafter termed as ‘all samples itemsets’. This approach of clustering all needed decision variables as samples and considering only the all samples itemsets for application of Apriori–gen algorithm follows the ‘top-down’ spanning tree data structure approach in mining knowledge from the data base intended to reduce time complexity. Now the scalar cardinality of each item on every all samples itemsets is verified and calculated for non – zero occurrence and stored in a temporary set C_r and if this count satisfies the minimum support then the attributes of such linguistic terms are moved to L_r, which is the set of potential itemsets. Starting with r = 1, the first iteration generates the first set of all samples itemsets (join step) and if the availability of all items in the samples itemsets is non-zero then the count is performed (prune step) and verified for support threshold. At this stage the model indulges the condition of supply of a 'Support threshold' value by the expert professional. The value can be low or high accordingly the cutting edge of knowledge discovery lies.

When r =2, the next all samples itemsets is generated to the cycle and for which the counts are calculated independently for availability of items and verified for support.

The above step is repeated for any number of cycles of iteration until no all samples itemsets is left in making cluster or in reaching the support threshold. Thus the algorithm is made to run on the all samples itemsets based on the theory of satisfying minimum support.

The procedure of AFA algorithm on all samples itemsets includes the following steps.

STEP: 1
Let A is the crisp data set.

STEP: 2

The crisp data A is transformed into fuzzy data using the triangular membership functions of ranges 'low', 'medium', and ' high' with appropriate mapping values.

STEP: 3

If the membership function value of the input crisp data A at the i_{th} record, j_{th} field and k_{th} term is $\mu^{(i)}jk$ then the j_{th} field member ship function of the input data A of record i is given as

$\mathbf{A} = \Sigma \left(\mu^{(i)}_{jk} / \mathbf{X}_{jk} \right)$ where X_{jk} is the k_{th} fuzzy region of Aj.

STEP: 4

Let r = 1.

C. Join step of all samples

Cluster all possible combinations of all needed samples of decision variables which form the all samples itemsets for knowledge discovery. The search stores the items in a temporary set C_r and scans for non-repetitive itemsets but framing all possible itemsets.

D. Prune step of all samples

The prune step scans for all non-zero cardinality count for all itemsets and fetches the minimum (μ) and finally sum up. Thus for a data base of n records the summed up minimum scalar cardinality of all itemsets called as Count is

$$\text{Count} = \sum_{i=1}^{n} \mu$$

STEP: 5

If Count > £..the minimum support threshold store the results in Lr.

STEP: 6

Repeat the above steps till no support reaches the threshold £.

STEP: 7

To check the level of confidence we can calculate the value of confidence rule (CR) for all discovered itemsets by framing association rules applying the probabilistic approach. The rule of confidence CR for any relative items A and B with respect to A or B is
P (A U B) / P (A) and
P (A U B) / P (B).

STEP: 8

λ is fixed as the threshold of confidence for fixing the degree of confidence of each rule.

STEP: 9

If degree of confidence is greater than λ, then the corresponding rule is fixed up to derive the knowledge for decision making.

Steps 1 – 6 are essential requisites of the AFA-Tree search algorithm but Steps 7 – 9 are highly optional depending upon the domain user.

This algorithm is patent to be the Adaptive Fuzzy Algorithm (AFA)-Tree search worked on all samples of itemsets.

STAGE: 4

In this stage the artificial neural network popularly known as the ANN [12] is proposed to validate the rules generated on the dominant itemsets to provide all possible predictions by which suggestions were made. The support and the confidence weighed itemsets are inducted in the artificial neural network rule generator as decision variables and compared with the weights to the rule and finally summed up to get the decision making results. Here the rule based linear algorithm is suggested to map on the single layer feed forward neural network. The trained and the tested data are validated to the accepted level of accuracy. The hyper interacted predictions and suggestions lead to an efficient solution which could be by expert professionals. Thus this fusion model works exploring knowledge discovery and hence supports decision making. The application of trained interrogations through this model results in better expertise, resulting in a better decision support system (DSS) [13, 14, and 15].

III. .PROCEDURE SEMANTICS OF THE HYBRID MODEL

To the proposed model the procedure semantics has to be incorporated, so that the application scenario of the fusion model can be understood more precisely and hence for the various stages explained in the section –III (section -3) the working algorithm is drafted as below.

A. Procedure 3.1 for Stage:1-Rule Induction Algorithm

Input D_t // Heterogeneous Crisp Database
Output Dc // Homogeneous Crisp Database
Algorithm RI
//Induct rule over D_t;
Segregate D_t;

Get one or a collection of pruned crisp homogeneous database D_c;

On running the procedure 3.1 any voluminous database is homogeously classified so as the target of data search and analysis is narrowed down. In this modern information era the capacity of storage whether the main memory or supplementary is of no matter but the very crucial concept of data mining, which is to asses and analyse the data serves to be the edge cutting factor in the applicatory aspects. This bottel neck event is been minimised by clustering databases of homogeneous kind by inducting supervised rule on the unpruned crisp database. The result is the pruned crisp database D_c.

B. Procedure 3.2 for Stage:2-Fuzzy Transformation Algorithm

Input D_c // Homogeneous Crisp Database
Output D // Fuzzy Database
//Triangular membership functions are used to transform D_c to D;

Since this stage tries to develop a fuzzy associated data mining methodology of illustrating the knowledge in the pertaining application which mostly deal with quantitative values building the methodology the fuzzy transformations using triangular membership functions were used to transpose the pruned crisp database D_c into a fuzzy database D.

C. Procedure 3.3.1 for Stage:3-Apriori Algorithm

Input: L_{i-1} // Large itemsets of size i-1
Output:C_i // Candidates of size i
Apriori Algorithm
C_i=0;
for each I € L_{i-1} do
  for each j ≠ I € L_{i-1} do
    if 1-2 of the elements in I and J are equal then
      $C_k = C_k \cup \{I \cup J\}$;

D. Procedure 3.3.2 for Stage:3-Adaptive Fuzzy Apriori (AFA) - Tree Search Algorithm on Samples

Input: I // Itemsets
D// Transformed Fuzzy Database
S // Support
Output: L // Potential itemsets

Algorithm AFA-Tree
K= 0; // k is used as the scan number
L = Ø;
C_1 = I; // Cardinality of items in the fuzzy database which are the initial candidates and are set to be the items
repeat
  k=k+1;
  L_k=Ø;
for each I_i € C_k do
  C_i=0; // Initial counts for all itemsets are 0
for each t_j € D do
  for each I_i € C_k do
    if I_i € t_j then
    $C_i = C_i + 1$;
for each I_i € C_k do
  if C_i >= (s * D) do
$L_k = L_k \cup I_i$;
$L = L \cup L_k$;
C_{k+1} = Apriori (L_k)
until C_{k+1} = Ø;

Hence to summarize on to the rule inducted sampling theory based AFA-tree search algorithm, the tested rules on the database improves the architecture of the system design and the use of fuzzy transformations not only explains the far continuum of the real world values but tries to minimizes the number of iterations run by the Apriori procedure. The rule of confidence is now framed by the probabilistic function to get the knowledge study of the dominant samples obtained by the procedure. Let A and B be the two example dominant samples derived by the algorithm. The rule of confidence (CR) is obtained by the probabilistic function as CR= P (A) U P(B)) / P (A) or P(B) .The confidence rule may be up to a maximum of 100%. To be more precise the expert professional needs to fix up a threshold say λ.

E. Procedure 3.4 for Stage:4-Single layer feed forward Artificial Neural Network Rule Based Algorithm

Input: // Training data
  N // Initial neural network
Output: R // Derived suggestive rules
Rule Based Algorithm
  // generate suggestive rules that describe the output values in terms of the hidden activation values;
   Generate rules that describe hidden output values in terms of inputs;
   Combine the rules for validation;
The potential samples derived applying the algorithm by fixing up threshold for support and confidence is now trained using the single layer supervised ANN by rule based algorithm and now the ANN is ready for testing. The results were summarized for discussion.

IV. Experiment

This hybrid model is now applied on a prototype case which studies the customer side risk factors in investment in the Indian National Stock Market. It is for an example case the Indian perspective is taken and the money value is taken in terms of 'Rupee'.

Measuring and managing risk in an institution is a continuous process and not a one time activity. Thus in any sector, factors of risk has to be identified, measured, and to be mitigated. Simply put, ***risk***, can be defined through Statistics as the '***Probability'*** of loss which is cent percent honest in case customer ***investment*** in stock market. In such a context this paper tries finding an application paradigm using the above described intelligent hybrid system by choosing the essential decision variables. The data is a pruned prototype database with the prescribed decision variable dataset. The crisp database is obtained by appropriate rule induction that is what called as the preprocessing of dataset or data cleaning done on the collective heterogeneous database. Now the crisp database is transformed into the fuzzy database applying triangular membership functions. At this stage it is needed to mention about the fuzzy transformation of the quantitative variables like age, amount of investment, term of investment, recurring income, and property asset of the investor involving three triangular membership values. As an example taking the variable age, its membership values are categorized as 'agelow', 'agemedium', and 'agehigh'. Similarly it is for the other variables also. The fuzzy data sets are collected in the form of a matrix and the AFA-Tree search algorithm is applied on the matrix modeled fuzzy data to work, generating the potential all samples itemsets satisfying the minimum support threshold. Thus a 5 by 3 matrix is framed as shown below and spanning the matrix is made from the first to the last row scanning each and every item of each column of the matrix. We say this to be the top-down tree search data structure approach of the spanning tree in which the independent leaf nodes were scanned for all probable combinations keeping the top spanned branch nodes constant. The structure of the spanning matrix is as below.

$$\begin{pmatrix} \textbf{(Age)} & \textbf{low} & \textbf{med} & \textbf{high} \\ \textbf{(Amt)} & \textbf{low} & \textbf{med} & \textbf{high} \\ \textbf{(Term)} & \textbf{low} & \textbf{med} & \textbf{high} \\ \textbf{(Income)} & \textbf{low} & \textbf{med} & \textbf{high} \\ \textbf{(Asset)} & \textbf{low} & \textbf{med} & \textbf{high} \end{pmatrix}$$

Figure: 2 Spanning Tree Matrix

Such a scan for combinations applying AFA-Tree search algorithm yields about 27 clusters each of 9 combinations. One such case as an example is (agelow, amountlow, termlow, incomelow, assetlow). The redundant itemset variables were eliminated. Totally about 243 combinations were framed and the algorithm scans for non zero fuzzy values pertaining to all combinations and fetches the minimum to make the count. If the count satisfies the minimum support the combination is termed to be 'potential'. Such potential all samples itemsets above the threshold are inducted into ANN for training and thereby testing.

The architecture of ANN includes input neurons like age, amount, recurring income, and the property asset where the hidden neurons make the calculations comparing the input age with a maximum limit say 65, and the other inputs like amount of investment and income are weighed by value of property asset and the average sum of these if found. Thus there forms two output neurons and are manipulated according to the training rules and the decisions are made. Nomenclature for neurons:

Inputs neurons: Age – N_1, Amount – N_2, Recurring Income -N_3, Property Asset – N_4.

Hidden neurons: Age / Max age – H_1, Amount / Asset – H_A, Recurring Income / Asset –H_R.

Output neurons: Output 1 – O_1, Output 2 – O_2, Where $O_1 = H_1$, and $O_2 = (H_A + H_R) / 2$.

Some of the imposed rules may be say if $O_1 <= 0.4$ and $O_2 <= 0.4$ then term = long & company profile =C, if $O_1 <= 0.4$ and $O_2 > 0.4$ & $<= 0.7$ then term = either short or long & company profile =B or C, if $O_1 > 0.7$ and $O_2 > 0.7$ then term = short and company profile = A and so on.

Here the terminology used is

Term: Term of investment may be short or long and in general the term 'short' refers to a period of 1 week or less than 3 months and 'long' refers to a period of or more than 6 months. Note that the same sort of tips can be also applied for daily trading customers. Company profile: This profile suitably categorizes the companies registered with the National Stock Exchange in turns with the Securities and Exchanges Board of India (SEBI) under three classifications as dominant (A), proven (B), and new ventures (C) based on the persistence, performance, and profitable launches and moves. Based on such issues, training the rules on ANN, the data is tested to summaries the results.

The correlation co-efficient between the trained and tested predictions are calculated by which the error count is estimated. The standard percentage of error for ANN training and testing has been calculated which is found to be bear minimum narrating that the paradigm is correct. Now concentrating on the system side advantage, fusing fuzzy transformation to the design of the proposed model is intended to make it as a lossless data compression system which results in lesser time search.

This aim is achieved by the high condition specified data search using AFA – Tree search algorithm which adds value to the model in turn on the application of stock market trading events making the system more intelligent and also utilizes a very compressed working memory making the IS oriented system more efficient. The role of ANN in the model is found to be very apt in making predictions based on supervised learning as this model is intended to work on linguistic quantitative values. The ANN predictions were

ISBN: 978-988-98671-8-8

sharp and precise based on the training rules and enable the system to give a highly definite solution as illustrated in the design paradigm. The frequency distribution made on the pruned data set is found progressive in the mean range of age 40, amount of investment below Rs 300000, recurring income in range of Rs 30000, and property asset value in range of Rs 500000. The ANN results of the model also identified the cases above and below this occurrence limit. Accordingly to which the cases were approached and predictions were made which may yield solutions for which the percentage of error is found to be the minimum, hence making the hybrid system more intelligent.

The results of the knowledge discovery module were comprehended. For investors carrying business purpose deals the AFA – Tree search algorithm discovers about 13 all samples itemsets to be potential mentioning that they are of least risk cases in the pruned business sub database. Those cases are

Ag	**Am**	**Te**	**In**	**Ass**
h	l	h	m	h
l	m	l	l	h
l	m	l	h	m
l	m	h	h	m
m	m	l	h	m
m	m	m	h	m
m	m	h	h	m
m	h	l	m	h
m	h	l	h	m
m	h	h	m	h
m	h	h	h	m
m	h	l	m	h
m	h	h	h	l

Thus identifying a customer of any one of the above mentioned types we infer the knowledge, that the particular customer faces least risk in trading in the stock market. Other wise the customer is counseled for any other type of financial management. Similarly for investors carrying investment purpose deals the application of the algorithm discovers 21 all samples itemsets to be potential mentioning that they are of least risk cases. The point to be mentioned here is that we supply the minimum support value to be equal to or greater than 0.04.The discovered potential cases are now fed into the ANN for making expert decisions through which the deal can be carried out of risk. The feed forward ANN is trained using rule based algorithm for a heterogeneous cluster of crisp data sets which makes the testing of the data fittest in predictions. The independent weight value say A = age / maximum age limit say 65 and the summed weight say D = ∑ [(amount of investment / property asset) + (recurring income / property asset)] / 2 forms the hidden neuron calculations by which the suggestions for the customers to undergo risk least investment is provided.

V. Conclusion and Scope for Future Work

The proposed fusion model takes into implementation of both factors, the efficient system architecture with the intelligent AFA-Tree search algorithm to discover knowledge and decision making using ANN. The able run of the application on that system make valued decisions as discussed in the above section. As the scope for future work, the work can be extended for fore casting the approximate revenue turn over of the investor for any proposed period on running the safe trade using the same hybrid design.

Acknowledgment

S.Bhuvaneswari, the corresponding author of the paper extends her thanks to the Chief Editor and the anonymous referees who valued the paper. This paper briefs the algorithm used in her Ph.D. thesis which has been applied on an example case of predicting the risk factors on the investor side in the sector of Indian National Stock Market and acceptance of the paper to the conference adds credits for its competency.

References

[1].'Brien, Introduction to Information Systems – Essentials for the e-Business Enterprise.
[2].M.P.Van der Aalst, B.F.Van Dongen, J.Herbst, L.Maruster, G.Schimm and A.J.M.M.Weijters, Workflow mining: a survey of issues and approach, Data Knowl. Eng. 47(2003) (2), 237-267.
[3].Grigori, F.casati, U.Dayal and M.C.Shan, Improving business process quality through exception understanding, prediction, and prevention, In: P.Apers, P.Atzeni, S.Ceri, S.Paraboschi, K.Ramamohanarao and R.Snodgrass, Editors, Proceedings of the 27th International Conference on Very Large Data Bases (VLDB'01), Morgan Kaufmann, Los Alamitos, CA (2001), 159-168.
[4].Sayal, F. Casati, U. Dayal and M.C shan, Business process cockpit, Proceedings of 28th International conference on Very Large Data Bases (VLDB'02), Morgan Kaufmann, Los Alamitos, CA (2002), 880-883.
[5]..P.Van der Aalst, H.A.Reijers, A.J.M.M.Weijters, B.F.Van Dongen, A.K.Alves de Medeiros, M.Song and H.M.W.Verbeek, Business Process Mining: An Industrial Application, Information Systems.
[6].Daskalaki, I. Kopanas, M. Goudara and N. Avouris, data mining for decision support on customer insolvency in telecommunications business, European Journal of Operational Research 145 (2003), 239-255.
[7].Bose and R.K. Mahapatra, Business data mining – a machine learning perspective, Information and Management 39 (2001), 211-225.
[8].T. Yamakawa, Pattern recognition hardware system employing a fuzzy neuron, in: Proceedings of the International Conference on Fuzzy Logic and Neural Networks (Izuka, Japan, 20–24), 1990, pp. 943–948.
[9]. L.A. Zadeh, Fuzzy sets, *Information and Control* **8** (1965), pp. 338–353.
[10]. R. Agarawal,T. Imielinksi, A Swami, *"Mining association rules between sets of items in large database,"* The 1993 ACM SIGMOD conference, Washington DC, USA, 1993.
[11]. .R. Agrawal, R. Srikant, Fast algorithm for mining association rules, The international Conference on Very Large Data Bases, 1994, pp.487-499.
[12].R.J. Schalkoff, Artificial Neural Networks, McGraw-Hill, New York (1997). [13].Kim and W.N. Street, An intelligent system for customer targeting: a data mining approach, *Decision Support Systems* **37** (2004), pp. 215–228.
[14].R.J. Kuo, P. Wu and C.P. Wang, An intelligent sales forecasting system through integration of Artificial Neural networks and fuzzy neural networks with weight elimination, *Neural Networks* **15** (2002), pp. 909–925.
[15].H. Mannila, Methods and problems in data mining, in: The International Conference on Database Theory, 1997.