An Effort to Developing the Knowledge Base in Data Mining by Factor Analysis and Soft Computing Methodology

Dharmpal Singh[1], J. Paul Choudhury[2], Mallika De[3]

Abstract - Huge amounts of data are collected nowadays from different application domains is not feasible to analyze all these data manually. Knowledge Discovery in Databases (KDD) is the nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data. With the help of data mining process, it is possible to looking for unknown relationships and patterns and extracting useful information volumes of data in data warehouse. Mining means to abstract the information or mode which is implicit, unknown and valuable in large database or data warehouse. Now a day, association rules mining from large databases is an active research field of data mining motivated by many application areas. Mining association rules (knowledge) is to searching out all exiting valuable relationship of items from the given database with the statistics principle. However, there are some problems in the strong association rules mining based on support-confidence framework. Firstly, there are a great number of redundant association rules generated, then it is difficult for user to find the interesting ones in them; secondly, the correlation between attributes of given application areas is ignored. Therefore, a new measure criterion, which is factor analysis, should be introduced to association rules mining, and the more important aspect is the use of factor analysis to reduce the amount of rules and then relate the data with the help different type of soft computing methodology to find the hidden pattern in the database.

In this paper an effort has been made to create a knowledge base using the available data items of Boston city. Initially, the factor analysis has been applied on the available data to form the resultant total effect of the data items. Thereafter, the concepts of fuzzy logic, neural network and particle swarm optimization have been applied on the data items. The neural network has been trained using the data values. A test data comprising all (A-I items data) values can be applied to the trained neural network to get the predicted value. The data item can also be tested by particle swarm optimization. Based on the minimum error, a particular model has been selected for the knowledge discovery in the data mining.

Keywords - Data mining, Association rule, Fuzzy logic, Neural Network, Factor Analysis, Particle Swarm optimization, Residual Analysis.

1 Introduction

Data Mining is a process of looking for unknown relationships and patterns and extracting useful information volumes of data in data warehouse. Mining means to abstract the information or mode which is implicit, unknown and valuable in large database or data warehouse. Nowadays, many research papers related to said topic has been published where J. E. Moreno, O. Castillo and J. Castro [1] have presented a clustering technique using K Means, Fuzzy K Means, etc. The authors have applied the technique on specific databases (Flower Classification and Mackey Glass Time Series) to identify most relevant and significant patterns in pattern recognition, to extract production rules using Mamdani and Takagi-Sugeno Kang fuzzy logic inference system type. The authors have shown the subtractive clustering technique in conjunction with FIS method (Mamdani and Takagi-Sugeno Kang), and opined that it has shown better performance than other techniques in all sample tests.

[1] Assistant Professor, Department of Computer Science & Engineering JIS College of Engineering, Kalyani, Nadia – 741235, West Bengal, India, Email: - singh_dharmpal@yahoo.co.in
[2]Professor, Department of Information Technology Kalyani Government Engineering College, Kalyani, Nadia – 741235, West Bengal, India, E-mail:-jnpc193@yahoo.com
[3]Professor, Department of Engineering and Technological Studies University of Kalyani, Kalyani, Nadia – 741235, West Bengal, India, E-mail:-demallika@yahoo.com

S. Dehuri, A. K. Jagadeva, A. Ghosh and R. Mall [2] have proposed a multi objective association rule for data mining technique. The authors have also used genetic algorithm for optimization. Under the objective function, the techniques confidence factor and comprehensibility has been used for making different association rule. Further the authors have opined that the fast scalability techniques using inherent parallel processing nature of genetic algorithm are suitable for homogeneous dedicated network of work stations.

P. Nagar and S. Srivastava [3] have tried to explore the genetic algorithm in its broader sense of simulated evolutionary system, as they have mentioned that genetic algorithm is generally more focused on optimization and search. The authors have encoded simple direct marketing problem and assumed certain things to make the problem easier. The authors have used artificial life simulating evolution technique and opined that the genetic algorithm provides optimal advantage in business applications.

H. Lu, R. Setiono and H. Liu [4] have presented an approach to discover symbolic classification rules using neural networks. Initially, trained the network to achieve the required accuracy rate and then removed the redundant connections of the network by a network pruning algorithm. To analyzed the activation values of the hidden units in the network and generated the classification rules using the result of the analysis. The authors have established an effective approach by the experimental results on a set of standard data mining test problems.

R. S. Segall and Q. Zhang [5] have presented preliminary research in the area of the application of modern heuristics and data mining techniques in knowledge discovery. They used data mining for neural networks using Neural Ware Predict software and genetic algorithm using Bio-discovery Gene Sight software for bioscience data sets of continuous numerical valued Abalone fish data.

B. Liu et al. [6] have presented a new technique for organizing discovered rules in different levels. The algorithm consists of two steps. The first one is to find top-level general rules, descending down the decision tree from the root node to find the nearest nodes whose majority classes can form significant rules. These rules have been termed as the top-level general rules. The second one is to find exceptions, exceptions of the exceptions and so on. They have made an effort to determine whether a tree node can form an exception rule or not using two criteria: significance and simplicity.

X. Ni [7] has proposed a research on data mining based on neural networks and initially explained the different models of neural networks and then divided the whole task into three steps, namely, a) Data mining process based on neural network (mentioning about how to prepare the data and then writing the rules and selecting the optimal rule), b) Data mining type base on neural network (explaining data mining with fuzzy neural networks and data mining with self organization maps), and, c) Key techniques and approaches of implementation (effective combination of neural and data mining techniques and effective combination of knowledge processing and neural combinations). It was also mentioned that combination of data mining methods and neural network model will greatly improve the effectiveness of data mining methods.

Y. Dhanalakshmi and T. R. Babu [8] have explored the possibility of integrating fuzzy logic with data mining methods, using genetic algorithm for intrusion detection. The authors have proposed the data mining algorithm to mine the fuzzy association rules, by extracting the best possible rule using genetic algorithm. They showed their comprehensive work by an experiment.

G.V.S.N.R.V. Prasad et al. [9] presented the clustering technique and used fuzzy association rules using multi-objective genetic algorithms. In the first phase, the data has been optimized to reduce the number of comparisons using clustering technique. In the second phase they have used multi-objective genetic algorithm to find the number of fuzzy association rules using threshold value and fitness function.

Association rules [10] mining is an important task of data mining, which describes potential relationships among data items in databases, the main idea of which was first proposed by R. Agrawal et al. in 1993, shortly after then realized by the well-known Apriori algorithm [11], which was an influential algorithm for mining frequent item sets for Boolean association rules. Usually most of other algorithms are improved on the basis of the Apriori algorithm. A lot of Apriori-like approaches [12-14] have achieved good performance. However, it is costly to handle a large number of candidate sets. Synchronously it needs to scan database repeatedly. FP-growth [14] that avoids the costly generation of a huge number of candidate sets is efficient and scalable for mining frequent patterns, and it runs faster than Apriori-like algorithms.

Clustering techniques have been applied by many researchers. Xiyu Liu et al. [15] have presented a survey on projection clustering. The authors have made extensive studies on the algorithms and applications of a new clustering technique based on grid architecture. Their new technique integrates minimum spanning tree and grid clustering together and by this integration of projection clustering with grid technique, the complexity of computing is lowered to $O(n \log n)$.

H. Mocian [16] has presented a survey on distributed clustering. The author has opined that the distributed clustering has been employed in a variety of distributed environments, from computer clusters to P2P networks with thousands of nodes, to wireless sensor networks etc.

S. Saha and S. Bandyopadhyay [17] have proposed fuzzy point symmetry based genetic clustering technique (Fuzzy-VGAPS) which can determine the number of clusters present in a data set as well as a good fuzzy partitioning of the data. A new fuzzy cluster validity index, FSym-index, which is based on the newly developed point symmetry based distance, was also proposed by them.

A. Faruq et al. [18] have presented an algorithm for clustering data in large datasets using image processing approaches. They mapped the dataset into a binary image plane and synthesized image is then processed utilizing efficient Image Processing techniques to cluster the data in the dataset. They have shown their work by an experiment.

P. Pal and B. Chanda [19] have proposed a clustering technique that extracts sub- and sup-clusters based on a simple measure of circular symmetry. These sub-clusters and sup-clusters are then used as building blocks to form final clusters of any arbiter shape including concave ones through merging and splitting iteratively. The proposed method is tested on multi-spectral satellite imagery by an experiment.

D. Karaboga and B. Basturk [20] have compared the performance of ABC algorithm with that of differential evolution (DE), particle swarm optimization (PSO) and evolutionary algorithm (EA) for multi-dimensional numeric problems and then shown that the performance of ABC algorithm can be efficiently employed to solve engineering problems with high dimensionality.

D. Karaboga and B. Basturk [21] have compared performance of the Artificial Bee Colony (ABC) algorithm for constrained optimization problems. They said that it has been first proposed for unconstrained optimization problems and showed that it has superior performance on these kinds of problems. They have mentioned that the ABC algorithm extended for solving constrained optimization problems and applied it to a set of constrained problems.

C. Ozturk and D. Karaboga [22] have compared Artificial Bee Colony (ABC) algorithm with Particle Swarm Optimization (PSO) algorithm and other nine classification techniques from the literature and showed that ABC algorithm can efficiently be used for multivariate data clustering. Thirteen typical test data sets from the UCI Machine Learning Repository were used to demonstrate the results of the techniques.

Wen et al. [23] have proposed an approach of particle swarm optimization PSO for compact planar microwave filter design and then finite element method (FEM) were combined together to allow optimal filter design with arbitrary geometries. Thereafter, with an example they showed that PSO–FEM approach is effective to make the structure variation converge to the desired target and the final optimal filter structure has much smaller size.

Xiangyang Wang et al. [24] have proposed a new feature selection strategy based on rough sets and Particle Swarm Optimization (PSO) and they carried out an experimentation using UCI data, which compares the deterministic rough set reduction algorithms with PSO and showed that the PSO is efficient for feature selection. In this paper, they have introduced new incremental learning algorithms based on harmony search.

J. Kennedy and R. Eberhart [25] have introduced particle swarm methodology for the optimization of nonlinear functions and then described the relation of genetic algorithm and neural network training with particle swarm optimization.

Z. Karimiet al. [26] have proposed a new classification algorithm for the classification of batch data called *harmony-based classifier* and then gave its incremental version for classification of data streams called *incremental harmony-based classifier*. Finally, they have improved it to reduce its computational overhead in absence of drifts and increased its robustness in presence of noise and called it *improved incremental harmony-based classifier*. The proposed methods were evaluated on some real world and synthetic data sets and the experimental results showed the robustness of improved incremental harmony-based classifier on the data.

J. Wang and G. Karypis [27] presented a new classifier, HARMONY, which directly mines the final set of classification rules. They have shown that HARMONY outperforms many well-known classifiers in terms of both accuracy and computational efficiency, and scales well with respect to the database size. HARMONY also has high efficiency and good scalability as compared to the other search strategies and pruning methods into the rule discovery process in large text and categorical databases.

H. Mohamad, et al. [28] have illustrated the ability of SA to develop an accurate fuzzy classifier and developed a SAFCS method. Experiments were performed with eight UCI data sets and the results indicated that the proposed SAFCS achieves competitive results in comparison with several well-known classification algorithms.

M. U. Shaikhet al. [29] have proposed a novel idea about the possibility of designing an intelligent decision support system by using the techniques of data mining as well as the differential evolution algorithm of artificial neural networks and used a pre-existing differential evolution algorithm with slight modification within the DSS environment. They have assumed that this merger will lead towards more development and advancement within the concept of DSS.

A. Daniel et al. [30] have developed initial results in scheduling procedure for an automated steel plate fabrication facility. They have used Taboo search and evaluated that it gave better performance than other optimal solution for small and large problems. Their results also showed that the Tabu search method works well for this problem and combining Tabu search with simulation allows the incorporation of more realistic constraints on system operation.

Mori et al. [31] have proposed a data mining based method that deal with short-term load forecasting in power systems. As a data mining technique, they used regression tree to extract some meaningful rules from a database so that the nonlinear relationship between input and output data is clarified. Tabu search has been applied to globally optimize structure of the regression tree to enhance the efficiency of data mining. They have also used a multi-layer perceptron network to predict one-step ahead daily maximum load with each learning data obtained by the optimal regression tree.

M. A. Tahir and J. E. Smith have proposed [32] a new ensemble technique to improve the performance of NN classifier where each classifier uses a different distance function and potentially a different set of features (feature vector). These feature vectors are determined for each distance metric using Simple Voting Scheme incorporated in Tabu Search (TS). The proposed ensemble classifier with different distance metrics and different feature vectors (TS-DF/NN) was evaluated using various benchmark data sets from UCI Machine Learning Repository and result showed a significant increase in the performance when compared with various well-known classifiers. Furthermore, their proposed ensemble method was also compared with ensemble classifier using different distance metrics but with same feature vector (with or without Feature Selection).

Q. Song and B. S. Chissom [33] have applied the fuzzy time series model to forecast the enrollments of the University of Alabama, where a first-order time invariant model has been developed and a step-by-step procedure has been provided.

They have used the following model

$A_i = A_{i-1} . R$----------------- (1)

where A_{i-1} is the enrollment of year i - 1 represented by a fuzzy set, A_i is the forecasted enrollment of year i represented by a fuzzy set. "." is the max-min composition operator, and R is a fuzzy relation indicating fuzzy relationships between fuzzy time series. But the method requires a large amount of computation to derive the fuzzy relation R of (1), and the max-min composition operations of (1) will take a large amount of computation time when the fuzzy relation R of (1) is of high dimension.

H. Bintley [34] has successfully applied fuzzy logic and approximate reasoning to a practical case of forecasting, but the concept of fuzzy time series was not applied on the method presented in [34].

Q. Song and B. S. Chissom [35] have used first order time variant models and utilized 3 layer back propagation neural network for defuzzification.

G. A. Tagliarini et al. [36] have demonstrated that artificial neural networks could achieve high computation rates by employing massive number of simple processing elements of high degree of connectivity between the elements. This paper presented a systematic approach to design neural networks for optimizing applications.

F. G. Donaldson and M. Kamstra [37] have investigated the use of artificial neural network (ANN) to combine time series forecasts of stock market volatility from USA, Canada, Japan and UK. The authors presented combining procedures to a particular class of nonlinear combining procedure based on artificial neural network.

J. V. Hansen and R. D. Nelson [38] have presented the neural network techniques, which provided valuable insights for forecasting tax revenues. The pattern finding ability of neural networks gave insightful and alternate views of the seasonal and cyclical components found in economic time series data. It was found that neural networks were stronger than exponential smoothing and ARIMA (autoregressive integrated moving average).

M. Ishikawa and T. Moriyama [39] have presented various methods of learning and the process of predicting time series analysis, which were ranged from traditional time series analysis to recent approaches using neural networks. It described that back propagation learning had a difficulty in interpreting hidden inputs. In order to solve these problems, a structural learning method was proposed which was based on an information criterion.

S. M. Chen [40] presented a new method to forecast university enrollments based on fuzzy time series. S. M. Chen [40] made groups after getting partitions (equal length intervals) of the historical data. After forming logical relationship groups, the forecasted output was calculated. It followed a procedure that when the group contained two or more fuzzy logical relationships, the predicted value would be the average of the midpoint of those two or more fuzzy logical relations and when the group contained a single fuzzy logical relationship, it would be the midpoint of logical relation partition.

Distinguished authors from diversified genres have used data mining [1], [2], [6] [10], [11], [12], [13], [14], [15] techniques for association rule generation and for selection of best rule among the extracted rules. Moreover, the authors have used soft computing techniques like neural network [5], [6], [7], [36], [37], [38], genetic algorithm [2], particle swarm optimization (PSO) [20] [22], [23], [24] and others swarm intelligence [20], [21], [22], [23], [24], [25], [26], [27], [28, [29], [30], [31], [32] techniques on singe data set [33], [34], [35], [36], [37], [38], [39], [40] for the prediction of the information which is already available. It has been observed that the said researchers have made a theoretical comparative study regarding the performance among various soft computing models. But the research work regarding the gathering of knowledge from set of number of information is still not available. Here, in research work, an effort is being made to extract knowledge from the set of data items already available in data warehouse. For this purpose, Boston city data set have been used.

In this paper an effort has been made to create a knowledge base using the available data items of Boston city. Initially, the factor analysis has been applied on the available data to form the resultant total effect value of the data items. Thereafter, the concepts of fuzzy logic, neural network and particle swarm optimization have been applied on the data items. The neural network has been trained using the data values. A test data comprising all (A-I items data) values can be applied to the trained neural network to get the predicted value.

The data item can also be tested by particle swarm optimization. The particular model has been selected by maximum number of minimum parameters of average error and residual analysis. That selected model has been

ISSN 2229-5518

used for the discovery of knowledge in the data base. This paper has been organized in different section. In first section, introduction and literature survey about the topics have been furnished. In second section, implementations of the methods have been discussed in detail. In third section and fourth section, reviews of result and steps of knowledge discovery in data mining have been furnished. The conclusion has been furnished in section five.

2 Implementation

Step 1:-

The data contains on the Boston city with following attribute per capita crime rate (A), proportion of residential land zoned for lots over 25,000 sq.ft. (B), proportion of non-retail business acres (C), Charles River dummy variable (= 1 if tract bounds river; 0 otherwise) (D), nitric oxides concentration (parts per 10 million) (E), average number of rooms per dwelling (F), proportion of owner-occupied units built prior to 1940(G), weighted distances to five Boston employment centers (H), index of accessibility to radial highways (I) and full-value property-tax rate per $10,000 (J) certain time series years as furnished in table 1.

Step 2:-

If the value of A is 6.32, B is 18000 a, C is 2310 and D have zero (ignore it from the table), E is 538, F is 6575, G is 65200, H is 4090, I is 1000 and J is 24200. Therefore it can be told that the value of (J) a particular year depends on the value of item A-I of the previous year. This can be used to form the association rule for the Data mining.

Table 1
Boston City Data

Serial Number	A	B	C	D	E	F	G	H	I	J
1	6.32	18000	2310	0	538	6575	65200	4090	1000	242000
2	273	0	7070	0	469	6421	78900	4967.1	2000	242000
3	273	0	7070	0	469	7185	61100	4967.1	2000	242000
4	324	0	2180	0	458	6998	45800	6062.2	3000	222000
5	691	0	2180	0	458	7147	54200	6062.2	3000	222000
6	299	0	2180	0	458	6430	58700	6062.2	3000	222000
7	883	12500	7870	0	524	6012	66600	5560.5	5000	311000
8	145	12500	7870	0	524	6172	96100	5950.5	5000	311000
9	211	12500	7870	0	524	5631	1E+05	6082.1	5000	311000
10	170	12500	7870	0	524	6004	85900	6592.1	5000	311000
11	225	12500	7870	0	524	6377	94300	6346.7	5000	311000
12	117	12500	7870	0	524	6009	82900	6226.7	5000	311000
13	938	12500	7870	0	524	5889	39000	5450.9	5000	311000
14	630	0	8140	0	538	5949	61800	4707.5	4000	307000
15	638	0	8140	0	538	6096	84500	4461.9	4000	307000
16	627	0	8140	0	538	5834	56500	4498.6	4000	307000
17	1054	0	8140	0	538	5935	29300	4498.6	4000	307000
18	784	0	8140	0	538	5990	81700	4257.9	4000	307000
19	803	0	8140	0	538	5456	36600	3796.5	4000	307000
20	725	0	8140	0	538	5727	69500	3796.5	4000	307000
21	1252	0	8140	0	538	5570	98100	3797.9	4000	307000
22	852	0	8140	0	538	5965	89200	4012.3	4000	307000
23	1232	0	8140	0	538	6142	91700	3976.9	4000	307000
24	988	0	8140	0	538	5813	1E+05	4095.2	4000	307000
25	750	0	8140	0	538	5924	94100	4399.6	4000	307000
26	841	0	8140	0	538	5599	85700	4454.6	4000	307000
27	672	0	8140	0	538	5813	90300	4682	4000	307000
28	956	0	8140	0	538	6047	88800	4453.4	4000	307000
29	773	0	8140	0	538	6495	94400	4454.7	4000	307000
30	1002	0	8140	0	538	6674	87300	4239	4000	307000

Step 3:-

The contribution of eigen value among all the eigen values have been calculated with the help of Matlab tool. All the Eigen values belong to each component have been calculated and it has been observed that element A and B have minimum contribution as compared to others. Therefore it has been ignored. The last six components eigen value are 0.266, 0.3117, 0.6858, 0.8692, 2.0565 and 3.7126 respectively. The major factors have been calculated as √eigen value × eigen vectors corresponding to that eigen value. The value of major factors is furnished in table 2.

Table 2
Contribution of Eigen Vector for Different Eigen Value

Data Item	0.266	0.3117	0.686	0.869	2.06	3.713
A	-0.25	-0.13	-0.13	-0.06	-0.47	0.81
B	-0.08	0.02	0.49	0.08	0.82	-0.23
C	0.21	0.39	-0.15	-0.02	0.23	0.85
E	-0.19	0.07	0.34	0.1	0.15	0.88
F	-0.28	0.34	-0.03	-0.34	-0.24	-0.79
G	0.07	-0.12	0.09	-0.85	0.26	0.43
H	-0.04	-0.08	-0.33	-0.06	0.69	-0.62
I	-0.19	-0.01	-0.41	0.1	0.69	0.53

Step 4:-

The cumulative effects for all these major factors have been calculated by adding the values row wise corresponding to each element of table 2. As for an example, the addition of six values for six major eigen values corresponding to A is

-0.22, that of B is 1.11 and accordingly the other cumulative effect have been calculated. Now a relation has been formed using the cumulative effect of all elements.

Total Effect Value = A × -0.22 + 1.11 × B+ 1.51 × C + 1.35 × E + F × -1.34 + G ×-0.12 + H ×-0.44 + I × 0.71. Now using the relation, a value has been computed and furnished in table 3 using the value of table 1.

Table 3
Total Effect Value Incorporating all Items

Serial Number	A	B	C	D	E	F	G	H	I	Total Effect Value
1	6.32	18000	2310	0	535	6575	65200	4090	1000	6468.91
2	27.31	0	7070	0	462	6421	78900	4967	2000	-7534.8
3	27.29	0	7070	0	462	7183	61100	4967	2000	-6422.6
4	32.37	0	2180	0	455	6995	45800	6062	3000	-11208
5	69.05	0	2180	0	453	7147	54200	6062	2000	-12723
6	298.3	0	2180	0	455	6430	58700	6062	3000	-12353
7	88.29	12500	7870	0	524	6012	66600	5561	5000	11502
8	144.55	12500	7870	0	524	6172	96100	5951	5000	7563.6
9	211.21	12500	7870	0	524	5631	1E+05	6082	5000	7747.97
10	170.04	12500	7870	0	524	6004	85900	6592	5000	8724.81
11	224.89	12500	7870	0	524	6377	94300	6347	5000	7312.9
12	117.47	12500	7870	0	524	6009	82900	6227	5000	9250.45
13	95.78	12500	7870	0	524	5889	59000	5451	5000	15025.8
14	629.76	0	8140	0	535	5949	61000	4708	4000	-1643.8
15	637.96	0	8140	0	535	6096	84500	4462	4000	-4554.5
16	627.39	0	8140	0	535	5834	56500	4499	4000	-857.27
17	1053.9	0	8140	0	533	5935	29300	4499	4000	2177.55
18	784.2	0	8140	0	533	5990	81700	4258	4000	-4018.9
19	802.71	0	8140	0	533	5456	36600	3797	4000	2307.6
20	725.6	0	8140	0	533	5727	69500	3797	4000	-1986.6
21	1251.8	0	8140	0	538	5570	98100	3798	4000	-5324.6
22	832.04	0	8140	0	535	5963	89200	4012	4000	-4792.3
23	1232.5	0	8140	0	538	6142	91700	3977	4000	-5397.6
24	988.43	0	8140	0	538	5873	1E+05	4095	4000	-6031.5
25	750.26	0	8140	0	538	5924	94100	4400	4000	-5473.3
26	840.54	0	8140	0	538	5500	85700	4455	4000	-4073.9
27	671.91	0	8140	0	538	5813	90300	4682	4000	-4975.6
28	955.77	0	8140	0	538	6047	88800	4453	4000	-5071.1
29	772.99	0	8140	0	538	6495	94400	4455	4000	-6303.7
30	1002.5	0	8140	0	538	6674	87300	4239	4000	-5647.2

Step 5:-

The total effect value and corresponding output have been furnished in table 4.

Table 4
Values of the Total Effect and the Actual Data

Serial Number	Total effect Value	Consequent Item
1	6468.91	242000
2	-7534.82	242000
3	-6422.58	242000
4	-11507.7	222000
5	-12723.4	222000
6	-12353.1	222000
7	11501.98	311000
8	7563.599	311000
9	7747.97	311000
10	8724.807	311000
11	7312.896	311000
12	9250.449	311000
13	15025.81	311000
14	-1643.81	307000
15	-4554.53	307000
16	-857.27	307000
17	2177.551	307000
18	-4018.9	307000
19	2307.604	307000
20	-1986.62	307000
21	-5324.57	307000
22	-4792.26	307000
23	-5397.56	307000
24	-6031.46	307000
25	-5473.34	307000
26	-4073.9	307000
27	-4975.62	307000
28	-5071.05	307000
29	-6303.73	307000
30	-5647.16	307000

2.1 Implementation by Fuzzy Logic

The total effect value (absolute value) has been sorted for the further application. The method of fuzzy logic has been applied on the total effect value. The range of values for total effect value is 857-15026. The universe U has been partitioned into five equal length intervals. The intervals are chosen as A1= [850, 3683], A2= [3683, 6522], A3= [6522, 9358], A4= [9358, 12194] and A5= [12194-15030]. Fuzzy sets have been defined on the universe and some linguistic values have also been determined. Let, A1 = (many) A2 = (many, many) A3 = (very many) A 4= (too many) be the possible value. All the Fuzzy sets Ai (i=1, 2, 3, 4) are expressed as follows.

The available data have been fuzzified based on the triangular function and sorted in ascending order. The concept of fuzzy logic [52] has been applied and estimated

ISSN 2229-5518

data have been furnished in table 5. The average error has been founded as 45.34%.

Table 5
Actual value, Estimated Value and Estimated Error

Serial No.	Available Data	Fuzzy input	Fuzzy Output	Estimated	Average Error (%)
1	857	[1, 0.0025, 0,0, 0]	[1 1 0.12 0.02 0.0025]	3686	330.11
2	1644	[1, 0.28, 0, 0, 0]	[1 1 0.28 0.27 0.04]	3686	124.21
3	1987	[1, 0.4, 0, 0, 0]	[1 1 0.4 0.27 0.04]	3686	85.51
4	2178	[1, 0.46, 0, 0, 0]	[1 1.46 0.27 0.04]	3686	69.24
5	2308	[1, 0.51, 0, 0, 0]	[1 1.51 0.27 0.04]	3686	59.71
6	4019	[0.88,1,0.12, 0, 0]	[0.88 1 1 0.27 0.12]	6522	62.28
7	4074	[0.86,1,0.14, 0, 0]	[0.86 1 1 0.27 0.14]	6522	60.09
8	4555	[0.69,1,0.31,0, 0]	[0.69 1 1 0.31 0.31]	6522	43.18
9	4792	[0.61,1,0.39, 0, 0]	[0.61 1 1 0.39 0.39]	6522	36.1
10	4976	[0.54 1,0.46, 0, 0]	[0.54 1 1 0.46 0.46]	6522	31.07
11	5071	[0.51,1,0.49, 0, 0]	[0.51 1 1 0.49 0.49]	6522	28.61
12	5325	[0.42 1,0.58, 0, 0]	[0.51 1 1 0.58 0.58]	6522	22.48
13	5398	[0.39,1,0.61, 0, 0]	[0.51 1 1 0.61 0.61]	6522	20.82
14	5473	[0.36 1,0.64, 0, 0]	[0.51 1 1 0.64 0.64]	6522	19.17
15	5647	[0.3, 1, 0.7, 0, 0]	[0.51 1 1 0.7 0.7]	6522	15.49
16	6031	[0.17,1,0.83, 0, 0]	[0.51 1 1 0.83 0.75]	6522	8.14
17	6304	[0.07 1,0.93, 0, 0]	[0.51 1 1 0.93 0.75]	6522	3.46
18	6423	[0.04, 1 0.96 0 0]	[0.51 1 1 0.96 0.75]	6522	1.54
19	6469	[0.02 1 0.98 0 0]	[0.51 1 1 0.98 0.75]	6522	0.82
20	7313	[0,0.73, 1, 0.27,0]	[0.51 0.96 1 1 0.75]	12194	66.74
21	7535	[0 0.64 1 0.36 0]	[0.51 0.96 1 1 0.75]	12194	61.83
22	7564	[0, 0.63, 1,0.37, 0]	[0.51 0.96 1 1 0.75]	12194	61.21
23	7748	[0, 0.56, 1, 0.44, 0]	[0.51 0.96 1 1 0.75]	12194	57.38
24	8725	[0,0.22,1 0.78, 0]	[0.22 0.96 1 1 0.75]	12194	39.76
25	9250	[0, 0.04, 1 0.96 0]	[0.02 0.96 1 1 0.75]	12194	31.83
26	11502	[0,0,0.25,1,0.75]	[0 0.25 0.78 1 1]	12194	6.02
27	11508	[0,0,0.24,1,0.76]	[0 0.24 0.78 1 1]	12194	5.96
28	12353	[0, 0, 0, 0.94, 1]	[0 0.04 0.78 0.94 1]	13612	10.19
29	12723	[0, 0, 0, 0.81, 1]	[0 0.04 0.78 0.81 1]	13612	6.99
30	15026	[0, 0, 0, 0.004, 1]	[0 0.004 0.78 0.24 1]	13612	-9.41

2.2 Implementation by Neural Network

It has been observed that the estimated error of fuzzy logic is high. Now it is necessary to minimize the error, therefore, the method of neural network has been applied on available data to minimize the error. The fuzzified input data has been fed into feed forward back propagation network. Here a 5 noded input layer, 5 noded output layer and 2 noded hidden layers have been used. The estimated data and estimated error (%) has been furnished in table 6. The average error has been founded as 7.16%.

Table 6
Actual data, Forecasted Data and Percentage Error based on FFBP Neural Network

Serial No.	Available Data	Estimated Data of NN	Average Error (%)
1	857		
2	1644	3686	124.21
3	1987	3686	85.51
4	2178	3686	69.24
5	2308	3686	59.71
6	4019	3686	-8.29
7	4074	5104	25.28
8	4555	5104	12.05
9	4792	5104	6.51
10	4976	5104	2.57
11	5071	5104	0.65
12	5325	5104	-4.15
13	5398	5104	-5.45
14	5473	5104	-6.74
15	5647	5104	-9.62
16	6031	5104	-15.37
17	6304	5104	-19.04
18	6423	5104	-20.54
19	6469	5104	-21.10
20	7313	5104	-30.21
21	7535	7940	5.37
22	7564	7940	4.97
23	7748	7940	2.48
24	8725	7940	-9.00
25	9250	7940	-14.16
26	11502	10776	-6.31
27	11508	10776	-6.36
28	12353	13612	10.19
29	12723	13612	6.99
30	15026	13612	-9.41

2.3 Implementation by Particle Swarm Optimization (PSO)

It has been observed that the estimated error of neural network is still high. Now it is necessary to minimize the error, therefore, the method of particle swarm optimization has been applied on available data to minimize the error. The concept and detail procedure has been furnished in paper [52]. The estimated data and estimated error have been furnished in table 7. The average error has been found as 8.856%.

Table 7
Actual data, Forecasted Data and Percentage Error based on PSO

Serial Number	Available Data	Estimated Data of PSO	Estimated Error (%)
1	857	5795	-3.91
2	1644	5432	-0.75
3	1987	6628	-9.37
4	2178	5973	-5.25
5	2308	6080	-6.01
6	4019	6050	-5.81
7	4074	9351	-18.70
8	4555	5546	-1.79

ISSN 2229-5518

9	4792	6792	-10.21
10	4976	9355	-18.71
11	5071	9904	-19.83
12	5325	6773	-10.11
13	5398	6911	-10.80
14	5473	7888	-14.72
15	5647	11642	-22.52
16	6031	7546	-13.51
17	6304	10145	-20.26
18	6423	3167	59.39
19	6469	2432	183.78
20	7313	5171	1.97
21	7535	2944	79.08
22	7564	4836	6.17
23	7748	4990	4.13
24	8725	5109	2.67
25	9250	5336	0.21
26	11502	4523	11.02
27	11508	3291	51.10
28	12353	4487	11.64
29	12723	3375	46.23
30	15026	5384	-0.26

3 Review of Result

The average error based on fuzzy, neural network and particle swarm optimization has been observed as 15.66%, 7.16% and 8.856 % respectively. Since neural network has given the minimum error on available data then it has been considered as preferable model for the data set. The validity of models has been cross check by the residual analysis (Sum of Absolute Residual (A), Maximum Residual (B), Mean Absolute Residual (C), Mean of Mean Absolute Residual (D), Median of Absolute Residual (E), and Standard Deviation of Absolute residual (F)). Here residual means relative difference between actual and expected. The residual analysis has been furnished in table 8.It has been observed that neural network has been preferred in five cases out of six whereas PSO preferred in one case. Therefore, it has verified our earlier conclusion.

Table 8
Residual Analysis

Serial No.	Method Name	A	B	C	D	E	F
1	Fuzzy logic	56176	4881	1379.34	45.97	5388	1380.64
2	Neural Network	28376	2829	931.57	31.05	875	680.8
3	PSO	126266	9642	2788.41	92.94	573	2309.04

4 Steps of Knowledge Discovery in Data Mining

Step 1:-

The main objective of data mining is to find hidden knowledge from the data base. Based on the estimated error, neural network has been preferable model as compared to others. Few unknown data have been taken from table 1 (as furnished in table 9) and total effect value has been calculated based on the formula as furnished in section 2(step 4).

Therefore, instructions as narrated in section 2.2 has been applied on tested data to calculate the estimated value. The estimated value has been furnished in table 10.

Table 9
Average data and Corresponding output

Serial Number	Tested Data	Corresponding Output
1	9532.79	311000
2	7655.11	311000
3	8235.72	311000
4	7910.18	311000
5	8281.84	311000
6	12138.13	311000

Table10
Tested Data and Output of NN

Serial Number	Tested Data	Output by NN
1	9532.79	10776
2	7655.11	7940
3	8235.72	7940
4	7910.18	7940
5	8281.84	7940
6	12138.13	10776

Step 2:-

The range has been calculated based on the output of neural network (table 6) and average of corresponding output value has been taken to assign the range value. As for an example, the range 3686-5104 of table (6), all data have the output 307000. Therefore its average value is 307000. That value has been termed as R.H.S for subsequent section. The range and corresponding value have been furnished in table 11.

Table 11
Range and Corresponding Value

Range	R.H.S

3686-5104	307000
5104-7940	291020
7940-10776	284015
10776-13612	274310

Step 3:-

From the table 11, it has been observed that if any value belongs to range 3686-5104 then it has corresponding output as 307000. Based on this concept, output of neural network (table 10) has been assigned its corresponding output. The output of neural network and assigned output have been furnished in table 12.

Table 12
NN and Corresponding Output

Serial Number	Tested Data	Output by NN	Corresponding Output
1	9532.79	10776	279163
2	7655.11	7940	287518
3	8235.72	7940	287518
4	7910.18	7940	287518
5	8281.84	7940	287518
6	12138.13	10776	279163

Step 4: -

The estimated error and estimated data based on the actual output has been furnished in table 13. The average error has been founded as 8.44%.

Table 13
Estimated Data and Estimated Data

Serial Number	Tested Data	Output by NN	Corresponding Output	Serial Number	Average Error (%)
1	9532.79	10776	279163*	311000	10.24
2	7655.11	7940	287518	311000	7.55
3	8235.72	7940	287518	311000	7.55
4	7910.18	7940	287518	311000	7.55
5	8281.84	7940	287518	311000	7.55
6	12138.13	10776	279163	311000	10.24

5 Conclusion

The said work has been undertaken on the available data of the Boston city. The application of rules as decided by factor analysis has been applied on new data set to form the total effect value for the new unknown data set and the algorithm of neural has been applied on the total effect value to form the estimated data. From previous range table (furnished in table 11), the necessary output of the estimated data by neural network may be decided, which is considered as it's predicted or inference output.

The instruction as narrated above has been used for the gathering of knowledge discovery (output extraction) for different unknown new data set. So in this ways, the extraction knowledge has been done from the data base.

If the said information is available in advance, necessary planning work can be decided by the Governments and various other agencies in the country in advance.

Acknowledgement

The authors express their deep sense of gratitude to the faculty members of the Department of Engineering & Technological Studies, University of Kalyani, West Bengal, India, where the work has been carried out. The work has been financially supported by DST, PURSE.

References:-

[1] J. E. Moreno, O. Castillo, J. Castro, "Data Mining for Extraction of fuzzy IF THEN Rules using Mamdani and Takagi-Sugeno-Kand FIS, " *Engineering Letters*, 15:1, EL-15-1-13, 2007.

[2] S. Dehuri, A.K.Jagadeva, A. Ghosh and R. Mall, "Multi Objective Genetic Algorithm for association rule mining using a homogenous dedicated cluster of Workstations," *American Journal of Applied Science*, vol. 3, no.1, pp. 2086-2095, Nov. 2006.

[3] P. Nagar and S. Srivastava, "Application of Genetic Algorithms in Data Mining," *Proceedings of 2nd National Conference on Challenges & Opportunities in Information Technology (COIT-2008) RIMT-IET*, pp. 52-55, March 29, 2008.

[4] H. Lu, R. Setiono, and H. Liu, "Effective Data Mining Using Neural Networks," *IEEE Transactions on Knowledge and data engineering*, vol. 8, no. 6, pp. 957-961, December 1996.

[5] R. S. Segall and Q. Zhang, "Applications of Neural Network and Genetic Algorithm Data Mining Techniques in Bioinformatics Knowledge Discovery-A Preliminary Study", pp. 278-285, 2006.

[6] B. Liu, M. Hu, W. Hsu, "Multi-level Organization and Summarization of the Discovered Rules", *Proceeding of 6th ACM SIGKDD International Conference on Knowledge Discovered & Data Mining (KDD– 2000)*, Boston, MA, USA, pp.208-217, 2000.

[7] Xianjun Ni, "Research of Data Mining Based on Neural Networks," *World Academy of Science, Engineering and Technology*, vol.39, pp. 381-384, 2008.

[8] Y. Dhanalakshmi and T. R. Babu, "Intrusion Detection Using Data Mining along Fuzzy Logic and Genetic Algorithms," *International Journal of Computer Science and Network Security*, vol. 8, no. 2, pp 27-32, February 2008.

[9] G.V.S.N.R. V. Prasad, Y. Dhanalakshmi, V. V. Kumar and I. R. Babu, "Mining for Optimized data using clustering along with fuzzy association rules and genetic algorithms," *International Journal of Artificial Intelligence and Applications (IJAIA)*, vol. 1, no. 2, pp 30-41, April 2010.

[10] R. Agrawal, T. Imielinski and A. Swami, "Mining Association Rules between Sets of Items in Large Databases," *Proceedings of ACM SIGMOD Conference on Management of Data*, Washington DC, pp. 207-216, May 1993.

[11] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases," Proceeding *of the 20th Intl. Conf. on Very Large Data Bases*, Santiago, Chile, pp. 487-499, Sep. 1994.

[12] R. Nag, L. V. S. Lakshmanan, J. Han, and A. Pang, "Exploratory mining and pruning optimizations of constrained association rules," *SIGMOD*, pp. l3-24, 98.

[13] S. Sarawagi, S. Thomas, and R. Agrawal, "Integrating association rule mining with relational database systems," *Alternatives and Implications in SIGMOD '98*, pp. 343-354, 98.

[14] R. Srikant, Q. Vu and R. Agrawal, "Mining association rules with item constraints," *in KDD*, pp. 67-73, 97.

[15] Xiyu LIU, Xinjiang XIE and W.WANG, "A Projection Clustering Technique Based on Projection," *Journal of Service Science and Management*, vol. 2, no. 4, pp. 362-367.

[16] Horatiu Mocian, "Survey of Distributed Clustering Techniques," *www.horatiumocian.com/papers/Distributed_Clustering_Survey.pdf*, October 29, 2010.

[17] S. Saha and S. Bandyopadhyay, "A Fuzzy Genetic Clustering Technique Using a New Symmetry Based Distance for Automatic Evolution of Clusters," *http:// www.isical.ac.in/~sriparna_r/p1092_camera.pdf*, March 5-7, 2007.

[18] Faruq A. Al-Omari and Nabeel I. Al-Fayoumi," An Image-Mapped Data Clustering Technique for Large Datasets," *International Conference on Computational Intelligence*, pp. 286-289, 2004.

[19] P. Pal and B. Chanda, "A Symmetry Based Clustering Technique for Multi-spectral Satellite Imagery," http://:www.ee.iitb.ac.in/~icvgip/PAPERS/252.pdf, January 1, 2003.

[20] D. Karaboga and B. Basturk, "On the performance of artificial bee colony ABC algorithm," *Applied Soft Computing Journal*, vol. 8, no. 1, pp. 687–697, 2008.

[21] D. Karaboga and B. Basturk, "Artificial bee colony ABC optimization algorithm for solving constrained optimization problems," *in Proceedings of the 12th International Fuzzy Systems Association World Congress on Foundations of Fuzzy Logic and Soft Computing (IFSA* '07), vol. 4529 of Lecture Notes in Artificial Intelligence, pp. 789–798, Springer, 2007.

[22] C. Ozturk and D. Karaboga, "A novel clustering approach: Artificial Bee Colony (ABC) algorithm," Proceedings *of Applied Soft computing of Science Direct*, vol. 11, no.1, pp. 652–657, January 2011.

[23] Wen Wang1, Yilong Lu1, Jeffrey S. Fu1, and Yong Zhong Xiong, "Particle Swarm Optimization and Finite-Element Based Approach for Microwave Filter Design," *IEEE Transactions on Magnetics*, vol. 41, no. 5, May 2005.

[24] Xiangyang Wang, Jie Yang, Xiaolong Teng, Weijun Xia and Richard Jensen," Feature selection based on rough sets and particle swarm optimization," *Pattern Recognition Letters 28*., vol. 4, pp. 459-471, 2007.

[25] J. Kennedy and R. Eberhart, "Particle swarm optimization," Proceedings *of the IEEE International Conference on Neural Networks*, pp. 1942–1948, December 1995.

[26] Z. Karimi, H. Abolhassani and Hamid Beigy, "A new method of mining data streams using harmony search," *Journal of Intelligent Information Systems*, vol. 39, no. 2, pp 491-511, October 2012.

[27] Jianyong Wang and George Karypis, "HARMONY: Efficiently Mining the Best Rules for Classification," *SIAM International Conference on Data Mining*, pp. 205-216, 2005.

[28] H. Mohamadi, J. Habibi, M. S. Abadeh and H. Saadi, "Data mining with a simulated annealing based fuzzy classification system," in *Science Direct*, vol. 41, no. 5, pp.1824–1833, May 2008.

[29] M. U. Shaikh, S. U. R. Malik, A. Qureshi and S. Yaqoob, "Intelligent Decision Making Based on Data Mining Using Differential Evolution Algorithms and Framework for ETL Workflow Management," *International Conference on Computer Engineering and Applications*, vol. 1, pp. 22-26, 2010.

[30] A. Daniel, D. Finke, J. Medeiros and Mark T. Traband, "Shop Scheduling using Tabu Search and Simulation" *Proceedings of the 2002 Winter Simulation Conference*, pp. 301-325, 2002.

[31] M. Hiroyuki, K. Noriyuki, I . S. Kenta and Kondo Tooru, "Application of Tabu Search to Data Mining Load Forecasting in Power Systems," *in Papers of Technical Meeting on Power Systems Engineering*, IEE Japan, **vol.** 1, **no.** 6, pp**.** 13-18, 2001.

[32] M. A. Tahir and J. Smith, "Improving Nearest Neighbor Classifier Using Tabu Search and Ensemble Distance Metrics," *International Conference on Data Mining, 2006, ICDM '06. Sixth Date of Conference*, pp. 1086-90, December 18-22, 2006.

[33] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series part I," *Fuzzy Sets and Systems*, vol. 54, pp. 1 – 9, 1993.

[34] H. Bintley,"Time Series analysis with REVEAL," *Fuzzy Sets and Systems.*, vol. 23, pp. 97-118, 1987.

[35] Q. Song and B. S. Chissom, "Forecasting enrollments with fuzzy time series - part II," *Fuzzy Sets and Systems*, vol. 62, pp. 1-8, 1994.

[36] G. A. Tagliarini, J. F. Christ and E. W. Page, "Optimization using Neural Networks," *IEEE Transactions on Computers*, vol.4, no.12, pp.1347-1358, December 1991.

[37] F. G. Donaldson and M. Kamstra, "Forecast combining with Neural Networks," *Journal of Forecasting*, vol.15, pp. 49-61, 1996.

[38] J. V. Hansen and R. D. Nelson, "Neural Networks and Traditional Time Series Methods: A Synergistic combination in state Economic Forecasts," *IEEE Transactions on Neural Networks*, vol.8, no. 4, pp. 863.73, July 1997.

[39] M. Ishikawa and T. Moriyama, "Prediction of Time Series by a Structural Learning of Neural Networks," *Fuzzy Sets and Systems*, vol. 82, pp. 167-176, 1996.

[40] S. M. Chen, "Forecasting enrollments based on fuzzy time series," *Fuzzy Sets and Systems*., vol. 81, pp. 311-319, 1996.

[41] D. P. Singh, J. P. Choudhury and Kalyan Chakrabarti, "Optimization using soft computing model," *Proceedings of 12th Annual Conference of Society of Operations Management, Indian Institute of Technology, Kanpur*, pp.33-34, December 2008.

[42] D. P. Singh, "Optimization using fuzzy logic membership function," *in DCIT-09 International Conference at Punjab College of Technical Education*, Ludhiana, Punjab, pp.36-46, May 2009.

[43] D. P. Singh, "Assessment of Exported Mango Quantity using fuzzy logic membership function," *in NCMTEE- 09, National Conference* at HETC, West Bengal, pp. 393-396, June 2009.

[44] D. P. Singh and J. P. Choudhury, "Assessment of Exported Mango Quantity By Soft Computing Model," *in IJITKM-09 International Journal*, Kurukshetra University, pp. 393-395, June-July 2009.

[45] D. P. Singh, J. P. Choudhury and Mallika De, "Performance Measurement of Neural Network Model Considering Various Membership Functions under Fuzzy Logic," *International Journal of Computer and Engineering*, vol. 1, no. 2, pp.1-5, 2010.

[46] D. P. Singh, J. P. Choudhury and Mallika De, "Prediction Based on Statistical and Fuzzy Logic Membership Function," *in PCTE, Journal Of Computer Sciences*, vol.8, no.1, pp. 86-90, June-July, 2010.

[47] D. P. Singh, J. P. Choudhury and Mallika De, "Optimization of Fruit Quantity by comparison between Statistical Model and Fuzzy Logic by Bayesian Network," *in PCTE Journal Of Computer Sciences*, vol.8, no.1, pp.90-96, June-July, 2010.

[48] D. P. Singh, J. P. Choudhury and Mallika De, "Performance measurement of Soft Computing models based on Residual Analysis," *in International Journal for Applied Engineering and Research, Kurukshetra University*, vol.5, pp 823-832, Jan-July, 2011.

[49] D. P. Singh, J. P. Choudhury and Mallika De, "Performance measurement of Soft Computing models based on Residual Analysis," *National Conference on Converging Technologies beyond 2020, (CTB-2020)*, pp 823-832, April 6-7, 2011.

[50] D. P. Singh, J. P. Choudhury and Mallika De, "A comparative study on the performance of Fuzzy Logic, Bayesian Logic and neural network towards Decision Making," *in International Journal of Data Analysis Techniques and Strategies*, vol.4, no.2, pp. 205-216, March, 2012.

[51] D. P. Singh, J. P. Choudhury and Mallika De, "Optimization of Fruit Quantity by different types of cluster techniques," *in PCTE Journal Of Computer Sciences*., vol.8, no.2, pp .90-96 , June-July, 2011.

[52] D. P. Singh, J. P. Choudhury and Mallika De, " A Comparative Study on the performance of Soft Computing models in the domain of Data Mining," *International Journal of Advancements in Computer Science and Information Technology*, vol. 1, no. 1, pp. 35-49, September, 2011.

[53] D. P. Singh, J. P. Choudhury and Mallika De," Comparative and Comprehensive Study to Select a Soft Computing Model in the Data Mining Domain," *in International Conference on Computation and Communication Advancement (ICA3)* , pp.1120.-116, Jan, 2013.

[54] D. P. Singh, J. P. Choudhury and Mallika De," A Comparative Study to Select a Soft Computing Model for Knowledge Discovery in Data Mining," *in International Journal of Artificial Intelligence and Knowledge Discovery*, Vol. 2, no. 2, pp. 6-19, April, 2012.

[55 D. P. Singh, J. P. Choudhury and Mallika De, "A Comparative Study to discover the knowledge in data mining by soft computing model," *in CCSN-2012 1st International Conference on Computing, Communication and Sensor Network*, pp. 42-46 Nov. 22-23, 2012.

[56] Export statement of APEDA. www.apeda.org.

Bibliography:-

Dharmpal Singh has received Bachelor of Computer Science & Engineering from West Bengal University of Technology and Master of Computer Science & Engineering also from West Bengal University of Technology. He has about six years of experience in teaching and research. At present he is with JIS College of Engineering, Kalyani, West Bengal, India as Assistant Professor. Now he is pursuing PhD in University of Kalyani. He has about 15 publications in National and International Journals and Conference Proceedings.

Dharmpal Singh

Jagannibas Paul Choudhury received his Bachelor of Electronics and Tele-Communication Engineering with Honours from Jadavpur University, Kolkata and Masters of Technology from Indian Institute of Technology Kharagpur. He received his PhD (Engineering) from Jadavpur University. He has about 24 years experience in teaching, research and administration. At present, he is with Department of Information Technology, Kalyani Government Engineering College, Kalyani, West Bengal, India as an Associate Professor and the Head in Information Technology. He has about 100 publications in national and international journals and conference proceedings. His research fields are soft computing, data mining, clustering and classification, routing a computer network, etc. He is a life member of Institution of Engineers, Institution of Electronics and Telecommunication Engineers, Computer Society of India and Operations Research Society of India.

Mallika De received her BSc in Physics from Calcutta University in 1973 and MSc in Applied Mathematics from Jadavpur University in 1976. She received her Advanced Diploma and MTech in Computer Science from Indian Statistical Institute, Calcutta in 1980 and 1985, respectively, and PhD in Engineering from Jadavpur University in 1997. She is currently the Head of the Department of Engineering and Technological Studies, University of Kalyani at West Bengal, India, where she is serving for the last 28 years as a faculty. Her research interests include parallel algorithms and architectures, fault-tolerant computing, image processing and soft computing. She has authored/co-authored 20 refereed journal articles and ten conference papers.