International Journal of Science and Research (IJSR)
ISSN (Online): 2319-7064
Index Copernicus Value (2013): 6.14 | Impact Factor (2015): 6.391

Effective Analysis of Multilayer Perceptron and Sequential Minimal Optimization in Prediction of Dyscalculia among Primary School Children

Sampada Margaj[1], Dr. Seema Purohit[2]

[1]Department of Computer Science, University of Mumbai, Kalina, Santacruz East, Mumbai, Maharashtra 400098

[2]Department of Mathematics, Kirti M. Doongursee College, Off. Veer Savarkar Marg, Dadar, Mumbai, Maharashtra 400028

Abstract: ***This study basically focuses on the two classification methods, Multilayer perceptron (MLP) and sequential minimal optimization (SMO), for the prediction of Dyscalculia among primary school children. Prediction of any of the categories of learning disability is not an easy task. Same is the case of dyscalculia. Detail knowledge of the subject is mandatory in accurate prediction of dyscalculia in any child. A sooner the detection faster we can overcome it which will help the child for bright future. Among above mentioned classifiers MLP gives us best accuracy results. This study will also reflect on determining the best classification method for our specific domain.***

Keywords: Dyscalculia,MLP, SMO, Classification

1 Introduction

The main aim of this work is to study the two classification methods, Multilayer Perceptron (MLP) and Support Vector Machines (SVM), for the prediction of Dyscalculia in school-age children. Prediction of any of the categories of learning disability is not an easy task. Same is the case of dyscalculia. Detail knowledge of the subject is mandatory in accurate prediction of dyscalculia in any child. The above two classifiers give us satisfactory results. This study will also reflect on determining the best classification method for our specific domain.

In our country there has not been much research and work in the field of learning disabilities. Also there isn't enough awareness or rehabilitation measures available for children with learning problems. Thus this study aims at targeting the children who are under tremendous pressure due to their bad performance in the school examinations. We have considered schools in and around Mumbai, which do not have computer facilities and other amenities unlike the private schools, where implementation and detection of dyscalculia is much easier.

Dyscalculia is a mathematical learning disorder where the mathematical ability is far below expected for a person's age, intelligence and education. Researchers have found evidence that such a disability exists and due to their findings there is a need to address dyscalculia as an important educational issue in mathematics.

In this study we have tested students of primary school with some specific attributes. Recent research has identified the heterogeneous nature of mathematical learning difficulties and, hence it is difficult to identify dyscalculia via a single diagnostic test. Diagnosis and assessment should use a range of measures, a test protocol, to identify which factors are creating problems for the learner. We have taken into consideration the views and ideas of certain doctors and teachers and accordingly prepared a questionnaire so that the students can be assessed on the different aspects of their mathematical abilities.

Thus on the basis of these scores we have trained the machine for prediction of dyscalculia on new sample test scores. The two classifiers MLP and SMO have given us some very good accuracy in prediction of the dyscalculia among children. Amongst the two of them MLP has given us a better accuracy in comparison to SMO.

In this section, we discuss the different literature surveys conducted in the field of learning disabilities and also various soft computing skills used for classification and predictions used along with the style of data processing mechanisms.

Researcher have studied about how to identify dyslexic students by using Artificial Neural Networks[1],the study actually proposed a systematic approach for identification of dyslexia and to classify important cases more accurately and easily by use of ANN. Tuang-Kuang also studied the similar application of Artificial Neural Network for the identification of students with learning disabilities [2],in this paper, they tried to adopt Artificial Neural Network technique, which has been applied successfully to solve problems in numerous fields, to the LD identification and diagnosis problem.

Study performed comparison of support vector machine and decision tree algorithms for the prediction of learning disabilities [3] with an emphasis on applications of data mining. In this study, Sequential Minimal Optimization algorithm is used in performing SVM and J48 algorithm is used in constructing decision trees.

Study focused on improving performance of Hybridized Predictors using suitable pre-processing techniques for the prediction of learning disability by Julie M. David [4]. In this research paper the relevance of various data pre-processing methods in classification is determined along with dimensionality reduction for the long list of attributes. The results obtained from this study have illustrated that the data pre- processing method has very good contribution in prediction system and capable of improving the performance of classifiers.

2 Proposed Work

2.1 Data Representation and Pre Processing

In this study we used 237 real world datasets from schools in and around Mumbai. The students were assessed on the parameters as discussed in the previous sections. The final sheet is arranged with names and the scores of all the subtests which are our attributes in this case. Below Table I has the list of attributes with their descriptions.

Table 1: List of Attributes

S. No	*Attribute*	*Signs & Symptoms of Dyscalculia*
1	DSR	Difficulty with Shape Recognition
2	DSD	Difficulty with Size Discrimination
3	DNA	Difficulty with Number Arrangements
4	DGS	Difficulty with Grouping Sets
5	DPV	Difficulty with Place Values
6	DNC	Difficulty with Numeric Calculations
7	DVA	Difficulty with Verbal Analysis
8	DCI	Difficulty with Counting Index

In this study we have used two different classifiers namely Multilayer Perceptron (MLP) classifier and Support Vector Machines (SVM) classifier for the comparison of Dyscalculia predictions.

2.2 Multilayer Perceptron Algorithm

Multilayer perceptron uses backpropagation to classify instances. This network can be created by an algorithm or built by hand or both. The network can be observed and modified during training time as well. The nodes in MLP network are all sigmoid.

A sigmoid function is a bounded differentiable real function that is defined for all real input values and has a positive derivative at each point and is mathematical function having an "S" shape (sigmoid curve). Defined by the formula -

$$S(t) = \frac{1}{1+e^{-t}} \qquad [11]$$

2.3 Sequential Optimization Method

SMO is basically a new form of SVM (Support Vector Machine) as SMO spends most of its time evaluating the decision function, rather than performing Quadratic Programming, it can exploit data sets which contain a substantial number of zero elements. SMO does particularly well for sparse data sets, with either binary or non-binary input data [10].

Sequential minimal optimization (SMO) is an algorithm used for solving optimization problems in minimum amount of time. Consider a binary classification problem with a dataset (x1, y1) ... (xn, yn), where xi is an input vector and yi∈ {-1, +1} is a binary label corresponding to it. The dual form of quadratic programming problem solved using support vector machine is as follows:

$$\max_{\propto} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} y_i y_j\, K(x_i x_j)\alpha_i \alpha_i,$$

subject to:

$0 \le \propto_i \le C,$ for i = 1, 2, …, n,

$$\sum_{i=1}^{n} y_i \propto_i = 0 \qquad [6]$$

where C is a Support Vector Machine hyper-parameter and K (xi, xj) is the kernel function, supplied by the user; and the variables are Lagrange multipliers.

SMO breaks the problem into a series of smallest possible sub-problems, which are then solved analytically. Since the linear equality constraint involving the Lagrange multipliers, the smallest possible problem involves two such multipliers. Then, for any two multipliers and, the constraints are reduced to:

$0 \le \propto_1, \propto_2 \le C,$

$$y_1 \propto_1 + y_2 \propto_2 = k, \qquad [7]$$

K,is the sum of the rest of terms in the equality constraint, which is fixed in each iteration.

The algorithm proceeds as follows:

- Find a Lagrange multiplier that violates the Karush–Kuhn–Tucker (KKT) conditions for the optimization problem.
- Pick a second multiplier and optimize the pair.
- Repeat steps 1 and 2 until convergence of multipliers
- The problem has been solved when all the Lagrange multipliers satisfy the KKT conditions within a user-defined tolerance

2.4 Measures used for performance evaluation

There are several different measures are used but we have considered whoever are appropriate for our dataset. Using these measures, the efficiency of classifiers is evaluated.

- **Classification Accuracy**

Classification results could have an error rate and it may fail to classify correctly. Classification accuracy can be calculated as follows:

$$Accuracy = \left(\frac{Instances\ \ Correctly\ \ Classified}{Total\ \ Number\ \ of\ \ Instances}\right) X\ 100\% \qquad [8]$$

- **Mean Absolute Error**

It is the average of difference between predicted and actual value in all test cases. The formula for calculating MAE is given in equation shown below:

$$MAE = \frac{(|a_1-c_1|+|a_2-c_2|+\cdots+|a_n-c_n|)}{n} \qquad [9]$$

Here, “a‟ is the actual output and “c‟ is the expected output.

- **Root Mean Square Error**

It is used to measure differences between values actually observed and the values predicted by the model. It is calculated by taking the square root of the mean square error as shown in equation given below:

$$RMSE = \frac{\sqrt{((a_1-c_1)^2+(a_2-c_2)^2+\cdots+(a_n-c_n)^2)}}{n} \qquad [10]$$

Here, "a" is the actual output and c is the expected output. The mean-squared error is the commonly used measure for numeric prediction.

- **Confusion Matrix**

A confusion matrix contains information about actual and predicted classifications done by a classification system.

3 Result and Discussion

In our study for the prediction of Dyscalculia and evaluation of the performances of both Multilayer Perceptron and Sequential Minimal Optimization we have used Waikato Environment for Knowledge Analysis (Weka).It is a popular suite of machine learning software written in Java, developed at the University of Waikato, New Zealand. It is free software licensed under the GNU General Public License.

Here we have checked the performance using the Training set itself and using different Cross validation and percentage split methods. The class (Dyscalculic & Non Dyscalculic) is obtained by considering the values of all the eight attributes.

3.1 Performance of Multilayer Perceptron (MLP) Classifier

The overall evaluation summary of Multilayer Perceptron Classifier (MPC) using training set and different cross validation methods is given in Table II. The classification summary of MPC for different percentage split is given in Table III. The confusion matrix for each different test mode is given in Table IIV to Table XIII. The chart showing the performance of Multilayer Perceptron Classifier with respect to Correctly Classified Instances and Classification Accuracy with different type of test modes are depicted in Figure 1, Figure 2 and Figure 3. Multilayer Perceptron gives 99.58% for the training dataset. But for evaluation testing with the test data is essential. So various cross validation and percentage split methods are used to test its actual performance. MLP outperforms than SMO during testing. On an average, it gives around 96% of classification accuracy for Dyscalculia prediction.

Table 2: MLP Classifier Overall Evaluation Summary

Test Mode	Correctly Classified Instances	Incorrectly Classified Instances	Accuracy	Mean Absolute Error	Root Mean Squared Error	Time Taken to Build Model(sec)
Training Set	236	1	99.58%	0.005	0.065	95.93
5 Fold CV	225	12	94.94%	0.0567	0.2116	95.38
10 Fold CV	228	9	96.20%	0.0445	0.1898	96.38
15 Fold CV	228	9	96.20%	0.0462	0.1928	104.18
20 Fold CV	225	12	94.94%	0.0531	0.2075	99.58
50 Fold CV	226	11	95.36%	0.0512	0.2098	100.88

Table 3: MLP Classifier Percentage Split Overall Evaluation Summary

Test Mode	Total Test Instances	Correctly Classified Instances	Incorrectly Classified Instances	Accuracy	Mean Absolute Error	Root Mean Squared Error	Time Taken to Build Model(sec)
66% Split	81	78	3	96.30%	0.0447	0.1793	101.01
33% Split	159	153	6	96.23%	0.0577	0.1956	101.05
75% Split	59	57	2	96.61%	0.0341	0.1573	100.21
80% Split	47	46	1	97.87%	0.0238	0.1428	97.16

Table 4. Confusion Matrix – MLP on Training Database

Class	Non Dyscalculic	Dyscalculic	Actual (Total)
Non Dyscalculic	205	0	205
Dyscalculic	1	31	32
Predicted (Total)	206	31	237

Table 5. Confusion Matrix – MLP for 5-fold CV

Class	Non Dyscalculic	Dyscalculic	Actual (Total)
Non Dyscalculic	202	3	205
Dyscalculic	9	23	32
Predicted (Total)	211	26	237

Table 6. Confusion Matrix – MLP for 10-fold CV

Class	Non Dyscalculic	Dyscalculic	Actual (Total)
Non Dyscalculic	203	2	205
Dyscalculic	7	25	32
Predicted (Total)	210	27	237

Table 7. Confusion Matrix – MLP for 15-fold CV

Class	Non Dyscalculic	Dyscalculic	Actual (Total)
Non Dyscalculic	204	1	205
Dyscalculic	8	24	32
Predicted (Total)	212	25	237

Table 8. Confusion Matrix – MLP for 20-fold CV

Class	Non Dyscalculic	Dyscalculic	Actual (Total)
Non Dyscalculic	202	3	205
Dyscalculic	9	23	32
Predicted (Total)	211	26	237

Table 9. Confusion Matrix – MLP for 50-fold CV

Class	Non Dyscalculic	Dyscalculic	Actual (Total)
Non Dyscalculic	203	2	205
Dyscalculic	9	23	32
Predicted (Total)	212	25	237

Table 10. Confusion Matrix - MLP for 66% Split

Class	Non Dyscalculic	Dyscalculic	Actual (Total)
Non Dyscalculic	71	1	72
Dyscalculic	2	7	9
Predicted (Total)	73	8	81

Table 11. Confusion Matrix - MLP for 33% Split

Class	Non Dyscalculic	Dyscalculic	Actual (Total)
Non Dyscalculic	139	1	140
Dyscalculic	5	14	19
Predicted (Total)	144	15	159

Table 12. Confusion Matrix - MLP for 75% Split

Class	Non Dyscalculic	Dyscalculic	Actual (Total)
Non Dyscalculic	50	0	50
Dyscalculic	2	7	9
Predicted (Total)	52	7	59

Table 13. Confusion Matrix - MLP for 80% Split

Class	Non Dyscalculic	Dyscalculic	Actual (Total)
Non Dyscalculic	42	0	42
Dyscalculic	1	4	5
Predicted (Total)	43	4	47

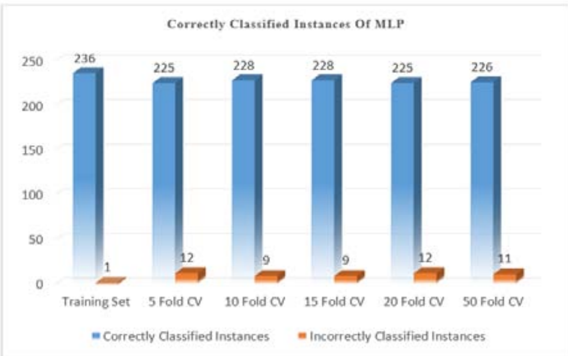

Figure 1. Correctly classified instances of Multilayer Perceptron

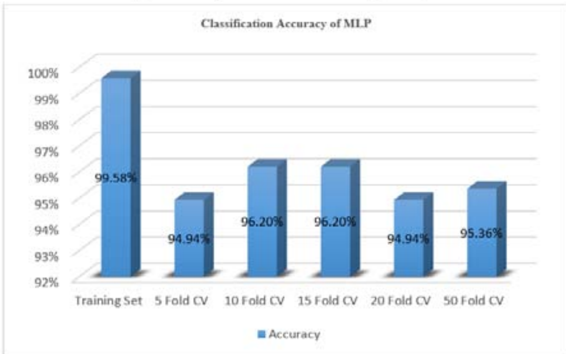

Figure 2. Classification Accuracy of Multilayer Perceptron

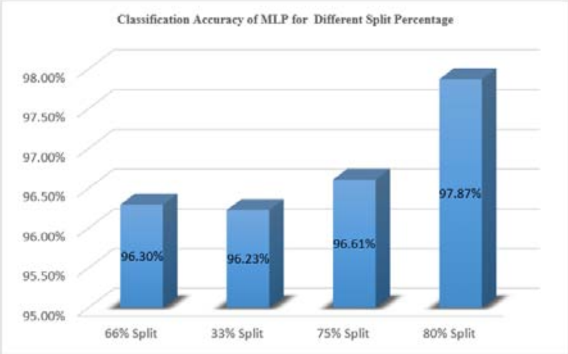

Figure 3. Classification Accuracy of Multilayer Perceptron Classifier for Different Split Percentage

3.2 Performance of Sequential Minimal Optimization (SMO) Classifier

The overall evaluation summary of Sequential Minimal Optimization (SMO) Classifier using training set and different cross validation methods is given in Table XIV. The classification summary of SMO Classifier for different percentage split is given in Table XV. The confusion matrix for each different test mode is given in Table XVI to Table XXV. The chart showing the performance of SMO Classifier with respect to Correctly Classified Instances and Classification Accuracy with different type of test modes are depicted in Fig. 4, Fig. 5 and Fig. 6. SMO classifier gives 99% for training data sets. But for testing various cross validation and percentage split methods, MLP gives a better output. On average, SMO Classifier gives around 95% of accuracy in prediction of Dyscalculia.

Table 14: SMO Classifier Overall Evaluation Summary

Test Mode	Correctly Classified Instances	Incorrectly Classified Instances	Accuracy	Mean Absolute Error	Root Mean Squared Error	Time Taken to Build Model(sec)
Training Set	236	1	99.58%	0.0042	0.065	0.07
5 Fold CV	226	11	95.36%	0.0464	0.2154	0.15
10 Fold CV	225	12	94.94%	0.0506	0.225	0.06
15 Fold CV	226	11	95.36%	0.0464	0.2154	0.03
20 Fold CV	225	12	94.94%	0.0506	0.225	0.05
50 Fold CV	225	12	94.94%	0.0506	0.225	0.05

Table 15: SMO Classifier Percentage Split Overall Evaluation Summary

Test Mode	Total Test Instances	Correctly Classified Instances	Incorrectly Classified Instances	Accuracy	Mean Absolute Error	Root Mean Squared Error	Time Taken to Build Model(sec)
66% Split	81	77	4	95.06%	0.0494	0.2222	0.05
33% Split	159	152	7	95.60%	0.044	0.2098	0.07
75% Split	59	54	5	91.53%	0.0847	0.2911	0.05
80% Split	47	46	1	97.87%	0.0213	0.1459	0.05

Table 16. Confusion Matrix – SMO on Training Database

Class	Non Dyscalculic	Dyscalculic	Actual (Total)
Non Dyscalculic	205	0	205
Dyscalculic	1	31	32
Predicted (Total)	206	31	237

Table 17. Confusion Matrix – SMO for 5 fold CV

Class	Non Dyscalculic	Dyscalculic	Actual (Total)
Non Dyscalculic	202	3	205
Dyscalculic	8	24	32
Predicted (Total)	210	27	237

Table 18. Confusion Matrix – SMO for 10 fold CV

Class	Non Dyscalculic	Dyscalculic	Actual (Total)
Non Dyscalculic	201	4	205
Dyscalculic	8	24	32
Predicted (Total)	209	28	237

Table 19. Confusion Matrix – SMO for 15 fold CV

Class	Non Dyscalculic	Dyscalculic	Actual (Total)
Non Dyscalculic	202	3	205
Dyscalculic	8	24	32
Predicted (Total)	210	27	237

Table 20. Confusion Matrix – SMO for 20 fold CV

Class	Non Dyscalculic	Dyscalculic	Actual (Total)
Non Dyscalculic	201	4	205
Dyscalculic	8	24	32
Predicted (Total)	209	28	237

Table 21. Confusion Matrix – SMO for 50 fold CV

Class	Non Dyscalculic	Dyscalculic	Actual (Total)
Non Dyscalculic	201	4	205
Dyscalculic	8	24	32
Predicted (Total)	209	28	237

Table 22. Confusion Matrix – SMO for 66% split

Class	Non Dyscalculic	Dyscalculic	Actual (Total)
Non Dyscalculic	69	3	72
Dyscalculic	1	8	9
Predicted (Total)	70	11	81

Table 23. Confusion Matrix – SMO for 33% split

Class	Non Dyscalculic	Dyscalculic	Actual (Total)
Non Dyscalculic	138	2	140
Dyscalculic	5	14	19
Predicted (Total)	143	16	159

Table 24. Confusion Matrix – SMO for 75% split

Class	Non Dyscalculic	Dyscalculic	Actual (Total)
Non Dyscalculic	49	1	50
Dyscalculic	4	5	9
Predicted (Total)	53	6	59

Table 25. Confusion Matrix – SMO for 80% split

Class	Non Dyscalculic	Dyscalculic	Actual (Total)
Non Dyscalculic	42	0	42
Dyscalculic	1	4	5
Predicted (Total)	43	4	47

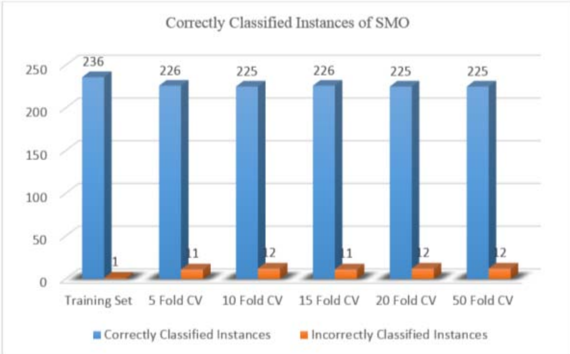

Figure 4. Correctly Classified Instances of SMO Classifier

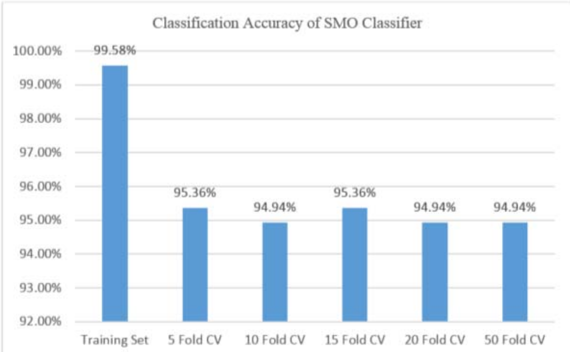

Figure 5. Classification Accuracy of SMO Classifier

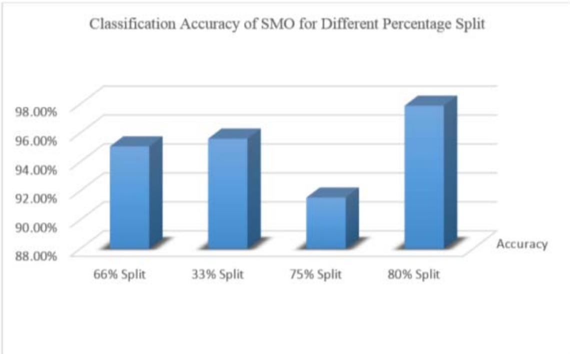

Figure 6. Classification Accuracy of SMO Classifier for different split percentage

4 Conclusion

In this study we analyzed the efficiency of two different classifiers namely, Multilayer Perceptron and Sequential Minimal Optimization (SMO) for the prediction of Dyscalculia among primary school children. The results obtained from these two classifiers help us to determine the relevance of quality of data as well as the significance of pre-processing in classification. Comparisons of both the classifiers have been done by considering different measures of performance evaluation. MLP takes more time to build model compare to other classifiers but in such crucial domain accuracy is considered to be more important factor than any other so that the right child should receive the right help. Overall, Multilayer Perceptron (MLP) performs better and gives maximum accurate results than Sequential Minimal Optimization (SMO) for the prediction of Dyscalculia.

References

[1] Ms.MaitreiKohli, Dr. T.V. Prasad, Member, IAENG:Proceedings of the World Congress on Engineering 2010 Vol IWCE 2010, June 30 - July 2, 2010, London, U.K. Identifying Dyslexic Students by Using Artificial Neural Networks

[2] Tung-Kuang Wu, Ying-Ru Meng ,Shian-Chang Huang:Application of Artificial Neural Network to the Identification of Students with Learning Disabilities .

[3] National Science Council of Taiwan, R.O.C. under Grant NSC 94-2520-S-018-004-

[4] Julie M. David et al, / (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 2 (2), 2011, 829-835

[5] Prediction of Learning Disabilities in School Age Children using SVM and Decision Tree

[6] Julie M. David, ISSN (Print): 2319-2526, Volume -4, Issue -5, 2015

[7] Improving Performance of Hybridized Predictors Using Suitable Preprocessing: Learning Disability Prediction

[8] John C. Platt, “Fast Training of Support Vector Machines using Sequential Minimal Optimization,” 2000.

[9] Germano C. Vasconcelos, Paulo J. L. Adeodato and Domingos S. M. P. Monteiro, “A Neural Network Based Solution for the Credit Risk Assessment Problem,” Proceedings of the IV Brazilian Conference on Neural Networks - IV CongressoBrasileiro de RedesNeurais pp. 269-274, July 20-22, 1999 - ITA, São José dos Campos - SP – Brazil.

[10] Tian-Shyug Lee, Chih-Chou Chiu, Chi-Jie Lu and I-Fei Chen, “Credit scoring using the hybrid neural discriminant technique,” Expert Systems with Applications (Elsevier) 23 (2002), pp. 245– 254.

[11] Zan Huang, Hsinchun Chena, Chia-Jung Hsu, Wun-Hwa Chen and Soushan Wu, “Credit rating analysis with support vector machines and neural networks: a market comparative study,” Decision Support Systems (Elsevier) 37 (2004) pp. 543– 558.

[12] John C. Platt, “Fast Training of Support Vector Machines using Sequential Minimal Optimization,” 2000

[13] Pang-Ning, T., Michael, S., Vipin, K: Introduction to Data Mining,Low price edn. Pearson Education, Inc., London, 2008

Author Profile

Sampada Margaj received the B.Sc. and M.Sc. degrees in Computer Science from University of Mumbai in 2002 and 2004, respectively. Further she did M.Phil. in Computer Science in 2008. She is carrying 12+ years of experience in teaching various subjects of Computer Science and Information Technology at graduate and postgraduate levels. Her expertise are in artificial intelligence and data mining.

Dr. Seema Purohit has a Post Graduate and Doctorate Degree in Computational Mathematics from University of Mumbai. With a rich experience of 25 years in the education field; she has an Administrative Experience of more than 14 years of heading the Department and an Institution. She has a unique combination of a Technological Expertise, Managerial and Operational Excellence by which understanding the needs of today’s Education System, IT Industries and Business Organizations becomes easy.She has also delivered several key note addresses and guest lecturers on the topics related to Technology and Management in many national and international conferences.