Predicting Student Performance for Early Intervention using Classification Algorithms in Machine Learning

Kalaivani K*
Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies, India
kalai.se@velsuniv.ac.in

Ulagapriya K
Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies, India
upriya.se@velsuniv.ac.in

Saritha A
Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies, India
saritha.se@velsuniv.ac.in

Ashutosh Kumar
Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies, India
ak214651@gmail.com

Abstract

Predicting Student's Performance System is to find students who may require early intervention before they fail to graduate. It is generally meant for the teaching faculty members to analyze Student's Performance and Results. It stores Student Details in a database and uses Machine Learning Model using i. Python Data Analysis tools like Pandas and ii. Data Visualization tools like Seaborn to analyze the overall Performance of the Class. The proposed system suggests student performance prediction through Machine Learning Algorithms and Data Mining Techniques. The Data Mining technique used here is classification, which classifies the students based on student's attributes. The Front end of the application is made using React JS Library with Data Visualization Charts and connected to a backend Database where all student's records are stored in MongoDB and the Machine Learning model is trained and deployed through Flask. In this process, the machine learning algorithm is trained using a dataset to create a model and predict the output on the basis of that model. Three different types of data used in Machine Learning are continuous, categorical and binary. In this study, a brief description and comparative analysis of various classification techniques is done using student performance dataset. The six different machine learning Classification algorithms, which have been compared, are Logistic Regression, Decision Tree, K-Nearest Neighbor, Naïve Bayes, Support Vector Machine and Random Forest. The results of Naïve Bayes classifier are comparatively higher than other techniques in terms of metrics such as precision, recall and F1 score. The values of precision, recall and F1 score are 0.93, 0.92 and 0.92 respectively.

1- Introduction

In India there are a large number of Universities which use the traditional methods to analyze the student's performance and find it difficult to manage hundreds of students and make use of their skills. There is a huge amount of data generated with student details and performance which can be used to improve the Education System like Identifying Problems, predict the performance, find out who needs an intervention etc.

Machine Learning is a subcategory of Artificial Intelligence, where a machine learns on its own from a given dataset without being programmed to make predictions. Machine Learning has become one of the most popular field of study today with a wide variety of applications in different domains. Machine Learning Algorithms can be used in recommendation system for customers, predicting stock prices or housing prices and clustering of customers etc., With the amount of data available today and the increasing computation speed of computers, the machine learning algorithms are able to tackle a variety of problems of high dimensional space. One of the best features of a machine learning algorithm is the ability to continually learn on its own and gradually increase its accuracy with time. If the prediction is not as expected, then the algorithm is re-trained multiple number of times until the desired output is found.

The Machine Learning process first requires a selection of algorithm and then a training data is used as input for

* Corresponding Author

the selected algorithm. If the output is unknown, it is called unsupervised learning and when the output is given it is called supervised learning. In supervised learning the algorithm takes both input and output for training and a model is generated which maps the inputs with the desired output. The model is then tested with a new set of input data where the predicted results are checked with the desired output.

Supervised learning algorithm can be further divided into two types: Regression and Classification. Both the algorithms can predict the output value from the given labeled input dataset with the only difference being that in regression the output variable is numerical and categorical in case of classification. Example for regression problem could be a situation where the output value is continuous such as salary, height or weight and age etc. In classification problems the output variable is a category such as yes or no, pass or fail etc. Some applications of classification algorithm are: Filtering of spam or not spam emails, student is going to pass the exam or not etc. [1].

H. Al-Shehri et al., used the popular dataset which consists of 395 data samples collected from the University of Minho in Portugal. They proposed that support vector machine-based prediction model accurately predicts the students' performance than K Nearest Neighbor [2]. S. Hossain et al., recommends that the Belief Rule based knowledge and Evidential Reasoning approach predicts thestudent performance with considering the personal and institutional parameters, they also stated that continuous performance analysis is necessary to find the skill and efficiency of students at various levels [3].E. S. Bhutto et al.,proposed that sequential minimal optimization algorithm greatly predicts the behavior of students by achieving improved accuracy than logistic regression. Their proposed system also suggests the measures to reduce the student's dropout ratio [4].

Ahammad, Khalil et al., stated that students who are at higher risk can be recognized by the use of machine learning models. They used the results of SSC exam and performed a comparative study by employing different machine learning techniques. They suggested that Multi-Layer Perceptron achieved higher accuracy for their chosen dataset and also showed that all other techniques were also yielded satisfactory accuracy in predicting the student's performance [5]. M. B. Shah et al., considered academic and other input variables like interests, attributes and opinions in predicting the performance. They explored various machine learning, deep learning modelsand basic exploratory data analysisto understand the correlations of student's performance using psychographic attributes [6].

Nurafifah Mohammad Suhaimiet al., proposed a model for the academic assessment to predict the student graduation time using Neural network and Support Vector Machine [7]. Fan Yang, Frederick W.B. Li, suggested that the Back Propagation based Neural Network outperforms

and they collected data from 60 schools to design the prediction model [8].Reynold A. Rustia et al., focused on a classification model using data mining techniques for predicting the probability of a student to pass the Licensure Examination for Teachers (LET) [9].

2- Methodology

Major Predicting students' performance from the current academic records is critical for adopting necessary pedagogical measures to make the students graduate on time and for satisfactory results. There are several challenges in predicting and to intervene for better performance. Some of the challenges are

- Students are from different backgrounds
- Lack of student evolving progress in making prediction.

The existing system is outdated and follows a traditional way of monitoring and generating student's reports

- Marks-based evaluation
- Exams have become "mugging-up" and memory tests
- Total reliance on pen and paper test
- Monitoring Hundreds of Students is difficult
- A weaker bond between assessment and learning outcomes

The drawbacks of the existing system are

- Focus was on rote learning and exams
- Missing of quick and timely Feedback on assessed work
- Not assessing Progress of Students or Analyzing Results or Scope of Improvement
- Project-based experimental learning is essential

To overcome these drawbacks, we make use of Classification Algorithms in Machine Learning to generate a model, which predicts the student's performance. There are a number of classification models and this paper describes and compares six different classification techniques with their advantages and disadvantages to analyse the student performance.

2-1- Logistic Regression

Logistic Regression can be used only when the output is categorical. It is similar to linear regression with a threshold. Based on the threshold value the classification is performed. Since Linear Regression cannot be used to solve classification problems, therefore in Logistic Regression, an activation function like sigmoid function to the Linear Regression model, which makes the value range from 0 to 1, can be added [10].

Advantage of Logistic Regression is that a threshold can be set once a value between 0 to 1 is arrived. For example, if there is a dataset to predict cancer is malignant or benign

based on its size and the predicted continuous value is 0.4 and the threshold value is 0.5, the data point will be classified as not malignant, which can lead to serious consequence. With Logistic Regression, a threshold value can be set as 0.6 or 0.7 accordingly. Logistic regression usually states where the boundary between the classes exists [1].

2-2- Naïve Bayes Classifier

Naïve Bayes is a classification technique based on Bayes Theorem. The classifier makes two assumptions: firstly, the attributes or features present in the dataset are independent and second is that each feature is given the same weight to predict the outcome. Naive Bayes model is easy to implement and works efficiently with both smaller and larger datasets. The dataset is divided into two parts, firstly input data which consists of the dependent features whose conditional probability is to be calculated based on the output class and the output data which contains the value of the class variable [11].

2-3- K-Nearest Neighbor

K-Nearest Neighbor is used to solve both classification and regression problems. KNN being a non-parametric technique, widely used in statistical estimation and pattern recognition. It is a lazy learning model, which is one of the easiest Machine Learning Classification techniques to implement. It first selects a group of labeled points and uses them to label other points. To predict a new point, it makes use of all the data points that are currently closest to that new point which are called the nearest neighbors and has those neighbors vote, here the "k" is the number of neighbors it checks. So whichever label has the most number of votes is the label for the new point [11].

K-Nearest Neighbor is also called a case-based learning method, where all the training data is used for classification. It is used in number of applications such as dynamic web mining, recommender systems etc. Further its efficiency can be improved with the help of some representatives to characterize the whole data i.e. implementing an inductive learning model from the training dataset and using that model for classification [12].

2-4- Support Vector Machine

A Support Vector Machine (SVM) is a classifier, which makes use of a separating hyperplane to classify points in higher dimensional space. In other words, when a labeled training dataset is given, the algorithm categorizes new points using an optimal hyperplane. In two-dimensional space the hyperplane can be visualized as a simple line, which divides the plane into two parts. Support Vector

Machine is very effective and can handle dataset with high dimensional spaces efficiently. It is used to classify non-linearly separable classes and this algorithm allows you to avoid overfitting due to its regularization parameter [13].

2-5- Decision Tree

A decision tree is a scenario-based tree structure, where the internal nodes represents a test on the attribute and each branch represents an outcome of the test, and each leaf node has a class label.

A Decision tree is easy to implement and the domain knowledge is not required for constructing decision tree classifiers. This classification algorithm can easily handle multidimensional data. The learning and classification phases of decision tree algorithms are simple and fast with promising accuracy. An attribute selection method is used to split the data based on the attribute selected; it divides the rows into distinct classes.

An attribute selection measure is a heuristic in choosing the splitting criterion to separate a given data partition. It suggests a ranking for each attribute describing the given training tuples. The splitting attribute for the given list of tuples attribute is selected based on the highest score. If the splitting attribute is continuous-valued or if it is restricted to binary trees, then a split point or a splitting subset must also be found respectively as a part of the splitting criterion. In the process of forming decision tress, most of the branches might expose noise or outliers in the training data. Tree pruning is done to find and remove those branches, in order to improve classification accuracy on unseen data.

One example of attribute selection measure is Information gain. Information Gain (IG) is normally used to measure the amount of information a feature provides about the class and can be calculated using the Eq. (1). While constructing a Decision Tree, Information gain is mainly used to find the key attribute, which splits the dataset into classes. Decision Tree algorithms constantly try to maximize the IG and an attribute with the highest IG will split first [2].

Information gain = entropy (parent) − [weights average] * entropy (children)   (1)

2-6- Random Forest

Random forest algorithm is an ensemble learning method for classification. This algorithm can be used to implement both classification and regression tasks. It uses various decision trees with different attributes as primary node. It is generally said to be a collection of decision trees and makes a mean prediction to the output. Random forests are better than decision trees as they avoid of over fitting to the training data and are very fast to train. Random Forest

algorithm involves two steps: first step is the creation of random forest and the second step is to make a prediction from the classifier [10][13].

3- Evaluation measures for Machine Learning Algorithms

There are number of measures or metrics [14][15] to evaluate the performance of the classification algorithms and they are described as follows:

Mean, Median and Mode
- Mean can be obtained by calculating the ratio between the sum of all values and the total number of observations.
- Mode is obtained by evaluating the most occurring value in the sample.
- Median is obtained by first sorting the numbers in increasing order and finding the middle value.

Variance
Variance is used to show the dispersion of the values around the mean. To find the variance, the first step is to calculate the mean and then sum the square difference between each value and finally divide the total number of observations.

Standard Deviation
Standard deviation gives us the variation of the values. It is obtained by taking the Square root of the variance.

Correlation
Correlation is used to find how the attributes are related to each other and how they contribute to the outcome. It is a measure of the relationship between attributes. Correlation value ranges between -1 and 1 where -1 specifies that the variables are negatively correlated and +1 confirms that the variables are positively correlated. 0 specifies that there is no correlation midst the target variables.

R-Squared
R-Squared is used to measure explained variation over total variation. Formula to calculate R squared is :
R squared = 1 - (Sum of Squared Residuals / Total Sum of Squares)

Confusion Matrix
Confusion matrix is in the form of a table that contains the results of classification algorithm. It is formed after

prediction is performed on the test set and the actual true values are known.

Precision and Recall
Precision is used to identify the relevant instances from the model and can be calculated as Eq. (2). It is given by:
Precision = true positives / (true positives + false positives)   (2)
Recall is used to identify all the relevant cases within a dataset and can be calculated as Eq. (3). It is given by:
Recall = true positives / (true positives + false negatives)
(3)

F1 Score
F1 uses both recall and precision in other words it provides a single metric that combines recall and precision using the harmonic mean. The results range between 1 and 0 where the values that are closer to 1 are considered the best whereas those, which lie towards 0, are considered as the worst.

Mean Absolute Error
Mean Absolute Error is used to find the average of the difference between the actual values and the predicted values. It is used to measure the distance between the predictions and the actual output [15].

Mean Squared Error
Mean Squared Error is used to find the average of the square of the dissimilarity between the original and the predicted values. It is similar to Mean Absolute Error but here the squared values are considered due to which the computation of gradient is easier [15].

Table 1: Advantages & Disadvantages of various Classification Algorithms

Machine Learning Algorithm	Advantages	Disadvantages	Applications
Logistic Regression	• Interpretability • Little tuning	• Correlated features may affect performance • Extensive feature engineering required	• Risk Assessment • Fraud Prevention
Naïve Bayes Classifier	• It performs well with categorical variables • Good when dataset contains several features	• Correlated features affect performance	• Sentiment Analysis • Document categorization
K-Nearest Neighbor	• Good performance	• Slow when predicting • Susceptible to high dimension (lots of features)	• Gene expression • Protein-protein interaction
Support Vector Machine (SVM)	• Good for datasets with more variables than observations • Good performance	• Long training time required • Tuning is required to determine which kernel is optimal for non-linear SVMs	• Stock market forecasting • Value at risk determination
Decision Trees & Random Forests	• Render feature importance • Less data pre-processing required • Does not overfit (in theory) • Good performance and accuracy • Robust to noise	• It does not predict beyond the range of the response values in the training data • Biased towards categorical variables with several categories	• Credit Risk Assessment • Assess probability of developing a chronic disease (healthcare)

Table 2: Comparison of various classification algorithms on training speed, feature scaling, missing data and outliers

Machine Learning Algorithm	Average Predictive Accuracy	Training Speed	Feature Scaling	Missing Data	Outliers
Logistic Regression	Medium	High	YES	Sensitive	Potentially sensitive
Naïve Bayes	Medium	High	NO	Can handle missing data	Robust to outliers
KNN	Medium	Low	YES	Sensitive	Sensitive
SVM	High	High	YES	Sensitive	Robust to outliers
Decision Tree	Medium	Medium	NO	Can handle missing data	Sensitive
Random Forest	High	Low	NO	Can handle missing data	Robust to outliers

	sex	failures	paid	activities	absences	G1	G2	G3
0	F	0	no	no	4	5	5	6
1	F	3	yes	no	10	7	8	10
2	F	0	yes	yes	2	15	14	15
3	F	0	yes	no	4	6	10	10
4	M	0	yes	yes	10	15	15	15
5	M	0	no	no	0	12	12	11
6	F	0	no	no	6	6	5	6
7	M	0	yes	no	0	16	18	19
8	M	0	yes	yes	0	14	15	15
9	F	0	yes	no	0	10	8	9
10	F	0	no	yes	4	10	12	12
11	M	0	yes	yes	2	14	14	14
12	M	0	yes	no	2	10	10	11
13	M	0	no	no	0	14	16	16
14	F	0	no	no	4	14	14	14
15	F	0	yes	yes	6	13	14	14

Fig. 1: Student Performance Dataset

4- Results and Discussion

The advantages, disadvantages and appropriate applications of different machine learning algorithm is shown in table 1. An analysis in terms of training speed, feature scaling, missing data and outliers is shown in table 2. Six different classification algorithms were employed on the Student performance dataset, taken fromUCI Machine Learning [16] – [21]

to analyze the performance. The dataset includes the student achievement in secondary education of two Portuguese schools. The data attributes include student grades, demographic, social and school related features) and it was collected by using school reports and questionnaires. The output is binary i.e. Pass (1) or Fail (0). The dataset contains the following information:

- Total number of students: 395
- Number of students who passed: 265
- Number of students who failed: 130

A comparison is done on student performance dataset with following 8 feature columns:

- sex: Male or Female (binary)

- failures: number of failures in previous classes or semesters (numeric)
- paid: enrolled in extra paid online courses (binary: yes, or no)
- activities: participating in extra-curricular activities (binary: yes, or no)
- absences: number of days absent in class (numeric: from 0 to 93)
- G1 G2 G3: Internal marks of students (numeric)
- passed: final result of exam (binary: yes, or no)

precision, recall and f1-score were chosen to test the accuracy of the model as these measures works best on classification problems. These metrics clearly shows how effectively the model has predicted or classified the test data [22 – 26].
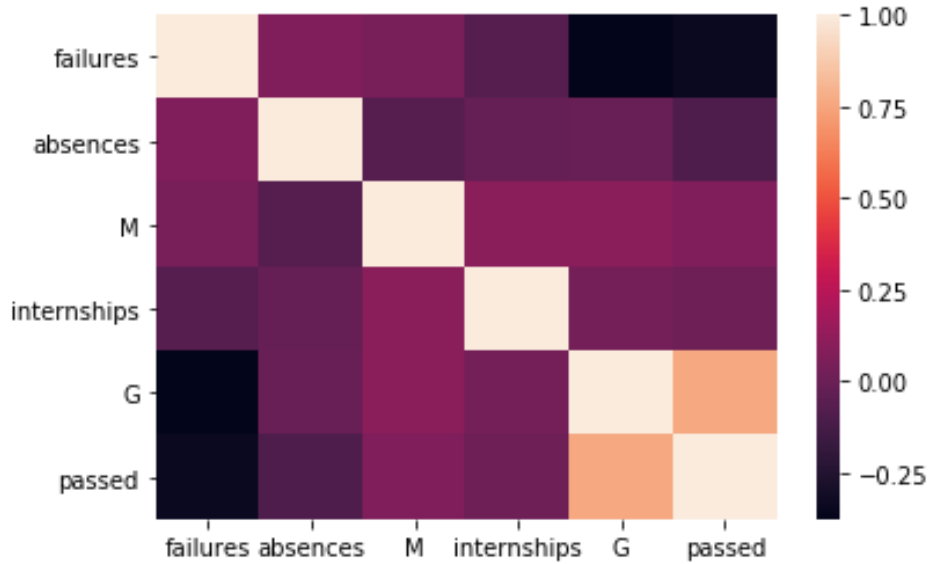

Fig. 2: Heat map of the Correlation Function

Table 3: Test Results of the Machine Learning Model

Machine Learning Algorithm	Precision	Recall	F1-Score
Logistic Regression	0.78	0.78	0.77
Naïve Bayes	**0.93**	**0.92**	**0.92**
KNN	0.81	0.73	0.65
SVM	0.91	0.91	0.91
Decision Tree	0.88	0.88	0.88
Random Forest	0.92	**0.92**	**0.92**

```
print(confusion_matrix(y_test,y_pred))
print('\n')
print(classification_report(y_test,y_pred))

[[19  6]
 [ 0 54]]


             precision    recall  f1-score   support

          0       1.00      0.76      0.86        25
          1       0.90      1.00      0.95        54

avg / total       0.93      0.92      0.92        79
```

Fig. 3: Test Results of Naïve Bayes Classifier

From the performance of various classification algorithms and the results achieved on Student performance dataset, it is evident that the performance of random forest and Naïve Bayes classifier is same in terms of recall and f1 score. The precision value of Naïve Bayes is slightly better than random forest. The Naïve Bayes and Random Forests are good at handling missing data and outliers.

5- Conclusions

In this paper, an analysis on student performance using different Classification Algorithms is done, to show whether the student will pass or fail based on the chosen attributes. The input data was trained on six classification algorithms and their test results were compared using metrics such as accuracy, precision, recall and F1 score. These measures were used to evaluate the accuracy of the classification models and as a result Naïve Bayes algorithm was found to be more effective. The study clearly shows that classification algorithm behaves differently with different attributes. Decision tree Algorithm shows high precision when the attributes are of binary type and not continuous and are prone to overfitting of data. Tree pruning process needs to be performed to avoid overfitting of data. Support Vector Machine, Naïve Bayes and Random Forests algorithms perform with high accuracy and precision regardless of the number of attributes.

The analysis was done on the merits and demerits of different algorithms at varied situations to understand their efficiency. After a better understanding of these algorithms for future prospects, how two algorithms can be combined together with their strengths and weaknesses should be investigated.

References

[1] F.Y. Osisanwo, J.E.T.Akinsola, O. Awodele, J.O. Hinmikaiye, O. Olakanmi, and J. Akinjobi, "Supervised Machine Learning Algorithms: Classification and Comparison", International Journal of Computer Trends and Technology, Vol. 48, No. 3, 2017, pp. 128-138.

[2] H. Al-Shehri et al., "Student performance prediction using Support Vector Machine and K-Nearest Neighbor," in IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE), 2017, pp. 1-4.

[3] S. Hossain, D. Sarma, F. Tuj-Johora, J. Bushra, S. Sen and M. Taher, "A Belief Rule Based Expert System to Predict Student Performance under Uncertainty", in 22nd International Conference on Computer and Information Technology (ICCIT), 2019, pp. 1-6.

[4] E. S. Bhutto, I. F. Siddiqui, Q. A. Arain and M. Anwar, "Predicting Students' Academic Performance Through Supervised Machine Learning",in International Conference on Information Science and Communication Technology (ICISCT), 2020, pp. 1-6.

[5] Khalil Ahammad, Partha Chakraborty, Evana Akter, Umme Fomey, and Saifur Rahman, "A Comparative Study of Different Machine Learning Techniques to Predict the Result of an Individual Student Using Previous Performances", International Journal of Computer Science and Information Security, Vol. 18, No. 1, 2021, pp. 5-10.

[6] M. B. Shah, M. Kaistha and Y. Gupta, "Student Performance Assessment and Prediction System using Machine Learning", in 4th International Conference on Information Systems and Computer Networks (ISCON), 2019, pp. 386-390.

[7] Nurafifah Mohammad Suhaimi, ShahAlam, Selangor, Shuzlina Abdul-Rahman, Sofianita Mutalib, Nurzeatul Hamimah Abdul Hamid, Ariff Md Ab Malik, "Review on Predicting Students', Graduation Time Using Machine Learning Algorithms", International Journal of Modern Education and Computer Science, Vol. 7, 2019, pp. 1-13.

[8] Fan Yang, Frederick W.B. Li, "Study on student performance estimation, student progress analysis, and student potential prediction based on data mining", Computers & Education, Vol. 123, 2018, pp. 97-108,

[9] Reynold A. Rustia, Ma. Melanie A. Cruz, Michael Angelo P. Burac, Thelma D. Palaoag, "Predicting Student's Board Examination Performance using Classification Algorithms", in 7th International Conference on Software and Computer Applications, 2018, pp. 233–237.

[10] Jiawei Han, Micheline Kamber, Jian Pei, "Data Mining Concepts and Techniques", USA: Morgan Kaufmann Publishers Elsevier, 2012.

[11] Saishruthi Swaminathan, "Logistic Regression – Detailed Overview", Canada: Towards Data Science – Medium Publication, 2018.

[12] Mandy Sidana, "Intro to types of Classification Algorithms in Machine Learning", USA: Medium Publication, 2017.

[13] Gongde Guo, Hui Wang, David Bell, Yaxin Bi, Kieran Greer, "KNN Model-Based Approach in Classification", Lecture Notes in Computer Science, Vol. 2888, 2003, pp. 986-996.

[14] Simon Tong, Daphne Koller, "Support Vector Machine Active Learning with Applications to Text Classification", Journal of Machine Learning Research, Vol. 2, No.1, 2001, pp. 45-66.

[15] H Zhang, J Zimmerman, D Nettleton, DJ Nordman, "Random forest prediction intervals", The American Statistician, Vol.74, Issue 4, 2019, pp. 392-406.

[16] Farhad Malik, "Must Know Mathematical Measures for Every Data Scientist", USA: Medium Publication, 2018.

[17] R.J. Howarth, "r2 (R Squared)", Dictionary of Mathematical Geosciences: With Historical Notes, Switzerland: Springer, 2017, pp. 503-527

[18] Aditya Mishra, "Metrics to Evaluate your Machine Learning Algorithm", Canada: Towards Data Science, Medium Publication, 2018.

[19] Haiyi Zhang, Di Li, "Naïve Bayes Text Classifier", in IEEE International Conference on Granular Computing, 2007, pp. 708.

[20] Nawzat Sadiq Ahmed, Mohammed Hikmat Sadiq, "Clarify of the Random Forest Algorithm in an Educational Field", in IEEE International Conference on Advanced Science and Engineering, 2018, pp.179-184.

[21] Dataset,https://archive.ics.uci.edu/ml/datasets/Student +Performance

[22] Sotiris Kotsiantis, "Supervised Machine Learning: a review of classification techniques", Informatica, Vol. 31, No. 3, 2007, pp. 249-268.

[23] S.B.Kotsiantis, I.D. Zaharakis, P.E.Pintelas, "Machine Learning: A Review of Classification Techniques and Combining Techniques", Artificial Intelligence Review, Vol. 26, No. 3, 2006, pp.159-190.

[24] R. Choudhary and H. K. Gianey, "Comprehensive Review on Supervised Machine Learning Algorithms," in IEEE International Conference on Machine Learning and Data Science (MLDS), 2017, pp. 37-43.

[25] C. A. Ul Hassan, M. S. Khan and M. A. Shah, "Comparison of Machine Learning Algorithms in Data classification," in 24th International Conference on Automation and Computing (ICAC), 2018, pp. 1-6.

[26] O. Obulesu, M. Mahendra and M. Thrilok Reddy, "Machine Learning Techniques and Tools: A Survey," in International Conference on Inventive Research in Computing Applications (ICIRCA), 2018, pp. 605-611.

Kalaivani K is an Assistant Professor in the Department of Computer Science and Engineering, Vels Institute of Science, Technology and Advanced Studies, Chennai, India. She received her B.E degree in Computer Science and Engineering from VIT University, India in 2001, and her M.E degree in Computer Science and Engineering from Vels Institute of Science, Technology and Advanced Studies, Chennai, India in 2012. She has teaching experience of more than 10 years and her research interests include Image Processing, Machine Learning and Big Data Analytics.

Ulagapriya K is an Assistant Professor in Department of Computer Science and Engineering in the School of Engineering at Vels Institute of Science, Technology and Advanced Studies, Chennai, India. She is having 12 years of IT experience and 8 years of teaching experience. Her research interests include Machine learning and Data Analytics.

Saritha A is an Assistant-Professor in Department of Computer Science and Engineering in the School of Engineering at Vels Institute of Science, Technology and Advanced Studies, Chennai, India. She is in teaching for more than 8 years and currently doing research in the area of Mobile Adhoc Networks.

Ashutosh Kumar is a UG scholar of Computer Science and Engineering in the School of Engineering at Vels Institute of Science, Technology and Advanced Studies, Chennai, India. His interests include Web Application and Machine Learning.