



# A study and review of applications of Machine Learning Classification Algorithms in Education Sector

**Prof. Vidhya Rao**

Assistant Professor

SIES College of Management Studies, Nerul, Navi Mumbai

**Abstract:** *The term machine learning is the study of algorithms that the computer systems use to perform tasks without explicitly instructions. It is a subset of Artificial Intelligence. The machine takes large amount of data, gets trained itself and learns from the data. It predicts outcomes after it has learnt from the past data. Machine learning algorithms are classified as supervised, unsupervised, semi-supervised and reinforcement types. The classification and regression algorithms fall under the supervised category. Clustering and association are grouped under the unsupervised learning. The main objective of this paper is to emphasize the types and study of machine learning classification algorithms. Educational sector produces huge amount of educational data. This data can be used for various purposes such as predicting the performance, dropout rates, retention, failure of students and evaluating their progress. This paper accumulates and refers to the literature related to machine learning classification algorithms, identifies implementation of machine learning classification algorithms and their applications in education domain.*

**Keyword:** Machine Learning, Education domain, Classification algorithms

## 2. INTRODUCTION

### 2.1 Data and Applications of Machine learning and its importance in education domain:

A collection of facts, such as values or measurements, observation or even just descriptions of things is referred to as data. Data collection is very important in any type of research study [30]. In the year, 2008, the world's servers process 9.57 zettabytes (9.57 trillion gigabytes) of information. Eric Schmidt, executive chairman of Google, tells as much data is now being created every two days, as was created from the beginning of human civilization to the year 2003 [26]. Wal-Mart company uploads 20 million point-of-sale transactions to an A&T massively parallel system with 483 processors running a centralized database (Alexander). Machine learning enables analysis of massive quantities of data. The system is trained using a large volume of data. Also, it acts based on experience. A report by Zion Market Research projects that spending around machine learning is growing sharply, from about \$1.58 billion in 2017 to \$20.83 billion in 2024 [15].

Machine learning is an application of Artificial Intelligence (AI) in which the systems have the ability to learn and get trained without being programmed explicitly. Machine learning is a instrument for turning information into knowledge. Since the machines have to undergo learning process and perform with different rules and learn from the inputted data that is why it's known as Machine Learning.

The applications of machine learning include self-driving cars, practical speech recognition, effective web search engine and page ranking, automatic friend tagging suggestions, transportation services like Uber and Ola, product recommendations like in Amazon, text to speech conversion, language translators, dynamic pricing to determine the best competitive prices, online video streaming like netflix, fraud detections, spam Detector that classifies emails as spam and moves the spam mails to spam folder, Traffic Alerts in google Maps, making predictions across varied domains, personalized medical treatment, sentiment analysis to determine the opinion of the speaker. Machine Learning and AI has been accelerating in many domains business, banking, communication, travel, health.

There has been a expeditious growth in higher education system. In a research conducted by McKinsey & Company, the potential economic value that might be expected from widespread use of big data could enable approximately \$890 billion to \$1.2 trillion in value annually [27]. According to Market Research Future (MRFR), there will be 38% i.e.

USD 2 billion dollars of growth in the education market by 2023. Almost all the countries will adopt the new and upgrading techniques of AI and ML [38]. There is lot of potential for research in the area of Machine Learning.

With the emergence of new academic institutions, there has been an increase in specific types of educational data sets that come from education sector. An effective utilization of machine learning techniques can assist the stakeholders in the educational sector to discover and analyze educational data and gain knowledge to make better academic decisions.

Machine learning in education sector is radically changing. The applications of machine learning in this sector include adaptive learning in which a student's performance is analyzed in real-time and thereby modifies teaching methods and the curriculum based on that data, predictive analytics like student's performance based on past data, personalized learning where the students can guide their own learning at their own pace, evaluating assessments where the machine is capable of grading students' assignment and exams are some of the applications that make use of machine learning.

## 2.2 CATEGORIZATION OF MACHINE LEARNING ALGORITHMS:

### *Supervised machine learning algorithms:*

The algorithms use known and labelled input data. These algorithms learn from the training dataset. The algorithm iteratively makes predictions on the training data. Learning stops when the algorithm achieves an acceptable level of performance. Classification and Regression are types of supervised learning.

### *Unsupervised machine learning algorithms:*

The algorithms use unknown input data and unlabeled data. The task of the algorithms is to group unsorted information according to similar characteristics, behaviors, patterns without any prior training of data. Clustering and Associations are types of unsupervised learning.

## 3. OBJECTIVES OF THE STUDY

1. To list popular machine learning classification algorithms.
2. To review earlier research work and find out the experimental research undertaken in the education domain using machine learning classification algorithms.

### 3.1. MACHINE LEARNING CLASSIFICATION ALGORITHMS:

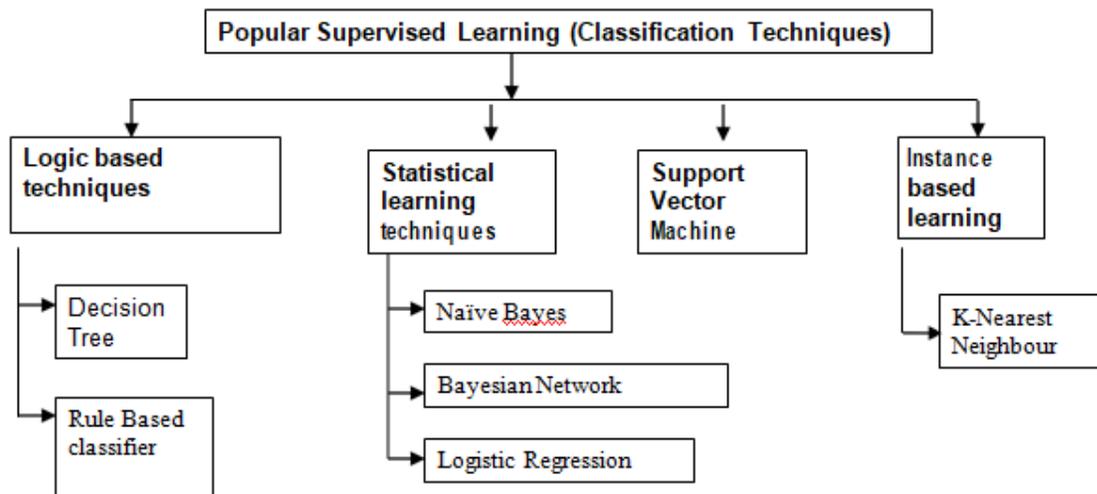


Figure 1: Supervised learning classification techniques



#### **Decision tree:**

A decision tree is a tree-like structure. It uses a greedy, top-down recursive partitioning strategy. Each internal node denotes an attribute and each leaf node represents a class label. The class label represents the decision taken after computing all features. Decision Trees are non-parametric meaning no prior assumptions are required regarding the type of probability distributions satisfied by the class and other attributes. They are used for both classification and regression tasks. In classification tree, the decision variable is categorical or discrete. In regression tree, the decision variable is continuous typically real numbers. The advantages of decision tree are it does not require data to be normalized, scaled, missing values does not affect the building decision tree considerably and can be simplified by visualization. The disadvantages of decision trees are calculations can become complex when there are many class labels, high probability of overfitting, involves higher time to train the model, training is relatively expensive as complexity and time taken is more. The algorithms used for decision tree are ID3, C4.5, C5.0 and CART.

#### **Rule based classifier:**

Rule-Based Classifier classifies records by using a collection of “if...then...” rules. Rules are easier to understand than large trees. One rule is created for each path from the root to a leaf. Rules and rule based classification systems have been widely applied in various expert systems, such as fault diagnosis for aerospace and manufacturing, medical diagnosis, highly interactive or conversational Q&A system, mortgage expert systems etc [25]. Rules can directly extract rules from data using algorithms like 1R, RIPPER(uses General-to-Specific Strategy), CN2(uses Specific-to-General Strategy) PRISM or from indirect method using decision trees [46]. The advantages of rule-based are they are easy to interpret, generate and can classify new instances rapidly. Other advantages of such rule languages include readability, and maintainability and foremost the possibility of directly transferring domain knowledge into rules [44]. The limitations of rule-based approach is that it is difficult and tedious to list all the rules. Rules can be arbitrarily complex. They can be memory intensive and computationally intensive. Sometimes they can be difficult to debug.

#### **Bayes Network:**

Bayesian classifiers are statistical classifiers. Bayesian classifiers have also exhibited high accuracy and speed when applied to large databases. Naïve Bayesian classifiers assume that the effect of an attribute value on a given class is independent of the values of the other attributes. This assumption is called class conditional independence [28]. For example, suppose a dataset comprises of columns named season (sunny or rainy), temperature (hot, mild and cold) and windy (yes, no). The temperature being ‘hot’ has nothing to do with the humidity or the season being ‘rainy’ has no effect on the winds. Hence, the features are assumed to be independent. Secondly, each feature or attribute is given the same weight (or importance) and all attributes are assumed to be contributing equally to the outcome. The advantage of Naïve Bayes classifier is that the training is fast, and consists of computing the priors and the likelihoods. Prediction on a new data point is quick. It easily handles missing feature values — by re-training and predicting without that feature [9]. For problems with a small amount of training data, it can achieve better results than other classifiers. It handles continuous and discrete data. The disadvantage of the Naive Bayes Algorithm is that it needs to calculate the prior probability. It is very sensitive to the form of input data. Due to the assumption of sample attribute independence and if the sample attributes are related, the effect is not good [45].

#### **BAYESIAN NETWORK:**

Bayesian network (BN), sometimes referred to as Bayes nets, belief networks and sometimes causal networks. It is a type of probabilistic graphical model that uses Bayesian inference for probability computations. Bayesian networks are graphical models, which unlike naïve Bayesian classifiers allow the representation of dependencies among subsets of attributes [28]. They can be used for a wide range of tasks including prediction, anomaly detection, diagnostics, automated insight, reasoning, time series prediction and decision making under uncertainty [5]. Compared to decision trees, Bayesian networks are usually more compact, easier to build, and easier to modify. The main weakness is that Bayesian networks require prior probability distributions. The graphical nature of a BN clearly displays the links between different system components. This can facilitate discussion of the system structure with people from a wide variety of backgrounds and can encourage interdisciplinary discussion and stakeholder participation. BNs provide a way to overcome data limitations by incorporating input data from different sources. It is a visual decision support tool and it can handle missing observations [23]. All branches must be calculated in order to calculate the probability of any one branch. The quality of the results of the network depends on the quality of the prior beliefs or model [47].



#### **LOGISTIC REGRESSION:**

Logistic Regression is used to predict the probability of a categorical dependent variable. In logistic regression, the dependent variable is a binary variable that contains data coded as 1 (spam, success, etc.) or 0 (not spam, failure, etc.). It is a statistical learning technique that attempts to predict a data value based on prior observations. The advantage is this algorithm is quick to train, very fast at classifying unknown records, Good accuracy for many simple data sets. It can be regularized to avoid overfitting. The limitations are if the number of observations is lesser than the number of features, Logistic Regression should not be used, otherwise it may lead to overfit. Logistic Regression can only be used to predict discrete functions. Therefore, the dependent variable of Logistic Regression is restricted to the discrete number set. This restriction itself is problematic, as it is prohibitive to the prediction of continuous data [24].

#### **SUPPORT VECTOR MACHINE:**

The objective of support vector machine is to find a hyper plane in an N-dimensional space (N- the number of features) that distinctly classifies the data points. For example, in two dimensional space this hyper plane is a line dividing a plane in two parts where in each class lay in either side [32]. The objects may or may not be linearly separable in which case complex mathematical functions called kernels are needed to separate the objects which are members of different classes. When the number of observations and features are high then SVM can be tried out [39]. The advantage of SVM is generalization is adopted in SVM so there is less probability of overfitting. It works really well with clear margin of separation. It is effective in high dimensional spaces. It is effective in cases where number of dimensions is greater than the number of samples. The disadvantage of SVM is it doesn't perform well, when we have large data set because the required training time is higher. It also doesn't perform very well when the data set has more noise i.e. target classes are overlapping. SVM doesn't directly provide probability estimates, these are calculated using an expensive five-fold cross-validation [41].

#### **K-NEAREST NEIGHBOUR:**

K -Nearest Neighbour (KNN) is a type of instance-based learning, or lazy learning where the function is only approximated locally and all computation is deferred until classification. KNN makes predictions using the training dataset directly. The KNN algorithm is done based on the distances between the training data and the testing data. The distances are calculated using distances such as the Euclidean Distance. Based on the similarity between the training and the testing data, the nearest k neighbors are selected. Here k is a positive integer. The label associated with these neighbors is taken as the reference. The testing data is associated to the class which has majority of the votes amongst the k - nearest neighbors [42]. The advantages of K-nearest neighbour algorithm are it is very simple to understand and equally easy to implement. K-NN is a non-parametric algorithm which means there are assumptions to be met to implement K-NN. K-NN does not explicitly build any model, it simply tags the new data entry based learning from historical data. It's an instance-based learning. The classifier immediately adapts as we collect new training data. One of the biggest advantages of K-NN is that it can be used both for classification and regression problems. The disadvantages of KNN are it might be very easy to implement but as the dataset grows, efficiency or speed of algorithm declines very fast. KNN works well with small number of input variables. KNN algorithm is very sensitive to outliers as it simply chooses the neighbors based on distance criteria. KNN inherently has no capability of dealing with missing value problem [18].

### **3. 2. TO REVIEW EARLIER RESEARCH WORK AND FIND OUT THE EXPERIMENTAL RESEARCH UNDERTAKEN IN THE EDUCATION DOMAIN USING MACHINE LEARNING CLASSIFICATION ALGORITHMS:**

The applications using decision tree algorithm such as ID3, CHAID and C4.5 include:

- a. Students' performance in end semester examination and identifying the dropouts and students who need special attention and allow the teacher to provide appropriate advising/counselling [3].
- b. Predicting under graduate Students' performance in job placement [22].
- c. To predict students' future performance in different semesters and act as an early- warning system for students at risk to prevent student dropout and act as a tool to detect outstanding students and help graduate students in choosing their post-graduate disciplines [10].
- d. Prediction of failure, Prediction of Dropout rates, grouping the students into bad, average, excellent, student retention [36].
- e. Predicting the likelihood of a vocational transfer student in Vocational High School [40].

The experimental research using Rule based classifiers include the following:

- a. A Rule-Based Expert System For Automatic Question classification In Mathematics Adaptive Assessment on Indonesian Elementary School Environment [48].

Naïve Bayes algorithm was used in the following applications:

- a. Predicting Students' performance[16].
- b. Analysing the result of Higher Secondary First Year Improvement Examination [28].
- c. Classification of sub-caste into the given set of quota classes [13].
- d. Predicting student academic performance in a Virtual Learning Environment [2].
- e. Predicting Grade Promotion Using Decision Tree and Naïve Bayes Classification algorithms [21].

The following applications used the Bayesian Networks algorithm:

- a. Publishing Student Grades[35].
- b. Predicting Students' final grade using e-book logs[29].
- c. Predicting group emotion in kindergarten classes[11].

Logistic regression algorithm was used in the following applications:

- a. Predict the learning performance of the learners and propose the suggestions on SPOC(Small Private Online Course )to guide the learners'learning behavior[37].
- b. Prediction of Students' Course Failure Based on Campus Card Data[20].
- c. Score prediction of final grades and identifying weak students[33].
- d. Predicting IT Employability Using Data Mining Techniques[34].

Using Support Vector Machines, the following research was carried out to determine the following:

- a. Predicting students' graduation outcomes[31].
- b. Predicting Students Academic Performance[7].
- c. Predicting Instructor Performance in Higher Education[1].
- d. predictive analytics on faculty performance evaluation[12].
- e. prediction model for the attrition rate of students[8].

The following research was done using K-Nearest Neighbour algorithm:

- a. Prediction of placement of a undergraduate student getting placed in an IT company[19].
- b. Alleviation of road traffic congestion during rush hours in school district planning[49].
- c. Predicting math test scores[6].
- d. Prediction of Education Success from the Edu720 Platform[14].
- e. Recommendation system for students when applying for graduate programs[4].

#### 4. CONCLUSION

The choice of which specific learning algorithm should be used is a decisive step. From the above list of applications, several attempts have been made to design and develop the predictions using classification algorithms in the education domain. The experts in the educational domain however play an significant role to determine the variety of data that should be collected in the specific problem domain. Although in many universities or educational institutes, massive amount of data are generated from varied educational activities, the data remains dispersed. These data can contain hidden knowledge and can be utilized in effective decision making and thereby improve the quality of education. Therefore, the universities/institutions need better assessment, analysis, and prediction tools. It can be provided by machine learning analytics.

In this paper, the advantages and disadvantages of the machine learning classification algorithms and research undertaken in the area of education have been reviewed. This study is likely to give insight to the readers to take an informed decision in identifying the available options of machine learning classification algorithms and then selecting the appropriate classification algorithm. Using various machine learning algorithms there should be an attempt and determination in the education domain to improve the decision making and quality of education.



**REFERENCES:**

- [1] Agaoglu, M. (2016). Predicting Instructor Performance Using Data Mining Techniques in Higher Education. *IEEE Access-Open Access Journal*, 2380-2387.
- [2] Azizah, E. N., Pujianto, U., & Nugraha, E. (2018). Comparative performance between C4.5 and Naive Bayes classifiers in Predicting student academic performance in a Virtual Learning Environment. *International Conference on Education and Technology (ICET)*. Malang Indonesia: IEEE.
- [3] Baradwaj, B. K., & Pal, S. (2011). Mining Educational Data to Analyze Students' Performance. *International Journal of Advanced Computer Science and Applications*, 63-69.
- [4] Baskota, A., & Ng, Y.-K. (2018). A Graduate School Recommendation System Using the Multi- Class Support Vector Machine and KNN Approaches. *International Conference on Information Reuse and Integration (IRI)* (pp. 277-284). IEEE.
- [5] *Bayesian networks - an introduction*. (n.d.). Retrieved January 9, 2020, from <https://www.bayesserver.com/https://www.bayesserver.com/docs/introduction/bayesian-networks>
- [6] Brown, J. M. (2017). Predicting math test scores using K-Nearest Neighbor. *IEEE Integrated STEM Education Conference (ISEC)* (pp. 104-106). IEEE.
- [7] Burman, I., & Som, S. (2019). Predicting Students Academic Performance Using Support Vector Machine. *Amity International Conference on Artificial Intelligence (AICAI)* (pp. 756-759). 2019.
- [8] Camba, J., David, R. E., Betan, A., Lagman, A., & Caro, J. D. (2016). Student analytics using support vector machines. *7th International Conference on Information, Intelligence, Systems & Applications (IISA)*. IEEE.
- [9] Catanzarite, J. (2018, December 14). *The Naïve Bayes Classifier*. Retrieved January 2020, 2020, from <https://towardsdatascience.com/https://towardsdatascience.com/the-naive-bayes-classifier-e92ea9f47523>
- [10] Chiheb, F., Boumahdi, F., & Bouarfa, H. (2017). Predicting Students Performance Using Decision Trees: Case of an Algerian University. *International Conference on Mathematics and Information Technology (ICMIT)*, 113-121.
- [11] Cho, S.-B., & Kim, J.-H. (2015). Predicting group emotion in kindergarten classes by modular Bayesian networks. *7th International Conference of Soft Computing and Pattern Recognition (SoCPaR)* (pp. 298-302). IEEE.
- [12] D. E., Sai, P. G., N. R., P. S., & K. K. (2016). SVM kernel based predictive analytics on faculty performance evaluation. *International Conference on Inventive Computation Technologies (ICICT)* IEEE.
- [13] Dangi, A., & Srivastava, S. (2018). Educational data classification using selective Naïve Bayes for quota categorization. *IEEE International Conference on MOOC, Innovation and Technology in Education*, 118-121.
- [14] Dervisevic, O., Zunic, E., Donko, D., & Buza, E. (2019). Application of KNN and Decision Tree Classification Algorithms in the Prediction of Education Success from the Edu720 Platform. *4th International Conference on Smart and Sustainable Technologies (SpliTech)* (p. 5). IEEE.
- [15] Desmond, M. (2019, 07 27). *Machine Learning Market to Grow to Nearly \$21 Billion by 2024*. Retrieved 1 5, 2020, from <https://pureai.com/https://pureai.com/articles/2019/07/23/nwsdes-machine-learning-market-growth.aspx>
- [16] Devasia, T., V. T., & Hegde, V. (2016). Prediction of Students Performance using Educational Data Mining . *International Conference on Data Mining and Advanced Computing (SAPIENCE)*.
- [17] Doug , A. (n.d.). <https://www.laits.utexas.edu/~anorman/BUS.FOR/course.mat/Alex/>. Retrieved March 2019,29,from <https://www.laits.utexas.edu/https://www.laits.utexas.edu/~anorman/BUS.FOR/course.mat/Alex/>
- [18] Genesis. (2018, September 25). *Pros and Cons of K-Nearest Neighbors*. Retrieved January 9, 2020, from <https://www.fromthegenesis.com/https://www.fromthegenesis.com/pros-and-cons-of-k-nearest-neighbors/>
- [19] Giri, A., V. M., B. V., Pruthvi, B., & Dubey, N. (2016). A Placement Prediction System using k- nearest neighbors classifier. *econd International Conference on Cognitive Computing and Information Processing (CCIP)* (p. 4). IEEE.
- [20] Guo, Q., Chen, M., Dong, Y., An, D., & Ye, L. (2019). Prediction of Students' Course Failure Based on Campus Card Data. *International Conference on Robots & Intelligent System (ICRIS)* (pp. 361-364). IEEE.
- [21] Irfiani, E., Elyana, I., Indriyani, F., Schadu, F. E., & Harmoko, D. D. (2018). Predicting Grade Promotion Using Decision Tree and Naïve Bayes Classification Algorithms. *Third International Conference on Informatics and Computing (ICIC)*. IEEE.
- [22] J. T., A. N., & Kumar, S. D. (2014). Performance Analysis of Undergraduate Students' Placement Selection using Decision Tree Algorithms. *International Journal of Computer Applications (0975 – 8887)* , 27-31.
- [23] Kragt, M. E. (2009). *A beginners guide to Bayesian network modelling for integrated catchment management*. Landscape Logic.



- [24] Kumar, N. (2019, March 2). *Advantages and Disadvantages of Logistic Regression in Machine Learning*. Retrieved January 9, 2020, from <http://theprofessionalspoint.blogspot.com: http://theprofessionalspoint.blogspot.com/2019/03/advantages-and-disadvantages- of.html>
- [25] Li, X.-L., & Liu, B. (2013). *Chapter 1 Rule-based Classification*. Retrieved January 9, 2020, from <https://www.semanticscholar.org: https://www.semanticscholar.org/paper/Chapter-1- Rule-based-Classification-Xiao-li/8557fde1e865f0dcc209468ccaf94f12a04b7835#paper- header>
- [26] Marr, B. (2015, February 24). A Brief History of Big Data Everyone Should Read. Retrieved March 27, 2017, from A LinkedIn website: <https://www.linkedin.com/pulse/brief-history-big- data-everyone-should-read-bernard-marr>.
- [27] Manyika, J., Chui, M., Groves, P., Farrell, D., Kuiken, S. V., & Doshi, E. A. (2013, October). *Open data: Unlocking innovation and performance with liquid information*. Retrieved Jan 01, 2020, from [https://www.mckinsey.com: https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Open%20data%20Unlocking%20innovation%20and%20performance%20with%20liquid%20information/MGI\\_Open\\_data\\_FullReport\\_Oct2013.ashx](https://www.mckinsey.com: https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Open%20data%20Unlocking%20innovation%20and%20performance%20with%20liquid%20information/MGI_Open_data_FullReport_Oct2013.ashx)
- [28] Mathew , S., & Abraham , D. T. (2016). Application of Data Mining in Higher Secondary Directorate of Kerala. *International Conference on Data Mining and Advanced Computing (SAPIENCE)*, 1-5.
- [29] Mouri, K., Okubo, F., Shimada, A., & Ogata, H. (2016). Bayesian Network for Predicting Students' Final Grade Using e-Book Logs in University Education. *16th International Conference on Advanced Learning Technologies (ICALT)* (pp. 86-89). IEEE.
- [30] Osang, J., Udoimuk A., Etta E., Ushie P., & Offiong N. (2013). Methods of Gathering Data for Research Purpose and Applications Using IJCSER Acceptance Rate of Monthly Paper Publication. *IOSR Journal of Computer Engineering (IOSR-JCE)* e-ISSN: 2278-0661, p- ISSN: 2278- 8727 Volume 15, Issue 2 , 1-7.
- [31] Pang, Y., Judd, N., O'Brien, J., & Ben-Avie, M. (2017). Predicting students' graduation outcomes through support vector machines. *Frontiers in Education Conference (FIE)*. IEEE.
- [32] Patel, S. (2017, May 3). Retrieved January 9, 2020, from <https://medium.com/: https://medium.com/machine-learning-101/chapter-2-svm-support-vector-machine-theory-f0812effc72>.
- [33] Pathak, R., Beig, M. T., Faraz, S., & Jain, M. (2019). Employing Data Analytics for Academic Improvement of Students. *International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)* (pp. 385-389). IEEE.
- [34] Piad , K. C., Dumlao, M., Ballera, M. A., & Ambat, S. C. (2016). Predicting IT employability using data mining techniques. *Third International Conference on Digital Information Processing, Data Mining, and Wireless Communications (DIPDMWC)* (pp. 26-30). IEEE.
- [35] Raigoza, J., & Wankhede, D. (2016). Publishing Student Grades while Preserving Individual Information Using Bayesian Networks. *International Conference on Computational Science and Computational Intelligence (CSCI)* (pp. 1399-1400). IEEE.
- [36] Rambola , D. K., Inamke, M., & Harne, S. (2018). Literature Review- Techniques and Algorithms Used for Various Applications of Educational Data Mining (EDM). *International Conference on Computing Communication and Automation (ICCCA)*, 1-4.
- [37] Ran, J., Zhang, G., Zheng, T., & Wang, W. (2018). Logistic Regression Analysis on Learning Behavior and Learning Effect Based on SPOC Data. *13th International Conference on Computer Science & Education (ICCSE)*. Colombo, SriLanka: IEEE.
- [38] Ray, S. (2019). A Quick Review of Machine Learning Algorithms. *International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con)* (pp. 35-39). Faridabad: IEEE.
- [39] Ray, T. (2019, January 16). *Current Trends and Future Scopes of Machine Learning and AI in the Education Sector*. Retrieved January 02, 2020, from <https://www.stoodnt.com: https://www.stoodnt.com/blog/machine-learning-ai-in-education/>
- [40] Roihan , A., Karwandi , & Listiyoko , L. (2018). Development of a website-based decision tree system in data mining concept. *Third International Conference on Informatics and Computing (ICIC)*, 1-4.
- [41] Strickland, J. (2016). *Data Analytics using Open Source Tools*. lulu.com.
- [42] V. M., B. V., B. P., & Dubey, N. (2016). A Placement Prediction System using k-nearest neighbors classifier. *Second International Conference on Cognitive Computing and Information Processing (CCIP)*. IEEE.
- [43] S. S., & S. S. (n.d.). *Introduction to Data Mining and its Applications*. Springer.
- [44] Waltl , D., Bonczek , G., & Matthes, F. (2018). *Rule-based Information Extraction: Advantages, Limitations, and Perspectives*. Retrieved January 9, 2020, from <https://wwwmatthes.in.tum.de>:



<https://www.matthes.in.tum.de/pages/1w12fy78ghug5/Rule-based-Information-Extraction-Advantages-Limitations-and-Perspectives>

[45] Xiaaoqiang. (July, 2019 18). *Naive Bayes classifier / NBC*. Retrieved January 9, 2020, from <https://easyai.tech/https://easyai.tech/en/ai-definition/naive-bayes-classifier/>

[46] Yaghini, D. (2010). *Data Mining: Rule Based Classification*. Retrieved January 9, 2020, from [http://webpages.iust.ac.ir: http://webpages.iust.ac.ir/yaghini/Courses/Data\\_Mining\\_882/DM\\_05\\_04\\_Rule-Based%20Classification.pdf](http://webpages.iust.ac.ir: http://webpages.iust.ac.ir/yaghini/Courses/Data_Mining_882/DM_05_04_Rule-Based%20Classification.pdf)

[47] Yahya, Y. (2017). *Bayesian Network Advantages, Limits*. Retrieved January 9, 2020, from <https://www.coursehero.com: https://www.coursehero.com/file/p7osgle/Bayesian-Network-Advantages-Limits-The-advantages-of-Bayesian-Networks-Visually/>

[48] Yuhana, U. L., Rochimah, S., Yuniarno, E. M., Rysbekova, A., Tormasi, A., Koczy, L. T., et al. (2019). A Rule-Based Expert System For Automatic Question classification In Mathematics Adaptive Assessment On Indonesian Elementary School Environment. *International Journal of Innovative Computing, Information and Control*, 143-161.

[49] Zhou, T., Guo, Y., Yang, Y., Mi, Z., & Obaidat, M. S. (2019). The School District Reorganization by Combining with Traffic Congestion Data. *International Conference on Computer, Information and Telecommunication Systems (CITS)* (p. 5). IEEE.