Predicting Student Academic Performance using Unsupervised Machine Learning

M. Tech. Scholar Simran Aliwal, Asst. Prof. Abhay Mundra
Department of Computer Science and Engineering,
Jawaharlal Institute of Technology,
Borawan Dist. Khargone, MP, India
simran2109aliwal@gmail.com, er.abhayjit@gmail.com

Abstract- Pupils' progress in academic performance may be tracked a fundamental problem preventing the academic community as a whole from making claims about increased intake. Depending on the amount of detail required, it may be possible to depict a system for analyzing students' results based on group analysis and the use of standard quantifiable computations. This study also includes the implementation of the k-mean clustering technique for analyzing students' outcome data. There's a good chance that the deterministic model was consolidated for the sake of convenience, and that it's recommended that the impacts of the model on the pupils be dissected using data from a private foundation that was clinched alongside% For a successful decision by academic organizers, it is helpful to look to Iberia as a standard for the development of academic execution on individuals for higher institutions.

Keywords- K – Mean, Clustering, Academic Performance, Algorithm.

I. INTRODUCTION

Machine Learning is the part of artificial intelligence (AI) that provide the system to ability to learn automatically, it also provide facility to improve from experience without being explicitly programmed.

It can be defined as a branch of artificial intelligence that provides computer with the ability to learn without having all the information with respect to a domain in the program itself. It has the study of Pattern Recognition. Machine Learning is related to Computational statistics, which also emphasis on prediction using the computer. Data mining is term in the machine learning, in which it focuses on the data analysis.

The feature of machine learning is significant because as the models are exposed to new data, they are able to independently adapt. They learn from earlier computations to produce consistent, repeatable decisions and result. Machine learning is field of AI in which there is lots of data and it fit the data into modules so that can be utilized by people.

Machine learning and traditional computational method both are different and also machine learning is field of computer science. In traditional computing, algorithm is sets of programmed instruction used by computers to calculate. Machine learning algorithm permit computer to train on data input and use statistical analysis for the output fall within specific range.

The main focus of any higher institution is to improve decision making at the administrative level and to impart education. Analysis of students' success in high institutions is one of the bases for improving the quality of education. Student's performance is an important and integral part in higher institutions.

This is because the quality of education in universities is based on its excellent record of academic achievements. Predicting students' performance has become an important task due to the huge amount of data in educational databases. So educational data mining is important and useful now days.

Machine learning is a field that can learn from past experience to improve the future performance. Here the learning means improvement of the algorithm and then use these algorithm in the future. The systems are there and it is cannot be consider intelligent if it does not have ability to learn. So learning ability is most important feature for the intelligent system [1].

The self-driving Google car, fraud detection, online recommendation system, friend suggestions on social network, Netflix showcasing the movies are the example of applied machine learning. Facial recognition technology is so used method now days which allows user to tag and share photos in social medial platform. OCR (Optical Character Recognition) technology converts images of text into movable type.

II. LITERATURE REVIEW

Fatma Chiheb [1]- Decision tree method is used in these paper. Decision tree is build using the J48 algorithm. Weka toolkit is used and CRISP-DM model is applied. They collect the data about graduates and post-graduates students. It is a case of an Algerian university. The data is taken from, from computer science department. They test decision tree and analysed the error rates in order to choose the best input and output. Different grades are taken as attribute and student's performance is predicated.

V.Shanmugarajeshwari [2]-They evaluate the students' performance using the classification techniques. The input data is collected from ayya nadir janaki Ammal College, sivakasi, from the computer science department. For feature selection number of methods is discussed.[22] Training data is applied on the data set and the classifier model has been developed. Decision tree classification was used to predict the students' performance.

M.Durairaj[3]-Educational details and performance is based upon various factors like personal details, social etc. WEKA toolkit is used they collect the data set of college students real time data that describe the relationship between learning behaviour of students and their academic performance, the data set contain students detail of different subject marks in semester which is subjected to the data mining process. [23]In these K-means clustering is used and from the total number of 300 student record dataset,

they choose 38 students record for our analysis .The confusion matrix is there to shows pass, fail, and absence for the exam. They compare the weighted average for decision tree and navie bayes techniques.

Mr.Shashikant pradip borgavakar[4]-Here the data clustering is used as k-means clustering to evaluate students' performance. Their performance is evaluated on the basic of class test, mid test, and final test.[24] In their model they measured by internal and external assessment, in which they tale class test marks, lab performance, quiz etc. and final grade of students is predicted They generate the graph which shows the percentage of students getting high, medium, low gpa.

Edin Osmanbegovic[5]-In these paper supervised data mining algorithm were applied. Different method of data mining was compared. The data were collected from the survey conducted during the summer semester at the University of Tuzla. Many variable like Gender, GPA, Scholarships, High school, [25] Entrance Exam, Grade, etc. are taken for the performance. Naive Bayes algorithm, multilayer Perceptron, J48 issued. The result indicates that the naïve Bayes classifier outperforms in predication decision tree and neural network method. These will help the student for future.

E.venkatasan et.al[6]-In this article the clustering and classification algorithm were compared using matrix laboratory software, for the initial data WEKA software is utilized. [26] Data set of students was picked up from private arts and science colleges from Chennai city. Near about 573 students are there in the database. In the details they take the internal exam and end semester exam details. Algorithm such as J48 were used allows the input attribute to get classification model. Matrix Laboratory is used for measuring the operational of several data mining algorithm. There is a table for error measure.

A.seetharam Nagesh[7]-Prediction of students' performance is so important but if it is predicted at early stage it become so useful for the students Here they applied k means clustering algorithm for analyzing the students result data and predicting the students' performance.[27 Unsupervised techniques are also called clustering techniques. The k means is partition based clustering algorithm. Euclidean distance is the distance which is measure in k means

clustering algorithm. Here the data set used was obtained from the information department of the engineering college. The attribute are aggregate and attendance for experiment. They create the final output after clustering, They shows by red, green, blue to differentiate the poor, average, good students

Qasem A. Al-Radelideh[8]-The title of the paper is "Mining student data using decision tree". They use data mining process for student performance in university courses to help the higher education management. [28] Many factors affect the performance. They use classification technique for building the reliable classification model, the CRISP-DM (cross-industry standard process for data mining) is adopted .These method consist of five steps i.e. collecting the relevant features of the problem, Preparing the data, Building the classification model, Evaluating the model and finally future prediction. [29] The data were collected in table in proper format, the classification model were building using the decision tree method. Many rules were applied. The WEKA toolkit is used Different classification methods were used like ID3, C4.5 and naïve Bayes and accuracy were in the table as result.

Mashael A[9]-These researches has applied decision tree for predicting students final GPA. It used WEKA toolkit .It collect the data from C.s. College at king save university in the year 2012 were collected from the institute. [30] Each student record is with different attributes i.e. Student name, student id, final GPA, semester of graduation etc. It is important to improve the final GPA of the student.

Ryan S.J.D.Baker10]-"The state of educational data mining in 2009:A review and future vision" In these paper author review the trend in 2009 in field of educational data mining. The year 2009 finds research communizing of EDM and these moment in EDM bring unique opportunity.[31] EDM categories in web mining, Statistics and Visualization, Clustering, Relationship mining i.e. Association rule mining and causal data mining. There are many application of edm. These papers discuss about the EDM.

Pooja M.Dhekankar[11]-"Analysis of student performance by using data mining concept "Data mining technique is used in many area and in the educational field it become so important for future of the students .Students classification is done on the

basic of students mark. Association rule, clustering outlier detection, classification is discussed in this paper.

Amjad Abu saa[12]-It applies c4.5, CART, ID5 algorithm for analysis of students' performance. It takes various parameters for the accuracy. Decision tree is build and based on it student performance is predict.[32] Naive Bayes classification is also applied which assumes that all given attribute in a dataset is independent. It create different quantities predictive model by using different data mining tasks which is effective to predict student grades .various decision tree algorithm were implemented . Finally we can say that it help the university as well as students.

Yoav Bergner[13] et.al-It used collaborative filtering analysis of student data. [33] There is logistic regression as collaborative filtering. There is parameter estimation. There is simulated skill response. It applied numerical method for analysing student response matrix with the goal of predicting response; it showed of naturally parameterizes series of models and multidimensional IRT.

Md.Hedayetul Islam Shovon[14]-Data clustering techniques i.e. K-means is applied. [34] Data clustering is process of extracting unknown, hidden patterns from large data set. In this model they use internal and external assessment for prediction. This model helps to weak students to identify their score before the exam. Graph shows the relationship between GPA and attendance and also number of students and percentage of student regarding to GPA. And from it they show the percentage of students getting high, medium, and low gpa.

J.K. Jothi Kalpana[15]-"Intellectual performance analysis of students by using data mining techniques "This paper focus on the prediction of school in different level such as primary, secondary, higher level. [35] Clustering method such as centroid based distribution based and density based clustering are used. The data were collected from Villupuram College. There method used for improving the performance as the students.

Cristobal Romero[16]- "Educational data mining; A Review of the state of the art". EDM i.e. educational data mining is emerging discipline. EDM process converts raw data coming from educational system into useful information.DM techniques are used i.e.

Simran Aliwal. International Journal of Science, Engineering and Technology, 2023, 11:3

International Journal of Science, Engineering and Technology

An Open Access Journal

association rule mining for selecting weak students.[36] Several classification algorithms were applied in order to group students. EDM tools were designed for educators.

Romero[17]-"Educational data mining survey from 1995 to 2005"There is also web-based education in the computer aided instruction in the specific location. Web based education is so popular now a days that predication its level is also become useful. Data processing is done for transform the original data into suitable shape. Web mining is there for extract knowledge from the web. Clustering, classification is used. In these it says that the predication of performance in e-learning is also so important.

S.Kotsiantis[18]-"Predicating students' performance in distance learning using machine learning techniques" Many university are giving distance learning education so predicating performance of students in that become so important. Machine learning algorithm is so effective for many types of learning tasks. [37] This paper Use ML techniques to predict students' performance in distance learning system. Set of rules are planned. Decision tree are used, ANN is also inductive learning based on computational models. Set of attribute are taken and divided into groups. There is ANOVA test result. It showed that best algorithm is naïve Bayes with 66.49% accuracy in the data it taken.

Pooja Thakar[19]-"Performance analysis and prediction in education data mining: A Research Travelogues'" Lots of data is collected in educational databases. [38] In order to get benefits from such big data tools are required. University produces lots of students and its performance predication is important. Set of weak students are taken and predication with data mining techniques is used. This paper says that many models are required for an instruction.

Ben Daniel [20]-It applied big data analysis in higher education.KDD is an interdisciplinary area focusing on method for identifying and extracting pattern from large data sets. Big data help provide insight to support students learning needs.

Tismy Devasaia[21]-It used classification technique to predict the student performance .Naive theorem is used various information like group action, class text,

semester and assignment marks were collected from the students previous information to predict performance of the student.

III. CONCLUSION

Analysis of student's performance is done with the help of modified k means algorithm. Machine learning is very emerging technology that every placed it used. Now a day in bank, labs, telecom, industrial each and every place machine learning is used. Data mining is part of it which helps in prediction, future prediction is very important and in the education system it becomes more important. Many algorithm is build and more and more research is going on every technology used the concept of it. We survey on many papers of both decision tree and k means and it is found that modified k means is more stable method.

REFERENCES

[1] Fatma Chiheb, Fatima Boumahdi, Hafida Bouarfa, Doulkifli Boukraa, "Predicting Students Performance using decision trees : Case of an Algerian University " IEEE, International Conference on Mathematics and information Technology, Adrar, Algeria December 4 - 5, 2017.

[2] V. Shanmugarajeshwari, R. Lawrance, "Analysis of Students' Performance evaluation using classification techniques" IEEE, 2016.

[3] M. Durairaj, C. Vijitha, "(IJCSIT) International Journal of Computer Science and Information Technologies", Vol. 5 (4), 5987-5991, 2014.

[4] Mr. Shashikant Pradip Borgavakar, Mr. Amit Shrivastava, "Evaluating student's performance using k-means clustering" International Journal of Engineering Research & Technology (IJERT), Vol. 6 Issue 05, May – 2017.

[5] Edin Osmanbegovic, Mirza Suljic, "Data mining approacg for predicting student performance" Economic Review – Journal of Economics and Business, Vol. X, Issue 1, May 2012.

[6] E.Venkatesan, S.Selvaragini, "Prediction of students' academic performance using classification and clustering algorithms" International Journal of Pure and Applied Mathematics, Volume 116 No. 16 , 327-333, 2017.

[7] A Seetharam Nagesh, Ch V S Satyamurty, "Application of clustering algorithm for analysis of student academic performance" IJCSE

Simran Aliwal. International Journal of Science, Engineering and Technology, 2023, 11:3

International Journal of Science, Engineering and Technology

An Open Access Journal

International Journal of Computer Sciences and Engineering, Volume-6, Issue-1, 31/Jan/2018.

[8] Qasem A. Al-Radaideh, Emad Al-Shawakfa, Mustafa I. Al-Najjar, "Mining students data using decision trees", The 2006 International Arab Conference on Information Technology (ACIT'2006), Jordan, Nov. 2006.

[9] Mashael A. Al-Barrak, Muna Al-Razgan, "Predicting students final GPA using decision trees: A case study" (IJIET), International Journal of Information and Education Technology, Vol. 6, No. 7, July 2016.

[10] Ryan S.J.D. baker, Kalina yacef, "The state of educational data mining in 2009 A review and future visions" Journal of Educational Data Mining, Article 1, Vol 1, No 1, Fall 2009.

[11] Pooja M .Dhekankar, Dinesh S. Datar, "Analysis of Student Performance by using Data Mining Concept" International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 3 Issue: 5, May 2015.

[12] Amjad Abu Saa, "Educational data mining & students' performance prediction" (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 7, No. 5, 2016.

[13] Yoav Bergner, Stefan droschlery, Gerd Kortemeyerz , Saif Rayyan, Daniel Seaton, David E. Pritchard, "Model based collaborative filtering analysis of students response data: machine learning item response theory" *International Educational Data Mining Society*, International Conference on Educational Data Mining (EDM) , 5th, Chania, Greece, Jun 19-21, 2012.

[14] Md. Hedayetul Islam Shovon, Mahfuza Haque, "Prediction of students academic performance by an application of k means clustering algorithm" International journal of advanced research in computer science and software engineering, Volume 2, Issue 7, July 2012.

[15] J.K. Jothi Kalpana, K. Venkatalakshmi, "Intellectual Performance analysis of students by using data mining techniques" (IJIRSET) International Journal of Innovative Research in Science, Engineering and Technology, Volume 3, Special Issue 3, March 2014.

[16] Cristobal Romero, Sebastián Ventura, "Educational data mining: A review of the state-of-the-art" IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) Volume: 40, Issue: 6 , Nov. 2010.

[17] C. Romero, S. Ventura, "Educational data mining: A survey from 1995 to 2005" Elsevier, Expert system with applications, volume 33, issue 1, pages 135-146, july, 2007.

[18] S. KOTSIANTIS, C. PIERRAKEAS, P. PINTELAS, "Predicting students' performance in distance learning using machine learning techniques" Taylor & Francis, Applied artificial intelligence, May- June , 2004.

[19] Pooja Thakar, Anil Mehta, Manisha, "Performance Analysis and prediction in educational data mining: A research Travelogue" International Journal of Computer Applications (0975 – 8887) Volume 110 – No. 15, January 2015.

[20] Ben Daniel, "Big data nad analytics in higher education: Opportunities and challenges" British Journal of Educational Technology, 2014.

[21] Tismy Devasia, Vinushree T, Vinayak Hegde, "Prediction of students' performance using educational data mining" IEEE, International Conference on Data Mining and Advanced Computing (SAPIENCE), 2016.

[22] A. Taiwade, N. Gupta, R. Tiwari, S. Kumar and U. Singh, "Hierarchical K-Means Clustering Method for Friend Recommendation System," 2022 International Conference on Inventive Computation Technologies (ICICT), 2022, pp. 89-95, doi: 10.1109/ICICT54344.2022.9850852.

[23] R. Baghel, P. Pahadiya and U. Singh, "Human Face Mask Identification using Deep Learning with OpenCV Techniques," 2022 7th International Conference on Communication and Electronics Systems (ICCES), 2022, pp. 1051-1057, doi: 10.1109/ICCES54183.2022.9835884.

[24] M. Ranjan, A. Shukla, K. Soni, S. Varma, M. Kuliha and U. Singh, "Cancer Prediction Using Random Forest and Deep Learning Techniques," 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), 2022, pp. 227-231, doi: 10.1109/CSNT54456.2022.9787608.

[25] Singh, Upendra, Gupta, Puja, and Shukla, Mukul. 'Activity Detection and Counting People Using Mask-RCNN with Bidirectional ConvLSTM'. 1 Jan. 2022: 6505 – 6520.

[26] Gupta, P., Shukla, M., Arya, N., Singh, U., Mishra, K. (2022). Let the Blind See: An AIIoT-Based Device for Real-Time Object Recognition with the Voice Conversion. In: Al-Turjman, F., Nayyar, A. (eds) Machine Learning for Critical Internet of Medical Things. Springer, Cham. https://doi.org/10.1007/978-3-030-80928-7_8

[27] Patidar, M., Singh, U., Shukla, S.K. et al. An ultra-area-efficient ALU design in QCA technology

using synchronized clock zone scheme. J Supercomput (2022). https://doi.org/10.1007/s11227-022-05012-2

[28] A. Taiwade, N. Gupta, R. Tiwari, S. Kumar and U. Singh, "Hierarchical K-Means Clustering Method for Friend Recommendation System," 2022 International Conference on Inventive Computation Technologies (ICICT), 2022, pp. 89-95, doi: 10.1109/ICICT54344.2022.9850852.

[29] Jindal, R, Kumar, N, Patidar, S. IoT streamed data handling model using delta encoding. Int J Commun Syst. 2022; 35(13):e5243. doi:10.1002/dac.5243

[30] S. Patidar, A. Jain and A. Gupta, "Comparative Analysis of Machine Learning Algorithms for Heart Disease Predictions," 2022 6th International Conference on Intelligent Computing and Control Systems (ICICCS), 2022, pp. 1340-1344, doi: 10.1109/ICICCS 537 18.2022.9788408.

[31] B. Jain, S. Patidar and D. Sudershan, "Heterogeneous Software Defect Prediction using Generative Models," 2022 IEEE 11th International Conference on Communication Systems and Network Technologies (CSNT), 2022, pp. 367-372, doi: 10.1109/CSNT 54456.2022.9787607.

[32] Singh, Shani Pratap and Shukla, Jayesh and Sharma, Shaili and Daga, khushhal and Bhalavi, Brahman Singh and Singh, Upendra, Face Mask Detection using Multi-Stage CNN Architecture (July 10, 2021). Proceedings of the International Conference on IoT Based Control Networks & Intelligent Systems - ICICNIS 2021,

[33] A. Saxena, A. Vyas, L. Parashar and U. Singh, "A Glaucoma Detection using Convolutional Neural Network," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 815-820, doi: 10.1109/ICESC48915.2020.9155930.

[34] B. Bamne, N. Shrivastava, L. Parashar and U. Singh, "Transfer learning-based Object Detection by using Convolutional Neural Networks," 2020 International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 2020, pp. 328-332, doi: 10.1109/ICESC48915.2020.9156060.

[35] Gupta, P., Shukla, M., Arya, N., Singh, U., Mishra, K. (2022). Let the Blind See: An AIIoT-Based Device for Real-Time Object Recognition with the Voice Conversion. In: Al-Turjman, F., Nayyar, A. (eds) Machine Learning for Critical Internet of Medical Things. Springer, Cham. Patidar, M., Singh, U., Shukla, S.K. et al. An ultra-area-efficient ALU design in QCA technology using synchronized clock zone scheme. J Supercomput (2022).

[36] Singh, Upendra, Gupta, Puja, and Shukla, Mukul. 'Activity Detection and Counting People Using Mask-RCNN with Bidirectional ConvLSTM'. 1 Jan. 2022 : 6505 – 6520.

[37] U. Singh and L. S. Songare, "Analysis and Detection of Monkeypox using the GoogLeNet Model," 2022 International Conference on Automation, Computing and Renewable Systems (ICACRS), Pudukkottai, India, 2022, pp. 1000-1008