STUDENT PERFORMANCE ANALYSIS SYSTEM

[1] Asst prof.Mohamed Sadiq B B[1], Spoorthy H P R[2]

[1] Department of Master of Computer Application,BIET,Davangere

[2] Spoorthy H P,PG Student,Department of Master of Computer Application,BIET,Davangere

ABSTRACT: This paper deals with Learning-based performance analysis of results a system that strives for excellence at different levels and dimensions in the area of student's interest. This system is designed to analyze and predict student performance only. The proposed framework analyzes student demographics, academic and psychological characteristics to extract all possible knowledge from students, teachers, and parents. In the proposed paper various machine learning algorithms accuracy in solving the problem statement has been compared.

I.   INTRODUCTION

Education plays an important role in the development of the country and a key factor for achieving long term economic progress. The statistics of literacy shows the percentage of failure rates due to, dropping out from the schools and student failing in particular subjects. There is an increase in the literacy rate throughout the world from the past years. In particular, failure in the core class such as mathematics is very serious, since they provide fundamental knowledge for the success in the remaining school subjects. An educational institution needs to maintain the records of all the students which results in a large database.

Educational data mining uses many techniques such as decision tree, Naïve Bayesian, neural networks, k-nearest neighbor and many others. By using these techniques, many kinds of knowledge can be discovered such as association rules, classifications and clustering. It showed what kind of data could be collected, how to pre-process the data, how to apply data mining methods on the data, and finally how to get benefited from the discovered knowledge. Many kinds of knowledge can be discovered from the extracted data. In this era of big data, huge amounts of structured and unstructured student data are being produced daily. Big Data is difficult to work with and requires massively parallel software running on a large number of computers. Previous works investigated the most common ones which are attribute selection, classification such as decision tree method.

The proposed work analysis various factors affecting the performance of students, data which is collected from domain experts such psychiatrist with respect to factors affecting students' performance are considered, before feeding the data to ML algorithms the exploratory data analysis is done , in this stage distribution nature of data, scaling of data to standard values, analyzing the missing data and compensating null values is done. Having done EDA data is fed to ML model to train the model. Later it is saved in the form of pickle file and performance analysis is done.

II.   LITERATURE SURVEY

IhsanA. Abu Amra, AshraY.A[2] says that data mining techniques are rapidly increasing in the hidden knowledge and patterns about student performance. This paper proposes a student performance prediction model by applying two classification algorithms: KNN and Naive Bayes on educational data set of secondary schools, collected from the ministry of education in Gaza Strip for 2015. The main objective of such classification may help the Ministry of education to improve performance due to early prediction of student performance. Teachers can also make appropriate assessments to improve student learning. Experimental results show that Naive Bayes is superior to KNN by obtaining the highest accuracy score of 93.6%.

Nikita Gorad, Ishani Zalte[3], choosing the right career is one of the most important decisions, and as the number of career paths and opportunities increases, it becomes very difficult for students to make this decision. According to a survey by the Scientific and Industrial Research Council (CSIR), about 40% of students are uncertain about their career options. This can lead to incorrect choices of profession and subsequent work in areas they did not intend, reducing talent productivity. Therefore, it is very important to make the right career choice at the right age to prevent the consequences of the wrong career choice.

[Type here]

Ganeshan et al. [4] proposed a web-based analysis system for student counseling and performance analysis. This system uses the techniques used in recommended systems. It divides students into similar groups. When a new student comes, this system assigns him a group by analyzing his features and also offers him similar courses.

Lopez et al. [5] proposed a data mining approach-based model for academic attrition (loss of academic status) at the University of Colombia. Two data mining models were defined to analyze the academic and nonacademic data, the models use two classification techniques, naive Bayes and a decision tree classifier, in order to acquire a better understanding of the attrition during the first enrolments and to assess the quality of the data for the classification task, which can be understood as the prediction of the loss of academic status due to low academic performance.

B.K Baradwaj, S.Pal [9] The model aims to predict the sales of the first four registrations of a student. First consider one of these periods, then at a particular registration. To train the model and measure their performance with various algorithms publically hosted dataset available in the kaggle repository have been considered.
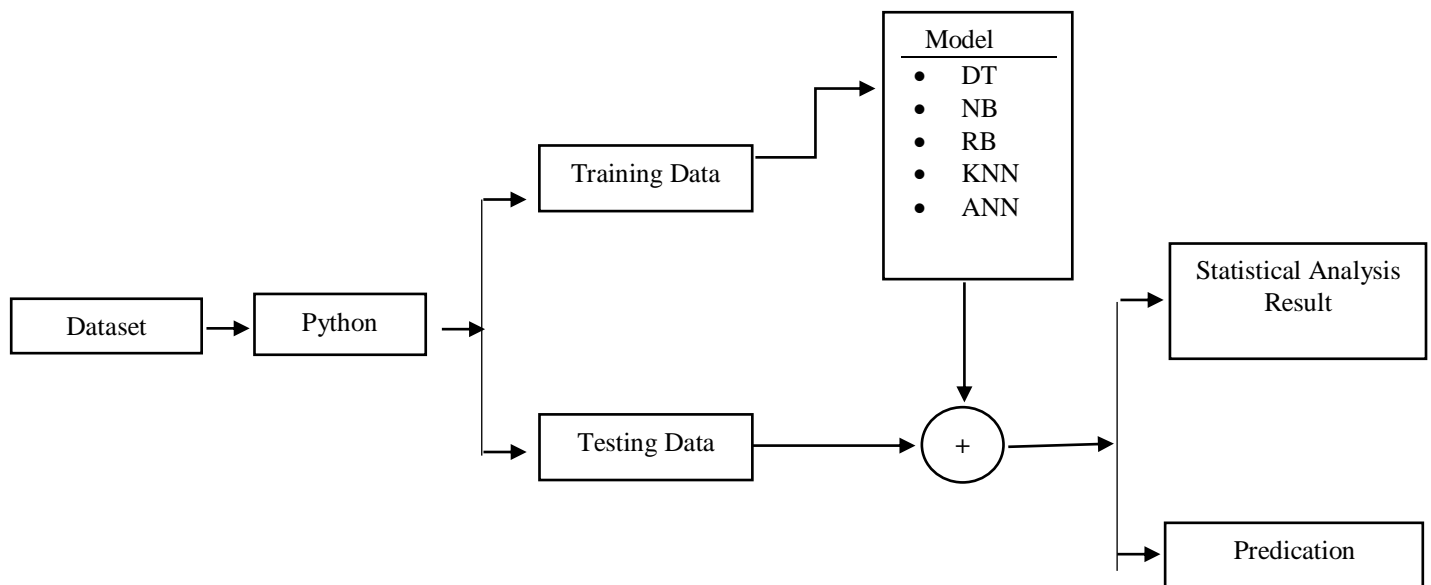
III. METHODOLOGY

Fig.1.: Schematic Diagram of Data Analysis and Prediction Model

The above diagrams shows the adopted methodology in the proposed system, dataset which plays key role on the end result is collected form the kaggle repository, it is analyzed using python libraries such as pandas, preprocessing is done using pandas, then it is split into training and test data to measure accuracy, this data is fed to ML algorithm to build the model, performance of each model backed by algorithms is measured using statistical analysis and end result is displayed.
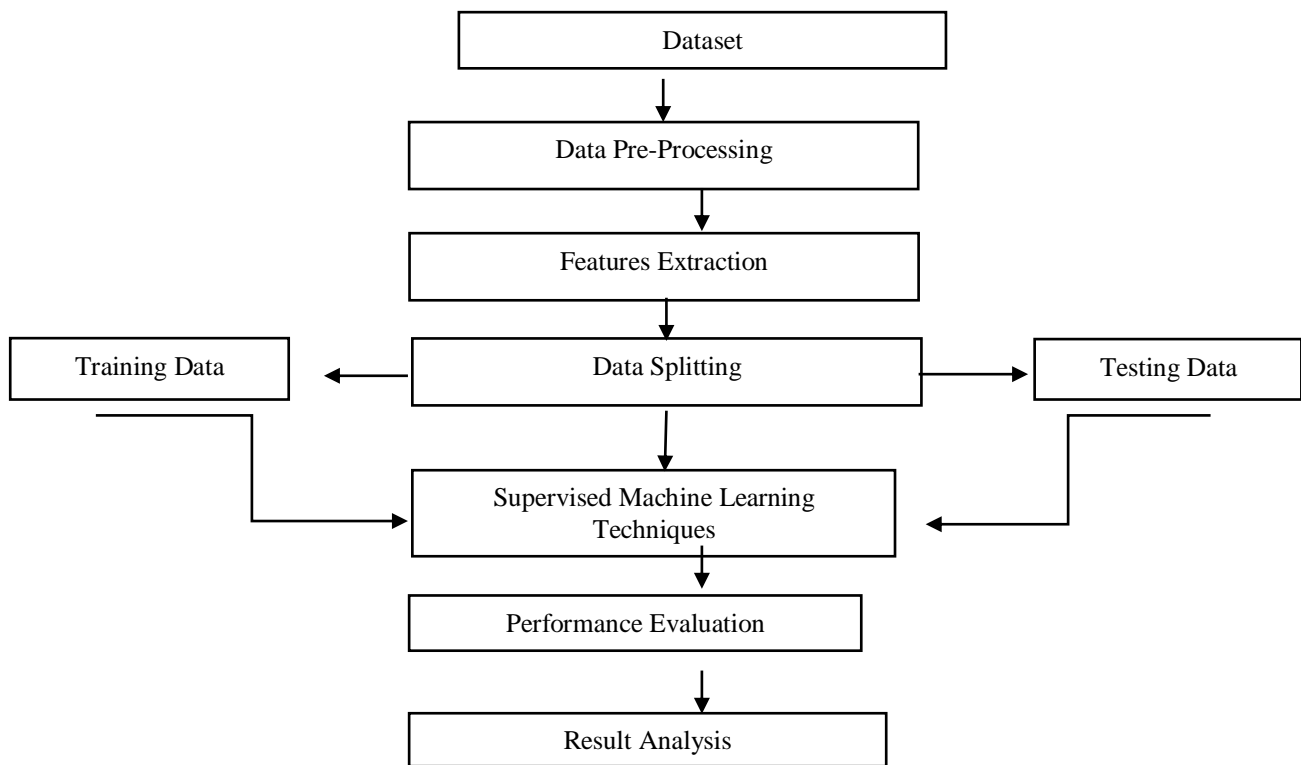
Fig.2.: Methodology in Student Performance Analysis

Dataset: A collection of data pieces that can be treated by a computer as a single unit for analytic and prediction purpose.

Data Pre-processing: It is the process of transforming raw data into an understandable format. It is also an important step in data mining as we can't work with raw data.

Feature Extraction: It refers to the process of transforming raw data into numerical features that can be processed while preserving the information in the original dataset.

Data Splitting: It is used to split data into a train, test, or validation set.

Training Data: It is an extremely large dataset that is used to teach machine learning model.

Testing Data: It is separate set of data used to test the model after completing the training.

Supervised Machine Learning Technique: Regression and classification, classification separates the data, regression fits the data.

Performance Evaluation: It is an important aspect of the machine learning process. However, it is complex task.

Result Analysis: It provides valuable things from mining so that it can provide information which helps in analysis of student performance and making career decisions and processes.

IV. DIFFERENT DATA MINING TECHNIQUES USED FOR PREDICTING STUDENT'S PERFORMANCE

Serial No.:	Author's	Attributes which affect prediction accuracy	Result (accuracy)				
			DT	NB	RB	KNN	ANN
1.	F Sarker et. al.	Internal attributes + students' first semester mark (Model1)	--	--	--	--	74.5
2.	F Sarker et. al.	Int + Ext attributes + students' first semester mark (Model2)	--	--	--	--	76.5
3.	F Ahmad et. al	Gender, race, hometown, GPA, family income, uni. mode entry, SPM grades	--	67.0	71.3	--	68.8
4.	Mashael A. et. al.	first midterm exam (Predict Students Failure)	--	91	55.0	--	89.8
5.	M Suljic et. al.	GPA, URK , MAT, VRI	--	76.6		--	73.9
6.	R Asif et. al.	HSC, MPC and HSC marks, pre-uni marks, marks in different courses	73	83.6	56.9	74	67.6
7.	El-Halees et. al.	SS_Type, HSC marks, City, Gender, Speciality	--	67.5	71.2	--	--

Serial No.:	Author's	Attributes which affect prediction accuracy	Result (accuracy)				
			DT	NB	RB	KNN	ANN
8.	A Aziz et. al.	Gender, race, Hometown Location, Uni Entry Mode, Family Income	68.8	63.3	68.8	--	--
9.	K D Kolo et.al.	status, gender	66.8	--	--	--	--
10.	Jyoti Bansode	SSC marks, SSC medium, Admission type, mother's occupation	85	--	--	--	--
11.	R. Sumitha et. al.	TWM, MOE, TOB, ATD ECUT, CGPA, arrears,	97.2	85.9	96.1	--	--
12.	S V. Shinde et. Al	Student's Internal Assessment	97.5				
13.	M Pandey et. al.	Academic information's, Demographic information	98.8	91.5	84.1	--	--
14.	N Goga et. al.	family, PEP, EES, end of the first session result	99.9	--	96.7	--	--
15.	G. S Josan et. al.	Sex, INS-High, TOB, MOI, TOS, PTUI, SArea, Mob, Com-HM, Netacs, Int-GR, Atdn.	69.7	65.1	--	--	--
16.	M Koutina et. al.	Another master, Comp literacy, Bachelor in	68.5	100	90.9	100	--
17.	D. Kanojia	Comparison of naïve basian and KNN classifier	--	93.70	--	63.65	--
18.	B. Harish Kumar Reddy	Gender, Age, Marital Status, No of children, Occu., Job associated with PC, Bachelor,	--	--	71.3	--	--

Table 1.: Different data mining techniques used for predicting student's performance.

With reference to the above table 1 to 15 [17]. In the field of educational data mining, predictions are usually made about the student's academic performance. To create predictive modeling, you need to consider various data mining techniques such as classification, clustering, association rule mining, and regression analysis. Almost all research papers consider the only classification algorithm for predicting student performance. There are numerous classification methods for prediction, but we are considering decision trees, naive bays, support vector machines (SVMs), artificial neural networks (ANNs), K-nearest neighbors, SMOs, linear regression, etc. Random forest, random tree, REPTree, LADTree, J48, etc. Table outlines the various research papers, their author names, and key attributes that help predict accuracy in the various data mining algorithms used.
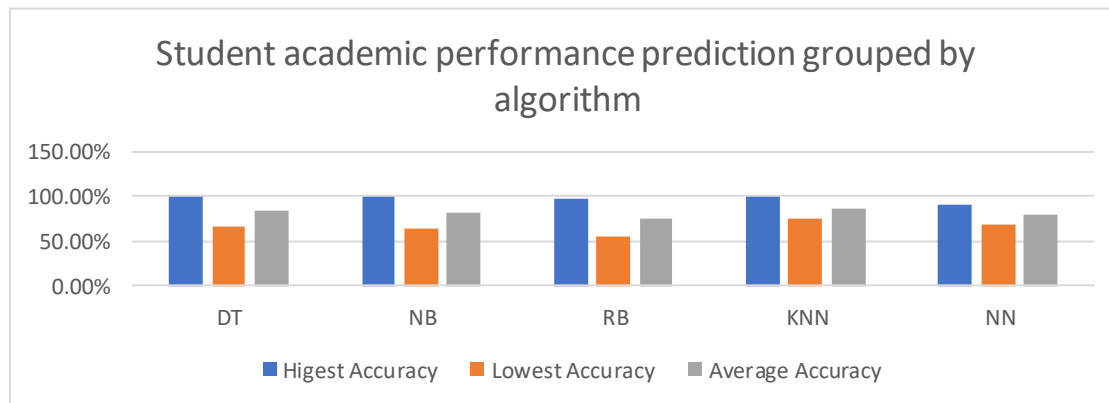
From the table above, we can see that Maria Koutina and Katia Lida Kermanidis found 100 ° accuracy in his work using the Naive Bayes and KNearest Neighbor algorithms. Results are shown in "Overall accuracy of resample data and feature selection (%)". Gender, age, marital status, number of children, profession, profession, computer-connected, bachelor's degree, another master's degree, computer skills, bachelor's degree in computer science to predict student academic performance.

V.　　DISCUSSION ON THIS PREDICTING STUDENT'S SURVEY

In this particular section, we will discuss the main findings of our analysis. In this analysis, we found that the most used data mining algorithms for SAP are Decision Tree (DT), Naive Bayes (NB), Artificial Neural Networks (ANN), Rulebased (RB) and KNearest Neighbor (KNN). In Decision tree algorithm, the maximum and minimum accuracy for predicting student's academic performance are 99.9% and 66.8% respectively. To find the maximum prediction accuracy, Maria Goga, Shade Kuyoro and Nicolae Goga used the combination of students' attributes like family, PEP, EES, end of first session result. . In Naive Bayes algorithm, the maximum and minimum accuracy for predicting a student's academic performance are 100% and 63.3% respectively. Maria Kotina and others. Maximum with different combinations of student attributes such as gender, age, marital status, number of children, profession, computer-related profession, bachelor's degree, another master's degree, computer skills, bachelor's degree in computer science, etc. Obtained the accuracy of With rule-based algorithms, the maximum and minimum accuracy for predicting student performance is 96.7% and 55.0%, respectively. Maria Gogaetal to find the maximum prediction accuracy. At the end of the first session, we used a combination of student attributes such as family, PEP, EES, and results. With the KNearest Neighbor algorithm, the maximum and minimum accuracy for predicting student performance is 100% and 74%, respectively. With Artificial Neural Networks (ANN), the maximum and minimum accuracy for predicting student performance is 89.8% and 67.6%, respectively. Mashael A. Al Barrak and Mona S to find the maximum prediction accuracy. AlRazgan used a combination of student attributes, such as the first midterm exam of the first-year course.

Data Mining Technique	Decision Tree	Naïve Bayes	Rule Based	K-Nearest Neighbor	Neural Network
Highest Accuracy	99.9%	100%	96.7%	100%	89.9%
Lowest Accuracy	66.8%	63.3%	55%	74%	67.6%
Average Accuracy	83.35%	81.65%	75.85%	87%	78.7%

Table 2.: Show the prediction accuracy that uses classification method grouped by algorithms for predicting student performance.

VI. CONCLUSION:

Education is very important in today's generation, and how to analyze the school's education system and predict the progress of educational institutions is very important. The focus of the proposed automation system is to predict the progress of the institution. The proposed automation system focuses on making predictions about students' academic performance and social behavior. The accuracy of the model is also calculated. Another area is to use a support vector machine (SVM) to create a classifier and analyze suitable classifiers to perform the classification.

VII. FUTURE WORK:

In the future, one could list out the areas on which should supposed to pay attention to,to improve the efficiency of student performance. You can derive a user interaction model to dynamically provide student records and alert staff about poor grade students. You can make predictions with neural networks and expect improvisational results. You can add non-academic attributes along with academic attributes. In the proposed system, a naive Bayes algorithm is used for classification and prediction. Similarly, many such algorithms can be used to classify and predict student behavior and academic performance. Due to the small volume of the dataset under consideration, future work will use larger datasets and compare the results with other classification algorithms.

VIII. REFERENCE

[1] Elbadrawy, Asmaa and Polyzou, Agoritsa and Ren, Zhiyun and Sweeney. "Predicting Student Performance Using Personalized Analytics", IEEE, vol. 49,pp. 61-69, Apr.2016.

[2] Ishan A,Abu Amra,Ashraf Y.A."Student performance prediction using KNN and Naïve Bayesian",IEEE,May 2017.

[3] Nikita Gorad, Ishani Zalte,Aishwary Nandi, Deepali Nayak, "Career Counselling Using Data Mining", International Journal of Innovative Research in Computer and Communication Engineering,2017.

[4] Ganeshan, Kathiravelu and Li, Xiaosong. "An intelligent student advising system using collaborative filtering", 2015 IEEE Frontiersin Education Conference (FIE), pp. 1-8, Oct. 2015.

[5] Lopez Guarin.Camilo Ernesto."A Model to predict Low Academic Performance at a Specific Enrollment Using Data Mining",IEEE Revista Iberoamericana de Tecnologies del Aprendizaje.Aug 2015.

[6] Mohammed M.Abu Tair and Alaa M.El-Halees, 'Mining Educational Data to Improve Students' Performance: A case Study', International journal of information and Communication Technology Research, ISSN: 2223-4985, vol.2 no.2, February 2012.

[7] F. H. D. Araujo, A. M. Santana and P. A. S. Neto, "Evaluation of Classifiers Based on Decision Tree for Learning Medical Claim Process,"in IEEE Latin America Transactions, vol. 13, no. 1, pp. 299306, Jan. 2015.

[8] H. Witten, E. Frank and M. A. Hall, Data mining: Practical Machine Learning Tools and Techniques.Burlington: Morgan Kaufmann, 2011.

[9] B. K. Baradwaj and S. Pal,"Mining Educational Data to Analyze Students' Performance," inInternational Journal of Advanced Computer Science and Applications – IJACSA, vol. 2, no. 6, pp. 63-69, 2011.

[10] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: an update," in ACM SIGKDD Explorations Newsletter, vol. 11, no. 1, pp. 10-18, Jun. 2009.

[11] F. H. D. Araujo, A. M. Santana and P. A. S. Neto, "Evaluation of Classifiers Based on Decision Tree for Learning Medical Claim Process," in IEEE Latin America Transactions, vol. 13, no. 1, pp. 299306, Jan. 2015.

[12] Sayali Rajesh Suyal and Mohini Mukund Mohod, "Quality improvisation of student performance using data mining techniques", International Journal of Scientific and Research Publications, vol 4, iss 4, April 2014.

[13] T.Jeevalatha, N. Ananthi and D.Saravana Kumar, "Performance analysis of undergraduate students placement selection using Decision Tree Algorithms", International Journal of Computer Applications (0975- 8887), vol 108, December 2012.

[14] Ryan J.D.B.Baker and Kalina Yacef , "The state of educational data mining in 2009: A review and future revisions", Journal of Educational Data Mining , Vol.1,No.1, February 2009.

[15] A.Dinesh Kumar and V.Radhika, "A survey on predicting student performance", International Journal of Computer Science and Information Technologies, Vol. 5, 2014.

[16] M.S. Mythili1 and A.R.Mohamed Shanavas , "An analysis of students' Performance using classification algorithms ", IOSR-JCE, Volume 16, iss1, Jan. 2014.

[17] Mukesh Kumar, A.J.Singh, Disha Handa,"Literature Survey on Student's Performance Prediction in Education using Data Mining Techniques", IJEME,Vol.7, June 2017.