IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2347-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 9, September 2022

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.118

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 9, September 2022||

| DOI:10.15680/IJIRSET.2022.1109031 |

Student Performance Prediction Using A Random Forest Regressor

Md.IRSHAD HUSSAIN B[1*], BASAVARAJGV[2]

Assistant Professor, Department of Master Applications, University B.D.T College of Engineering, Davanagere, Karnataka, India

Student, Department of Master of Applications, University B.D.T College of Engineering, Davanagere, Karnataka, India

ABSTRACT: It is impossible to exaggerate the value of education for both a person and the nation as a whole. Many students withdraw from various academic courses each year. This study explores the relationship between student demographics and academic achievement. The ultimate performance of the pupils is predicted using the Random Forest regressor method. These variables are assessed along with their impact on student performance using readily available datasets that include a variety of demographic data.A training data set is necessary to apply this approach, and the needed dataset has been obtained from reliable sources. Each student's test results, which are their particular student parameters, is compared to the trained data set and a prediction is made. Web technologies can be used with an application that has been made. It was produced as a product for a school and has students and administrators as actors. The product was created using the Python programming language and tested locally using the Flask web framework.

Key words: performance prediction,data mining for schooling and a random forest regressor.

I. INTRODUCTION

Education is one of a person's core needs. Elementary, middle, and lower secondary schools are some of the levels. School dropout at any level is another common occurrence. However, the dropout rate increases as pupils go to higher levels, notably at the primary level. Rising dropout rates might be greatly influenced by a variety of factors. In the majority of cases, when kids' grades start to slip, it is the main element that raises the risk that they will be yanked out. Finding the several factors that affect first-year students' performance at the primary level is the major objective of this study [1].

Numerous approaches, including K-Nearest Neighbor, Decision Tree, Random Forest, Naive Bayesand many more, are used in educational data mining. A variety of data kinds, including regression, classification, and association systems, may be disclosed by abusing these approaches. They showed how to gather information in a reasonable manner, how to prepare it, how to use data processing techniques on it, and how to use the information that was eventually found. A number of data styles can be discovered using the obtained data. In this era of vast information, a lot of structured and unstructured student data is generated every day. Analyzing large volumes of data is difficult andrequire software that runs on many machines in high parallel. Prior studies looked at the most common ones, such as attribute choice and classifications. Data mining techniques can be used to anticipate rates of school failure and dropout, according to preliminary study in a manner similar to the decision-tree method. However, this task cannot handle very large amounts of data. Additionally, the categorization operation takes a long time to conclude. As a result, the temporal complexity is high [2].

The suggested solution uses a technique known as artificial intelligence, which takes use of current technological developments. In this technique, scientific data gathered from subject matter specialists is examined, a model is created, then deployed at a web server for prediction purposes. EDA is used to create a data collection including a variety of facts, after which a model is produced. Random Forest regressor technique will be used to forecast the outcome.

II. LITERATURE SURVEY

S.J.Romero [3] A new field of study called "educational data mining" combines data mining with education to glean useful knowledge from vast educational databases. Educational institutions and students can improve their academic performance with the help of educational data analysis. With the use of educational technology, it is now possible to

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 9, September 2022||

| DOI:10.15680/IJIRSET.2022.1109031 |

analyse virtual learning environments, students' interactions with e-learning systems, and crucial elements of the learning process as well as to forecast students' test results.

S.Vranes.[4] Academic performance of students is significantly influenced by their final test scores. In order to stop kids from dropping out, it is crucial to identify their weak areas early. Early warning systems might benefit both teachers and students by enhancing their final test marks and learning.Various machine learning techniques have been applied in recent years to forecast student performance. However, different studies employed academic and non-academic characteristics of individual pupils to make predictions about the outcomes. To forecast how students would perform on their final exams, a research is undertakenDecision Tree, Support Vector Machine (SVM), K-Nearest Neighbors (KNN), Artificial Neural Network (ANN), and Linear Regression are seven supervised learning classifiers.are compared to determine which prediction approach is the best. Performance on the final exam is predicted using the demographic, assessment, and click data from the Open University Learning Analytics Dataset.

P,Kavipriya.[5] The requirement to forecast students' academic achievement has emerged as a crucial element in raising the standard of academic curricula, helping students while they study, and giving tutors additional options when instructing their pupils. There have been several publications released recently that deal with this topic. We located multiple studies of the literature that explored various aspects of the modelling of student academic performance. Utilizing various data mining techniques, the review focuses on the prediction, analysis, early warning, and assessment of the students' performance.
Jadhav and Channe [6] two experiments were carried out to forecast the student's ultimate grade. In the first experiment, three datasets were used to test classification methods. The findings showed that including all available data was more accurate than filtering. In the second trial, they came to the conclusion that using a classification model for both numerical and categorical data can yield the greatest accuracy.VinayakHegde[7]In the commercial sector, data mining is crucial, and it aids educational institutions in forecasting and making judgments on the academic standing of their students. Higher education has seen an increase in student dropout rates recently, which has an impact not just on the student's career but also on the institution's image.

III. METHODOLOGY

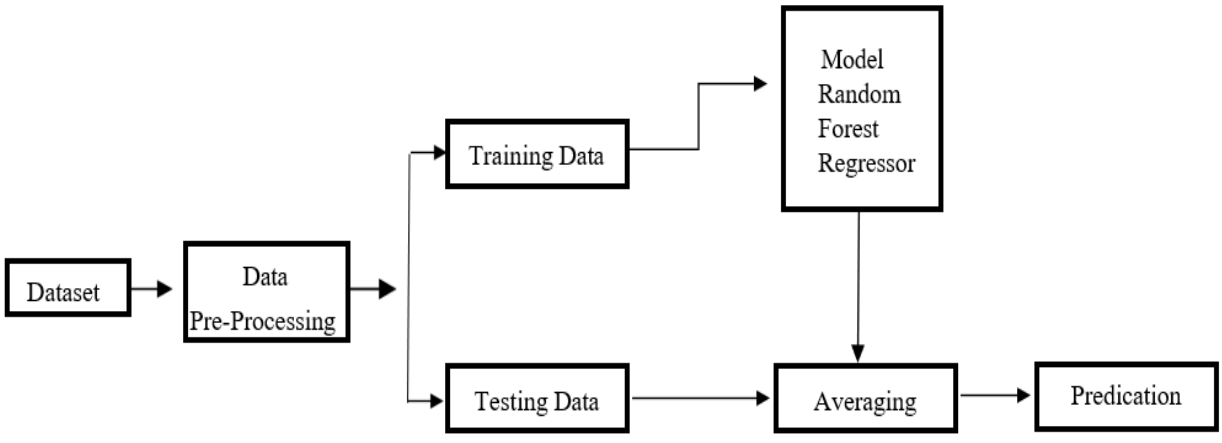

Fig.1.: Model for Data Analysis and Prediction Schematic [9]

The graphic above illustrates the system's technique. Important datasets are pulled from the Kaggle repository, analysed with pandas and other python packages, then divided into training and test data to gauge accuracy. Random Forest Regression, a supervised learning methodology that employs ensemble learning for model construction, is then fed this data to create the model. We employ the flask web framework, which offers practical features and tools that make it simpler to develop online applications in Python. Because you can simply construct a web application using just one Python file, it allows developers flexibility and is a more approachable framework for beginning developers.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 9, September 2022||

| DOI:10.15680/IJIRSET.2022.1109031 |

	A	B	C	D	E	F	G	H	I
1	gender	race/ethni	parental le	lunch	test prepa	math scor	reading sc	writing score	
2	female	group B	bachelor's	standard	none	72	72	74	
3	female	group C	some colle	standard	completed	69	90	88	
4	female	group B	master's d	standard	none	90	95	93	
5	male	group A	associate's	free/reduc	none	47	57	44	
6	male	group C	some colle	standard	none	76	78	75	
7	female	group B	associate's	standard	none	71	83	78	
8	female	group B	some colle	standard	completed	88	95	92	
9	male	group B	some colle	free/reduc	none	40	43	39	
10	male	group D	high schoo	free/reduc	completed	64	64	67	
11	female	group B	high schoo	free/reduc	none	38	60	50	
12	male	group C	associate's	standard	none	58	54	52	
13	male	group D	associate's	standard	none	40	52	43	
14	female	group B	high schoo	standard	none	65	81	73	
15	male	group A	some colle	standard	completed	78	72	70	
16	female	group A	master's d	standard	none	50	53	58	
17	female	group C	some high	standard	none	69	75	78	
18	male	group C	high schoo	standard	none	88	89	86	
19	female	group B	some high	free/reduc	none	18	32	28	
20	male	group C	master's d	free/reduc	completed	46	42	46	
21	female	group C	associate's	free/reduc	none	54	58	61	
22	male	group D	high schoo	standard	none	66	69	63	
23	female	group B	some colle	free/reduc	completed	65	75	70	
24	male	group D	some colle	standard	none	44	54	53	
25	female	group C	some high	standard	none	69	73	73	
26	male	group D	bachelor's	free/reduc	completed	74	71	80	

Fig.2.: STUDENT DATASET

The student dataset collected from the Kaggle website is shown in the above image.a list of students from a certain institution who scored well on tests in math, reading, and writing, together with information about their gender, colour and ethnicity, parental education level, meal payment type, and test preparation course description. 1000 rows and 8 columns make up the structure.

IV. RANDOM FOREST ALGORITHM

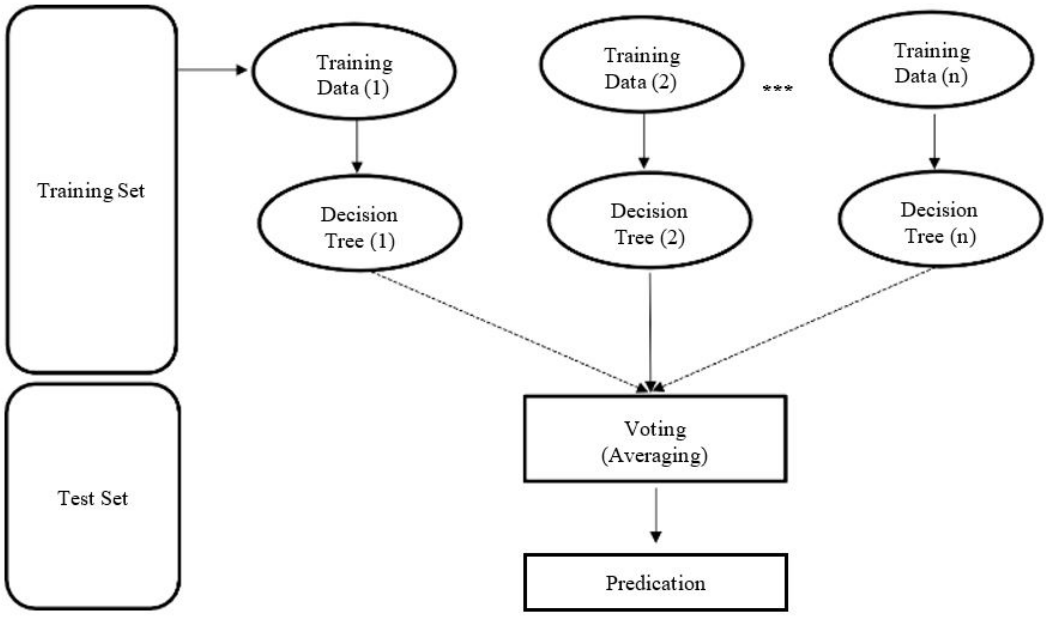

Fig 3.: Random Forest Algorithm

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 9, September 2022||

| DOI:10.15680/IJIRSET.2022.1109031 |

This method combines many trees to predict the class of the dataset. There is a chance that certain decision trees will provide the correct outcomes while others won't.In order for the feature variable of the dataset to predict accurate outcomes rather than assumed results, there should be some real variables in it. Each tree's prediction must have very little correlation. When compared to other algorithms, using this one requires less training time. It predicts the results with high accuracy, performs well with huge datasets, and retains accuracy even when significant amounts of data are absent.

The following stages may be used to understand how the random forest algorithm operates:

- **Step 1:** Begin by choosing random samples from the provided dataset.
- **Step 2:** Next, the algorithm will form a decision tree for every sample**.**
- **Stage 3:** Voting will be done for each expected outcome in this step**.**
- **Step 4:** Finally, choose the prediction result that received the most votes as the final prediction result**.**

Random Forest Regression When compared to other types of regression approaches, algorithm is a powerful algorithm. A function y = mx + c is used to represent a linear regression technique [18]. These basic functions are not adequate to represent a random forest algorithm. When compared to other kinds of algorithms, they generate superior outcomes. They work well with huge datasets and generate estimates to deal with missing data. The main danger posed by random forest regression is that they cannot be used outside of their typical range.

A random forest is just a collection of decision trees that have been joined together at random to form a random forest algorithm. A different sample of rows is used to construct each tree. At each node, a unique collection of features is picked for splitting. Every tree generates a distinct, independent prediction. The aggregate of all of these outcomes yields the final result.

When compared to other algorithms, the random forest method performs better thanks to the averaging process. Accuracy is increased, and overfitting is decreased, via the averaging process. It is an average of the forecasts made by all the forest's trees.

There are few options in order to deal with the extrapolation problem. Instead of random forest regression, Because deep learning models can handle the extrapolation problem, we may employ alternative regression models such support vector machine regression, linear regression, etc. There are several stacking approaches that may be used to combine prediction findings. A linear model and random forest regressor can be used to build a stacking regressor. A completely nonparametric prediction algorithm is the Random Forest algorithm. It does not effectively take into account how the response and predictors are related.

Advantages:

We may use the random forest regression approach when there is a nonlinear trend in the data. We can use the random forest regression approach if extrapolation outside the training set is not crucial.

Disadvantages:

If the data is in the form of a time series, this approach cannot be employed. A random forest regressor is unable to identify an upward or downward trend, which is a need for any time series problem.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 9, September 2022||

| DOI:10.15680/IJIRSET.2022.1109031 |

V. RESULT& DISCUSSION

Input	Result
1. Gender 2. Race 3. Parents education 4. Payment mode of lunch 5. Test preparation course	The input parameters are analyzed and prediction is done

Table.1.: Student prediction form

In the above table random forest regressor is used to analyze the data and make prediction.

Input Parameters					Prediction Result		
Gender	Race /Ethnicity	Parental level education	Type of Lunch	Test preparation Course	Maths Score	Reading Score	Writing Score
Male	Group A	Masters'	Standard	Complete	77.0	75.0	75.0
Female	Group A	Masters'	Standard	Complete	72.0	82.0	84.0
Male	Group B	Bachelors'	Standard	Complete	77.0	75.0	76.0
Female	Group C	Associate	Free	None	54.0	66.0	65.0
Male	Group D	College	Standard	Complete	79.0	75.0	77.0
Female	Group E	High School	Standard	Complete	75.0	79.0	79.0
Male	Group D	Bachelors'	Standard	None	76.0	71.0	71.0
Female	Group C	High School	Free	None	51.0	63.0	61.0
Male	Group B	College	Standard	None	70.0	64.0	62.0
Male	Group E	Masters'	Standard	Complete	85.0	79.0	79.0
Male	Group D	High School	Free	None	57.0	56.0	59.0
Male	Group A	College	Free	None	59.0	58.0	54.0

Table2.: Student Marks Prediction Result

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 9, September 2022||

| DOI:10.15680/IJIRSET.2022.1109031 |

VISUALIZATION

Fig.4.: SHOWS VISUALIZATION OF EXPLORATORY DATA ANALYSIS

Above figure we use With just two lines of code, the free source Python module Sweetviz creates stunning, high-density visuals to jump-start exploratory data analysis. The final product is a totally independent HTML application. The technology is designed to compare datasets and instantly see desired values. Exploratory data analysis is the process of summarising the data using statistical and graphical approaches in order to focus on key elements of the data for additional study.

VI. CONCLUSION

In today's manufacturing, education is absolutely vital, and it's crucial to assess school teaching practises and make projections for business growth. The intended automated technique places a lot of emphasis on forecasting the behaviour of pupils in order to grow the company. The intended machine-driven approach stresses student prediction. Academic success and overall conduct. The form's precision is based on calculated boots.Any range should create a classifier using a support vector machine and research how well the classifiers work for classifications.

VII. FUTUREWORK

The projected method, Random Forest regressor algorithmic rule program is used for assortment and prophecy. the same, many such algorithm area unit typically use for classifications and predictions of student performance and his instructional presentation. The dataset thought-about has less volume, thus the longerterm work should use huge dataset and compare the results victimization different algorithm for classifying things.

REFERENCES

[1] P.D. Gil et al., a method based on statistics that forecasts first-year college students' performance in the classroom. Information technologies and education, 2020.
[2] V.E.J. Chaves et al. The International Conference on European Transnational Education will feature an analysis of student achievement scores using cluster analysis in 2020.
[3] S.J. Romero, S. Pamplona, and J. Bravo-Agapito A five-year research on the early prediction of undergraduate students' academic achievement in fully online learning. Technology and Human Behavior.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 9, September 2022||

| DOI:10.15680/IJIRSET.2022.1109031 |

[4] Vranes, S. a description and comparison of supervised data mining methods for forecasting student test performance. 2020: Computers in Education.
[5] P. Kavitha, "A Review on Predicting Students' Academic Performance Earlier, Using Data Mining Techniques," International Journal of Advanced Research in Computer Science and Software Engineering, Volume 6, Issue 12, December 2016, ISSN: 2277 128XS.D.JadhavandH.P.Channe,"ComparativeStudyofK-NN,NaiveBayesandDecisionTree Classification Techniques," vol. 5, no. 1, pp. 2014–2017, 2016.
[6] Vinayak Hegde and S. Pal,"Mining Educational Data to Analyze Students' Performance," in InternationalJournalofAdvancedComputerScienceandApplications –IJACSA,vol.2,no.6,pp. 63-69, 2011.
[7] H.Witten,E.FrankandM.A.Hall,Datamining:PracticalMachineLearningToolsand Techniques.Burlington: Morgan Kaufmann, 2011.
[8] BasavarajGV,Md.IrshadHussainB,"Surveyonstudentperformanceanalysissystembasedon the machine learning" in IJIRCCE Volume 10, Issue 5, May 2022.
[9] M.Hall,E.Frank,G.Holmes,B.Pfahringer,P.Reutemann,andI.H.Witten,"TheWEKAdata mining software: an update," in ACM SIGKDD Explorations Newsletter, vol. 11, no. 1, pp. 10-18, Jun. 2009.
[10] F.H.D.Araujo,A.M.SantanaandP.A.S.Neto,"EvaluationofClassifiersBasedonDecision Tree for Learning Medical Claim Process," in IEEE Latin America Transactions, vol. 13, no. 1, pp. 299306, Jan. 2015.
[11] Sayali Rajesh Suyal and Mohini Mukund Mohod, "Quality improvisation of student performanceusingdataminingtechniques",InternationalJournalofScientificandResearch Publications, vol 4, iss 4, April 2014.
[12] T.Jeevalatha, N. Ananthi and D.Saravana Kumar, "Performance analysis of undergraduate studentsplacementselectionusingDecisionTreeAlgorithms",InternationalJournalofComputer Applications (0975- 8887), vol 108, December 2012.
[13] RyanJ.D.B.BakerandKalinaYacef,"Thestateofeducationaldataminingin2009:Areview and future revisions", Journal of Educational Data Mining , Vol.1,No.1, February 2009.
[14] A.DineshKumarandV.Radhika,"Asurveyonpredictingstudentperformance",International Journal of Computer Science and Information Technologies, Vol. 5, 2014.
[15] M.S.Mythili1andA.R.MohamedShanavas,"Ananalysisofstudents'Performanceusing classification algorithms ", IOSR-JCE, Volume 16, iss1, Jan. 2014.
[16] MukeshKumar,A.J.Singh,DishaHanda,"LiteratureSurveyonStudent'sPerformancePrediction in Education using Data Mining Techniques", IJEME,Vol.7, June 2017.
[17] Panteche-learning,Masterclass"PredictionRandomForestRegressor" ;::https://youtu.be/y6RippbvA48.

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details