Forecast Pupil Performance using SVM & Logistic Regression

UPPULA MUKUL KUMAR
M.Tech Student, Malla Reddy College of
Engineering and Technology, Hyderabad,T.S, India.

Mr. M SANDEEP
Associate Professor CSE, Malla Reddy College of
Engineering and Technology, Hyderabad,T.S, India.

Dr. M JAYAPAL
Associate Professor CSE, Malla Reddy College of Engineering and Technology, Hyderabad,T.S, India.

Abstract: - **The primary reason why machine learning has gained so much prominence nowadays is that it enables accurate and reliable decision making by extracting hidden relationships between various features present in the data. For this purpose the technique such as supervised methodologies and unsupervised methodologies are used. For this reason, machine learning can be used in almost any area of work to help in proper decision making and predictions. In our current project, we are trying to predict this student performance that utilizes supervised machine learning methodologies like support vector machines , logistic regression, random forests etc. We have also tried to publish this model to a web application so that it can be used by the academic community. The information extracted and the knowledge gained by extracting information from the educational data set would be helpful for predicting the student grades and their future performance. The main intention of the project used to predict student performance beforehand and help them get good grades in future. This would help in increasing the motivation levels of the students, improving their grades, decreasing their dropout ratio, and preparing better students for a better world.**

I INTRODUCTION

Data analytics in the area of learning and education has a significant role in refining the present education system so that the underlying problems with the present education system can be discovered from various perspectives like administration perspective , perspective of the faculty, perspective of the students. Discoveries of data analytics helps in proper planning, understanding the real time issues, and taking appropriate decisions that help in refining the system [1]. Due to the immense advantages that it offers, data analytics has paved its way into the educational systems for activities like knowledge discovery from the existing databases, making predictions, text mining, artificial intelligence etc. The primary objective of using learning analytics used to bring a shift from the existing learning practices to a technological paradigm.

Lot of student data has been gathered over the years and has been stored in the database but one is unable to bring out any useful information from the existing data and use it for analytical purposes. The data present in various source systems have to be integrated together so that the power of analytics could be utilized in the field of education. these factors have increased the research conducted in the field of learning analytics as there are various avenues to explore.

Applying data analytics over educational data helps in understanding the week areas of the students at an individual level, gathering information about the performance of the faculties, proper allocation of resources and planning appropriately to mitigate the pain points and take necessary actions improve the existing scenarios. It is important to understand that the results of these analytics help in pinpointing the issues from various perspectives but things do not improve unless and otherwise all the people in the system like parents, students, teachers and the administrators come together to take appropriate decisions and implement them.

In this project, we have tried to analyze various features that are impacting the grades of the students in the educational data set in order to understand the underlying situations and help the teachers to take necessary actions I got that uplift the performance of the students [2]. We have also tried to publish this model to a web application students and faculty can use it for future predictions and improving the grades of the students.

AIM OF THE THESIS

The main aim of the project is to forecast pupil performance by applying machine learning techniques in the field of education. We have used the Kalboard data set and fed it to the proposed model that is built using advanced machine learning techniques.

SCOPE OF THE THESIS

The scope of the project is limited to the compute the accuracy of the proposed model and predict pupil performance. The admin of the system trains the proposed model with training data and tests the model with test data. The users of the system can be students

or faculty who can register themselves on the system and utilize the prediction model.

II LITERATURE SURVEY

2.1 "Mining Students Data To Analyze Learning Behavior: A Case Study"

The process of acquiring knowledge from the educational data from educational organizations is called educational data mining. In this paper a case study has been presented where students data from the database course has been acquired and preprocessed in order to gain insights into the data and make appropriate predictions and take associated decisions in order to improve student performance. data preprocessing is followed by application of various machine learning techniques such as discovery of associations between the features, classification, clustering of data points, detection of outliers that impact the predictions and predict the student performance.

Educational data mining is proved to be an emerging area of research because of the vast scope of improvement that it offers in the field of education. For the purpose of educational mining, data can be collected from any educational institutions are e-learning systems. The data can come from multiple source systems such as personal data academic data etc and all these would be required to do a complete exploratory data analysis. In this paper many machine learning techniques have been applied to discover the associations between the features, classification common knowledge discovery, outlier detection and prediction [3]. Some of such techniques are decision trees , KNN, artificial neural networks etc.

2.2 "Educational Data Mining: A Review of the State of the Art"

Educational data miningis a field of analytics which amalgamates machine learning, data mining and statistical techniques over educational data gathered from various sources like online educational platforms, educational institutions, Details from course data etc. The main goal or primary objective off educational data mining is to discover underlying issues by analyzing patterns in the data and provide solutions in order to revive the existing educational system and thereby improve the performance of students.

This paper elaborately describes the different kinds of machine learning techniques used to solve problems from various scenarios and has concluded that machine learning isan effective tool that can be used to bring technological advancements in the field of education and improve and refine the existing system.

III SYSTEM ANALYSIS

Existing system:

With the tremendous growth in technology, there has been a paradigm shift in the field of education also. Most of the processes have been digitized in order to make a system more effective and efficient [4]. Nowadays, lots of data is collected in educational databases, but it remains unutilized. In order to get required benefits from such a big data, powerful tools are required. The traditional educational systems are incapable of making any predictions with respect to the student performance.

Disadvantages:

- Not efficient
- Not capable of making predictions
- Does not employ advanced techniques to utilize the educational data.

Problem Statement:

Lot of student data has been gathered over the years and has been stored in the database but one is unable to bring out any useful information from the existing data and use it for analytical purposes [5]. The data present in various source systems have to be integrated together so that the power of analytics could be utilized in the field of education. these factors have increased the research conducted in the field of learning analytics as there are various avenues to explore.

Proposed System:

We propose to develop and deploy a web application which can predict the academic performance of a student based on various parameters [6]. The machine learning model would be deployed to web where the student can upload and view his academic parameters and receive predictions. We aim to build an automated prediction system which can accurately classify if the academic grade of the student and make this information available to students, parents and teachers.

Advantages:

- Automated system
- Can predict student grade before the beginning of the course
- Utilizes the large amount of data that is generated through educational systems.

IV IMPLEMENTATION

Below is the proposed modular implementation of the project. It consists of two modules:

1. Admin
2. Student

Admin Module:

The admin of the system is responsible for the activities like:

1. Uploading the dataset

2. Data Analysis of the dataset

3. Splitting the dataset for training and testing

4. Training the model for logistic regression and SVM

5. Review the performance of the algorithms on the given dataset

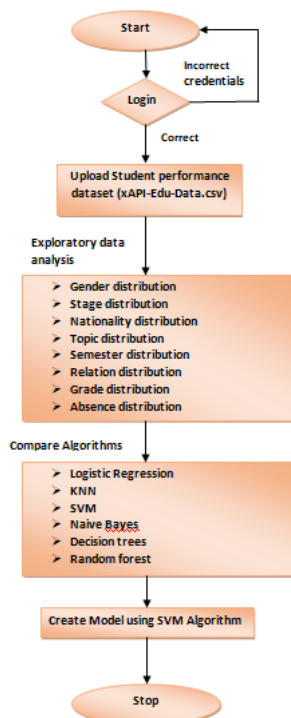6. View student details and predictions

Student Module:

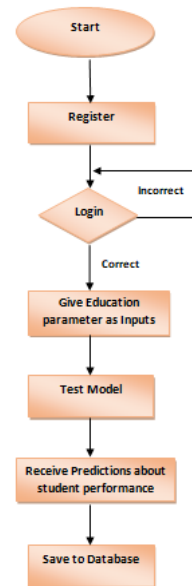The user of the system can utilize the machine learning services that are offered like:

1. Logging into the system

2. Save one's academic parameters to the database

3. Receive predictions about the academic performance.

V SYSTEM DESIGN

Data Flow Diagram: Admin

Data flow Diagram: User

Use Case Diagram: Admin
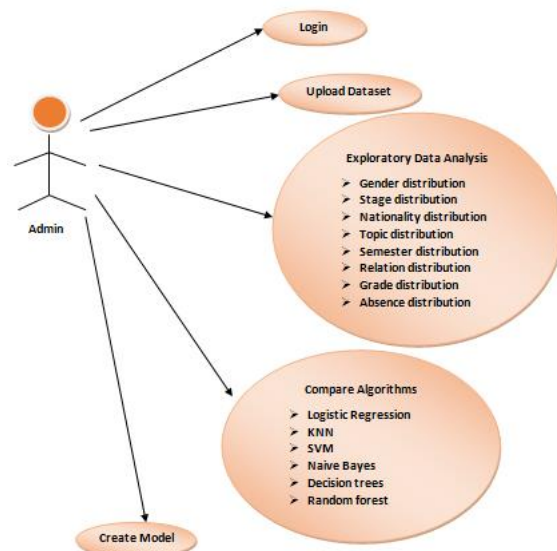
Use Case: User

Sequence Diagram:
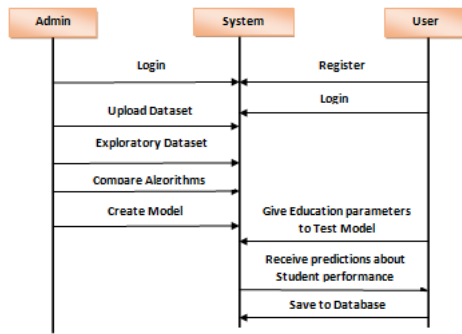
VI PROJECT EXECUTION

Upload Dataset:
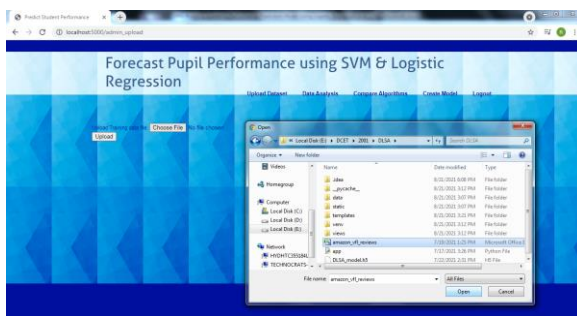
Data Analysis:
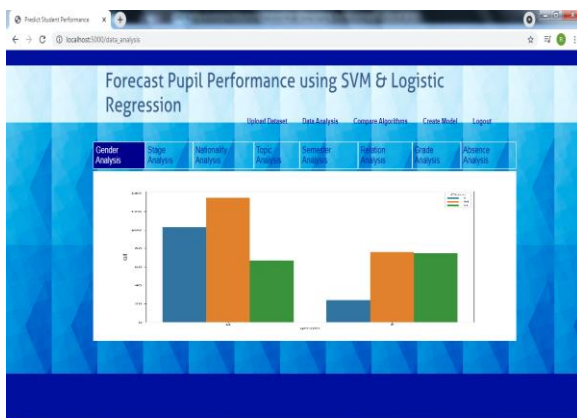
Stage Analysis:

Nationality Analysis:
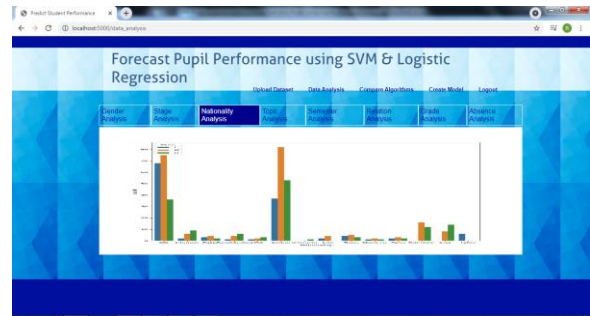
Topic Analysis:
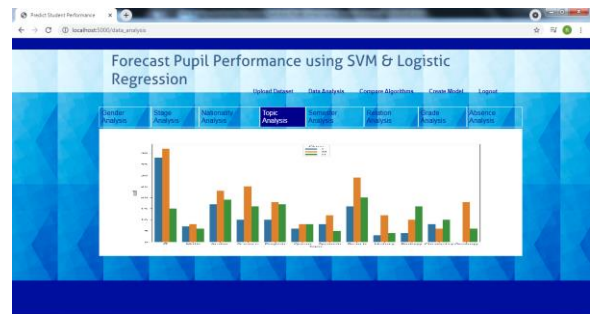
Semester Analysis:
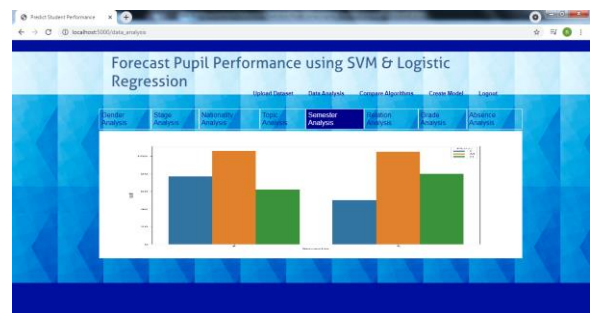
Relation Analysis:
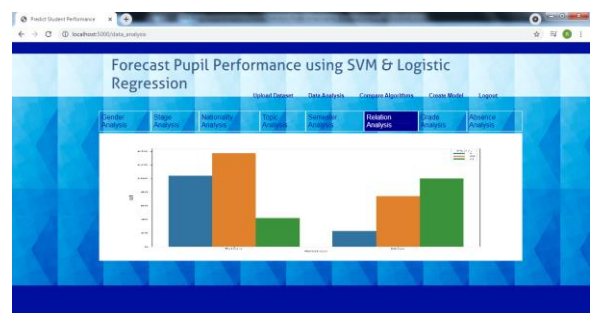
Grade Analysis:
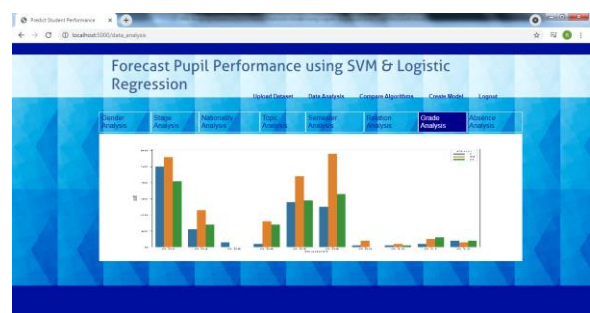
Absence Analysis:
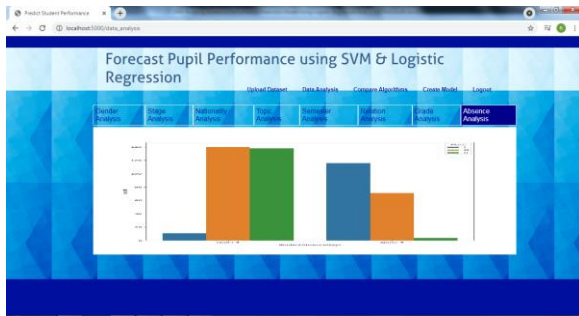
Decision Trees:
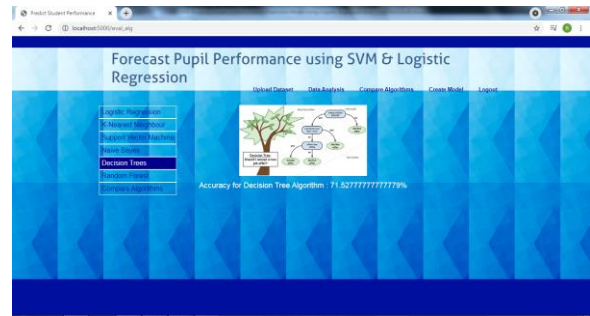
Compare Algorithms:
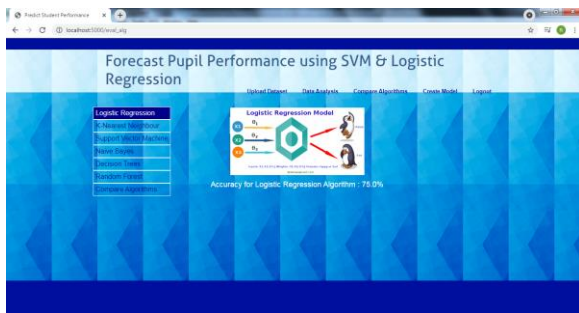
Random Forest:
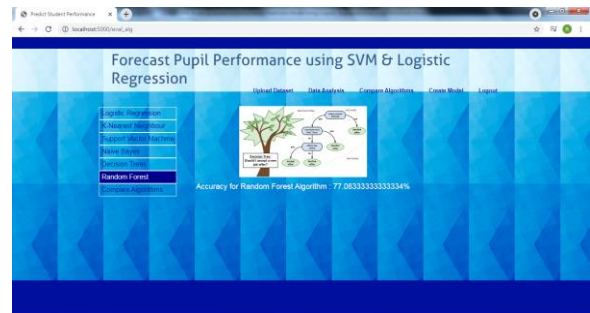
K-Nearest Neighbour:
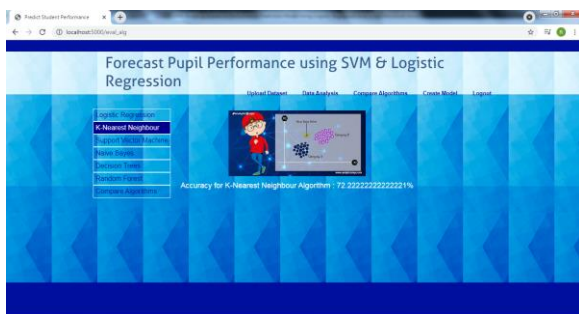
Compare Algorithms:
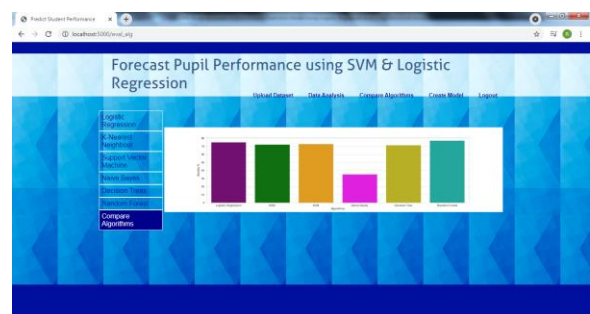
Support Vector Machine:
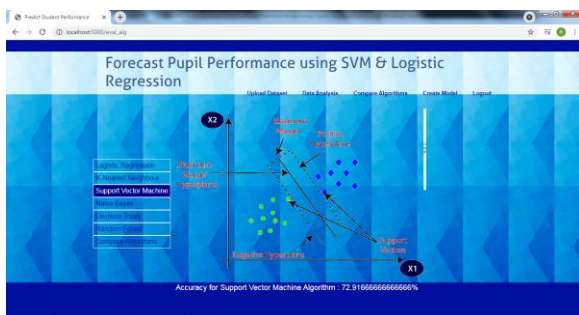
Create Model:
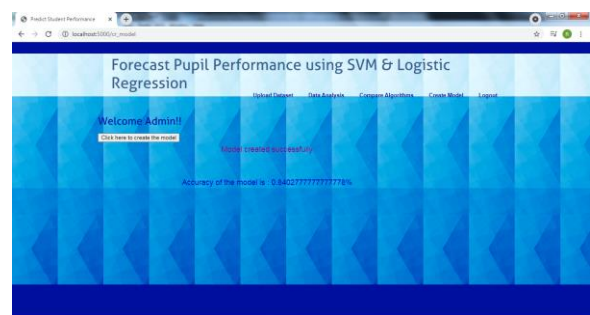
Naive Bayes:
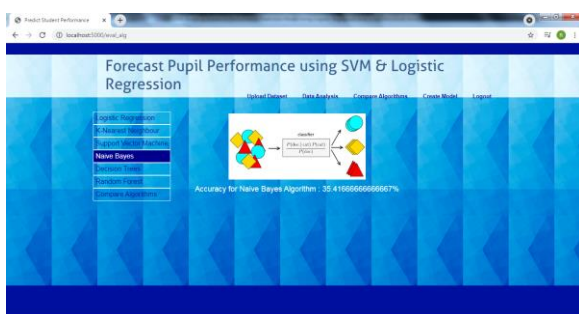
User Registration:

User Login:
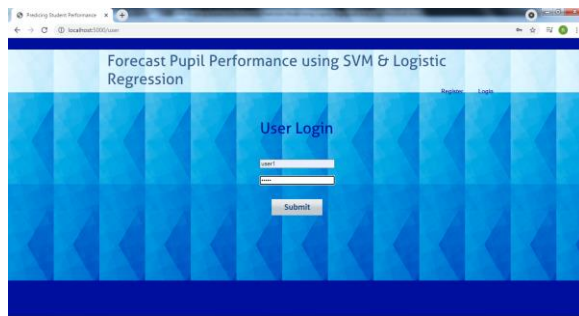
VII CONCLUSION

A successful career of a student mainly depends upon their excellence in academics. Most of the countries moved their typical educational system into e-learning systems. Thus, during distance learning, it is very hard to analyze student's behaviour towards their studies, participation in quizzes or discussion groups and taking part in classroom activities and instructors. By analyzing and extracting the hidden knowledge that is obtained through these learning management systems can give very good benefits to academic institutions. And by utilizing this type of knowledge to work for the improvement of the student's academic excellence by predicting their future performance and taking proper interventions. In the following research, we have studied on student model and the two classifiers that help evaluate the model performance is logistic regression and support vector machine. These classifiers are selected by using two different Technologies based on the features that are selected through the feature selection method. The three major categories which are the factors that influence the student's academic grades are satisfaction level of student interaction with the system and punctuality in the classroom. The results for the about three factors are obtained through gain ratio feature selection and support vector machine. The results that are obtained for the prediction of student's future performance will give better results if we show the support vector machine algorithm using sequential minimal Optimization. We also propose to deploy the model to a web application and make it available to the students, parents and faculty in order to minimize dropout rates and help the students in their weak areas.

REFERENCES

[1]   N.Padhy , D.Mishra ,R. Panigrahi , "The survey of data mining applications and feature scope," arXiv preprint arXiv:1211.5723, November, 2012.

[2]   R.S.Baker, S.Gowda, and A.Corbett, "Automatically Detecting a Student's Preparation for Future Learning: Help Use Is Key," In Proceedings of the 4th International Conference on Educational Data Mining, pp. 179–188,2011.

[3]   R.S.Baker and K. Yacef, "The State of Educational Data Mining in 2009: A Review and Future Visions," Journal of Educational Data Mining, vol.1(1) , pp. 3–17,2009.

[4]   A.Merceron,K. Yacef, C. Romero, S.Ventura and M. Pechenizkiy, "Measuring correlation of strong symmetric association rules in educational data," Handbook of educational data mining, pp. 245-256, 2010.

[5]   A.El-Halees, "Mining Students Data to Analyze Learning Behavior: A Case Study," The 2008 international Arab Conference of Information Technology (ACIT2008) Conference Proceedings, University of Sfax, Tunisia, December , 2008.

[6]   C.R.Romero and S. Ventura, "Educational Data Mining: A Review of the State of the Art," IEEE Transactions on Systems, Man and Cybernetics, Part C: Applications and Reviews, vol.40 (6), pp.601– 618, 2010.

[7]   A.Wolff, Z.Zdrahal, D.Herrmannova, J. Kuzilek and M. Hlosta, "Developing predictive models for early detection of at-risk students on distance learning modules," 2014.

[8]   E.A.Amrieh, H.Thair, A.Ibrahim, "Preprocessing and analyzing educational data set using X-API for improving student's performance," 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT), December, 2015.

[9]   H.Agrawal and H.Mavani, " In Student Performance Prediction using Machine Learning," International Journal of Engineering Research and Technology,2015.

[10]   E.AAmrieh, H.Thair, A.Ibrahim, "Mining Educational Data to predict student's academic performance using ensemble methods," International Journal of Database Theory And Application , Vol. 9 , 2016.