

Predicting Academic Future Courses for Students Using Machine Learning

Preetham S.¹, Dr. Chandrashekhar M C², Dr M N Eshwarappa³

¹Department of Electronics and Communication Engineering Sri Siddhartha Institute of Technology, Tumkur-05

²Professor, Department of Electronics and Communication Engineering Sri Siddhartha Institute of Technology, Tumkur-05

Abstract: "Predicting Academic Future Courses for Students Using Machine Learning" should provide an overview of the research objectives, methodologies, findings, and implications. This study explores the application of machine learning techniques in predicting the most suitable academic future courses for students based on their past academic performances, interests, and career aspirations. The primary objective is to develop a predictive model that helps academic advisors and institutions provide tailored course recommendations, thereby improving students' academic success and satisfaction. Various machine learning algorithms, including decision trees, support vector machines, and neural networks, are evaluated to determine the optimal approach for accurate course predictions. The dataset includes a range of student data, such as grades in prerequisite courses, previous course selection patterns, extracurricular interests, and feedback on prior courses. Feature engineering and data preprocessing are emphasized to handle challenges like imbalanced data and missing values. The research shows that machine learning models can achieve high predictive accuracy, with the neural network model performing best under specific conditions. The findings suggest that machine learning-based course recommendation systems can significantly enhance educational planning by aligning course selections with students' strengths and preferences, ultimately leading to improved educational outcomes. This study contributes to the growing field of educational data mining and demonstrates the potential for AI-driven approaches to personalize academic pathways in higher education.

Keywords: E-learning, Machine Learning, Naïve Bayes, Evaluation, Predictive modelling, Support vector machines, Data preprocessing

1. Introduction

Exploiting the ceaselessly enhancing, online learning frameworks assumes a paramount part for adjusting toward oneself, particularly on account of working individuals. All things considered, learning frameworks don't for the most part adjust to learners' profiles. Learners need to invest a ton of time before arriving at the learning objective that is perfect with their insight foundation. This paper investigates Machine learning and its applications in E-Learning frameworks. Machine learning, we may say, is a sort of artificial intelligence (AI) that gives machines the capacity to learn without being expressly customized. Machine learning concentrates on the advancement of machine projects that can develop themselves and change when presented to new information. Distance learning, at times called e-learning, is a formalized showing and learning framework particularly intended to be completed remotely by utilizing electronic correspondence (Mayer, 2001). The expanding interest of web learning and learners has required the customary configuration of E-learning interfaces to be extended past the center of the conventional visual channel, to join intelligent allegories as well as excitement components. Edutainment is a paramount element connected with enhancing learning results, in light of the fact that it incorporates diversion as a supporting piece of training to ease students fatigue and expand excitement and engagement. The quick development of the Internet and its applications, instructive organizations started utilizing Information and Communication Technologies (ICT) to support conventional education. There is significant proof of development in the E-learning business and in the quantity of E-learning organizations in a few areas, for example, Asia and Western Europe. The benefit of learning online lies in its adaptability that permits the understudies to learn at once and spot of their decision. Inside the setting of E-learning, the learner experience and view of

engineering is affected not just by their sex and their recognition of the innovations utilized, but also their learning styles focused around four measurements, specifically data handling, data observation (sensing/ intuitive), input mode and understanding [1]. This makes space for usage of manufactured smart projects to enhance E- Learning frameworks.

2. Objective

Register student information to suggest the Online Degree Course based upon marks and percentage. To predict and suggest degree courses based on students 10th and 12th marks is to provide academic guidance to students and provide decision-marking regarding higher education. The E-learning platform is made to be easy to use, so students can quickly enter their academic information and get recommendations without any hassle. By training the value as input data and in the backend, process the test case to predict the outcome. Thereby obtain the predictable outcome as the Degree course name which student can take up.

3. Proposed System

To predict the level of student knowledge after undergoing the online learning mode, the supervised back propagation learning technique reaches a high state. Another interesting ANN type, namely, a feed-forward neural network with a batch gradient descent, is preferred for classifying student learning styles [7]. This study [8] suggests versatile learning in demonstrating English as a Second Language (TESL) to an e-learning system (AL-TESL-e-learning system) that takes into account the student characteristics. This examination inspects the learning-results of different understudies, such as intuitive, active, and sequential learning-results. The ANN model achieved a 69.3% accuracy by applying a learning

algorithm to obtain the output layer of the students' learning behaviors. A supervised learning models, i.e., a feed-forward neural network, was developed, and a back propagation algorithm was implemented, which adopted a classification

technique to predict the level of student knowledge from different types of learning approaches.

Implementation:

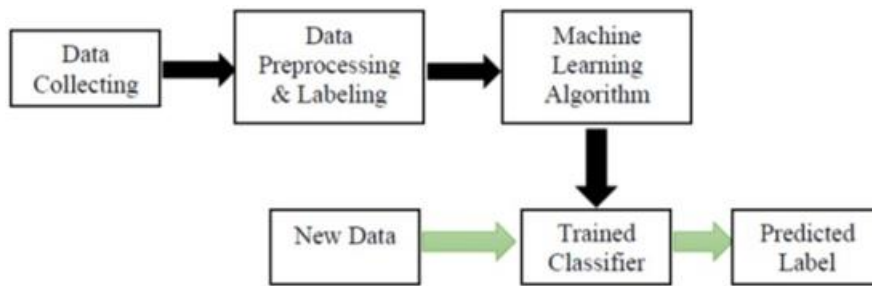


Figure 3.1: Block diagram for implementation of prediction model for students using machine learning

The first step is collecting the data from the data sources. In our case, the data has been collected using a survey given to the students and the students' grade book. The second step is pre-processing the data in order to get a normalized dataset and then labeling the data rows. In the third step, the result of the second step, the training and testing dataset, is fed to the Machine Learning algorithm. The Machine Learning Algorithm builds a model using the training data and tests the model using the test data. Finally, the Machine Learning Algorithm produces a trained model or a trained classifier that can take as an input a new data row and predicts its label.

3.1 Data Collecting

- 1) **Role:** The first step involves gathering all relevant data on students. This data can come from various sources such as academic records, demographic information, attendance records, and behavioral data.
- 2) **Types of Data:**
 - **Academic Performance:** Grades, scores on assignments, quizzes, and exams.
 - **Behavioral Data:** Attendance records, participation in extracurricular activities, engagement in class, etc.
 - **Demographic Data:** Age, gender, socioeconomic background, family support structure.
 - **Psychometric Data (optional):** Survey responses, personality assessments, etc.
- 3) **Importance:** High-quality data collection is crucial, as the performance and accuracy of the machine learning model heavily depend on the relevance and quality of the input data. Any missing or inaccurate data could lead to misleading predictions.

3.2. Data Preprocessing & Labelling

- 1) **Data Cleaning:**
 - **Handling Missing Data:** Many datasets have missing values due to incomplete records. Common techniques include filling missing values with mean/median values, using algorithms to estimate missing values, or dropping incomplete records.
 - **Removing Outliers:** Outliers or abnormal values can distort model predictions. Outliers are typically removed or adjusted based on statistical techniques.
 - **Normalizing and Scaling:** Different features might have

varying scales (e.g., scores out of 100 versus attendance percentages). Scaling ensures that all features contribute equally to the model.

2) Feature Engineering:

- **Creating New Features:** Sometimes, creating new, meaningful features (such as a combined engagement score based on attendance and class participation) can improve model performance.
- **Encoding Categorical Variables:** If data contains categories (e.g., "Passed" vs. "Failed"), it needs to be converted into numerical format for machine learning models. Encoding techniques, such as one-hot encoding or label encoding, are used to transform these values.

3) Labeling:

This involves assigning target labels or outcomes to each data point, representing what the model is intended to predict. In the case of student performance prediction, labels could include:

- **Course outcome**
- **Grade Levels (A, B, C, etc.)**
- **Risk Categories (e.g., high risk, moderate risk, low risk)**
- **Importance:** Preprocessing and labeling ensure that the data is ready for the machine learning algorithm. Proper labeling enables supervised learning models to identify patterns associated with the outcomes you want to predict.

3.3. Machine Learning Algorithm

- **Model Selection:** Based on the prediction objective, an appropriate machine learning algorithm is chosen. Common algorithms for student performance prediction include:
- **Classification Algorithms (for categorical predictions like "Pass" or "Fail"):** Decision Trees, Logistic Regression, Support Vector Machines (SVM), and K-Nearest Neighbors (KNN), Navie Bayes algorithm

4. System Design

Collect information about students, including their progress, and performance. Use machine learning to analyze this data and create user profiles. Set up recommendation systems to suggest relevant courses with marks input data. Design algorithms that adjust learning paths based on how well users

are doing and their specific needs. Based on the data from students, the system will suggest future courses for their degree.

5. Methodology

Machine learning models can help predict student outcomes, like identifying students who might be struggling or finding difficulties. The system creates a learning journey to each student's needs. Learning platforms adjust to the students, pace and progress, offering support where needed. Getting better recommendations for courses based on a student's

learning history and preferences. Enhances access to education for all students. Using the Naïve Bayes algorithm, can improve accuracy when recommending the right courses for students.

6. Flow Chart

The flowchart depicts a student login process for an educational platform, but it does not directly illustrate the implementation of a prediction model for students using machine learning. However, we can infer how such a model might be integrated into this system.

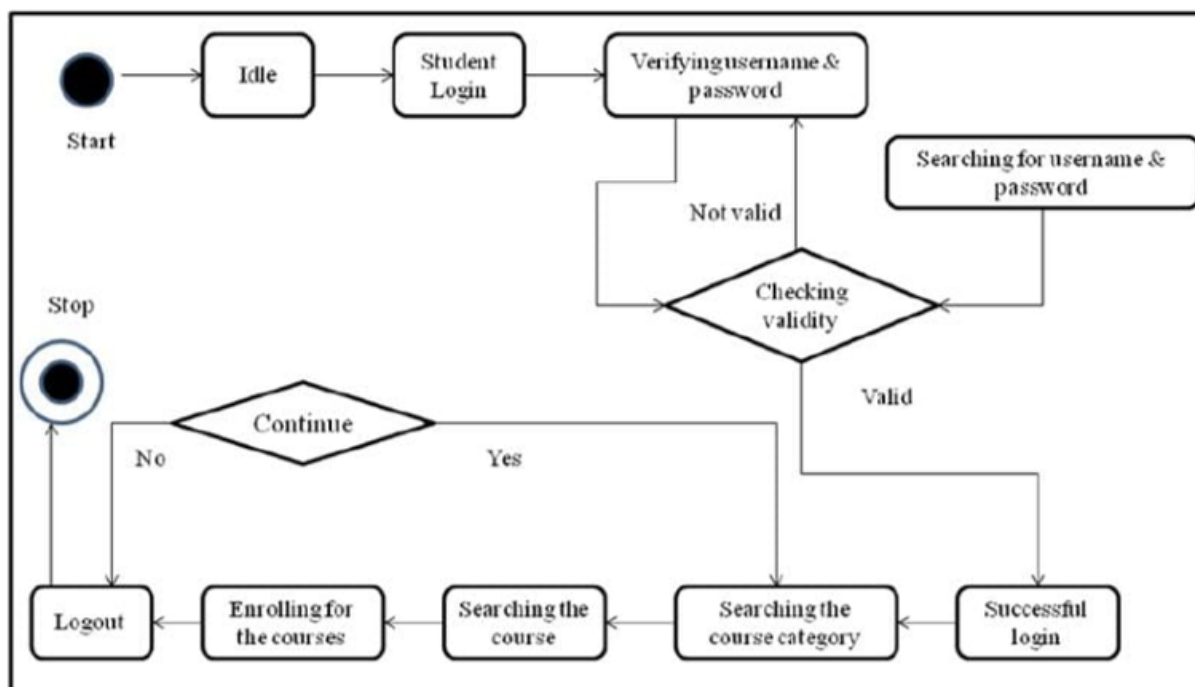


Figure 3.2: Flow chart that depicts student future prediction course

Here's a detailed flowchart explanation, including the potential integration of a prediction model:

- 1) Start: The process begins.
- 2) Idle: The system waits for user interaction.
- 3) Student Login: The student initiates the login process.
- 4) Verifying Username & Password: The system checks the entered credentials against the stored database.
- 5) Searching for Username & Password: The system searches for the entered credentials in the database.
- 6) Checking Validity: The system validates the credentials.
 - If valid: Proceeds to "Successful Login".
 - If not valid: Prompts the user to re-enter credentials.
- 7) Successful Login: The student successfully logs in.
- 8) Enrolling for the Courses: The student can now enroll in courses.
- 9) Searching the Course: The student can search for specific courses.
- 10) Searching the Course Category: The student can search for courses by category.
- 11) Continue: The student can continue interacting with the system.
- 12) Logout: The student logs out of the system.
- 13) Stop: The process ends.

7. System Specification

- Operating system: windows 10.
- Coding Language: Java/J2EE
- Tool: Eclipse Juno
- Data base: MYSQL
- Webserver: Apache Tomcat 7.0

8. System Implementation

- Set up the java EE Environment in the Eclipse Lab.
- Use java import Package to link the Backend code with the front end.
- Start the server
- After the server has started , open the application or the e learning platform.
- Input the training data values into the system.
- Utilize the naïve bayes algorithm to test the trained dataset or values
- Naive bayes algorithm is a simple and effective method used for classifying data.it predicts the category based on the features it has, using probabilities.
- The final result will indicate which course the student can take.

9. Results

In this project, we successfully implemented a machine learning-based prediction model for student performance. This model leverages historical student data, processes it, and applies machine learning algorithms to predict future performance outcomes. The development of such a model

holds significant value for educational institutions, as it enables proactive interventions, optimized resource allocation, and targeted support for at-risk students.

Flow way of result:

university login>student login>user login>student detail view>prediction result.

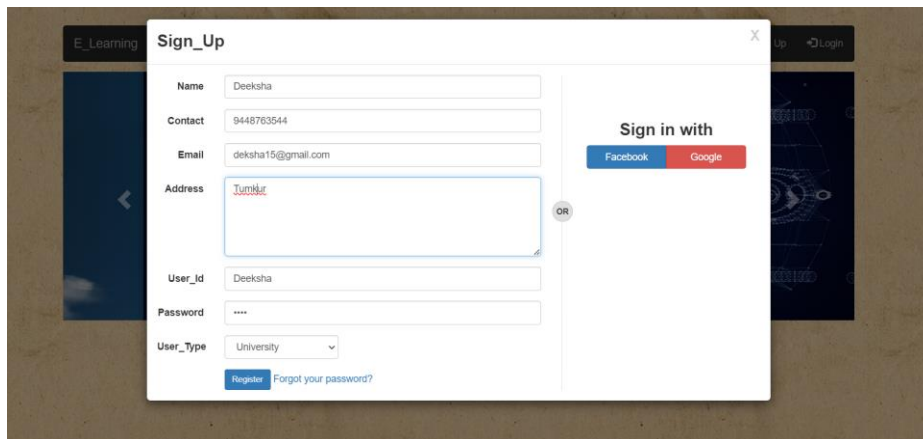


Figure 9.1: University login

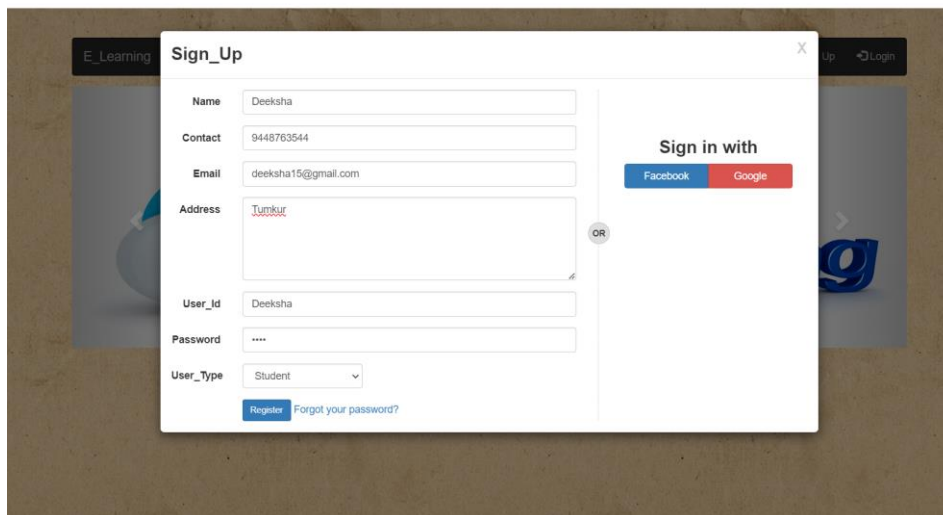


Figure 9.2: Student login

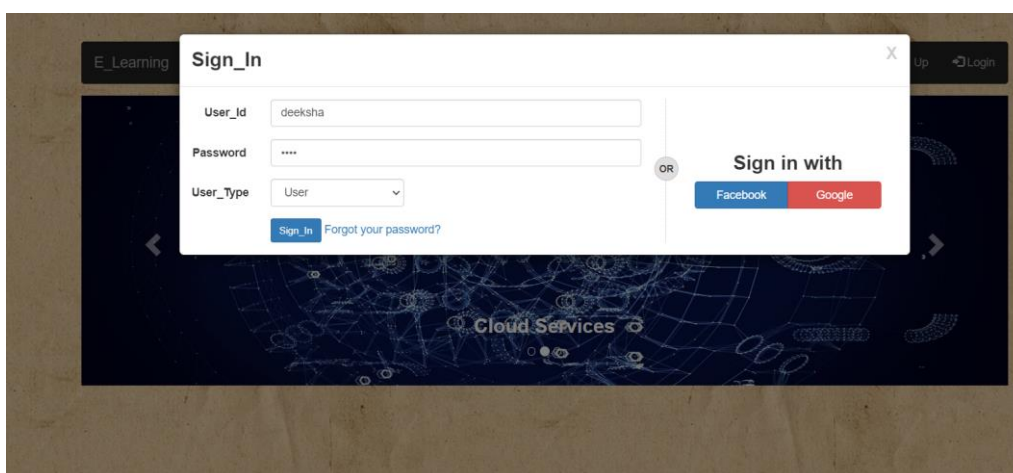


Figure 9.3: User login

Figure 9.4: Student detailed view

PName	Contact	A_Score	Grade	SCHOOL	SSLC_PER	PUC_MATHS	PUC_SCIENCE	Result
Deeksha	9448763544	1	HIGH	government	70	56	72	ENGINEERING

Figure 9.5: Final Predicted results

10. Advantages

- Machine learning generate more accurate decisions.
- Helps students to build their carrier
- Auto suggestion of course selection based on previous performance.
- It reduces cost associated with Traditional Methods.
- Identify skills gap and suggest relevant course suggestion.

11. Disadvantages

- If the data used to train the model is inaccurate or incomplete, the predictions can be unreliable.
- Creating and keeping a prediction model can be complicated.
- Machine learning models can accidentally learn unfair pattern from the training data, which can result in unfair predictions for some student groups.
- It might not work well with new data, which makes it less helpful in real- life situations.
- Gathering and using student data can cause privacy problems.

12. Conclusion

Student data is gathered to help with course selection and predicting which courses might suit them best. Machine learning in e- learning provides a more personalized, efficient and adaptable learning experience for students. E- learning platforms can predict student performance and outcomes

based on educational content or input data. This prediction method will help students choose the right courses based on their abilities and progress.

References

- [1] M. Xia, H. Wei, M. Xu, L. Yu Ho Lo, Y. Wang, R. Zhang, and H. Qu, "Visual analytics of student learning behaviours on K-12 mathematics E-learning platforms," 2019, arXiv:1909.04749. [Online]. Available: <http://arxiv.org/abs/1909.04749>
- [2] X. Wang, P. Wu, G. Liu, Q. Huang, X. Hu, and H. Xu, "Learning performance prediction via convolutional GRU and explainable neural networks in e-learning environments," *Computing*, vol. 101, no. 6, pp. 587–604, Jun. 2019
- [3] A. M. de Moraes, J. M. F. R. Araujo, and E. B. Costa, "Monitoring student performance using data clustering and predictive modelling," in *Proc. IEEE Frontiers Educ. Conf. (FIE)*, Oct. 2014, pp.
- [4] M. Injadat, A. Moubayed, A. B. Nassif, and A. Shami, "Multi-split optimized bagging ensemble model selection for multi-class educational data mining," *Int. J. Speech Technol.*, vol. 50, no. 12, pp. 4506–4528, Dec. 2020.