

STUDENT'S COURSE RECOMMENDATION SYSTEM USING MACHINE LEARNING

¹Rukesh. S

¹Department of Data Science, St. Joseph's College (Autonomous)

Tiruchirappalli, India

¹rukeshcs@gmail.com

ABSTRACT

Machine learning (ML) is an important research field of artificial intelligence that assists computers in modeling based on experiences and accurately predicting future events[1]. Machine learning enables computers to learn from data and make predictions and helps in revolutionizing various industries by automating tasks, improving efficiency, and uncovering insights from vast datasets. Its significance lies in its ability to enhance decision-making processes. Machine learning plays a crucial role in developing student course recommendation systems by analysing student data, their interests, and preferences to suggest course selections, ultimately enhancing academic success, personalized learning experiences, and student satisfaction.

Keywords: Machine learning, Classification algorithms, Non-Linear Models, Data Preprocessing, Model Evaluation

INTRODUCTION

Machine learning (ML) is used to teach machines how to handle the data more efficiently. Sometimes after viewing the data, we cannot interpret the extract information from the data. In that case, we apply machine learning.[3]. Machine learning is a subfield of artificial intelligence that gives computers the ability to learn without explicitly being programmed. A machine learning system builds prediction models, learns from previous data, and predicts the output of new data whenever it receives it. The amount of data helps to build a better model that accurately predicts the output, which in turn affects the accuracy of the predicted output.

Based on the methods and way of learning, machine learning is divided into mainly four types such as Supervised Machine Learning (labelled Data), Unsupervised Machine Learning (unlabelled Data), Semi-Supervised Machine Learning (labelled as well as unlabelled data), Reinforcement Learning (Trial and error), Self-supervised learning is a form of unsupervised learning where the training data is autonomously (or automatically) labelled.

Supervised machine learning performs two tasks based on the labelled data such as Prediction and Classification. Regression algorithms are used to predict continuous output variables, such as market trends, weather prediction, etc. In student Course Recommendation System, classification task is performed

Classification algorithms are used to solve the classification problems in which the output variable is categorical, the classification algorithms predict the categories present in the dataset.

This analysis is based on supervised learning methods, i.e., Naive Bayes, SVM, Logistic regression, Decision Tree Classifier, Random Forest, and K- Nearest Neighbor.[2]

A learning path is the implementation of a curriculum design. It consists of a set of learning activities that help users achieve particular learning goals. Personalizing these paths became a significant task due to differences in users' limitations, backgrounds, goals, etc.[5]

In this proposed paper, we suggest post graduate course for the students based on their past courses that they have studied to ensure their eligibility and consider their interests and pave the way for them that are best suited for them

3. Methodology

DATASET

Dataset can be created manually based on the successful student approach towards their career. Dataset includes columns such as Undergraduate course, Higher Secondary course, Allied course and their interests, Post Graduate course(which we should recommend).

DATA PREPROCESSING

After the creation of dataset, Data can be preprocessed such as removing null values, Normalizing Data etc., All the categorical data should be encoded using encoding methods as machine learning needs all its data to be in numerical form

CLASSIFICATION ALGORITHMS

For, student course recommendation system, **Non-linear Models** can be used which are explained as follows:

- **K-Nearest Neighbors (KNN):** KNN assigns class labels to new data points based on the majority class of their K nearest neighbors, memorizing the training dataset without constructing a model. Its simplicity and reliance on similarity make it intuitive but computationally intensive for large datasets.
- **Kernel Support Vector Machine (Kernel SVM):** Kernel SVM finds optimal hyperplanes to separate classes in higher-dimensional spaces using kernel functions, enabling it to handle nonlinear data effectively. It's powerful for complex classification tasks but may be computationally expensive.
- **Naïve Bayes:** Naïve Bayes is a probabilistic classifier based on Bayes' theorem with an assumption of feature independence. Despite its simplicity, it's effective for text classification and requires minimal training data, being computationally efficient.
- **Decision Tree Classification:** Decision trees create tree-like structures where each node represents a feature-based decision, suitable for interpretability and handling both numerical and categorical data. However, they can overfit with complex trees and noisy data.
- **Random Forest Classification:** Random forests combine multiple decision trees to improve accuracy and reduce overfitting through ensemble learning, making them robust for classification tasks across various domains while being less sensitive to hyperparameters.

Evaluating a Classification model:

Once our model is completed, it is necessary to evaluate its performance; either it is a Classification or Regression model. So for evaluating a Classification model, we have the following ways:

1. Log Loss or Cross-Entropy Loss:

It is used for evaluating the performance of a classifier, whose output is a probability value between the 0 and 1. For a good binary Classification model, the value of log loss should be near to 0. The lower log loss represents the higher accuracy of the model.

2. Confusion Matrix:

The confusion matrix provides us a matrix/table as output and describes the performance of the model. It is also known as the error matrix. The matrix consists of predictions result in a summarized form, which has a total number of correct predictions and incorrect predictions. The matrix looks like as below table:

	Actual Positive	Actual Negative
Predicted Positive	True Positive	False Positive
Predicted Negative	False Negative	True Negative

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{Total Population}}$$

3. AUC-ROC curve:

ROC curve stands for **Receiver Operating Characteristics Curve** and AUC stands for **Area Under the Curve**. It is a graph that shows the performance of the classification model at different thresholds. To visualize the performance of the multi-class classification model, we use the AUC-ROC Curve.

RESULT:

The model with more accuracy can be chosen as the model to recommend student course. The model with more accuracy will give more precise and correct output.

CONCLUSION:

Our Study showcases a study on how a multifaceted student recommendation system integrating enhancing accuracy and user satisfaction can be built. We emphasize the importance of considering diverse factors such as academic performance and their interests for personalized recommendations. Our work contributes valuable insights for educators and developers aiming to tailor learning experiences effectively. Ultimately, our system offers a promising approach to address the diverse needs of students in educational settings.

REFERENCES

1. Alican Dogan, Derya Birant, Machine learning and data mining in manufacturing, Expert Systems with Applications, Volume 166, 2021, 114060, ISSN 0957-4174.
2. TR, R. ., Lilhore, U. K., M, P. ., Simaiya, S. ., Kaur, . A. ., & Hamdi, M. . (2022). PREDICTIVE ANALYSIS OF HEART DISEASES WITH MACHINE LEARNING APPROACHES. Malaysian Journal of Computer Science, 132–148.
3. Batta Mahesh, Machine learning algorithms – A review, International Journal of science and research, 2319-704
4. Alian Dogan, Derya Brint b, Data Mining and manufacturing
5. Amir Hossein Nabizadeh, José Paulo Leal, Hamed N. Rafsanjani, Rajiv Ratn Shah, Learning path personalization and recommendation methods: A survey of the state-of-the-art, Expert Systems with Applications, Volume 159, 2020, 113596, ISSN 0957-4174,