ML Course Recommender

Vaibhav Jha
Bachelors of Engineering in Computer
Science
Chandigarh University
Unnao, India

Abstract — **This study presents a detailed analysis of different models in Machine Learning, aiming to find the best possible model for the Course Recommendation System. AI Educational System has completely changed the student-teacher bond in the past couple of years and with time it enhances the teaching methodology. AI is acting as a strong pillar between both student and teacher as well as student and his learning. With such an advancement in the growth of AI, one of the popular recommendation systems Course Recommendation Systems is becoming an essential part of today's learning, especially for students who want the perfect course according to the subject they want to learn. So this study shows the best machine learning model to classify the best course for recommending. For this four most popular machine learning models are been compared - Multinomial Naïve Bayes, Support Vector Machine, Random forest, and Logistic Regression. On comparing, the model with the best accuracy comes to the best-fit model for our recommendation system.**

Keywords— Recommendation, Multinomial Naïve Bayes, Support Vector Machine, Random forest, Logistic Regression

I. INTRODUCTION

Artificial intelligence (AI) is a multidisciplinary field whose aim is to automate tasks that currently need human intelligence. [1] In past couple of years, Artificial Learning has reached great heights and has immensely changed the way we live our lives. Unprecedented growth in artificial intelligence with its subsets like machine learning, deep learning, artificial general intelligence, and big data results in the fact that today there is no sector that is untouched by them. From an automatic bulb that glows with human sensors to the entire robotics field, from large-scale industries to small-scale businesses, from agriculture to healthcare, from Fintech to cyber security, and in many more sectors artificial intelligence has transformed the way everything works. It improved our lives, growth, and learning by offering its impeccable services. One of the crucial and important innovative approaches happened is in the field of Education and this resulted in the creation of the AI Educational Recommendation System.

Machine learning is one of the one of the core section of AI and is growing at the full pace in last couple of years. It is one of the main branch of artificial intelligence which focuses on development of new learning algorithms to enable the machine to learn automatically though experience. The advancement of this technology is seen in many walk of lives including healthcare, educational field, marketing, manufacturing, to name a few. [2]

A. AI in Education

In education, AI has played a significant role for teachers and students. They can utilize artificial intelligence and its subsets in two main ways –

1. By creating a dynamic learning environment which facilitate practical learning experience

2. By developing a computational tutoring system that enhances tutorial methodology by making certain concepts understood to the students based on their ability to learn new things in the easiest yet best conceivable way. [3]

Thus, both teachers and students directly or indirectly depends on artificial intelligence. After seeing the importance of Artificial Intelligence in the field of Education System, more research happened in the past some years to make the bond of student and teacher more flexible yet extraordinarily strong on the bridge of AI. According to a report, in 2021 AI market size was valued at USD 1.82 billion in the field of education which is expected to expand with 36% growth rate from 2022 to 2030. [4]

One of those AI Educational Recommendation system which become crucial for both students as well as teachers these days is AI Course Recommendation System. University's student is highly benefited by this recommendation system.

B. Concept of Course Recommendation System

At the university level and during the journey at corporate sector, a student learn many courses whether that course belongs to his domain or other, his core subjects or interpersonal skills. At such level student want such recommendation system which help them by reducing their time, effort and provide them best available course based upon the necessity of student. Apart from

that, at university the curriculum is divided into two type of courses - mandatory and elective courses. Mandatory subjects are compulsory subjects that students must take while elective subjects are optional subjects or course groups. [5] Though mandatory subjects are predefined subjects based on college's set curriculum, optional subjects are where this recommendation system comes. Based upon the skill a student has, from where he want do course, his area of interest, the credit his optional subject contain and certain other important factor this recommendation system suggest you the courses which you do. This study propose the best model among some major Machine Leaning techniques to make such a recommendation system.

C.   *Overview of models*

A crucial aspect of this research is the comparison of different models of machine learning and check the best among them based upon their accuracy to recommend the courses. The comparison of effectiveness is made among the models like Multinomial Naive Bayes, Random Forest, Support Vector Machine and Logistic Regression and check best fit for AI Course Recommendation System.

TABLE I: MODEL OVERVIEW

Model	Overview
Multinomial Naïve Bayes	Probability based text classification
Random Forest	Ensemble of decision trees.
Support Vector Machine	Maximize margin classifier
Logistic Regression	Binary classification using probabilities

II. LITERATURE REVIEW

A.   *Related Projects*

Several related projects have contributed till now showing different types of recommendation system in the field if education especially with AI. However specifically talking about AI Course Recommendation System there are few but their research contribution is invaluable in this field. Some of the key projects are –

a) Methods for building course recommendation systems

This paper presents several models which can be used to build course recommendation system. These models are compared and analyzed to check their performance using real educational dataset and thereby selecting the best model to perform Course Recommendation System. [6]

b) Recommender systems in e-learning environments

It gives an outline of the requirements and challenges in making a recommendation system in e-learning environments. This paper gives an idea about the limitations in the current recommendation system and possible extensions. [7]

Apart from above mentioned studies some other important and innovative researches have done in the field of AI Recommendation system which results in enhancing the level of education.

B.   *Evolution of AI and AI Recommendation in Education*

During the period of 1940-50 the theoretical roots of computer science and learning of machines was laid by a mathematical and computer scientist name Allen Turing. And since then the growth of AI is climbing a never ending peak. During the mid of 20th century, the growth of AI was seen in the field of education. Computers at that time were used by many educational institutions and simple exercises content were delivered with it. The limitations of this was the lack of interactivity and grain size and thus used for programmed responses.

By late ninety's AI started gripping educational field. At that time some tutor models were being made. GUIDON and MYCIN were made to teach to the medical students about different diagnosis diseases and their treatment [8]. After that the evolution of Neural Processing Unit (NLP) and Chabot during the starting of 21th century has changed the way of education. The rise of NLP, Machine Learning and Big data during 2010 has entirely changed the teaching methodology. New software are made which provide free learning to students of any subjects. Further some software's were made to clear the doubts of students.

The growth was seen during COVID 19 when entire world was restricted at their homes. According to the survey conducted during March and April of 2022 by International Association of Universities in 424 Universities and other Higher Educational Institutions, due to COVID 19 pandemic 67% said that classroom teaching has completely been replaced by Online teaching while only 7% said that teaching has been cancelled. [9] This shows how deeply AI has rooted itself in the field of education.

Today one of the popular application of Artificial Intelligence, digital assistance is creating a huge impact in teaching methodology. They teach both teachers as well as students in number of ways, including a vast range of learning materials on each specific topics or subject user wants. [10]

This growth of AI after the massive growth and research of AI subsets like machine learning, deep learning, neural language processing, big data, etc Course Recommendation System is highly been used by many software and applications. Many applications like Coursera, YouTube, Edx, Google's AI Education, including many other websites using Course Recommendation. They uses different types of algorithms and methodologies to predict the best possible course for the students.

C.   Types of Recommender Systems:

There are several types of recommendation system like Collaborative Filtering, Content-based Filtering, Hybrid Recommendation System, Contextual Recommendation, Popularity-based Recommendation, Item-based Recommendation, etc. However this research paper specially work with Contextual Recommendation System.

Contextual Recommendation is a types of recommendation system that select and recommend items or content t users based on their current situation or context. In course recommendation system, we have worked with a database. So our database suggest and recommend the course to the students or the users based on the following columns as the recommendation parameters. These are –

- *Course Name* – Name of the course which is   present in the database
- *University* – Different universities which are offering different courses
- *Difficulty* – How much student has already done whether student wants a beginner level or advanced level course in a particular course
- *Course Rating* – If student specifically want course above a certain rating
- *Course URL* – Link of the course
- *Course Description* – Tells what is going to be learn
- *Skills* – Judging parameter. Using skills a student want he gets the course.

Now in this recommendation using different parameters we will provide the students best possible course using contextual learning. The recommendation is based on the skills student wants and other parameter and with the help of contextual learning we provide them the course which is meaningful and engaging for them.

III. METHODOLOGY

A.   Model Selection and Justification

This research paper works on 4 machine learning models - Multinomial Naive Bayes, Random Forest, Support Vector Machine, and Logistic Regression. Here these models are compared to find the best and the most accurate models to recommend course to the user.

- *Multinomial Naïve Bayes*

Multinomial Naïve Bayes or MultinomialNB is a popular supervised machine learning algorithm. It is one of the naïve bayes classifier variant that is used to compare specialized event models. Because MNB is easy to implement and is relevantly fast, it is often used as a baseline in text classification and document classification task. [11] It is a probabilistic algorithm which estimates the probability of a particular class based on the frequencies of words or features.

Suppose a text document t_i which belong to a particular class C having N as the size of vocabulary. The multinomial naive Bayes of the document assigns to the class label with highest probability $Pr(c|t_i)$, as the predicted class for that document. [12] Now to calculate probability we Bayes theorem is used which can be calculated using the formula -

$$P(C|t_i) = \frac{P(C)\,P(t_i|C)}{P(t_i)}$$

It works very well with large datasets and with high dimensional data set and because of all these important functionalities, Multinomial Naïve Bayes is one of the model for comparison in this paper.

- *Random Forest*

Random forest, an ensemble machine learning method, is a combination of large number of individual decision trees. Multiple decision trees results in improving accuracy and reducing overfitting. It is used for both classification and regression where in classification, random forest combines the predictions of all decision trees by majority voting (the class predicted by the most trees) whereas in regression tasks, it combines predictions by averaging the outputs of all trees. It is highly effective for the dataset having mix of feature types.

It uses base classifiers of tree shape $\{h(X,\Theta_n),\ N=1,2,3,\dots L\}$, where X is input data and $\{\Theta_n\}$ is independent identically distributed random vector. [13]

- *Support Vector Machine*

Support Vector Machine or SVM in short form – is a popular supervised machine learning algorithm which is used for both classification as well as regression however in machine learning it is primary used as a classification model. It built on the basis

of a limited number of samples in the information contained in the existing training text to get the best classification results. [14]

The main goal of support vector machine is to find the best fit line know as maximum margin hyperplane that can differentiate n-dimensional hyperplane into classes so that new input data can be put into that correct category.
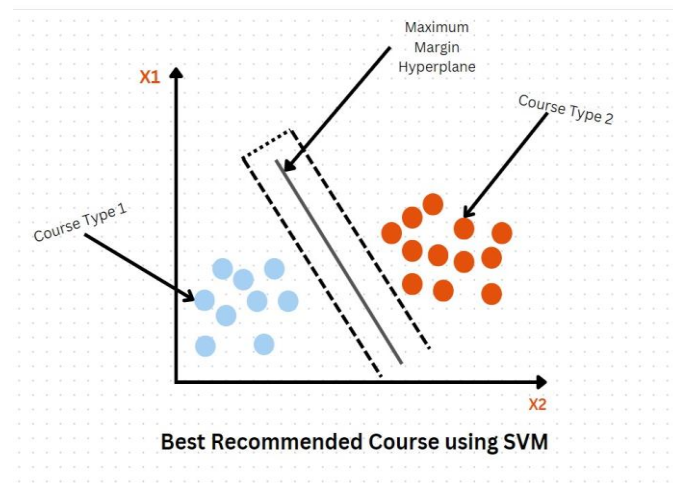
i) SVM Model

- *Logistic Regression*

One of the most popular machine learning algorithm, Logistic Regression is used for classification. It is used for binary classification problem, where the binary outcome is 0 or 1 (or Yes or No) for a given set of independent variables. It predicts that the given input belongs to one of all the given classes. The dependent variable outcome is discrete. In order to convert the linear combination of input features into a value between 0 and 1, logistic regression uses sigmoid function which is–

$$p(x) = \frac{1}{1+e^{-(B_0 + B_1 x)}}$$

B.　Libraries Used

For implementing and analyzing all the models and choose which model is the best fit for Course Recommendation, certain Python libraries are used. These libraries contain various predefined functions which is used for both analyze and visualize the result obtained.

The libraries which are used to implement Course Recommendation system are shown in the following table

TABLE II:　　LIBRARIES USED

Libraries	Description
Pandas	Use for data analysis and cleaning
Matplotlib	Use for data visualization. Provide object oriented API
Seaborn	Use for making statistical graphics in python
Scikit-learn	In implementation of classification, regression and clustering algorithm

In this Scikit-learn is a library from which all different models can be imported. So all the four models - Multinomial Naïve Bayes, Random forest, Support Vector Machine and Logistic Regression are been imported from this library. Further PANDAS, Python-data Analysis Library is used for data clearing and data analysis.

C.　Model Architecture

The workflow of Course recommendation is shown as –
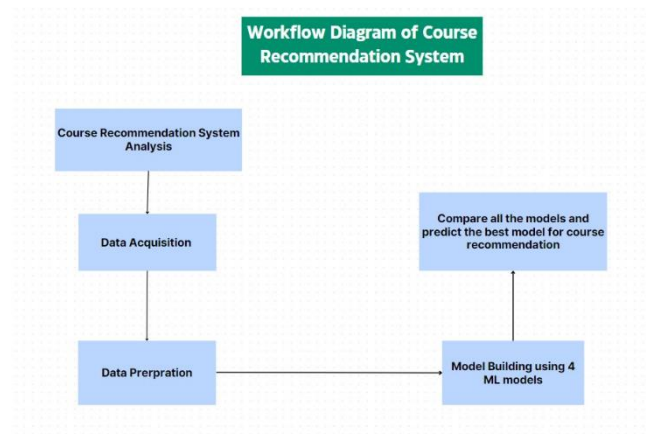
ii) Workflow Diagram

After analysis of Course Recommendation system, we gather and scrape the data from multiple source like from kaggle. Then data cleaning and transformation is performed to avoid empty row values or prevent from outliers. And after preparing the data, all the four models are been compared by their accuracies. And using seaborn the result is been visualized.

D.   Model Evaluation Metrics – Confusion Matrix

To check the best model out of the four machine learning model for Course Recommendation System, a popular evaluation metrics are being used, Confusion Matrix. This is used to find the best possible model.

A Confusion matrix is table layout showing the predicted outcome of classification model. It gives detail breakdown of correctly and wrongly predicted outcomes and help to visualize the result. It is mostly used in binary classification problem having two classes positive and negative. [15]

Confusion matrix table is divided into four blocks showing its four components –
a)   *False Negative* – The number of instances our model wrongly predicts positive value as negative values.
b)   *True Positive* – The number of instances our actual positive values are equal to predicted positive values.
c)   *False Positive* – The number of instances our model wrongly predict negative values as positive values.
d)   *True Negative* – The number of instances actual negative is predicted as predicted negative
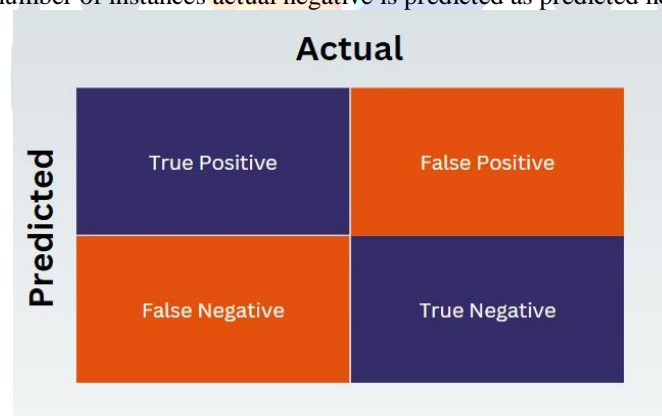
iii) Confusion Matrix table

There are four main metrics in Confusion metrics which are used to measure the performance. These are – Accuracy, Precision, Recall, F1-score. Among which this paper focuses on one of those metrics which is Accuracy. Accuracy is use to find the portion   of   correctly   classified   values

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Where –
TP – True Positive   FP – False Positive
TN – True Negative   FN – False Negative

So, accuracy uses TP, TN, FP, and FN values from the confusion matrix to measure the overall correctness of a classification model. In Course Recommendation System, the model having highest accuracy will be the best suited model.

IV. IMPLEMENTATION AND OUTPUT

A. Implementation

For implementation Python language is been used. Python is a simple high level language, which also supports Object Oriented features. This language has the main use is in the implementation of Artificial Intelligence models in this paper. It is being implemented on Jupyter Notebook which is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations, and narrative text.

B. Result and Performance Evaluationt

TABLE III: ACCURACY OF MODELS

Model	Accuracies (in %)
Multinomial Naïve Bayes	48.51
Random Forest	52.77
Support Vector Machine	55.74
Logistic Regression	54.18
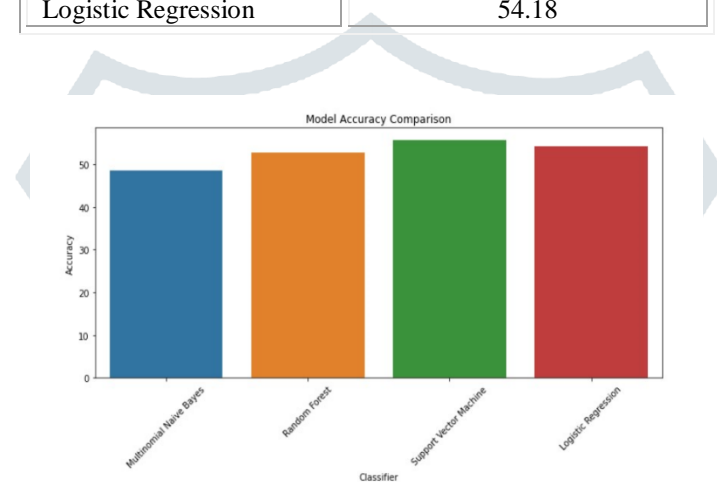
V. CONCLUSION AND FUTURE SCOPE

A. Conclusion

The conclusion of this paper is that Support Vector Machine is the best model among all the four chosen models having the accuracy of 55.74%. However Logistic Regression also shown 54.18% accuracy which is relevantly same as Support Vector Machine However, Multinomial Naïve Bayes shown the lowest accuracy of 48.51%. Therefore for Course Recommendation System Support Vector Machine is the best model.

B. Future Recommendation

This research paper output is certain limitation which opens the door for future work on this area. Some of them are:

a) Adoption:

Still a lot of scope in AI Education industry is left behind for the Course Recommendation System. Many student still not know about any such recommendation.

b) Limitation of Dataset:

Due to small and limited dataset out model can predict only the courser course recommendation system. However in future we will work upon expanding of dataset by including maximum possible courses in dataset of different websites and after such a huge trained model, the model will predict the best possible course out of all the courses it has.

c) Research Expansion:

Further research could explore for increasing the accuracy for this recommendation by using neural networks and deep learning methodology. As Machine learning not provide satisfactory accuracy in future other AI area should be explored.

d) Lack of Software:

Today there are no or very few software which only provide best recommendation. Some applications are using but it work only for a particular application not working for all the courses available for learning and then find the best among that.

VI. Dataset

The dataset used to compare and analyze different models is present in Kaggle with the name of Coursera Recommendation System dataset.

References

[1]. Kaur, P. (2023). Artificial Intelligence. *International Journal for Research in Applied Science & Engineering Technology (IJRASET)*, 11(X), https://doi.org/10.22214/ijraset.2023.56025

[2]. Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*. https://www.science.org/doi/abs/10.1126/science.aaa8415

[3]. Chounta, Irene-Angelica, et al. "Exploring Teachers' Perceptions of Artificial Intelligence as a Tool to Support their Practice in Estonian K-12 Education." *International Journal of Artificial Intelligence in Education*, 02 June 2021 https://link.springer.com/article/10.1007/s40593-021-00243-5

[4]. AI In Education Market Size, Share & Trends Analysis Report By Component (Solutions, Services), By Deployment, By Technology, By Application, By End-use, By Region, And Segment Forecasts, 2022 – 2030 https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-education-market-report

[5]. Huynh-Ly Thanh-Nhan, Huu-Hoa Nguyen, and Nguyen Thai-Nghe. "Methods for building course recommendation systems." *2016 Eighth International Conference on Knowledge and Systems Engineering (KSE)*, October2016 https://www.researchgate.net/publication/311313309_Methods_for_building_course_recommendation_systems

[6]. Huynh-Ly Thanh-Nhan, Huu-Hoa Nguyen, and Nguyen Thai-Nghe. "Methods for building course recommendation systems." *2016 Eighth International Conference on Knowledge and Systems Engineering (KSE)*, October2016 https://www.researchgate.net/publication/311313309_Methods_for_building_course_recommendation_systems

[7]. Klašnja-Milićević, Aleksandra, Mirjana Ivanovic, and Alexandros Nanopoulos. "Recommender systems in e-learning environments: a survey of the state-of-the-art and possible extensions." DOI 10.1007/s10462-015-9440-z https://sci-hub.se/10.1007/s10462-015-9440-z

[8]. Woolf, B. P. (1991). *AI in Education* (COINS Technical Report 91-37) https://web.cs.umass.edu/publication/docs/1991/UM-CS-1991-037.pdf

[9]. *Sustainability*, vol. 13, no. 23, December 2021 https://www.mdpi.com/2071-1050/13/23/13501

[10]. Limna, Pongsakorn, et al. "A Review of Artificial Intelligence (AI) in Education during the Digital Era." *Advance Knowledge for Executives*, 3, 2022 URL- https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4160798

[11]. Xu, Shuo, Yan Li, and Zheng Wang. "Bayesian Multinomial Naïve Bayes Classifier to Text Classification." https://sci-hub.se/10.1007/978-981-10-5041-1_57

[12]. Kibriya, Ashraf M., et al. "Multinomial Naive Bayes for Text Categorization Revisited." https://sci-hub.se/10.1007/978-3-540-30549-1_43

[13]. Ali, Jehad, Rehanullah Khan, and Nasir Ahmad. "Random Forests and Decision Trees." *Journal Name*, vol. Volume, no. Issue, September 2012 https://www.researchgate.net/publication/259235118_Random_Forests_and_Decision_Trees

[14]. Zhang, Yongli. "Support Vector Machine Classification Algorithm and Its Application." *Conference Name*, https://link.springer.com/chapter/10.1007/978-3-642-34041-3_27

[15]. Karimi, Zohreh. "Confusion Matrix." *Journal Name*, vol. Volume, no. Issue, October 2021 https://www.researchgate.net/publication/355096788_Confusion_Matrix