ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 9, Issue 3, April 2024

Machine Learning-Based Learning Effect Prediction For Improving Student Information Literacy

[1] B. Sujai Mathew, [2] S. Sugashini
[1] Student, Department Of Cse, Kingston Engineering College, Vellore - 632 059, Tamil Nadu, India
[2] Assistant Professor, Department Of Cse, Kingston Engineering College, Vellore - 632 059, Tamil Nadu, India

Abstract: ***This project investigates the relationship between students' preadmission academic profile and final academic performance. Predicting Students performance beforehand can be very beneficial for educational institutions to improve their teaching quality. Further, the importance of several different attributes, or "features" is considered to determine which of these correlate with student performance. This project proposes to predict students' performance by evaluating their academic details. Data preprocessing was done to remove the results of rusticated and expelled students. Results obtained by comparing SVM with other ML techniques such as KNN, Decision trees, and linear Regression show that SVM outperforms other ML algorithms. The parameters of the SVM algorithm(kernel) were also tuned to improve its accuracy.***

I. INTRODUCTION

Every educational institution handles and deals with large amounts of student data which can be beneficial for several reasons. One of the important applications of such data is predicting student performance. Such a prediction can be useful not only for the students but also for teachers/mentors. Mentors can provide special assistance to students who are on the verge of failing.
Any prominent school or college can use this application used to predict the pointer ranges or percentage ranges for future semester exams.One of the challenges in organizing teaching interventions is that any change is likely to affect students for whom the prevalent situation is more suitable. For example, if a student is already at a stage where she could work on more challenging projects on their own, mandatory excessively structured learning activities that everyone needs to follow may even be counterproductive for her.
The aim of this project is the selection of features that show a strong relationship with a target attribute that is to be predicted from a high-dimensional dataset. We have evaluated and compared several algorithms such as decision tree, random forest, Support Vector Machine, naive Bayes and neural networks by applying them to the dataset The rest of the project provides an explanation of the nature of neural networks along with the results of our evaluation.

II. SOFTWARE ANALYSIS

PyCharm is the most popular IDE used for Python scripting language. This chapter will give you an introduction to PyCharm and explain its features. PyCharm offers some of the best features to its users and developers in the following aspects:

- Code completion and inspection
- Advanced debugging
- Support for web programming and frameworks such as Django and Flask

ANACONDA (PYTHON DISTRIBUTION)

Anaconda is a free and open-source distribution of the Python and R programming languages for scientific computing that aims to simplify package management and deployment. Package versions are managed by the package management system conda.

ANACONDA NAVIGATOR

Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda distribution that allows users to launch applications and manage conda packages, environments and channels without using command-line commands. Navigator can search for packages on Anaconda Cloud or in a local Anaconda Repository, install them in an environment, run the packages and update them. It is available for Windows, macOS and Linux.

The following applications are available by default in Navigator:

- JupyterLab
- Jupyter Notebook
- QtConsole
- Spyder
- Glue
- Orange
- RStudio
- Visual Studio Code
- Conda

JUPYTER NOTEBOOK

As a server-client application, the Jupyter Notebook App allows you to edit and run your notebooks via a web browser. The application can be executed on a PC without Internet access, or it can be installed on a remote server, where you can access it through the Internet.

This explains how to install, run, and use Jupyter Notebooks for data science, including tips, best practices, and examples. As a web application in which you can create and share documents that contain live code, equations, visualizations as well as text, the Jupyter Notebook is one of the ideal tools to help you to gain the data science skills you need.

This will cover the following topics:

- A basic overview of the Jupyter Notebook App and its components,
- The History of Jupyter Project to show how it's connected to IPython,
- An overview of the three most popular ways to run your notebooks: with the help of a Python distribution, with pip or in a Docker container,
- The best practices and tips that will help you to make your notebook an added value to any data science project

III. EXISTING SYSTEM

Methods for automatically identifying students in need of assistance were based on somewhat static factors such as students educational background and results from various questionnaires, while more recently, constantly accumulating data such as progress with course assignments and behavior in lectures has gained attention.

When combining source code snapshot data that is recorded from students' programming process with machine learning methods, we can detect high- and low-performing students with high accuracy already after the very first week of an introductory programming course. A comparison of our results to the prominent methods for predicting students' performance using source code snapshot data is also provided.

The accessed course materials, 40.28% of the students would post messages on the discussion board, and, there was more than 70% probability they would post again the same day. However, this assumption is hard to follow in our real lives. A useful method to achieve independence is to apply a feature selection strategy.

PROPOSED SYSTEM

This project investigates the relationship between students' preadmission academic profile and final academic performance. Predicting Students performance can be very beneficial for educational institutions to improve their teaching quality. Further, the importance of several different attributes, or "features" is considered to determine which of these correlates with student performance.

This project proposes to predict students performance by considering their academic details. Data preprocessing was done to remove the results of rusticated and expelled student. Results obtained by comparing SVM with other ML techniques such as KNN, Decision trees, linear Regression shows that SVM outperforms other ML algorithms.
The immediate feedback from self-assessment allows students to examine their understandings promptly; the sustainable learning purpose is obvious. Teachers can also gather and analyze learners' assessment records to diagnose their learning problems. In addition, through regular online self-assessments, students have the opportunities to enhance their self-efficacy beliefs and further, to improve their learning ultimately.

IV. MODULES

Importing the packages
For this project, our primary packages are going to be Pandas to work with data, NumPy to work with arrays, scikit-learn for data split, building and evaluating the classification models.
Encode the categorical variables
To check the categorical variables in the data, you can use the train data types () function. This will give you a list of the data types against each variable. Label Encoding refers to converting the labels into numeric form to convert it into the machine-readable form. Machine learning algorithms can then decide in a better way how those labels must be operated. It is an important pre-processing step for the structured dataset in supervised learning.
Identifying features to build the ML pipeline
As discussed initially, the most important part of designing a machine learning pipeline is defining its structure, and we are almost there. We are now familiar with the data, we have performed the required preprocessing steps and built a machine-learning model on the data. At this stage, we must list down the final set of features and necessary preprocessing steps (for each of them) to be used in the machine learning pipeline
Splitting of data
In this process, we are going to define the independent (X) and the dependent variables (Y). Using the defined variables, we will split the data into a training set and testing set which is further used for modeling and evaluating. We can split the data easily using the 'train_test_split' algorithm in python.
Performance Prediction using SVM
SVM is a supervised machine learning algorithm which works based on the concept of decision planes that defines decision boundaries. A decision boundary separates the objects of one class from the object of another class. Support vectors are the data points which are nearest to the hyper-plane. Kernel function is used to separate non-linear data by transforming input to a higher dimensional space. Gaussian radial basis function kernel is used in our proposed method.

$$K(X_i, X_j) = e^{-\|X_i, X_j\|^2 / 2\varphi^2}$$

ARCHITECTURE DIAGRAM

ISSN (Online): 2456-0448
International Journal Of Innovative Research In Management, Engineering And Technology
Vol. 9, Issue 3, April 2024

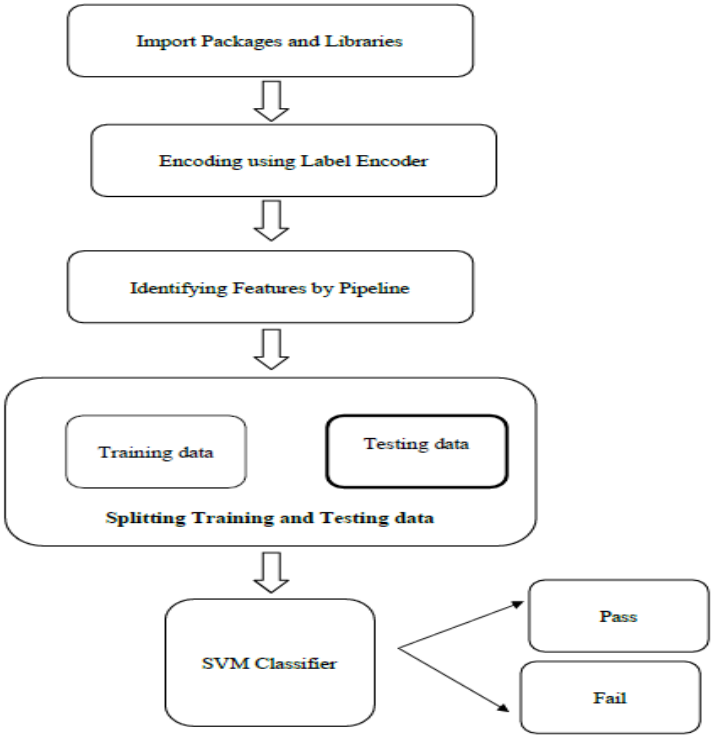

Fig.1. Students Information Literacy

V. RESULT

Information literacy is a crucial skill for students in today's digital age. With the vast amount of information available online, students need to be able to critically evaluate and make sense of the information they consume. Machine learning algorithms can play a significant role in predicting the effectiveness of different learning strategies for improving student information literacy.

By analyzing data on student behavior, engagement, and performance, machine learning algorithms can provide insights into which learning activities and resources are most effective in enhancing information literacy skills. These algorithms can predict the impact of different interventions, such as online tutorials, workshops, or interactive exercises, on student information literacy outcomes.

By leveraging machine learning-based learning effect prediction, educators can tailor their instructional approaches to better meet the needs of individual students. This personalized learning approach can lead to improved student engagement, motivation, and ultimately, better information literacy skills. Additionally, by continuously refining and updating the predictive models based on student feedback and outcomes, educators can ensure that their interventions remain effective and relevant in an ever-evolving educational landscape

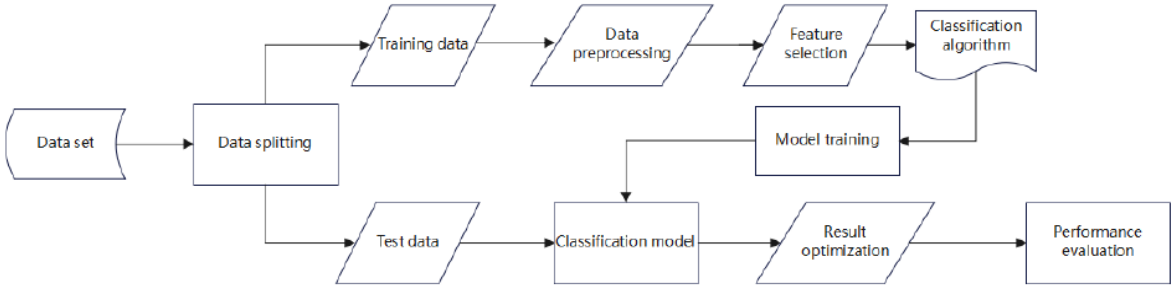

Fig.2 Technical Road map

VI. CONCLUSION

Predicting Students performance can be very beneficial for educational institutions to improve their teaching quality. Further, the importance of several different attributes, or "features" is considered, to determine which of these are correlated with student performance. This project proposes to predict students performance by considering their academic details. Data preprocessing was done to remove the results of rusticated and expelled student. Results obtained by comparing SVM with other ML techniques such as KNN, Decision trees, and linear Regression shows that SVM outperforms other ML algorithms. We explore methods for early identification of students to guide naturally accumulating programming process data. Such information can be useful for instructors and course designers and can be used to create targeted interventions and to adjust materials accordingly. For example, the students who are performing well in the course may benefit from additional, more challenging tasks. In contrast, the students who are performing poorly are likely to benefit from rehearsal tasks and other activities typically used to help at-risk students. conducting interviews and including some extra attention to the student's surroundings regarding their family and situations which matters to them most that hopefully will shed further light on students' working practices and those who were misclassified. We are also performing targeted interventions within the studied context.

REFERENCE

1. D. Yan and G. Li, ''A heterogeneity study on the effect of digital education technology on the sustainability of cognitive ability for middle school students,'' *Sustainability*, vol. 15, no. 3, pp. 2784–2786, Feb. 2023,

2. J. Li, ''Machine learning-based evaluation of information literacy enhancement among college teachers,'' *Int. J. Emerg. Technol. Learn.*, vol. 17, no. 22, pp. 116–131, Nov. 2022,

3. C. Y. Ko and F. Y. Leu, "Analyzing attributes of successful learners by using machine learning in an undergraduate computer course," in Proc. 32nd IEEE Int. Conf. Adv. Inf. Netw. Appl. (AINA-2018), Krakow, Poland, 2018, pp. 801–806.

4. S. Kotsiantis and D. Kanellopoulos, "Association rules mining: A recent overview," Int. Trans. Comput. Sci. Eng., vol. 32, no. 1, pp. 71–82, 2006.

5. C. Pei, ''The construction of a prediction model for the teaching effect of two courses education in colleges and universities based on machine learning algorithms,'' *Wireless Commun. Mobile Comput.*, vol. 2022, Aug. 2022

6. A. Ezen-Can, K. E. Boyer, S. Kellogg, and S. Booth, "Unsupervised modeling for understanding MOOC discussion forums: A learning analytics approach," in Proc. 5th Int. Conf. Learn. Anal. Knowl., 2015, pp. 146–150.

7. J.-L. Hung, M. C. Wang, S. Wang, M. Abdelrasoul, Y. Li, and W. He, "Identifying at-risk students for early interventions—A time-series clustering approach," IEEE Trans. Emerg. Topics Comput., vol. 5, no. 1, pp. 45–55, Jan.–Mar. 2017.

8. C. Romero, M.-I. López, J.-M. Luna, S. Ventura, "Predicting students' final performance from participation in on-line discussion forums," Comput. Educ. vol. 68, pp. 458–472, Oct. 2013.

9. R. Asif, A. Merceron, S. A. Ali, and N. G. Haider, "Analyzing undergraduate students' performance using educational data mining," Comput. Educ., vol. 113, pp. 177–194, Oct. 2017.

1. J. Li, ''Machine learning-based evaluation of information literacy enhancement among college teachers,'' *Int. J. Emerg. Technol. Learn.*, vol. 17, no. 22, pp. 116–131, Nov. 2022,