

A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs

Ashtami Prasad

Department of Computer Application
Amal Jyothi College of Engineering Kanjirappally, India
ashtamiprasad2022a@mca.ajce.in

Shelly Shiju George

Department of Computer Application
Amal Jyothi College of Engineering Kanjirappally, India
shellyshijugeorge@amaljyothi.ac.in

Abstract— Predicting students' future performance based on their current academic records is important for ensuring essential pedagogical interventions and assuring students' prompt and efficient graduation. Although there is a considerable literature on using data-driven methodologies to anticipate student performance whether solving problems or studying for classes, forecasting student performance is still a challenge. when finishing degrees is far less researched. Existing techniques have mostly used elements linked to academic performance and family financial assets, whereas features related to family expenditures and personal information of students have been largely overlooked. The purpose of this study is to acquire data from students in various classrooms in order to analyze the aforementioned feature sets.

Keywords— Academic records, predicting.

I. INTRODUCTION

Analyzing a vast amount of data to provide usable summaries can be a difficult undertaking for humans. Grading students' academic performance is a more difficult and time-consuming task that will help instructors keep track of students' progress.. For all educational institutes, improving the quality of education in order to improve student performance is more vital. Predicting student outcomes in advance is advantageous; knowing this ahead of time will affect student performance. Predicting student success within a degree program, on the other hand, is quite different and presents different problems.

Students can have vastly diverse backgrounds and areas of interest, resulting in varied course selections and sequences. Students from other places, on the other hand, might study the same course. One of the most difficult components of designing an effective predictor is finding out how to deal with heterogeneous student data due to varied areas and interests, because forecasting student performance in a certain course is predicated on the student's previous performance in other courses. Similarly, in-course tests that are supposed to be the same for all students are frequently used to predict students' performance in classes.

Predicting student performance with high accuracy is more important since it aids in the identifying of pupils who are academically underperforming. The delivery of better and higher education is the primary goal of all educational institutions.

II. LITERATURE REVIEW

Kasthuriarachchi and S. Liyanage [1] Developed a system that correlates with students' attendance, as well as innovative strategies for encouraging students to improve their attendance and reporting on their success in the classroom. Academicians, educators, and parents will be able to make better decisions as a result of this. In addition, the institute's administration provides financial assistance to students from low-income households through scholarships.

Furthermore, students' performance would suffer as a result of their parents' negligence, therefore parents must keep a constant eye on their children's progress. These variables are beneficial in the process of making key decisions in educational institutions.

Tripathi et al. [2] suggested a study that uses a naive bayes technique to analyse student performance forecasts. In terms of accuracy and overall execution time, the proposed model's results are compared to those of existing systems. According to this study, forecasting students' performance using prediction analysis for complex datasets is a more difficult task. The new model was proved to be more accurate than the previous model and to take less time to execute.

Ching-Chieh Kiu [3] created a model for predicting academic success in students. The authors undertake research to see if student background, coursework success, and social activities have an impact on performance prediction. In predicting student achievement and identifying poor individuals, the author emphasises the importance of a student's past as well as their social activities. Early prediction by the model helps students do better in school and teachers improve the teaching learning process through inventiveness in the teaching process, which improves student performance.

Zhang et al. [4] conducted a study on current teaching learning backgrounds using Technology Mediated Learning (TML). The DA (Dynamic Assessment) method has a bigger influence on teaching and learning approaches. This study solves two research gaps on how DA influences academic achievement in students. For starters, pre- and post-test assessments have been utilised to focus previous research on DA. It is vital to understand the influence of DA based on computer for projecting students' performance over time and to employ information technology to do so. Second, because many students benefit from remote TML support, solutions are being developed based on TML assessments. This system has a few flaws, such as a small class size and the inability to specify external parts.

The methodology for analysing student performance was created by Chew Li Sa et al. This model focuses on creating a system that analyses student performance. For the study, the author considers the course "TMC1013 System Analysis and Design." Various data mining techniques, such as classification algorithms, are employed to ensure the accuracy of performance forecasts throughout the course. The system's key benefit is that it assists teachers and academicians in analysing student performance and identifying poor pupils who may fail the course.

III. METHODOLOGY

The goal of this work is to develop a Machine Learning model for determining student performance.. In this system, the PyCharm is used that makes it easy to set up and access in python.

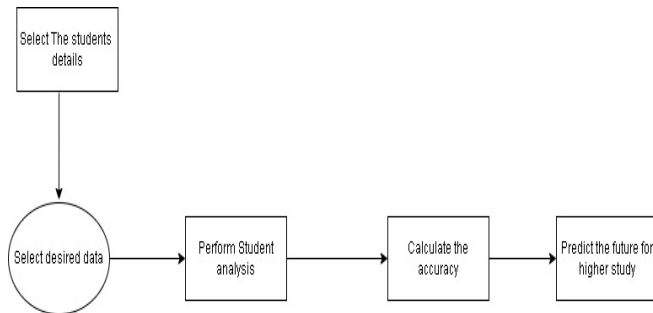


Fig 1: Students performance prediction The steps that require to be followed are:

1. Data Gathering
2. Data Preparation
3. Model Building
4. Analyzing
5. Predicting
6. Result

A. Data Gathering: The user has two datasets for student data, which they must integrate into one.

B. Data Preparation: This is the first stage in any machine learning algorithm. This procedure entails data. Cleaning, data transformation, and data reduction are all things that may be done with data. All of this is done to improve the data's effectiveness. The data can be analysed to improve the accuracy of our model. In order for the categorization to be correct.

1. **Cleaning Data** – First, the data is cleansed. Users are eliminating the unnecessary entries in this area. The dataset is also possible cleaned up by removing the noisy data.
2. **Removing Stop words** – Stop words are removed from the datasets to tidy them up. These are the ones that pop up the most and aren't really important in the text. This can be done to improve the model's accuracy.
3. **Splitting of Data** – The data is divided into training and testing datasets after it has been cleaned. After that, the data chosen for training to train our model is used.

C. Machine Learning- Here machine learning is used to analyze and predict the students performance.

IV. IMPLEMENTATION

The model building is the main step in the Students performance analysis and prediction.

1. First step is open a Pycharm project then, Import the packages that are necessary.

```

import matplotlib
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
from pasta.augment import inline
# %matplotlib inline
import seaborn as sns

sns.set_style('whitegrid')
import statsmodels.api as sm
  
```

Fig 2:

2. Read Datasets

```

mat_ds = pd.read_csv("student-mat.csv", sep=';')
por_ds = pd.read_csv("student-por.csv", sep=';')
  
```

Fig 3:

3. Combine the datasets

```

df = pd.concat([mat_ds, por_ds])
  
```

Fig 4:

4. Calculate the Grade

```

# columns = ['school', 'sex', 'age', 'address', 'family_size', 'parents_status', 'mother_education', 'father_education', 'mother_to',
# final_grade'] = 'na'
# Loc[(df.final_score >= 15) & (df.final_score <= 20), 'final_grade'] = 'good'
# Loc[(df.final_score >= 10) & (df.final_score <= 14), 'final_grade'] = 'fair'
# Loc[(df.final_score >= 0) & (df.final_score <= 9), 'final_grade'] = 'poor'

# isnull().any()
# index.is_unique
# index.duplicated()

# df[~df.index.duplicated()]
  
```

Fig 5:

5. Find the final grade by frequency of going out.

```

out_tab = pd.crosstab(index=df.final_grade, columns=df.go_out)
out_perc = out_tab.apply(perc).reindex(index)

out_perc.plot.bar(colormap="Blues", fontsize=16, figsize=(14,6))
plt.title('Final Grade By Frequency of Going Out', fontsize=20)
plt.ylabel('Percentage of Student', fontsize=16)
plt.xlabel('Final Grade', fontsize=16)

out_table = sm.stats.Table(out_tab)
out_rslt = out_table.test_nominal_association()
out_rslt.pvalue
  
```

Fig 6:

6. Find Final Grade By Desire to Receive Higher Education

```
higher_tab = pd.crosstab(index=df.final_grade, columns=df.desire_higher_edu)
higher_perc = higher_tab.apply(perc).reindex(index)

higher_perc.plot.bar(colormap="YlGnBu", figsize=(14,6), fontsize=16)
plt.title('Final Grade By Desire to Receive Higher Education', fontsize=20)
plt.xlabel('Final Grade', fontsize=16)
plt.ylabel('Percentage of Student', fontsize=16)
```

Fig 7:

```
ks.where(ks==ks.max()).dropna()

# final model
sk = SelectKBest(chi2, k=8)
x_new = sk.fit_transform(X_train, y_train)
x_new_test = sk.fit_transform(X_test, y_test)
lr = lr.fit(x_new, y_train)
print("Logistic Regression Model Score" ,":", lr.score(x_new, y_train) ,":")
"Accuracy" ,":", lr.score(x_new_test, y_test))
```

Fig 8:

```
# final model
tree = DecisionTreeClassifier(min_samples_leaf=17)
t = tree.fit(X_train, y_train)
print("Decision Tree Model Score" ,":", t.score(X_train, y_train) ,":")
"Accuracy" ,":", t.score(X_test, y_test))
```

Fig 9:

V. RESULT

In this work two algorithms are used, that is logistic regression and decision tree. Using these algorithms the accuracy and model scores can be determined .

```
Decision Tree Model Score : 0.8920704845814978 , Accuracy : 0.8512820512820513
Logistic Regression Model Score : 0.8876651982378855 , Accuracy : 0.6564102564102564

Process finished with exit code 0
```

Fig 10:

Then result shows that the Students performance and predict the number of students goes to future studies.

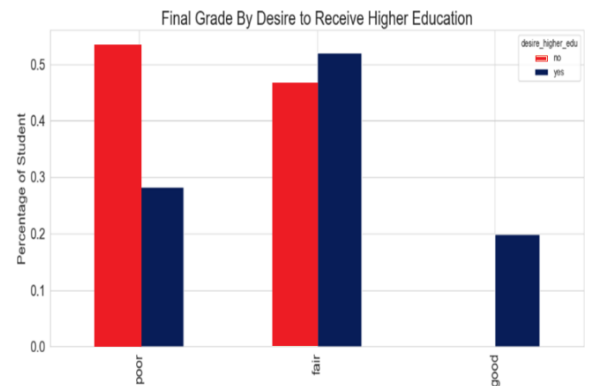


Fig 11:

VI. CONCLUSION

A new technique for predicting students' future performance in degree programmes based on prior academic achievement is offered in this study. Performance both now and in the past. A latent component model-based course clustering method was developed to discover acceptable courses for building base predictors. An ensemble-based progressive prediction architecture was created to incorporate students' evolving performance into the prediction. These data-driven strategies can be used in conjunction with other pedagogical methods to assess students' performance and provide useful information for academic advisors in recommending the number of students who should continue their studies and implementing pedagogical intervention measures as needed. This work will have an impact on the curriculum design of degree programmes as well as education policy planning in general.

VII. REFERENCES

- [1].Kasthuriarachchi, K. T. S., & Liyanage, S. R. (2019), "Predicting Students' Academic Performance Using Utility Based Educational Data Mining", Frontier Computing, pp. 29-39.
- [2].A. Tripathi, S. Yadav and R. Rajan (2019), "Naive Bayes Classification Model for the Student Performance Prediction," 2nd International Conference on Intelligent Computing, Instrumentation.
- [3].C. Kiu (2018), "Data Mining Analysis on Student's Academic Performance through Exploration of Student's Background and Social Activities," Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA), Subang Jaya, Malaysia, pp. 1-5.
- [4]. Zhang, R.-C., Lai, H.-M., Cheng, P.-W., & Chen, C.-P. (2017), "Longitudinal effect of a computer-based graduated prompting assessment on students' academic performance", Computers & Education, , 110, pp. 181-194
- [5]. C. L. Sa, D. H. b. Abang Ibrahim, E. Dahliana Hossain and M. bin Hossin (2014), "Student performance analysis system (SPAS)," The 5th International Conference on Information and Communication Technology for The Muslim (ICT4M), Kuching, pp. 1-6.