Student Performance Prediction

A MINI PROJECT REPORT

18CSC305J - ARTIFICIAL INTELLIGENCE

Submitted by
TANUMAY GHOSH [RA2011027010101]

OMISHA SINGAL [RA2011027010103]

SANTHANA LAKSHMI [RA2011027010129]

Under the guidance of

Dr. Premalatha G

Assistant Professor, Department of Computer Science and Engineering

in partial fulfillment for the award of the degree

of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE & ENGINEERING

of

FACULTY OF ENGINEERING AND TECHNOLOGY

SRM INSTITUTE OF SCIENCE & TECHNOLOGY
Deemed to be University u/s 3 of UGC Act, 1956

S.R.M. Nagar, Kattankulathur, Chengalpattu District

MAY 2023

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

(Under Section 3 of UGC Act, 1956)

BONAFIDE CERTIFICATE

Certified that Mini project report titled **“Student Performance Prediction”** is the bonafide work of **Tanumay Ghosh (RA2011027010101), Omisha Singal (RA2011027010103), Santhana Lakshmi (RA2011027010129)** who carried out the minor project under my supervision. Certified further, that to the best of my knowledge, the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

Dr. Premalatha G
GUIDE
Assistant Professor
Department of Computing Technologies

SIGNATURE

Dr. Lakshmi M
HEAD OF THE DEPARTMENT
Professor & Head
Department of Computing Technologies

ABSTRACT

Student performance prediction is a critical task in the field of education that aims to forecast a student's academic achievement based on various attributes. It involves utilizing data mining and machine learning techniques to extract relevant features from a student's profile, such as their previous academic performance, socio-economic status, and demographic information. By analyzing these attributes, predictive models can be built to anticipate a student's future academic outcomes, such as their grades, dropout rates, and career prospects.

Predictive models have various applications, including early intervention programs, personalized learning, and educational policy-making. These models can help educators identify struggling students early on, allowing them to provide timely support and guidance to improve their academic performance. Additionally, predictive models can be used to personalize learning by tailoring teaching methods and materials to each student's learning style, preferences, and abilities.

However, the development of accurate predictive models requires high-quality data, careful feature selection, and robust modeling techniques. Furthermore, ethical considerations, such as the potential for bias and discrimination, must be addressed to ensure that predictive models do not reinforce existing inequalities in the education system. Despite these challenges, student performance prediction is a rapidly evolving field with significant potential to improve educational outcomes for students.

TABLE OF CONTENTS

LIST OF FIGURES

ABBREVIATIONS

CNN	Convolution Neural Network
ML	Machine Learning
AI	Artificial Intelligence
LSTM	Long Short Term Memory
DL	Deep Learning
SVM	Support Vector Machine
AUC	Area Under Curve

CHAPTER 1

INTRODUCTION

In today's highly competitive education system, student performance prediction has become an essential tool for educators and policymakers alike. The ability to forecast a student's academic performance based on various attributes can help educators provide timely interventions and support to struggling students, personalize learning, and inform policy decisions.

Predictive models that use data mining and machine learning techniques to analyze a student's academic history, demographic information, and socio-economic status can accurately forecast their academic performance. These models can predict a range of outcomes, including grades, dropout rates, and career prospects.

Student performance prediction has several practical applications in the education sector. For instance, it can help identify at-risk students early on and provide them with targeted support to improve their academic performance. It can also help personalize teaching methods and materials to suit each student's learning style and preferences, leading to better learning outcomes.

Despite its many benefits, student performance prediction is not without its challenges. Developing accurate predictive models requires high-quality data and robust modeling techniques, which can be time-consuming and resource-intensive. Moreover, ethical considerations, such as the potential for bias and discrimination, must be addressed to ensure that predictive models do not reinforce existing inequalities in the education system.

Despite these challenges, student performance prediction has enormous potential to transform the education sector by providing actionable insights into student performance and helping educators make data-driven decisions.

Student performance prediction refers to the use of data mining and machine learning techniques to forecast a student's academic performance based on various attributes. These attributes may include a student's previous academic performance, demographic information, socio-economic status, and other relevant factors.

Predictive models are developed by analyzing historical data to identify patterns and relationships between a student's attributes and their academic performance. These models can predict a range of outcomes, such as grades, dropout rates, and career prospects.

Student performance prediction has several practical applications in the education sector. For example, it can help educators identify struggling students early on and provide them with targeted support and interventions to improve their academic performance. It can also be used to personalize teaching methods and materials to suit each student's learning style and preferences.

The development of accurate predictive models requires high-quality data and robust modeling techniques. It is crucial to use appropriate data preprocessing techniques and feature selection methods to ensure that the models are accurate and unbiased. Moreover, ethical considerations, such as the potential for bias and discrimination, must be addressed to ensure that predictive models do not reinforce existing inequalities in the education system.

Despite its potential benefits, student performance prediction is not without its challenges. There are concerns about data privacy and the potential for predictive models to be used to stigmatize or label students unfairly. It is essential to ensure that the use of predictive models is transparent, explainable, and aligned with the values and goals of the education system.

Overall, student performance prediction is a rapidly evolving field with significant potential to improve educational outcomes for students. It involves the use of sophisticated data analysis techniques to identify patterns and relationships that can inform educational policy and support the development of effective interventions to improve academic performance.

CHAPTER 2

LITERATURE SURVEY

Student performance analysis projects typically involve using data and machine learning techniques to predict and analyze the academic performance of students.
We have worked upon the Literature survey and here are some of the key findings:

- Predicting Student Performance using Personalized Long Short-Term Memory Recurrent Neural Networks" by Mohammad Aliannejadi, et al. (2019): The authors used personalized long short-term memory (LSTM) recurrent neural networks to predict the performance of individual students. The results showed that the model had higher accuracy compared to other models.

- A Hybrid Deep Learning Model for Predicting Students Academic Performance" by Rupesh Kumar Gupta, et al. (2021): The authors used a hybrid deep learning model that combined a convolutional neural network (CNN) and LSTM to predict the academic performance of students. The model achieved an accuracy of 89.46%.

- Predicting Student Performance in MOOCs using Clickstream Data and Probabilistic Graphical Models" by Behrouz Haji Soleimani, et al. (2019): The authors used clickstream data from massive open online courses (MOOCs) to predict the performance of students. They used probabilistic graphical models and achieved an accuracy of 89.3%.

- A Comparative Study of Machine Learning Techniques for Predicting Student Academic Performance" by Moustafa Mahmoud Eissa, et al. (2021): The authors compared the performance of different machine learning techniques for predicting student academic performance. They found that support vector machines (SVM) and random forests had the highest accuracy.

- An Intelligent Tutoring System for Predicting Student Performance" by Ahmed Hammad, et al. (2018): The authors developed an intelligent tutoring system that used Bayesian networks to predict student performance. The system achieved an accuracy of 84.6%.

Overall, these papers demonstrate the effectiveness of various machine learning and deep learning techniques for predicting student performance. Personalized models, hybrid models, and models that use clickstream data show particular promise.

CHAPTER 3

SYSTEM ARCHITECTURE AND DESIGN

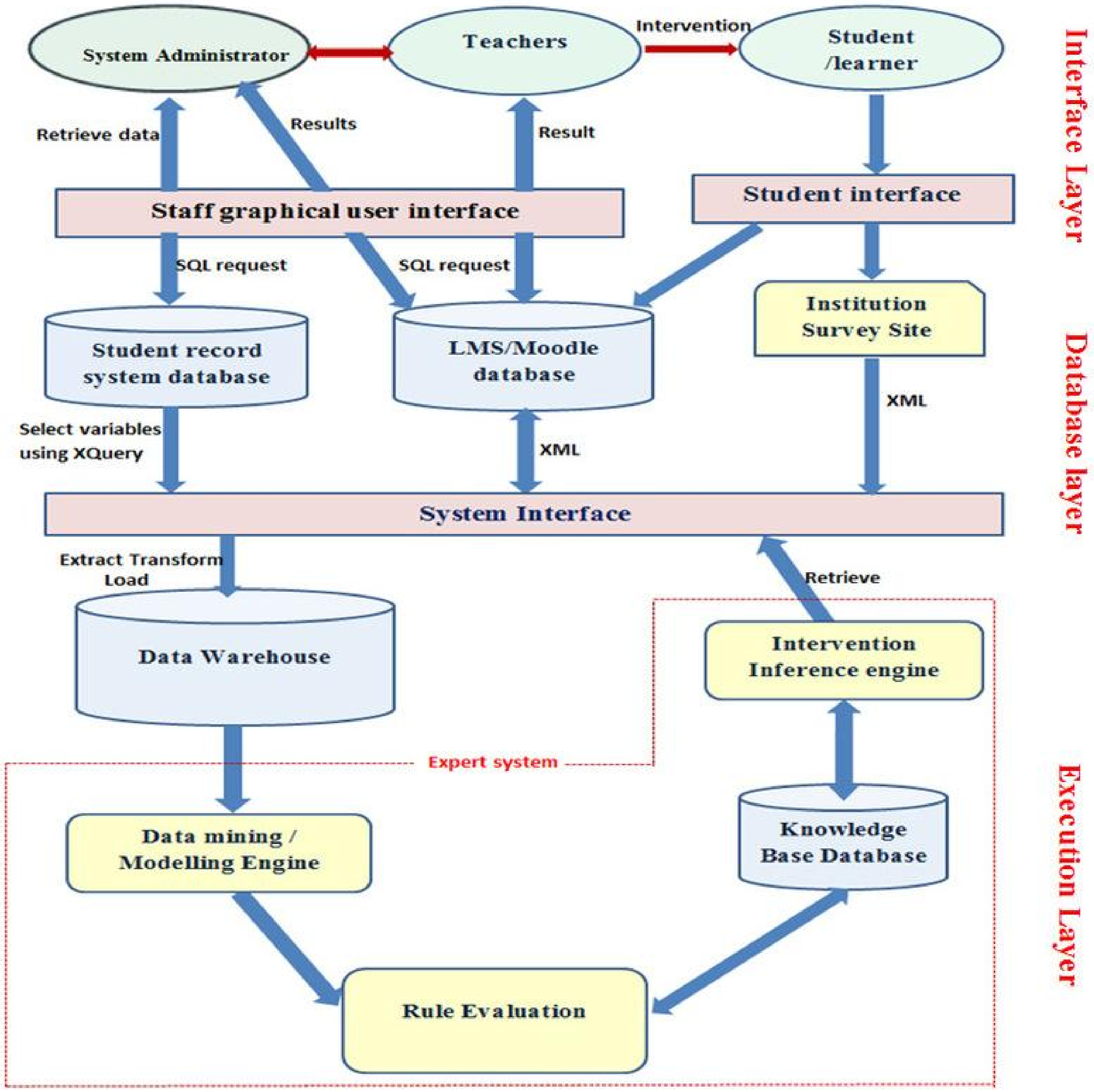

fig 3.1

CHAPTER 4

METHODOLOGY

- Data collection: The first step is to collect data on student performance, such as grades, attendance, and test scores. This data may come from various sources, such as student information systems, learning management systems, and assessments.

- Data preprocessing: The collected data may need to be cleaned and processed before it can be used for analysis. This includes tasks such as removing duplicates, dealing with missing values, and normalizing the data.

- Feature engineering: The next step is to identify features that are relevant for predicting student performance. This may include demographic data, socio-economic status, previous academic history, and other factors.

- Model selection: Depending on the problem statement, different machine learning models can be selected. Some popular models for student performance prediction are decision trees, random forests, neural networks, and support vector machines.

- Model training: The selected model is trained using the preprocessed data. This involves using a portion of the data to train the model and another portion to validate its performance.

- Hyperparameter tuning: Hyperparameters are parameters that are set before training the model, such as the learning rate and regularization. These parameters can be tuned to improve the performance of the model.

- Model evaluation: The performance of the model is evaluated using various metrics such as accuracy, precision, recall, and F1 score. This step helps to determine the effectiveness of the model in predicting student performance.

- Deployment: Once the model is trained and evaluated, it can be deployed in a real-world setting to predict student performance. The deployment may involve integrating the model into an existing system or developing a new application for educators to use.

CHAPTER 5

CODING

```
# Assigning grades to the grades according to the following criteria :

# 0  - 40 marks : grade E

# 41 - 60 marks : grade D

# 60 - 70 marks : grade C

# 70 - 80 marks : grade B

# 80 - 90 marks : grade A

# 90 - 100 marks : grade O


def getgrade(percentage, status):

  if status == 'Fail':

    return 'E'

  if(percentage >= 90):

    return 'O'

  if(percentage >= 80):

    return 'A'

  if(percentage >= 70):

    return 'B'

  if(percentage >= 60):

    return 'C'

  if(percentage >= 40):

    return 'D'

  else :

    return 'E'


data['grades'] = data.apply(lambda x: getgrade(x['percentage'], x['status']),
```

```
axis = 1 )


data['grades'].value_counts()

from sklearn.preprocessing import LabelEncoder


# creating an encoder

le = LabelEncoder()


# label encoding for test preparation course

data['test preparation course'] = le.fit_transform(data['test preparation
course'])

data['test preparation course'].value_counts()

# label encoding for race/ethnicity

# we have to map values to each of the categories


data['race/ethnicity'] = data['race/ethnicity'].replace('group A', 1)

data['race/ethnicity'] = data['race/ethnicity'].replace('group B', 2)

data['race/ethnicity'] = data['race/ethnicity'].replace('group C', 3)

data['race/ethnicity'] = data['race/ethnicity'].replace('group D', 4)

data['race/ethnicity'] = data['race/ethnicity'].replace('group E', 5)


data['race/ethnicity'].value_counts()

# label encoding for grades

# we have to map values to each of the categories


data['grades'] = data['grades'].replace('O', 0)

data['grades'] = data['grades'].replace('A', 1)

data['grades'] = data['grades'].replace('B', 2)
```

```
data['grades'] = data['grades'].replace('C', 3)

data['grades'] = data['grades'].replace('D', 4)

data['grades'] = data['grades'].replace('E', 5)


data['race/ethnicity'].value_counts()

# splitting the dataset into training and test sets


from sklearn.model_selection import train_test_split


x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.25,
random_state = 45)


print(x_train.shape)

print(y_train.shape)

print(x_test.shape)

print(y_test.shape)

# importing the MinMaxScaler

from sklearn.preprocessing import MinMaxScaler


# creating a scaler

mm = MinMaxScaler()


# feeding the independent variable into the scaler

x_train = mm.fit_transform(x_train)

x_test = mm.transform(x_test)


from sklearn.linear_model import  LogisticRegression
```

```python
# creating a model
model = LogisticRegression()


# feeding the training data to the model
model.fit(x_train, y_train)


# predicting the test set results
y_pred = model.predict(x_test)


# calculating the classification accuracies
print("Training Accuracy :", model.score(x_train, y_train))
print("Testing Accuracy :", model.score(x_test, y_test))


# printing the confusion matrix


from sklearn.metrics import confusion_matrix


# creating a confusion matrix
cm = confusion_matrix(y_test, y_pred)


# printing the confusion matrix
print(cm)
```

TESTING

Testing on student performance prediction involves evaluating the accuracy and effectiveness of predictive models in forecasting a student's academic performance. This involves dividing the available data into a training set and a testing set to assess the performance of the model on unseen data.

Various performance metrics can be used to evaluate the effectiveness of predictive models, including accuracy, precision, recall, F1 score, and area under the curve (AUC). These metrics provide a quantitative measure of how well the model performs in predicting student performance.

To ensure that predictive models are accurate and unbiased, it is essential to use appropriate data preprocessing techniques and feature selection methods. It is also crucial to consider ethical considerations, such as ensuring that the predictive models do not reinforce existing biases or inequalities in the education system.

In addition to evaluating the performance of predictive models, it is essential to consider their practical applications and potential impact on the education system. Predictive models should be used to support, rather than replace, the expertise of educators and should be transparent and explainable to ensure that they are trusted and accepted by stakeholders.

Overall, testing on student performance prediction plays a critical role in evaluating the effectiveness of predictive models and ensuring that they are accurate, unbiased, and aligned with the needs of the education system.

- **Unit testing:** Test each individual component of your algorithm to ensure they are working as expected.

- **Integration testing:** Test how well different components of your algorithm work together.

- **Regression testing:** Test the algorithm's performance over time to ensure it maintains accuracy.

- **Performance testing:** Test how well the algorithm can handle large amounts of data and ensure it provides results in a timely manner.

- **Cross-validation testing:** Test the algorithm's ability to make accurate predictions on new and unseen data.

- **Accuracy testing:** Test the accuracy of the algorithm's predictions by comparing them against actual outcomes.

- **Error testing:** Test how the algorithm handles errors and unexpected inputs, ensuring it provides appropriate error messages or fallbacks.

- **User acceptance testing:** Test the algorithm's overall usability and user experience, ensuring it meets user expectations and is easy to use.

Hence, here the algorithm validates and fulfills all the requirements for testing our modules. Different testing techniques are tested and approved.

CHAPTER 6

SCREENSHOTS AND RESULTS

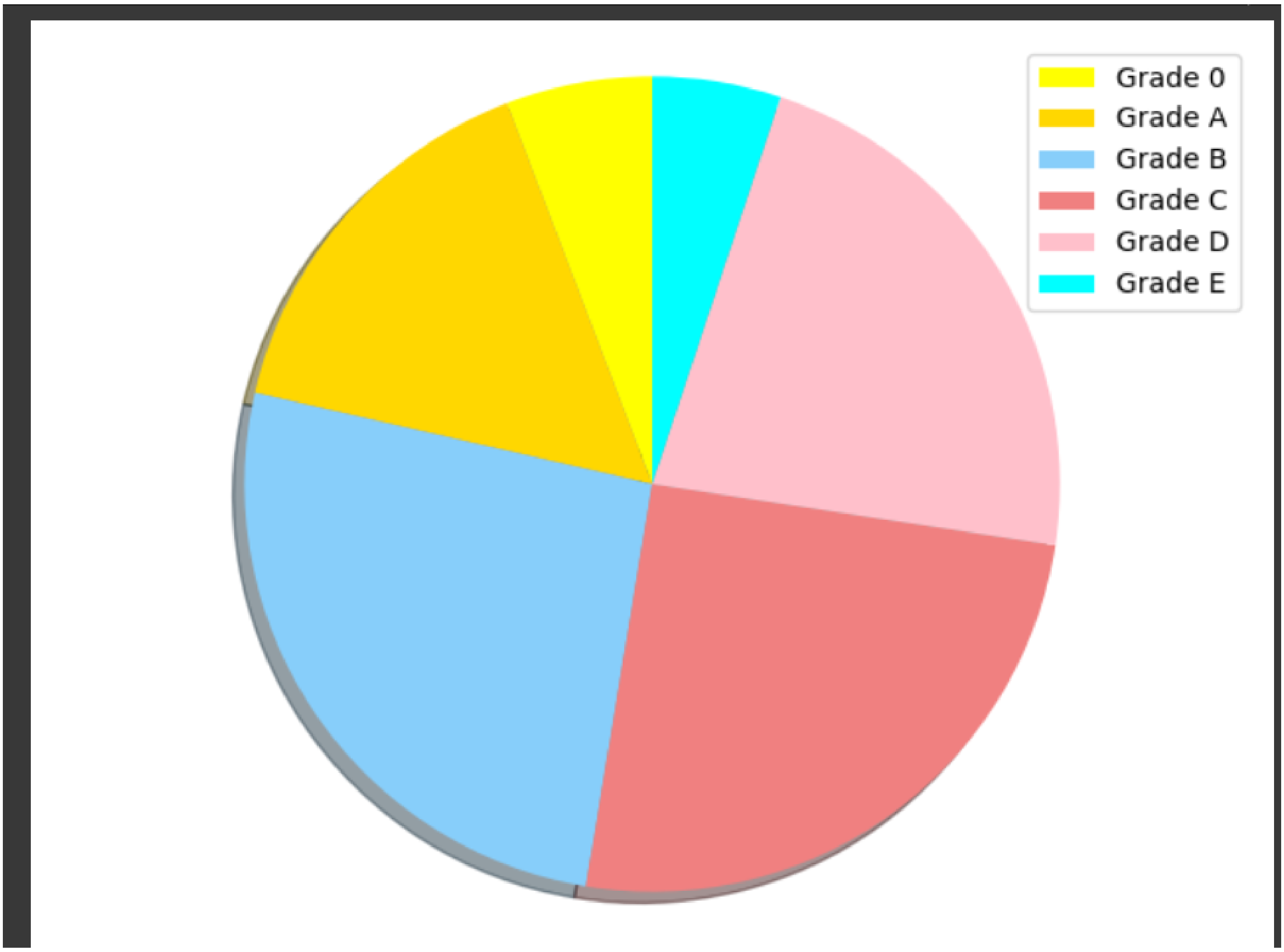

fig 6.1

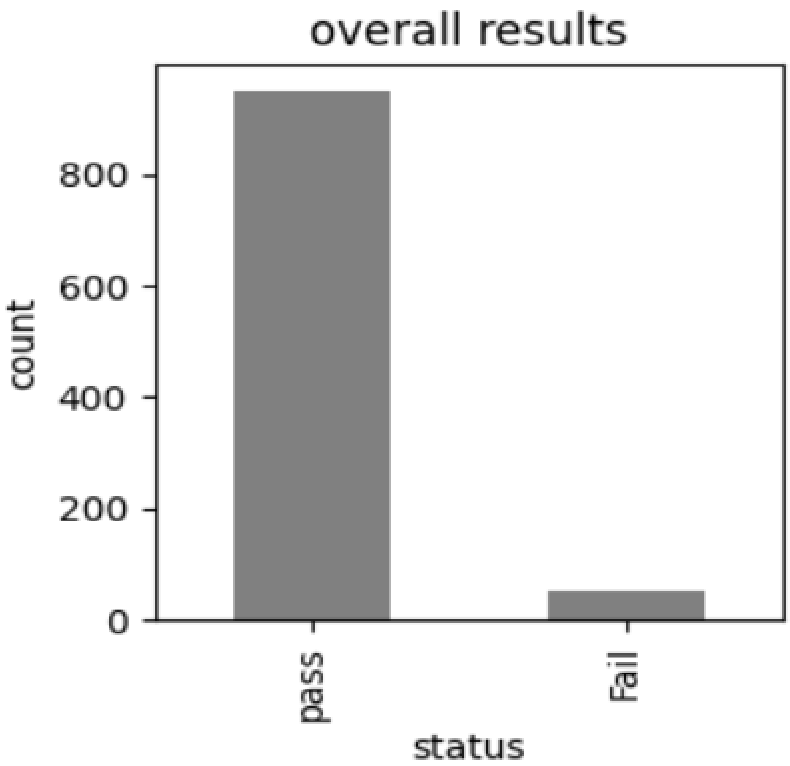
overall results
count
800
600
400
200
0
pass
Fail
status

fig 6.2

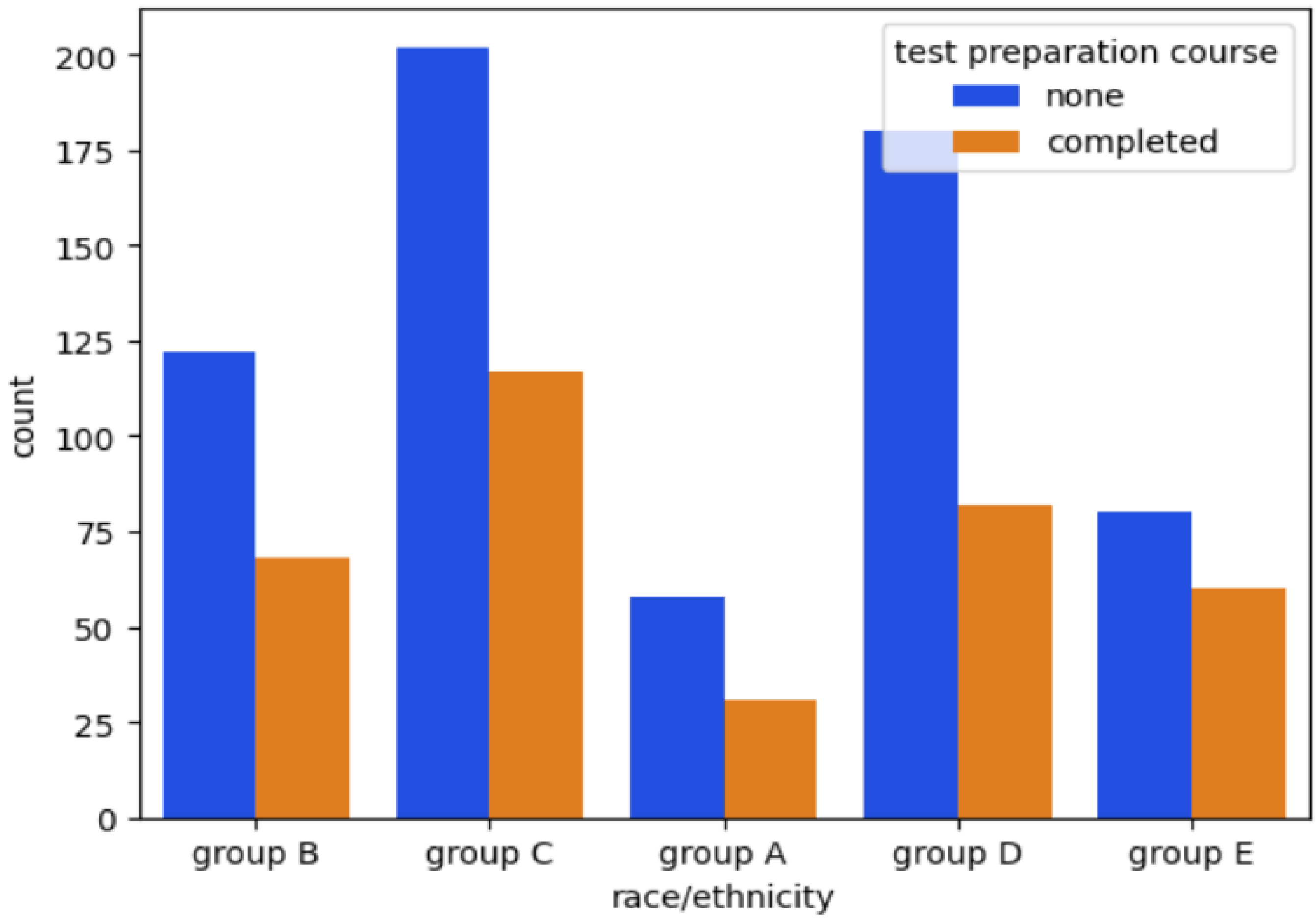
test preparation course
none
completed
count
200
175
150
125
100
75
50
25
0
group B
group C
group A
group D
group E
race/ethnicity

fig 6.3

CHAPTER 7

CONCLUSION AND FUTURE ENHANCEMENTS

CONCLUSION:

In conclusion, a student performance prediction project using AI/ML has the potential to provide valuable insights into the academic performance of students. By analyzing data on student performance and identifying relevant factors, machine learning models can be trained to predict the future performance of individual students. This can help educators identify at-risk students and provide targeted interventions to help them succeed.

The methodology for a student performance prediction project typically involves data collection, preprocessing, feature engineering, model selection, model training, hyperparameter tuning, model evaluation, and deployment. Future enhancements to the project may include incorporating more data sources, developing personalized models, using dynamic models, incorporating explainable AI techniques, and implementing continuous learning.

Overall, a student performance prediction project has the potential to improve educational outcomes by identifying at-risk students and providing targeted interventions to help them succeed. With the ongoing development of AI and machine learning techniques, the potential for these projects to have a positive impact on education is only increasing.

FUTURE ENHANCEMENTS

There are several potential future enhancements that can be made on the project are as follows:

- Incorporating more data sources: The performance of the model can be improved by incorporating additional data sources such as student behavior, social interactions, and other factors that may affect academic performance.
- Personalized models: Instead of predicting the academic performance of a group of students, personalized models can be developed to predict the performance of individual students. This can provide more accurate predictions and targeted interventions.
- Dynamic models: Instead of using historical data to make predictions, dynamic models can be developed that can adjust their predictions in real-time based on new data. This can help educators identify students who are falling behind and provide timely interventions.
- Explainable AI: In some cases, the output of the model may not be immediately interpretable. Incorporating explainable AI techniques can help to make the output more transparent and interpretable to educators.
- Continuous learning: The model can be updated and retrained on a regular basis using new data. This can help to improve the accuracy of the model over time and ensure that it stays up-to-date with changing educational environments.

Overall, these enhancements have the potential to make student performance prediction models more accurate, personalized, and effective in helping educators to identify students who are at risk and provide timely interventions.

REFERENCES

https://www.researchgate.net/figure/Method-proposed-for-student-performance-prediction_fig1_261761241

https://www.researchgate.net/figure/Architectural-framework-of-the-holistic-student-performance-prediction-system

https://ieeexplore.ieee.org/document/9360749

https://www.analyticsvidhya.com/blog/2023/04/student-performance-analysis-and-prediction/