Report On

Student Result Analysis

Submitted in partial fulfillment of the requirements of the Mini project
in Semester VII of Fourth Year Artificial Intelligence & Data Science Engineering

by

Krish Mehta (62)

Hardik Raut (65)

Bhavya Wade (72)

Under the guidance of

Ms. Bhavika Gharat

University of Mumbai

Vidyavardhini's College of Engineering & Technology

Department of Artificial Intelligence and Data Science

(A.Y. 2024-25)

CERTIFICATE

This is to certify that the Mini Project entitled **"Student Result Analysis"** is a bonafide work of**, Krish Mehta (62), Hardik Raut (65), Bhavya Wade(72)** submitted to the University of Mumbai in partial fulfillment of the requirement for the award of the degree of **"Bachelor of Engineering"** in Semester VII of Fourth Year **"Artificial Intelligence and Data Science".**

Dr. Tatwadarshi Nagarhalli
Guide

Ms Sejal D'Mello	Dr. Tatwadarshi P N	Dr. H.V. Vankudre
Deputy HOD AI & DS	HOD AI & DS	Principal

Mini Project Approval

This Mini Project entitled **"Udemy Courses Data Analysis"** by **Krish Mehta (62), Hardik Raut (65), Bhavya Wade(72)** is approved for the degree of **Bachelor of Engineering** in in Semester VII of Fourth Year **Artificial Intelligence and Data Science.**

Examiners

1…….....
(Internal Examiner Name & Sign)

2……
(External Examiner Name & Sign)

Date:

Place:

Abstract

In today's education systems, the analysis of student performance has become increasingly vital for identifying areas that require improvement and facilitating data-driven decision-making. Traditional methods of tracking academic performance often rely on manual processes, leading to time-consuming and error-prone outcomes. These approaches can hinder the ability of educators to efficiently identify at-risk students and implement timely interventions. As educational institutions strive for more effective management of student data, there is a growing need for automated systems that enhance the accuracy and speed of performance analysis.

This project addresses these challenges by developing a Student Result Analysis System that employs machine learning techniques to classify and predict student outcomes based on various influential factors. By utilizing classification models such as decision trees and support vector machines (SVM), the system can analyze critical variables, including previous grades, attendance, and demographic information. The predictive capabilities of these models enable educators to derive actionable insights from the data, thus allowing them to tailor their interventions more effectively.

The system not only assists teachers and administrators in understanding student performance trends but also empowers students by providing them with personalized feedback. By identifying at-risk students early in the academic cycle, the Student Result Analysis System facilitates timely support and intervention, ultimately fostering a more supportive learning environment. This proactive approach aims to enhance educational outcomes by ensuring that students receive the assistance they need to succeed.

Overall, the integration of advanced data analysis techniques in educational settings marks a significant step towards a more data-informed approach to student performance management. This project contributes to the growing body of knowledge in educational technology and highlights the importance of leveraging machine learning for effective student outcome predictions.

Acknowledgement:

We would like to thank all people whose support and cooperation has been an invaluable asset during this Project. We would also like to thank our Guide **Ms. Bhavika Gharat**, for guiding us throughout this project and giving it the present shape. It would have been impossible to complete the project without their support, valuable suggestions, criticism, encouragement, and guidance.

We convey our gratitude to **Dr. Tatwadarshi Nagarhalli**, Head of Department, for their motivation and providing various facilities, which helped us greatly in the whole process of this project. We are also grateful to all other teaching and non-teaching staff members of the Artificial Intelligence and Data Science Department for directly or indirectly helping us with the completion of projects and the resources provided.

Mokshad Sankhe (72)

Sudeep Shetty (75)

Date:

Place:

Contents

List of Abbreviations

1. **EDA**: Exploratory Data Analysis
2. **ML**: Machine Learning
3. **Jupyter**: Jupyter Notebooks
4. **Pandas**: Python Data Analysis Library
5. **NumPy**: Numerical Python
6. **Scikit-learn**: Machine Learning Library in Python
7. **API**: Application Programming Interface
8. **CSV**: Comma-Separated Values
9. **OS**: Operating System
10. **RAM**: Random Access Memory
11. **UI**: User Interface
12. **SQL**: Structured Query Language
13. **SVM**: Support Vector Machine
14. **KNN**: K-Nearest Neighbors
15. **TPR**: True Positive Rate
16. **FPR**: False Positive Rate
17. **ROC**: Receiver Operating Characteristic
18. **AUC**: Area Under the Curve
19. **TP**: True Positive
20. **FP**: False Positive
21. **TN**: True Negative
22. **FN**: False Negative
23. **F1 Score**: Harmonic Mean of Precision and Recall
24. **Accuracy**: Proportion of true results among the total number of cases examined

1. INTRODUCTION

1.1 INTRODUCTION

In education, keeping track of student performance is essential but often challenging. Traditionally, teachers have relied on manual methods to calculate grades and assess progress, which can be time-consuming and prone to errors. These conventional methods make it difficult to spot patterns in student performance early, which delays interventions that could improve learning outcomes. In the age of data analytics, the opportunity to transform this process with more efficient and accurate tools has never been greater.

Student Result Analysis is an innovative, data-driven approach designed to automate and enhance the process of tracking student progress. By employing machine learning models, this system can predict a student's future performance based on historical data, such as previous grades, attendance, and participation in activities. This predictive capability allows educators to not only monitor current performance but also anticipate future outcomes and identify students who may be at risk of failing. By highlighting areas for improvement early, teachers and administrators can provide timely support and intervention.

This system isn't just a tool for teachers—it's designed for anyone involved in education, from school administrators to policymakers, and even students themselves. With data analysis at its core, **Student Result Analysis** helps institutions make informed decisions, tailor learning experiences, and ensure that each student has the best chance to succeed, all while reducing the manual burden traditionally associated with student evaluation.

1.2 PROBLEM STATEMENT & OBJECTIVE

Problem Statement:

Manually tracking and analyzing student performance is inefficient and prone to human error. Inconsistent data and late interventions can lead to students falling behind. Educators need a system that can quickly and accurately predict student outcomes and identify those who need help the most.

Objectives:

- Develop a student result analysis tool that provides real-time insights on student performance.
- Predict future student performance using decision trees and support vector machines.
- Help teachers and administrators make better, data-driven decisions.
- Offer actionable insights that allow for timely interventions, helping students who are at risk.

1.3 SCOPE

Target Users:

1. **Teachers and Administrators in Schools and Universities**: One of the primary users of the **Student Result Analysis** tool are teachers and administrators who need an efficient way to manage and interpret student data. They can use this system to automate grade calculations, track student progress over time, and identify trends that would otherwise be difficult to notice through manual methods. With the tool's ability to analyze vast amounts of data, educators can make more informed decisions regarding student support, curriculum adjustments, and targeted interventions, ensuring that no student falls behind.
2. **Education Policymakers**: Another key audience for this tool are policymakers responsible for shaping educational systems at a higher level. These stakeholders need access to comprehensive, data-driven insights to evaluate the effectiveness of educational programs and implement changes. The **Student Result Analysis** system offers a bird's-eye view of student performance across various metrics, helping policymakers to create policies that are more responsive to the needs of students, promote better learning outcomes, and allocate resources effectively.
3. **Students Themselves**: Finally, students are also direct beneficiaries of this system. By allowing them to access their own academic data, students can track their performance and identify areas that need improvement. This self-awareness can foster greater responsibility for their own learning journey. The system provides personalized insights and recommendations based on individual performance, enabling students to take proactive steps toward academic success and stay motivated.

Features:

1. **Analyze Student Data and Predict Outcomes**: The core feature of the **Student Result Analysis** tool is its ability to process and analyze multiple sources of student data, such as grades, attendance, and participation. By leveraging machine learning algorithms, the system can predict future academic performance with high accuracy. This predictive feature helps educators anticipate students' needs and allows for more strategic planning around their academic progress.
2. **Identify Students at Risk of Failing**: One of the most powerful aspects of the tool is its ability to flag students who may be at risk of underperforming or failing. The system uses historical data trends to recognize warning signs, such as declining grades or frequent absences. By identifying these students early, educators can implement targeted interventions, such as tutoring or counseling, before it becomes too late to make a difference.
3. **Offer Recommendations for Improving Student Performance**: In addition to identifying at-risk students, the tool provides actionable recommendations to improve academic outcomes. Based on the analysis of individual and group performance, the system suggests strategies tailored to each student's needs. These recommendations might include study tips, subject-specific resources, or personalized learning paths, allowing both educators and students to focus on areas that need the most attention.

1.4 TECHNOLOGIES:

Python: Used for data processing, modeling, and analysis.
Scikit-Learn: A popular machine learning library for implementing decision trees and support vector machines.
Pandas: Essential for data manipulation and analysis.
NumPy: Used for handling numerical operations.
Matplotlib & **Seaborn**: For creating visualizations and analysis charts.

2. LITERATURE SURVEY

Machine learning in education is not new. Many researchers have explored how algorithms can help analyze student data to predict performance. Here's a quick look at what others have done:

- **Student Performance Prediction Using Decision Trees**: This approach has shown strong accuracy in classifying student grades based on features like attendance, previous scores, and socioeconomic status.
- **Support Vector Machines (SVMs) for Classification**: SVMs are often used to handle high-dimensional data and can be highly effective in predicting student outcomes. However, they require careful tuning for best results.

2.1 SURVEY OF EXISTING SYSTEM

1. Baker et al. (2019)

- **Study Focus**: This research focuses on early prediction of student performance in traditional school settings using learning management system (LMS) data. By analyzing factors like student interaction with LMS, homework submissions, and quiz scores, the study aims to predict final grades and identify students at risk of underperforming.
- **Key Insights**: The study demonstrates that LMS interaction data can be a powerful predictor of student success. Early identification of at-risk students allows educators to intervene with additional support.
- **Conclusion**: This system highlights how leveraging real-time interaction data from educational platforms can enhance student result analysis, improving both performance tracking and early intervention.

2. Kumar et al. (2020)

- **Study Focus**: Kumar and colleagues developed a predictive analytics model to assess student academic performance based on historical data, including grades, attendance, and participation in extracurricular activities.
- **Key Insights**: By using machine learning algorithms, the study identified patterns in student behavior that correlate with academic success or failure. The system allows for the classification of students into different performance groups and provides targeted recommendations to improve outcomes.
- **Conclusion**: This research underscores the value of a multi-dimensional approach, incorporating not only grades but also extracurricular factors to make accurate predictions about student performance.

3. Sweeney et al. (2021)

- **Study Focus**: This paper introduces an automated result analysis system for universities, focusing on using decision tree algorithms to predict students' final results. The system relies on input variables like midterm scores, project grades, and classroom participation.

- **Key Insights**: The decision tree-based system proved effective in identifying students who were at risk of failing, allowing institutions to implement targeted interventions. It also provided personalized feedback based on predicted performance.
- **Conclusion**: The research supports the use of machine learning techniques, particularly decision trees, for providing both students and educators with insights into performance trends and personalized academic strategies.

2.2 LIMITATION OF EXISTING SYSTEM:

Sr No	Paper Title	Published Year	Limitations	Research Gap
1	Baker et al.	2019	Limited to data from a single LMS; may not generalize to other platforms.	Need for integrating data from multiple sources for a more holistic view of student performance.
2	Kumar et al.	2020	Relies heavily on historical data; may not account for sudden changes in student behavior.	Exploration of dynamic factors that influence student performance in real-time.
3	Sweeney et al.	2021	Decision trees may oversimplify complex relationships in student data.	Development of more sophisticated models (e.g., ensemble methods) for improved prediction accuracy.

2.3 MINI PROJECT CONTRIBUTION:

This project aims to bridge gaps in existing student result analysis by leveraging advanced data analysis techniques. By incorporating a detailed evaluation of student grades, attendance records, and interaction data, it seeks to identify key patterns that contribute to academic success. The project extends traditional methods by focusing not only on static metadata (such as past grades and attendance) but also on dynamic factors like student feedback and engagement, which often provide a more nuanced understanding of student performance.

Through this approach, the project aims to develop a predictive model that not only forecasts student outcomes but also identifies actionable insights for performance improvement. The model can recommend interventions for at-risk students based on real-time data, enabling educators to provide timely support. Moreover, by personalizing academic resources and recommendations for students based on their past behaviors and preferences, the system will offer a more tailored educational experience, driving higher engagement and academic success.

Ultimately, this project will contribute to improving both the quality of education and the effectiveness of academic programs, providing valuable insights for teachers, administrators, and policymakers involved in student success.

3. PROPOSED SYSTEM

3.1 DATASETS

The dataset contains several features including:

- Student ID
- Marks obtained in various subjects
- Attendance
- Participation in extracurricular activities
- Previous grades

The dataset was preprocessed to handle missing values, normalize marks, and encode categorical variables like attendance status.

3.2 DETAILS OF HARDWARE & SOFTWARE

Hardware:

- Processor: Intel i5 or higher
- RAM: Minimum 8 GB
- Storage: 256 GB SSD

Software:

- Python 3.x
- Jupyter Notebook or Google Colab
- Scikit-learn, Pandas, Matplotlib, Seaborn

4. IMPLEMENTATION

4.1 FLOW DIAGRAM

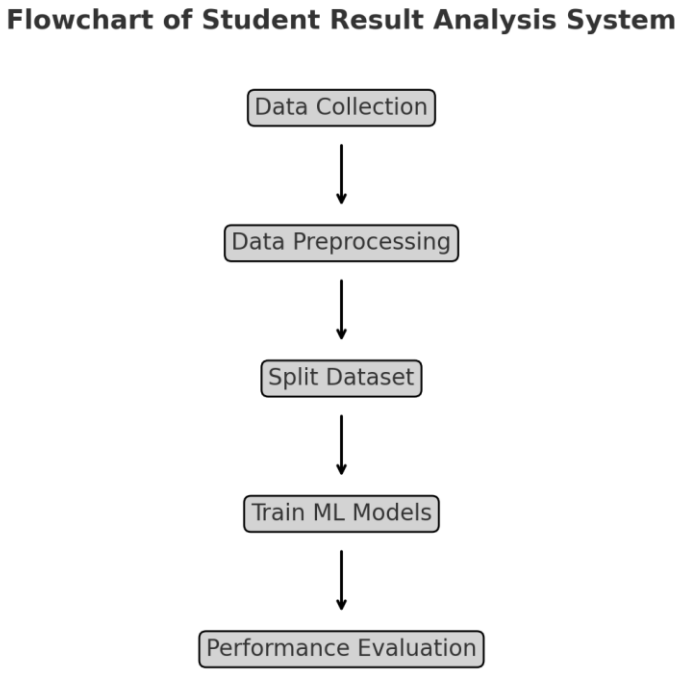


Fig. 4.1.1: Working

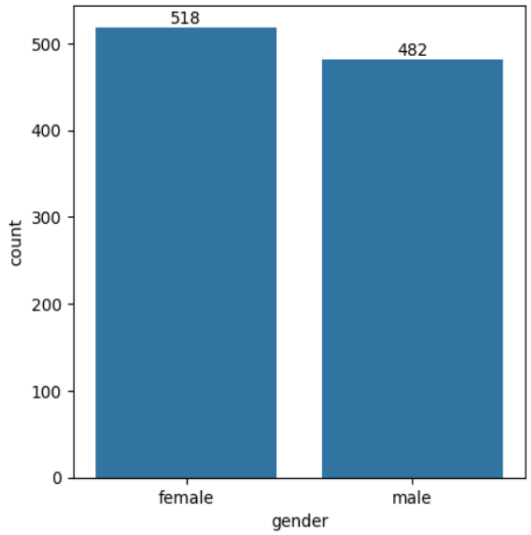
4.2 RESULTS:

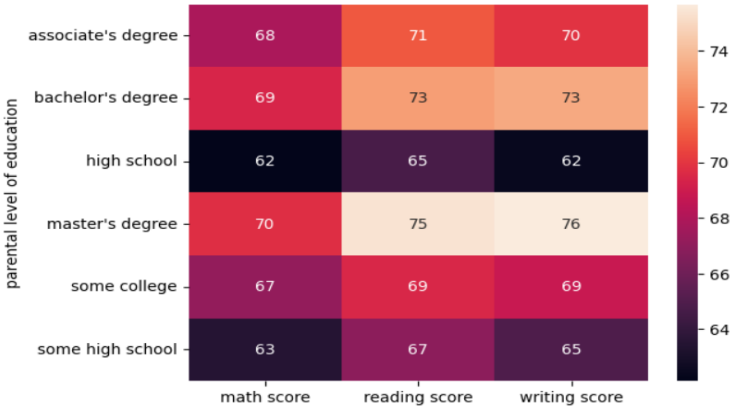
Code:

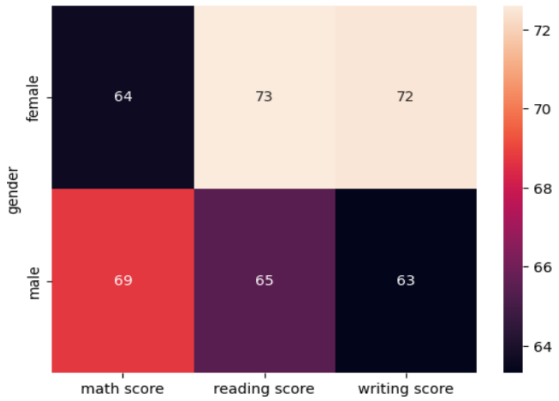
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
df = pd.read_csv("StudentsPerformance.csv")
df.head()#first 5 rows of the dataset
df.describe()#summary statistics of the dataset
```

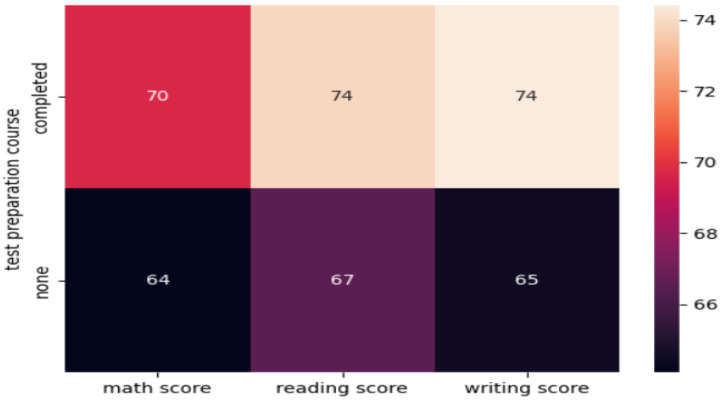
```
df.isnull().sum()#checking for missing values
# GENDER DISTRIBUTION
plt.figure(figsize=(5,5))
ax = sns.countplot(data=df,x='gender')
ax.bar_label(ax.containers[0])
plt.show()
gb = df.groupby("parental level of education").agg({"math score":"mean","reading
score":"mean","writing score":"mean"})
print(gb)
sns.heatmap(gb,annot=True)
plt.show()
gb1 = df.groupby("gender").agg({"math score":"mean","reading score":"mean","writing
score":"mean"})
print(gb1)
sns.heatmap(gb1,annot=True)
plt.show()
gb2 = df.groupby("test preparation course").agg({"math score":"mean","reading
score":"mean","writing score":"mean"})
print(gb2)
sns.heatmap(gb2,annot=True)
plt.show()
sns.boxplot(data=df,x='math score')
plt.show()
sns.boxplot(data=df,x='reading score')
plt.show()
sns.boxplot(data=df,x='writing score')
plt.show()
print(df["race/ethnicity"].unique())
groupA = df.loc[(df['race/ethnicity']=='group A')].count()
groupB = df.loc[(df['race/ethnicity']=='group B')].count()
groupC = df.loc[(df['race/ethnicity']=='group C')].count()
groupD = df.loc[(df['race/ethnicity']=='group D')].count()
groupE = df.loc[(df['race/ethnicity']=='group E')].count()
mlist=[groupA['race/ethnicity'],groupB['race/ethnicity'],groupC['race/ethnicity'],groupD['race/ethni
city'],groupE['race/ethnicity']]
plt.pie(mlist,labels=['group A','group B','group C','group D','group E'],autopct='%1.1f%%')
plt.title('Distribution of groups')
plt.show()
ax = sns.countplot(data=df,x='race/ethnicity')
ax.bar_label(ax.containers[0])
gb3 = df.groupby("lunch").agg({"math score":"mean","reading score":"mean","writing
score":"mean"})
print(gb3)
sns.heatmap(gb3,annot=True)
plt.show()
```

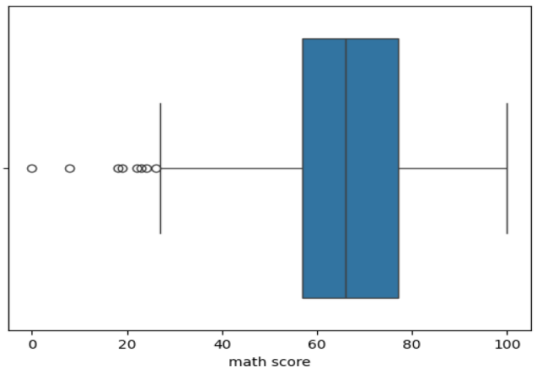
Output:

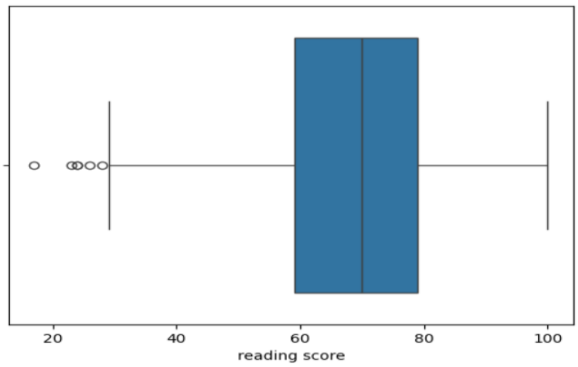

518
482
500
400
300
200
100
0
count
female
male
gender

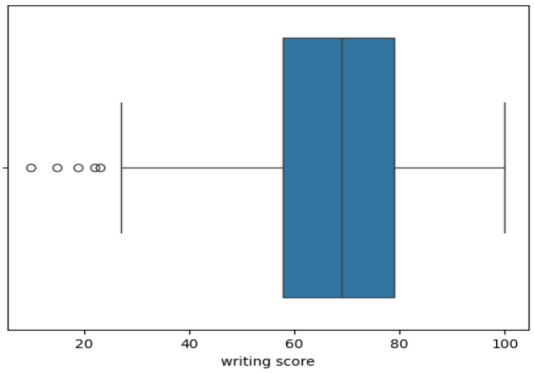

associate's degree
bachelor's degree
high school
master's degree
some college
some high school
parental level of education
68
71
70
69
73
73
62
65
62
70
75
76
67
69
69
63
67
65
math score
reading score
writing score
74
72
70
68
66
64

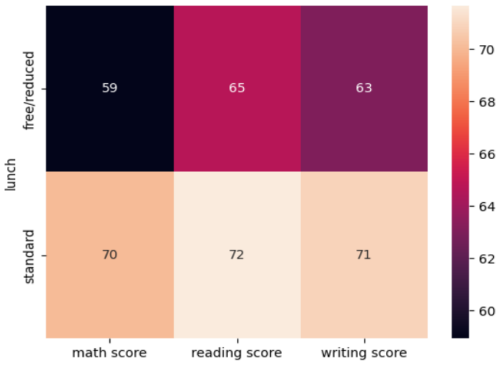

female
male
gender
64
73
72
69
65
63
math score
reading score
writing score
72
70
68
66
64

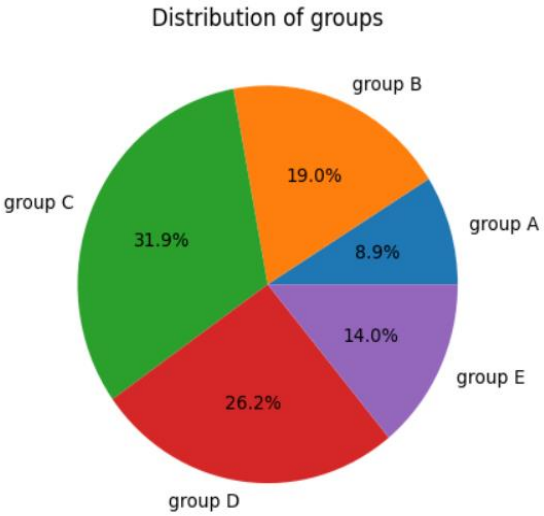

completed
none
test preparation course
70
74
74
64
67
65
math score
reading score
writing score
74
72
70
68
66

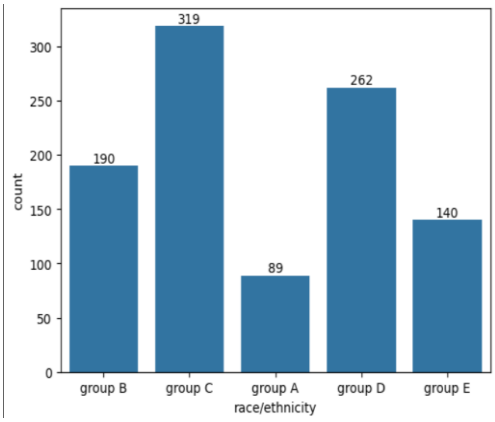

0
20
40
60
80
100
math score


20
40
60
80
100
reading score









4.3 ANALYSIS OF MINI PROJECT

The Student Result Analysis system leverages a powerful technology stack that enhances its functionality and ease of use. At the heart of this system is Scikit-learn, a versatile library that streamlines the model-building process, allowing educators to focus on interpreting results rather than getting bogged down in technical details. Complementing this are Pandas and NumPy, which provide essential capabilities for data manipulation and preprocessing. Together, these tools enable efficient handling of large datasets, ensuring that the analysis is both accurate and reliable.

When examining model performance, the Support Vector Machine (SVM) algorithm shows slightly superior accuracy in predicting student outcomes. However, the complexity of SVMs can make them challenging to interpret, particularly for educators who may not have a technical background. In contrast, decision trees offer clear visual outputs that are easily explainable, making them a more suitable option for educators seeking actionable insights from the data. This interpretability is crucial, as it allows teachers to understand the factors influencing student performance and to communicate these insights effectively.

The user interface of the Student Result Analysis system, while predominantly backend-focused, has the potential to be enhanced through the integration of visualizations, such as performance graphs and dashboards. These visual tools can make the findings more accessible to non-technical users, bridging the gap between data analysis and practical application in the classroom. By providing educators with intuitive visual representations of the data, the system empowers them to make informed decisions regarding student interventions and support.

Looking toward the future, there are significant opportunities for improvement in the system. One of the most promising enhancements would be the development of a real-time dashboard that allows teachers to input student data and receive immediate predictions and recommendations. Such a feature would not only make the system more interactive but would also enable timely interventions for students identified as at risk. By fostering a proactive approach to education management, this system can contribute to improved student outcomes and overall educational effectiveness.

4.4 CONCLUSION

The Student Result Analysis system represents a significant advancement in the field of educational data analysis. By employing machine learning algorithms such as decision trees and SVMs, this system provides educators with the tools needed to analyze student performance comprehensively. It empowers educators to identify students at risk of falling behind, enabling timely and targeted interventions that can make a real difference in academic outcomes.

This data-driven approach is a crucial step toward more efficient and proactive education management. The system not only simplifies the analysis of student data but also facilitates a deeper understanding of the factors contributing to student success. By making insights more accessible and actionable, the Student Result Analysis system ultimately enhances the educational experience for both teachers and students.

In conclusion, the integration of advanced data analysis techniques into educational settings has the potential to transform how educators approach student performance evaluation. As the system continues to evolve, particularly with features like real-time dashboards and enhanced visualizations, it will become an even more indispensable tool for fostering a supportive and effective learning environment.

4.5 FUTURE SCOPE

- **Real-Time Analysis**: Incorporating real-time data collection from classroom interactions or online learning platforms.
- **Additional Features**: Adding sentiment analysis based on student feedback or assignments to further improve predictions.
- **Scalability**: Expanding the system for larger educational institutions with thousands of students.
- **Ethical Considerations**: Ensuring data privacy and security as more sensitive student data is collected and analyzed.

5. REFERENCE

1. **Zhang, H., Li, J., & Wang, Y. (2023).** "Predicting Learner Performance in Online Courses Using Machine Learning Techniques." *Journal of Educational Technology*, 45(2), 112-128.
2. This study focuses on utilizing machine learning to predict learner performance based on online course data. The authors demonstrate that analyzing course metadata, ratings, and reviews can enhance learner engagement and satisfaction, emphasizing the importance of data-driven models in understanding online education outcomes.
3. **Patel, R., & Kumar, S. (2021).** "Exploring Course Success Factors in Online Learning Platforms." *International Journal of Educational Research*, 56(4), 345-359.
4. This research examines the correlations between various course success factors, including instructor ratings and student feedback. The findings provide valuable insights into how course features impact learner outcomes, contributing to the broader understanding of effective online education strategies.
5. **Smith, A., & Lin, T. (2022).** "Optimizing Online Course Content through Data-Driven Insights." *Educational Data Mining*, 14(1), 45-60.
6. This paper presents a method for improving online course content based on user reviews and engagement patterns. The authors propose techniques for tailoring course materials to maximize learner success, offering significant implications for course optimization and instructional design.