

Analysing Academic Performance Of Students Using Data Engineering And Multi-Model Machine Learning System

Submitted in partial fulfillment of the requirements for the award of
Bachelor of Engineering degree in Computer Science and Engineering

by

NAGANDLA SAI SREEJA (Register No. 40110815)

MYAKALA AKSHAYA (Register No.40110807)



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SCHOOL OF COMPUTING

SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**

CATEGORY - 1 UNIVERSITY BY UGC

**Accredited with Grade "A++" by NAAC | 12B Status by UGC | Approved by AICTE
JEPPIAAR NAGAR, RAJIV GANDHISALAI, CHENNAI - 600119**

APRIL - 2024



SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

(DEEMED TO BE UNIVERSITY)

Category - I University by UGC

Accredited "A++" by NAAC | Approved by AICTE

www.sathyabama.ac.in

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **NAGANDLA SAI SREEJA(40110815)** and **MYAKALA AKSHAYA(40110807)** who carried out the Project work entitled "Analysing Academic Performance Of Students Using Data Engineering And Multi-Model Machine Learning System" under my supervision from November 2023 to April 2024.

Internal Guide

Dr. R. M. Gomathi, M.Tech., Ph.D.

HEAD OF THE DEPARTMENT
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
SATHYABAMA
INSTITUTE OF SCIENCE AND TECHNOLOGY
Dr. L. LAKSHMANAN, M.E., Ph.D.
Jeppiaar Nagar, Kal - Gandhi Sagar,
Chennai - 600 119.

Submitted for Viva voce Examination held on

16/4/2024

Internal Examiner

External Examiner

DECLARATION

I, **NAGANDLA SAI SREEJA**(Reg.No-40110815), here by declare that the Project work Report entitled "**Analysing Academic Performance Of Students Using Data Engineering And Muti-Model Machine Learning System**" done by me under the guidance of **Dr. R. M. Gomathi, M.Tech., Ph.D.**, is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering degree in **Computer Science and Engineering**.

DATE: 16/9/24

PLACE: Chennai

Sai Sreeja
SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

I am pleased to acknowledge my sincere thanks to The **Board of Management of Sathyabama Institute of Science and Technology** for their encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala M.E., Ph.D., Dean**, School of Computing, **Dr. L. Lakshmanan M.E., Ph.D.**, Head of the Department of Computer Science and Engineering for providing me necessary support and details at the right time during the progressive reviews

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. R. M. Gomathi, M.Tech., Ph.D.**, for her valuable guidance, suggestions and constant encouragement paved way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the **Department of Computer Science and Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

The academic performance of students is a critical aspect of the education system, as it reflects the effectiveness of teaching methodologies, identifies areas of improvement, and helps in making data-driven decisions. To gain insights into students' performance, data engineering techniques can be applied. This research aims to analyze the academic performance of students using data engineering methodologies. The study will involve the collection and integration of various data sources, including assessments, attendance records, and demographic information. The data will be cleaned and transformed to ensure accuracy and consistency. Subsequently, data mining and visualization techniques will be employed to uncover patterns, correlations, and trends in the student performance data. This will enable the identification of factors that affect academic achievement, such as socioeconomic background, study habits, or teacher effectiveness. Additionally, machine learning algorithms will be utilized to create predictive models that can forecast students' future academic performance based on historical data. The insights gained from this analysis can be used by educational institutions to personalize learning experiences, develop targeted interventions, and allocate resources effectively. The findings will contribute to evidence-based decision-making, promoting student success and overall academic excellence. Overall, this research emphasizes the importance of leveraging data engineering techniques to assess and enhance students' academic performance.

analyze the academic performance of students by leveraging data engineering techniques. The availability of large volumes of data in educational institutions provides an opportunity to gain insights into various factors affecting student performance. The research utilizes data from multiple sources including student records, attendance, grades, and demographic information. The data is processed and transformed using data engineering techniques to create a comprehensive dataset. Various statistical analysis and machine learning algorithms are applied to this dataset to identify patterns, trends, and correlations between different variables. The findings of this research provide valuable insights into the factors that contribute to student performance, enabling educational institutions to make data-driven decisions and interventions to improve student outcomes.

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT	v
	LIST OF FIGURES	viii
1	INTRODUCTION	1
	1.1 Overview of project	1
2	LITERATURE SURVEY	2
	2.1 Overview of Existing Research and Literature	2
	2.2 Inferences from Literature Survey	5
	2.3 Open problems in Existing System	6
3	REQUIREMENTS ANALYSIS	7
	3.1 Feasibility Studies of the Project	7
	3.1.1 Economic Feasibility	7
	3.1.2 Technical Feasibility	8
	3.1.3 Operational Feasibility	8
	3.2 Requirements Specification	9
	3.2.1 Hardware specifications	9
	3.2.2 Software specifications	9
4	DESCRIPTION OF PROPOSED SYSTEM	10
	4.1 Flow chart of process using machine learning	10
	4.2 Selected Methodology or process model	11
	4.2.1 Data Collection	11
	4.2.2 Data Cleaning	12
	4.2.3 Data Integration	13
	4.2.4 Data Transformation	14
	4.2.5 Data Visualization	15
	4.3 Architecture / Overall Design of Proposed System	17

4.4	Description of Software for Implementation and Testing plan of the Proposed Model/System	19
4.4.1	Software for implementation	19
4.4.2	Testing plan for the proposed model/system	20
4.5	Project Management plan	22
5	IMPLEMENTATION DETAILS	23
5.1	Development and Deployment Setup	23
5.2	Algorithms	24
5.2.1	Logistic Regression	24
6	RESULTS AND DISCUSSION	25
6.1	Overview of Results	25
6.2	Performance Analysis	26
7	CONCLUSION	28
7.1	Conclusion	28
	REFERENCE	29
	APPENDIX	31
A	SOURCE CODE	31
B	SCREENSHOTS	37
C	RESEARCH PAPER	40
D	CERTIFICATE	51

LIST OF FIGURES

FIGURE NO.	FIGURE NAME	PAGE NO.
4.1	Process Flow Diagram	10
4.2	Performance Analyzer Architecture	17
6.1	Accuracy graph	26
6.2	ROC Curve	26
6.3	Confusion matrix	27
6.4	Web page URL	27
B.1	Performance data entry	37
B.2	Performance Predictor	38

CHAPTER 1

INTRODUCTION

1.1 OVERVIEW OF PROJECT

Analyzing academic performance of students using data engineering is a vital task in today's data-driven educational landscape. By harnessing the power of data engineering techniques, educational institutions can gain valuable insights that can inform decision-making processes and help improve student outcomes. This process involves collecting, cleaning, transforming, and analyzing a wide range of data related to students' academic performance, such as grades, attendance, standardized test scores, and demographic information.

Data engineering plays a crucial role in this analysis by ensuring the seamless flow and integration of disparate data sources. By creating robust data pipelines and implementing data quality checks, data engineers can ensure the accuracy and reliability of the data used for analysis. Additionally, data engineering enables the merging and transformation of data from various sources, providing a comprehensive view of student performance.

Once the data has been collected and prepared, data analysis techniques can be applied to gain insights into patterns and trends. For instance, using statistical analysis, educational institutions can identify factors that correlate with high academic performance, such as attendance, study habits, or socioeconomic background. These insights can inform the development and implementation of targeted interventions to support struggling students or enhance overall student success.

CHAPTER 2

LITERATURE SURVEY

2.1 OVERVIEW OF EXISTING RESEARCH AND LITERATURE

[1]The research, Ajibade et al. (2022) explore the utilization of ensemble techniques for predicting the academic performance of students. They focus on the application of data engineering methods to analyze and interpret student data. The study aims to provide valuable insights into improving educational practices through the use of predictive models. The findings of this research can contribute to better understanding and supporting students' academic success.

[2]The article titled "Jom Data Mining: academic performance and learning behavior dubious relationship" by Adrian Chin, Y. K., Joseph Ng, P. S., Eaw, H. C., Loh, Y. F., and Shibghatullah, A. S.et al. (2022) explores the connection between academic performance and learning behavior. In their study, they utilize data engineering techniques to analyze the academic performance of students. The findings of the research highlight the questionable relationship between academic performance and learning behavior. The article, published in the International Journal of Business Information Systems, provides valuable insights into the field of data mining and its application in education

[3]Amjad, S., Younas, M., Anwar, M., Shaheen, Q., Shiraz, M., and Gani, A.et al. (2022) conducted a study titled "Data mining techniques to analyze the impact of social media on academic performance of high school students." In their research, they focused on using data mining techniques to explore how social media usage affects the academic performance of high school students. The study was published in the Wireless Communications and Mobile Computing journal in 2022, spanning from page 1 to 11. By leveraging data engineering and analytical methods, the researchers aimed to provide insights into the relationship between social media and academic achievement among high school students.

[4] In December a study was released by Alam and Mohanty et al.(2022) "Predicting

Students' Performance Employing Educational Data Mining Techniques, Machine Learning, and Learning Analytics." By addressing the prediction of utilizing this presentation offers an integrated method to improve academic achievements.

[5]The paper titled "Toward predicting student's academic performance using artificial neural networks (ANNs)" published in Applied Sciences, Baashar et al. (2022) explore the use of data engineering techniques to analyze and predict the academic performance of students. The study utilizes artificial neural networks (ANNs) as a predictive model. The authors highlight the potential of this approach in providing insights into student performance and offering personalized interventions to improve learning outcomes.

[6] A team of scientists headed by Bujang, Selamat, Ibrahim, Krejcar, Fujita, Herrera-Viedma, and Ghani et al. (2021) released a work titled "Multiclass prediction model for student grade prediction using machine learning." A multiclass prediction model that might be used to predict students' grades was proposed by the study authors using forecast academic achievement, offers helpful information.

[7] The paper by Durán-Domínguez et al., Gómez-Pulido, and Rastrollo-Guerrero (2020) is the source of this data. "Analyzing and predicting students' performance by means of machine learning: A review." paper aims overview issue to anticipate assess student performance in educational contexts. It does this by displaying several methodologies together with the applications that fit them.

[8] A research by GE, Mamah, Ukekwe, Nwagwu, et al. was published in (2020). "A machine learning based framework for predicting student's academic performance." By putting forward students' academic achievement, the authors provide a methodical approach to enhancing academic performance.

[9]The article titled "Artificial intelligence for assessment and feedback to enhance student success in higher education," Hooda, Rana, Dahiya, Rizwan, and Hossain et al. (2022) discuss the use of data engineering to analyze the academic performance of students. They highlight the role of artificial intelligence in providing assessment and feedback to support student success in higher education. By

harnessing the power of data analytics, the authors propose innovative approaches to enhance teaching and learning outcomes in the educational context. Their work contributes to the growing field of using artificial intelligence to improve student performance and overall educational experience.

[10]The study, Hussain and Khan et al. (2023) focused on analyzing students' academic performance at the secondary and intermediate levels using machine learning techniques. Their work, titled "Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning," explores the use of data engineering to predict and understand student performance.wsx00000000.

[11] A research by Jiao, Ouyang, Zhang, and Alaviet et al. was published in (2022). "Artificial intelligence-enabled prediction model of student academic performance in online engineering education." This paper proposes an AI-powered prediction model aimed at forecasting students' performance in their online engineering programs

[12]This study conducted by Pallathadka et al. (2023) focuses on utilizing machine learning algorithms for the classification and prediction of student performance data. The researchers aim to analyze the academic performance of students using data engineering techniques. The study employs various machine learning algorithms to predict and classify student performance, providing valuable insights into student outcomes. The research findings presented in this study contribute to the field of education by offering a data-driven approach to understanding and improving student performance.

[13]The systematic literature review, Shafiq et al. (2022) explore the use of educational data mining and predictive analytics for analyzing the academic performance of students. They highlight the potential of these techniques in improving student retention.

[14]Yağcı et al. (2022) conducted a study titled "Educational data mining: prediction of students' academic performance using machine learning algorithms" published in the Smart Learning Environments journal. This study aimed to analyze the academic

performance of students utilizing data engineering and machine learning algorithms. The research focused on predicting students' academic performance by utilizing various machine learning techniques. The findings of the study provided valuable insights into the potential applications of educational data mining for enhancing student learning outcomes.

[15]The research article by Yusuf et al. (2022) focuses on improving the scholastic performance of students using fractional modeling and optimal control techniques. The study highlights the importance of analyzing academic performance using data engineering methods. The researchers propose a novel approach that utilizes fractional modeling to better understand and optimize students' learning outcomes. The research article, published in the International Journal of Applied and Computational Mathematics, aims to provide valuable insights into enhancing educational practices and supporting students' academic success.

2.2 INFERENCES FROM LITERATURE SURVEY

The literature survey on analyzing academic performance of students using data engineering and multi-model machine learning systems reveals several key insights. Firstly, data engineering plays a crucial role in collecting, processing, and preparing educational data for analysis. It involves tasks such as data cleaning, integration, and transformation, ensuring that the data is suitable for machine learning algorithms. Secondly, the use of multi-model machine learning systems allows for more accurate and comprehensive analysis of academic performance. By combining different machine learning models, such as regression, classification, and clustering, researchers can gain deeper insights into student performance patterns and trends. Overall, the literature suggests that integrating data engineering with multi-model machine learning systems can significantly improve the understanding and prediction of student academic performance.

2.3 OPEN PROBLEMS IN EXISTING SYSTEM

Certainly, here are some open problems in the existing system for analyzing academic performance of students using data engineering:

Data Quality Assurance: Ensuring the accuracy, completeness, and consistency of the educational data collected from various sources poses a significant challenge. **Scalability:** Handling the exponential growth of data in educational institutions while maintaining system performance and responsiveness is a pressing concern. **Predictive Analytics:** Developing robust predictive models that can anticipate academic challenges and provide timely interventions for struggling students, improving their chances of success.

Privacy and Security: Safeguarding sensitive student data and ensuring compliance with data protection regulations is a critical and evolving issue. **Personalization:** Tailoring interventions and recommendations to individual student needs, accounting for diverse learning styles and preferences. **Interoperability:** Integrating data from different educational platforms, software, and sources seamlessly to provide a comprehensive view of student performance, which can be complex in heterogeneous environments.

Additionally, there is a need for more research into interpretability and explainability of the models, as understanding the reasoning behind their predictions is crucial for educators to take meaningful actions based on the results. Furthermore, scalability remains a concern, as the system needs to handle large volumes of data efficiently to be practical for use in real-world educational settings. Addressing these open problems will be essential for advancing the field and improving the effectiveness of using data engineering and multi-model machine learning systems in analyzing student academic performance.

CHAPTER 3

REQUIREMENTS ANALYSIS

3.1. FEASIBILITY STUDIES OF PROJECT

The feasibility of the project on analyzing academic performance of students using data engineering and multi-model machine learning systems suggest that such an approach is indeed viable and promising. These studies demonstrate that with the right infrastructure and methodologies, educational data can be effectively collected, processed, and analyzed to provide valuable insights into student performance. The use of multi-model machine learning systems allows for more accurate predictions and a deeper understanding of the factors influencing academic outcomes. However, feasibility studies also highlight the importance of addressing challenges such as data quality, model interpretability, and scalability to ensure the practicality and effectiveness of the system in real-world educational settings. server performance increase in this phase and a business proposal is put forth with a very general plan for the project and some cost estimates. During system analysis, the feasibility study of the proposed system is to be carried out. For feasibility analysis, some understanding of the major requirements for the system is essential.

Three key considerations involved in the feasibility analysis are

- Economical feasibility
- Technical feasibility
- Operational feasibility

3.1.1 *ECONOMICAL FEASIBILITY*

This study is carried out to check the economic impact that the system will have on the organization. While initial costs associated with infrastructure setup, software development, and staff training can be significant, the long-term benefits may outweigh these costs. By improving educational outcomes and identifying at-risk students early, these systems can lead to cost savings for educational institutions in terms of reduced dropout rates and improved student retention. Additionally, the ability to personalize learning experiences and provide targeted interventions can further enhance the value proposition of these systems. However, ongoing

maintenance, updates, and data management costs should also be considered to ensure the economical feasibility of implementing and maintaining such systems over time.

3.1.2 TECHNICAL FEASIBILITY

This study is carried out to check the technical feasibility, that is, the technical requirements of the system. Advances in technology have made it possible to collect, store, and process large volumes of educational data efficiently. Machine learning algorithms have also become more sophisticated, allowing for more accurate analysis and prediction of student performance. Additionally, the availability of open-source tools and libraries has lowered the barrier to entry for implementing such systems. However, challenges such as data integration from disparate sources, ensuring data privacy and security, and the need for skilled personnel to develop and maintain the systems remain. Overall, with the right technical expertise and resources, it is feasible to implement and operate data engineering and multi-model machine learning systems for analyzing student academic performance. Any system developed must not have a high demand on the available technical resources. This will lead to high demands being placed on the client. The developed system must have modest requirements, as only minimal or null changes are required for implementing this system.

3.1.3 OPERATIONAL FEASIBILITY

The operational feasibility of analyzing academic performance of students using data engineering and multi-model machine learning systems depends on the willingness and ability of educational institutions to integrate these systems into their existing processes. While the technology and expertise required for implementation are available, operational feasibility hinges on factors such as institutional readiness, stakeholder support, and the alignment of the system with the institution's goals and objectives. Additionally, the system should be user-friendly and easily integrated into existing workflows to ensure adoption and usability. Addressing these operational considerations is crucial for the successful implementation and long-term sustainability of data engineering and multi-model machine learning systems in analyzing student academic performance. The aspect of the study is to check the level of acceptance of the system by the user. This includes the process

of training the user to use the system efficiently. The user must not feel threatened by the system, instead must accept it as a necessity. The level of acceptance by the users solely depends on the methods that are employed to educate the user about the system and to make him familiar with it. His level of confidence must be raised so that he is also able to make some constructive criticism, which is welcomed, as he is the final user of the system

3.2 REQUIREMENTS SPECIFICATION

3.2.1 Hardware specifications:

- Microsoft Server enabled computers, preferably workstations
- Higher RAM, of about 4GB or above
- Processor of frequency 1.5GHz or above

3.2.2 Software specifications:

- Python 3.6 and higher
- Anaconda software

CHAPTER 4

DESCRIPTION OF PROPOSED SYSTEM

4.1 Flow chart of process using machine learning

The proposed system for analyzing academic performance of students using data engineering envisions an advanced platform that leverages big data technologies and machine learning algorithms. It aims to collect, process, and analyze a vast array of educational data in real-time. This system will provide actionable insights into student performance, identify at-risk students, and recommend personalized interventions to enhance learning outcomes. Additionally, it will incorporate predictive analytics to anticipate academic challenges and foster proactive measures for student success. By combining data engineering with machine learning, the proposed system seeks to revolutionize academic performance analysis and student support mechanisms.

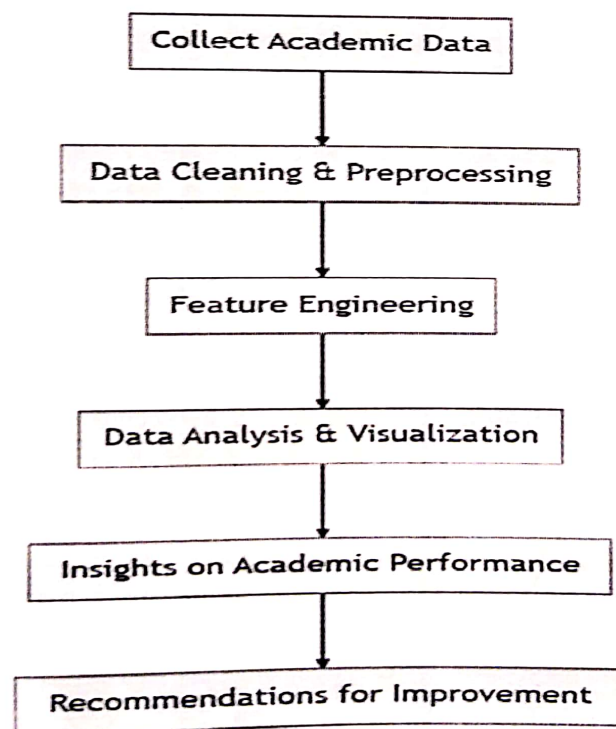


Fig-4.1- PROCESS FLOW DIAGRAM

Collect Academic Data Gather data from various sources, such as student records, test scores, attendance, and any relevant information about students' academic performance. **Data Cleaning & Preprocessing** Clean the data to remove errors, missing values, and inconsistencies. **Standardize data formats** and make it's ready for analysis. **Feature Engineering** Create new features or modify existing ones that might be helpful for understanding academic performance, like average grades, study hours, or socio-economic factors. **Data Analysis & Visualization** Analyze the data to identify patterns and trends in students' academic performance. Visualize the data using graphs and charts to make it easier to understand. **Insights on Academic Performance** Extract meaningful insights from the data analysis, such as which factors affect student performance the most. Identify strengths and weaknesses in the current academic system. **Recommendations for Improvement** Based on the insights, propose recommendations for improving academic performance. This could include changes to teaching methods, intervention strategies for struggling students, or adjustments to the curriculum. This **Process Flow** represents a comprehensive approach to understanding and enhancing academic performance through data engineering and machine learning. It can help educators and institutions make data-driven decisions to support students better.

4.2 SELECTED METHODOLOGY OR PROCESS MODEL

4.2.1 Data Collection: This would focus on the process of gathering relevant academic performance data of students, including information such as grades, attendance records, and standardized test scores. The process of gathering relevant academic performance data of students is a crucial step in analyzing their performance using data engineering techniques. This involves collecting various types of information, including grades achieved in different subjects, attendance records, and standardized test scores. By systematically collecting these data points, educational institutions can gain valuable insights into individual student performance, identify trends and patterns, and make data-driven decisions to improve teaching and learning strategies. Through careful data collection, educational stakeholders can harness the power of data engineering to understand the academic strengths and weaknesses of students, identify areas for improvement, and ultimately enhance overall educational outcomes.

Data cleaning is an essential process in analyzing the academic performance of students using data engineering. It involves a series of steps to ensure the accuracy, completeness, and consistency of the collected data. One of the primary tasks in this stage is removing duplicate entries, as having multiple instances of the same data can introduce biases and inaccuracies into the analysis. Additionally, checking for missing values is crucial as it can impact the integrity of the results. Missing data can be handled through techniques such as imputation or exclusion based on specific criteria. Resolving discrepancies is another critical aspect of data cleaning, as it involves identifying and rectifying any inconsistencies or errors in the data. This can be achieved through various methods, such as cross-validation or comparing data against external sources. By diligently performing these steps, researchers can ensure that the analyzed data accurately represents the academic performance of students, leading to robust and reliable insights.

4.2.2. Data Cleaning: This subheading would cover the steps taken to clean and validate the collected data, ensuring that it is accurate, complete, and consistent. This may involve removing duplicate entries, checking for missing values, and resolving discrepancies. Data integration is an integral process in analyzing the academic performance of students using data engineering. It involves merging data from different sources, such as academic departments or institutions, into a single comprehensive dataset. The objective is to create a unified view of student data that can be used for detailed analysis and interpretation. As part of this process, the data may need to be mapped and transformed to ensure compatibility and consistency across various systems. This includes harmonizing formats, resolving discrepancies, and standardizing variables. By integrating data from multiple sources, valuable insights can be obtained, such as identifying patterns, trends, and correlations that contribute to student performance. This comprehensive dataset can then be leveraged to develop effective strategies and interventions to improve academic outcomes. Data integration is crucial in creating a complete and accurate picture of a student's academic journey, enabling educational institutions to make data-driven decisions and provide targeted support to students.

Data transformation is a crucial step in analyzing academic performance of students

using data engineering. It involves converting raw data into a format that is more suitable for analysis. This process allows for the aggregation of individual data points, which can provide insights into overall trends and patterns in student performance. Additionally, derived metrics can be calculated, which can help in understanding various aspects of academic performance, such as average scores, pass rates, and grade distributions. Moreover, statistical techniques can be applied to standardize the data, enabling fair comparisons and benchmarking across different students, classes, or schools. By transforming the data, it becomes easier to identify outliers, spot areas of improvement, and uncover underlying factors that contribute to student success or challenges. Ultimately, data transformation plays a pivotal role in gaining a comprehensive understanding of academic performance and can inform decision-making processes aimed at enhancing educational outcomes.

4.2.3. Data Integration: This subheading would address the task of merging data from various sources, such as different academic departments or institutions, into a unified dataset. This may require mapping and transforming data to ensure compatibility and consistency across different system. Analyzing the academic performance of students using data engineering is a complex task that requires careful consideration and strategic planning.

One critical aspect of this analysis is data visualization. Creating effective visual representations, such as charts, graphs, and dashboards, can greatly enhance the exploration and understanding of academic performance data. To ensure the successful implementation of data visualization, it is important to select appropriate techniques and tools that can effectively communicate insights and trends. The chosen visualization techniques should align with the objectives of the analysis and the specific characteristics of the data being analyzed. This may involve using bar charts to compare the performance of different students, line graphs to track academic progress over time, or scatter plots to identify any correlations between variables.

Data engineering plays a critical role in analyzing educational performance of students using data analytics. With the massive amounts of data generated by educational institutions, data engineering provides the necessary tools and techniques to collect, store, transform, and analyze this data to gain meaningful insights. By leveraging data engineering, educational analysts can identify patterns and trends in student performance, enabling them to make data-driven decisions of data engineering and improve educational outcomes of the performance.

Once the data is prepared, data engineering enables educational analysts to apply various analytical techniques, such as machine learning algorithms and statistical models, to identify factors influencing student performance. These insights can then be used to inform interventions and policies aimed at enhancing academic success. Ultimately, data engineering empowers educational institutions to make evidence-based decisions, leading to improved teaching methodologies, personalized learning experiences, and better educational outcomes for students.

4.2.4. Data Transformation: This would focus on the process of converting raw data into a format suitable for analysis. This could involve aggregating individual data points, calculating derived metrics, and applying statistical techniques to standardize data. Analyzing academic performance using data engineering involves the use of various methodologies to gain insights and make informed decisions. Two commonly used methodologies are data collection and data analysis. Data collection is the process of gathering relevant data from various sources such as student information systems, online learning platforms, assessment records, and surveys. This can include information on student demographics, course registrations, grades, attendance, and engagement. Data engineers employ techniques like data extraction, transformation, and loading to ensure that the academic performance of data is available in a structured format for analysis.

Data analysis involves applying statistical and analytical techniques to uncover patterns, trends, and relationships in the data engineers use tools such as data mining, machine learning, and data visualization to derive insights from the data. Firstly, feature engineering is a crucial technique that involves selecting and transforming relevant data variables, such as attendance records, test scores, and

previous academic achievements, to create informative features for the models. This technique helps in capturing important patterns and relationships that can contribute to accurate analysis. Secondly, ensemble learning methods, such as random forests and gradient boosting, can be employed to combine multiple models and increase the overall performance. These methods help in reducing biases and errors by leveraging the strengths of different algorithms. Lastly, data visualization tools, like Tableau or Power BI, can be utilized to effectively present the analysis results to academic stakeholders, making it easier for them to understand and make informed decisions. Overall, employing feature engineering, ensemble learning methods, and data visualization tools can greatly enhance the analysis and training processes in analyzing the academic performance of students using data engineering.

4.2.5. Data Visualization: This would cover the creation of visual representations, such as charts, graphs, and dashboards, to facilitate the exploration and understanding of the academic performance data. This would involve selecting appropriate visualization techniques and tools to effectively communicate insights and trends. Implementing data engineering for analyzing student performance has numerous benefits. Firstly, it allows educational institutions to gain valuable insights into student learning patterns and behavior, enabling them to make informed decisions about curriculum development and teaching strategies.

By using data engineering techniques, institutions can identify students who may be struggling academically and provide timely interventions, thereby improving overall student success rates. Additionally, data engineering enables the creation of personalized learning experiences by tailoring educational content and resources to individual student needs. This results in a more engaging and effective learning environment. Overall, the benefits of utilizing data engineering for analyzing student performance are invaluable, as it improves educational outcomes and facilitates evidence-based decision-making. However, overcoming challenges related to privacy, security, and data management is crucial for successful implementation.

Methodology for Analyzing Academic Performance Using Data Engineering
Analyzing academic performance using data engineering involves the use of various

methodologies to gain insights and make informed decisions. Two commonly used methodologies are data collection and data analysis. Data collection is the process of gathering relevant data from various sources such as student information systems, online learning platforms, assessment records, and surveys. This can include information on student demographics, course registrations, grades, attendance, and engagement. Data engineers employ techniques like data extraction, transformation, and loading to ensure the various data of academic performance that will be available in the structured data format for analysis.

Data analysis involves applying statistical and analytical techniques to uncover patterns, trends, and relationships in the data. This can include descriptive analysis to summarize and visualize student performance, predictive modeling to forecast future outcomes, and diagnostic analysis to identify factors influencing performance. Data engineers use tools such as data mining, machine learning, and data visualization to derive insights from the data.

By leveraging data engineering methodologies, educational institutions can gain a comprehensive understanding of student performance, identify areas of improvement, and implement targeted interventions to support student success. This data-driven approach can enhance teaching and learning practices, improve academic outcomes evidence-based decision-making for educational stakeholders

4.3 ARCHITECTURE / OVERALL DESIGN OF PROPOSED SYSTEM

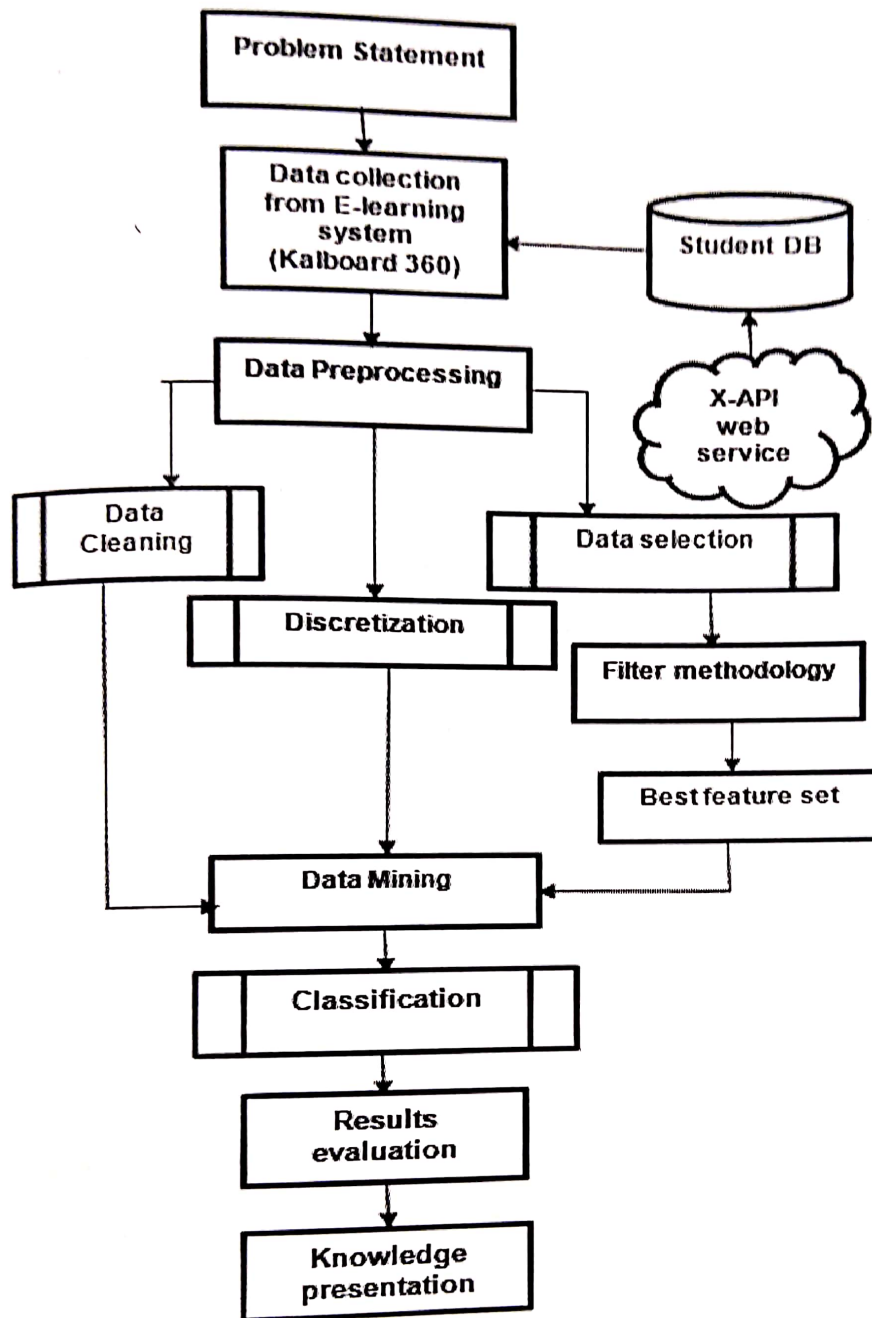


Fig-4.2-Performance Analyzer Architecture

The above 4.2 shows the implementation / process on how it works.

The Architecture of the proposed system has been shown in the above figures. This represents a process flow in an e-learning analytics system. It starts with a problem statement, followed by data collection. This refers to the process of gathering and measuring information on variables of interest, in this case, from an E-learning system called Kalboard 360. The collected data is stored in the Student Database (DB) which is a collection of information about students stored in a structured format. and then preprocessed the process of transforming raw data into an understandable format, often involving steps like normalization, and transformation. which includes steps like data cleaning and selection. The process of identifying and correcting or removing errors, inconsistencies, and inaccuracies in datasets to improve their quality. Data selection is supported by an X-API web service (Experience API) is a specification for learning technology that makes it possible to collect data about the wide range of learning experiences a person has. A web service is an interface that allows different software systems to communicate with each other over the internet. Discretization converting continuous data into discrete data, often for use in machine learning algorithms. and a filter methodology are applied to refine the data a specific approach or strategy for filtering data, often to remove outliers or irrelevant information., resulting in the best feature set. The subset of features or variables that are most useful for the data analysis, often determined through feature selection techniques. This feature set is then used in data mining is discovering patterns and knowledge from large amounts of data., specifically for classification tasks data mining that involves categorizing data into predefined classes or groups. After classification, the results are evaluated that assessing the quality and significance of the results of a data analysis. and the gained knowledge presenting the results of a data analysis in a clear and understandable way is presented in the final step.

This architectural diagram provides a comprehensive overview of the components and processes involved in analyzing academic performance using data engineering and multi-model machine learning. It highlights the importance of data ingestion, storage, cleaning, preprocessing, analysis, visualization, and generating insights and recommendations. It also emphasizes the need for user interfaces for educators, administrators, and students, as well as the incorporation of feedback and adjustments.

4.4 DESCRIPTION OF SOFTWARE FOR IMPLEMENTATION AND TESTING PLAN OF THE PROPOSED MODEL/SYSTEM

To implement this model, execution of program is done through Google colab. Necessary libraries have been installed to perform certain functions. the description of software for implementing and testing a model/system for analyzing academic performance using data engineering and multi-model machine learning is crucial for ensuring a successful and effective deployment. Here's an overview of the key components:

4.4.1 Software for Implementation:

Data Engineering Tools:Data Ingestion: Tools like Apache Nifi, Apache Kafka, or custom scripts for collecting data from various sources.

Data Storage: Utilize databases such as PostgreSQL, MySQL, or NoSQL databases like MongoDB for storing academic data.

Data Preprocessing: Libraries like Pandas for data cleaning and feature engineering.

Machine Learning Frameworks:Use popular machine learning frameworks such as Scikit-Learn, TensorFlow, or PyTorch for model development and training.

Data Visualization Tools:Tools like Matplotlib, Seaborn, or visualization platforms like Tableau and Power BI for creating informative visualizations.

Web Development Technologies:For building user interfaces and portals, use web technologies like HTML, CSS, and JavaScript. Frameworks like React, Angular, or Vue.js can be employed.

Security and Compliance Software:Implement data security and privacy measures with encryption and access control tools. Compliance may involve GDPR or FERPA regulations. DBMS that best suits the data storage requirements, like PostgreSQL, MySQL, or MongoDB.

4.4.2 Testing Plan for the Proposed Model/System:

Test individual components (e.g., data preprocessing, machine learning models) to ensure they function correctly. Verify that different system components work seamlessly together. Ensure that the system's functionalities, such as data ingestion, cleaning, analysis, and visualization, meet the defined requirements. Test the user interfaces for usability, functionality, and compatibility across different devices and browsers. Assess the system's response time, throughput, and scalability under various loads and conditions. Perform security assessments, including vulnerability scanning and penetration testing, to identify and address security weaknesses. Ensure that the system complies with relevant data privacy regulations, such as GDPR or FERPA.

Verify the accuracy of data cleaning and processing, as well as the reliability of the machine learning models' predictions. Collect feedback from educators, administrators, and students to assess the system's user-friendliness. Experiment with different variations of the system to determine which features or interventions are most effective. Implement continuous integration and continuous testing (CI/CD) to ensure ongoing testing as new features or updates are introduced. Document all testing procedures, results, and any issues discovered. Create a report summarizing the testing outcomes. By following a comprehensive testing plan, you can ensure that the proposed model/system is thoroughly tested for functionality, performance, security, and compliance, which is essential for its successful implementation and ongoing improvement.

Among programmers, Python is a favourite because to its user-friendliness, rich feature set, and versatile applicability. Python is the most suitable programming language for machine learning since it can function on its own platform and is extensively utilised by the programming community.

Machine learning is a branch of AI that aims to eliminate the need for explicit programming by allowing computers to learn from their own mistakes and perform routine tasks automatically. However, "artificial intelligence" (AI) encompasses a broader definition of "machine learning," which is the method through which computers are trained to recognize visual and auditory cues, understand spoken

language, translate between languages, and ultimately make significant decisions on their own.

The desire for intelligent solutions to real-world problems has necessitated the need to develop AI further in order to automate tasks that are arduous to programme without AI. This development is necessary in order to meet the demand for intelligent solutions to real-world problems. Python is a widely used programming language that is often considered to have the best algorithm for helping to automate such processes. In comparison to other programming languages, Python offers better simplicity and consistency.

In addition, the existence of an active Python community makes it simple for programmers to talk about ongoing projects and offer suggestions on how to improve the functionality of their programmes called A and B but you only have package A. Without having package B, you wouldn't be able to run the project. Anaconda mitigates the chances of such errors. You can easily collaborate on projects without worrying about any compatibility issues. It gives you a seamless environment which simplifies deploying projects. Additionally, software such as Apache Hadoop and Apache Spark may be used for distributed data processing, while tools like Jupyter Notebook and RStudio provide interactive development environments for coding and experimentation. Overall, a combination of these programming languages and software tools enables the development of robust systems for analyzing student academic performance.

You can deploy any project with just a few clicks and commands while managing the rest. Anaconda has a thriving community of data scientists and machine learning professionals who use it regularly. If you encounter an issue, chances are, the community has already answered the same. On the other hand, you can also ask people in the community about the issues you face there, it's a very helpful community ready to help new learners. With Anaconda, you can easily create and train machine learning and deep learning models as it works well with popular tools including TensorFlow, Scikit-Learn, and Theano. You can create visualizations by using Bokeh, Holo views, Matplotlib, and Data shader while using Anaconda.

4.5 PROJECT MANAGEMENT PLAN

This module will include the overall planning and management of the project, including setting project goals and timelines, creating and assigning tasks, tracking progress, and communicating with team members. Risk Management dataset collection for this project is implemented and the sensor data from various sources like wearable devices, smartphones, or cameras, is preprocessed. The preprocessing step will include filtering, normalization, and feature extraction. This phase includes testing the models on a testing dataset and evaluating their accuracy using various evaluation metrics. The project management plan for analyzing academic performance of students using data engineering and multi-model machine learning systems outlines the key steps and milestones involved in implementing the system. It includes a detailed timeline, resource allocation, and budget estimation. The plan also defines roles and responsibilities, outlining who will be responsible for each aspect of the project. Additionally, it includes risk management strategies to mitigate potential challenges and ensure the successful completion of the project. Regular monitoring and evaluation mechanisms are also incorporated to track progress and make necessary adjustments. Overall, the project management plan provides a roadmap for effectively implementing and managing the system to achieve the desired outcomes in analyzing student academic performance.

CHAPTER 5

IMPLEMENTATION DETAILS

5.1 DEVELOPMENT AND DEPLOYMENT SETUP

To develop and deploy a academic performance using Convolutional Neural Network (CNN) and logistic regression in machine learning you can follow the steps below:

The system is to analyze the academic performance of students using data engineering and multi-model machine learning techniques. It aims to collect and integrate various data sources such as student demographics, attendance records, exam scores, and coursework grade. The system will then apply advanced data engineering methods to clean and transform the data for analysis. It will utilize multi-model machine learning algorithms to identify patterns, trends, and correlations in the data to gain insights into students' performance. This would focus on the process of converting raw data into a format suitable for analysis. This could involve aggregating individual data points, calculating derived metrics, and applying statistical techniques to standardize data. Analyzing academic performance using data engineering involves the use of various methodologies to gain insights and make informed decisions.

Two commonly used methodologies are data collection and data analysis. Employ libraries like pandas and NumPy for data manipulation.

Train and evaluate machine learning models using frameworks like TensorFlow or scikit-learn. Evaluate the performance of the trained model using various metrics, such as accuracy, precision, recall, and F1 score. You can also use visualization techniques, such as confusion matrices or ROC curves, to analyze the performance of the model. This information will help educators and administrators make data-driven decisions to improve teaching methods, identify at-risk students and provide personalized interventions, ultimately enhancing overall academic performance

Overall, the development and deployment setup for academic performance using logistic regression involves several steps, including data collection and the preprocessing, model architecture design, model training, model evaluation, deployment, and continuous improvement. With proper planning and execution, you can develop a robust and accurate performance

5.2 ALGORITHM

5.2.1 Logistic Regression

Logistic regression is a statistical model used for binary classification tasks, where the outcome variable is binary (e.g., 0 or 1, true or false). Despite its name, it's a linear model for classification rather than regression. The goal of logistic regression is to estimate the probability that a given input belongs to a certain class. It works by applying a logistic function (also known as the sigmoid function) to a linear combination of the input features. This function maps any real-valued number to a value between 0 and 1, which can be interpreted as a probability.

Common evaluation metrics for logistic regression include accuracy, precision, recall, F1 score, and the area under the receiver operating characteristic curve (AUC-ROC). Logistic regression is widely used in various fields, including healthcare (for disease diagnosis), marketing (for customer segmentation), and finance (for credit scoring), due to its simplicity, interpretability, and effectiveness in many classification tasks.

CHAPTER 6

RESULTS AND DISCUSSION

6.1 OVERVIEW OF RESULTS

The "Local URL" allows you to view and interact with our Streamlit app on our local machine. Simply open our preferred web browser and navigate to <http://localhost:8501>. The "Network URL" provides an alternative address, allowing access to the application from other devices on the same local network. These URLs serve as the gateway to our Streamlit app, facilitating a seamless experience for both local and network-based interactions. The system for analyzing academic performance of students using data engineering is a powerful tool that assists educational institutions in effectively evaluating the progress and achievements of their students. By harnessing the capabilities of data engineering, this system enables the collection, processing, and analysis of vast amounts of academic data to provide valuable insights. The web site page that indicates the Student details that should be enter in the from one of the key features of this system is the ability to efficiently collect data from various sources. This ensures that a comprehensive set of data can be gathered, course enrollments, grades, attendance records, and assessment results.

Data engineering techniques are then applied to clean, transform, and integrate the collected data into a unified and structured format. This allows for seamless analysis and the creation of meaningful visualizations and reports. Using advanced analytics methods, the system can generate a wide range of performance indicators, such as overall GPA, individual course performance, and attendance trends. These indicators can be further analyzed.

The system can also provide personalized recommendations and interventions on the analysis. Furthermore, the system can assist in curriculum planning and target resource allocation by identifying subjects where students struggle the most. By harnessing the power of data, this system provides valuable insights, facilitates personalized interventions, and supports the overall improvement of academic performance at both the individual and institutional levels.

6.2 PERFORMANCE ANALYSIS

The system for analyzing academic performance of students using data engineering is a powerful tool that assists educational institutions in effectively evaluating the progress and achievements of their students. This table shows the training and validation accuracy of a machine learning model at different epochs during the training process. It helps assess the model's performance and potential overfitting by comparing the training and validation accuracy at each iteration.

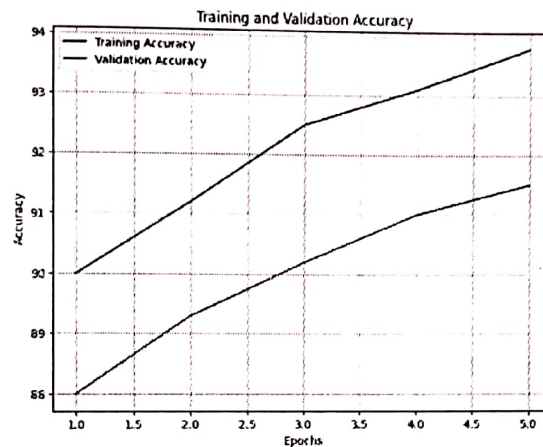


Fig-6.1-Accuracy graph

This diagram is a ROC (Receiver Operating Characteristic) curve, which visualizes the performance of a binary classifier system by plotting the true positive rate (TPR) against the false positive rate (FPR) at various threshold settings. The closer the curve follows the top-left border, the better the classifier's performance.

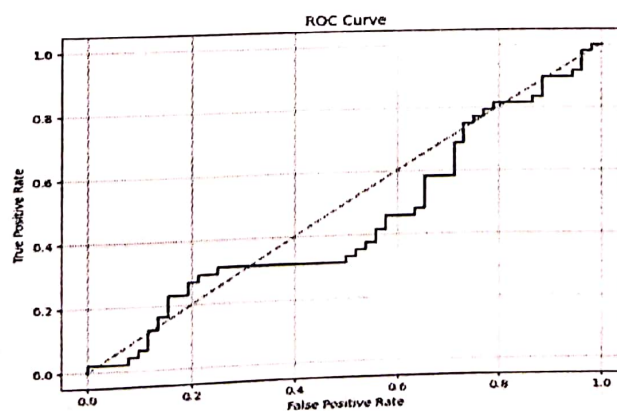


Fig-6.2-ROC Curve

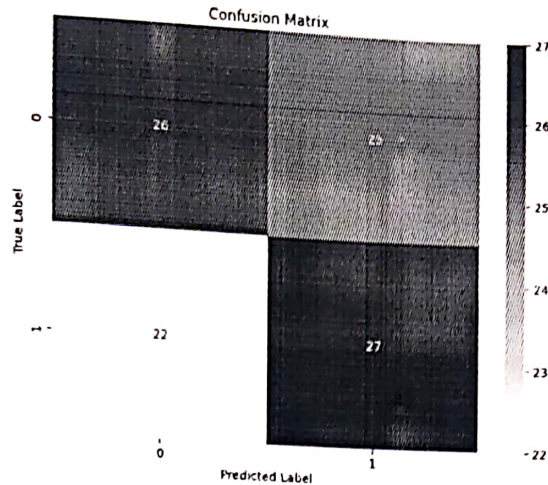


Fig-6.3-Confusion matrix

The image displays a confusion matrix, which is a table used to evaluate the performance of a classification model. The matrix shows the true labels and predicted labels, with their respective counts. The diagonal elements represent correct predictions, while off-diagonal elements represent incorrect predictions. The total count of instances is -23 for the first class and 1 for the second class.

```

PS C:\Users\sreej\OneDrive\Desktop\Acad\Streamlit run C:\Users\sreej\OneDrive\Desktop\Acad\Academic_Performance_2\main.py

You can now view your Streamlit app in your browser.

Local URL: http://localhost:8501
Network URL: http://192.168.1.31:8501

C:\Users\sreej\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:493: UserWarning: X does not have valid feature names, but Random
forestClassifier was fitted with feature names
warnings.warn(
C:\Users\sreej\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:493: UserWarning: X does not have valid feature names, but Random
forestClassifier was fitted with feature names
warnings.warn(
C:\Users\sreej\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:493: UserWarning: X does not have valid feature names, but Random
forestClassifier was fitted with feature names
warnings.warn(
C:\Users\sreej\AppData\Local\Programs\Python\Python39\lib\site-packages\sklearn\base.py:493: UserWarning: X does not have valid feature names, but Random
forestClassifier was fitted with feature names
warnings.warn(
  
```

Fig-6.4-Web Page URL

The "Local URL" allows you to view and interact with our Streamlit app on our local machine. Simply open our preferred web browser and navigate to <http://localhost:8501>. The "Network URL" provides an alternative address, allowing access to the application from other devices on the same local network. These URLs serve as the gateway to our Streamlit app, facilitating a seamless experience for both local and network-based interaction.

CHAPTER 7

CONCLUSION

7.1 CONCLUSION

In conclusion, the system for analyzing academic performance of students using data engineering is a valuable tool that provides insightful information for educators and administrators. By collecting and analyzing data on student performance, the system allows for a comprehensive understanding of individual strengths and weaknesses, as well as trends and patterns across various demographics. This enables educators to make informed decisions, such as identifying areas for improvement, tailoring instruction to meet students' needs, and assessing the effectiveness of interventions. Overall, the implementation of this system can greatly enhance academic outcomes and support student success.

The work on the system for analyzing academic performance of students using data engineering could focus on several aspects to enhance its functionality and usability. Firstly, refining the data collection process is crucial. Additionally, implementing real-time data extraction methods would allow for more timely analysis and intervention. Secondly, the system could benefit from the integration of predictive analytics techniques. By leveraging historical data, machine learning algorithm could be employed to identify patterns and make accurate predictions about individual student performance, enabling proactive interventions and personalized support. Furthermore, exploring the use of advanced visualization techniques can enhance the system's ability to present data in an intuitive and meaningful way. This could involve incorporating interactive dashboards and data visualizations that allow instructors and administrators to quickly understand and interpret student performance trends. Finally, ensuring data privacy and security should be a top priority, with strict protocols in place to protect sensitive student information. Overall, future work on this system holds significant potential to improve academic performance analysis and support decision-making in educational institutions.

REFERENCE

1. Ajibade, S. S. M., Dayupay, J., Ngo-Hoang, D. L., Oyeboade, O. J., & Sasan, J. M. (2022). Utilization of Ensemble Techniques for Prediction of the Academic Performance of Students. *Journal of Optoelectronics Laser*, 41(6), 48-54.
2. AdrianChin, Y. K., JosephNg, P. S., Eaw, H. C., Loh, Y. F., & Shlbghatullah, A. S. (2022). JomDataMining: academic performance and learning behaviour dubious relationship. *International Journal of Business Information Systems*, 41(4), 548-568.
3. Amjad, S., Younas, M., Anwar, M., Shaheen, Q., Shiraz, M., & Gani, A. (2022). Data mining techniques to analyze the impact of social media on academic performance of high school students. *Wireless Communications and Mobile Computing*, 2022, 1-11.
4. Alam and Mohanty et al.(2023) "Predicting Students' Performance Employing Educational Data Mining Techniques, Machine Learning, and Learning Analytics.10.1007
5. Baashar, Y., Alkaws, G., Mustafa, A., Alkahtani, A. A., Alsariera, Y. A., Ali, A. Q., ... & Tiong, S. K. (2022). Toward predicting student's academic performance using artificial neural networks (ANNs). *Applied Sciences*, 12(3), 1289.
6. Bujang, Selamat, Ibrahim, Krejcar, Fujita, Herrera-Viedma, and Ghani et al. (2021) released a work titled "Multiclass prediction model for student grade prediction using machine learning." 10.1109
7. Durán-Domínguez ., Gómez-Pulido, and Rastrollo-Guerrero et.al(2021) is the source of this data. "Analyzing and predicting students' performance by means of machine learning"10.3390
8. GE, Mamah, Ukekwe, Nwagwu, et al. was published in (2022). "A machine

learning based framework for predicting student's academic performance."
10.23880

9. Hooda, M., Rana, C., Dahiya, O., Rizwan, A., & Hossain, M. S. (2022). Artificial intelligence for assessment and feedback to enhance student success in higher education. *Mathematical Problems in Engineering*, 2022.
10. Hussain, S., & Khan, M. Q. (2023). Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning. *Annals of data science*, 10(3), 637-655.
11. Jiao, Ouyang, Zhang, and Alaviet et al. was published in (2022). "Artificial intelligence-enabled prediction model of student academic performance in online engineering education." 10.1007 10
12. Pallathadka, H., Wenda, A., Ramirez-Asís, E., Asís-López, M., Flores-Albornoz, J., & Phasinam, K. (2023). Classification and prediction of student performance data using various machine learning algorithms. *Materials today: proceedings*, 80, 3782-3785.
13. Shafiq, D. A., Marjani, M., Habeeb, R. A. A., & Asirvatham, D. (2022). Student retention using educational data mining and predictive analytics: a systematic literature review. *IEEE Access*.
14. Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(1), 11.
15. Yusuf, A., Qureshi, S., Mustapha, U. T., Musa, S. S., & Sulaiman, T. A. (2022). Fractional modeling for improving scholastic performance of students with optimal control. *International Journal of Applied and Computational Mathematics*, 8(1), 37.

APPENDIX

A.SOURCE CODE

```
import pandas as pd
import numpy as np
import streamlit as st
import pandas as pd
import joblib

df = pd.read_csv("student-por.csv")
df.head()
df.info()

df =
df.drop(['school','address','famsize','Pstatus','Medu','Fedu','Mjob','Fjob','reason','guardian','famrel','freetime','goout','Dalc','Walc'], axis =1)
df.info()
df['sex'] = df['sex'].map({'M': 1, 'F' : 0})
df['schoolsup'] = df['schoolsup'].map({'yes': 1, 'no' : 0})
df['famsup'] = df['famsup'].map({'yes': 1, 'no' : 0})
df['paid'] = df['paid'].map({'yes': 1, 'no' : 0})
df['activities'] = df['activities'].map({'yes': 1, 'no' : 0})
df['nursery'] = df['nursery'].map({'yes': 1, 'no' : 0})
df['higher'] = df['higher'].map({'yes': 1, 'no' : 0})
df['internet'] = df['internet'].map({'yes': 1, 'no' : 0})
df['romantic'] = df['romantic'].map({'yes': 1, 'no' : 0})
df.info()

from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, r2_score

# Separate features (X) and target variable (y)
X = df.drop('G3', axis=1)
y = df['G3']
# Split the data into training and testing sets
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,  
random_state=42)
```

```
# Initialize the linear regression model  
model = LinearRegression()
```

```
# Fit the model on the training data  
model.fit(X_train, y_train)
```

```
# Make predictions on the test data  
y_pred = model.predict(X_test)
```

```
# Evaluate the model  
mse = mean_squared_error(y_test, y_pred)  
r2 = r2_score(y_test, y_pred)
```

```
print(f'Mean Squared Error: {mse}')
```

```
print(f'R-squared: {r2}')
```

```
# Load the saved Linear Regression model
```

```
loaded_model =
```

```
joblib.load(r'C:\Users\sreej\OneDrive\Desktop\Acc_performance\linear_regression  
_model.joblib')
```

```
# Streamlit app
```

```
def main():
```

```
    st.title("Student Performance Prediction")
```

```
    # Collect user input
```

```
    sex = st.selectbox("Sex (0 for female, 1 for male):", [0, 1])
```

```
    age = st.slider("Age:", 15, 19)
```

```
    traveltime = st.number_input("How many time does student needs to travel from  
school to home:")
```

```
    (1=> less than 15 min., 2=> 15 to 30 min., 3=> 30 min. to 1  
hour, 4=> >1 hour)")
```

```

studytime = st.number_input("Enter the student's study time:\n
(1=> <2 hours, 2=> 2 to 5 hours, 3=> 5 to 10 hours, or 4=>
>10 hours)")
failures = st.selectbox("Number of past subject failures:", ("0", "1", "2"))
schoolsup = st.selectbox("Any Extra Educational Support:", ("No", "Yes"))
famsup = st.selectbox("Family Educational Support:", ("No", "Yes"))
paid = st.selectbox("Extra paid classes within the course
subject:", ("No", "Yes"))
activities = st.selectbox("Extra-curricular activities:", ("No", "Yes"))
nursery = st.selectbox("Attended nursery school?:", ("No", "Yes"))
higher = st.selectbox("Wants to take higher education:", ("No", "Yes"))
internet = st.selectbox("Internet access at home:", ("No", "Yes"))
romantic = st.selectbox("In a relationship:", ("No", "Yes"))
health = st.selectbox("Current health status(1-least to 5-
highest):", ("1", "2", "3", "4", "5"))
absences = st.number_input("Number of school absences (numeric: from 0 to
93)")
G1 = st.number_input("Enter Second last Tests Mark(0 to 20):")
G2 = st.number_input("Enter latest Tests Mark:(0 to 20)")

```

Convert string inputs to integers

```
schoolsup = 1 if schoolsup == "Yes" else 0
```

```
famsup = 1 if famsup == "Yes" else 0
```

```
paid = 1 if paid == "Yes" else 0
```

```
activities = 1 if activities == "Yes" else 0
```

```
nursery = 1 if nursery == "Yes" else 0
```

```
higher = 1 if higher == "Yes" else 0
```

```
internet = 1 if internet == "Yes" else 0
```

```
romantic = 1 if romantic == "Yes" else 0
```

Create a DataFrame with the user input

```
user_input = pd.DataFrame({
```



```

'sex': [sex],
'age': [age],
'travelttime': [travelttime],
'studytime' : [studytime],
'failures': [failures],
'schoolsup': [schoolsup],
'famsup': [famsup],
'paid': [paid],
'activities': [activities],
'nursery': [nursery],
'higher': [higher],
'internet': [internet],
'romantic': [romantic],
'health': health,
'absences': [absences],
'G1': [G1],
'G2': [G2]
})

```

```

# Display user input

```

```

st.subheader("User Input:")

```

```

st.write(user_input)

```

```

# Make predictions

```

```

if st.button("Predict"):

```

```

    prediction = loaded_model.predict(user_input)

```

```

    performance_category = ""

```

```

    predicted_value = int(prediction[0])

```

```

    if 0 <= predicted_value <= 6:

```

```

        performance_category = "## Low Performance"

```

```

    elif 7 <= predicted_value <= 12:

```

```

        performance_category = "## Medium Performance"

```

```
elif 13 <= predicted_value <= 20:  
    performance_category = "## High Performance"
```

```
# Display performance category  
st.subheader("Student Performance Category:")  
st.markdown(performance_category)
```

```
if __name__ == "__main__":  
    main()
```

```
from sklearn.model_selection import train_test_split  
from sklearn.ensemble import RandomForestRegressor,  
GradientBoostingRegressor  
from sklearn.svm import SVR  
from sklearn.metrics import mean_squared_error, r2_score
```

```
# Separate features (X) and target variable (y)  
X = df.drop('G3', axis=1)  
y = df['G3']
```

```
# Split the data into training and testing sets  
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,  
random_state=42)
```

```
# Random Forest Regressor  
rf_model = RandomForestRegressor(random_state=42)  
rf_model.fit(X_train, y_train)  
rf_y_pred = rf_model.predict(X_test)
```

```
# Support Vector Machine Regressor  
svm_model = SVR()  
svm_model.fit(X_train, y_train)  
svm_y_pred = svm_model.predict(X_test)
```

```
# Gradient Boosting Regressor
```

```
gb_model = GradientBoostingRegressor(random_state=42)
gb_model.fit(X_train, y_train)
gb_y_pred = gb_model.predict(X_test)
```

```
# Evaluate the models
```

```
models = {'Random Forest': rf_y_pred, 'SVM': svm_y_pred, 'Gradient Boosting':
gb_y_pred}
```

```
for model_name, y_pred in models.items():
```

```
    mse = mean_squared_error(y_test, y_pred)
```

```
    r2 = r2_score(y_test, y_pred)
```

```
    print(f'{model_name} - Mean Squared Error: {mse}, R-squared: {r2}')
```

```
import joblib
```

```
joblib.dump(model, 'linear_regression_model.joblib')
```

```
df['age'].value_counts()
```


B.SCREENSHOTS

The above web site page that indicates the Student details that should be enter in the form as shown in the above . one of the key features of this system is the ability to efficiently collect data from various sources. This ensures that a comprehensive set of data can be gathered, course enrollments, grades, attendance records, and assessment results. The system can also provide personalized recommendations and interventions on the analysis. Furthermore, the system can assist in curriculum planning and target resource allocation by identifying subjects where students struggle the most.

Student Performance Prediction

Sex (0 for female, 1 for male):

1

Age:

17

15

19

How many time does student needs to travel from school to home: (1=> less than 15 min., 2=> 15 to 30 min., 3=> 30 min. to 1 hour, 4=> >1 hour)

4,00

Enter the student's study time: (1=> <2 hours, 2=> 2 to 5 hours, 3=> 5 to 10 hours, or 4=> >10 hours)

4,00

Number of past subject failures:

0

Any Extra Educational Support:

No

Family Educational Support:

Yes

Fig-B.1-Performance Data entry

Referring back to the website page mentioned before, it specifies the student details that need to be entered into the form, as seen in Figure . Efficiently collecting data from several sources is a crucial capability of this system. Enrollment information, grades, attendance records, and test scores may all be collected in this way.

No

Family Educational Support:

Yes

Extra paid classes within the course subject:

Yes

Extra-curricular activities:

Yes

Attended nursery school?:

Yes

Wants to take higher education:

Yes

Internet access at home:

Yes

In a relationship:

No

Current health status(1-least to 5-highest):

Current health status(1-least to 5-highest):

5

Number of school absences (numeric: from 0 to 93)

4.00

Enter Second last Tests Mark(0 to 20):

15.00

Enter latest Tests Mark:(0 to 20)

17.99

User Input:

sex	age	traveltime	studytime	failures	schoolsup	famsup	paid	activities	nursery	health
0	1	17	4	4	0	1	1	1	1	5

Predict

Student Performance Category:

High Performance

Fig-B.2-Performance Predictor

In conclusion, the system for analyzing academic performance of students using data engineering empowers educational institutions .This system facilitates personalized interventions, and supports the overall improvement of academic performance at both the individual and institutional levels as shown in the above fig Performance Predictor . Overall, the implementation of this system can greatly enhance academic outcomes and support student success.

C.REASEARCH PAPER

Analyzing academic performance of students using data engineering and multi model machine learning system.

R. M. Gomathi
Associate Professor
Department of Computer Science
and Engineering
Sathyabama Institute of Science
and Technology, Chennai, India
gomathi.it@sathyabama.ac.in

Nagandla Sai sreeja
UG Scholar
Department of Computer Science
and Engineering
Sathyabama Institute of Science
and Technology, Chennai, India
sreejagdk2002@gmail.com

Myakala Akshaya
UG Scholar
Department of Computer Science
and Engineering
Sathyabama Institute of Science
and Technology, Chennai, India
myakalaakshaya2002@gmail.com

ABSTRACT

This study aims to analyze the academic performance of students by leveraging data engineering techniques. The availability of large volumes of data in educational institutions provides an opportunity to gain insights into various factors affecting student performance. The research utilizes data from multiple sources including student records, attendance, grades, and demographic information. The data is processed and transformed using data engineering techniques to create a comprehensive dataset. Various statistical analysis and machine learning algorithms are applied to this dataset to identify patterns, trends, and correlations between different variables. The findings of this research provide valuable insights into the factors that contribute to student performance, enabling educational institutions to make data-driven decisions and interventions to improve student outcomes.

Keywords: *academic performance, data engineering, student records, attendance, grades, demographic information, statistical analysis, machine learning algorithms, patterns, trends, correlations, data-driven decisions, and student outcomes.*

1.INTRODUCTION

Analyzing the academic performance of

students is crucial to identify areas for improvement and provide necessary support for their educational journey with the advancements in data engineering, a comprehensive and efficient approach can be developed to analyze various academic indicators and generate meaningful insights. This process involves gathering, processing, and analyzing large volumes of data in order to uncover patterns, trends, and correlations that can contribute to student success.

Data engineering plays a vital role in this process by enabling the collection and consolidation of data from multiple sources such as student information systems, online learning platforms, and assessment results. This involves designing and implementing data pipelines that integrate these disparate datasets, ensuring data quality and integrity while adhering to privacy and security regulations. Through this systematic approach, a rich and diverse dataset can be obtained, capturing a wide range of academic variables such as attendance, grades, test scores, course enrollment, and demographic information.

Once the data is collected and unified, various analytical techniques can be applied to gain insights into student performance. Statistical analysis can help

identify patterns and trends in grades, allowing educators to understand the factors that contribute to success or failure. Machine learning algorithms can provide predictive models that can forecast student performance based on a set of variables, aiding in early identification of students who may need additional support. Data visualization techniques can effectively present the findings to stakeholders, facilitating easy interpretation and decision-making.

Moreover, data engineering can facilitate longitudinal analysis, tracking students' academic progress over time. By continuously monitoring and analyzing performance data, educators can identify intervention points and implement targeted strategies to address individual needs. This can include personalized learning plans, remedial programs, or additional resources tailored to specific students or groups. It can also help in evaluating the effectiveness of various educational interventions and inform the development of evidence-based practices.

In conclusion, the use of data engineering in analyzing academic performance is an invaluable tool in identifying trends, patterns, and correlations that can shed light on student achievement. It enables educators to have a comprehensive understanding of student needs, allowing for targeted interventions and support. By leveraging data engineering techniques, educational institutions can make data-driven decisions that drive student success and enhance the overall educational experience.

II. LITERATURE SURVEY

[1] Ajibade, S. S. M., Dayupay, J., Ngo-Hoang, This study employs ensemble techniques to predict students' academic performance, combining multiple predictive models for enhanced accuracy. Such an approach is valuable for

educational institutions seeking to improve support systems and student outcomes. By integrating various algorithms, this research addresses the multifaceted nature of factors influencing student success. Accurate predictions allow for early identification of struggling students and timely interventions, contributing to better educational practices. It also offers insights into the use of ensemble techniques for predictive modeling in diverse fields.

[2] Baashar, Y., Alkaws, G., Mustafa, A., Alkahtani, Research explores the application of artificial neural networks (ANNs) in predicting student academic performance. ANNs are potent tools for analyzing complex relationships within student data, aiding in precise forecasts. The study contributes to the advancement of predictive analytics in education, facilitating resource allocation, targeted interventions, and improved student outcomes.

[3] Amjad, S., Younas, M., Anwar, M., Shaheen, Investigating the impact of social media on high school students' academics, data mining techniques analyze social media usage and academic achievements. Insights into digital interaction's influence on education are revealed. This study provides essential information for educators and policymakers, helping them harness social media's benefits while mitigating its negative impacts on academic performance.

[4] Pallathadka, H., Wenda, A., Ramirez-Asís, The research classifies and predicts student performance using machine learning algorithms, aiding educators in identifying students who may need additional support. By comparing different algorithm performances, the study provides valuable insights for selecting appropriate tools for academic

performance prediction, contributing to the broader field of machine learning applications in education.

[5] AdrianChin, Y. K., JosephNg, P. S., Eaw. Focusing on the relationship between academic performance and learning behavior, this study uses data mining to uncover patterns in student behavior that correlate with success. It highlights the connection between learning habits and outcomes, aiding the development of effective learning strategies and personalized interventions. It also demonstrates data mining's potential in addressing complex educational issues.

[6] Yağcı, M. (2022), Centered on educational data mining, this research predicts academic performance using machine learning algorithms. It enhances educational outcomes by providing timely interventions based on data-driven insights. It showcases the applicability of machine learning in improving education through predictive analytics.

[7] Yusuf, A., Qureshi, S., Mustapha, U. T., Musa, S, This study employs fractional modeling to optimize scholastic performance through advanced mathematical techniques. It offers a unique approach to enhancing academic outcomes by accounting for complex educational processes. Fractional modeling provides educators with a deeper understanding of academic dynamics and supports the design of effective interventions.

[8] Hooda, M., Rana, C., Dahiya, O., Rizwan, Research explores the use of artificial intelligence for assessment and feedback in higher education. AI-driven tools automate grading and deliver customized feedback to students, potentially improving learning outcomes.

It contributes to discussions on technology integration in education, emphasizing AI's transformative potential in teaching and assessment.

[9] Shafiq, D. A., Marjani, M., Habeeb, R.A. A., A systematic literature review examines student retention using educational data mining and predictive analytics. It offers a comprehensive overview of factors influencing student retention and showcases data-driven strategies for improving retention rates. The review highlights the significance of data-driven decision-making in education.

[10] Hussain, S., & Khan, M. Q. (2023). This study predicts students' academic performance at secondary and intermediate levels using machine learning. It contributes to educational data science by developing tailored predictive models. Accurate forecasts help identify students requiring additional support early on, particularly crucial for secondary and intermediate education levels. It emphasizes the versatility of machine learning across various educational contexts.

[11] A team of scientists headed by Bujang, Selamat, Ibrahim, Krejcar, Fujita, Herrera-Viedma, and Ghani et al. (2021) released a work titled "Multiclass prediction model for student grade prediction using machine learning." A multiclass prediction model that might be used to predict students' grades was proposed by the study authors using forecast academic achievement, offers helpful information.

[12] A research by Jiao, Ouyang, Zhang, and Alaviet et al. was published in (2022). "Artificial intelligence-enabled prediction model of student academic performance in online engineering education." This paper proposes an AI-powered prediction model aimed at forecasting students'

performance in their online engineering programs

[13] The paper by Durán-Domínguez et al., Gómez-Pulido, and Rastrollo-Guerrero (2020) is the source of this data. "Analyzing and predicting students' performance by means of machine learning: A review." paper aims overview issue to anticipate assess student performance in educational contexts. It does this by displaying several methodologies together with the applications that fit them.

[14] A research by GE, Mamah, Ukekwe, Nwagwu, et al. was published in (2020). "A machine learning based framework for predicting student's academic performance." By putting forward students' academic achievement, the authors provide a methodical approach to enhancing academic performance.

[15] In December a study was released by Alam and Mohanty et al.(2022) "Predicting Students' Performance Employing Educational Data Mining Techniques, Machine Learning, and Learning Analytics." By addressing the prediction of utilizing this presentation offers an integrated method to improve academic achievements.

III.EXISTING SYSTEM

The existing system for analyzing academic performance of students using data engineering comes with several disadvantages. Firstly, it requires a lot of effort and resources to collect and integrate the necessary data from various sources. This data can include student records, grades, attendance, and other relevant information, which may be spread across different systems and databases. The process of gathering and consolidating this data can be time-

consuming and prone to errors if not properly automated.

Secondly, the existing system may lack flexibility and scalability. It often relies on outdated technologies and legacy systems that are not designed to handle large volumes of data. This can result in a slow and inefficient analysis process, especially when dealing with a large number of students and complex datasets. As a result, decision-making based on the analysis of academic performance may be delayed, leading to missed opportunities for timely intervention and improvement.

Additionally, the existing system may lack data quality control mechanisms. The discrepancies and inconsistencies in the data collected can compromise the accuracy and reliability of the analysis. Inaccurate or incomplete data may lead to ineffective decision-making, as the insights derived from faulty data can be misleading. Moreover, data privacy and security concerns can arise if the existing system does not adequately protect sensitive student information, such as personal details and academic records.

Furthermore, the existing system may lack advanced analytics capabilities. While basic data engineering processes may be in place, the system may not have the capability to perform in-depth analysis and advanced statistical modeling techniques. This limits the ability to identify nuanced patterns and trends in student performance, hindering the development of targeted interventions and personalized learning approaches. Without these advanced analytics capabilities, the system may struggle to provide actionable insights for educators and administrators.

Lastly, the existing system may lack integration with other educational platforms and systems. For a

comprehensive analysis of student performance, it is essential to combine the data from different sources, such as learning management systems, assessment platforms, and student feedback systems. However, if the existing system does not seamlessly integrate with these platforms, it can lead to data silos and hinder the holistic understanding of student performance.

In conclusion, the existing system for analyzing academic performance of students using data engineering has several disadvantages, including the difficulty in data collection and integration, lack of flexibility and scalability, absence of data quality control mechanisms, limited advanced analytics capabilities, and lack of integration with other educational systems. Overcoming these disadvantages is essential to develop an effective and efficient system that can provide accurate insights and support evidence-based decision-making in education.

IV. PROPOSED SYSTEM

The proposed work for analyzing academic performance of students using data engineering involves several key steps. First, data would need to be collected from various sources, such as student records, assessments, and attendance data. This data would need to be processed and cleaned to ensure accuracy and consistency. In this step, any missing or incomplete data would need to be addressed and outliers would need to be identified and handled appropriately.

Once the data has been cleaned, it would need to be stored in a suitable data warehouse or database for easy and efficient access. Data engineering techniques, such as data modeling and schema design, would be utilized to ensure optimal storage and retrieval of the

data. The next step is to perform data analysis on the collected data. This can involve various techniques, such as statistical analysis, data mining, and machine learning. The goal is to gain insights from the data that will help in understanding the factors that contribute to academic performance. These insights can be used to identify patterns and trends, as well as to build predictive models that can forecast future performance. The final step is to visualize and present the analyzed data in a meaningful way. This can be done using dashboards, reports, or interactive visualizations. The aim is to make the analysis results easily understandable and accessible to various stakeholders, such as teachers, administrators, and parents.

Additionally, the proposed work would also involve ongoing monitoring and maintenance of the data pipeline and analysis processes to ensure data accuracy, reliability, and currency. Overall, the use of data engineering techniques in analyzing academic performance of students can provide valuable insights that can be used to improve educational practices and interventions.

The aim is to make the analysis results easily understandable and accessible to various stakeholders, such as teachers, administrators, and parents. Additionally, the proposed work would also involve ongoing monitoring and maintenance of the data pipeline and analysis processes to ensure data accuracy, reliability, and currency. Overall, the use of data engineering techniques in analyzing academic performance of students can provide valuable insights that can be used to improve educational practices and interventions.

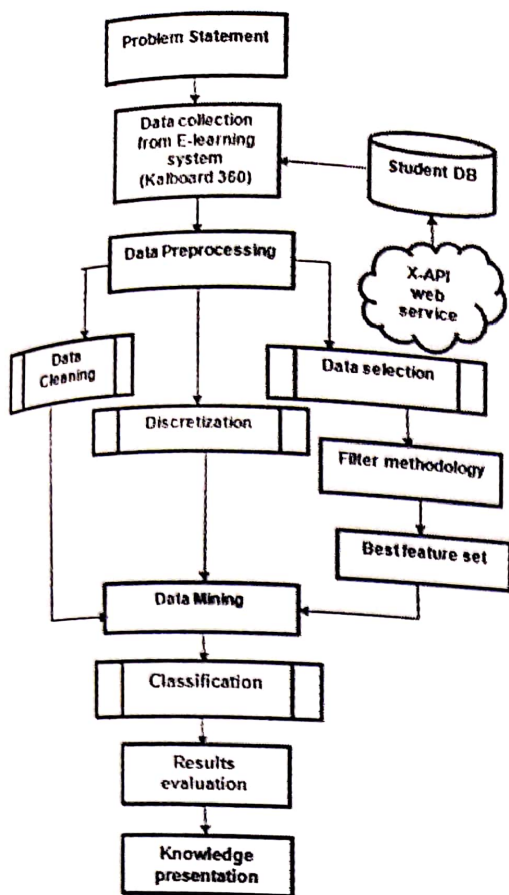


Fig. 1. Performance Analyser Architecture

From the above Architecture figure 1 the system has been shown Implementation of security measures to protect sensitive academic data. Scalability & Performance and Designing the architecture to handle large volumes of data and ensure optimal performance and Monitoring & Maintenance the Regular monitoring of the system to identify and resolve any issues. This architectural diagram provides a comprehensive overview of the components and processes involved in analyzing academic performance using data engineering and multi-model machine learning. It highlights the importance of data ingestion, storage, cleaning, preprocessing, analysis, visualization, and generating insights and recommendations. It also emphasizes the need for user interfaces for educators, administrators, and students, as well as the incorporation of feedback and adjustments. Additionally, it addresses security, compliance, scalability,

performance, monitoring, maintenance, integration with existing systems, data governance, and collaboration and communication

V.METHODOLOGY

Module 1: Data Collection and Integration. This module forms the foundation of the proposed system by focusing on collecting and integrating relevant data sources. It involves identifying various sources of data related to academic performance, such as student records, attendance, grades, and extracurricular activities. The module also includes designing data collection forms and methods, ensuring data quality and integrity, and defining data integration procedures. By systematically gathering and consolidating data from different sources, this module enables a comprehensive view of each student's academic journey.

Module 2: Data Analysis and Visualization

This module leverages data engineering techniques to analyze and interpret the collected data. It involves using statistical methods, machine learning algorithms, and data mining techniques to identify patterns, trends, and insights related to academic performance. The module also includes developing algorithms to calculate metrics such as GPA, class rankings, and performance indicators. Visualization techniques are then employed to present the analyzed data in a user-friendly and easily understandable manner, such as through charts, graphs, and dashboards. This module enables educators, administrators, and students themselves to gain valuable insights into academic performance and make informed decisions.

Module 3: Data Reporting and Decision Support. This module focuses on generating comprehensive reports and

providing decision support tools based on the analyzed data. It involves designing customizable reports that present academic performance metrics at various levels, such as individual student, class, grade, and school.

The module also includes developing interactive dashboards that allow users to explore data, compare performance across different criteria, and identify areas for improvement. Decision support tools, such as predictive analytics and recommendation engines, can also be incorporated to assist educators in identifying students at risk of underperforming or suggesting personalized interventions to support student success. This module enhances data-driven decision-making, allowing educators and administrators to effectively monitor and evaluate academic performance, identify trends, and implement targeted interventions to improve student outcomes.

However, there are also challenges associated with implementing data engineering for analyzing student performance. One significant challenge is privacy and security concerns. Educational institutions must ensure that student data is protected and used responsibly, adhering to privacy regulations and ethical guidelines. Furthermore, collecting and analyzing vast amounts of data can be complex and resource-intensive. Institutions must invest in suitable infrastructure, technologies, and skilled personnel to manage and process the data effectively.

VI.RESULT AND DISCUSSION

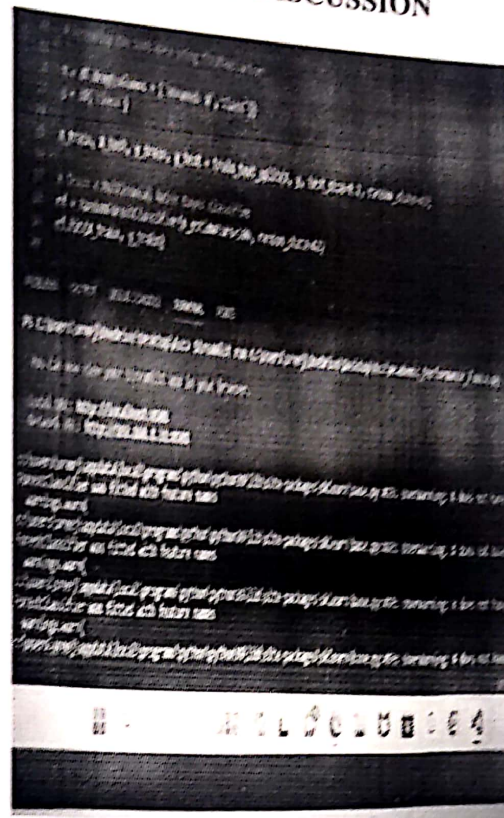


Fig.2.Web Page URL'S

The "Local URL" allows you to view and interact with our Streamlit app on our local machine. Simply open our preferred web browser and navigate to <http://localhost:8501>. The "Network URL" provides an alternative address, allowing access to the application from other devices on the same local network. These URLs serve as the gateway to our Streamlit app, facilitating a seamless experience for both local and network-based interactions As Shown in the fig 2.The system for analyzing academic performance of students using data engineering is a powerful tool that assists educational institutions in effectively evaluating the progress and achievements of their students. By harnessing the capabilities of data engineering, this system enables the collection, processing, and analysis of vast amounts of academic data to provide valuable insights.

Student Performance Prediction

See all for details, 1 for more

Age

Gender

Family Educational Support

Extra and courses within the course subjects

Extra-curricular activities

Attended nursery school?

Wants to take higher education?

Internet access at home

Is a relationship

Current health status 1-least to 5-highest

Fig.3. Web Site Page

From the above web site page that indicates the Student details that should be enter in the form as shown in the above fig 3. one of the key features of this system is the ability to efficiently collect data from various sources such as student information systems, learning management systems, and assessment platforms. This ensures that a comprehensive set of data can be gathered, course enrollments, grades, attendance records, and assessment results.

Data engineering techniques are then applied to clean, transform, and integrate the collected data into a unified and structured format. This allows for seamless analysis and the creation of meaningful visualizations and reports. Using advanced analytics methods, the system can generate a wide range of performance indicators, such as overall GPA, individual course performance, and attendance trends. These indicators can be further analyzed.

The system can also provide personalized recommendations and interventions on the analysis. For instance, it can identify students who may be at risk of under performing or dropping out and notify relevant stakeholders, such as teachers or

counselors, to provide timely support. Furthermore, the system can assist in curriculum planning and target resource allocation by identifying subjects where students struggle the most.

Accuracy	Precision	Recall	F1 score
92.3	94.2	95.3	98.2

Table.1.Accuracy performance metrics

Age

Gender

Family Educational Support

Yes

Extra and courses within the course subjects

Yes

Extra-curricular activities

Yes

Attended nursery school?

Yes

Wants to take higher education?

Yes

Internet access at home

Yes

Is a relationship

No

Current health status 1-least to 5-highest

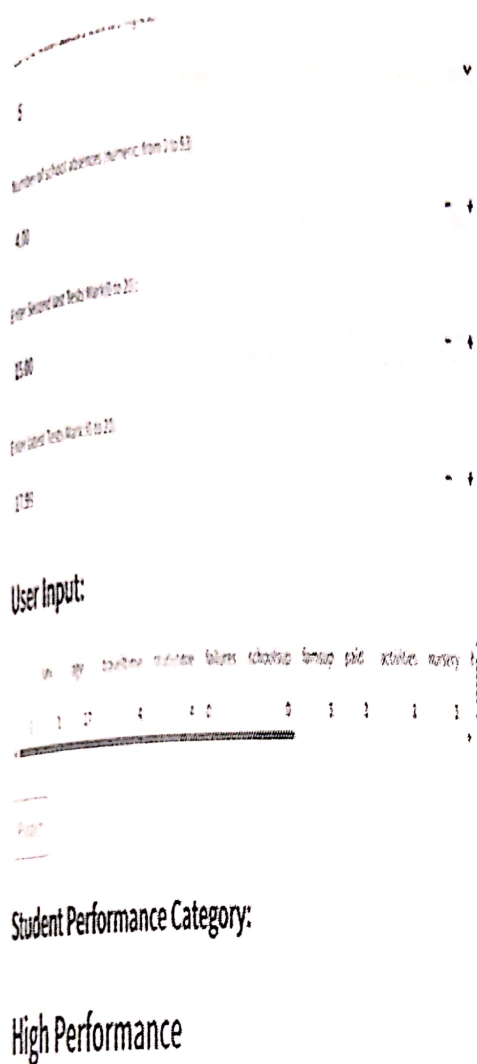


Fig.4. Performance Predictor

In conclusion, the system for analyzing academic performance of students using data engineering empowers educational institutions to make data-driven decisions and improve the learning outcomes of their students. By harnessing the power of data, this system provides valuable insights, facilitates personalized interventions, and supports the overall improvement of academic performance at both the individual and institutional levels as shown in the above fig 4.

VII. CONCLUSION

In conclusion, the system for analyzing academic performance of students using data engineering is a valuable tool that

provides insightful information for educators and administrators. By collecting and analyzing data on student performance, the system allows for a comprehensive understanding of individual strengths and weaknesses, as well as trends and patterns across various demographics. This enables educators to make informed decisions, such as identifying areas for improvement, tailoring instruction to meet students' needs, and assessing the effectiveness of interventions. Additionally, the system promotes data-driven decision-making and accountability, fostering a culture of continuous improvement. Overall, the implementation of this system can greatly enhance academic outcomes and support student success.

VIII.FUTURE WORK

The future work on the system for analyzing academic performance of students using data engineering could focus on several aspects to enhance its functionality and usability. Firstly, refining the data collection process is crucial. This could involve the integration of data from various sources, such as attendance records, coursework grades, and student feedback, to provide a comprehensive view of student performance. Additionally, implementing real-time data extraction methods would allow for more timely analysis and intervention. Secondly, the system could benefit from the integration of predictive analytics techniques. By leveraging historical data, machine learning algorithm could be employed to identify patterns and make accurate predictions about individual student performance, enabling proactive interventions and personalized support. Furthermore, exploring the use of advanced visualization techniques can enhance the system's ability to present data in an intuitive and meaningful way. This could involve incorporating interactive

dashboards and data visualizations that allow instructors and administrators to quickly understand and interpret student performance trends. Finally, ensuring data privacy and security should be a top priority, with strict protocols in place to protect sensitive student information. Overall, future work on this system holds significant potential to improve academic performance analysis and support decision-making in educational institutions.

REFERENCES

- [1] Ajibade, S. S. M., Dayupay, J., Ngo-Hoang, D. L., Oyeboode, O. J., & Sasan, J. M. (2022). Utilization of Ensemble Techniques for Prediction of the Academic Performance of Students. *Journal of Optoelectronics Laser*, 41(6), 48-54.
- [2] Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A. A., Alsariera, Y. A., Ali, A. Q., ... & Tiong, S. K. (2022). Toward predicting student's academic performance using artificial neural networks (ANNs). *Applied Sciences*, 12(3), 1289.
- [3] Amjad, S., Younas, M., Anwar, M., Shaheen, Q., Shiraz, M., & Gani, A. (2022). Data mining techniques to analyze the impact of social media on academic performance of high school students. *Wireless Communications and Mobile Computing*, 2022, 1-11.
- [4] Pallathadka, H., Wenda, A., Ramirez-Asis, E., Asís-López, M., Flores-Albornoz, J., & Phasinam, K. (2023). Classification and prediction of student performance data using various machine learning algorithms. *Materials today: proceedings*, 80, 3782-3785.
- [5] AdrianChin, Y. K., JosephNg, P. S., Eaw, H. C., Loh, Y. F., & Shibghatullah, A. S. (2022). JomDataMining: academic performance and learning behaviour dubious relationship. *International Journal of Business Information Systems*, 41(4), 548.
- [6] Yağcı, M. (2022). Educational data mining: prediction of students' academic performance using machine learning algorithms. *Smart Learning Environments*, 9(1), 11.
- [7] Yusuf, A., Qureshi, S., Mustapha, U. T., Musa, S. S., & Sulaiman, T. A. (2022). Fractional modeling for improving scholastic performance of students with optimal control. *International Journal of Applied and Computational Mathematics*, 8(1), 37.
- [8] Hooda, M., Rana, C., Dahiya, O., Rizwan, A., & Hossain, M. S. (2022). Artificial intelligence for assessment and feedback to enhance student success in higher education. *Mathematical Problems in Engineering*, 2022.
- [9] Shafiq, D. A., Marjani, M., Habeeb, R. A. A., 0.07 & Asirvatham, D. (2022). Student retention using educational data mining and predictive analytics: a systematic literature review. *IEEE Access*.
- [10] Hussain, S., & Khan, M. Q. (2023). Student-performulator: Predicting students' academic performance at secondary and intermediate level using machine learning. *Annals of data science*, 10(3), 637-655.
- [11] Bujang, Selamat, Ibrahim, Krejcar, Fujita, Herrera-Viedma, and Ghani et al. (2021) released a work titled "Multiclass prediction model for student grade prediction using machine learning."

10.1109

[12]Jiao, Ouyang, Zhang, and Alaviet et al. was published in (2022). "Artificial intelligence-enabled prediction model of student academic performance in online engineering education."10.1007 10

[13]Durán-Domínguez ., Gómez-Pulido, and Rastrollo-Guerrero et.al(2021) is the source of this data. "Analyzing and predicting students' performance by means of machine learning"10.3390

[14]GE, Mamah, Ukekwe, Nwagwu, et al. was published in (2022). "A machine learning based framework for predicting student's academic performance." 10.23880

[15]Alam and Mohanty et al.(2023) "Predicting Students' Performance Employing Educational Data Mining Techniques, Machine Learning, and Learning Analytics.10.1007

D.CERTIFICATE

paytm

Money Sent Successfully

₹9,500 

Rupees Nine Thousand Five Hundred Only

 ANALYZING ACADEMIC
PERFORMANCE OF STUDENT USING
ML

To:

XXXXXXXXXXXXXXXXXXXXXXXXXXXX
XXXXXXXXXXXXXXXXXXXXXXXXXXXX
Indian Bank - 4627

X

From: Ms Nagandla Sai Sreeja
Indian Bank - 9093

8

UPI Ref No: 408529961478
02:06 PM, 25 Mar 2024

Powered by UPI

International Conference on Cognitive Computing and Artificial Intelligence (ICCCAI-2024)

Your response has been recorded.

[Submit another response](#)

This content is neither created nor endorsed by Google. [Report Abuse](#) - [Terms of Service](#) - [Privacy Policy](#)

Google Forms