PROJECT DOCUMENTATION

Improving Academic Performance at Lander School through Data-Driven Insights

Introduction

This project was designed to improve academic performance among Lander School students by collecting and analyzing key data points related to their study habits, support structures, mock exam participation, use of technology, and career guidance. By leveraging a comprehensive Google Form survey, this project sought to identify the factors that contribute to student success and provide actionable recommendations for enhancing their exam readiness and overall academic experience. It also streamline data collection, integrate sources, ensure automated processing, and provide insightful reports for decision-making.

Data Collection Process

Survey Design

A Google Form survey was designed to collect a wide range of data from Lander School students. The survey covered multiple aspects of student life that influence exam performance, including:

- **Study habits**: Questions about the number of hours spent studying daily.
- **Support structures**: Questions about parental and teacher involvement in exam preparation.
- **Career guidance**: Questions to assess whether students had clear career goals and their perceived impact on exam performance.
- **Technology use**: Questions about the use of mobile apps and online platforms for studying.
- **Mock exam participation**: Questions to measure the frequency of mock exam usage and its effect on students' confidence.
- **Confidence and stress levels**: Questions to gauge students' confidence in their exam preparedness and levels of exam-related stress.

Link to the Google survey form:
https://docs.google.com/forms/d/e/1FAIpQLSdaucj6vf5KddPoqb8AkTrkvov2eRM871uLzN5DK5A64rnEVQ/viewform

The survey was distributed to the entire student body, ensuring responses from various performance groups (e.g., high performers, average students, and struggling students).

A separate survey was designed for Teachers to collect the student Grade and remarks about the students.

The survey responses were collected over a period of time. Students were asked to provide honest answers about their academic habits, challenges, and emotional well-being.

Over 100 responses were collected, representing a diverse sample of students across different academic departments (Science, Commercial, and Arts)

Data Preprocessing

Data Cleaning:

Before analyzing the data, several preprocessing steps were conducted to ensure the dataset's accuracy and consistency:

- **Handling missing values**: Incomplete responses were either filled where possible or removed to maintain data integrity.
- **Standardizing responses**: Responses for categorical data such as "confidence levels" and "mock exam frequency" were standardized to ensure uniformity.
- **Converting categorical data**: Certain survey responses (e.g., "Yes/No" answers, study hours) were converted into numerical or binary formats to facilitate analysis.

Data Transformation:

- **Categorizing Confidence Levels**: Responses to questions like "How confident do you feel about your exams?" were categorized into "Low", "Medium", and "High" confidence levels based on a 1-5 scale.

- **Mock Exam Participation Frequency**: Responses about mock exam frequency were categorized into intervals ("Never", "Sometimes", "Often") for clearer analysis.
- **Study Habits**: The question about daily study hours was grouped into categories like "Less than 4 hours", "4-6 hours", "6-8 hours", and "More than 8 hours".
- **Data Splitting:** Some responses like the Exam being prepared for were comma-separated (ie WAEC, NECO, JAMB, School Exams) were separated for analysis.

DATA ANALYSIS

Study Habits Analysis:

Findings: A significant portion of students performing below average reported studying for less than 1 hour per day, indicating insufficient exam preparation time for many students. Also based on the sleep habits of students during exam season a larger portion of High Performing students tend to sleep less than 4 hours while below-average students tend to sleep for a longer duration this period.

Insight: Students with poor study habits are likely at a disadvantage and may require structured study plans or workshops to improve their time management and study efficiency. Adapting the study habits of the Straight A Student may have a positive impact on the performance of Below Average Students.

Recommendations: Implement structured study plans and academic workshops to help students with poor study habits manage their time more effectively.

Support Structures Analysis:

Findings: There were noticeable disparities in the level of guidance involvement. Some students reported receiving strong support, while others most especially the Below Average students reported little to no involvement.

Insight: Students with higher parental and teacher engagement performed better in their studies. Lack of support is a common challenge among struggling students, particularly those in the "Below Average" performance category.

Recommendation: Increase parental engagement through regular communication and progress monitoring. Guidance should be enlightened on the role they play in helping their ward improve their academic performance.

Career Guidance and Goal Clarity:

Findings: A large number of students lacked clear career goals. Some of them felt that having a clearer career goal could positively impact their exam performance.

Insight: Career uncertainty may contribute to reduced motivation and focus, leading to poorer academic outcomes. A larger portion of the high performing student has clarity on their future plan after school.

Recommendation: Career counseling should be provided to help students clarify their career paths, boosting motivation and focus on their studies. Providing career guidance could help students align their academic goals with future aspirations.

Technology Use for Studying:

Findings: Many students reported using mobile apps and online platforms for studying, though the frequency of use varied widely.

Insight: Students who frequently used online study tools reported higher confidence in their exam preparation. Promoting effective digital study tools could further enhance exam readiness.

Recommendation: Based on the student study style preference accessibility to those study material through digitized means could boost their performance. Encouraging the use of effective study apps and online platforms to improve exam readiness.

Mock Exam and Practice Tests Analysis:

Findings: Students who took mock exams more frequently had higher confidence levels compared to those who rarely or never participated in mock exams.

Insight: Regular mock exams are strongly correlated with improved student confidence, and therefore, higher exam performance. Encouraging more frequent mock exam participation is crucial for student success.

Recommendation: Schedule more regular mock exams and provide feedback to students to help them improve their exam-taking strategies.

Confidence and Stress Analysis:

Findings: Confidence levels were lower among students who were rated as "Below Average". Exam stress was also higher in this group.

Insight: Students with lower confidence need additional academic and emotional support to reduce exam-related anxiety and boost their performance.

Recommendation: Introduction of stress management workshops and offering emotional support for students struggling with exam-related anxiety. Teachers and Guidance should also be oriented on pressure they mount on students that could lead to stress and depression that could affect their performance.

Predictive Modeling Using Teacher Feedback

Teacher Feedback Dataset

The **teacher feedback dataset** contained qualitative and quantitative assessments of students, including:

- **Teacher-predicted JAMB scores**.
- **Observations on student study habits and classroom behavior**.
- **Assessment of student preparedness for final exams**.

This feedback was used in combination with the student survey data to build a predictive model that forecasts students' performance in their JAMB exams.

Predictive Model Development

Model Objective: The objective is to build a machine learning model to predict candidates' JAMB (Joint Admissions and Matriculation Board) scores based on their previous academic performances. The model maps grades obtained in various subjects to numerical values and uses these values to predict a student's potential score in the JAMB exam.

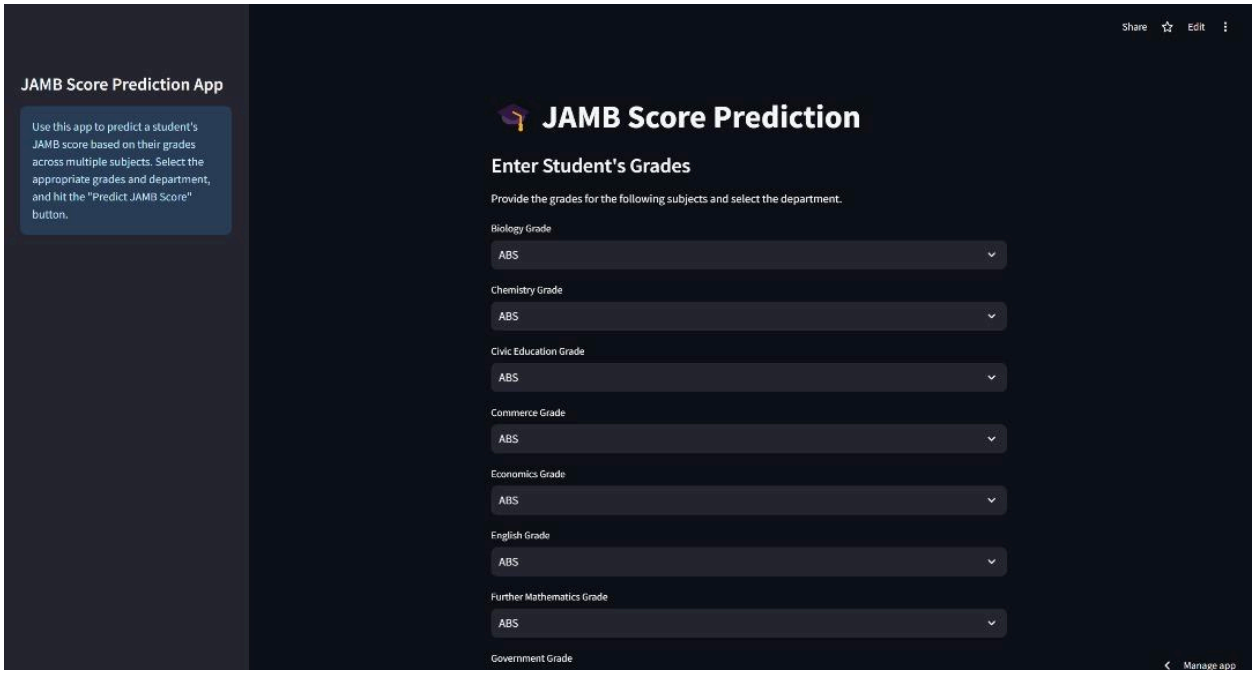
Model Features: The key features used in the JAMB score prediction model are the grades obtained in various subjects by the students. These grades are first mapped to numerical values, and the resulting numerical data serves as input features for the machine learning model.

Model Type: A regression model was used to predict continuous JAMB scores based on the input features.

Model Results and Accuracy:

The Random Forest model performed the best, achieving the lowest RMSE. This suggests that Random Forest would be the most effective model for predicting synthetic JAMB scores from the available data.

The predictive model was deployed to Streamlit app and can be accessed via this link: https://jambpredict.streamlit.app/



Insights from the Predictive Model:

By predicting students' future performance in standardized exams like JAMB, educational institutions can tailor personalized learning plans. For students predicted to score lower, more attention could be given to subjects where their performance might need improvement, which could potentially uplift their overall JAMB score.

Designing an Enterprise Data Solution for Lander School

The objective is to streamline data collection, integrate sources, ensure automated processing, and provide insightful reports for decision-making.

1. Data Collection

1.1. Sources of Data

- **Student Surveys**: Data on study habits, technology usage, career guidance, mock exam participation and other relevant data of students collected via Google Forms.
- **Teacher Feedback**: Teacher evaluations of student performance.
- **Student Academic Performance**: Data on exam scores (JAMB, mock exams, etc.).

- **Other School Systems**: Administrative systems that track attendance, activities, behavior, etc.

1.2. Tools for Data Collection

- **Google Forms/Sheets**: For surveys and feedback collection.
- **Excel/CSV Imports**: For historical data (exam scores, attendance, etc.).

2. Data Pipeline Architecture

2.1. Pipeline Components

The pipeline handles the movement and transformation of data from various sources into a central data warehouse:

Data Ingestion:

Batch Ingestion: Import student survey data and teacher feedback at regular intervals (by term).

Data Transformation:

Data Cleaning: Handle missing values, normalize fields (e.g., "Yes/No" to binary 1/0), and ensure data consistency.

Feature Engineering: Create new variables such as confidence levels, study habits categorization, or derived metrics (e.g., mock exam participation rates).

Data Storage:

A **data warehouse** built on cloud services (e.g., Azure SQL Database) serves as the central storage location for all collected data.

Data from multiple sources will be **ETL'd** (Extract, Transform, Load) into structured tables for easier querying.

Tools and Technologies

- **ETL Tools**: **Apache NiFi** platform is used for seamless extraction, transformation, and loading of data into the warehouse.
- **Data Governance**: Rules Implementation for managing data integrity, security, and privacy (especially for student records).

3. Automation

3.1. Scheduling and Triggering

- **Automated Data Pipelines**: Set up scheduled batch jobs using **Apache Airflow** to ingest data every term.
- **Real-time Processing**: Implementation of a real-time streaming platform like **Apache Kafka** for immediate data ingestion and processing (e.g., capturing new student records).

3.2. Automated Alerts

- **Trigger-based alerts** for key events:

Performance Alerts: If a student's predicted score falls below a threshold, send automated notifications to teachers and parents.

Study Habit Alerts: Identify students studying less than a recommended threshold and trigger notifications to academic counsellors.

4. Data Warehousing

4.1. Centralized Data Warehouse

- A **cloud-based data warehouse** will store data from all sources (student surveys, teacher feedback, academic performance).
- Implement a **star schema** with fact tables for student performance and dimension tables for student profiles, study habits, and support structures.

4.2. Security and Access Control

- **Role-based access**: Ensure different user groups (e.g., teachers, administrators) have the appropriate level of access to the data.
- **Encryption and Compliance**: Encrypt sensitive student data (e.g., grades) and ensure compliance with educational data privacy laws.

5. Reporting

5.1. Visualization Tools

- **Power BI**: Create automated dashboards and reports showing student performance, mock exam participation, support structures, and career guidance trends..

5.2. Key Reports and Dashboards

- **Performance Dashboards**: Highlight top performers, students at risk of failing, and predicted JAMB scores based on the predictive model.
- **Behavior and Support Dashboards**: Show the correlation between teacher feedback, parental involvement, and student performance.
- **Study Habit Reports**: Display trends in study habits and their relationship with exam success.
- **Mock Exam Participation**: Track the number of mock exams taken by students and the resulting confidence levels.

5.3. Automated Reporting

- Set up **automated report generation** to send performance updates at intervals to teachers and school administrators.
- **Scheduled email reports**: Automatically send parents and guardians reports about their child's academic progress and exam readiness.

6. Scalability and Future Expansion

6.1. Scalability

The data pipeline will be designed to scale as the school grows:

Add new data sources (e.g., student behavioural data) without modifying the existing pipeline.

Scale storage and processing power based on the number of students or the complexity of reports using cloud platforms.

6.2. Machine Learning Integration

- As the pipeline matures, integrate more sophisticated **machine learning models** to predict long-term academic outcomes, identify patterns in behaviour, or personalize learning recommendations.

Conclusion

The integration of student survey data and teacher feedback provided a comprehensive understanding of the factors influencing academic performance at Lander School. The predictive model, built on these insights, offers valuable foresight into students' likely JAMB scores, allowing for proactive interventions to improve student outcomes. By implementing the recommendations, the school can expect to see improved academic performance, higher confidence levels, and better exam preparedness among its students. The enterprise data solution will provide Lander School with a robust, automated system for collecting, processing, storing, and reporting student data. By integrating data from surveys, teacher feedback, academic performance records, and other systems, the school will gain powerful insights to guide interventions and improve student outcomes. The solution will be flexible enough to scale and expand as new data sources and reporting needs emerge.