Student Performance prediction
DEPI

- **Project members**

  - Yousef Mohamed
  - Solange Hany
  - Mohamed Emad
  - Aliyah Mostafa
  - Nouran Mostfa
  - Mahmoud Mahdy

Introduction

This project leverages **student data analytics** to gain meaningful insights into academic performance and potential predictors of success. The dataset integrates key indicators such as **IQ scores, student grades,** and **parents' academic degrees**, providing a comprehensive view of the various factors that may influence learning outcomes.

The objective of the project is to explore correlations and patterns within these data points to better understand how cognitive abilities, academic achievements, and family educational backgrounds contribute to student performance. These insights can be used to **identify at-risk students**, **optimize teaching strategies**, and **enhance academic support systems** to foster improved outcomes across educational environments.

This project combines **data engineering techniques with analytics**, laying the foundation for **evidence-based decision-making** in the education sector and presenting new opportunities for data-driven education policies

● Objectives

1. **Analyze correlations** between students' IQ levels, grades and parents' academic degrees to understand how these factors influence academic performance.

2. **Identify patterns** that can predict student success or underperformance based on the provided data.

3. **Develop insights** to assist educators and administrators in improving teaching strategies and academic support systems.

4. **Build a machine learning model** to predict students' future performance or identify those at risk of academic failure, enabling proactive interventions.

5. **Provide recommendations** for personalized support strategies to improve individual student outcomes.

6. **Create a scalable and efficient data pipeline** that integrates multiple datasets while maintaining data quality and integrity.

7. **Visualize findings** through dashboards or reports to ensure stakeholders can easily interpret the results.

8. **Ensure privacy and compliance** by managing student data responsibly, following ethical data-handling practices.

● Stakeholders

1. School Administrators

- **Use insights to shape academic policies and improve school performance.**
- **Make data-driven decisions for resource allocation and student support programs.**

2. Teachers and Educators

- **Understand students' learning needs and adjust teaching strategies accordingly.**
- **Identify at-risk students early and provide targeted interventions.**

3. Students

- **Benefit from personalized learning plans based on predictive analytics.**
- **Gain better academic support through the identification of areas for improvement.**

4. Parents and Guardians

- **Receive insights about their child's academic strengths and areas needing support.**
- **Collaborate with educators to create an environment that promotes academic success.**

5. Education Policy Makers

- **Utilize aggregated data insights to inform regional or national education policies.**
- **Promote evidence-based practices across different schools and districts.**

- **Programming languages**
  1- Python
  2- SQL

- **Frameworks**

  - SQL Server Management System (SSMS)
  - SQL Server Integration System (SSIS)
  - Azure Blob Storage
  - Azure Workspace
  - Power BI

• *Week 1*

Planning the Database:

- I started by **designing a SQL database schema** that captures all the necessary data points: students' personal information, IQ scores, grades, and their parents' academic degrees.
- I ensured that the relationships between tables were clear, linking them using **Primary Keys** (e.g., Student_ID) and **Foreign Keys** (e.g., Parent_Degree_ID).

Implementing the Database:

- I used **Microsoft SQL Server** to create the schema and set up the tables.
- After setting up the structure, I populated the tables with **sample data** to test how the queries and relationships would function.

Writing SQL Queries:

- I created **SQL queries** to extract essential information such as average grades and explored correlations between IQ scores and student performance.

Overcoming Challenges:

- I spent time ensuring data integrity by debugging **relationships between tables** and fixing issues with missing or incorrect data.

Tools Used:

- Microsoft SQL Server, SQL Management Studio.

Deliverables

- A well-designed SQL schema with populated tables.
- SQL queries for basic data extraction and analysis.

- # Week 2

Creating a Data Warehouse:

- I set up a **Data Warehouse** to consolidate the different datasets, making it easier to run analyses across multiple variables.
- I used **ETL techniques** to extract data from the SQL database, transform it (e.g., cleaning data), and load it into the warehouse.

Writing Python Scripts:

- I wrote **Python scripts** to connect to the SQL database using **SQLAlchemy**. These scripts automated the process of extracting, cleaning, and preparing the data.
- Using **Pandas**, I handled missing values, normalized some columns (like grades), and ensured all the data was consistent.

Challenges and Solutions:

- A key challenge was ensuring **data consistency** between the warehouse and the SQL database. I tackled this by adding **data validation checks** at each step.

Tools Used:

- Microsoft SQL Data Warehouse, Python (Pandas,SQLAlchemy).

Deliverables:

- A functioning data warehouse with integrated student data.
- Python scripts for data extraction and transformation.

- Week 3

Performing EDA:

- I used **Python libraries** such as **Matplotlib** and to visualize trends in the data, like how IQ scores or parental education affect students' grades.
- This analysis gave me a better understanding of the patterns in the data, which I later used for building the machine learning model.

Building a Machine Learning Model:

- Based on the EDA, I chose to build a **Regression model** to predict student performance.
- I trained the model on historical data (e.g., IQ scores and grades) and tested it using different evaluation metrics, such as **accuracy and precision**, to ensure its reliability.

Model Evaluation:

- I compared the model's predictions to real data to check how well it identified at-risk students. I made adjustments to **hyperparameters** to improve the model’s performance.

Tools Used:

- Python (Scikit-learn, Matplotlib), Jupyter Notebooks, Azure Machine Learning.

Deliverables:

- EDA report with charts and insights.
- A trained machine learning model with performance metrics.

- Week 4
  - **Creating Dashboards in Power BI:**
    - I designed **interactive dashboards** in **Power BI** to present key metrics, such as the distribution of students' grades and the percentage of students identified as at-risk.
    - I focused on making the visualizations clear and actionable for stakeholders, such as teachers or school administrators.
  - **Preparing the Final Report and Presentation:**
    - I compiled all my findings into a **comprehensive report**, summarizing the data preparation, model development, and results.
    - I created a **presentation** to showcase the project's outcomes, using visualizations from **Power BI** to engage the audience and demonstrate the impact of the project.
  - **Challenges Faced:**
    - One of the main challenges was **integrating the machine learning model with Auzer Workspace** I resolved this by learning how to deployment the machine learning model into Azure workspace

Tools Used:

- Azure, Power BI

Deliverables:

- Deployed machine learning model.
- Power BI dashboards for data visualization.
- Final project report and presentation.

Conclusion

In this project, we successfully analyzed students' performance data by integrating various factors such as IQ scores, academic grades, and parents' educational backgrounds. Through a combination of **SQL database management, data warehousing, Python-based analytics, and machine learning**, we gained meaningful insights into the relationships between cognitive abilities, academic achievements, and parental education.

The development of a **predictive machine learning model** allowed us to forecast student performance and **identify at-risk students**, providing an opportunity for proactive interventions. Additionally, the use of **Power BI dashboards** enabled us to visualize key metrics effectively, making the findings accessible and actionable for educators, administrators, and other stakeholders.

Throughout the project, we faced challenges in **data integration, model optimization, and deployment**, but by leveraging **Azure services ,** we ensured that the solutions were scalable, maintainable, and efficient.

This project not only demonstrates the potential of **data science and machine learning** in the education sector but also provides a foundation for further research and development. Moving forward, the model and insights generated here can be expanded to include additional variables, making it a valuable tool for **personalized learning strategies** and **policy-making** in education.

Thank you