PROJECT CLOSURE REPORT

By Blessing Ilesanmi

18th October, 2024

PROJECT OBJECTIVES AND ACHIEVEMENTS

Objective
The primary objective of the School Dropout Prediction Project was to develop a predictive model that could identify students at risk of dropping out. This project aimed to assist educational institutions in early intervention by providing insights into the factors contributing to dropout rates, allowing for timely and personalized support to students.

Key Achievements:

- **Data Cleaning and Preprocessing:** The raw dataset provided was unstructured, initially in a semicolon-delimited format. Data cleaning involved restructuring the data into appropriate columns, addressing missing values, correcting inconsistent column names, and converting data types as needed for analysis.
- **Feature Engineering:** Significant features such as academic grades, and financial data were identified and engineered for optimal model performance.
- **Model Development:** Various machine learning models, including logistic regression, decision trees, and random forests, were developed. The best-performing model achieved a predictive accuracy of approximately 77%, indicating reasonable effectiveness in identifying at-risk students.
- **Dropout Dashboard:** A dashboard was created to visualize the data
- **Deployment:** The predictive model was deployed using Streamlit, providing an intuitive app where educators can predict at-risk students. The dashboard allows for real-time insights into student data, facilitating early intervention.

CHALLENGES FACED AND SOLUTIONS IMPLEMENTED

- **Data Quality Issues:** The project encountered several data quality issues during preprocessing. The initial dataset was in a semicolon-delimited format, requiring restructuring into proper columns. The first row was mistakenly treated as data instead of header information and was reassigned accordingly. One column had values in double quotation marks, which were removed for consistency. Additionally, the misspelled column 'Nacionality' was corrected to 'Nationality.' Lastly, several columns initially read as object types were converted to the appropriate numerical types (int or float) to ensure accurate analysis. Addressing these issues was essential for the reliability of the predictive model.

- **Deep Learning Model Proficiency:**
  Another challenge was working with deep learning models, as there was a lack of comfort and proficiency in this area.

FUTURE WORK AND POTENTIAL IMPROVEMENTS

Model Accuracy and Performance:
Future iterations should focus on enhancing model accuracy. Although the current model achieved a 77% accuracy, 23% of the predictions may still be incorrect. Exploring advanced deep learning models could help improve both the accuracy and recall rates.

Incorporating Additional Factors:
The current model lacks certain factors that could contribute to dropout rates, such as mental health, personal circumstances, and peer relationships. Expanding the dataset to capture these variables may provide more comprehensive insights and improve the generalization of the model.

Real-Time Data and Feedback Loop:
In future work, integrating real-time data, such as attendance, behavior, and health information, can enhance the predictive capability of the model. Incorporating a feedback loop to regularly update and retrain the model with new data can improve the model's accuracy over time

CONCLUSION

The School Dropout Prediction Project successfully met its objectives by developing a predictive model that identifies at-risk students with high accuracy. The project faced several challenges, but effective solutions were implemented to overcome these hurdles. Future work will focus on enhancing the model and expanding datasets,