Personal Contribution Report – Erick Banegas

Final Machine Learning Project – Student Academic Performance Prediction
ITAI 1371 – December 2025

1. Introduction

This personal report summarizes my individual contributions to the team project *"Student Academic Performance Prediction System."* My responsibilities focused on four major areas: **Dataset Cleaning and Preprocessing**, **Exploratory Data Analysis Visualization (EDA)**, **Model Training and Evaluation**, and **Overall Documentation Structure**. Each task was essential to ensuring the technical quality, clarity, and reliability of the final machine learning solution.

2. Dataset Cleaning and Preprocessing

I led the initial preparation of the dataset, ensuring that all variables were correctly formatted and ready for modeling. My contributions included:

Data Integrity Verification

- Checked for missing values using .isnull().sum().

- Confirmed the dataset had consistent data types.

- Validated ranges for numerical values such as scores and attendance.

Feature Separation

I categorized features into **numerical** and **categorical**, which later allowed the team to create a clean preprocessing pipeline.

Encoding and Scaling Decisions

I helped design and configure the **ColumnTransformer**, which included:

- **One-Hot Encoding** of categorical variables such as
  gender, lunch_type, parent_education, internet_access, extra_activities

- **Numerical scaling** (StandardScaler) for:
  math_score, reading_score, writing_score, attendance_rate, study_hours

Ensuring Reproducibility

I helped build the preprocessing steps into a pipeline so that the preprocessing was automatically applied during both training and testing.

Impact:
This preprocessing strategy ensured a clean, consistent, and professional workflow, preventing data leakage and guaranteeing fairness across all models.

3. EDA Visualizations

I took responsibility for creating the **Exploratory Data Analysis visualizations**, which allowed the team to understand the dataset deeply.

Visualizations I Created

- Histograms for academic scores, attendance, and study hours

- Boxplots to observe distribution and identify outliers

- Correlation heatmap to study linear relationships

- Violin/boxplots comparing Pass vs Fail categories

- Scatterplots analyzing the correlation between study_hours and scores

- Barplots showing the effects of parent education and gender

Analytical Insights I Identified

- Academic scores were the strongest predictors of final results

- Attendance rate had a strong positive influence on performance

- Higher study hours correlated with higher scores

- Students with parents holding higher education degrees tended to perform better

Impact:
These insights guided correct model selection and helped interpret the final results. The EDA made the project look professional and academically strong.

4. Model Training and Evaluation

I collaborated on the design and execution of the machine learning experiments, contributing to:

Model Setup and Training

I trained several models using pipelines, including:

- Logistic Regression

- Random Forest

- Gradient Boosting

- KNN

- SVM

I helped ensure that each classifier was correctly combined with the preprocessing pipeline.

Evaluation Metrics

I worked on computing:

- Accuracy

- Precision

- Recall

- F1-score

- ROC-AUC

Results Interpretation

I helped analyze and compare model performance.
My key contribution was identifying that:

- Logistic Regression, Gradient Boosting, and XGBoost were the most stable and accurate models.

- The dataset was highly separable, explaining the near-perfect performance.

Impact:
My work ensured that the evaluation was scientifically valid, consistent across all models, and easy to compare.

5. Documentation Structure

I was responsible for structuring and organizing the overall project documentation, including:

Organizing notebook sections

I helped arrange the notebook with clear, professional markdown sections:

- Project overview

- Dataset description

- EDA

- Preprocessing

- Modeling

- Results

- Conclusions

Ensuring readability and consistency

I reviewed markdown text, titles, code separation, and explanations to ensure the notebook was clean and easy to follow.

Contribution to the Team Report

I prepared key parts of the final structure and contributed to ensuring the team report followed proper academic and technical standards.

Impact:
My involvement guaranteed that the final submission was polished, cohesive, and communicated our work clearly and professionally.

6. Conclusion

Throughout this project, I played a central role in the technical and structural execution of the machine learning solution. My contributions to dataset cleaning, preprocessing, visualization, model development, and documentation were essential to the success of the team. I ensured that our workflow was professional, reproducible, and aligned with industry best practices.

This experience strengthened my skills in:

- Data preprocessing

- EDA visualization

- Machine learning pipelines

- Model evaluation

- Technical report writing

Overall, my contributions were instrumental in guiding the team toward a high-quality final project.