**Project Documentation**

**Project Title: *Investigating the Causes of JAMB Failure Rates and Predicting Future Performance Trends (2020–2030)*  
Prepared by:**

* **Odi Chibuzor Greg**
* **Amankwe Amarachi Francisca**
* **Amakor Euphemia Chima**
* **Tina Rufai**
* **Emmanuella Agyapong**

**Date: June 2025**

**1. Workflow and Methodology**

**This project was conducted in the following structured phases:**

1. **Survey Design and Data Collection**
   * **A custom Google Form survey was created to gather data from individuals who sat for the JAMB exam between 2020 and 2025.**
   * **The survey collected details on demographics (gender, year), academic background (school type, study hours), and technology access (device usage).**
   * **Responses totalled 110 students, with a calculated pass mark of 200 to distinguish success from failure.**
2. **Data Cleaning and Transformation**
   * **The raw data was exported as a CSV and cleaned in Excel.**
   * **Features were encoded for modeling: categorical variables like gender and school type were transformed appropriately.**
   * **The outcome variable was binary: *Pass (Score ≥ 200)* or *Fail (Score < 200)*.**
3. **Exploratory Data Analysis (EDA)**
   * **Descriptive statistics and visualizations were used to highlight patterns:**
     + **Pass Rate: 78.2% (86 students)**
     + **Fail Rate: 21.8% (24 students)**
     + **Average Score: 227.7**
     + **Females slightly outperformed males (77.1% vs. 71.7%)**
     + **Mission/Religious schools had the highest pass rate (77.8%)**
     + **Study time >20 hours/week correlated with the highest pass rate (84.6%)**
   * **Device access (laptop/tablet) showed minimal impact on performance.**
4. **Predictive Modeling**
   * **A Random Forest Classifier was used for prediction due to its flexibility and feature importance insight.**
   * **Model training and testing were performed in a Jupyter Notebook environment.**
   * **Top predictive features:**
     + **Exam Score (most dominant)**
     + **Study Hours**
     + **School Type**
   * **Performance metrics (Precision, Recall, F1-Score) were to be confirmed in final model evaluation.**
5. **Reporting and Recommendations**
   * **Insights were summarized in a PowerPoint presentation, with clear visualizations on gender, year, school type, study behavior, and device usage.**
   * **Recommendations were based on the data and model outcomes.**

**2. Tools and Technologies Used**

| **Component** | **Technology/Tool** |
| --- | --- |
| **Survey Collection** | **Google Forms** |
| **Data Analysis** | **Python (pandas, numpy)** |
| **Visualization** | **seaborn, matplotlib** |
| **Machine Learning Model** | **scikit-learn (Random Forest)** |
| **IDE/Notebook** | **Jupyter Notebook** |
| **Presentation** | **Microsoft PowerPoint** |

**3. Key Decisions Made**

* **Score Threshold of 200 was used to classify Pass vs. Fail.**
* **Random Forest was selected for its robustness and ease of interpretation.**
* **Study hours and school type were prioritized for deeper exploration based on initial correlation with results.**
* **Visual storytelling in PowerPoint was emphasized to improve communication of technical findings.**

**4. Challenges and Mitigations**

| **Challenge** | **Mitigation** |
| --- | --- |
| **Low survey participation early on** | **Encouraged broader participation via social media/student groups** |
| **Slight gender imbalance in responses** | **Included gender-neutral and "Other" categories for fairness** |
| **Missing values or inconsistencies** | **Cleaned data using pandas; filled missing fields when appropriate** |
| **Model performance tuning required** | **Applied grid search and feature selection strategies** |

**5. Future Improvements**

* **Expand Sample Size: To include a wider demographic and regional representation.**
* **Integrate Official JAMB Data: Partner with educational institutions or JAMB body.**
* **Deploy a Web Dashboard: For real-time prediction and trend updates for stakeholders.**
* **Improve Model Evaluation: Finalize confusion matrix and performance metrics.**