Zaio Technology (Pty) Ltd
9 Lower Burg Street,
Cape Town City Centre, Cape Town,
South Africa, 8000

Objective

The objective of this project is to analyse student habits and their effect on academic performance using machine learning techniques. Students will work with the student_habits_performance.csv dataset to perform data preprocessing, exploratory data analysis, predictive modelling, and interpretability through explainable AI (XAI). The project will reinforce concepts such as data cleaning, visualisation, classification, model evaluation, and explainability.

Timeline and Milestones (3 Weeks)

Week	Focus	Milestones
Week 1	**Data collection and preparation**	- Import dataset into Jupyter Notebook - Handle missing values - Encode categorical variables - Standardise/normalise numeric variables - Create cleaned dataset (students_clean.csv)
Week 2	**Data exploration and model building**	- Perform EDA with 2D & 3D plots (scatter, stacked bar, box/violin, heatmap, parallel coordinates, radar) - Train Decision Tree and Logistic Regression models - Compare model performance (accuracy, classification reports)
Week 3	**Model evaluation and explainability**	- Evaluate models using confusion matrices, ROC curves, and AUC - Apply SHAP (global importance), LIME (local explanations) - Summarise insights and prepare presentation

Week 4	**Finalisation and submission**	- Finalise cleaned dataset, notebook, and slides - Deliver presentation - Submit deliverables in required format

---

Tools & Technologies

- **Programming Language**: Python
- **Environment**: Jupyter Notebook/ Visual studio code
- **Libraries**: pandas, numpy, matplotlib, seaborn, plotly, scikit-learn, shap, lime
- **Collaboration Tools**: Google Classroom
- **Dataset Source**: student_habits_performance.csv

Team Structure

- **Team Size:** Individual assignment
- **Team Roles:** You will take on the following roles:
    - **Data Engineer:** Handles cleaning and preprocessing
    - **Model Engineer:** Focuses on training and evaluating models
    - **Analyst:** Exploratory data analysis and visualisation
    - **Presenter/Documenter:** Prepares final report and/or demo

---

Deliverables

1. **Technical Notebook / Report**
    - Clean and documented Jupyter Notebook (.ipynb)/Visual studio code
    - PDF report including:
        - Data Collection and preparation
        - Data Exploration
        - Model building and evaluation
        - Explainability
        - Conclusion
    - Exported cleaned dataset (students_clean.csv)

2. **Final Presentation (5-10 minutes)**
    - Overview of dataset and problem
    - Key findings from analysis
    - Model performance results
    - Lessons learned and recommendations
3. **Submission Package**
    - ZIP file containing: Notebook, PDF report, cleaned dataset
4. Bonus points
    - GitHub repo with organized code
    - GitHub repo with all the documentation

Evaluation Criteria

Assessment will be based on:

- **Data Collection**: Correct import, dataset summary, missing values identified
- **Data Preparation**: Proper handling of missing data, encoding, normalisation, justification provided
- **Data Exploration**: Quality of visualisations and insights from 2D and 3D plots
- **Model Building**: Decision Tree and Logistic Regression trained, evaluated, and compared
- **Model Evaluation**: Confusion matrices, ROC curves, and AUC with interpretations
- **Explainability**: Use of SHAP, LIME with clear insights
- **Report**: Clear, structured report with conclusions and reflection

---

Submission Guidelines

- Final notebook/report: Upload to Google Classroom
- Attach the following files:
    - Jupyter Notebook (.ipynb)/ Visual Studio code
    - PDF Report
    - Cleaned dataset (students_clean.csv)
    - ZIP folder containing all work (Project_YourName).
- Presentation: Delivered live on final day
- All work must be submitted before the final session for evaluation.