EGT207 Machine Learning Project

Objective:

To develop ML models to solve a real-world problem with the given dataset, and to apply the relevant techniques to evaluate and optimize the model and efficiency.

- Develop ML models for the given dataset and problem definition.
- Evaluate ML models using appropriate metrics for efficiency and effectiveness.
- Apply model tuning and optimization to improve the performance of ML models.
- Compare the different ML models on their performance with respect to the given dataset and problem definition.

Project Tasks:

1. Dataset: **parkinson_disease_assignment.csv**
2. Develop **Two ML models** for the classification task in a **single Colab notebook**.
   - The first model is **Decision Tree**.
   - The second model is a **self-selected Ensemble Method**. You are expected to select an appropriate Ensemble Method based on the given dataset and task.
   - Add your adminNo as the prefix of all ML models' name.
3. Manage the Colab notebook with below sections:
   - Data preparation
   - Build Decision Tree model
   - Summary of Decision Tree model development (<150 words): to describe your key considerations and concerns on design, evaluation, optimization, regularization, etc.
   - Build Ensemble Method model
   - Summary of Ensemble Method model development (<150 words): to describe your key considerations and concerns on model selection, design, evaluation, optimization, regularization, etc.
   - Comparison of the 2 models (Decision Tree and Ensemble Method) (<200 words): to compare and explain their results and performance, pros/cons, suitability for the given task, any other concerns.
   - Keep all running results within the notebook.

Submission:

- Submit a single Colab notebook (admin_project.ipynb) via Project Submission (Week17) link in Brightspace.
- Deadline: By end of Week 17 (23:59 on 11 Aug 2024).

*This is an individual project. ZERO mark will be given if fully copy is found!

Assessment Rubrics

Marks will be allocated based on the task completeness, model development, techniques, model performance and the summary and comparison:

Items	Need Improvement	Satisfactory	Excellent
Data Preparation (15%)	Poor preparation: ● data review and visualization ● data correction ● data splitting ● data engineering	Satisfactory preparation: ● data review and visualization ● data correction ● data splitting ● data engineering	Excellent preparation: ● data review and visualization ● data correction ● data splitting ● data engineering
Decision Tree (20%)	Poor development: ● training ● evaluation ● optimization and regularization ● performance and results	Satisfactory development: ● training ● evaluation ● optimization and regularization ● performance and results	Excellent development: ● training ● evaluation ● optimization and regularization ● performance and results
Summary of Decision Tree (10%)	Insufficient and unclear description and explanation: ● evaluation and optimization ● overfitting techniques ● results and performance	Insufficient and clear description and explanation: ● evaluation and optimization ● overfitting techniques ● results and performance	Sufficient and clear description and explanation: ● evaluation and optimization ● overfitting techniques ● results and performance
Ensemble Method Model (20%)	Poor development: ● selection ● training ● evaluation ● optimization ● overfitting techniques ● performance and results	Satisfactory development: ● selection ● training ● evaluation ● optimization ● overfitting techniques ● performance and results	Excellent development: ● selection ● training ● evaluation ● optimization ● overfitting techniques ● performance and results
Summary of Ensemble Method model (10%)	Insufficient and unclear description and explanation: ● model selection ● evaluation and optimization ● overfitting techniques ● results and performance	Insufficient and clear description and explanation: ● model selection ● evaluation and optimization ● overfitting techniques ● results and performance	Sufficient and clear description and explanation: ● model selection ● evaluation and optimization ● overfitting techniques ● results and performance
Comparison of 2 models (15%)	Comparisons are not clear: ● results and performance ● pros/cons ● reasoning for decisions ● comprehensive views and supporting evidence ● technical knowledge ● understanding of problem	Comparisons are insufficient: ● results and performance ● pros/cons ● reasoning for decisions ● comprehensive views and supporting evidence ● technical knowledge ● understanding of problem	Comparisons are clearly organized and conveyed: ● results and performance ● pros/cons ● reasoning for decisions ● comprehensive views and supporting evidence ● technical knowledge ● understanding of problem
Execution Performance (10%)	Poor execution performance: ● errors in notebook ● execute correctly ● notebook developed in sequences ● originality & creativity ● new ideas or insights	Satisfactory execution performance: ● errors in notebook ● execute correctly ● notebook developed in sequences ● originality & creativity ● new ideas or insights	Excellent execution performance: ● errors in notebook ● execute correctly ● notebook developed in sequences ● originality & creativity ● new ideas or insights