Project Milestone-2

April 8, 2025

Overview:
Building on the initial decision tree baseline, this phase focused on strengthening preprocessing and expanding the model repertoire to improve classification of malicious Android apps using system call fingerprints.

Preprocessing Enhancements:
Implemented two robust preprocessing pipelines:

- **Log-transform + Standard Scaler**: Stabilizes variance and centers features.
- **Log-transform + MinMax Scaler**: Scales features to a fixed range while preserving relative distances.

These pipelines significantly improved model convergence and performance.

New Models Implemented:

- **Perceptron Variants**: All 3 variants of perceptron implemented.
- **AdaBoost**: Delivered the best test performance with strong generalization.
- **Support Vector Machine (SVM)**: Balanced training and test performance.

Evaluation Results:

- **Margin-Perceptron**
  Test Precision: 0.758 Recall: 0.748 F1-score: 0.753
- **AdaBoost**
  Test Precision: 0.861 Recall: 0.851 F1-score: 0.856
  Kaggle Score: 0.86
- **SVM**
  Test Precision: 0.769 Recall: 0.835 F1-score: 0.801

Challenges Faced:

- Integrating pipeline-based preprocessing into the existing hyperparameter tuning framework.
- F1-score plateauing around 0.85 despite extensive model and preprocessing upgrades.

Next Steps:

- Implement and evaluate a **feed-forward neural network** using established ML libraries to further boost classification performance.