Project Milestone 2

1. Updates After the First Milestone

Since the initial milestone, I have focused on experimenting with Support Vector Machines (SVM) and evaluating different feature selection strategies. Specifically, I applied mRMR (Minimum Redundancy Maximum Relevance) and PCA (Principal Component Analysis) and tested various kernels, including linear, polynomial, and RBF. Through these experiments, I found that using an RBF kernel with mRMR (selecting 130 features) achieved the highest Kaggle score so far (0.81).

2. Challenges Faced

- **Feature Selection:** Selecting a suitable number of features proved challenging. Striking a balance between reducing dimensionality and preserving informative features required multiple trials.
- **Kernel Tuning:** Identifying the best kernel for SVM involved tuning several hyperparameters (e.g., C, gamma) and comparing performance across linear, polynomial, and RBF kernels.
- **Computational Costs:** Repeated model training with large feature sets and multiple methods (mRMR, PCA, etc.) was time-consuming and required efficient use of computational resources.

3. Plan for the Rest of the Semester

- **Model Ensemble:** I plan to combine Decision Trees, Decaying Perceptrons, and SVM into an AdaBoost ensemble. This approach will allow me to leverage the strengths of each model for improved classification accuracy.
- **Further Feature Engineering:** I will continue to refine feature selection methods, as well as investigate potential data preprocessing strategies (e.g., normalization or scaling) to ensure robust model performance.
- **Evaluation and Optimization:** In addition to Kaggle scores, I will incorporate cross-validation and detailed error analysis. This will help identify overfitting, guide hyperparameter tuning, and reveal specific areas for further improvement.