Dev Makwana & Kaavya Borra: Project Proposal

**Motivation:**

The Random Forest algorithm treats all attributes equally when building decision trees, which can reduce performance on datasets containing attributes of varying importance. By implementing attribute weighting, we aim to tackle this limitation by assigning a higher weightage to more critical attributes and reducing the influence of less relevant ones. This can lead to more accurate predictions and faster model convergence.

**Method:**

We plan to improve the Random Forest algorithm by implementing an attribute weighting method based on a pre-training analysis of the attributes. Before building the decision trees, we will use statistical methods, such as Gini importance. Each attribute will be assigned a relevance score, which will be used to adjust attribute selection probabilities, valuing higher weighted attributes over lower weighted whens when choosing an attribute to split a branch on.

**Intended Experiments:**

1. Baseline Comparison:
   1. Train a standard Random Forest model without any attribute weighting on benchmark datasets.
   2. Train the improved model using the proposed feature-weighting approach.
2. Evaluation Metrics:
   1. For classification models, we will use accuracy, F1-score, precision, recall, and AUC/ROC.
   2. For regression models, we will use root mean squared error (RMSE), mean absolute error (MAE), and R-squared.
3. Ablation Study:
   1. We will measure the change in model performance by removing the weighting mechanism to assess its direct impact.
4. Computational Efficiency:
   1. We will measure the training time and memory usage to ensure that the approach is scalable and that the improved model creates decision trees that converge faster than standard decision trees.