Ensemble 2 - Implementation

Junaid Mahmud

Data Import

Data was uploaded to the drive and then imported

Data Preprocessing

The following preprocessing was done.

1. Data Encoding of the Categorical Variables.
2. Data Shuffling
3. Scaling
    a. Standard Scaling
    b. Min Max Scaling
4. No Null Value Found
5. Outlier Removal

Ensemble Building

The following estimators were used:

1. Random Forest
2. AdaBoost

Experimentation Details

The experiment was done on the data as follows:

1. Data with no scaling
2. Standard Scaled Data
3. Min Max Scaled Data

The **F1 Score** of the ensemble along with the estimators are as follows

Before Outlier Removal

	Random Forest Classifier	Adaptive Boost
No Scaling	79.93%	85.66%
Standard Scaler	94.11%	88.48%
Min Max Scaler	80.00%	88.33%

After Outlier Removal

	Random Forest Classifier	Adaptive Boost
No Scaling	88.56%	85.66%
Standard Scaler	76.66%	91.42%
Min Max Scaler	82.73%	91.42%

Result Discussion

- When outliers were not removed
    - Random forest was better with 94.11% F1, and AdaBoost was with 88.48% F1. When a Standard scaler was used.
- With outliers removed
    - Random Forest did well when no scaling was done.
    - Adaptive Boost outperformed Random Forest Classifier in all cases. It did the best with a scaled data.