Nikunj Wadgeri, Brayan Paredes, Octavio Morales

CS 559-WS GROUP 5

Final Project Report

I pledge my honor that I have abided by the Stevens Honor System.

**– Section 2: Report the number of subgroups and their unique characteristics.**

There are three subgroups for the three group members. Octavio is doing subgroup 0, which has a k value of 7 and using 35 features. Brayan is doing Subgroup 1, which has a k value of 5 and using 38 features. Nikunj is doing subgroup 2, which has a value of 3 and using 40 features.

**– Section 3 and 4: Construct a table as shown in Table 2. In the notebook file, present the**

**length of the train data column and the confusion matrix.**

| Subgroup ID | Name of Student | Average Accuracy Score Base Models [TT(TF)] | Accuracy Score Meta Model [TT(TF)] | Accuracy Score K-Fold CV [TT(TF)] |  |
| --- | --- | --- | --- | --- | --- |
| 0 (K=7) | Octavio | 0.9432 [298(492)] | 0.9681 [1117(8)] | 0.8788 [986(32)] | 35 |
| 1 (K=5) | Brayan | 0.9371  [228(640)] | 0.9690  [1119(7)] | 0.8784  [988(32)] | 38 |
| 2 (K=3) | Nikunj | 0.8850 | 0.9964 | 0.8794 | 40 |

**– Section 5: Explain the best model from Sections 3 and 4.**

In the context of the results provided, Nikunj’s model showcases a K-Fold CV accuracy close to 1 (0.8794), suggesting that his model has the best generalization capabilities in comparison to Octavio’s and Brayan’s accuracies (~0.878). This high score indicates that the model accurately predicts outcomes on various subsets of data, showing how it is better for data variations and maintaining consistent performance across different data folds. Such a performance level speaks to the model’s reliability and its potential effectiveness in real-world applications where similar data conditions are expected. Although the Score Meta Model is consistent, doing K-Folds shows how increasing the number of features can drastically increase the accuracy, showing the better utilization of data. In general, K-Folds CV can be seen as advantageous due to their reduction in bias, as the model is less likely to be biased towards any particular segment of the data.

**– Section 6: Report the accuracy scores, acctrain, the result of the best model selected in Section 5, and the number of features N features, used in the best k-fold cross validation model in Section 5. These values will be used for the ranking score.**

FOR NIKUNJ:

Accuracy score of stack model (section 3) - 0.9664

Accuracy score of k fold cv model (section 4) - 0.8794 {**k = 3, n\_features = 40**}

FOR OCTAVIO:

Accuracy score of stack model (section 3) - 0.9681

Accuracy score of k fold cv model (section 4) - 0.8788 {**k = 7, n\_features = 35**}

FOR BRAYAN:

Accuracy score of stack model (section 3) - 0.9690

Accuracy score of k fold cv model (section 4) - 0.8784 {**k = 5, n\_features = 38**}