**Overall Report for my Project:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **S. No** | **Model** | **Accuracy**  **(In Sample)** | **Kaggle Score (Out sample)** | **Comments** |
|  | Logistic Regression | 82.27% | 76.06% | For the first model this gave a pretty good start, So I submitted this on Kaggle to set a benchmark. |
|  | K-Nearest Neighbors | 73.625% | NA | Not that great accuracy scores so avoided it. |
|  | Naïve Bayes | 46.58% | NA | Lowest accuracy score of all. |
|  | Stochastic Gradient Descent | 71.92% | NA | Not that great accuracy scores so avoided it. |
|  | Decision Tree Classifier | 79.91% | NA | Not that great accuracy scores so avoided it. |
|  | AdaBoost with Decision Tree | 80.35% | NA | Although it had good accuracy, but it was not the best. So, I did not submit it on Kaggle. |
|  | Random Forest | 83.27% | 77.99% | I submitted various times in Kaggle using different sets of hyper parameters using grid search and came to this result. |
|  | **Extra Trees classifier** | **82.49%** | **80.38%** | **This had the best result on Kaggle website. I submitted various times in Kaggle using different sets of hyper parameters using grid search and came to this result.** |
|  | SVM | 81.59% | NA | Although it had good accuracy, but it was not the best. So, I did not submit it on Kaggle. |
|  | **Ensemble Modeling (Random Forest, Extra Trees classifier, Logistic Regression)** | **85.63%** | **79.90%** | **I tried combinations of different models but the best result I got was from these three. Although it’s not the best I think this reduced the high bias between the models and gave the reasonable model.** |