Justifications for choosing Classifiers

Naïve Bayes : Because it is super simple, requires less computations and it needs less training data.

Logistic Regression : Because it can be scaled in case we are able to find more data.

SVM : High accuracy.

KNN : Since we didn’t have priori knowledge of the classifier parameters.

Parameters for the Classifiers

KNN: Chose k=3, because it was not too large, hence did not require too many computations and gave a better accuracy than higher k values.

Naïve Bayes: All the parameters of the dataset were used because the dataset was relatively small and reducing the parameters further would make the data a little inadequate to have a good model. (It gave us a lower accuracy when we reduced the number of parameters).

Logistic Regression: Retained the parameters with the lowest coefficients of correlation because higher coefficients indicate that the two parameters are somewhat identical and represent similar information.

SVM performed better than all the other classifiers because:

1. It is more robust
2. It is computationally efficient
3. This was a two-class classifier problem and SVM generally performs best for these data.
4. The number of features are less as well.

KNN performed somewhere in the middle because:

1. It has a low bias.

Naïve Bayes performed the worst among all the classifiers because:

1. The classifier parameters were very few in number. (They perform well when there are more parameters).