The features created from the data set are

Fakeness = no of occurrence of a word in fake review / no of total occurrence of the word in the data set

Fakeness without bow = no. of occurrence of a non-stop word in fake review / total number of occurrence of the non-stop word

Fakeness of review content = product of fakeness of each words (fake ness of stop words = 1)

Review length = total number of words present in the review content

No of Capitalized review words = total numbers of words with beginning capitalized / total no of words in the review content

The features related to the hotels were reducing the prediction score in cross validation as well as in the

Classifiers

Following Classifiers were used:

SVM

Adaboost

Random Forest

Gaussian Naïve Bayes

Random Forest showed a better result against all other methods. On pruning the parameters of the Random Forest it was found that 68 trees could provide the best score.