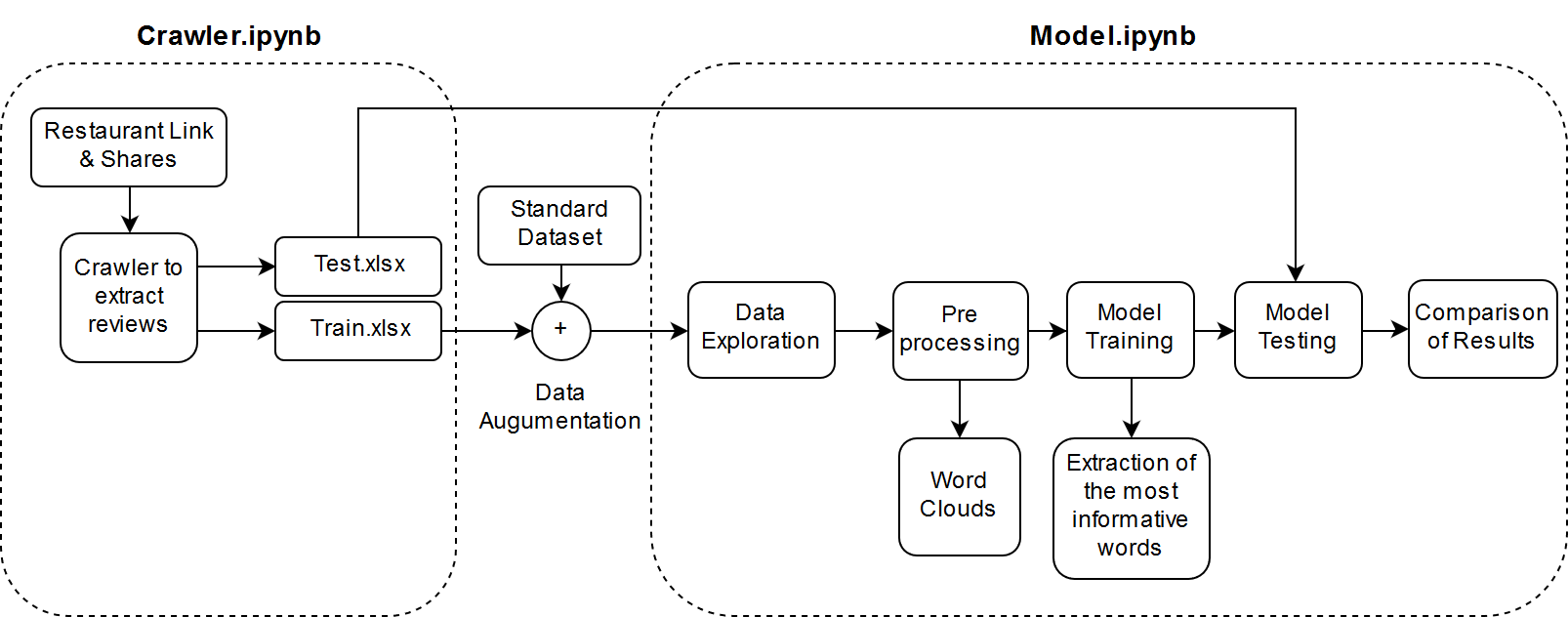
Approach: We formulated two different python files, one for the scraping & one for the model building.



Crawler.ipynb: Takes in the restaurant links & shares, Crawls all the reviews for the restaurants; It outputs two files Train.xlsx & Test.xlsx. Train got 6161 reviews & Test got 4045 reviews.

Train set is augmented with the Standard set and resulted into 26161 reviews for the train set.

Model.ipynb: Takes in the augmented file and train the model

The most informative words were computed from the Naive Bayes Classifier are as follows:

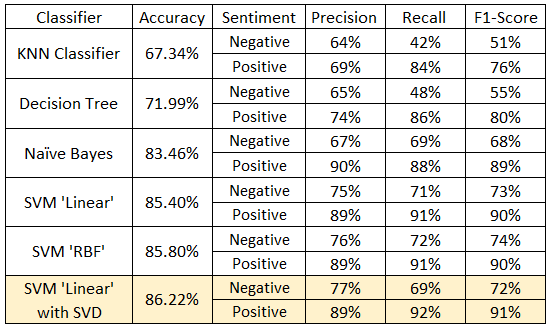
  
Clearly, words such as flavourless, unacceptable, tasteless, insult, worst, downhill, disgusting, bleh, ugh, rudely impact the negativity greatly and got a great predictive power.

After that we ran the model on the following classifier:

* KNN Classifier
* Decision Tree
* Naïve Bayes
* SVM ‘Linear Kernel’
* SVM ‘RBF Kernel’

SVM ‘RBF Kernel’ resulted into the highest Accuracy and F1 Score; Then, we performed Single Value Decomposition on the Test and Test vector to see the improvement with ‘SVD’. Model improved with the SVD, however improvement wasn’t that significant.

Below Table Summarises the different Models:



The SVM linear with SVD resulted into the best accuracy and has the following confusion matrix

|  |  |  |  |
| --- | --- | --- | --- |
|  | | Predicted | |
| Negative | Positive |
| Actual | Negative | 733 | 224 |
| Positive | 333 | 2755 |