Rating Prediction Project Report

We have a website where people write different reviews for technical products.
Now they are adding a new feature to their website i.e. The reviewer will have to add stars(rating) as well with the review. The rating is out 5 stars and it only has 5 options available 1 star, 2 stars, 3 stars, 4 stars, 5 stars.
Now they want to predict ratings for the reviews which were written in the past and they don't have a rating.
So, we have to build an application which can predict the rating by seeing the review.

Steps taken:

1. We check the data inside the dataset to understand the type of data.
2. We see that there are null values present in the dataset and we clean them.
3. Now, we look into the comment feature of the data closely.
4. There are a lot of data cleaning to be done, hence we do that.
5. After the data cleaning, we look into the data again to infer insights.
6. Now we do a general data analysis to understand the classification.
7. Now we do an extensive research using a correlation matrix to grade each column to the weightage to the target column.
8. We remove the features that has very little correlation.
9. Now we have reduced our dataset to the features that produces a weightage to the target column and hence ready to train the model.
10. We test different models to short list the best performing one.

Conclusion:

From the project we get a better understanding of the words that are included in the sentence that grade the comment to the rating along it. On this basis and understanding, we can easily infer that sentences containing such words and on the frequency of the appearance of these words classifies the comment as various ratings.