**Group 149: Yelp Review Classification using Multi -Label and Multi Class Classification**

|  |  |  |
| --- | --- | --- |
| First Name | Last Name | Share project with ITMD 527? (Y or N) |
| Sadaa Sree | Ravichandar | N |

Table of Contents

[**1. Introduction** 2](#_Toc498878359)

[**2. Data** 2](#_Toc498878360)

[**3. Problems to be Solved** 2](#_Toc498878361)

[**4. KDD** 3](#_Toc498878362)

[4.1. Data Processing 3](#_Toc498878363)

[4.2. Data Mining Methods and Processes 5](#_Toc498878364)

[**5. Evaluations and Results** 7](#_Toc498878365)

[5.1. Evaluation Methods 7](#_Toc498878366)

[5.2. Results and Findings 7](#_Toc498878367)

[**6. Conclusions and Future Work** 15](#_Toc498878368)

[6.1. Conclusions 15](#_Toc498878369)

[6.2. Limitations 15](#_Toc498878370)

[6.3. Potential Improvements or Future Work 15](#_Toc498878371)

# **Introduction**

Yelp users give ratings and write reviews about businesses and services on Yelp. These reviews and ratings help other Yelp users to evaluate a business or a service and make a choice.

The aim of this project is to build a classifier that can classify the businesses into four defined categories, based on the information present in the reviews and to classify the sentiments of the reviews.

This project formulates the task of classifying a review into relevant categories since a review can be associated with multiple categories at the same time (Multi-Label classification) and the project also includes the task of predicting and classifying the sentiment of the text review that is whether it is bad, good, excellent (Multi-Class Classification).

This type of categorization of reviews into relevant categories can help user to understand why the reviewer rated the restaurant as good or bad. This information can help other yelpers to make a personalized choice, especially when one does not have much time to spend on reading the reviews. Moreover, such categorization can also be used to rank restaurants according to these categories.

The project is done in Jupyter Notebook using python 3.

# **2. Data**

The yelp review dataset from Kaggle has 10,000 business reviews. The attributes present in this dataset include “business\_id”, ”date”,”review\_id”,”stars”,”text”,”type”,”user\_id”,”cool”,”useful” and “funny”.

For the **first task** to classify the reviews into relevant categories only food or restaurant data is considered. This reduces the number of reviews to 4290.These reviews are further categorized under four categories namely “Food”, “Service”, “Ambience” and “Worthiness”. The final dataset is used to predict and classify the text reviews into the above mentioned four categories.

***Here the data was carefully analyzed and categorized into Food, Service, Ambience and Worthiness using excel sheet sort/filter. "Food" and "Service" categories are easy to interpret. "Ambience" category relates to the décor and look and feel of the place. “Worthiness” category can be summarized as value for money. The reviews were filtered with various text for to categorize the ambience like neat, clean, beautiful, patio, painting and many combination of words.***

For the **second and third task** where we classify the sentiment and ratings of the text reviews based on the label/class “Stars”, the entire 10,000 reviews are taken into consideration for this task.

We then split the data into 80% train and 20% test and proceed with the data mining process.

***Source:***

<https://www.kaggle.com/astandrik/yelp-review/data>

# **3. Problems to be Solved**

**Research Problem 1: Classify and predict the restaurant reviews into relevant categories(Food, Service, Ambience, Worthiness)**

The first task in this project is to predict and classify the restaurant data reviews into four labels named “Food”, “Service”, “Ambience” and “Worthiness”. We use Multi-label classification task to solve this problem since a customer can write a review based on more than one category (for eg: a restaurant review may be classified under both “food” and “service”).

**Research Problem 2: Classify and predict the sentiments of the yelp reviews (Excellent, good, average, bad)**

The second research problem is where the sentiments of the text reviews are predicted and classified as bad, average, good and excellent with star rating 1 as “bad”, 2 & 3 as “average”,

4 as “good” and 5 as “excellent”. Since the yelper would have rated the business based on the number of stars and since a review can fall under any one of the mentioned class/label, we use Multi-Class Classification task to classify the sentiments of the yelp reviews.

**Research Problem 3: Classify and predict the ratings of the yelp reviews (1,2,3,4,5)**

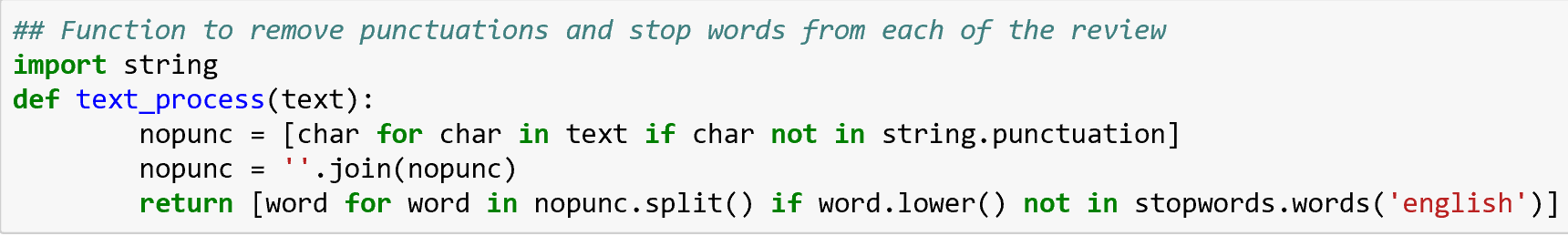
The third research problem is where the ratings of the text reviews are predicted and classified. Since the yelper would have rated the business based on the number of stars and since a review can fall under any one of the mentioned class/label, we use Multi-Class Classification task to classify the ratings of the yelp reviews.

# **4. KDD**

## 4.1. Data Processing

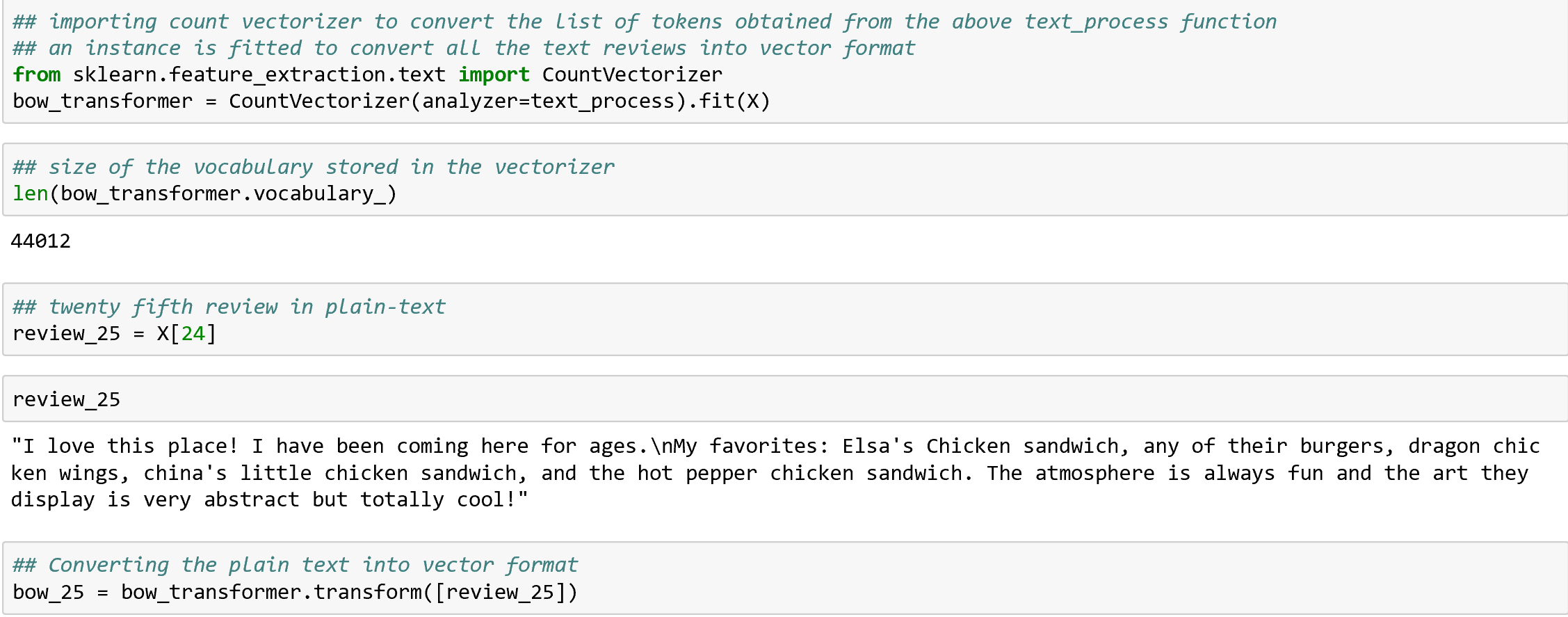
**Feature Extraction**

The classification algorithm will need some sort of **feature vector** to perform the classification task. The simplest way to convert a corpus to a vector format is the **bag-of-words** approach, where each unique word in a text will be represented by one number. Taking advantage of the NLTK library I have removed the stop words

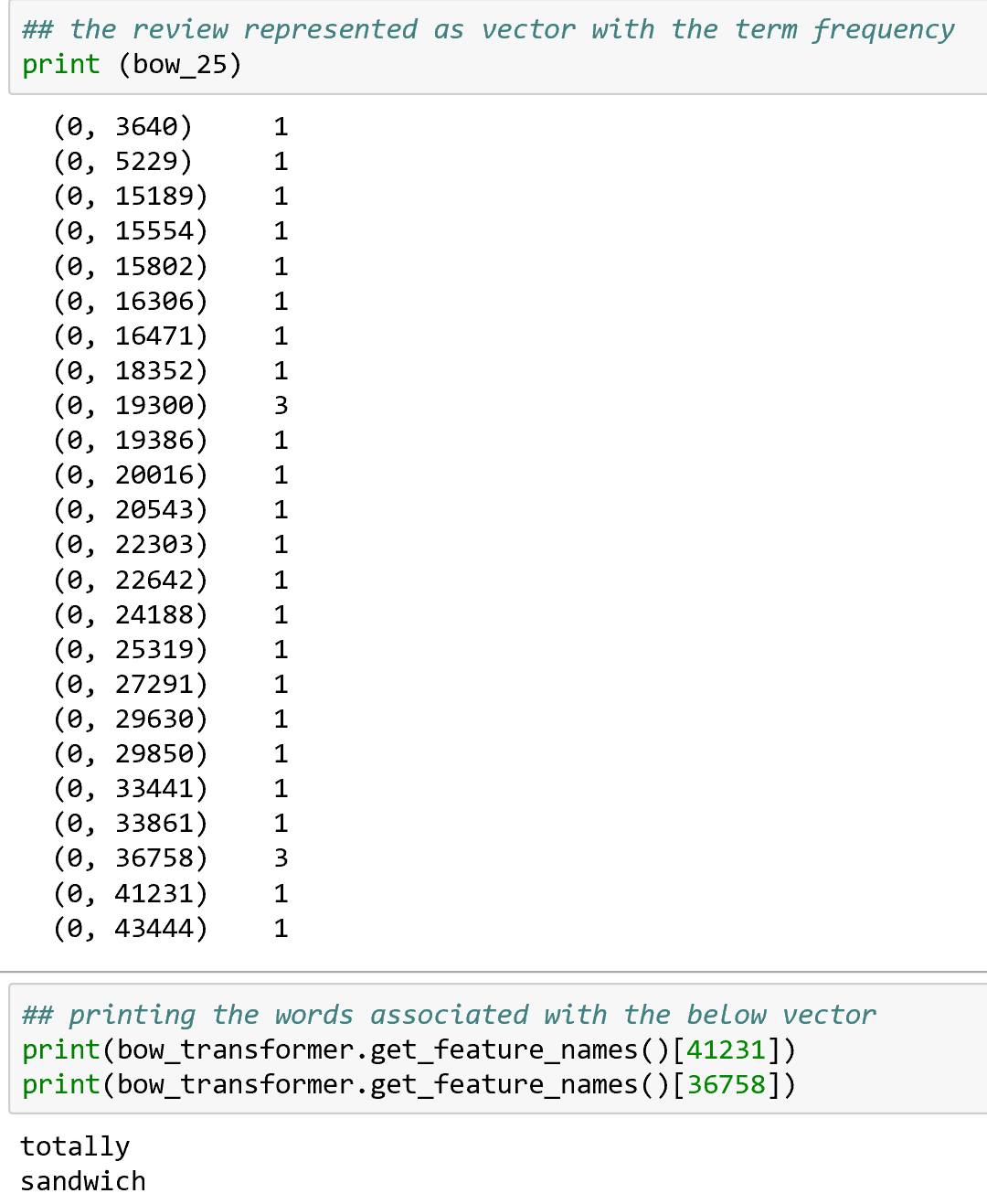


**Vectorization**

At the moment, the reviews are a lists of tokens (also known as [lemmas](https://nlp.stanford.edu/IR-book/html/htmledition/stemming-and-lemmatization-1.html)). To enable Scikit-learn algorithms to work on these text, each review should be converted into a vector. We can use Scikit-learn’s [Count Vectorizer](http://scikit-learn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html) to convert the text collection into a matrix of token counts. Sickit learn’s Count Vectorizer is imported an d an instance is fitted to convert all the reviews into vector.

Transforming the plain text of a sample review into a list of words and then into vector using the Bag Of words Approach

The sample review where the words in the review have been transformed into vector and their frequency in the review



As we can find there are 24 unique words present in the above review with two words occurring 3 times (term frequency)

## 4.2. Data Mining Methods and Processes

**Multi-Label Classification**

To classify and predict the reviews into four relevant categories say Food, service, ambience and worthiness. I formulated the task of classifying a review into relevant categories as a learning problem. However, since a review is in exclusively associated with multiple categories at the same time, it is not a simple binary classification or a multi-class classification. It is rather a multi-label classification problem.

There are three methods to do a multi label classification

1.**Problem Transformation**

This method can be carried out in three ways

i.***Binary Relevance*** where each label is treated as a separate single class classification

ii. ***Classifier Chain*** where the first classifier is trained on the input data and then each next classifier is trained on the input space and all the previous classifiers in the chain

iii. ***Label Powerset*** where we transform the problem into a multi class classifier is trained on all unique label combinations found in the training data

Classifiers like Linear SVC, MLP, Extra Tree, Decision Tree, KNN, Gaussian Naïve Bayes, Multinomial Naïve Bayes were each performed on all the above methods in problem transformation method.

2.**Adapted Algorithm**

This method is nothing but an adapted version of multi class classifiers to perform multi label classification. Classifiers MLKNN (Multi Label KNN) with different K values were performed and classifier BRKNN(Binary relevance KNN) was also performed

3.**Ensemble Approach**

Ensemble of Label power set Random forest, Binary Relevance Random Forest, Classifier Chain Random forest, Label power set Extra trees, Binary Relevance Extra Trees, Classifier Chain Extra Trees were performed to find the best model in the Multi-Label classification.

**Multi-class Classification**

To classify and predict the sentiments of the yelp reviews into Excellent, good, average and bad and to classify and predict the ratings of the reviews from 1 to 5 Multi -class classification method is used.

Different Classifiers say MLP classifier, Logistic Regression, Extra tree, Linear SVC , Decision Tree , KNN with different K values, Bernoulli Naïve Bayes, Gaussian Naïve Bayes, Multinomial Naïve Bayes and many Ensemble approaches like Gradient boosting, Adaboost, Bagging & KNN ,Extra Trees, Random Forest were also performed to determine the best model based on the accuracy

# **5. Evaluations and Results**

## 5.1. Evaluation Methods

On this dataset we will follow the standard process of supervised learning, since the dataset is large we will perform hold-out evaluation with 80% as training data and 20% as test data to predict our results in the test data and determine the accuracy to find the best model.

For Multi-Label classification we simply find the accuracy of the predicted data with the test data and in Multi -class classification we use precision, recall, f1-score and support in addition to accuracy to determine the accuracy of the model.

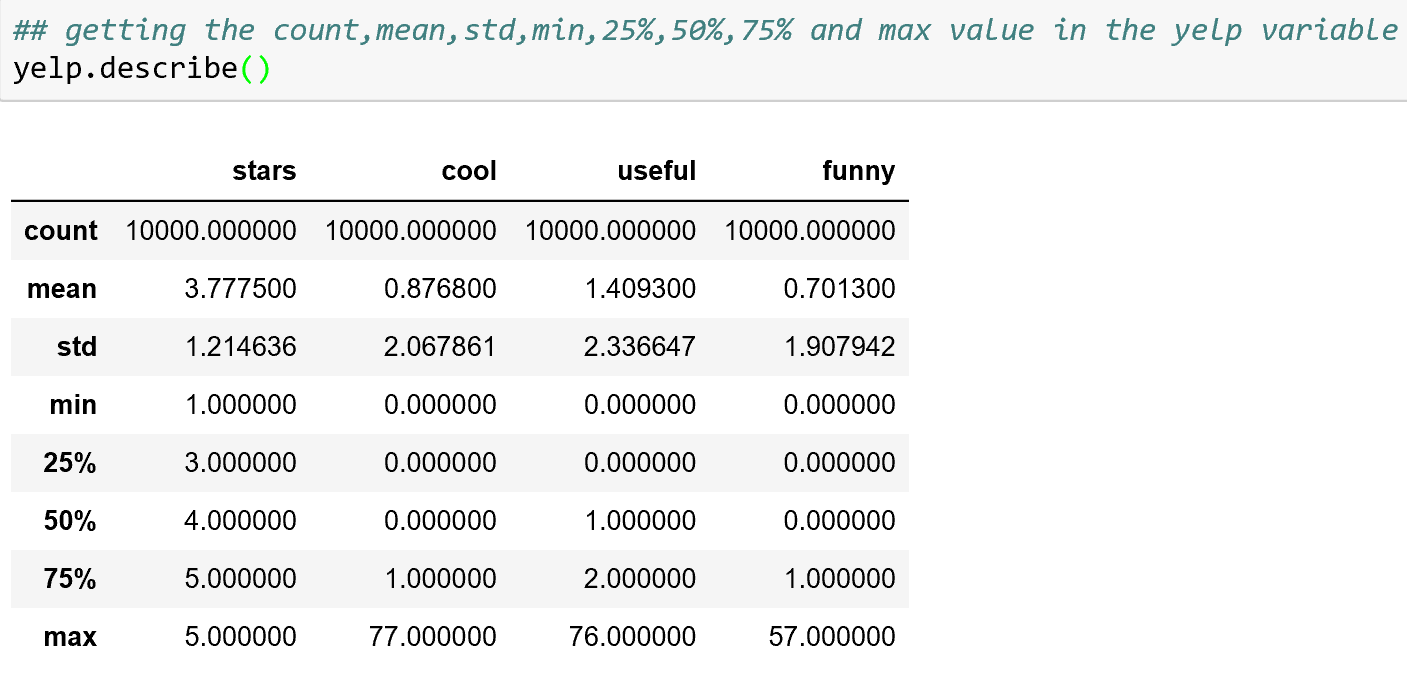
The accuracy of each model is then compared to select the nest model which is used to predict future unknown labels

## 5.2. Results and Findings

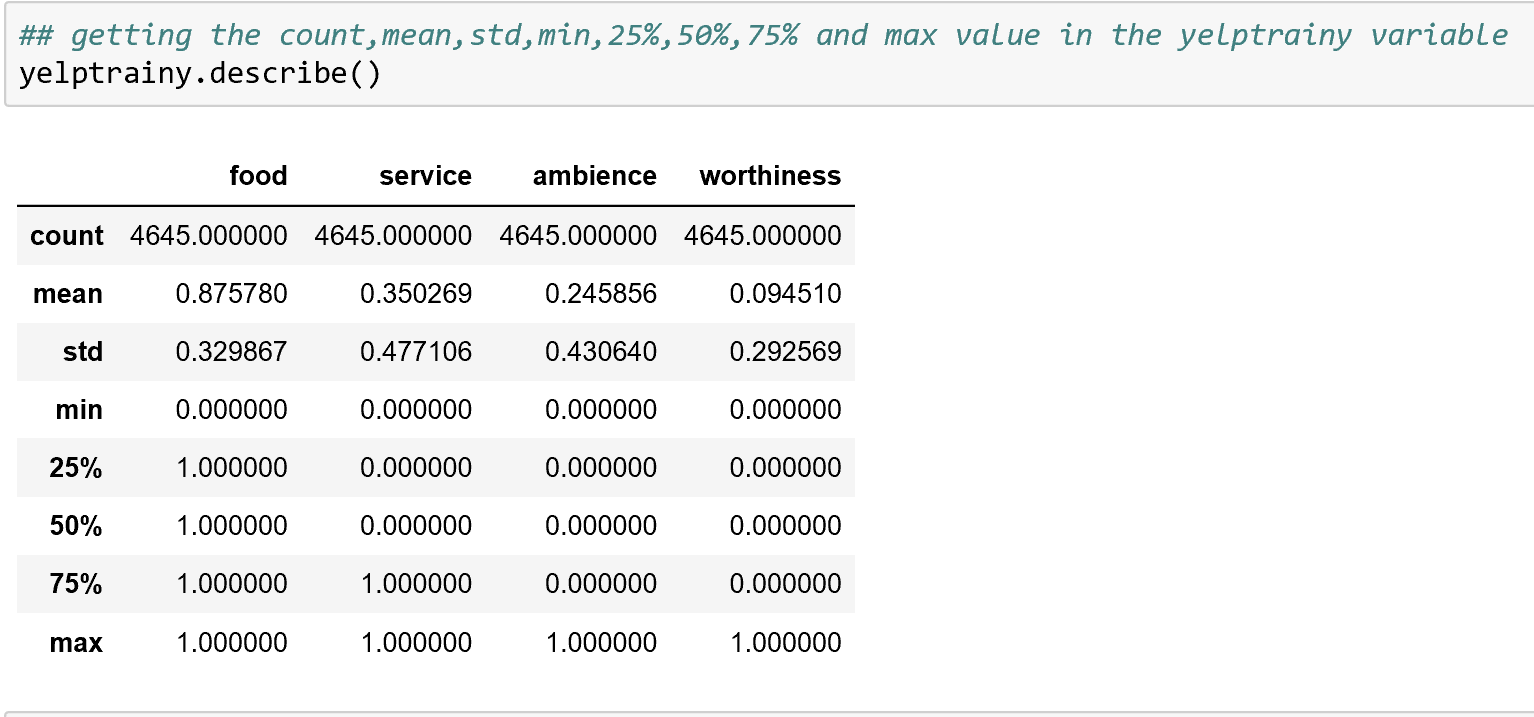
**To Classify and Predict the yelp reviews into four Categories (food, Ambience, service, Worthiness)**

After performing the feature extraction and vectorization on 4600 reviews (Food or Restaurant reviews), the data was later then split into 80% training and 20% testing. Here the labels to be predicted are food, ambience, service and worthiness. The data mining method and process were performed on the training data and each classifier was predicted for the test data and then compared with the original label of the test data to determine the accuracy of the model.

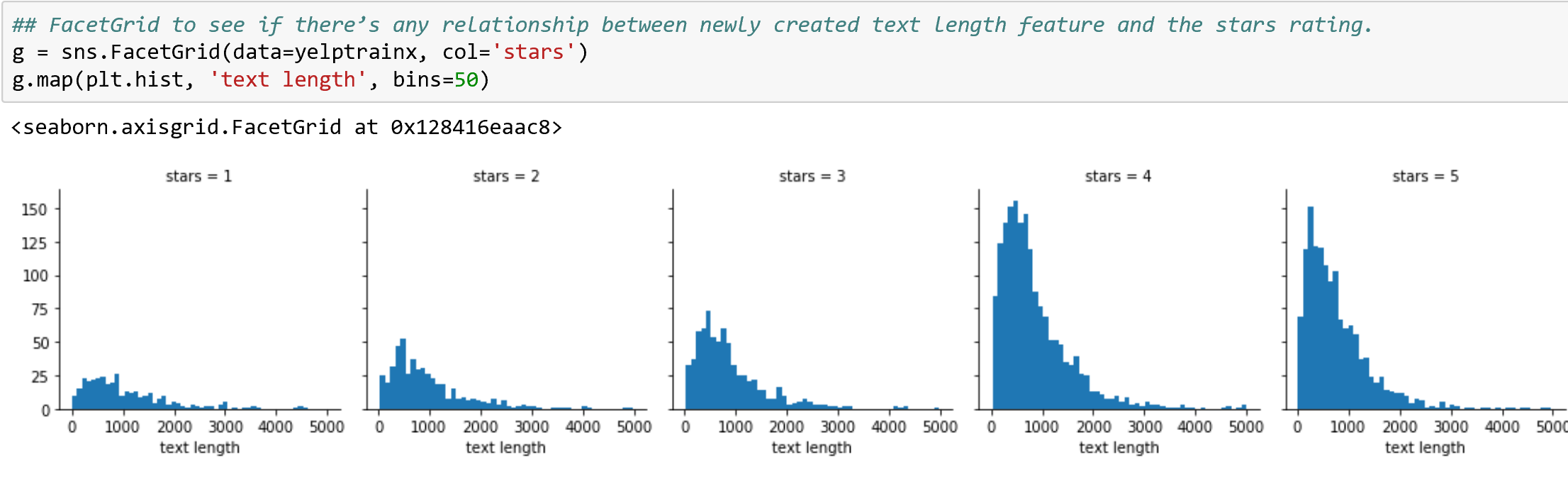
Determining the statistics of the yelp restaurant review dataset



Determining the statistics of the y variable



A new column text length was added in the data to see if there is any relationship between text length and the star rating



The distribution of text length is similar across all five ratings. However, the number of text reviews seems to be skewed a lot higher towards the 4-star and 5-star ratings.

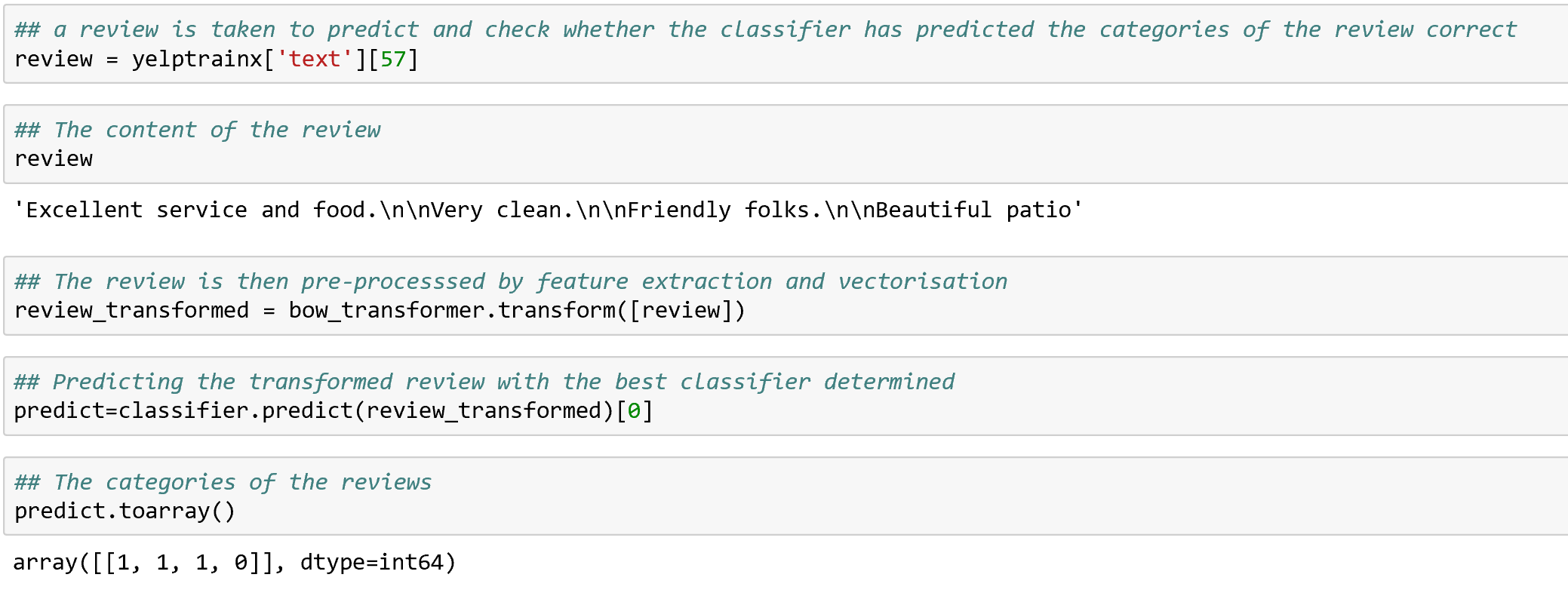
Accuracy from Problem Transformation

Accuracy from Adapted algorithm method

accuracy from ensemble approach

So based on the above results we find that the decision tree from binary relevance has given the best accuracy of about 96 %. Later this model was used to predict the unknown labels that is whether a review is written based on Food, Service, Ambience and Worthiness.

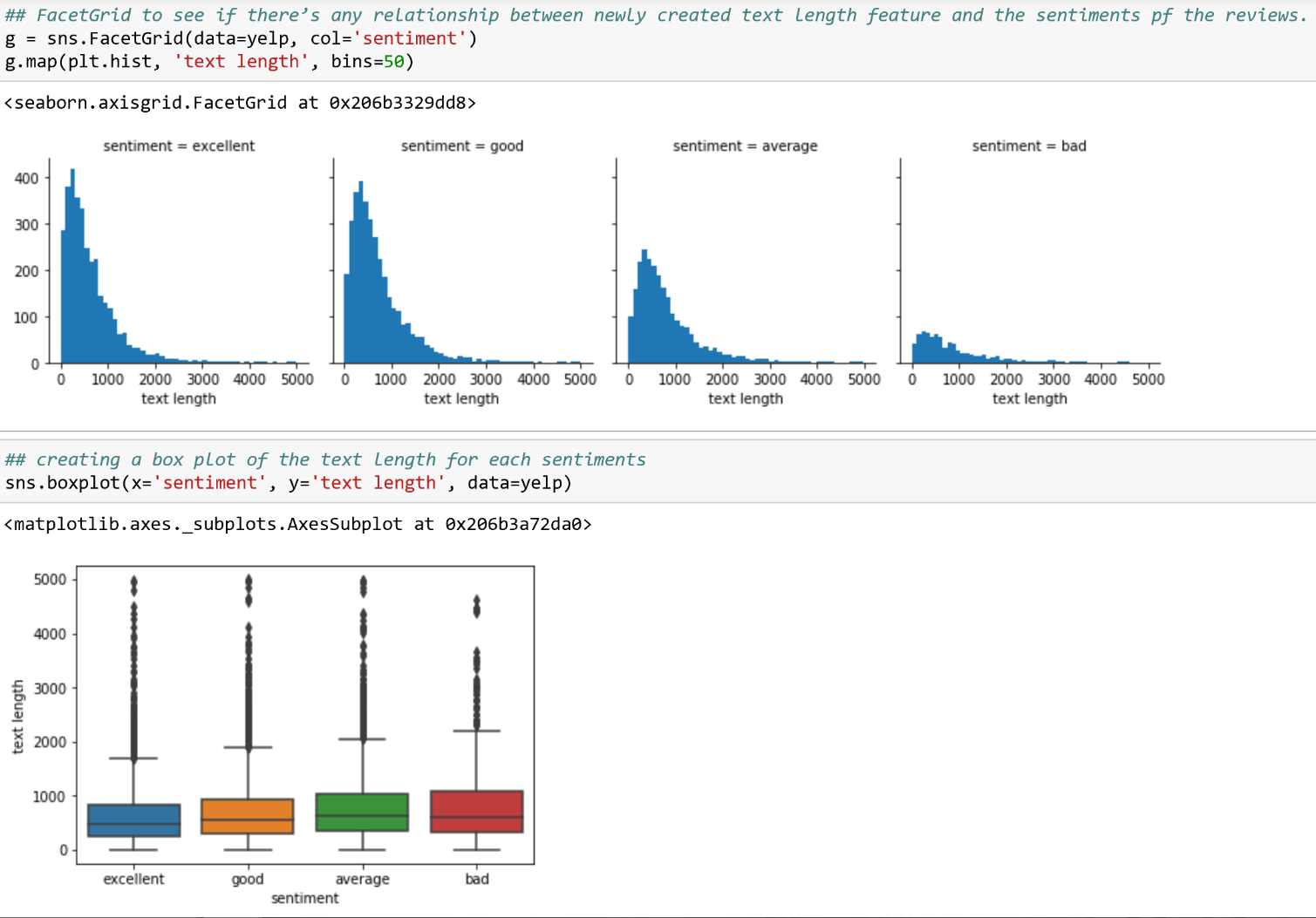
An unknown Label is taken and predicted to see based on what categories the review was written



When we read the review, it is clear that the yelper has written the reviwe based on his experience on the food, service and the appearance of the place. As expected our classifier has predicted that the review falls under food, service and ambience category.

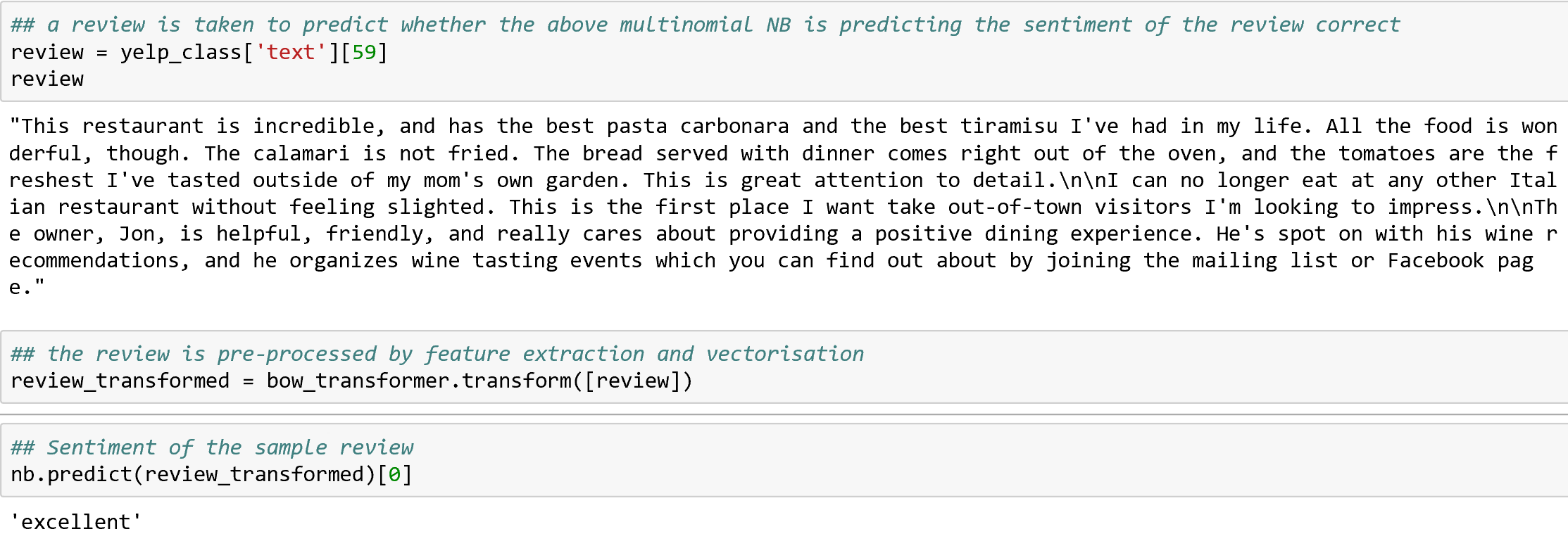
**To Classify and Predict the sentiments of the yelp reviews into (Excellent, Good, Average and Bad)**

After performing the feature extraction and vectorization on overall 10000 yelp reviews, the data was later then split into 80% training and 20% testing. Here the labels to be predicted are excellent , good, average and bad. The data mining method and process were performed on the training data and each classifier was predicted for the test data and then compared with the original label of the test data to determine the accuracy of the model.

Determining the relationship between text length and the sentiment of the reviews

Sentimental Analysis- accuracy

From the above result we find the Multinomial Naïve Bayes has the highest accuracy of 57 % and now this classifier is used to predict and classify the sentiment of the review given the plain text review.



When we read the review, we find that the reviewer was very much satisfied with business service and as expected our classifier has the **predicted and classified the sentiment of the review to be excellent.** Hence the classifier which was built can predict the sentiments of the review given a review in a plain text format

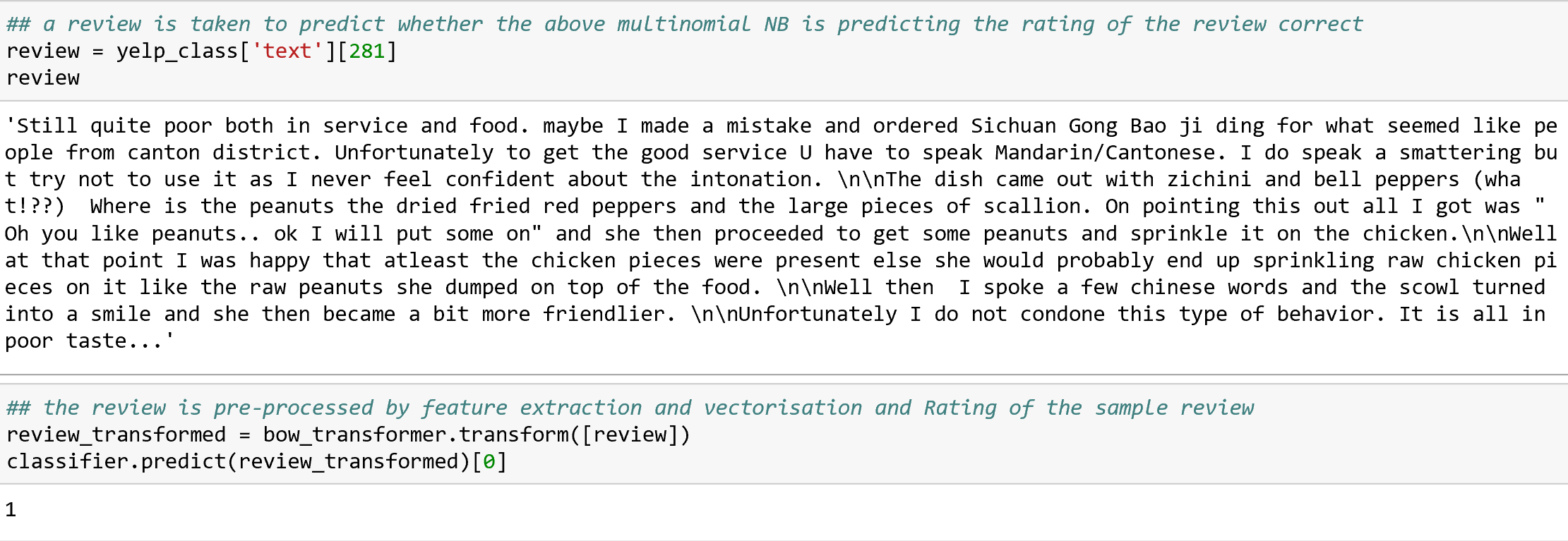
**To Classify and Predict the ratings of the yelp reviews into (1 to 5)**

After performing the feature extraction and vectorization on overall 10000 yelp reviews, the data was later then split into 80% training and 20% testing. Here the labels to be predicted 1,2,3,4,5.The data mining method and process were performed on the training data and each classifier was predicted for the test data and then compared with the original label of the test data to determine the accuracy of the model.

Determining the correlation between star ratings and the other features in the yelp data

Classifying review ratings – accuracy

From the above result we find the Logistic Regression has the highest accuracy of close to 50 % and now this classifier is used to predict and classify the ratings of the review given the plain text review.



When we read the review, we find that the reviewer was not satisfied with business service and as expected our classifier has the **predicted and classified the rating of the review to be 1**. Hence the classifier which was built can predict the ratings of the review given a review in a plain text format.

# **6. Conclusions and Future Work**

## 6.1. Conclusions

## Yelp reviews and ratings are important source of information to make informed decisions about a venue. I conjecture that further classification of yelp reviews into relevant categories can help users to make an informed decision based on their personal preferences for categories. Moreover, this aspect is especially useful when users do not have time to read many reviews to infer the popularity of venues across these categories. In this paper, I have demonstrated how reviews for restaurants can be automatically classified into four relevant categories with an accuracy of 96 % respectively.

The sentimental analysis can be automatically classified into excellent, good, average and bad and the rating of the review can be also automatically classified from 1 to 5. Thus a model/classifier was built to determine whether the user liked a business or not based on what they have typed.

## 6.2. Limitations

We can find that the accuracy of the sentimental analysis and the prediction of the review rating does not give us a good accuracy and the project was not checked for overfitting

This might be because while doing the text pre-processing to retrieve the information tf-idf could have been used since about 83% of text-based recommender systems use tf-dif.

I have planned to continue with this project using tf-dif in the near future to get better results.

## 6.3. Potential Improvements or Future Work

As mentioned in the limitation the accuracy for the rating prediction is around 50% which is not super great. So, in the future, feature extraction using tf-idf (term frequency–inverse document frequency) could be done to improve the classifier that has been built and also to check the overfitting problem and to overcome the overfitting problem if any exists.