**Amazon Fine Food Review Analysis**

The Amzon Fine Food reviews dataset is available in Kaggle.The dataset contains reviews given by the customer for a perticular product.

**Objective :**

The main objective of this project is to find whether the given review is +ve or -ve review.This kind of problem can be called as sentiment analysis problem.

**Preprocessing** :

Even though the dataset contains several columns, we will try to predict only by looking the review given by the customer.

In this step, based on the score column I have classified the points. If score <3 then its negative.If score >3 then its positive. If score = 3 then we simply ignore the data points for which score=3.

I have also performed some regular text preprocesing like removing spaces, unwanted symbols or charecters etc..Cleaning the text is a very important because if data cleaning is not properly done then the model trainng will suffer.

**Vectorization :**

Here in this step we convert the text into vectors. I have mainly done 3 kinds of vectorization.

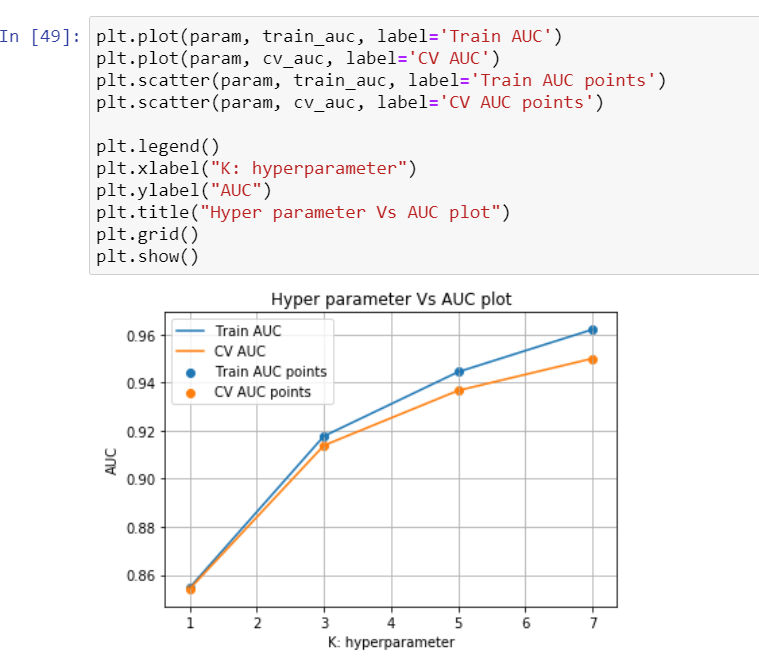
1. Bag of Words
2. TF-IDF
3. TF-IDF-Word2Vec

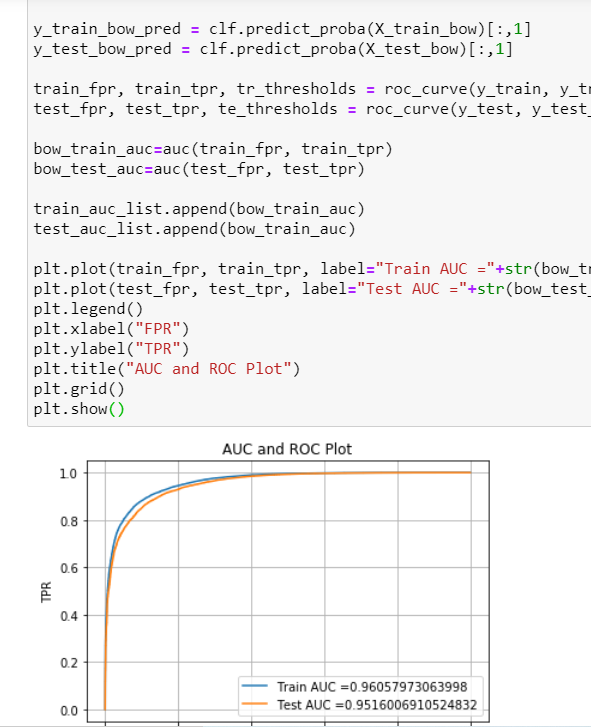
The above mentioned are the one of the most popular vectorization techniques.

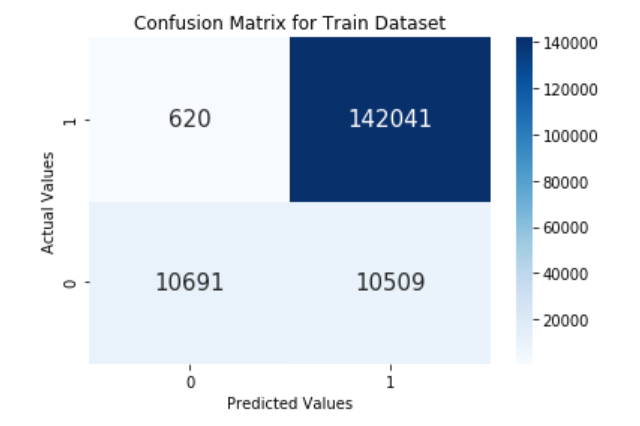
**Modelling**:

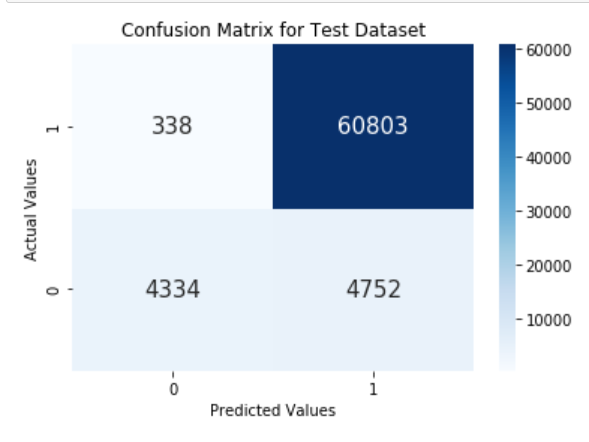
For the above mentioned 3 vectorizations I have used the XGBoost Classifier with hyper parameter tuning.

1. XGBoost with Bag of Words :





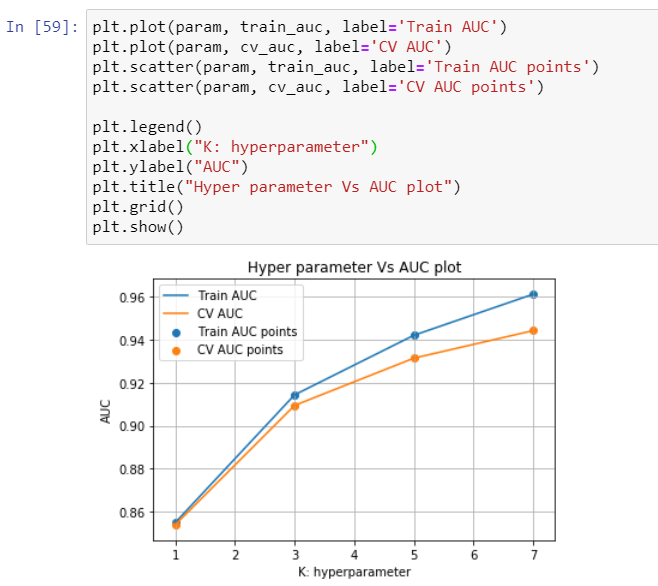


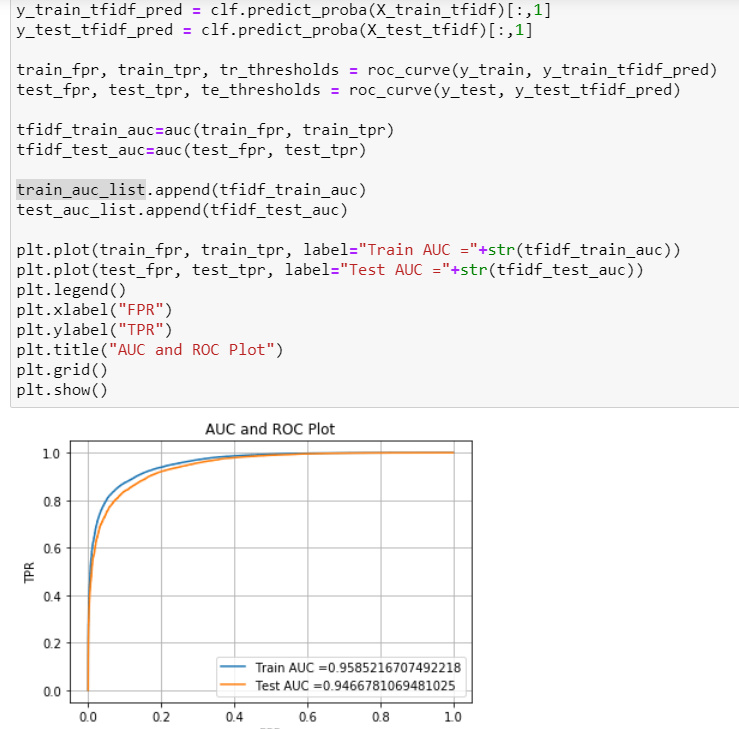


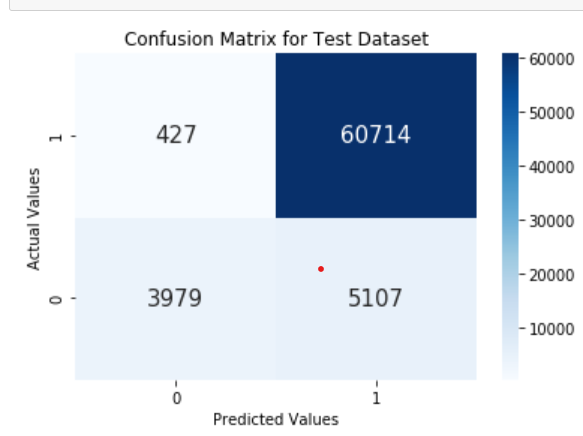
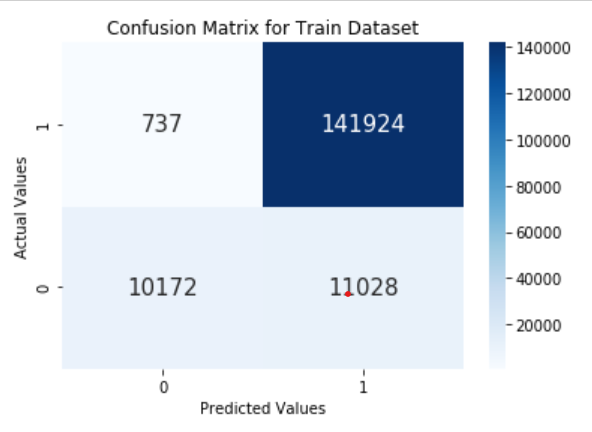
**Training Accuracy: 93.20826798322969**

**test accuracy: 92.7520754125906**

1. **TF-IDF :**



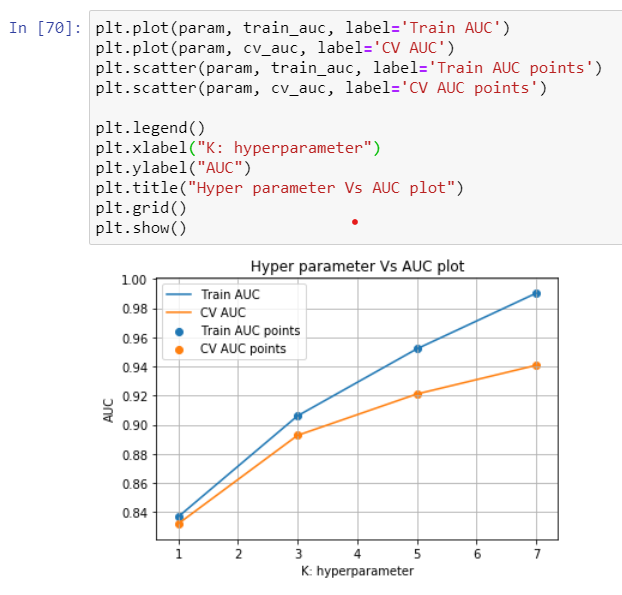


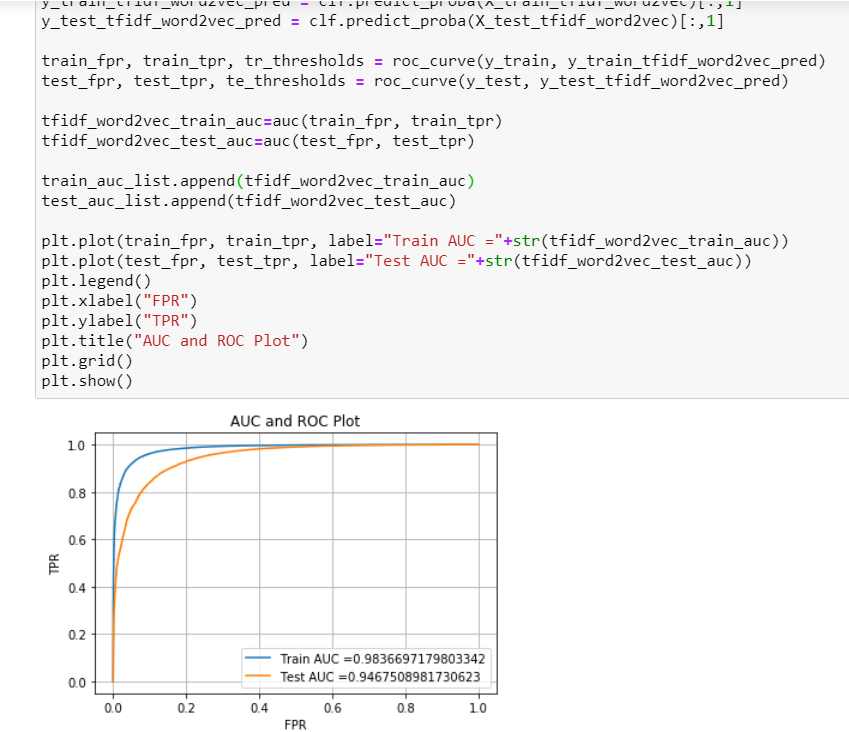


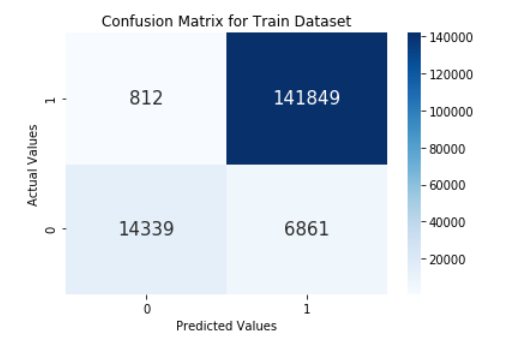
Training Accuracy: 92.82013413808045

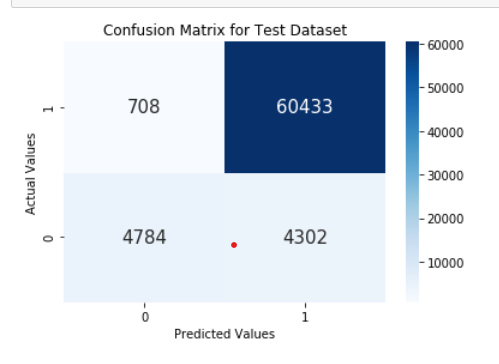
test accuracy: 92.1198399475985

1. **TF-IDF-Word2Vec :**



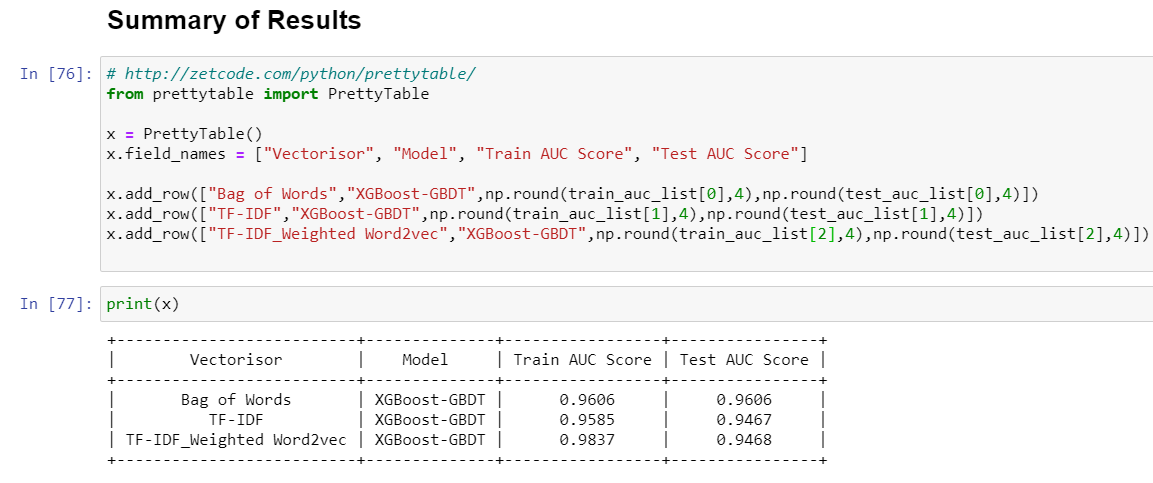






Training Accuracy: 95.31737265121048

test accuracy: 92.86599171258918



**Conclusion :**

In all 3 Vectorizations, if we observe the test AUC is almost same and accuracy is also almost same. Hence we can choose bag of words model because its computationally efficient compared to other techniques.