Naive Bayes Algorithm is applied on Reviews of Customer using Bag of Words Model

OBJECTIVES:

Objectives are listed below.

1. Visit the given links:
   https://www.gartner.com/reviews/market/meeting-solutions-web-conferencing/vendor/att/product/att-connect
   https://www.gartner.com/reviews/market/meeting-solutions-web-conferencing/vendor/att/product/att-connect?pageNum=2
   https://www.gartner.com/reviews/market/meeting-solutions-web-conferencing/vendor/att/product/att-connect?pageNum=3
   https://www.gartner.com/reviews/market/meeting-solutions-web-conferencing/vendor/att/product/att-connect?pageNum=4
   https://www.gartner.com/reviews/market/network-services/vendor/att
   https://www.gartner.com/reviews/market/network-services/vendor/att?pageNum=2
   https://www.gartner.com/reviews/market/network-services/vendor/att?pageNum=3
   https://www.gartner.com/reviews/market/network-services/vendor/att?pageNum=4
   https://www.gartner.com/reviews/market/network-services/vendor/att?pageNum=5
   https://www.gartner.com/reviews/market/network-services/vendor/att?pageNum=6
   https://www.gartner.com/reviews/market/network-services/vendor/att?pageNum=7
   https://www.gartner.com/reviews/market/network-services/vendor/att?pageNum=8
2. Scrape Reviews and Ratings, and then save the reviews in a csv format file
3. Apply NLP practices for cleaning the data
4. Apply Naive Bayes Algorithm

OVERVIEW:

Naive Bayes classifiers are a collection of classification algorithms based on Bayes' Theorem. It is not a single algorithm but a family of algorithms where all of them share a common principle, i.e. every pair of features being classified is independent of each other.

DATA AND MODEL:

I fit a Naïve Bayes Algorithm on Bag of Words model to predict Rating that included a Reviews of the Customers as predictors

PROCEDURE:

1. Gathering all the essential link(s) and store them on 'Links.csv' file
2. Scrape the given links for reviews and ratings using Selenium and BeautifulSoup
3. Apply preprocessing, Stemming and Lemming on data (or reviews) retrieved
4. Split the Model into Train set and Test set
5. Fit and transform those Train set and Test Set models
6. Check scores on Train data, Test data and also the Actual data

RESULTS:

I achieved a score of 99% and sometimes 100% because of two following reasons:

1. Because the data is very sufficient
2. This sufficient data leads to overfitting

CONCLUSION:

Naïve Bayes Algorithm gives best result if the Feature Matrix 'X' has no correlation among any other feature

Naïve Bayes uses a Bag of Words model to predict either 'Positive' or 'Negative'