MAJOR PROJECT

Name – Sharad Kumar

Topic:: Restaurant Review

Brief about it..

Firstly I start preprocessing part.

1. Firstly , I upload the tsv file using df . Here I use import pandas for reading the file
2. I describe the file using head which defines review and Liked defines 0 or 1.
3. Now I checked what null data is present in the file and also check dataset consist of the file
4. ### Data cleaning
5. We use regular expression for cleaning additional symbol. By using importing re
6. Now, I convert review into lower (small alphabets). For that I used porter stemmer which is in importing nltk. Here I download stopwords too.
7. Now, I convert all the stopwords using import corpus. I use set too for the collection of the stopwords .
8. Afterwards it seems to arrange the words. So, I use join for all collection. Therefore I use review=’ ’.join(review) for creating the space in the words.

Here preprocessing part ends.

1. Now applying for loops to get all the words store using corpus..
2. It seems that creating bag of words.
3. Now I use count Vectorizer for data analysis.
4. As x=vect.fit\_transform(corpus).toarray()
5. As y= dataset.iloc[:,1].values
6. For convert data and then after splitting the data using train\_test\_split. We take 20%of test size for that hence we get 80% train data.
7. Now I apply different model for checking accuracy.

A) SVC(Support classifier)

B) Gaussian NB

C) Multinomial NB

Steps that are used in every model.

1. Firstly support classifier
2. do classification report
3. Confusing matrix
4. All accuracy

By applying the same thing for different model

* For Gaussian NB

I got Confusion matrix

[ 55 42

12 91]

Accuracy = 73 %

* For Naïve Bayes Multinomial

I got confusion matrix

[ 72 75

22 81]

Accuracy 76.5%

* K Neighbours classifiers

Confusion Matrix

[ 74 75

22 81]

Accuracy 61%

* Logistic Regression

Confusion Matrix

[76 25

22 61]

Accuracy 71%

Now Below are the screen shot of the Sentiments..

Some bar Graphs and Plot using seaborn,sckit learn library.



















