USE CASE 3.2

IDENTIFYING KEY TOPICS OR AREAS OF CONCERN

FLOW CHART:

Reading Data Pre-Process Data Sentiment Analysis

Model Training and Evaluation Feature Extraction

Topic Modelling Preparing and displaying Overall Key Topics

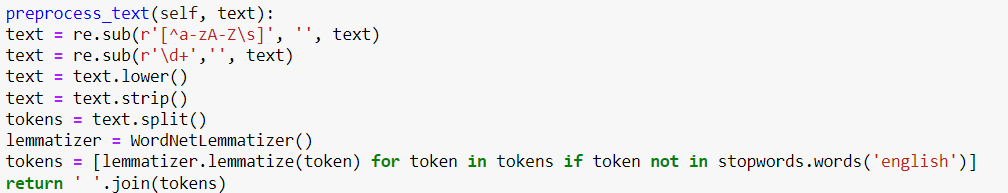
The End Preparing and displaying Key Negative Topics

Process Details:

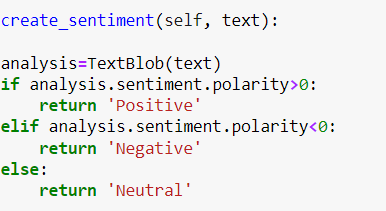
Reading Data: Reading the data from the given source

Pre-Process Data: The steps involved in the pre process are

* Removing Special characters and digits
* Convert text to lower case
* Tokenization and lemmatization
* Removing Stop Words

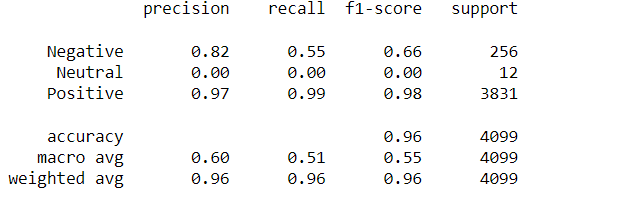


Sentiment Analysis: Since the review written are long I have selected TextBlob instead of sentiment IntensityAnalyzer since it is best for short reviews.



Feature Extraction: Split the data into training and testing sets, then perform TF-IDF vectorization on the text data.

Model Training and Evaluation: Train a Support Vector Machine (SVM) classifier using the TF-IDF features and evaluate its performance



Topic Modelling : Use Non-negative Matrix Factorization (NMF) to identify key aspects across over all reviews and negative reviews and displaying them



Considering the size of your dataset and ease of interpretation, LDA(Latent Dirichlet Allocation) or NMF would be good choices but NMF works more efficiently with the size of our data

Model Selection: Support Vector Machines(SVC)

Solution Architecture: On- Premisis Deployment

Deployment Steps: Copy the code to any IDE, update the data\_path before running

Output Screenshot:

