CUSTOMER SATISFACTION ANALYSIS USING MACHINE LEARNING AND SENTIMENT ANALYSIS TECHNIQUE

Group 9

Sharanya Sriram 11597311

Sreekar Thanda 11636306

Srinivas Rao Sankula 11667743

INTRODUCTION

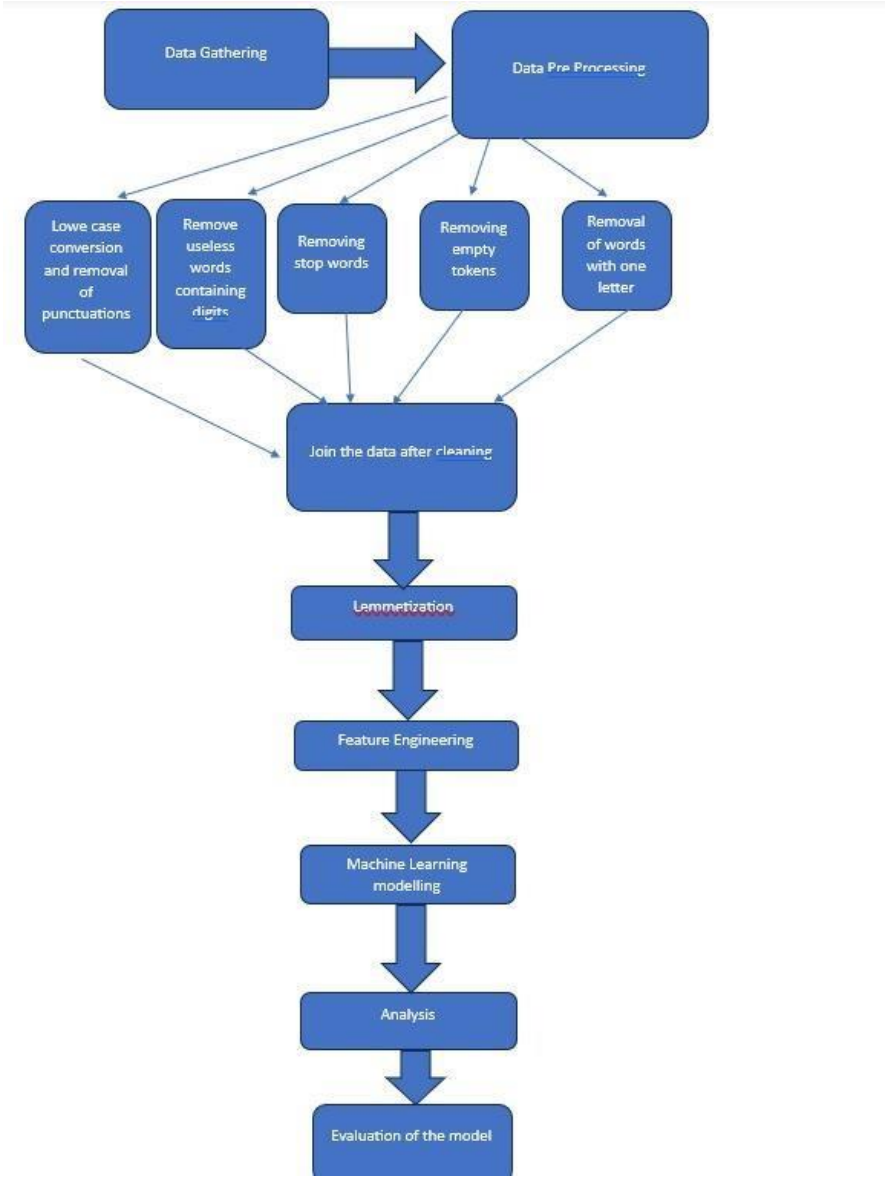
- A comprehensive approach to analyze hotel reviews using natural language processing techniques in Machine learning.
- We are using Sentiment analysis / Opinion Mining NLP technique to know the opinions of the customers.
- The goal of this project is to analyze hotel reviews sentiment and modeling reviewer scores as customer feedback plays a vital role to improve their services .

PROBLEM STATEMENT

- In today's competitive business landscape, understanding customer satisfaction is essential for success. Traditional methods, such as surveys and focus groups, are often impractical. The vast amount of online customer feedback presents an untapped source of valuable insights, but manually analyzing this data is not feasible. We are trying to overcome this in this project.

METHODOLOGY

- Some of the Python programming concepts and libraries that we will be using are Numpy, Pandas, NLTK, GENISM package, ROC curve, Precision-Recall curve for Tokenization, Stemming / Lemmatization, Exploratory Data Analysis, Feature Engineering, Unigram, Bi-Gram and Tri-Gram analysis, Sentiment analysis, Model Evaluation, and Visualization techniques.

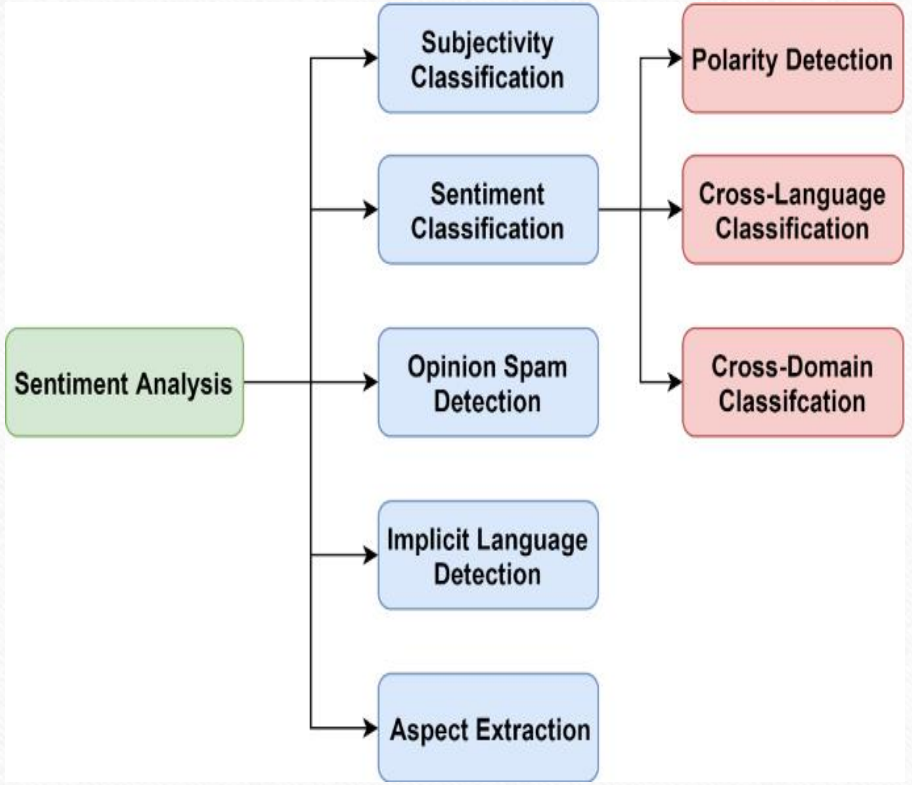


ARCHITECTURE OF MODELS

- Sentiment Intensity Analyzer
- Random Forest Classifier

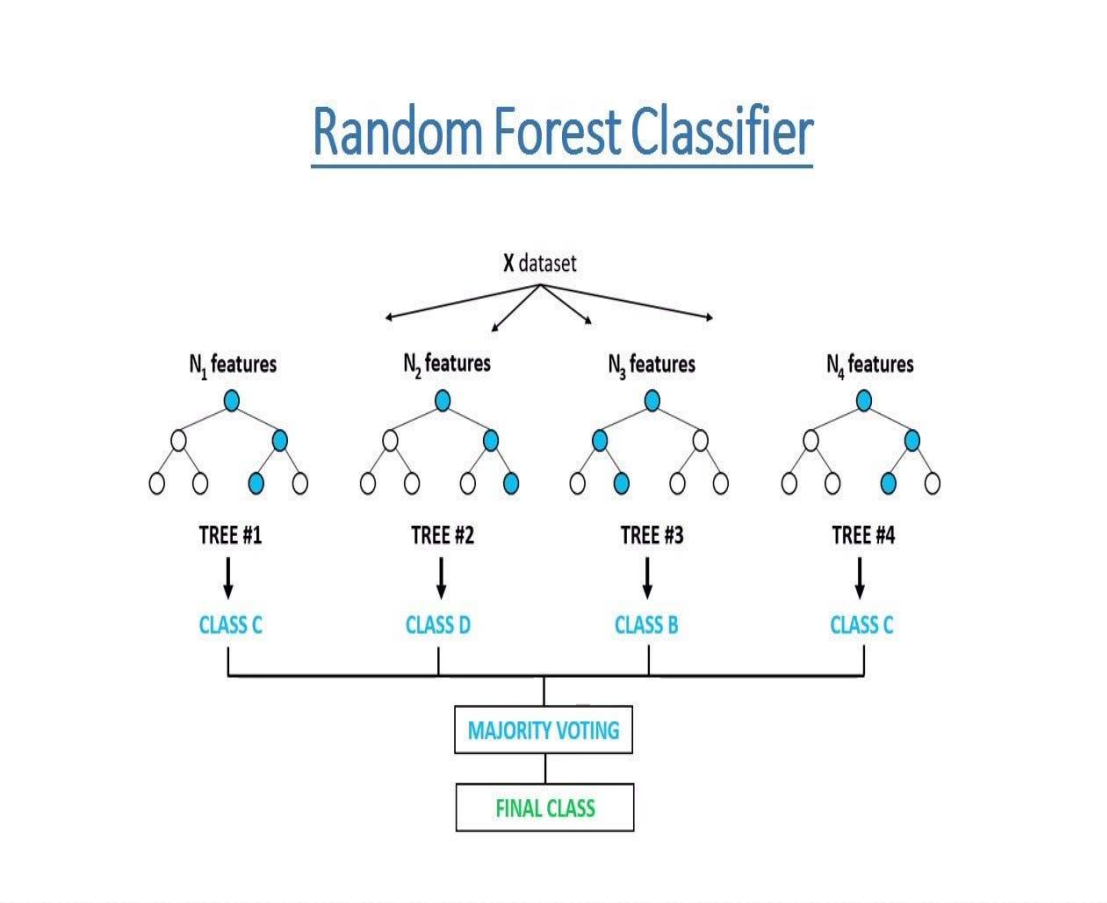
SENTIMENT INTENSITY ANALYSER

- Sentiment analysis is the computational study of various textual elements such as opinions, thoughts, judgements, interests, points of view, feelings, and subjectivity.
- VADER, a sentiment analysis tool, is specifically designed to analyze the sentiment expressed in social media. It utilizes a lexicon of words and their associated sentiments, providing both a positivity and negativity score for each piece of text. The Compound Score, a cumulative measure of all lexical ratings, ranges from -1 (most negative) to +1 (most positive).



RANDOM FOREST CLASSIFIER

- An effective machine learning approach called a random forest classifier uses a group of decision trees to jointly predict results.
- The classification power of this ensemble learning technique is well known, and it comes from its capacity to precisely capture complex relationships present in the data.
- Furthermore, random forest classifiers are strong and dependable because they show resistance to noise and outliers.
- Their adaptability to a wide range of applications is further demonstrated by their ability to handle both numerical and categorical data.



DATASET

- The dataset we are using here is Hotel_reviews dataset. (source: Kaggle)
- This dataset is a valuable resource for anyone who wants to learn more about customer satisfaction in the hotel industry or develop applications to improve the guest experience.
- It consists of various columns such as Hotel Address, Review date, Title , Review Text, Reviewer username, positive review tags, negative review tags and average rating .
- This dataset has 5,15,738 rows and 17 features.

DESIGN OF FEATURES

- The Hotel Reviews dataset is a relational dataset, with the following tables:
  - reviews: (review_id, user_id, hotel_id, review, ratings)
  - users: (user_id, username)
  - hotels: (hotel_id, hotel_address, review_date)
- Labels:
  - Negative_Review, Review_Total_Negative_Word_Counts, Positive_Review, Review_Total_Positive_Word_Counts.

DATA PREPROCESSING

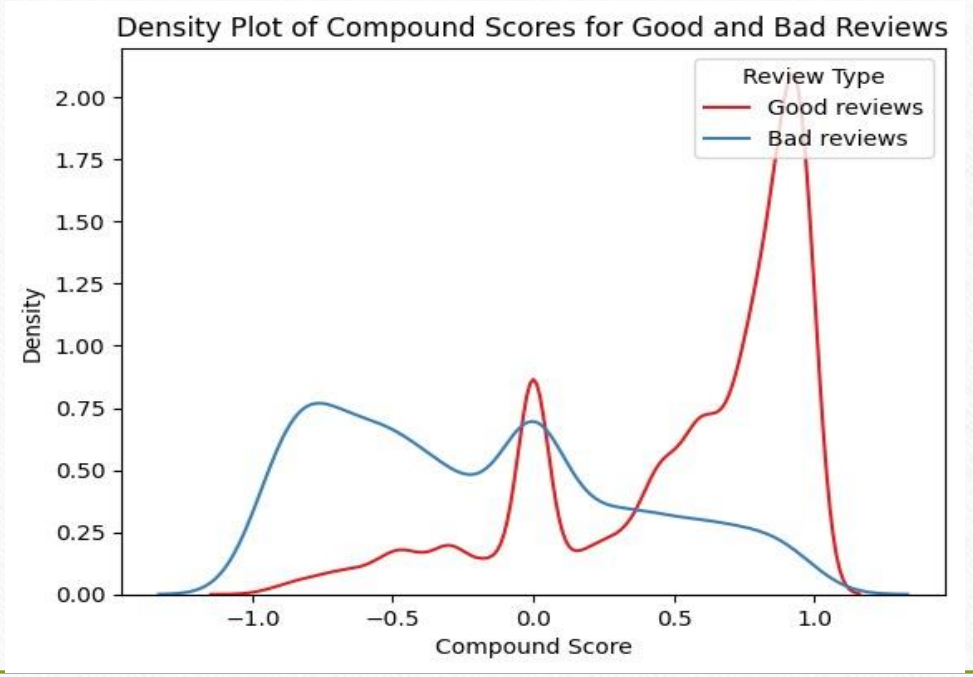
- Conversion to Lower Case
- Tokenization and removal of punctuations
- Stop-words removal
- Removing empty tokens
- Lemmatization

EXPLORATORY DATA ANALYSIS

- Visualization of the most common appeared words.

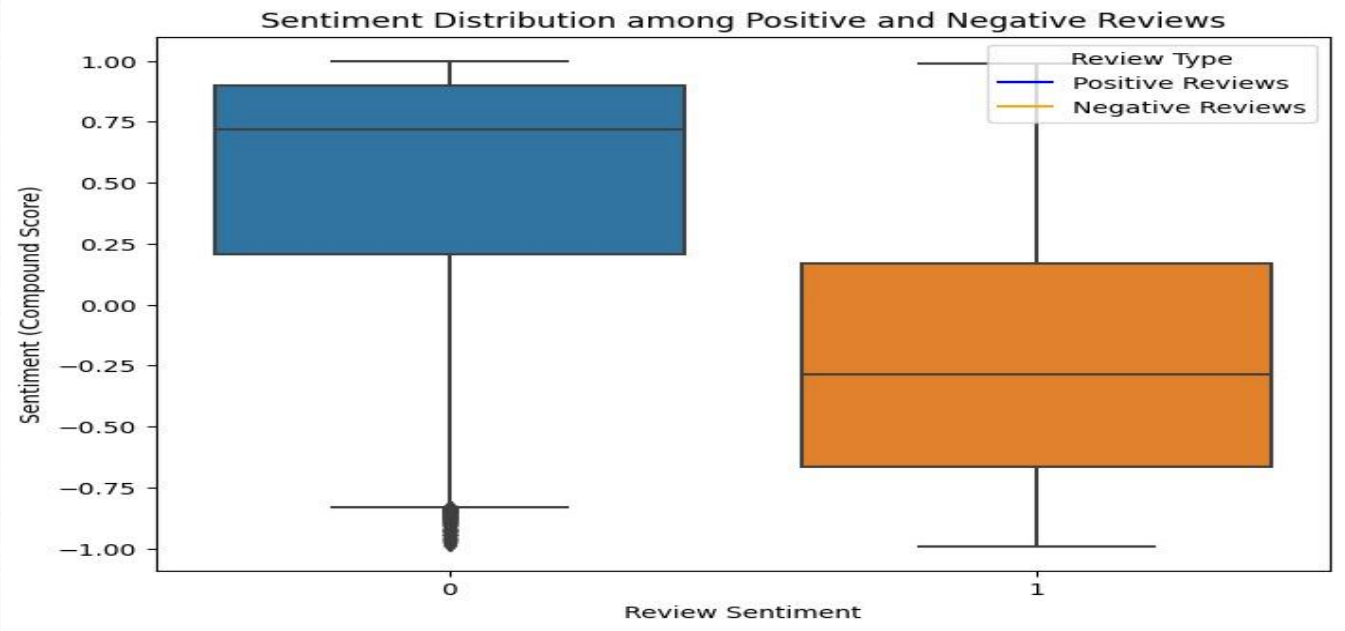
EXPLORATORY DATA ANALYSIS(Cntd…)

- Line Chart depicting the distribution of density of compound scores of the Good and Bad Reviews.
- Red Line depicts Good Reviews
- Blue Line depicts Bad Reviews.



EXPLORATORY DATA ANALYSIS(Cntd..)

- Distribution of Positive and Negative Reviews using Box Plot.



IMPLEMENTATION

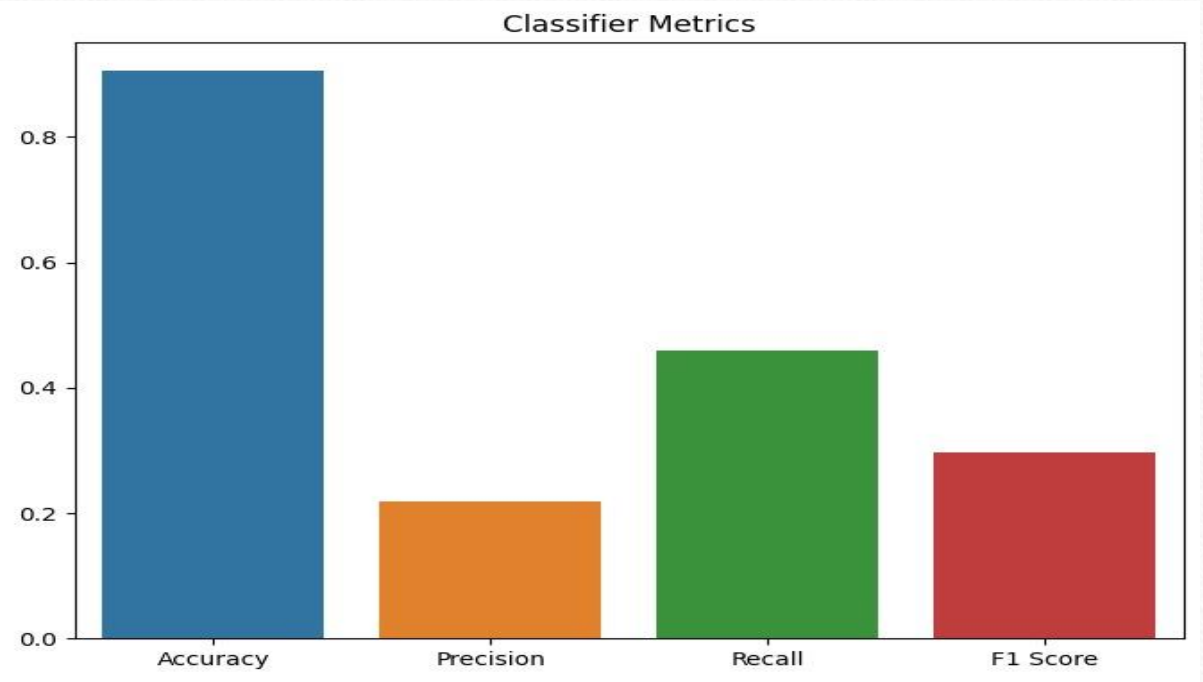
- Algorithm: Random Classifier
  - The Hotel Review is preprocessed, analyzed.
  - The Hotel Review dataset is fed to model for train and access the random forest classifier in this implementation.
  - Data is randomly fetched to train and test data with random sampling parameter as 42.
  - Model evaluation is performed by using evaluation metrix, confusion matrix and also Roc Curve and PR-Curve.

IMPLEMENTATION LIBRARIES

- Pandas
- Sklearn
- Matplot, seaborn
- NLTK
- Popular
- Vader_lexicon
- Random Forest Classifier

RESULTS

- Random Forest Results:
- Accuracy: 90%
- Presicion: 22%
- Recall: 46%
- F1 Score: 30%



IMPLEMENTATION STATUS REPORT

Work completed:

- Successfully implemented machine learning model Random Forest for Hotel Reviews classification.

Description:

- The Good Reviews and Bad Reviews are classified based on NLP Techniques.

Responsibility(Task, Person):

- Data Preprocessing : Sharanya Sriram
- Exploratory Analysis of TextData(visualization) : Sreekar Thanda
- Model Training and Evaluation : Srinivas Sanku

Contributions:

- Sharanya Sriram : 32%
- Sreekar Thanda : 35%
- Srinivas Sankula : 33%

Issues/Concerns:

- Class imbalance.
- High Dimensionality.

REFERENCES

- JavaTpoint. "Decision Tree Classification Algorithm." *JavaTpoint.*, 2021, https://www.javatpoint.com/machine-learning-decision-tree-classification-algorithm.
- Niklas Donges. *Random Forest: A Complete Guide for Machine Learning*. 2023, https://builtin.com/data-science/random-forest-algorithm.
- Tetsuya Nasukawa, Jeonghee Yi. **"entiment analysis: capturing favorability using natural language processing"** [PDF] Sentiment analysis: capturing favorability using natural language processing | Semantic Scholar
- Ch. Sarath Chandra Reddy; K. Uday Kumar; J. Dheeraj Keshav; Bakshi Rohit Prasad; Sonali Agarwal **"Prediction of star ratings from online reviews".** Prediction of star ratings from online reviews | IEEE Conference Publication | IEEE Xplore

Github Link:
https://github.com/ThandaSreekar/NLP-Project/blob/main/analysis_of_hotel_customer_sentiments_new.ipynb

Presentation Video

NLP Project Presentation-20231128_230826-Meeting Recording.mp4 (sharepoint.com)

THANK YOU