



VIT[®]

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

SCHOOL OF COMPUTER SCIENCE AND ENGINEERING

M.Tech Computer Science and Engineering [Integrated]

A Project report on

Analyzing Sentiments in Restaurant Reviews

(Natural Language Processing and Computational Linguistics)

Submitted By

HARSHITA SOROUT (19MIC0047)

INTRODUCTION: -

Sentiment analysis by natural language processing (NLP) is a project aimed at understanding and categorizing the emotions expressed in textual data. Using machine learning algorithms, it analyzes the sentiment behind words and phrases, determining whether they convey positive and negative feelings. This project holds significance in various fields, including social media monitoring, customer feedback analysis, and market research, providing valuable insights into public opinion and sentiment trends.

In today's digital age, online reviews play a crucial role in shaping consumer perceptions and decisions. For restaurants, in particular, customer reviews on platforms like Yelp, TripAdvisor, or Google Reviews can significantly impact their reputation and business success. Sentiment analysis, a branch of natural language processing (NLP), offers a powerful tool to analyze and understand the sentiment expressed in these reviews.

In this context, sentiment analysis refers to the automated process of determining whether a piece of text expresses positive, negative, or neutral sentiment. By harnessing the power of machine learning and linguistic analysis, sentiment analysis algorithms can systematically analyze vast amounts of textual data to extract insights about customer opinions and emotions.

In this project, we focus on sentiment analysis for restaurant reviews using a variety of techniques. We leverage both traditional machine learning algorithms and modern NLP tools to extract meaningful insights from textual data.

PROBLEM STATEMENT

In the era of digital communication, online reviews play a pivotal role in shaping consumer decisions and influencing business reputations. For restaurants, understanding the sentiment expressed in customer reviews is crucial for maintaining and improving their brand image, customer satisfaction, and overall business success. However, manually analyzing a large volume of reviews is time-consuming and inefficient.

Therefore, the problem we aim to address in this project is to develop a sentiment analysis system tailored specifically for restaurant reviews. The system will leverage natural language processing (NLP) techniques and machine learning algorithms to automatically classify reviews into positive, negative, or neutral sentiment categories.

OBJECTIVE

The goal of this project is to develop a machine learning model that can accurately classify the sentiment of restaurant reviews as either positive, negative and neutral.

LITERATURE SURVEY: -

Title	Year	Journal name	Keyword	Abstract
Sentiment analysis using product review data	16 June 2015	Journal of Big Data	Naive Bayesian classifier, Random forest, Support vector machine	<p>Sentiment analysis or opinion mining is one of the major tasks of NLP (Natural Language Processing). Sentiment analysis has gained much attention in recent years.</p> <p>In this paper, we aim to tackle the problem of sentiment polarity categorization, which is one of the fundamental problems of sentiment analysis.</p> <p>A general process for sentiment polarity categorization is proposed with detailed process descriptions.</p> <p>Data used in this study are online product reviews collected from Amazon.com.</p>

Title	Year	Journal Name	Keyword	Abstract
A survey on sentiment analysis methods, applications, and challenges	OCT 2022	COMPUTER SCIENCE, ARTIFICIAL INTELLIGENCE <i>in SCIE edition</i> Source: Journal Citation Reports.	<p>Sentiment analysis Text analysis Word embedding Machine learning Social media</p> <p>Keywords Plus CONVOLUTIONAL NEURAL-NETWORKSUPPORT VECTOR MACHINEFEATURE-SELECTIONASPECT EXTRACTIONHYBRID APPROACHSOCIAL MEDIACLASSIFICATIONTEXTAL GORITHMSFRAMEWORK</p>	<p>This article discusses a complete overview of the method for completing this task as well as the applications of sentiment analysis.</p> <p>Then, it evaluates, compares, and investigates the approaches used to gain a comprehensive understanding of their advantages and disadvantages.</p> <p>Finally, the challenges of sentiment analysis are examined in order to define future directions.</p>

Title	Year	Journal Name	Keyword	Abstract
Sentiment Analysis in Social Media and Its Application: Systematic Literature Review	2019	<p>The Fifth Information Systems International Conference 2019.</p> <p>article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)</p>	Sentiment analysis; Big data; Social media	<p>its application. Social media contain a large amount of raw data that has been uploaded by users in the form of text, videos, photos and audio.</p> <p>The data can be converted into valuable information by using sentiment analysis.</p> <p>A systematic review of studies published between 2014 to 2019 was undertaken using the following trusted and credible database including ACM, Emerald Insight, IEEE Xplore, Science Direct and Scopus.</p>

Title	Year	Journal Name	Keyword	Abstract
Sentiment analysis of social networking sites (SNS) data using machine learning approach for the measurement of depression	2017	2017 <i>International Conference on Information and Communication Technology Convergence (ICTC)</i> Publisher: IEEE	Support vector machines; Social network services; Feature extraction ;Entropy ;Sentiment analysis ;Machine learning algorithms; Speech ;Sentiment Analysis ;Social Networking Sites (SNS) ;Depression Measurements.	In this paper, we present how to find the depression level of a person by observing and extracting emotions from the text, using emotion theories, machine learning techniques, and natural language processing techniques on different social media platforms.

Title	Year	Journal Name	Keyword	Abstract
Multi-kernel SVM based depression recognition using social media data	2019	<i>International Journal of Machine Learning and Cybernetics</i>	Multi- kernal SVM, Single kernal SVM, Naive Bayes, Decision Trees, KNN.	<p>Depression has become the world's fourth major disease. Compared with the high incidence, however, the rate of depression medical treatment is very low because of the difficulty of diagnosis of mental problems.</p> <p>The social media opens one window to evaluate the users' mental status. With the rapid development of Internet, people are accustomed to express their thoughts and feelings through social media.</p> <p>Thus social media provides a new way to find out the potential depressed people. In this paper, we propose a multi-kernel SVM based model to recognize the depressed people.</p> <p>Three categories of features, user microblog text, user profile and user behaviors, are extracted from their social media to describe users' situations. According to the new characteristics of social media language,</p>

Title	Year	Journal Name	Keyword	Abstract
Sentiment Analysis of Social Networking Sites (SNS) Data using Machine Learning Approach for the Measurement of Depression	October 2017	DOI: 10.1109/ICTC.2017.8190959	Sentiment Analysis, Social Networking Sites (SNS), Depression Measurements	In this paper, we present how to find the depression level of a person by observing and extracting emotions from the text, using emotion theories, machine learning techniques, and natural language processing techniques on different social media platforms.

Title	Year	Journal Name	Keyword	Abstract
Sentiment analysis in education research: a review of journal publications	2023	Interactive Learning Environments Volume 31, 2023 - Issue 3	Sentiment analysis , literature review , emotions, education, learning.	Sentiment analysis (SA) is widespread across all fields and has become one of the most active topics in education research, and there is a growing body of papers published. So far, however, there has been little discussion about comprehensive literature reviews in SA in education. Therefore, this study aims to review the high-qualified scientific literature of SA in education and reveals the future research prospects of SA based on the reviewed papers. After systematically searching five online bibliographic databases, 41 relevant articles were located and included in the study.

Title	Year	Journal Name	Keyword	Abstract
Depression detection from social network data using machine learning techniques	2018	Health Inf Sci Syst , AAPS Journal, National library of medicine (National Center for Biotechnology Information).	Social network, Emotions, Depression, Sentiment analysis	In this study, we aim to analyze Facebook data to detect any factors that may reflect the depression of relevant Facebook's users. Various machine learning techniques are employed for such purpose. Considering the key objective of this study, the following are subsequent research challenges addressed in paper.

METHODOLOGY: -

Data Collection and Preprocessing:

We begin by collecting restaurant reviews from various sources and preprocessing the data to ensure consistency and quality. This involves tasks such as removing irrelevant information, handling missing values, and standardizing the text format.

Feature Engineering:

To effectively analyze text data, we transform the raw text into numerical features that machine learning algorithms can understand. This typically involves techniques like tokenization, vectorization, and feature extraction.

Model Training:

We train machine learning models, such as decision trees, using labeled data to learn patterns and associations between textual features and sentiment labels. These models are then capable of predicting the sentiment of unseen reviews.

Sentiment Analysis with Emojis:

Emojis have become an integral part of online communication, adding nuance and emotional context to text-based messages. In our sentiment analysis, we also take into account the presence of emojis in reviews, as they can influence the overall sentiment expressed.

Evaluation and Interpretation:

Finally, we evaluate the performance of our sentiment analysis model and interpret the results to gain insights into customer sentiment towards different restaurants. This allows restaurant owners and managers to identify areas for improvement, capitalize on strengths, and make data-driven decisions to enhance customer satisfaction and loyalty.

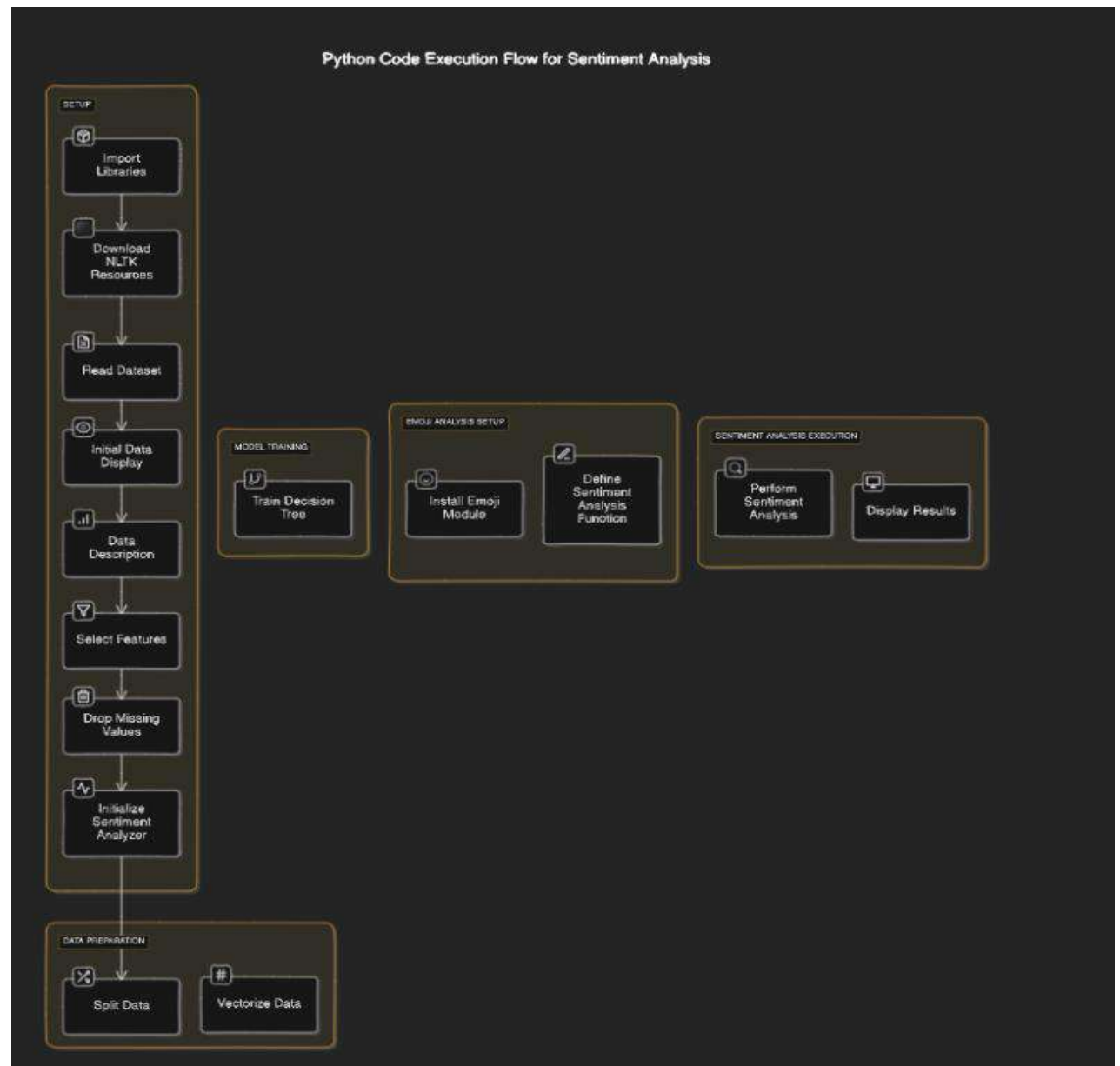
HARDWARE USE: -

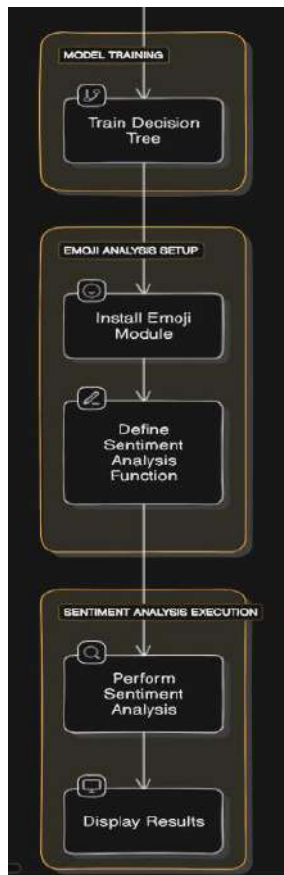
- **Computing Devices:** The system should be compatible with various computing devices such as desktop computers, laptops.
- **Processor:** The hardware should have a processor with sufficient computational power to handle data processing and machine learning model training and inference efficiently.
- **Memory (RAM):** Adequate RAM is required to store and manipulate datasets and perform computations during model training and prediction.
- **Storage:** Sufficient storage space is necessary to store datasets, machine learning models, and application files.
- **Network Connectivity:** Internet connectivity may be required for accessing online resources, downloading updates, and communicating with external servers if applicable.

SOFTWARE USE: -

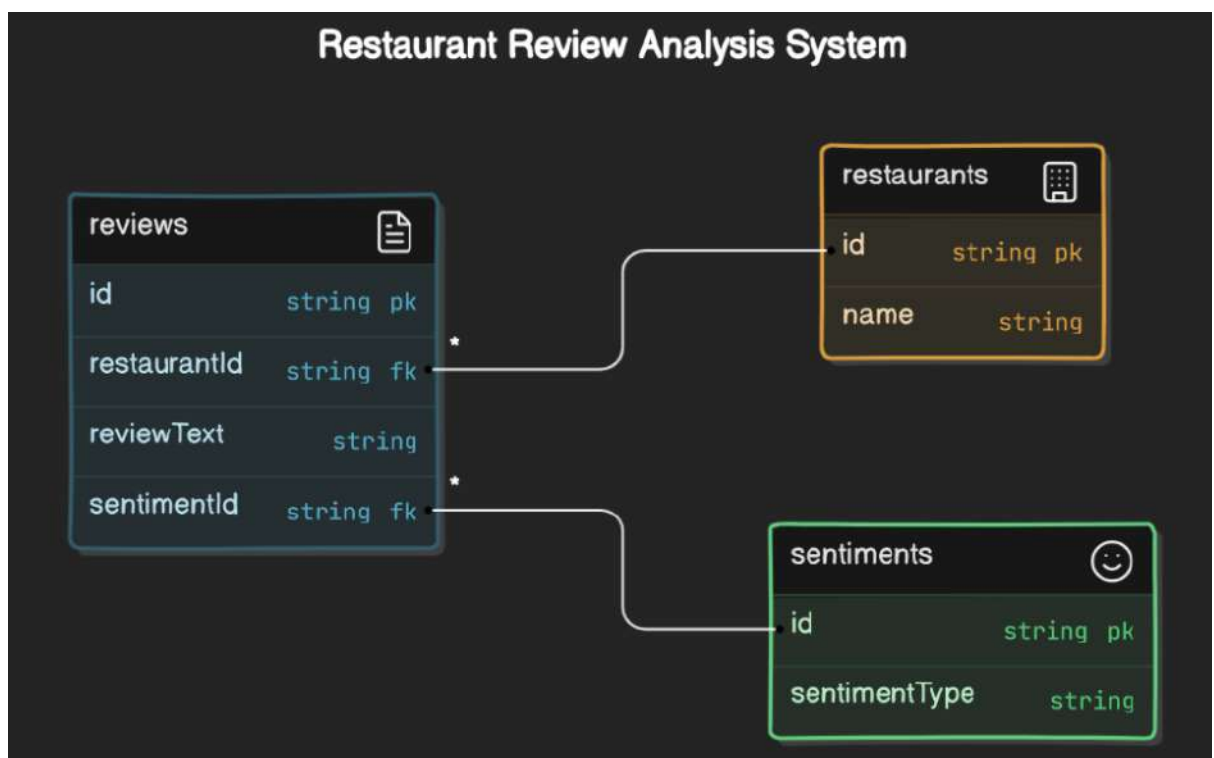
- Goggle colab
- Libraries- emoji, nltk, panda, numpy
- Python language

WORK FLOW DIAGRAM





Entity relationship diagram



IMPLEMENTATION

Code: -

```
import nltk

# Download NLTK resources
nltk.download('vader_lexicon')

#importing functions
import pandas as pd
import numpy as np
import nltk
from nltk.sentiment.vader import SentimentIntensityAnalyzer
import emoji
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.pipeline import make_pipeline
import warnings

#Reding dataset
data = pd.read_csv("Restaurant reviews.csv")
warnings.filterwarnings("ignore")
data.head(4)

data
data.describe()

#Select only required features
data = data[['Restaurant', 'Review']]

# Drop rows with missing values
data.dropna(inplace=True)

# Initialize Sentiment Analyzer
sid = SentimentIntensityAnalyzer()
```

#Split the data into training and testing set

```
X_train, X_test, y_train, y_test = train_test_split(data['Review'], data['Restaurant'],
test_size=0.2, random_state=42)
```

#Convert the text into numerical vector

Vectorize the reviews using CountVectorizer

```
vectorizer = CountVectorizer()
X_train_vectorized = vectorizer.fit_transform(X_train)
X_test_vectorized = vectorizer.transform(X_test)
```

#Train the decision tree

```
clf = DecisionTreeClassifier()
clf.fit(X_train_vectorized, y_train)
```

Install emoji module

```
!pip install emoji
```

#Function to classify sentiment based on emojis

```
import emoji

def sentiment_analysis(review):
    positive_words = ['good', 'great', 'excellent', 'delicious', 'fantastic']
    negative_words = ['bad', 'poor', 'terrible', 'awful', 'disappointing']
    review_lower = review.lower()

    # Check for positive and negative words
    positive_count = sum(word in review_lower for word in positive_words)
    negative_count = sum(word in review_lower for word in negative_words)
    if positive_count > negative_count:
        return "positive review " + "😊 "
    elif negative_count > positive_count:
        return "negative review " + "😞 "
    else:
        # Check for emojis
        has_emojis = any(emoji.is_emoji(c) for c in review)
        if has_emojis:
```

```
        return "neutral review with emojis " + " 😊 "
```

else:

```
        return "neutral review " + " 😊 "
```

Perform sentiment analysis for test reviews

```
sentiments = [sentiment_analysis(review) for review in X_test]
```

Display the sentiment analysis result for each review

```
results = pd.DataFrame({'Review': X_test, 'Sentiment': sentiments})  
print(results)
```

REFERENCES: -

1. <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-015-0015-2>
2. <https://www.webofscience.com/wos/woscc/full-record/WOS:000752164300001?SID=EUW1ED0C3Fo0Hn6v3bieZLLsxiC98>
3. <https://pdf.sciencedirectassets.com>
4. <https://ieeexplore.ieee.org/document/8190959>
5. https://scholar.google.com/scholar?as_q=Multi-kernel+SVM+based+depression+recognition+using+social+media+data&as_occt=title&hl=en&as_sdt=0%2C31
6. https://www.researchgate.net/publication/321289496_Sentiment_Analysis_of_Social_Networking_Sites_SNS_Data_using_Machine_Learning_Approach_for_the_Measurement_of_Depression/citations
7. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6111060/#Sec1title>
8. <https://www.tandfonline.com/doi/abs/10.1080/10494820.2020.1826985>