***Title - Restaurant Review using NLP***

***ABSTRACT***:

The aim of this project is to analyze the restaurant reviews provided by customers using Natural Language Processing (NLP). We will be building a model that can classify a review as positive or negative based on the text provided. The model will be trained using a dataset of restaurant reviews and evaluated on a separate test dataset. The goal is to develop an accurate model that can help restaurant owners to identify areas of improvement and make better decisions for their business.

***OBJECTIVE:***

The objective of this project is to build a machine learning model using NLP techniques to classify restaurant reviews as positive or negative. We will be using Python and various libraries such as scikit-learn, NLTK, and pandas to preprocess the data, extract relevant features, and train the model. The accuracy of the model will be evaluated using metrics such as precision, recall, and F1 score.

***INTRODUCTION:***

In the era of social media and online platforms, reviews are an essential part of any business, especially for restaurants. Customers share their opinions and experiences on various platforms, including review websites, social media, and mobile apps. Analyzing these reviews can help restaurant owners to understand their customers' expectations and preferences and make better decisions for their business.

NLP is a field of Artificial Intelligence that focuses on the interaction between computers and human languages. With the help of NLP, we can extract useful information from large volumes of text data, including sentiments, emotions, and topics. In this project, we will use NLP techniques to analyze restaurant reviews and classify them as positive or negative.

***METHODOLOGY:***

The methodology of this project can be divided into the following steps:

Data collection: We will collect a dataset of restaurant reviews from various sources, including Yelp, TripAdvisor, and Google Reviews. The dataset will include text reviews and their corresponding ratings.

Data preprocessing: We will preprocess the data by removing stop words, punctuations, and special characters. We will also convert the text to lowercase and perform stemming and lemmatization to reduce the number of unique words.

Feature extraction: We will extract relevant features from the preprocessed data using techniques such as Bag of Words, TF-IDF, and Word2Vec. These features will be used as inputs to the machine learning model.

Model training: We will train a machine learning model using the preprocessed and feature-extracted data. We will use algorithms such as Naive Bayes, Logistic Regression, and Support Vector Machines (SVM) to train the model.

Model evaluation: We will evaluate the performance of the model using metrics such as precision, recall, and F1 score. We will also use techniques such as cross-validation and grid search to fine-tune the model and improve its accuracy.

***CODE:***

The code for this project will be written in Python. We will be using various libraries such as scikit-learn, NLTK, pandas, and matplotlib. The code will be divided into different modules, including data preprocessing, feature extraction, model training, and evaluation. We will also create a separate module for the user interface, where the user can input a restaurant review, and the model will classify it as positive or negative.

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

import nltk

from nltk.corpus import stopwords

from nltk.tokenize import word\_tokenize

from nltk.stem import WordNetLemmatizer

from sklearn.feature\_extraction.text import CountVectorizer

from sklearn.feature\_extraction.text import TfidfVectorizer

from sklearn.model\_selection import train\_test\_split

from sklearn.naive\_bayes import MultinomialNB

from sklearn.metrics import classification\_report, confusion\_matrix, accuracy\_score

df = pd.read\_csv('restaurant\_reviews.csv', encoding='ISO-8859-1')

df.drop(['Category'], axis=1, inplace=True)

df['Liked'] = df['Liked'].map({1: 0, 2: 1})

stop\_words = set(stopwords.words('english'))

lemmatizer = WordNetLemmatizer()

def preprocess(text):

text = text.lower()

words = word\_tokenize(text)

words = [word for word in words if word not in stop\_words]

words = [lemmatizer.lemmatize(word) for word in words]

return ' '.join(words)

df['Review'] = df['Review'].apply(preprocess)

X\_train, X\_test, y\_train, y\_test = train\_test\_split(df['Review'], df['Liked'], test\_size=0.2, random\_state=42)

cv = CountVectorizer()

X\_train\_cv = cv.fit\_transform(X\_train)

X\_test\_cv = cv.transform(X\_test)

nb = MultinomialNB()

nb.fit(X\_train\_cv, y\_train)

y\_pred = nb.predict(X\_test\_cv)

print('Confusion matrix:')

print(confusion\_matrix(y\_test, y\_pred))

print('Classification report:')

print(classification\_report(y\_test, y\_pred))

print('Accuracy score:', accuracy\_score(y\_test, y\_pred))

***CONCLUSION:***

In conclusion, this project aims to analyze restaurant reviews using NLP techniques and build a machine learning model that can classify a review as positive or negative. The model can help restaurant owners to identify areas of improvement and make better decisions for their business. The accuracy of the model can be improved by using advanced techniques such as deep learning and by using a larger and more diverse dataset. Overall, this project demonstrates the potential of NLP and machine learning in analyzing customer feedback and improving business operations.