**RESTAURANT REVIEW USING NLP**

ABSTRACT

This project aims at developing a machine learning model to classify restaurant reviews as good or bad. The model uses a labelled set of reviews to predict whether a set of test reviews are good or bad, followed by which, it gives the rating of the restaurant out of 10.

This model uses Naïve Bayes classifier, by Sklearn, to perform analysis on the data, and it was evaluated using its accuracy score and its confusion matrix.

This model was found to have an accuracy of around 74%.

OBJECTIVES

The main goal of this project is to be automate the process of recognising of positive and negative reviews, to understand consumer feedback. This would help businesses to quickly identify areas for improvement and improve consumer experience.

INTRODUCTION

Consumers play a key role in the growth of businesses, as their purchase decisions are what keep the businesses alive. Businesses are therefore accountable to customers and it is essential to improve or tailor the quality of services as per consumer needs.

Keeping this requirement in mind, this model has been trained to recognise positive and negative reviews, so as to analyse customer requirements.

METHODOLOGY

First, a restaurant review dataset of 1000 reviews (source: Kaggle) was imported. It was a labelled dataset that contained one feature; the review, and a column that said whether the respective review was good or bad.

To avoid the unnecessary use of too much of memory, punctuation marks were replaced by a space. All the reviews were converted into lower case to avoid bias based on case, and the words that do not carry much significance to the meaning of the review were removed, and each word was stemmed to convert the words into their shortest forms.

The resulting sentences were added to a corpus of reviews. Count Vectorizer was used to create a bag of words. 80% of the data was used to train the model using Gaussian Naïve Bayes classifier.

The model was evaluated based on the confusion matrix and the accuracy score.

CODE

# Importing required packages

import pandas as pd

import numpy as np

# Importing dataset

dataset = pd.read\_csv("Restaurant\_Reviews.tsv", delimiter = "\t", quoting = 3)

import re

import nltk

#nltk.download('stopwords')

import nltk

from nltk.corpus import stopwords

print(stopwords.words('english'))

from nltk.corpus import stopwords

from nltk.stem.porter import PorterStemmer

corpus = []

for i in range(0,1000):

review = re.sub('[^a-zA-Z]', ' ', dataset['Review'][i])

print(review)

review = review.lower()

review = review.split()

ps = PorterStemmer()

stopwordlist = stopwords.words('english')

stopwordlist.remove('not')

stopwordlist.remove("wasn")

stopwordlist.remove("didn")

# Stemming the words

review = [ps.stem(word) for word in review if not word in set(stopwordlist)]

review = ' '.join(review)

print(review)

corpus.append(review)

print(corpus)

# Creating a Bag of Words

from sklearn.feature\_extraction.text import CountVectorizer

cv = CountVectorizer(max\_features = 1500)

x = cv.fit\_transform(corpus).toarray()

y = dataset.iloc[:,-1].values

np.random.seed(852)

# Splitting into train and test datasets

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

x\_train, x\_test, y\_train, y\_test = train\_test\_split(x, y, test\_size = 0.2)

bad = 0; good = 0

for i in y\_train:

if i == 0:

bad += 1

else:

good += 1

print("No. of bad reviews: ",bad,"\nNo.of good reviews: ",good)

from sklearn.naive\_bayes import GaussianNB

gnb = GaussianNB()

# Training model using Naive Bayes classifier

gnb.fit(x\_train, y\_train)

y\_pred = gnb.predict(x\_test)

# Confusion Matrix

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

cm = confusion\_matrix(y\_test, y\_pred)

print(cm)

# Acuracy of the model

accuracy\_score(y\_test, y\_pred)

# Get the average rating of the restaurant

bad = 0; good = 0

for i in y\_pred:

if i == 0:

bad += 1

else:

good += 1

print("Rating out of 10: ", (good/(good+bad))\*10)

CONCLUSION

This project will help businesses analyse customer feedback and improve their services.

This model was found to have an accuracy score of 0.74 on this dataset, and the average rating of this restaurant turned out to be 6.7 out of 10.

Further scopes of improvement for this project would be to summarise the positive and negative reviews of the customers, to let the companies know where exactly do they have to improve.