

STUDENT DETAILS

Name: Shruti Khare

IBM SkillsBuild ID: shrutikhare612@gmail.com

College Name: JIMS Engineering Management and
Technical Campus, Gr. Noida, Uttar Pradesh

Internship Domain: Artificial Intelligence

Internship Time Period: 18th,Aug 23 to 30th,Sept 23



CONTENT

- 01 STUDENT DETAILS**
- 02 PROBLEM STATEMENT**
- 03 AGENDA**
- 04 PROJECT OVERVIEW**
- 05 WHO ARE THE END USERS ?**
- 06 SOLUTION AND ITS VALUE PROPOSITION**
- 07 CUSTOMIZATION**
- 08 MODELLING TECHNIQUES**
- 09 RESULTS OF THE PROJECT**
- 10 STATISTICS**
- 11 ESSENTIAL LINKS**

PROBLEM STATEMENT



TOPIC

SENTIMENT ANALYSIS OF RESTRAUNT REVIEWS

- Sentimental reviews of customers towards restruant services and their, likes, dislikes are in 0 and 1.
- Python comes with huge number of libraries, many of them are for AI, ML, DL.
- With the help of that we will understand the data and implement step-wise code to solve NLP Problem and apply suitable ML Models in it.



AGENDA

Mission



Sentiment analysis, also referred to as opinion mining, serves as a vital business intelligence tool, facilitating product and service enhancement for companies. This analytical approach empowers businesses to glean actionable insights in diverse domains, offering the following advanced benefits:

- Enhanced Customer Service
- Real-time Analytics
- Brand Monitoring
- Tracking Campaign Performance

In the context of most restaurants, the task of manually recording and analyzing customer reviews is time-consuming and impractical in today's fast-paced environment. My project leverages Machine Learning models and Data Analytics visualization techniques to streamline this process, empowering end-users to make informed decisions and derive actionable insights efficiently.

Vision



PROJECT OVERVIEW

This venture entails undertaking sentiment analysis on restaurant reviews.

The foremost purpose is to **understand the sentiment or feelings** expressed in those critiques. Are they positive, bad, or neutral? This information can assist eating places enhance their offerings based totally on consumer feedback.

PURPOSE

This project focuses on **analyzing written reviews from customers about their dining experiences** in restaurants. The scope doesn't encompass studying different types of comments or aspects of restaurant operations.

SCOPE

- Gain insights into what clients **like and dislike** about their restaurant experiences.
- Provide **actionable suggestions** to restaurants for reinforcing consumer delight.
- Measure **the accuracy** of the sentiment evaluation to **ensure reliable results**.

OBJECTIVE

WHO ARE THE END USERS?



Using this evaluation it's going to help them to beautify these precious insights into their offerings. And thru this they will be capable of make enhancements in the fields of food pride, ambience, main to better customer pride and extended commercial enterprise

RESTRAUNT MANAGERS AND PROPRIETORS



They can be the folks that may be seeking dining studies. Other clients reliable feedback will help them to make a better decision. This assignment-the evaluation will help them make dependable selections based on different clients reports.

CUSTOMERS



This will target that audience who will in charge of promoting restaunt. They may need this statistics to devise advertising campaigns, and it will advantage them to tailoring effective advertising and marketing messages.

MARKETING TEAMS

SOLUTION AND ITS VALUE PROPOSITION

My project is mainly divided into 5 sub sections, following :-

Importing Libraries

At the beginning, I imported essential Python libraries for various tasks:

1. Numpy for linear algebra.
2. Pandas for data manipulation.
3. NLTK for Natural Lang processing.
4. Seaborn, Matplotlib, and Warnings for managing warnings, among others.

Exploratory Data Analysis

Following I've applied a series of code steps to process the data, which involved cleaning the dataset using techniques such as tokenization, stemming, and removing special characters. Also Descriptive Data Analysis have been performed.

Train, Test, Split

Next, I divided the data into two parts: one for training the model and the other for testing its accuracy on data it hasn't seen before.

ML Model Training

In my project, I utilized the various machine learning algorithms like: AdaBoost Classifier, Multinomial Naive Bayes, Random Forest, etc. These algorithms played a key role in efficiently training my data.

Predictions

Finally, I implemented the necessary code to make predictions on sample reviews and assessed the accuracy of these predictions.

CUSTOMIZATION

What distinguishes my project from others lies in its incorporation of three key components:

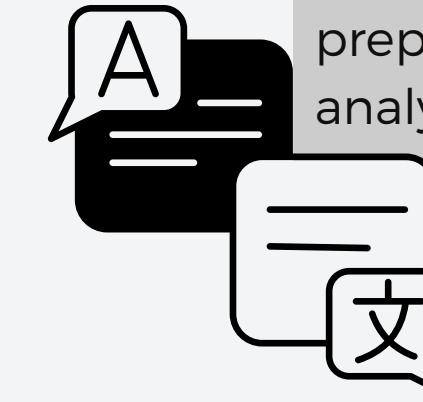
Data Analytics:

This component is essential in our contemporary world, especially when dealing with vast quantities of reviews. It offers an easily comprehensible means of extracting valuable insights from extensive datasets.



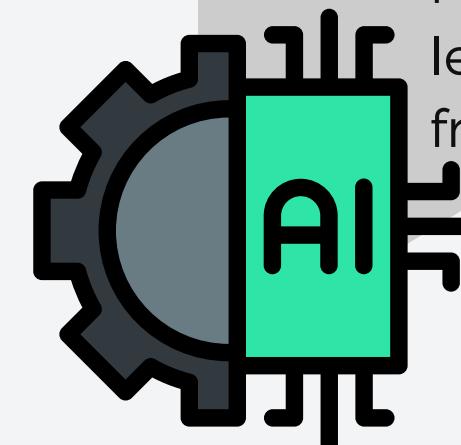
Natural Language Processing (NLP) Libraries:

The project features meticulously structured code for NLP processing, encompassing critical stages such as data cleaning, stemming, lemmatization, stop word removal, and more. This comprehensive approach to text data preprocessing optimizes the accuracy and quality of the analysis results.



Machine Learning Algorithms:

My project integrates a diverse array of machine learning algorithms, ensuring a robust and adaptable framework for making informed decisions.



MODELLING TECHNIQUES

01

02

03

DATA IMPORT AND EXPLORATORY DATA ANALYSIS (EDA):

In the initial phase of the project, the data import process was executed, followed by the integration of Google Drive with Google Colab. Subsequently, a comprehensive descriptive and exploratory data analysis was conducted to gain valuable insights from the dataset.

DATA PREPROCESSING:

Following the data import and initial analysis, critical data preprocessing steps were executed. These encompassed essential procedures such as stemming, stopword removal, tokenization, among others, to enhance data quality and prepare it for further analysis.

MODEL TRAINING:

The project leveraged a diverse array of machine learning algorithms, including Random Forest and Multinomial Naive Bayes. Subsequently, through rigorous training and testing phases, it was determined that the optimal model for this specific task was the AdaBoost Classifier, employing a Stratified K-Fold technique to ensure robust performance.



RESULTS OF PROJECT



MultinomialNB: Training Accuracy=93.38%

MultinomialNB : Testing Accuracy=76.50%

MultinomialNB : Accuracy Difference=16.88%

RandomForestClassifier: Training Accuracy=99.69%

RandomForestClassifier : Testing Accuracy=74.20%

RandomForestClassifier : Accuracy Difference=25.49%

AdaBoostClassifier: Training Accuracy=81.31%

AdaBoostClassifier : Testing Accuracy=74.90%

AdaBoostClassifier : Accuracy Difference=6.41%

SVC: Training Accuracy=96.09%

SVC : Testing Accuracy=78.00%

SVC : Accuracy Difference=18.09%

LinearSVC: Training Accuracy=99.26%

LinearSVC : Testing Accuracy=77.50%

LinearSVC : Accuracy Difference=21.76%

NuSVC: Training Accuracy=98.60%

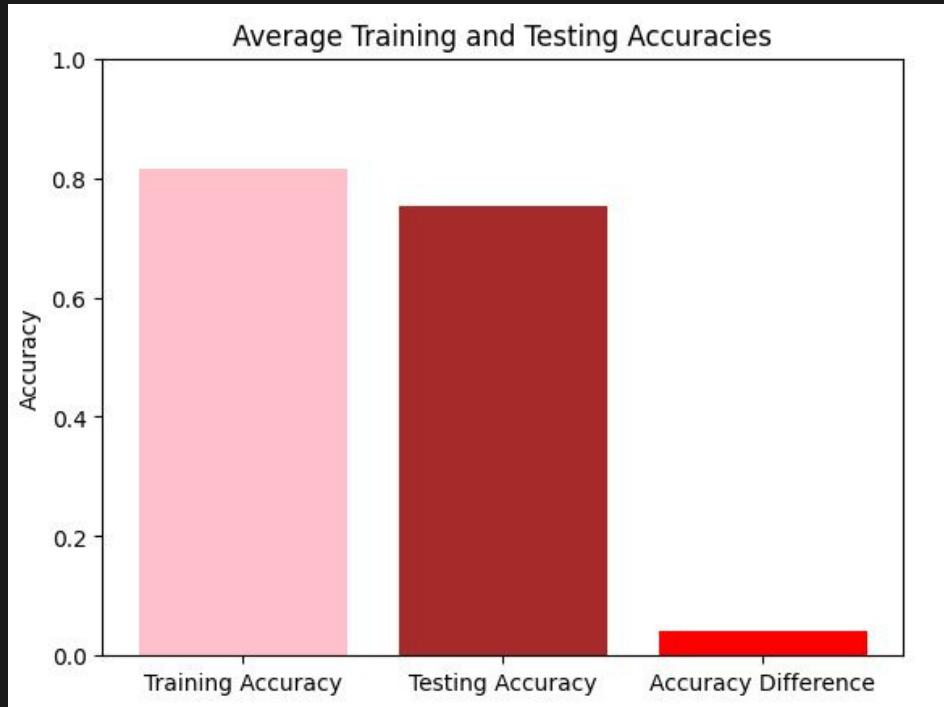
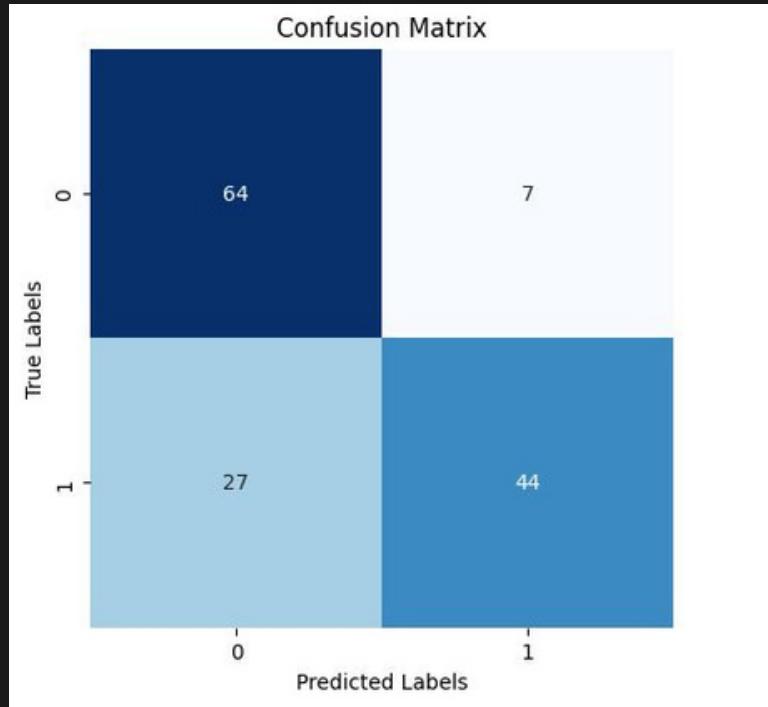
NuSVC : Testing Accuracy=78.40%

NuSVC : Accuracy Difference=20.20%

LogisticRegression: Training Accuracy=95.93%

LogisticRegression : Testing Accuracy=77.90%

LogisticRegression : Accuracy Difference=18.03%



Evaluation metrics used here are:

Using AdaBoostClassifier,

- Training Accuracy: 81.31 %
- Testing Accuracy: 74.90 %
- Difference b/w both accuracies: 5.88 %
- Precision Score: 0.86
- Recall score: 0.60

Using AdaBoostClassifier with Stratified K-Fold,

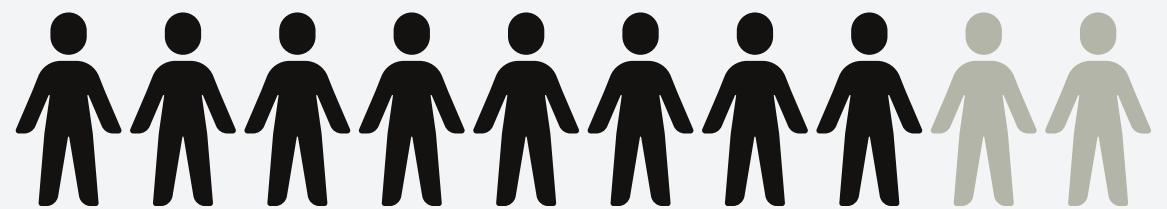
- Training Accuracy: 81.55 %
- Testing Accuracy: 75.30 %
- Difference b/w both accuracies: 4.07 %
- Precision Score: 0.85
- Recall score: 0.61



STATISTICS

AdaBoostClassifier with Stratified K-FoldCross Validation Technique is slightly better method compared to others, with 81.55% as Training accuracy and 75.30% as Testing Accuracy which means the model built for the prediction of sentiment analysis of the restaurant reviews gives 75.30% right prediction.

81.55%



ESSENTIAL LINKS

LinkedIn

shrutikhare
4496xyz

DataSet

Restraunt
Reviews

Gmail

shrutikhare
612

Google Colab

GitHub

Khare
Shruti

Google Colab

My
Project

THANK YOU

I wish to express my profound gratitude to IBM Skillsbuild and the Edunet Foundation for granting me the invaluable opportunity to partake in an Artificial Intelligence internship, which has proven to be a golden learning experience.

I am sincerely thankful to my mentor, Mr. Lakshman, whose commitment to knowledge dissemination was exemplified through his diligent scheduling of weekly classes over the course of six weeks. His guidance and expertise were instrumental in shaping my understanding of Artificial Intelligence.

Furthermore, I extend my heartfelt appreciation to Mr. Channabasva for his exceptional coordination and unwavering support throughout the entirety of the internship. His dedication in addressing queries and providing guidance was instrumental in my growth.

This internship has provided me with an in-depth understanding of cutting-edge technologies and the vast domain of Artificial Intelligence. It has been a transformative journey, and I am genuinely grateful for this enriching experience.

