Details

**Student Name:- Gayatri Pramod Chaudhari**

**Seat No:-**

**Project Name:- Restaurant Review Analysis System.**

**Organization Name:- MySoft Services.**

**Start Date:- 27Feb 2022**

**End Date:- 31st May 2022**

A

Project Report

on

Restaurant Review Analysis System

at



Submitted By:

Gayatri Pramod Chaudhari

to

School of Computer Sciences,

Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon

In the partial fulfillment of the requirement for the award of the degree of Master of Science (Computer Science)

June 2022

|  |  |
| --- | --- |
|  | || अंतरी पेटवू ज्ञानज्योत ||  **कवयित्री बहिणाबाई चौधरी उत्तर महाराष्ट्र विद्यापीठ, जळगाव**  **संगणकशास्त्र प्रशाळा,** |
| **KAVAYITRI BAHINABAI CHAUDHARI NORTH MAHARASHTRA UNIVERITY, JALGAON**  **SCHOOL OF COMPUTER SCIENCES,**  **Prof. S. R. Kolhe प्रा. स. र. कोल्हे**  **Director संचालक**  CERTIFICATE  This is to certify that **Gayatri Pramod Chaudhari**, a final year student of 'Master of Science (Computer Science) from School of Computer Sciences, Kavayitri Bahinabai Chaudhari North Maharashtra University, Jalgaon has successfully completed the project entitled **(Restaurant Review Analysis System )** as a part of six months full-time industrial training during the academic year 2021-22.  **Head,**  Department of Computer Science,  School of Computer Sciences,  KBCNMU Jalgaon | |



**CERTIFICATE**

**MYSoft Services**

**TO WHOMSOEVER IT MAY CONCERN**

Dear Gayatri Pramod Chaudhari,

We are delighted to welcome you for the internship in software Development in our fellowship. This internship is observed by MYSoft Services s being a learning opportunity for you.

This passion is scheduled is being 27/02/2022 and will be a four-month internship opportunity ending on 31/05/2022

In essence, your internship will embrace orientation and give emphasis on learning new skills with a deeper understanding of concepts through the hands-on application of the knowledge you gained as an intern. Our Team is confident that you will acknowledge your obligation to perform all work allocated to you to the best of your ability within the lawful and reasonable direction given to you.

We look forward to a worthwhile and fruitful association that will make you make you equipped for future projects. Wid=shing you the most enjoyable and truly meaningful internship program experience.

With Best Wishes, Pramod M. Chaudhari

MYSoft Services Authorized Signatory



**ACKNOWLEDGEMENT**

The project “**Restaurant Review Analysis System** ” was the result of persistent efforts, which helped me to bring out my creativity in me to the fore. During the few months that had gone into executing this project, I have interacted with a number of people who have guided me through various stages of the project. I take this opportunity to express my gratitude to them, whose guidance helped me bridge the gap between academics and industry.

I express my profound thanks to my project manager **Mr. Parmod Chaudhari**. Provided me with the necessary resources and facilities and their undivided attention and support without which this project could not be a success.

I also would like to thank the staff members of my Department without whom I would not have achieved this level. My thanks go again to all staff members of **MYSoft Sevices**. Who help me directly or indirectly. I would like to thank my Head of Department **Dr. S.R.Kolhe** our Department teachers as well as the non-teaching staff who helped me indirectly.

I would also like to thank all those who have directly or indirectly helped me in this project.

Last but not the least; I would like to extend my gratitude to my parents, who have always been my source of inspiration and motivation.

Miss.Gayatri Pramod Chaudhar

School of Computer Science, Kavayatri Bahinabai Chaudhari North Maharashtra University, Jalgoan.

**ABSTRACT**

*One of the most effective tools any restaurant has is the ability to track food and*

*beverage sales daily. Currently, Recommender systems play an important role in both*

*academia and industry. These are very helpful for managing information overload. In*

*this paper, we applied machine learning techniques for user reviews and analyze*

*valuable information in the reviews. Reviews are useful for making decisions for both*

*customers and owners. We build a machine learning model with Natural Language*

*Processing techniques that can capture the user's opinions from users’ reviews. For*

*experimentation, the python language was used.*

***Keywords:*** *recommender systems, machine learning, python*

**INDEX**

|  |  |  |  |
| --- | --- | --- | --- |
| **Sr.No** | **Topic** | | |
| 1 | Introduction | | |
|  |  | | |
|  | 1.1 |  | Company Profile |
|  | 1.2 |  | Introduction to project |
|  |  | 1.2.1 | Aim Of the Project |
|  |  | 1.2.2 | Problem Definition |
| 2 | Import Dataset | | |
|  | 2.1 |  | Importing necessary Packages |
|  | 2.2 |  | Dataset |
| 3 | Text Processing | | |
| 4 | Approach | | |
| 5 | Model Training Using MultinomialNB | | |
| 6 | Model Performance | | |
| 7 | Technical use. | | |
| 8 | Deployment Using GUI | | |
| 9 | Result | | |
| 10 | Conclusion | | |
| 11 | Future Scope | | |
| 12 | Reference | | |

* 1. **Company profile-**
  2. **Introduction to Project-**

This project is related to the Restaurant Rating prediction.

Restaurant customers give their ratings and write reviews based on their satisfaction

levels. These ratings and reviews help the other customers to make a decision on going to

those restaurants. These ratings are also helpful for the restaurant owners to make

changes based on their reviews for improving their business Restaurant reviews contains

textual information. But most machine learning algorithms work with numerical

data only. Machine learning can be considered one of the applications of artificial

intelligence (AI).ML provides a way to learn the systems without being explicitly

programmed and this learning can be used for solving problems. Machine learning takes

data as input and it learns some important relations from data to make decisions as per

user requirements. The learning process starts with the observations like samples, and direct

experience and then finding patterns in that data to make better decisions to predict or

classify new things in the future. For text, processing machine learning provides Natural

processing (NLP) capabilities. We can easily analyze our textual datasets

through NLP methodologies.NLP provides an opportunity for data analysts to apply

machine learning and deep learning algorithms to our textual datasets. We make use of

machine learning algorithms for classifying reviews and recommending the best restaurant.

In general, the methods implemented in a recommender system are three types namely

Content-based Methods, Collaborative Methods, and Hybrid Methods. content-based

methods depend on likenesses between the reviews of the users. It prescribes items to a

client dependent on recently evaluated most noteworthy things by a similar client.

Generally, we need to construct customer-profile data and item-profile data by using the

content of shared attribute space. For example, consider a movie, we can represent it

with the movie stars in it and the genres. For customer profiles, we can do the same thing

based on the user’s likes some movie stars/genres, etc. For calculating how good a movie

is, we may use cosine similarity. Collaborative techniques are based on user behaviour

for a recommendation of items. These methods don’t need anything else except users’

historical preference on a set of items. Because it’s based on historical data, the core

assumption here is that the users who have agreed in the past tend to also agree in the

future. In terms of user preference, it is usually expressed in two categories. The hybrid

method comprises both the features of content-based methods and collaborative methods.

**1.2.1 Aim Of The Project –**

Restaurant customers give their ratings and write reviews based on their satisfaction

levels. These ratings and reviews help the other customers to make a decision on going to

those restaurants. These ratings are also helpful for the restaurant owners to make

changes based on their reviews for improving their business Restaurant reviews contains

textual information

**Literature Survey**

Machine Learning is not a new technique for text processing. Various researchers

applied machine learning techniques for restaurant revreviewassification. M.

Govindarajan [1] et.al proposed a hybrid classification model for sentiment analysis of

restaurant reviews. They proposed an ensemble classifier com comprising a support vector

machine and Naive Bayes models. With their model, they achieved an accuracy of 90%.

Sasikala. proposed a model for classifying restaurant reviews using sentiments

in the words. Their model is based on the score combined with existing text analyzing

packages. Most people use 'yelp' for finding good resrestaurantselp reviews are very

helpful for finding a good restaurant. Boya proposed support vectormachines

for analyzing Restaurant Features using Sentiment Analysis on Yelp Reviews

Kirange[4] et.al also proposed a Support Vector classifier for Emotion Classification of

Restaurant Reviews. They compared their model with Naive Bayes, K-NN ad neural

network models and Shoshone at SVM achieved good results. Tri Doan [5] et.al

proposed a variant of online random forest classifiers for performing sentiment analysis

on user reviews. They showed that their model achieved an accuracy similar to offline

methods. Ekaterina Pronoza[6] et.al proposed a restaurant information extractionmethod

for the restaurant recommendation system. Veda Waikul [7] et.al proposed an SM

classifier for classifying restaurant reviews. With their model, they achieved an accuracy

of 77%.

**Natur**a**l Language Generation:-**

Natural Language Generation (NLG) is the process of producing phrases, sentences, and paragraphs that are meaningful from an internal representation. It is a part of Natural Language Processing and happens in four phases: identifying the goals, planning on goals may be achieved by evaluating the situation and available communicative sources, and realizing the plans as a text.



Component of NLG

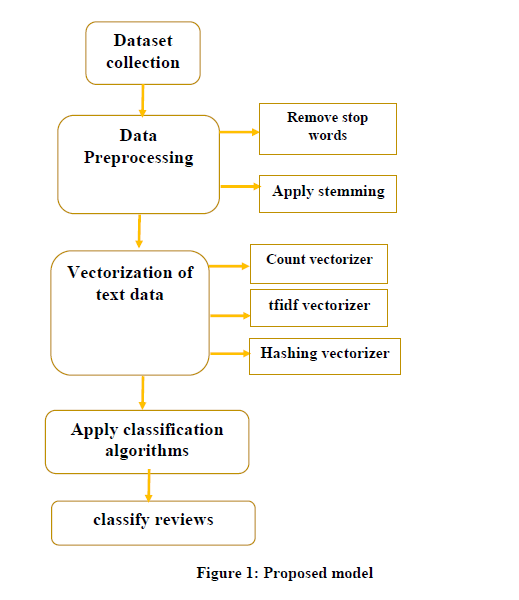
1. **Research Methodology**

The dataset is downloaded from superdatascience.com. The dataset contains reviews in the

text format. It also contains a feature that indicates whether the review is positive or negative.

A positive review can de indicate by 1 and a negative review is indicated by 0. This dataset contains 1000 reviews

**Structure of Proposed model:**



Generally, we need a procedure for representing text information for the ML algorithm.

Bag-of-words are useful to complete this task. This model is simple to implement. It is one

of the methods to extract features from the given text for machine learning models. The Bag

of Words model is used to preprocess the input text by changing it into a bag of

words. The bow can be represented using a table, which contains the count of words

corresponding to the word itself.

**Problem Definition:-**

Normally, a lot of businesses are remained as failures due to lack of profit, lack of proper improvement measures. Mostly, restaurant owners face a lot of difficulties to improve their productivity. This project really helps those who want to increase their productivity, which in turn increases their business profits. This is the main objective of this project.

What the project does is that the restaurant owner gets to know about drawbacks of his restaurant such as most disliked food items of his restaurant by customer’s text review which is processed with ML classification algorithm(Naive Bayes).

**Motivation:-**

As Natural language processing is used in many artificial intelligence applications, and it is an area where lots of work is already done, and lotswordsword need to be done in the future. Natural language processing can be beneficial for both consumers and businesses. The work given by SemEval2014[3], is a similar kind of work, where customers and owners of the restaurant can use our application to find their beneficial information. Earlier applications using the unsupervised method are not able to give a usable system, which cannot be used to find the right information. I found myself that I can contribute to this method in a positive manner. The unsupervised method is usable, where the labeled data is not available to use supervised learning methods for classifying row data.

**Import Dataset**

**Importing necessary Packages:-**

**Import pandas as pd:-**

**pandas** (all lowercase) is a popular Python-based data analysis toolkit which can be imported using import pandas as pd. It presents a diverse range of utilities, ranging from parsing multiple file-formats to converting an entire data table into a [NumPy](https://www.educative.io/edpresso/how-to-create-an-array-in-numpy) matrix array. This makes pandas a trusted ally in data science and machine learning.

Similar to NumPy, pandas deals primarily with data in 1-D and 2-D arrays; however, pandas handle the two differently.

**Import nltk:-**

**Natural Language Processing (NLP)** is a process of manipulating or understanding the text or speech by any software or machine. An analogy is that humans interact and understand each other’s views and respond with the appropriate answer. In NLP, this interaction, understanding, and response are made by a computer instead of a human.

NLTK is Natural Language Tool Kit. It is used to build python programming. It helps to work with human languages data. It gives a very easy user interface. It supports classification, steaming, tagging, etc.

**Import re**

A RegEx, or Regular Expression, is a sequence of characters that forms a search pattern.RegEx can be used to check if a string contains the specified search pattern.

Python has a built-in package called re, which can be used to work with Regular Expressions.

Import the re module:

import re

Regular expressions are a powerful language for matching text patterns. This page gives a basic introduction to regular expressions themselves sufficient for our Python exercises and shows how regular expressions work in Python. The Python "re" module provides regular expression support.

**Dataset**:- Our dataset consists of 457,023 restaurant reviews from OpenTable.com, a service for finding restaurants and making reservations online. The reviews span 11,067 restaurants. The reviews range

in length from 1 to 750 characters, with an average of 248 characters per review. Because the reviews are quite short on average, we need efficient techniques that make ample use of the little available data.

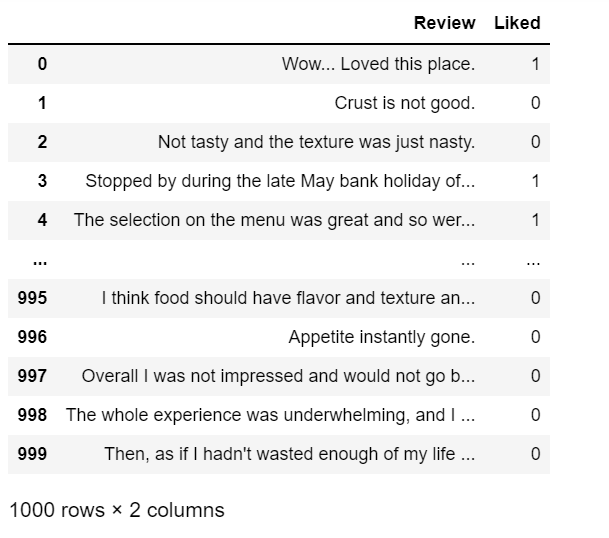
For the purpose of development, we use only a small portion of these reviews when implementing and testing our models. We select 5,000 reviews from

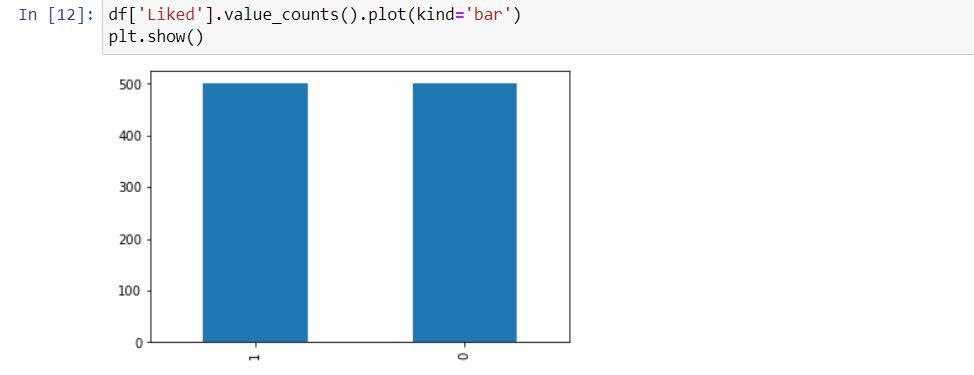
the data as a test set and use an additional 15,000 reviews as a training set. For our best model, we experimented with varying amount of training data to

see if we could further improve performance. As a final note, some reviews in our data set are quite short (i.e. only a few characters or less than a

sentence). Because we want to focus on actually predicting the sentiment of reviews based on their linguistic content, we exclude from both training and

testing any reviews that have fewer than 100 characters, which we estimate to be the length of a short sentence.

****

****

**Implementation:-**

We adapted the Maximum Entropy classifier implemented in a previous assignment for a use with

the restaurant review dataset. As a first pass at feature selection, we simply included, for each review a

feature for every unigram contained in that review. This simple implementation brought precision up to

0.537, well above the baseline level. The mean average error also improved to 0.572, far better than the

baseline established above. We experimented with different methods of preprocessing

the data. Because the reviews are unstructured in terms of user input, reviews can look

like anything from a paragraph of well-formatted text to a jumble of seemingly unrelated words to a run-on

sentence with no apparent regard for grammar or punctuation. Our initial pass over the data simply

tokenized the reviews based on whitespace and treated each token as a unigram, but we were able to

improve performance by removing punctuation in addition to the whitespace and converting all letters

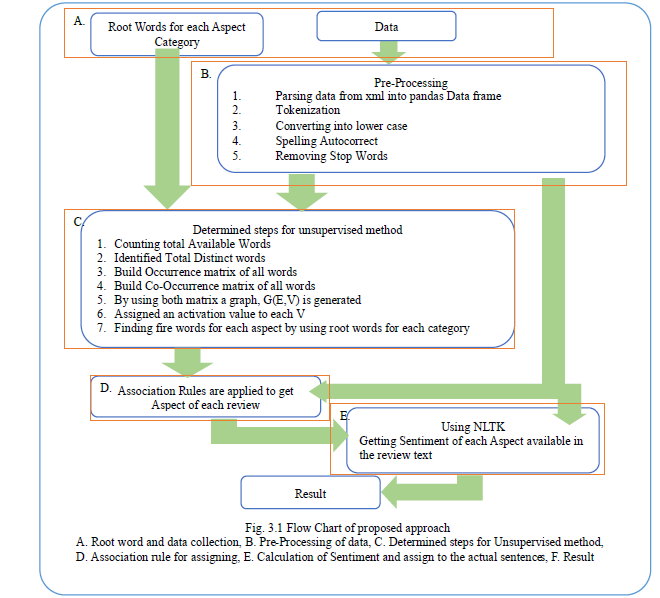
to lowercase. In this way, we treat the occurrences of “good”, “Good”, and “good.” all as the same, which

gives better predictive power to any test set review containing any of these three forms. Through pre-processing, we were able to improve precision by

1.5% to 0.552.

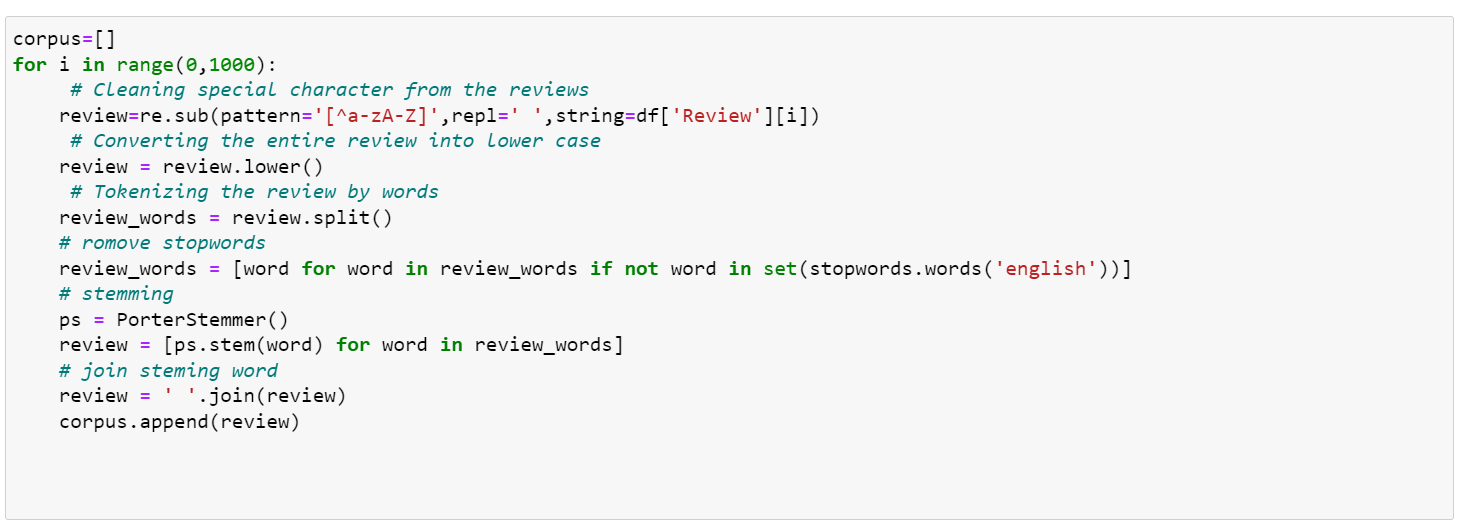
**Approach:-**

A system that can classify the textual data(3044 restaurant reviews) under given aspect categories, Service, Food, Ambience, Price, and Anecdotes/Miscellaneous need to be very accurate because many systems are already been introduced and the last best one is[1] the work is published in IEEE Transaction paper, April 2017. The idea to solve this aspect of category detection problem by unsupervised machine learning, that the paper used a method called unsupervised, its accuracy



**Text Processing by using NLTK:-**

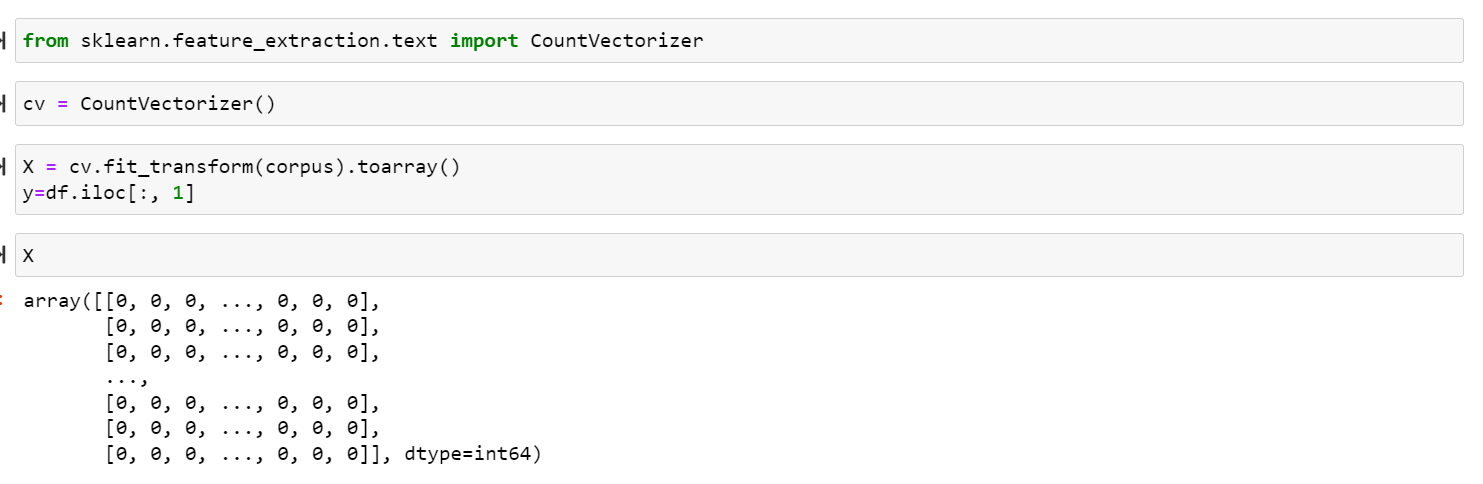
NLTK consists of the most common algorithms such as **tokenizing, part-of-speech tagging, stemming, sentiment analysis, topic segmentation, and named entity recognition**. NLTK helps the computer to analysis, preprocess, and understand the written text.

****

**Countvectorizer:-**

CountVectorizer means **breaking down a sentence or any text into words by performing preprocessing tasks like converting all words to lowercase, thus removing special characters**. In NLP models can't understand textual data they only accept numbers, so this textual data needs to be vectorized.

CountVectorizer is a great tool provided by the sci-kit-learn library in Python. It is **used to transform a given text into a vector on the basis of the frequency (count) of each word that occurs in the entire text**

****

**Train Test split:-**

The train-test split is a technique for evaluating the performance of a machine learning algorithm.

It can be used for classification or regression problems and can be used for any supervised learning algorithm.

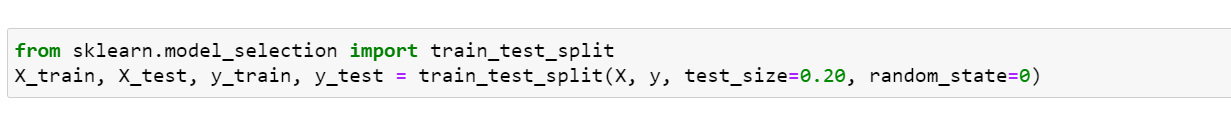
The procedure involves taking a dataset and dividing it into two subsets. The first subset is used to fit the model and is referred to as the training dataset. The second subset is not used to train the model; instead, the input element of the dataset is provided to the model, then predictions are made and compared to the expected values. This second dataset is referred to as the test dataset.

* **Train Dataset**: Used to fit the machine learning model.
* **Test Dataset**: Used to evaluate the fit machine learning model.

The objective is to estimate the performance of the machine learning model on new data: data not used to train the model.T

This is how we expect to use the model in practice. Namely, to fit it on available data with known inputs and outputs, then make predictions on new examples in the future where we do not have the expected output or target values.

The train-test procedure is appropriate when there is a sufficiently large dataset available

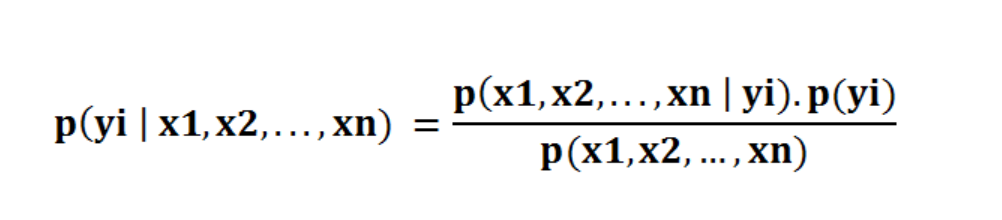
****

**Model Training Using MultinomialNB**

**1 Naive Bayes:-**

Bayes’ theorem which was given by Thomas Bayes, a British Mathematician, in 1763 provides a means for calculating the probability of an event given some information.

Naive Bayes is a supervised learning algorithm for classification so the task is to find the class of observation (data point) given the values of features. A naive Bayes classifier calculates the probability of a class given a set of feature values (i.e. p(yi | x1, x2 , … , xn)).Input this into Bayes’ theorem:



Bayes’ rule provides us with the formula for the probability of Y given some feature X. In real-world problems, we hardly find any case where there is only one feature.

When the features are independent, we can extend Bayes’ rule to what is called Naive Bayes which assumes that the features are independent that means changing the value of one feature doesn’t influence the values of other variables and this is why we call this algorithm “*NAIVE*”

# Probability and conditional probability

We’ve learned where the “naive” comes from. How about the “Bayes”? Bayes comes from the famous [Bayes Theorem](https://en.wikipedia.org/wiki/Bayes%27_theorem) of Thomas Bayes. To get a comprehensive understanding of Bayes’ Theorem, we should talk about probability and conditional probability first.

Bayes’ rule provides us with the formula for the probability of Y given some feature X. In real-world problems, we hardly find any case where there is only one feature.

Probability simply means the likelihood of an event to occur and always takes a value between 0 and 1 (0 and 1 inclusive). The probability of event A is denoted as **p(A)** and calculated as the number of the desired outcome divided by the number of all outcomes. For example, when you roll a die, the probability of getting a number less than three is 2 / 6. The number of desired outcomes is 2 (1 and 2); the number of total outcomes is 6.

When the features are independent, we can extend Bayes’ rule to what is called Naive Bayes which assumes that the features are independent which means changing the value of one feature doesn’t influence the values of other variables and this is why we call this algorithm “*NAIVE*”

Naive Bayes can be used for various things like face recognition, weather prediction, Medical Diagnosis, News classification, Sentiment Analysis, and a lot more.

# Naive Bayes Classifier

Naive Bayes is a supervised learning algorithm used for classification tasks. Hence, it is also called Naive Bayes Classifier.

Like other supervised learning algorithms, naive Bayes uses features to make a prediction on a target variable. The key difference is that naive Bayes assume that features are independent of each other and there is no correlation between features. However, this is not the case in real life. This naive assumption of features being uncorrelated is the reason why this algorithm is called “naive”.

## Assumptions of Naive Bayes

· All the variables are independent. That is if the animal is Dog that doesn’t mean that the Size will be Medium

· All the predictors have an equal effect on the outcome. That is, the animal being dog does not have more importance in deciding If we can pet him or not. All the features have equal importance.

We should try to apply the Naive Bayes formula to the above dataset however before that, we need to do some precomputations on our dataset.

1. **Application:-**

This is only for maintaining the model of the situation. Here the speaker just initiates the process and s doesn’t take part in the language generation. It stores the history, structures the content that is potentiallrelevantva,nt and deploys a representation of what it actually knows. All these form the situation while selecting the g subset of propositions that hee t speaker has. The only requirement is the speaker has to make sense of the situation.

**Model Performance:-**

* **Accuracy score:-**

Accuracy is pretty high but this is a simple task used to show the concept and go through the steps of implementation. One important thing to remember is that the accuracy on the training set should not be much higher than the accuracy on the test set which indicates our model is too specific and not generalized well. This results in overfitting which is a serious problem for any machine learning algorithm.

* **Classification report:-**

A Classification report is used to measure the quality of predictions from a classification algorithm. How many predictions are True and how many are False. More specifically, True Positives, False Positives, True negatives, and False Negatives are used to predict the metrics of a classification report as shown below. The report is copied from related to K-Means on Iris Dataset.

The report shows the main classification metrics precision, recall, and f1-score on a per-class basis. The metrics are calculated by using true and false positives, and true and false negatives. Positive and negative in this case are generic names for the predicted classes. There are four ways to check if the predictions are right or wrong:

1. **TN / True Negative:**when a case was negative and predicted negative
2. **TP / True Positive:**when a case was positive and predicted positive
3. **FN / False Negative:**when a case was positive but predicted negative
4. **FP / False Positive:**when a case was negative but predicted positive

### ****Precision – What percent of your predictions were correct?****

Precision is the ability of a classifier not to label an instance positive that is actually negative. For each class, it is defined as the ratio of true positives to the sum of true and false positives.

**TP – True Positives**

**FP – False Positives**

**Precision – Accuracy of positive predictions.**

**Precision = TP/(TP + FP)**

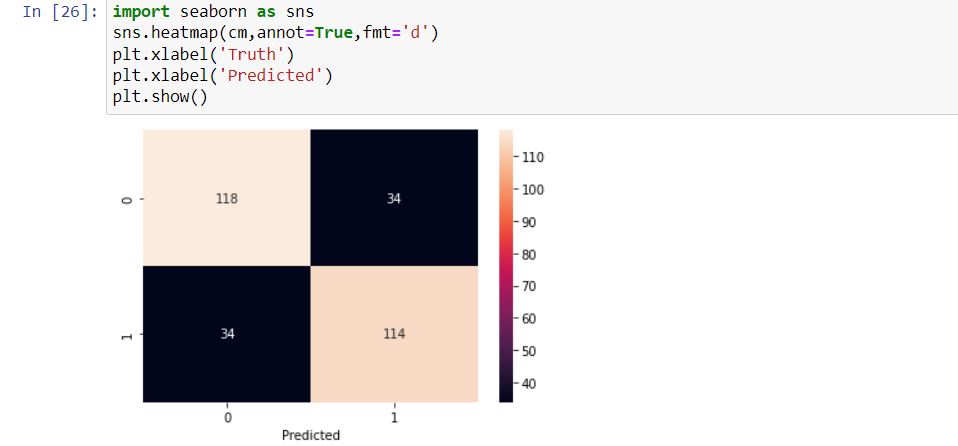
### ****Recall – What percent of the positive cases did you catch?****

The recall is the ability of a classifier to find all positive instances. For each class, it is defined as the ratio of true positives to the sum of true positives and false negatives.

**FN – False Negatives**

Recall Fraction of positives that were correctly identified.  
Recall = TP/(TP+FN)

**Confusion Matrix:-**

****

**Tools & Technologies Used:**

**NLTK:-**

NLTK is a leading platform for building Python programs to work with human language data. It provides easy-to-use interfaces to over 50 corpora and lexical resources such as WordNet, along with a suite of text processing libraries for classification, tokenization, stemming, tagging, parsing, and semantic reasoning, wrappers for industrial-strength NLP libraries, and an active discussion forum.

Thanks to a hands-on guide introducing programming fundamentals alongside topics in computational linguistics, plus comprehensive API documentation, NLTK is suitable for linguists, engineers, students, educators, researchers, and industry users alike. NLTK is available for Windows, Mac OS X, and Linux. Best of all, NLTK is a free, open-source, community-driven project.

NLTK has been called “a wonderful tool for teaching and working in, computational linguistics using Python,” and “an amazing library to play with natural language.”

**Machine Learning:-**

Machine learning (ML) is a type of artificial intelligence ([AI](https://www.techtarget.com/searchenterpriseai/definition/AI-Artificial-Intelligence)) that allows software applications to become more accurate at predicting outcomes without being explicitly programmed to do so. Machine learning [algorithms](https://www.techtarget.com/whatis/definition/algorithm) use historical data as input to predict new output values.

Machine learning is important because it gives enterprises a view of trends in customer behavior and business operational patterns, as well as supports the development of new products. Many of today's leading companies, such as Facebook, Google and Uber, make machine learning a central part of their operations. Machine learning has become a significant competitive differentiator for many companies.

**Python:-**

Python is an interpreted, object-oriented, high-level programming language with dynamic semantics. Its high-level built in data structures, combined with dynamic typing and dynamic binding, make it very attractive for Rapid Application Development, as well as for use as a scripting or glue language to connect existing components together. Python's simple, easy to learn syntax emphasizes readability and therefore reduces the cost of program maintenance. Python supports modules and packages, which encourages program modularity and code reuse. The Python interpreter and the extensive standard library are available in source or binary form without charge for all major platforms and can be freely distributed.

**Tkinter:-**

Tkinter is the standard GUI library for Python. Python when combined with Tkinter provides a fast and easy way to create GUI applications. Tkinter provides a powerful object-oriented interface to the Tk GUI toolkit.

Python offers multiple options for developing GUI (Graphical User Interface). Out of all the GUI methods, Tkinter is the most commonly used method. It is a standard Python interface to the Tk GUI toolkit shipped with Python. Python with Tkinter is the fastest and easiest way to create GUI applications. Creating a GUI using Tkinter is an easy task.

**MySQL:-**

MySQL, the most popular Open Source SQL database management system, is developed, distributed, and supported by Oracle Corporation.

The SQL part of “MySQL” stands for “Structured Query Language”. SQL is the most common standardized language used to access databases. Depending on your programming environment, you might enter SQL directly (for example, to generate reports), embed SQL statements into code written in another language, or use a language-specific API that hides the SQL syntax.

* A database is a structured collection of data. It may be anything from a simple shopping list to a picture gallery or the vast amounts of information in a corporate network. To add, access, and process data stored in a computer database, you need a database management system such as MySQL Server. Since computers are very good at handling large amounts of data, database management systems play a central role in computing, as standalone utilities, or as parts of other applications.

**Pandas:-**

**pandas** is a [software library](https://en.wikipedia.org/wiki/Software_library) written for the [Python programming language](https://en.wikipedia.org/wiki/Python_(programming_language)) for data manipulation and [analysis](https://en.wikipedia.org/wiki/Data_analysis). In particular, it offers [data structures](https://en.wikipedia.org/wiki/Data_structure) and operations for manipulating numerical tables and [time series](https://en.wikipedia.org/wiki/Time_series). It is [free software](https://en.wikipedia.org/wiki/Free_software) released under the [three-clause BSD license](https://en.wikipedia.org/wiki/3-clause_BSD_license).[[2]](https://en.wikipedia.org/wiki/Pandas_(software)#cite_note-2) The name is derived from the term "[**pan**el **da**ta](https://en.wikipedia.org/wiki/Panel_data)", an [econometrics](https://en.wikipedia.org/wiki/Econometrics) term for [data sets](https://en.wikipedia.org/wiki/Data_set) that include observations over multiple time periods for the same individuals.[[3]](https://en.wikipedia.org/wiki/Pandas_(software)#cite_note-3) Its name is a play on the phrase "Python data analysis" itself.[[4]](https://en.wikipedia.org/wiki/Pandas_(software)#cite_note-4) [Wes McKinney](https://en.wikipedia.org/wiki/Wes_McKinney) started building what would become pandas at [AOR](https://en.wikipedia.org/wiki/AQR_Capital) while he was a researcher there from 2007 to 2010.

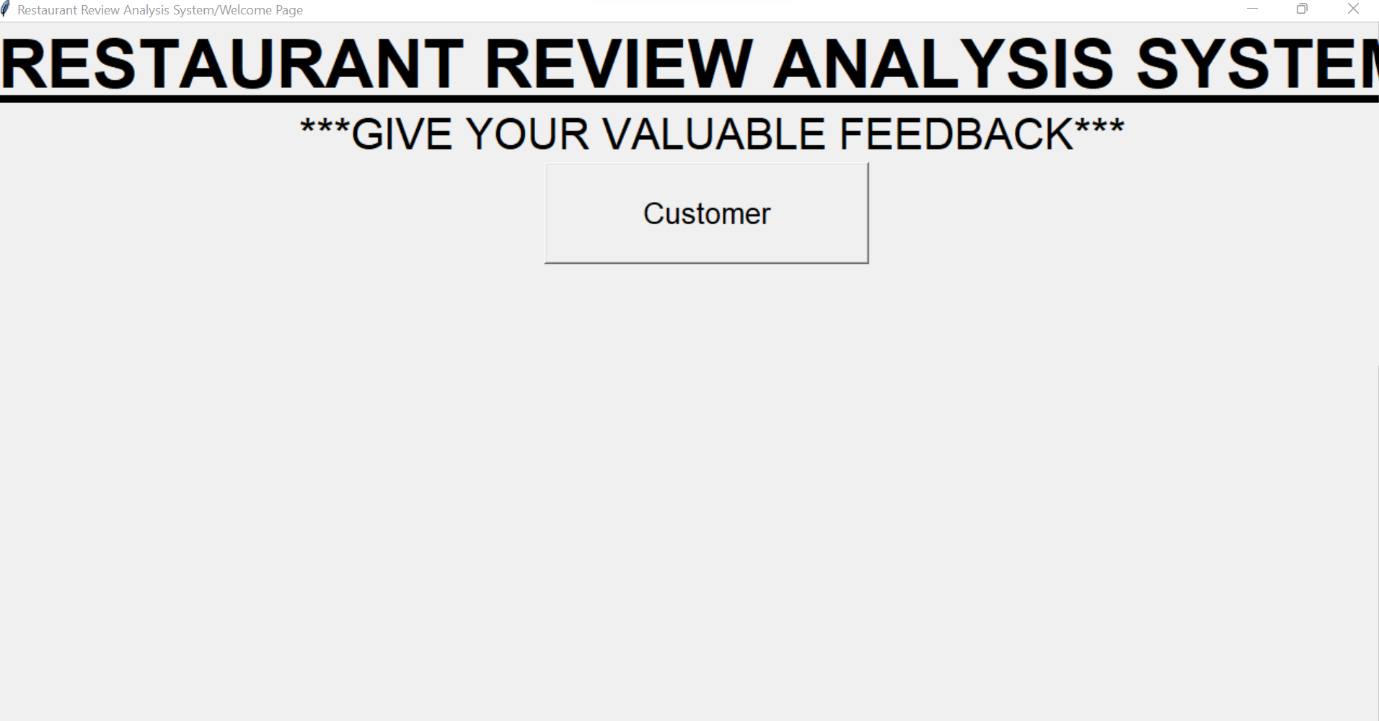
Library features

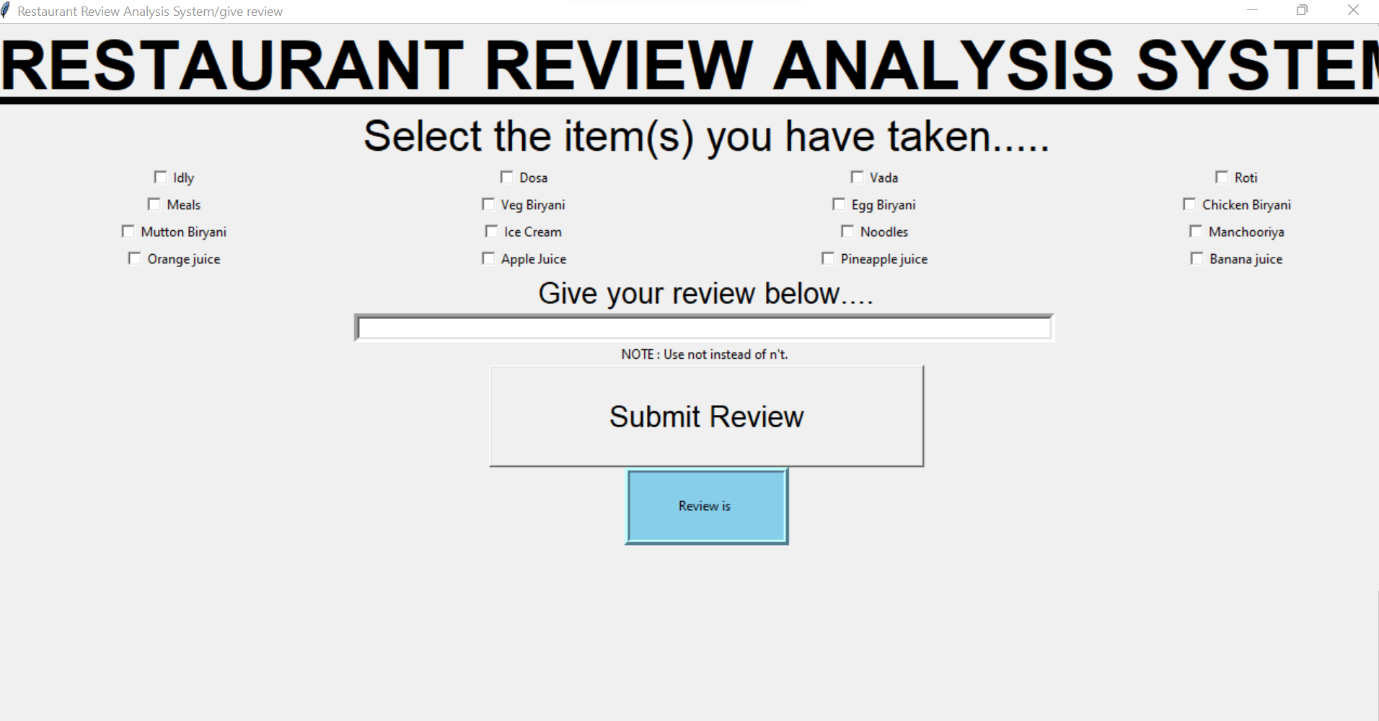
* DataFrame [object](https://en.wikipedia.org/wiki/Object-oriented_programming) for data manipulation with integrated indexing.
* Tools for reading and writing data between in-memory [data structures](https://en.wikipedia.org/wiki/Data_structure) and different [file formats](https://en.wikipedia.org/wiki/File_format).
* Data alignment and integrated handling of missing data.
* Reshaping and pivoting of data sets.
* Label-based slicing, fancy indexing, and subsetting of large data sets.
* Data structure column insertion and deletion.
* Group by engine allowing split-apply-combine operations on data sets.
* Data set merging and joining.
* Hierarchical axis indexing to work with high-dimensional data in a lower-dimensional data structure.
* Time series-functionality: Date range generation[[6]](https://en.wikipedia.org/wiki/Pandas_(software)#cite_note-6) and frequency conversions, moving window [statistics](https://en.wikipedia.org/wiki/Statistics), moving window [linear regressions](https://en.wikipedia.org/wiki/Linear_regression), date shifting and lagging.
* Provides data filtration.

**Deployment Using GUI:-**

In [machine learning](https://analyticsindiamag.com/step-by-step-building-block-for-machine-learning-models/), while building a predictive model we follow several different steps. We first do exploratory data analysis to understand the data well and do the required preprocessing. After the data gets ready we do modeling and develop a predictive model. This model is then used to compute prediction on the testing data and the results are evaluated using different [error metrics](https://analyticsindiamag.com/practical-guide-to-machine-learning-model-evaluation-and-error-metrics/). But what to do next? Do you know how you can use this model and check real-time predictions? It is said you can validate the model performance when you compute prediction in real-time. As we have already seen how we can do model deployment using flask.

In this article, we will be exploring Tkinter – python GUI programming tool. We will explore how we can deploy a machine learning model and check real-time predictions using Tkinter. For this experiment, we will be using the Pima Indians Diabetes Data set that is available on [Kaggle](https://www.kaggle.com/uciml/pima-indians-diabetes-database). We will first build a [classification model](https://analyticsindiamag.com/how-to-implement-lstm-rnn-network-for-sentiment-analysis/) that will classify whether a patient is diabetic or not. Then we will make a GUI using Tkinter and will check predictions on new data points.

****

****

**Customer Review:-**

Natural language processing in artificial intelligence applications makes it easy to gather product reviews from a website and understand what consumers are actually saying as well as their sentiment in reference to a specific product. Companies with a large volume of reviews can actually understand them and use the data collected to recommend new products or services based on customer preferences. This application helps companies discover relevant information for their business, improve customer satisfaction, suggest more relevant products or services, and better understand the customer’s needs.

**Database Connection:-**

**Mysql theory:-**

MySQL, the most popular Open Source SQL database management system, is developed, distributed, and supported by Oracle Corporation.

***MySQL is a database management system.***

A database is a structured collection of data. It may be anything from a simple shopping list to a picture gallery or the vast amounts of information in a corporate network. To add, access, and process data stored in a computer database, you need a database management system such as MySQL Server. Since computers are very good at handling large amounts of data, database management systems play a central role in computing, as standalone utilities, or as parts of other applications

A relational database stores data in separate tables rather than putting all the data in one big storeroom. The database structures are organized into physical files optimized for speed. The logical model, with objects such as databases, tables, views, rows, and columns, offers a flexible programming environment. You set up rules governing the relationships between different data fields, such as one-to-one, one-to-many, unique, required or optional, and “pointers” between different tables. The database enforces these rules, so that with a well-designed database, your application never sees inconsistent, duplicate, orphan, out-of-date, or missing data.

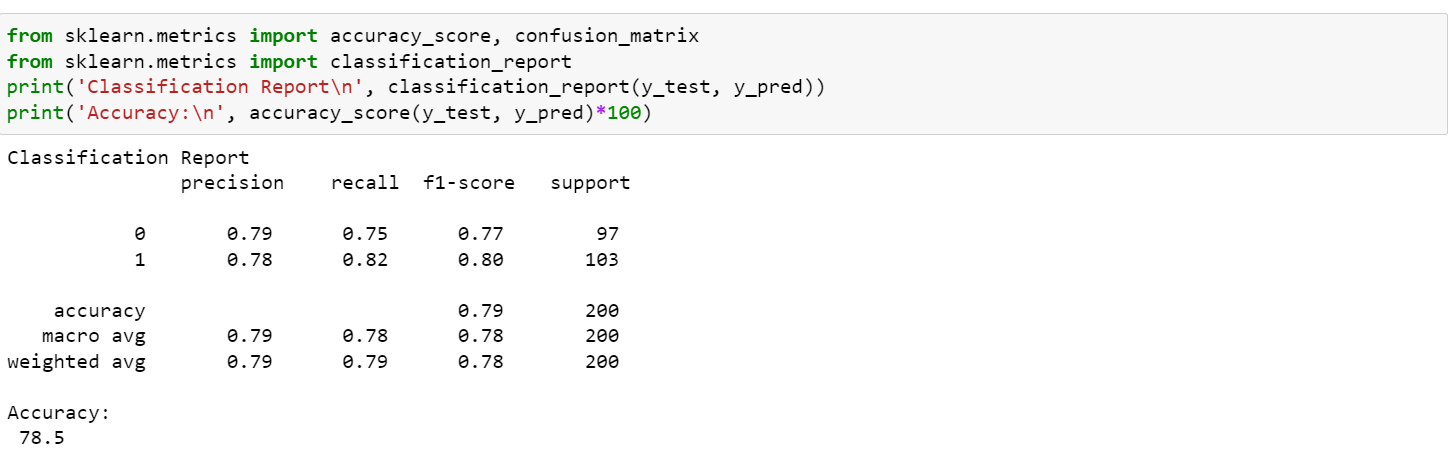
**Result :-**

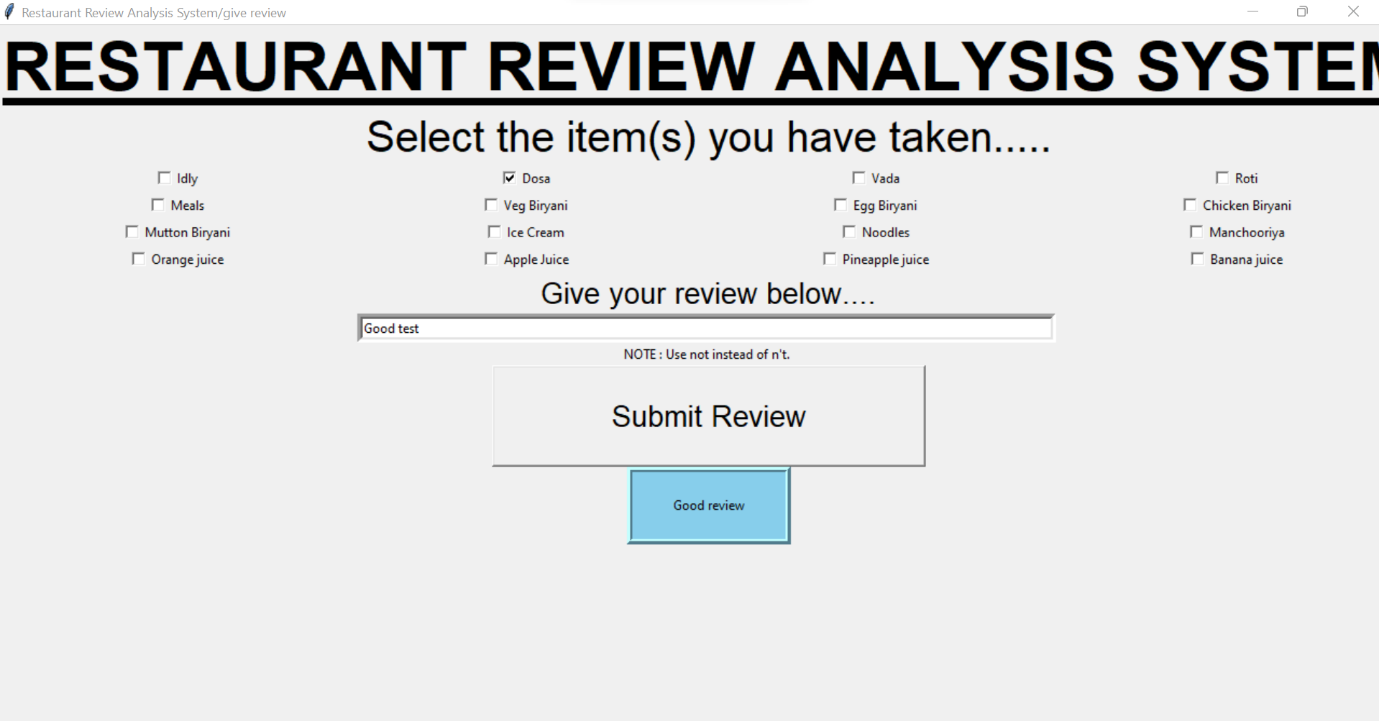
We used python language for implementing classification algorithms. Python provides

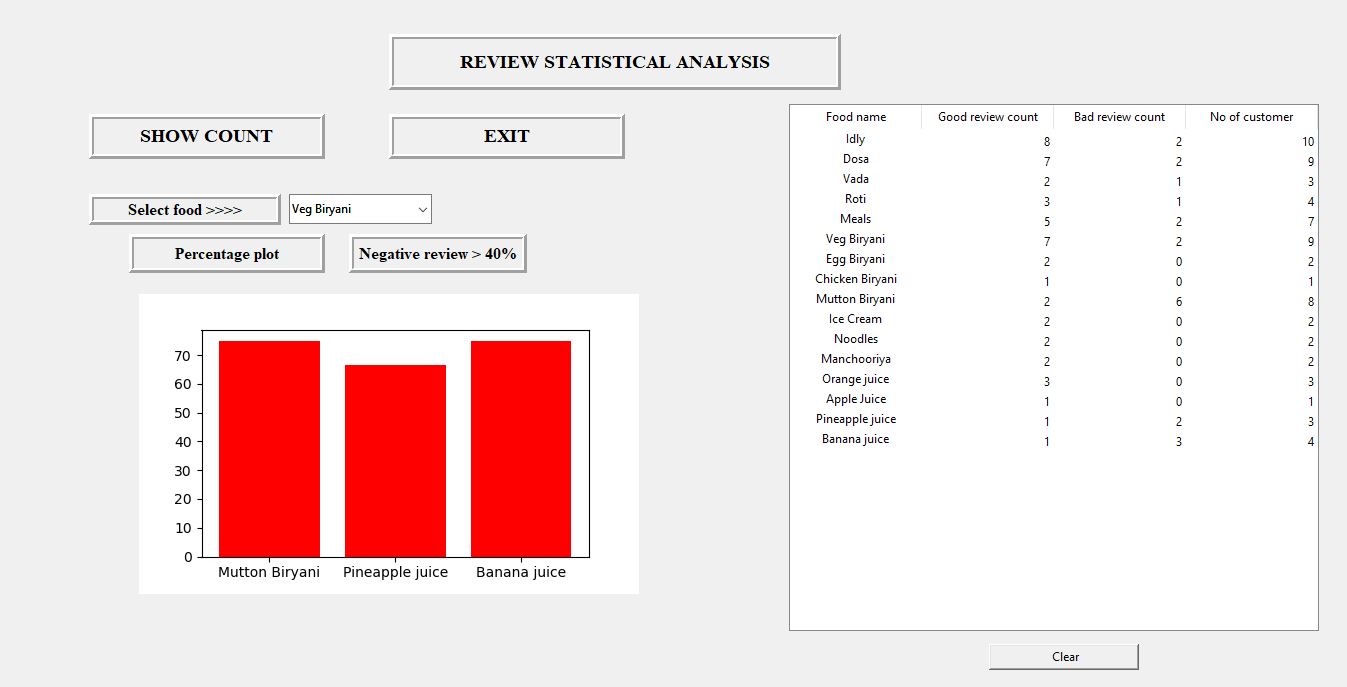
vast number of packages for implementing ML models. python’s scikit-learn library is

used for machine learning tasks. For applying NLP techniques, python provides

NLTK(Natural Language Processing Toolkit).

****

****

****

**Conclusion:-**

How we landed to this work, well we were searching for a problem where we can create a model that will be useful for general public, which will help to take decisions for purchasing news product or services and it will give them a brief explanation of the product/service so that they need not to waste much time to take decision. Finally, we reached to this SemEval2014 task. In this work we have to do aspect-based sentiment analysis, for the general public and to the owner of the restaurant as well. This will help both parties. For the owner of the restaurant, it will help to find their negative and positive aspects. And they can improve their services/product quality by knowing the status of the respective aspect. Here we have four major aspects [service, food, ambiance, price] and one default category [Anecdotes/Miscellaneous] if a review does not have any aspect then it will go to default category

**Future Work:-**

While calculating the sentiment, the sentiment is taken from sentiwordnet3.0, it is for the English language only, we will use different techniques and different libraries like TextBlob to get the sentiment.

This work is done only in the English language, the work can be extended to Indian languages, we have to study more research paper on sentiment analysis on Indian language, and aspect category detection for Indian language, as far as concerned about my knowledge, there are not any sentiwordnet3.0 and Wordnet for Indian language, so we need to implement the work from the scratch. It would be a great

**Reference:-**

**https://github.com/mohamedsgap/restaurant-reviews**

**https://github.com/abhi7585/Restaurant-Review-Sentiment-Analysis**

**https://www.kaggle.com/code/apekshakom/sentiment-analysis-of-restaurant-reviews**

## https://www.kaggle.com/code/apekshakom/sentiment-analysis-of-restaurant-reviews