



**Shivnagar Vidya Prasarak Mandal's
College of Engineering, Malegaon (Bk.)**

Department of Computer Engineering

A

Mini Project Report On

“FOOD REVIEWS PREDICATION”

BY

Ms. Kajal Vilas Bankar Roll No: 402

Ms. Vidya Shivaji Kudale Roll No: 444

Ms. Harshda Rajendra Sable Roll No: 465

UNDER THE GUIDANCE OF

Prof .Y.R. Khalate

B. E. Computer(Sem-VII) 2021-2022

INDEX

1.1	Name Of Project.....	3
1.2	Dataset used in its link.....	3
1.3	Outcomes.....	4
2	Introduction.....	5
3	Software Requirement Specification.....	6
3.1	Purpose.....	6
3.2	Hardware Requirements.....	6
3.3	Software Requirements.....	6
3.4	Dependencies.....	6
3.5	Work Flow.....	7
3.2.1	Data Extraction.....	7
3.2.2	Data Pre-Processing.....	7
3.2.3	Data Visualization.....	8
	Classification Algorithms.....	9
4.1	Simple Neural Network.....	9
4.2	Naive Bayes.....	10
4.3	LogisticRegression.....	11
5	Output Scene.....	12
	Comparison Of Classification.....	17
7	Conclusion.....	18
8	Bibliography.....	18

1.1 Name of project:-

Predication of Food according their reviews in particular Restaurant dataset using Machine learning Alogorithms .

1.2 Dataset used and its link:

- Restaurant Reviews Data Set
- data.gov.in

[Restaurant-reviews | Kaggle](#)

Input Variables:

- a) Review
- b) Liked

- **Number of attributes Variable:- 3** (ID,Review, Liked)

1.3 Outcomes:-

- To understand the various factors which are responsible for identification food reviews predication.
- To measure the accuracy of classification techniques in food reviews dataset.

- To identify the better classifier among Different Algorithm
- Random Forest Classifier, Naïve Bayes and Logistic Regression.

2. Introduction:-

- Restaurant Rating has become the most commonly used parameter for judging a restaurant for any individual positive or negative. The Purpose of this analysis is to a predication model to predict whether a review on the restaurant is positive A lot of research has been done on different restaurants and the quality of food it serves. Rating of a restaurant depends on factors like reviews, area situated, average cost for two people, votes, cuisines and the type of restaurant. The main goal of this is to get insights on restaurants which people like visit and to identify the rating of the restaurant. With this article we study different predictive models like Random forest and Logistic Regression, Naïve Bayes Algorithm and In that we found accuracy Score by using different algorithm and have achieved a maximum score of 82.5% with Logistic Regression. Also use Confusion Matrix, Confusion Matrix is a table that allows to visualize the performance of a classification Model.
- To build a model to predict if review is positive or negative, following steps are performed:
 - Importing Dataset
 - Preprocessing Dataset
 - Vectorization
 - Training and Classification
 - Analysis

3. Software Requirement Specification:-

3.1 Purpose:-

The work is carried out on different stages, starting with the management of the data, and then we move on to visualization, then analysis, classification modeling. Exploration explains the type reviews on food. The study context is to try different classification algorithms on the data and examine the results.

3.2 Hardware Requirement:-

□ laptop/ PC

3.3 Software Requirement:-

- Jupyter Notebook
- Github

3.4 Dependencies:-

You can use pip or conda to install the dependencies:

- Pandas.
- NumPy
- Seaborn.
- Scikit-learn
- Matplotlib

3.5 Work Flow:-

The work involves the following steps:

3.5.1 Data Extraction:-

- Downloading a publicly available dataset.
- Describing the dataset.
- Describing the research

3.5.2 Data Pre-Processing:-




- Cleaning/removing invalid values from rows.
- Cleaning up columns.
- Removing/filling missing data.
- Label Encoding
- Min-Max normalization

3.5.3 Data Visualization:-

- Data Exploratory Analysis.
- Descriptive Analytics.
- Prediction and Model Selection.
- Classification.
- Deriving Conclusion/Insights from the data.

Sample Data:

Detail Compact Column

 Review 	# Liked 
977 unique values	705 total values
Wow... Loved this place.	1
Crust is not good.	0
Not tasty and the texture was just nasty.	0
Stopped by during the late May bank holiday off Rick Steve recommendation and loved it.	1
The selection on the menu was great and so were the prices.	1
Now I am getting angry and I want my damn pho.	0

• **Random Forest Classifier Algorithm:**

The random forest is a classification algorithm consisting of many decision trees. It uses bagging and feature randomness when building each individual tree to try to create an uncorrelated forest of trees whose prediction by committee is more accurate than that of any individual tree. It can perform both regression and classification tasks. A random forest produces good predictions that can handle large datasets efficiently. The random forest algorithm provides a higher level of accuracy in predicting outcomes over the decision tree algorithm.

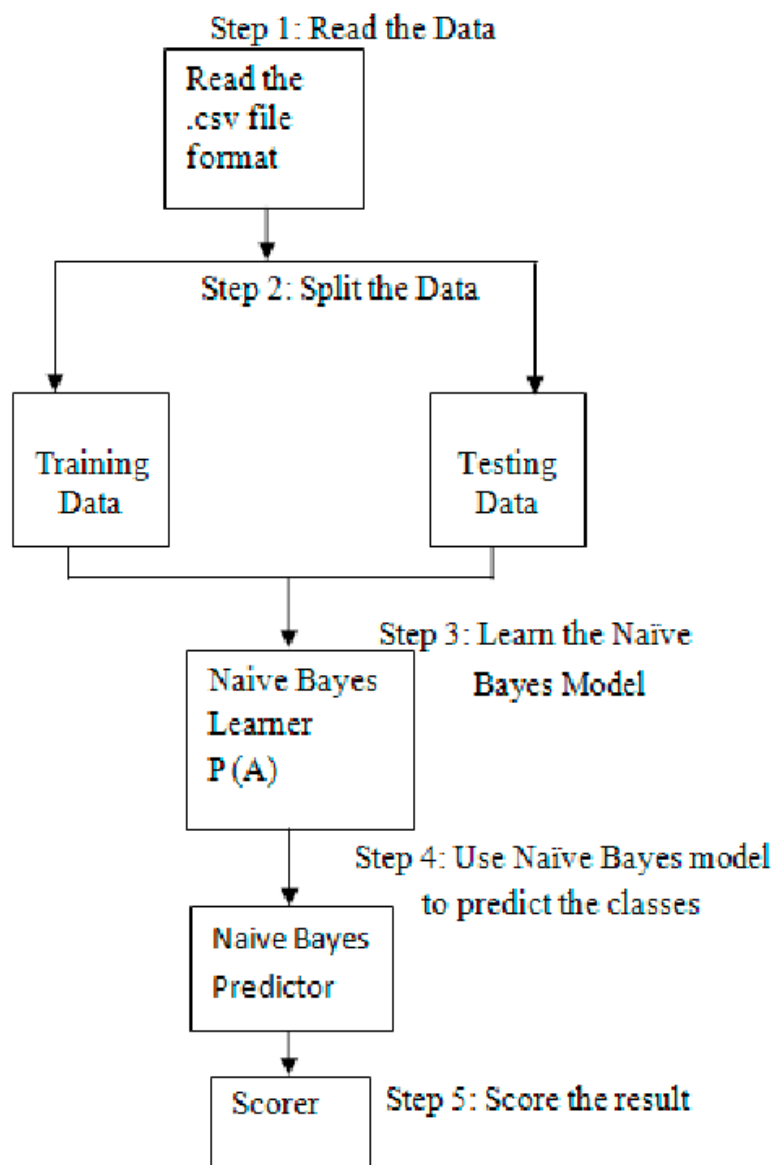
Random forests are an ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time. For classification tasks, the output of the random forest is the class selected by most trees

• **Naive-Bayes**

Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. Naive Bayes is a probabilistic machine learning algorithm that can be used in wide variety of classification tasks. Typical applications include filtering spam, classifying documents, sentiment prediction etc. The name naïve is used because it assumes the features that go into the model are independent of each other.

There is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle: all naive Bayes classifiers assume that the value of a particular feature is independent of the value

of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 10 cm in diameter. A naive Bayes classifier considers each of these features to contribute independently that this fruit is an apple. Regardless any possible correlation between the color, roundness, and diameter features.



regardless of any possible correlations between the color, roundness.

For some types of probability models, naive Bayes classifiers can be trained very efficiently in a supervised learning setting. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood ; in other words, one can work with the naive Bayes model without accepting Bayesian probability or using any Bayesian methods.

Despite their naive design and apparently oversimplified assumptions, naive Bayes classifiers have worked quite well in many complex real-world situations. In 2004, an analysis of the Bayesian classification problem showed that there are sound theoretical reasons for the apparently implausible efficacy of naive Bayes classifiers.^[5] Still, a comprehensive comparison with other classification algorithms in 2006 showed that Bayes classification is outperformed by other approaches, such as boosted trees or random forests.

An advantage of naive Bayes is that it only requires a small number of training data to estimate the parameters necessary for classification.

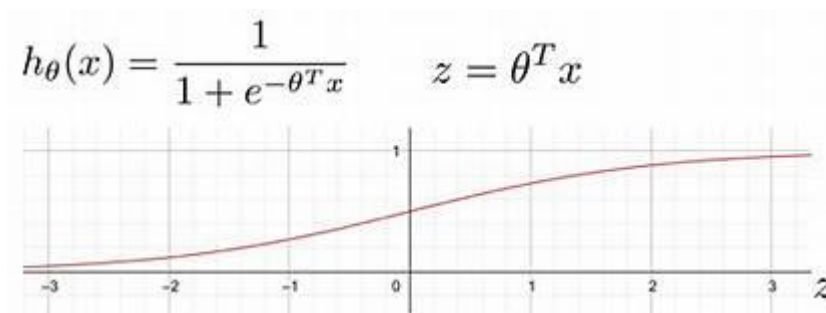
4.3 Logistic Regression :

It is a machine learning algorithm which is used for the classification problems, it is predictive analysis algorithm and based on the concept of probability. Some of the examples of classification problems are Email spam or not spam, online transaction fraud or not fraud. Logistic regression transforms its output using the logistic sigmoid function to return a probability value. The hypothesis of logistic regression tends it to limit the cost function

between 0 and 1. Therefore linear regression fails to represent it as it can have a value greater than 1 or less than 0 which is not possible as per the hypothesis of logistic regression.

1. **Binary logistic** : examples. Food review predication
2. **Multi-linear functions fails class**: eg. Cats, dogs or sheep

Hypothesis Representation –



5. Output Scene:-

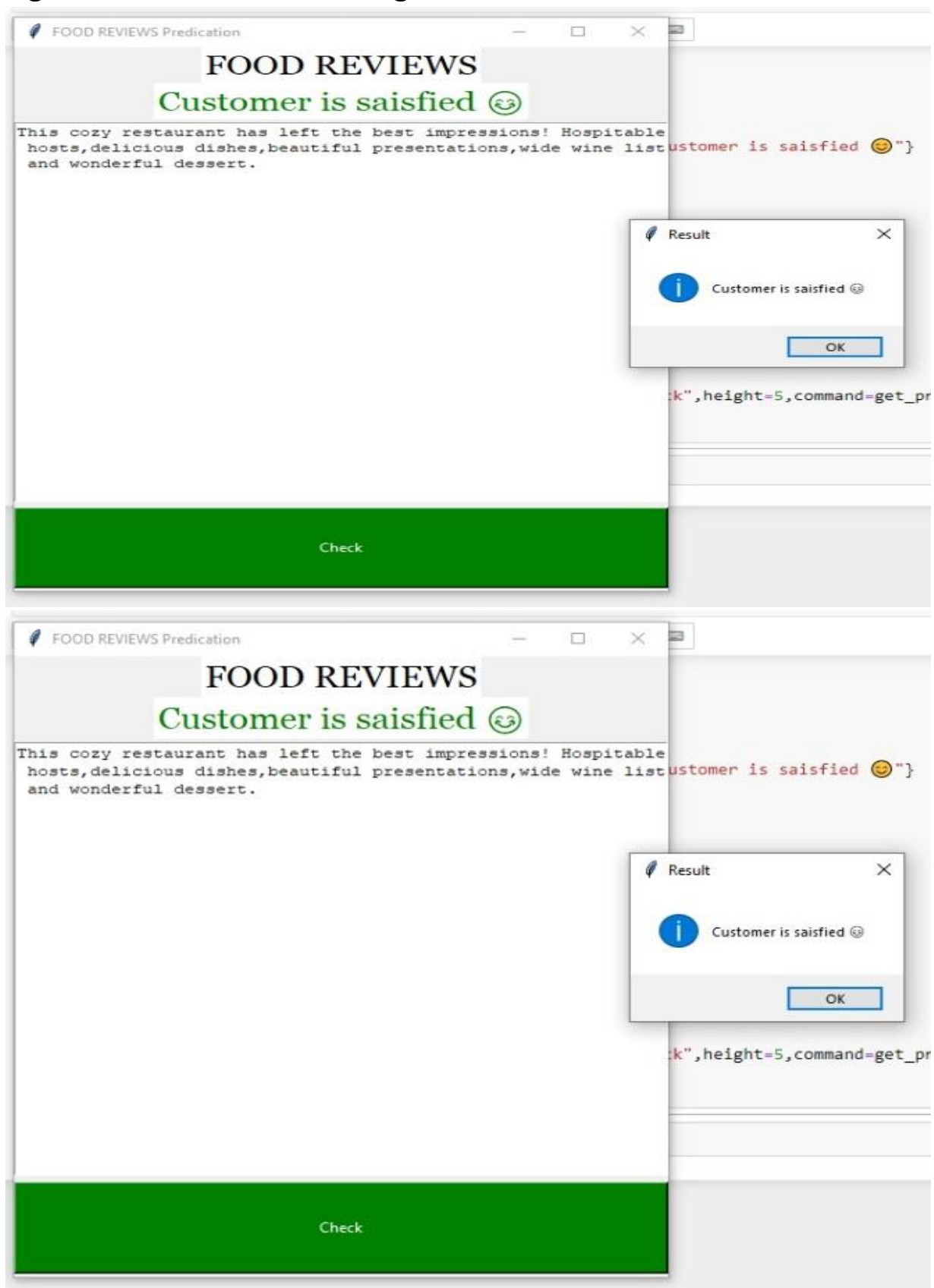
GUI For :-

1) Positive Review: If review is positive way it shows that customer is satisfied.

2) Negative Review: If review is negative way it shows that customer is not satisfied.

- This model able to check any kind of review in following picture we take review from google and check it .

Fig: 1 Positive Review 2 Negative Review



6.Data Analysis graph:

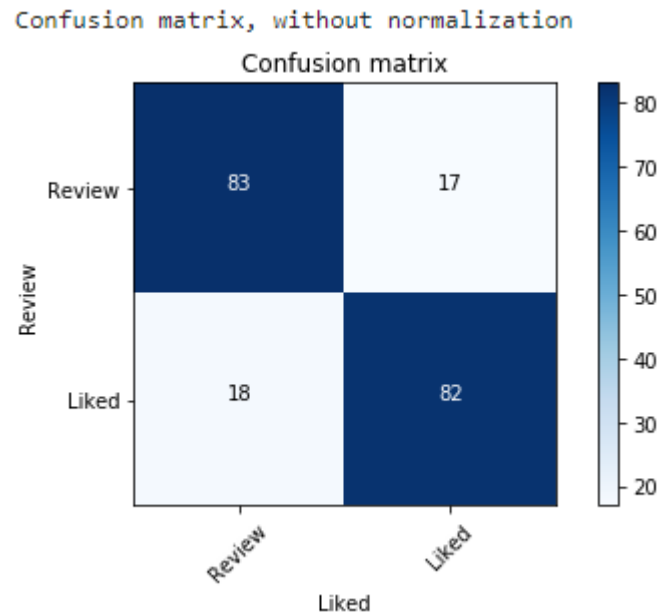


Fig: Logistic Regression Algorithm

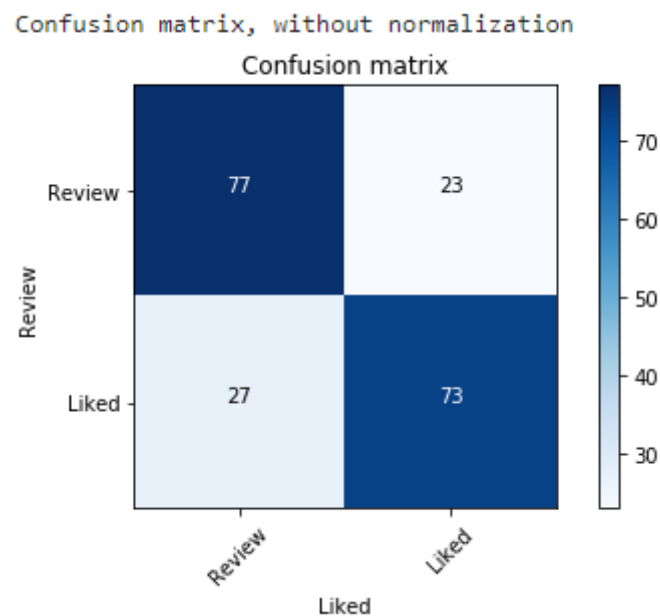


Fig: Random Forest Classification

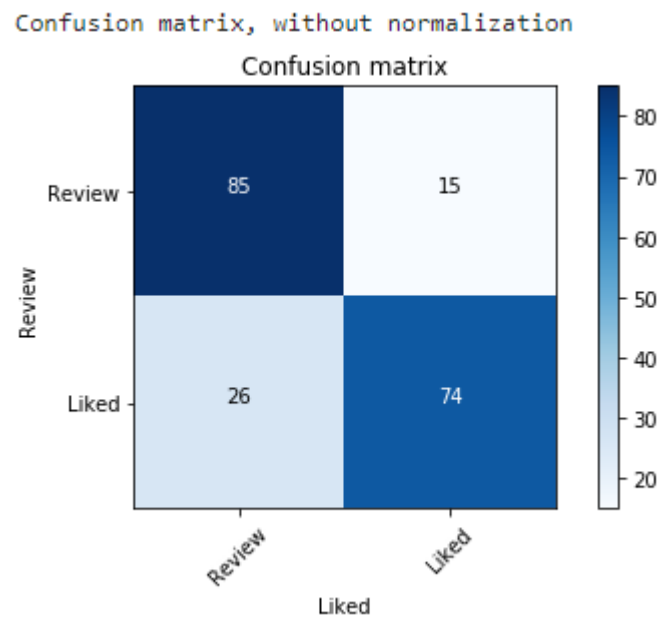


Fig. Naive Bayes Algorithm

6. Comparison Of Classification:-

We worked through a classification predictive modelling in machine learning. We have met our objective which is to evaluate and investigate three selected classification algorithms. The best algorithm based on the data is Logistic regression according to algorithm tuning. Machine Learning classification requires thorough fine tuning of the parameters and at the same time sizeable number of instances for the dataset. It is not a matter of time to build the model for the algorithm only but precision and correct classification.

Therefore, the best learning algorithm for a particular data set does not guarantee the precision and accuracy for another set of data whose attribute are logically different from the other.

Comparison of accuracy Vs test/train ratio of various classification methods

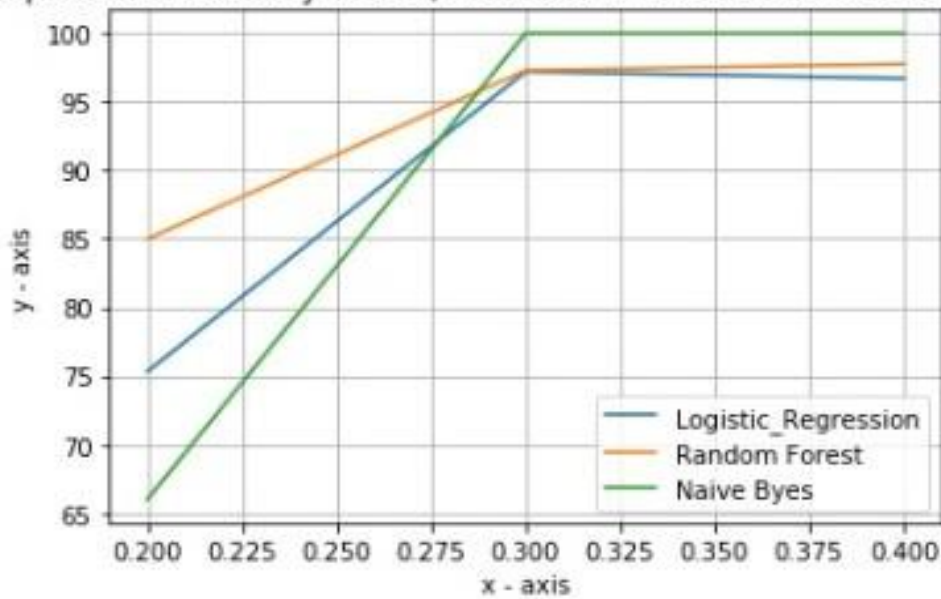


Fig. Comparison Of Logistic Regression , Random Forest, Naïve Bayes Algorithm

- Logistic regression performs well above other techniques for test ratio of 0.2 and above.

7.CONCLUSION

- In this model, we have considered various restaurants records with features like Review and rating .This model will help business owners predict their rating on the parameters considered in our model and improve the customer experience. Different algorithms were used but in the end the final model is selected on Logistic regression which gives the highest accuracy compared to others

8 .Bibliography:-

- 1)<http://www.Wikipedia.com>
- 2)http://www.tutorialspoint.com/machine_learning_with_python