

ZOMATO REVIEW ANALYSIS USING NLP

A Project Report

Submitted in the partial fulfillment of the requirements for the
award of the degree of

Bachelor of Technology

in

Department of Computer Science and Engineering

By

Vallabhaneni Himaja (170330039)

Vaishnavi Gopaluni (170030398)

under the supervision of

Dr. Shanthi Makka

Associate professor



Department of Computer Science and Engineering

K L University Hyderabad,

Aziz Nagar, Moinabad Road, Hyderabad – 500 075, Telangana, India.

Declaration

The Project Report entitled “Zomato Review Analysis” is a record of bonafide work of Vallabhaneni Himaja(170330039) and Vaishnavi Gopaluni(170030398), submitted in partial fulfillment for the award of B.Tech in the Department of Computer Science and Engineering to the K L University, Hyderabad. The results embodied in this report have not been copied from any other Departments/ University/ Institute.

Vallabhaneni Himaja

Vaishnavi Gopaluni

Certificate

This is to certify that the Project Report entitled “Zomato Review Analysis” is being submitted by Vallabhaneni Himaja and Vaishnavi Gopaluni submitted in partial fulfillment for the award of B.Tech in computer science engineering to the K L University, Hyderabad is a record of bonafide work carried out under our guidance and supervision.

The results embodied in this report have not been copied from any other departments/
University/Institute.

Signature of the Supervisor

Dr. Shanthi Makka

Signature of the HOD

Signature of the External Examiner

ACKNOWLEDGEMENT

First and foremost, we thank the lord almighty for all his grace & mercy showered upon us, for completing this project successfully.

I am highly indebted to **Dr. Shanthi Makka** our Guide for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project. Also, We are grateful to our Principal **Dr. L. Koteswara Rao** who has been constantly bearing the torch for all the curricular activities undertaken by us.

We pay our grateful acknowledgement & sincere thanks to our Head of the Department **Dr. E. Anupriya** for her exemplary guidance, monitoring and constant encouragement throughout the course of the project. We pay our sincere thanks to our Project Coordinator **Dr. K. G. Suma** who has been constantly inculcating logistic support and has given constant encouragement during our project. I would like to express my gratitude towards my parents & Dr. Shanthi Makka for their kind co-operation and encouragement which help me in completion of this project.

I would like to express my special gratitude and thanks to industry persons for giving me such attention and time.

My thanks and appreciations also go to my colleague in developing the project and people who have willingly helped me out with their abilities.

We whole heartedly thank all the teaching and non-teaching staff of our department without whom we won't have made this project a reality. We would like to extend our sincere thanks especially to our parent, our family members and friends who have supported us to make this project a grand success.

ABSTRACT

Zomato is unarguably the most popular and trusted company by millions of customers for food reviews as it gives the floor to share customers' review on a particular restaurant.

Think of a situation where machines can understand the reviews and classify them into various categories! We have developed an application using Deep Learning Concepts wherein the machine can comprehend the wordy sentences and comments from customers and place the restaurants into appropriate categories such as good, poor, average, etc.

With an overwhelming demand of restaurants it has therefore become important to study the demography of a location. What kind of a food is more popular in a locality? What people prefer, and what need to be improvised to predict these all we have build a model analysing reviews of the Restaurant given by Customers as Positive or Negative or Neutral on the basis of Collected. Zomato is an application for rating restaurants. The rating has a review of the restaurant which can be used for sentiment analysis.

Table of Contents

| | |
|--|----|
| Chapter 1)INTRODUCTION..... | 7 |
| 1.1) OVERVIEW | 8 |
| 1.2) PURPOSE..... | 8 |
| Chapter 2)LITERATURE SURVEY | 9 |
| 2.1) EXISTING PROBLEM | 11 |
| 2.2) PROPOSED SOLUTION..... | 12 |
| 2.3) SOFTWARE AND HARDWARE REQUIREMENTS..... | 12 |
| 2.4) FUNCTIONAL AND NON-FUNCTIONAL..... | 13 |
| Chapter 3)THEORITICAL ANALYSIS..... | 15 |
| 3.1) BLOCK DIAGRAM..... | 15 |
| 3.2) SOFTWARE DESIGNING | 18 |
| Chapter 4)EXPERIMENTAL INVESTIGATION | 18 |
| 4.1)IMPLEMENTATION..... | 19 |
| Chapter 5)ZOMATO REVIEW ANALYSIS PROCESS FLOWCHART | 21 |
| Chapter 6)RESULT | 22 |
| 6.1)PLAN OF ACTION DOCUMENT..... | 22 |
| 6.2)WHAT IS NLP?..... | 24 |
| 6.3)NLP PIPELINE..... | 24 |
| 6.4)PROCESS..... | 26 |
| Chapter 7)ADVANTAGES AND DISADVANTAGES | 29 |
| Chapter 8)APPLICATIONS | 30 |
| Chapter 9)FUTURE SCOPE | 30 |

| | |
|-------------------------------------|----|
| Chapter 10)BIBLIOGRAPHY | 30 |
| Chapter 11)APPENDIX | 30 |
| Chapter 12)SOURCE CODE | 31 |
| Chapter 13)OUTPUT SCREENSHOTS | 44 |
| Chapter 14)DISSCUSSION..... | 48 |
| Chapter 15)CONCLUSION..... | 49 |
| Chapter 16)REFERENCES..... | 50 |

CH-1) INTRODUCTION

Natural Language Processing is one a part of Artificial Intelligence and Machine Learning to make an know-how of the interactions among computer systems and human (natural) languages. Sentiment evaluation is one a part of Natural Language Processing, that frequently used to analyze phrases primarily based totally at the styles of humans in writing to discover high quality, bad, or impartial sentiments. Sentiment evaluation is beneficial for understanding how customers like some thing or not. Zomato is an utility for score restaurants. The score has a assessment of the restaurant which may be used for sentiment evaluation. Based on this, writers need to talk about the sentiment of the assessment to be predicted. The approach used for preprocessing the assessment is to make all phrases lowercase, tokenization, eliminate numbers and punctuation, prevent phrases, and lemmatization. Then after that, we create phrase to vector with the time period frequency-inverse file frequency (TF-IDF). The statistics that we system are 150,000 critiques. After that make high quality with critiques which have a score of three and above, bad with critiques which have a score of three and below, and impartial who've a score of three. The writer makes use of Split Test, 80ta Training and 20ta Testing. The metrics used to decide random wooded area classifiers are precision, take into account, and accuracy. The accuracy of this studies is 92%. The precision of high quality, bad, and impartial sentiment are 92%, 93%, 96%. The take into account of high quality, bad, and impartial sentiment are 99%, 89%, 73%. Average precision and take into account are 93% and 87%. The 10 phrases that have an effect on the outcomes are: “bad”, “good”, “average”, “best”, “place”, “love”, “order”, “food”, “try”, and “nice”

1.1) OVERVIEW

Zomato is an online website which provides restaurant search and discovery service. It provides its customers a platform to evaluate choices for great places to eat. It has set foot on 22 countries including India. The Gurgaon headquartered company Zomato was named among the top 25 most promising internet companies in India by Smart Techie Magazine. Because of its consistent performance and success, it has been successful in getting regular investments from Info Edge (India). **Zomato** is an application for rating restaurants. The rating has a **review** of the restaurant which can be used for sentiment **analysis**. ... The precision of positive, negative, and neutral sentiment are 92%, 93%, 96%. The recall of positive, negative, and neutral sentiment are 99%, 89%, 73%.

1.2) PURPOSE

The basic idea of analyzing the Zomato dataset is to get a fair idea about the factors affecting the aggregate rating of each restaurant. We also plan to find out the strengths, weaknesses, future opportunities and threats of their business model. This research paper includes analysis of client ratings and reviews in Zomato utilizing content mining. Utilizing content mining, break down the content audits / reviews from the client with a specific end goal to create productive results and legit surveys.

Sentiment **analysis** is useful for knowing how users like something or not. **Zomato** is an application for rating restaurants. The rating has a **review** of the restaurant which can be used for sentiment **analysis**. Based on this, writers want to discuss the sentiment of the **review** to be predicted.

CH-2) LITERATURE SURVEY

[1] **Laksono, R. A.,** Sungkono, K. R., Sarno, R., & Wahyuni, C. S. Zomato is an online website which provides restaurant search and discovery service. It provides its customers a platform to evaluate choices for great places to eat. It has set foot on 22 countries including India. The Gurgaon headquartered company Zomato was named among the top 25 most promising internet companies in India by SmartTechie Magazine. Because of its consistent performance and success, it has been successful in getting regular investments from Info Edge (India).

[2] **Liu, B,** Sentiment analysis and opinion mining is the field of study that analyzes people's opinions, sentiments, evaluations, attitudes, and emotions from written language. It is one of the most active research areas in natural language processing and is also widely studied in data mining, Web mining, and text mining. In fact, this research has spread outside of computer science to the management sciences and social sciences due to its importance to business and society as a whole. The growing importance of sentiment analysis coincides with the growth of social media such as reviews, forum discussions, blogs, micro-blogs, Twitter, and social networks. For the first time in human history, we now have a huge volume of opinionated data recorded in digital form for analysis.

[3] **Porntrakoon, P., & Moemeng, C,** Sentiment analysis is a recent research field in Natural Language Processing (NLP). Text mining and computational techniques determine the sentiment discovered from text. This paper proposes a sentiment analysis using the Text-Representing Centroid (TRC). TRC is a method to determine minimum average distance to all words of the respective document, it also deploys a co-occurrence graph to represent existing relationships among terms in a customer's reviews on particular products and services. A corpus that contains 800 randomly selected hotel reviews from TripAdvisor website is used to evaluate performance by comparison between TRC method and expert's judgment review. The results show 75% accuracy over Thai customer's reviews.

[4] **Putraranti, N. D., & Winarko, E.** Sentiment analysis is useful for knowing how users like something or not. Zomato is an application for rating restaurants. The rating has a review of the restaurant which can be used for sentiment analysis. Based on this, writers want to discuss the sentiment of the review to be predicted. The method used for preprocessing the review is to make all words lowercase, tokenization, remove numbers and punctuation, stop words, and lemmatization. Then after that, we create word to vector with the term frequency-inverse document frequency (TF-IDF). The data that we process are 150,000 reviews. After that make positive with reviews that have a rating of 3 and above, negative with reviews that have a rating of 3 and below, and neutral who have a rating of 3. The author uses Split Test, 80% Data Training and 20% Data Testing. The metrics used to determine random forest classifiers are precision, recall, and accuracy.

[5] **Samuel, Y. T., Hutapea, J. J., & Jonathan,** Some start-ups in the Indian Food & Beverages industry have revolutionized the way we look around for places to dine. The scenario has changed from previous practice when we had to actually go to restaurants to select the restaurant and have the experience. Today, choice for the best places to have food is just a click away. One can choose the best rated place and then decide to enjoy there with friends and family.

[6] **Santoso, V. I., Virginia, G., & Lukito, Y,** Zomato provides **information**, menus and user-reviews of restaurants, and also has food delivery options from partner restaurants in select cities.

[7] **Zhou, Y., Guo, J., Fu, L., & Liang, T.** The Project focuses on Customers and Company, we have analyzed the data and made some useful conclusion in the form of Visualizations. The data is vizualized as it becomes easy to analyse data at instant. The Analysis also solve some of the business cases that can directly help the customers finding the Best restaurant in their locality.

2.1) EXISTING PROBLEM

Analyzing huge data and interpreting it requires many tools and technologies. So our aim is to analyse huge data collected from reviews and opinions of clients and interpret it using nlp. ZOMATO is the brainchild of Mr. Deepinder Goyal and Pankaj Chadda who have been colleagues in

‘Bain And Company’

. The concept to begin ZOMATO got here from the call for for menu playing cards to order meals amongst their colleagues. Mr. Deepinder Goyal and Pankaj Chadda to start with began out a

database for meals menu named “Foodiebay” in 2008 which quickly had g

one stay with menus of 1200 eating places in Delhi NCR (INDIA) with the aid of using July 2008. Later the call became modified into ZOMATO (to keep away from any legal implication as the call Foodiebay has Ebay annexed with it) in November 2010 which is now the biggest eating place detecting platform in India indexed with greater than 4200 eating place throughout 12 towns within the country. And the world over it has lists of approximately 1.4 million eating place throughout 10,000 towns in and at gift it operates in 23 countries, including India, Australia and the United States. ZOMATO were given its first investment of \$1 million from Info

Edge and it’s additionally

ZOMATO

’s biggest shareholder.

extensively used with the aid of using internet customers who want facts for eating and shipping options. ZOMATO is capturing clients with the aid of using offering all of the facts a patron can want earlier than going to a eating place or ordering meals on-line. ZOMATO is following associate advertising and marketing version with internet site and cellular apps. Customers who’re ordering meals on-line will pay thru debit playing cards, credit score playing cards and net banking. ZOMATO has additionally partnered with Paytm, PayPal, and Freecharge to enable on-line payments. For advertising and marketing purpose, ZOMATO

’s key strategy

is it’s pleasant internet site with up

-to-date facts of the eating places. It additionally has a worldwide cellular utility from wherein it gets greater than 50% of its general traffic. ZOMATO has partnered with Delhivery and has invested in Pickingo and Grab to quick up its shipping process

2.2) PROPOSED SOLUTION

With an overwhelming demand of restaurants it has therefore become important to study the demography of a location. What kind of a food is more popular in a locality? What people prefer, and what need to be improvised to predict these all we have build a model analysing reviews of the Restaurant given by Customers as Positive or Negative or Neutral on the basis of Collected Dataset of Zomato System using nlp. All through this procedure look at the client audit premise on their content setting and it demonstrates that how they feel about their visit to that place.

2.3) HARDWARE & SOFTWARE REQUIREMENTS

Software Requirements:-

- 1.Front end: HTML,CSS
- 2.Backend :Flask Framework
- 3.Jupyter Notebook
- 4.Spyder
- 5.Python 3.6 programming language
- 6.Anaconda Navigator

Hardware Requirements:-

1. Processors: Intel Atom® processor or Intel® Core™ i3 processor
2. RAM: 6 GB

3. Operating Systems: 64-bit operating system, x-64 based processor

4. Windows 10 Home

5. Python* versions: 2.7.X, 3.6.X

2.4) FUNCTIONAL AND NON-FUNCTIONAL REQUIREMENTS

Functional

Requirements:- Tools

used:-

- python 3.6
- Numpy
- Pandas
- Matplotlib
- Seaborn
- Data science
- Machine learning

Non-Functional Requirements:-

Safety Requirements:- For the safety requirements nothing but an operation of weekly backups for the data base should take place.

Security and Privacy Requirements:- There are no specific security requirements, anyone can access and use the portal but only authorized persons who are allowed to use and access the database, web pages and the product engine.

Reliability:- The solution should provide reliability to the user that the product will run with all the features mentioned in this document are available and executing perfectly. It should be tested and debugged completely. All exceptions should be well handled.

Accuracy:- The solution should be able to reach the desired level of accuracy. But also keeping in mind that this prototype version is for proving the concept of the project.

CH-3) THEORITICAL ANALYSIS

3.1)BLOCK DIAGRAM

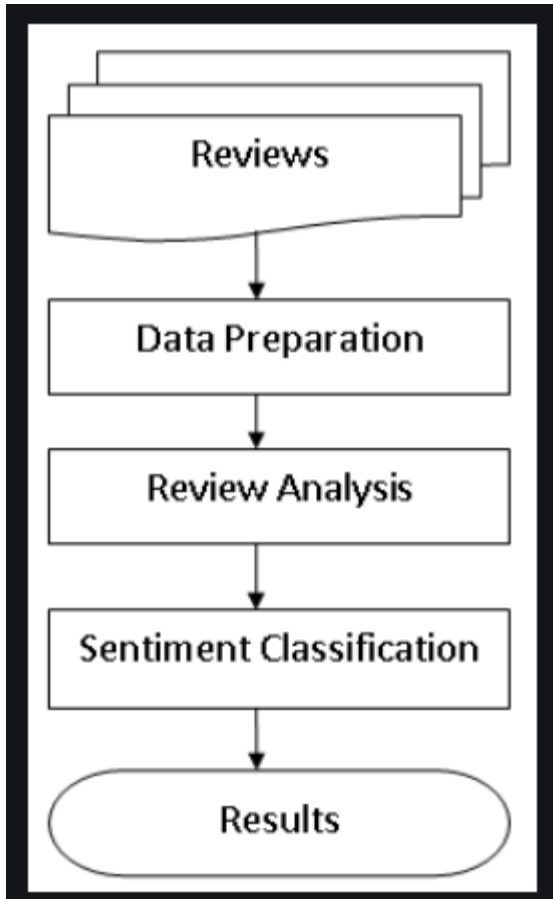


Fig 1: block diagram

Based on this, writers want to discuss the sentiment of the review to be predicted. The method used for preprocessing the review is to make all words lowercase, tokenization, remove numbers and punctuation, stop words, and lemmatization. Then after that, we create word to vector with the term frequency-inverse document frequency. Sharing on the internet is something we usually do. Giving a review is also a useful activity so that other people on the internet can find out something else and see opinions about things. The usual things reviewed by someone in the form of experiences, places, objects, and others. Give a review we usually use text to explain something that we experience with an item, place, or event that we normally

experience. Review on Zomato is still in the form of text and can be classified with positive, negative, or neutral with their ratings. Zomato doesn't have an analysis of how users interact with the reviews and what words will indicate they like or not it. We need to extract the words in review and analysis it so we can know how users interact in Zomato and get customers satisfaction by their review. In this paper, we purpose a method to analyze user's sentiment of Zomato Restaurants and focusing review in Bangalore for study case. We are using Random Forest Classifier to classify the sentiments of users based on their review. We also find words that affects the classifier model. Acquiring public and consumer opinions has long been a huge business itself for marketing, public relations, and political campaign companies. With the explosive growth of social media on the Web, individuals and organizations are increasingly using the content in these media for decision making using its text

Sentiment analysis, is that field of study people's opinions, sentiments, evaluations, judgments, attitudes, and emotions towards entities such as products, services, tourism, movies, organizations, political issues, individuals, problems, events, topics, etc (Liu, 2012). Sometimes, organization or a business needed public or consumer opinions, it conducted surveys, opinion polls, and focus groups. Sharing on the internet is something we usually do. Giving a review is also a useful activity so that other people on the internet can find out something else and see opinions about things. The usual things reviewed by someone in the form of experiences, places, objects, and others. Give a review we usually use text to explain something that we experience with an item, place, or event that we normally experience. Natural Language Processing is one part of Artificial Intelligence and Machine Learning to make an understanding of the interactions between computers and human (natural) languages. Sentiment analysis is one part of Natural Language Processing, that often used to analyze words based on the patterns of people in writing to find positive, negative, or neutral sentiments. Sentiment analysis is useful for knowing how users like something or not. Zomato is an application for rating restaurants. The rating has a review of the restaurant which can be used for sentiment analysis. Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The generalization error for forests converges as. to a limit as the number of trees in the forest becomes large. The generalization error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correlation between them. Using a random selection of features to split each node yields error rates that compare favorably to Adaboost, but are more robust with respect to noise. Internal estimates

features used in the splitting. The precisions of positive, negative, and neutral sentiment are 92%, 93%, 96%. The recalls of positive, negative, and neutral sentiment are 99%, 89%, 73%. Average precision and recall are 93% and 87%.Confusion Matrix.

3.2)SOFTWARE DESIGNING

Software requirements:

Compatible OS for python, javascript, HTML.

Software needed:

tensor, keras, flask.

CH-4) EXPERIMENTAL INVESTIGATION

- 1. Data collection:** we collected the reviews data set from online source (kaggle.com). It is of size 547mb.
- 2. Text pre-processing:** After acquiring the dataset we have imported it, removed punctuations and numbers, converted to lower case, stemming and splitting into training and test set have been implemented.
- 3. Model Building:** Importing the model adding input, hidden and output layers are done and optimising the model.
- 4. Application Building:** we made web application using html and javascript.
- 5. Experiment and testing methods:** Our application model tested 97.7% accurate.

4.1)IMPLEMENTATION

A Web application is developed to predict the rating or category for a given input review using the following concepts:

→ NLP

→ ANN

→ FLASK

NLP is a subset of AI where computers are provided with the ability to understand our language and perform a specific task.

NLP is widely used in sentiment analysis to determine the customer's say on a company or a service or to reflect a person's mood.

There are a lot of models to implement the above but here we use Bag of Words model where the word occurrence is used as a input feature for the model.

As machine learning practitioners interested in working with text data, we are concerned with the tools and methods from the field of Natural Language Processing.

We have seen the path from linguistics to NLP in the previous section. Now, let's take a look at how modern researchers and practitioners define what NLP is all about.

In perhaps one of the more widely textbooks written by top researchers in the field, they refer to the subject as "*linguistic science*," permitting discussion of both classical linguistics and modern statistical methods.

Natural Language Processing, or NLP for short, is broadly defined as the automatic manipulation of natural language, like speech and text, by software.

The study of natural language processing has been around for more than 50 years and grew out of the field of linguistics with the rise of computers.

In this post, you will discover what natural language processing is and why it is so important.

Artificial neural networks are built like the human brain, with neuron nodes interconnected like a web. The human brain has hundreds of billions of cells called neurons. Each neuron is made up of a cell body that is responsible.

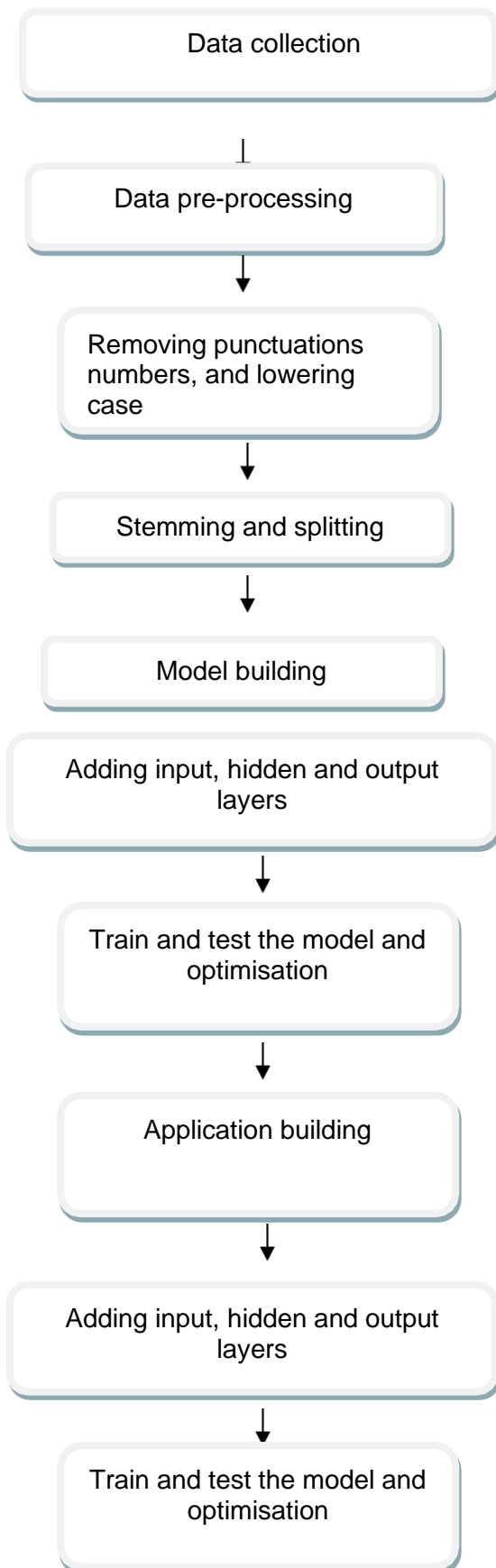
An ANN has hundreds or thousands of artificial neurons called processing units, which are interconnected by nodes. These processing units are made up of input and output units. The input units receive various forms and structures of information based on an internal weighting system, and the neural network attempts to learn about the information presented to produce one output report.

Just like humans need rules and guidelines to come up with a result or output, ANNs also use a set of learning rules called backpropagation, an abbreviation for backward propagation of error, to perfect their output results. An ANN initially goes through a training phase where it learns to recognize patterns in data, whether visually, aurally, or textually. During this supervised phase, the network compares its actual output produced with what it was meant to produce—the desired output. The difference between both outcomes is adjusted using backpropagation.

This means that the network works backward, going from the output unit to the input units to adjust the weight of its connections between the units until the difference between the actual and desired outcome produces the lowest possible error. Artificial neural networks are paving the way for life-changing applications to be developed for use in all sectors of the economy.

Artificial intelligence platforms that are built on ANNs are disrupting the traditional ways of doing things. From translating web pages into other languages to having a virtual assistant order groceries online to conversing with chatbots to solve problems, AI platforms are simplifying transactions and making services accessible to all at negligible costs.

CH-5) ZOMATO REVIE ANALYSIS PROCESS FLOWCHART



CH-6) RESULT

6.1)PLAN OF ACTION DOCUMENT

Given a review our model displays the performance appraisal on scale of (1 to 5 stars).

1 - Poor, 2 – fair, 3 – Average, 4 – good, 5 – excellent.

Our present model displays the performance appraisal based on the review or costumer opinion given. Our future enhancement would be prompting the costumers what went well or asking suggestions and how likely they would recommend this to others.

From split data 80% of training data, we made a machine model to predict. The machine model predicts the other 20 % of data as data testing to see how our model work. We use Random Forest Classifier with tree estimators to 100 trees for make the machine model good enough to predict. 1. Accuracy The results of this machine model are having 92,43667% accuracy. It is explained from 20% data, or 30.000 data to predict the machine correct around 27.740 data. 2. Precision-Recall Precision is the level of accuracy between the information requested by the user and the answers provided by the system. Recall is the level of success of the system in rediscovering information.

Table 1. Precision- Recall

| | Precision | Recall |
|----------|-----------|--------|
| Positive | 92% | 99% |
| Negative | 93% | 89% |
| Neutral | 96% | 73% |
| Average | 93% | 87% |

The precisions of positive, negative, and neutral sentiment are 92%, 93%, 96%. The recalls of positive, negative, and neutral sentiment are 99%, 89%, 73%. Average precision and recall are 93% and 87%.Confusion Matrix Confusion Matrix is a metric that shows the true and prediction errors of data from

From 30.000 data test, we split it the actual data are 18933 positive, 5589 negative, 5478 neutral. From the confusion matrix, we can explain the precision-recall from. 4. Feature Importance Feature importance is part of random forest to extract any features that affect a machine model. From this machine model we got the 10 words that affect the results are: “bad”, “good”, “average”, “best”, “place”, “love”, “order”, “food”, “try”, and “nice”.

6.2) WHAT IS NLP?

NLP is a subset of AI where computers are provided with the ability to understand our language and perform a specific task.

NLP is widely used in sentiment analysis to determine the customer's say on a company or a service or to reflect a person's mood.

There are a lot of models to implement the above but here we use Bag of Words model where the word occurrence is used as a input feature for the model.

Natural language processing (NLP) is a field of artificial intelligence in which computers analyze, understand, and derive meaning from human language in a smart and useful way. By utilizing NLP, developers can organize and structure knowledge to perform tasks such as automatic summarization, translation, named entity recognition, relationship extraction, speech recognition, and topic segmentation. NLP algorithms are typically based on machine learning algorithms. Instead of hand-coding large sets of rules, NLP can rely on machine learning to automatically learn these rules by analyzing a set of examples (i.e. a large corpus, like a book, down to a collection of sentences), and making a statistical inference.

NLP PIPELINE

STEP 1:

Lower case conversion and retain
only the characters a-z

STEP 2:

Tokenization

STEP 3:

Stop words removal

STEP 4:

Stemming / Lemmatization

STEP 5:

Vectorization

6.3) PROCESS

ACQUIRING

DATA SET

NLP- TEXT

PREPROCESSING

MODEL BUILDING

USING ANN

WEB

APPLICATION

The dataset is acquired and the
aforementioned steps in nlp are
performed.

The vectorised text is the input to the ANN model.

The model is trained and tested.

The output falls into one of the below categories:

Excellent

Good

Average

Fair

Poor

Given a review our model displays the performance appraisal on scale of (1 to 5 stars).

1 - Poor, 2 – fair, 3 – Average, 4 – good, 5 – excellent.

Our present model displays the performance appraisal based on the review or costumer opinion given. Our future enhancement would be prompting the costumers what went well or asking suggestions and how likely they would recommend this to others.

Our project tested 97.7% accurate, which is good measure. Given a review our model displays the performance appraisal on scale of (1 to 5 stars).

- Poor, 2 – fair, 3 – Average, 4 – good, 5 – excellent.

This paper takes restaurant review data on Kaggle, specific restaurants in Bangalore, and will be analyzed using the Random Forest in Python Scikit-Learn and analyze it with accuracy and precision-recall. There are several steps of the research method.

1. Data Collection:-

Data that we got from Kaggle is a platform for predictive modeling and analytics competitions in which companies and researchers post data and statisticians and data miners compete to produce the best models for predicting and describing the data. Data that we collect specific is about reviews on Zomato Bangalore and collect 150.000 reviews to be analyzed.

From 150.000 reviews we split them by its ratings. Make positive with reviews that have a rating of 3 and above, negative with reviews that have a rating of 3 and below, and neutral who has a rating of 3. The results of this split, we got 94.743 positive reviews, 28.001 negative reviews, and 27.256 neutral reviews. There are imbalanced datasets but, from research by Yusran, Juliana, and Bern imbalanced data set didn't affect significant accuracy (Samuel, Hutapea, & Jonathan, 2019). So in this research, we are not handling the imbalanced data set.

2.Workflow Process:- The study was conducted and processed in Python 3.6 and with the Scikit-Learn library using the Random Forest method to implement the Sentiment Analysis program.

From data review we are pre-processing text then if the data clean, we split it to 80% training data and 20% test data. Then the 80% data training data we train using Random Forest. After the machine model finish trained, we are testing it to data testing and evaluate the accuracy and precision-recall to see how the metrics of our machine model.

3. Text Preprocessing:- Text preprocessing is the first stage of text mining. The purpose of text preprocessing is to prepare unstructured text documents into structured data that is ready to be used for processes then by eliminating noise, homogenize word forms and reduce word volume (Putraranti & Stages of preprocessing text used in this study are lowercase, tokenization, remove punctuation, stopwords removal, pos tags, lemmatized, and using Tf-Idf Vectorizer to vectorize text to number.

CH-7) ADVANTAGES AND DISADVANTAGES

Advantages:

1. Improved **customer** trust and confidence in your business.
2. Improved customer credibility by allowing contrary reviews.
3. Reduction in complaints and returns as customers can see any negative points before purchasing.
4. The ability to highlight areas for improvement in your order process.

Disadvantages:

1. One negative review of a product or business can skew a potential customer's view of them.
2. You need to keep reviews current and up to date. Otherwise they will seem out of date and irrelevant.
3. A lack of negative reviews could potentially affect a customer's opinion of your business and their decision to make a purchase.

CH-8) APPLICATIONS

Our model helps costumers to decide which is better or reliable to buy based on the review analysis provided. Hence there is more scope for costumers to order an item based on the user reviews and helps them to easily decide. These food delivery apps have made the life of foodies exciting and more straightforward. That's why social media platforms are filled with cheesy platters on the weekends. Moreover, Food delivery has become really popular among youth and millennials.

CH-9) FUTURE SCOPE

Our present model displays the performance appraisal based on the review or costumer opinion given. Our future enhancement would be prompting the costumers what went well or asking suggestions and how likely they would recommend this to others.

CH-10) BIBLIOGRAPHY

CH-11) APPENDIX

Kaggle data set link: <https://www.kaggle.com/himanshupoddar/zomato-bangalore-restaurants>

Further references :

<https://smartbridge.teachable.com/>

CH-12) SOURCE CODE

Main code:

zomato_nlp.py

```
# -*- coding: utf-8 -*-  
"""zomato_nlp.ipynb  
  
Automatically generated by Colaboratory.  
  
Original file is located at  
    https://colab.research.google.com/drive/1chQe3rHW4bhvoNtO4forHMEb\_8tj0q0M  
"""  
  
from google.colab import files  
  
"""# New Section"""  
  
uploaded = files.upload()  
  
import pandas as pd  
import numpy as np  
import matplotlib.pyplot as plt  
  
uploaded = files.upload()
```

```

dataset = pd.read_csv("Restaurant reviews.csv")
dataset = dataset.dropna()
dataset = dataset.drop(dataset[dataset['Rating']=='Like'].index.values)
dataset.reset_index(inplace=True)

data = pd.DataFrame(dataset[['Review','Rating']])

data

import re
import nltk
nltk.download('stopwords')
from nltk.corpus import stopwords
from nltk.stem.porter import PorterStemmer
ps = PorterStemmer()
data_review=[]

import math
for i in range(0,9954):
    p1=data["Review"][i]
    p1=re.sub('[^a-zA-Z]', " ",p1)
    p1= p1.lower()
    p1 = p1.split()
    p1 = [ps.stem(word) for word in p1 if not word in set(stopwords.words('english'))]
    p1= ' '.join(p1)
    data_review.append(p1)
    temp = data['Rating'][i].split(".")
    if(temp[0]!='1'):
        data['Rating'][i]="Poor"
    elif(temp[0]=='2'):
        data['Rating'][i]="Fair"
    elif(temp[0]=='3'):
        data['Rating'][i]="Average"
    elif(temp[0]=='4') :
        data['Rating'][i]="Good"
    elif(temp[0]=='5'):
        data['Rating'][i]="Excellent"

```

```

from sklearn.preprocessing import LabelEncoder, OneHotEncoder
le = LabelEncoder()
data['Rating'] = le.fit_transform(data['Rating'])
print(data)
y = data.iloc[:,1:2].values
print(y)

one_hot_encoder = OneHotEncoder()
z = one_hot_encoder.fit_transform(y[:,0:1]).toarray()
y = np.delete(y,[0],axis=1)
y = np.concatenate((z,y),axis=1)
print(y.shape)

#count vector
from sklearn.feature_extraction.text import CountVectorizer
cv = CountVectorizer(max_features = 11000)
x = cv.fit_transform(data_review).toarray()
import pickle
with open('CountVectorizer','wb') as file:
    pickle.dump(cv,file)
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,random_state = 0,test_size = 0.2)
x_train.shape

from keras.models import Sequential
from keras.layers import Dense

model = Sequential()
model.add(Dense(units = 10772,init = 'random_uniform',activation = 'relu'))
model.add(Dense(units = 11000,init = 'random_uniform',activation = 'relu'))
model.add(Dense(units = 12000,init = 'random_uniform',activation = 'relu'))
model.add(Dense(units = 5,init = 'random_uniform',activation = 'softmax'))
model.compile(optimizer = 'adam',loss = 'categorical_crossentropy',metrics = ['accuracy'])
model.fit(x_train,y_train,epochs = 10 ,batch_size = 64)

```

```

y_pred = model.predict_classes(x_test)
print(y_pred)
print(y_test)

model.save('zomato.h5')

y_p = model.predict_classes(cv.transform(["the food was okay"]))
names = ["average", "Excellent", "Fair", "Good", "Poor"]
print(names[y_p[0]])

temp

```

```

from flask import render_template, Flask, request, url_for
from keras.models import load_model
import pickle
import tensorflow as tf
graph = tf.get_default_graph()
with open(r'CountVectorizer', 'rb') as file:
    cv=pickle.load(file)

cla = load_model('zomato.h5', compile=False)
#cla.compile(optimizer='adam', loss='categorical_crossentropy')
app = Flask(__name__)
@app.route('/')
def index():
    return render_template('index2.html')

```

```

@app.route('/tpredict',methods = ["POST"])
def page2():

    if request.method == "POST":

        topic = request.form['review']
        topic=cv.transform([topic])
        nameslist=["Average","Excellent" , "Fair","Good","Poor"]
        with graph.as_default():
            y_pred = cla.predict_classes(topic)

    return render_template('index2.html',y = str(nameslist[y_pred[0]]))

if __name__ == '__main__':
    app.run(host = 'localhost', debug = True , threaded = False)

```

```

import json
import math
import re
import sys
import imp
import sendgrid
import mysql.connector
reload(sys)
sys.setdefaultencoding('utf-8')
filenameAFINN = '/Users/ramanathansubramanian/Documents/AFINN-111.txt'
afinn = dict(map(lambda (w, s): (w, int(s)), [ws.strip().split('\t') for ws in
open(filenameAFINN)]))
pattern_split = re.compile(r"\W+")
Connection = mysql.connector.connect(user='root',password='',host='127.0.0.1',database='Zomato')
global cursor
cursor = Connection.cursor()
cursor.execute("USE Zomato")
global InsertIntoSQL
def InsertIntoSQL(pyline):
    sql = "INSERT INTO ZomatoFeedback(PythonLine) VALUES('"+str(pyline)+"')"

```

```

class ZomatoReview(object):
    InsertIntoSQL('Declaring Global Function')
    global zomato_review
    global review_list
    global zomato_ratings
    global zomato_ratings_text
    global sendMail
    global sentiment
    InsertIntoSQL('global zomato_review()')
    InsertIntoSQL('global review_list()')
    InsertIntoSQL('global zomato_ratings()')
    InsertIntoSQL('global zomato_ratings_text()')
    InsertIntoSQL('global sendMail()')
    InsertIntoSQL('global sentiment()')
    #status_rating_review = [] * 5
    #status_rating_text_review = [] * 5
    def zomato_review(res_id):
        #api_key = '9da6a6b0c48bf3929eca72380659bfb5'
        InsertIntoSQL('Assign Zomato Developer URL in Function zomato_review()')
        Zomato_url = 'https://developers.zomato.com/api/v2.1/reviews?res_id='+str(res_id)
        InsertIntoSQL('Assign Zomato URL Header with API KEY in Function zomato_review()')
        headers = {'user-key':'9da6a6b0c48bf3929ecaxxxxxxxxxxxxxx','Accept':'application/json'}
        InsertIntoSQL('Assign Zomato URL with Header to Form Request Url in zomato_review()')
        request = urllib2.Request(Zomato_url,None,headers)
        InsertIntoSQL('Assign request urls response')
        response = urllib2.urlopen(request)
        InsertIntoSQL('Convert response into Json Format')
        review_data = json.load(response)
        InsertIntoSQL('Declaring List of Json Parameters')
        rating = [] * 5
        review_text = [] * 5
        rating_color = [] * 5
        rating_text = [] * 5
        name = [] * 5
        profile_url = [] * 5
        InsertIntoSQL('rating, review_text, rating_text, rating_color')
        InsertIntoSQL('Iterating through for loop to fetch each Json Object value and Store it in
the list declared previously')
        for userreview in review_data['user_reviews']:
            for review in userreview:
                #print userreview[review]['rating']
                #print userreview[review]['rating_color']

```



```

        rating.append(userreview[review]['rating'])
        review_text.append(userreview[review]['review_text'])
        rating_color.append(userreview[review]['rating_color'])
        rating_text.append(userreview[review]['rating_text'])
        name.append(userreview[review]['user']['name'])
        profile_url.append(userreview[review]['user']['profile_url'])
    #print rating
    #print review_text
    #print rating_color
    #print rating_text
    #res_id = 16774318,18387851
    global name
    global profile_url
    global rating_text
    return rating, review_text, rating_text, rating_color
InsertIntoSQL('Assigning the Restaurant ID and Calling the function zomato_review()')
review_list = zomato_review(16774318)
InsertIntoSQL('Defining the function zomato_ratings()')
def zomato_ratings(r):
    #r = int(r)
    InsertIntoSQL('Based on the Json Value for the element - rating, review is assigned')
    if (r <= 2.5):
        bR = 'Bad Review'
        return bR
    elif (r > 2.5 and r < 3.5):
        nR = 'Neutral Review'
        return nR
    elif (r >= 3.5 and r <= 5):
        gR = 'Good Review'
        return gR
    else:
        nRA = 'No Review Available'
        return nRA
InsertIntoSQL('Defining the function zomato_ratings()')
def zomato_ratings_text(text_rating):
    text_rating = float(text_rating)
    InsertIntoSQL('Based on the Json Value for the element - rating_text, review is assigned')
    if text_rating < 0.00:
        bR = 'Bad Review'
        return bR
    elif text_rating == 0.00:
        nR = 'Neutral Review'
        return nR
    elif text_rating > 0.00:
        gR = 'Good Review'
        return gR

```

```

        nRA = 'No Review Available'
        return nRA
def sentiment(text):
    words = pattern_split.split(text.lower())
    sentiments = map(lambda word: afinn.get(word, 0), words)
    if sentiments:
        sentiment = float(sum(sentiments))/math.sqrt(len(sentiments))
    else:
        sentiment = 0
    return sentiment
    InsertIntoSQL('Defining the function reviewed rating, we assign a text based on rating and
making it global for comparsion with the text fetched from text_rating')
def reviewed_rating(self):
    #review_list = zomato_review(16774318)
    #for rating, review_text, rating_text, rating_color in review_list.iteritems():
    #for rating, review_text, rating_text, rating_color in zip(review_list):
    #rating, review_text, rating_text, rating_color = review_list
    rating, review_text, rating_text, rating_color = review_list
    status_rating_review = [] * 5
    ratingValue = [] * 5
    #for r, ra_t in zip(rating, rating_text):
    for r in rating:
        #if __name__ == '__main__':
        #print(zomato_ratings(r),r)#,"%6.2f %s" % (sentiment(ra_t), ra_t)
        ratingValue.append(r)
        status_rating_review.append(zomato_ratings(r))
        #print("%6.2f %s" % (sentiment(ra_t), ra_t))
    global status_rating_review
    global ratingValue
    InsertIntoSQL('Defining the function reviewed rating text, we assign a text based on rating
value of sentiment text and making it global for comparsion with the text fetched from rating')
def reviewed_rating_text(self):
    #review_list = zomato_review(16774318)
    status_rating_text_review = [] * 5
    sentimentValue = [] * 5
    rating, review_text, rating_text, rating_color = review_list
    for ra_t in rating_text:
        #print("%6.2f %s" % (sentiment(ra_t), ra_t))
        text_rating = "%6.2f" % sentiment(ra_t)
        #print(text_rating)
        #print(zomato_ratings_text("%6.2f %s" % (sentiment(ra_t))))
        #print(zomato_ratings_text(text_rating))
        sentimentValue.append(text_rating)
        status_rating_text_review.append(zomato_ratings_text(text_rating))
    global status_rating_text_review
    global sentimentValue

```

```

InsertIntoSQL('Passing a sentiment text through the function sentiment')
InsertIntoSQL('Split the text with words and assigning the value for each word from -5 to
+5')
InsertIntoSQL('assigning value to word is done by affin.txt where the word is compared
assigns itself a value.')
def sendMail():
    sg = sendgrid.SendGridAPIClient(apikey='SG.82nZPqHvSRy26dR-
kBYTIg.1qxBWLjobvYQ7u0G9ZXXXXXXXXXXXXXXXXX')
    userName = 'User Name '
    profileurl = 'Profile Url of the User '
    data = {
        "personalizations": [
            {
                "to": [
                    {
                        "email": "mayurachennaiite@gmail.com"
                    }
                ],
                "subject": "NEGATIVE FEEDBACK DETECTED!"
            }
        ],
        "from": {
            "email": "mayur87545@gmail.com"
        },
        "content": [
            {
                "type": "text/plain",
                "value": "Hello, Please find the negative review received from the user !"
            }
        ]
    }
    response = sg.client.mail.send.post(request_body=data)
    print(str(response.status_code)+' Mail Sent Successfully')
    print('Response Body \n'+str(response.body))
    print(response.headers)
def final_review(self):
    InsertIntoSQL('Defining the Function Final_review,')
    InsertIntoSQL('Here, we compare the text assigned to Sentiment Value and Rating Value')
    InsertIntoSQL('When the compared text matches for the -Bad Review mail is sent- or else we
just print the statement as NO Negative Reviews')
    count = 0
    for zR,zRT in zip(status_rating_review,status_rating_text_review):
        try:
            if(zR == 'Bad Review' and zRT == 'Bad Review'):
                print('Review is '+zR)

```

```

InsertIntoSQL('Using the Inbuilt attributes, to store Mail ID of To and From of user.')
    sendMail()
    print('User Name '+str(name[count])+ ' with profile Url '+profile_url[count])
    print('The Rating Text is '+str(rating_text[count]))
    print('The Rating System Review is '+str(ratingValue[count])+ ' & Sentiment Value
is '+str(sentimentValue[count])+ ' of the '+zR)
    print('- - - - - Review Completed - - - - -')
    count += 1
elif(zR == 'Bad Review' and zRT == 'Neutral Review'):
    print('Review is '+zR)
    InsertIntoSQL('Defining the Function sendMail, Assigning the API Key to
Variable')

    InsertIntoSQL('Using the Inbuilt attributes, to store Mail ID of To and From of
user.')

    sendMail()
    print('User Name '+str(name[count])+ ' with profile Url '+profile_url[count])
    print('The Rating Text is '+str(rating_text[count]))
    print('The Rating System Review is '+str(ratingValue[count])+ ' & Sentiment Value
is '+str(sentimentValue[count])+ ' of the '+zR)
    print('- - - - - Review Completed - - - - -')
    count +=1
elif(zR == 'Bad Review' and zRT == 'Good Review'):
    print('Review is '+zR)
    InsertIntoSQL('Defining the Function sendMail, Assigning the API Key to
Variable')

    InsertIntoSQL('Using the Inbuilt attributes, to store Mail ID of To and From of
user.')

    sendMail()
    print('User Name '+str(name[count])+ ' with profile Url '+profile_url[count])
    print('The Rating Text is '+str(rating_text[count]))
    print('The Rating System Review is '+str(ratingValue[count])+ ' & Sentiment Value
is '+str(sentimentValue[count])+ ' of the '+zR)
    print('- - - - - Review Completed - - - - -')
    count +=1
elif(zR == 'Neutral Review' and zRT == 'Bad Review'):
    print('Review is '+zR)
    InsertIntoSQL('Defining the Function sendMail, Assigning the API Key to
Variable')

    InsertIntoSQL('Using the Inbuilt attributes, to store Mail ID of To and From of
user.')

    sendMail()
    print('The Rating System Review is '+str(ratingValue[count])+ ' & Sentiment Value
is '+str(sentimentValue[count]))
    print('The Rating Text is '+str(rating_text[count]))
    print('- - - - - Review Completed - - - - -')
    count +=1

```

```

        print('Review is '+zR)
        InsertIntoSQL('Defining the Function sendMail, Assigning the API Key to
Variable')
        InsertIntoSQL('Using the Inbuilt attributes, to store Mail ID of To and From of
user.')
        sendMail()
        print('The Rating System Review is '+str(ratingValue[count])+' & Sentiment Value
is '+str(sentimentValue[count]))
        print('The Rating Text is '+str(rating_text[count]))
        print('- - - - - Review Completed - - - - -')
        count +=1
    elif(zR == 'Neutral Review' and zRT == 'Good Review'):
        print('Review is '+zR)
        print('The Rating System Review is '+str(ratingValue[count])+' & Sentiment Value
is '+str(sentimentValue[count]))
        print('The Rating Text is '+str(rating_text[count]))
        print('- - - - - Review Completed - - - - -')
        count +=1
    elif(zR == 'Good Review' and zRT == 'Bad Review'):
        print('Review is '+zR)
        print('The Rating System Review is '+str(ratingValue[count])+' & Sentiment Value
is '+str(sentimentValue[count]))
        print('The Rating Text is '+str(rating_text[count]))
        print('- - - - - Review Completed - - - - -')
        count +=1
    elif(zR == 'Good Review' and zRT == 'Neutral Review'):
        print('Review is '+zR)
        print('The Rating System Review is '+str(ratingValue[count])+' & Sentiment Value
is '+str(sentimentValue[count]))
        print('The Rating Text is '+str(rating_text[count]))
        print('- - - - - Review Completed - - - - -')
        count +=1
    elif(zR == 'Good Review' and zRT == 'Good Review'):
        print('Review is '+zR)
        print('The Rating System Review is '+str(ratingValue[count])+' & Sentiment Value
is '+str(sentimentValue[count]))
        print('The Rating Text is '+str(rating_text[count]))
        print('- - - - - Review Completed - - - - -')
        count +=1
    else:
        count +=1
        print('No Negative Feedback')
except ValueError:
    print('Oops! No Reviews Found !')

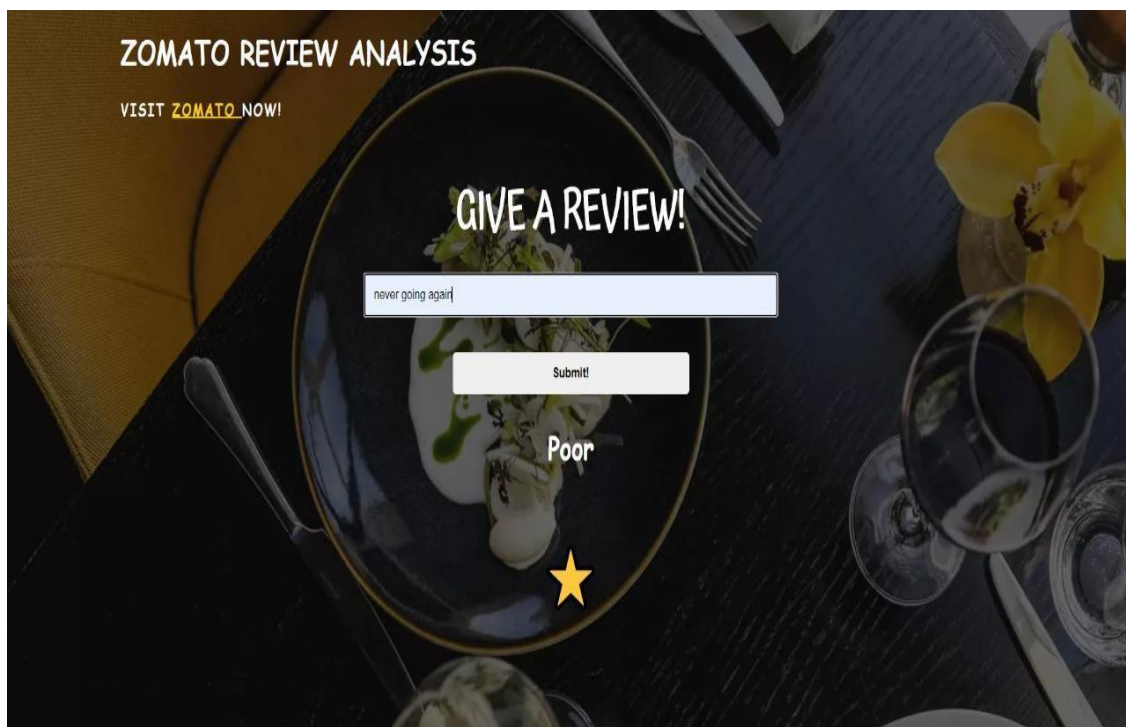
```

```

        print('Review is '+zR)
        InsertIntoSQL('Defining the Function sendMail, Assigning the API Key to
Variable')
        InsertIntoSQL('Using the Inbuilt attributes, to store Mail ID of To and From of
user.')
        sendMail()
        print('The Rating System Review is '+str(ratingValue[count])+' & Sentiment Value
is '+str(sentimentValue[count]))
        print('The Rating Text is '+str(rating_text[count]))
        print('- - - - - Review Completed - - - - -')
        count +=1
    elif(zR == 'Neutral Review' and zRT == 'Good Review'):
        print('Review is '+zR)
        print('The Rating System Review is '+str(ratingValue[count])+' & Sentiment Value
is '+str(sentimentValue[count]))
        print('The Rating Text is '+str(rating_text[count]))
        print('- - - - - Review Completed - - - - -')
        count +=1
    elif(zR == 'Good Review' and zRT == 'Bad Review'):
        print('Review is '+zR)
        print('The Rating System Review is '+str(ratingValue[count])+' & Sentiment Value
is '+str(sentimentValue[count]))
        print('The Rating Text is '+str(rating_text[count]))
        print('- - - - - Review Completed - - - - -')
        count +=1
    elif(zR == 'Good Review' and zRT == 'Neutral Review'):
        print('Review is '+zR)
        print('The Rating System Review is '+str(ratingValue[count])+' & Sentiment Value
is '+str(sentimentValue[count]))
        print('The Rating Text is '+str(rating_text[count]))
        print('- - - - - Review Completed - - - - -')
        count +=1
    elif(zR == 'Good Review' and zRT == 'Good Review'):
        print('Review is '+zR)
        print('The Rating System Review is '+str(ratingValue[count])+' & Sentiment Value
is '+str(sentimentValue[count]))
        print('The Rating Text is '+str(rating_text[count]))
        print('- - - - - Review Completed - - - - -')
        count +=1
    else:
        count +=1
        print('No Negative Feedback')
except ValueError:
    print('Oops! No Reviews Found !')

```

CH-13) OUTPUT SCREENSHOTS



ZOMATO REVIEW ANALYSIS

VISIT [ZOMATO_NOW!](#)

GIVE A REVIEW!

Excellent



ZOMATO REVIEW ANALYSIS

VISIT [ZOMATO NOW!](#)

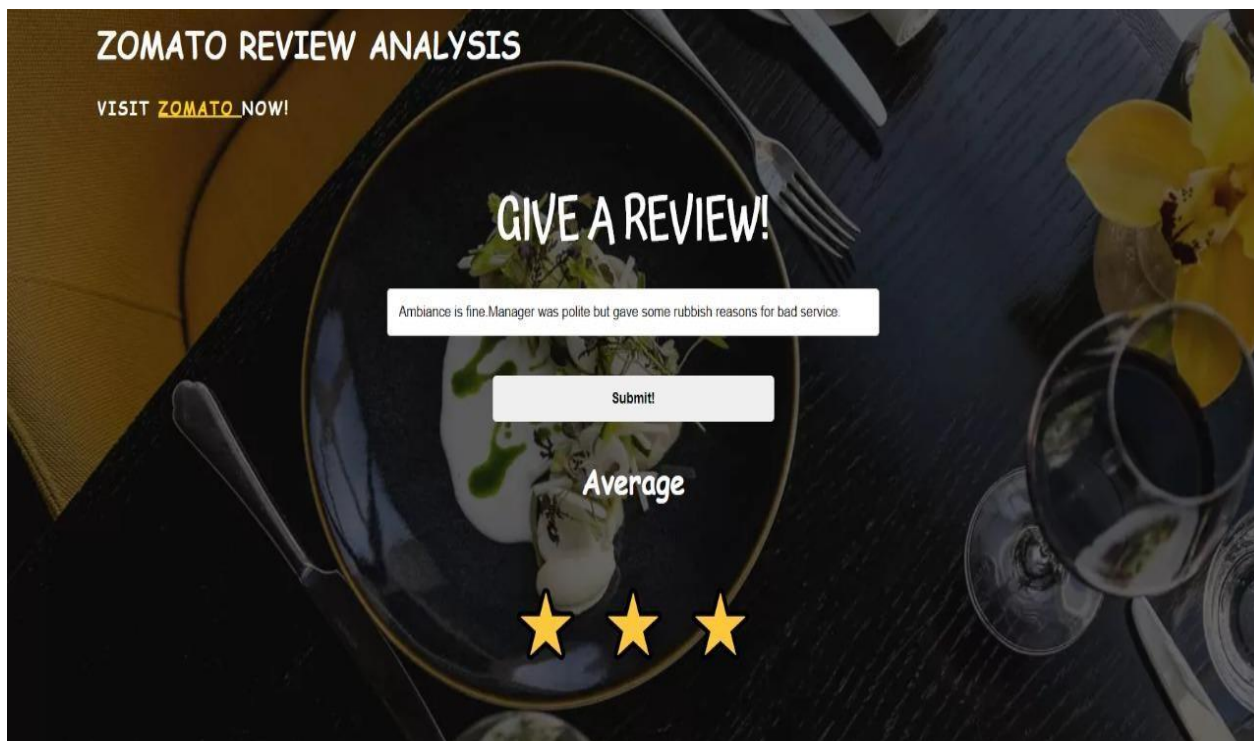
GIVE A REVIEW!

Food quality and taste is not good. Super slow service. Nice ambience but very poor in cus.

Submit!

Fair





CH-14) DISCUSSION

Based on the results of the analysis and testing that has been carried out in this study, there are some conclusions that can be given:

1. The Accuracy of this machine model are 92%
2. The best precision is neutral with 96% accuracy and the good recall is positive with 99% accuracy.
3. The least percentage of recall is neutral. It is indicating the machine model is least to predict neutral than the others.
4. From the feature importance we got the “bad” word is more indicative to predict the sentiment of the people trying to say.

Suggestions given from this research and for the development of further research are as follows:

1. There are imbalanced data in positive, negative, and neutral data. We can use imbalanced dataset algorithms to improve the results.
2. We can use the word data of sentiments to find the sentiments rather than see their ratings.

This paper takes restaurant review data on Kaggle, specific restaurants in Bangalore, and will be analyzed using the Random Forest in Python Scikit-Learn and analyze it with accuracy and precision-recall. There are several steps of the research method. 1. Data Collection Data that we got from Kaggle is a platform for predictive modeling and analytics competitions in which companies and researchers post data and statisticians and data miners compete to produce the best models for predicting and describing the data. Data that we collect specific is about reviews on Zomato Bangalore and collect 150.000 reviews to be analyzed. From 150.000 reviews we split them by its ratings. Make positive with reviews that have a rating of 3 and above, negative with reviews that have a rating of 3 and below, and neutral who has a rating of 3. The results of this split, we got 94.743 positive reviews, 28.001 negative reviews, and 27.256 neutral reviews. There are imbalanced datasets but, from research by Abstract Proceedings International Scholars Conference, Yusran, Juliana, and Bern imbalanced data set didn't affect significant accuracy (Samuel, Hutapea, & Jonathan, 2019). So in this research, we are not handling the imbalanced data set. 2. Workflow Process The study was conducted and processed in Python 3.6 and with the Scikit-Learn library using the Random Forest method to implement the Sentiment Analysis program. The following figure is a block diagram of the stages of research.

CH-15) Conclusion

Based on the results of the analysis and testing that has been carried out in this study, there are some conclusions that can be given: 1. The Accuracy of this machine model are 92% 2. The best precision is neutral with 96% accuracy and the good recall is positive with 99% accuracy. 3. The least percentage of recall is neutral. It is indicating the machine model is least to predict neutral than the others. 4. From the feature importance we got the “bad” word is more indicative to predict the sentiment of the people trying to say. Suggestions given from this research and for the development of further research are as follows: 1. There are imbalanced data in positive, negative, and neutral data. We can use imbalanced dataset algorithms to improve the results. 2. We can use the word data of sentiments to find the sentiments rather than see their ratings.

CH-16)References

Breiman, L. (2001). Random Forest. Farhadloo, M., & Rolland, E. (2016). Fundamentals of sentiment analysis and its applications. In Studies in Computational Intelligence (Vol. 639, hal. 1–24).[1]

Ilieska, K. (2013). Importance of Customer Satisfaction. Diambil dari.[2]

Laksono, R. A., Sungkono, K. R., Sarno, R., & Wahyuni, C. S. (2019). Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naïve Bayes. 12th International Conference on Information & Communication Technology and System (ICTS) 2019, 54–59.[3]

Liu, B. (2012). Sentiment Analysis and Opinion Mining. Morgan & Claypool Publishers. Porntrakoon, P., & Moemeng, C. (2019). Thai sentiment analysis for consumer's review in multiple dimensions using sentiment compensation technique (SenSecomp). ECTICON 2018 - 15th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, 25–2.[4]

Putraranti, N. D., & Winarko, E. (2014). Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan Maximum Entropy dan Support Vector Machine. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 8(1), 91–100. <https://doi.org/10.22146/ijccs.3499>. [5]

Samuel, Y. T., Hutapea, J. J., & Jonathan, B. (2019). Predicting the Timeliness of Student Graduation Using Decision Tree C4 . 5 Algorithm in Universitas Advent Indonesia. *12th International Conference on Information & Communication Technology and System (ICTS) 2019*, 281–285. [6]

Santoso, V. I., Virginia, G., & Lukito, Y. (2017). PENERAPAN SENTIMENT ANALYSIS PADA HASIL EVALUASI DOSEN DENGAN METODE SUPPORT VECTOR MACHINE. *Jurnal Transformatika*, 14(2), 72. [7]

Zhou, Y., Guo, J., Fu, L., & Liang, T. (2019). Research on Aero-Engine Maintenance Level Decision Based on Improved Artificial Fish-Swarm Optimization Random Forest Algorithm. *Proceedings - 2018 International Conference on Sensing, , Prognostics, and Control, SDPC 2018*, 606–610. [8]

