

zomato review system

Artificial intelligence



Smartbridge, Hyderabad

Summer Internship Project Report

*Submitted By,*

Shaik.Nawaz Hussain,

P.Hemalatha,

Sai Veda,

Guru.

**Certificate**

This is to certify that **Shaik.Nawaz Hussain,P.Hemalatha,Sai Veda,Guru,** has completed a four-week **Internship** cum hands-on training program conducted by **Smartbridge** in collaboration with **IBM** at National Institute of Management and Hospitality in **“Artificial Intelligence”** during the time period of 3rd June, 2019 to 22th June, 2019.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Supervisor

ABSTRACT

In today’s digital world, food app like Zomato is widely used because it provides a platform for people to share their opinion about the restaurants and cafes they have visited. 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 result and legit surveys. Gathering surveys dataset and handling it to check the reliability of the rating given and audit composed by client. Ascertaining reliability of the eatery subsequent to dissecting the surveys as indicated by the administration gave and cost estimation. 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.

CONTENTS

1. Title of the project 4
   1. Introduction 4
   2. Objectives of Research 5
   3. Problem Statement 9
   4. Industry Profile 9
2. Review of Literature 10
3. Data Collection 12
4. Methodology 12
   1. Import Libraries 12
   2. Import Data 14
   3. Exploratory Data Analysis 14
   4. Cleaning Of data 17
   5. Splitting of Data 17
   6. Algorithm 18
   7. Evaluation 23
5. Findings and Suggestions 23
6. Conclusion 25
7. References 25
8. **ZOMATO REVIEW SYSTEM**
   1. **INTRODUCTION**

Artificial intelligence (AI) is an area of computer science that emphasizes the creation of intelligent machines that work and react like humans. Some of the activities computers with artificial intelligence are designed for include:

* Speech recognition
* Learning
* Planning
* Problem solving

Artificial intelligence is a branch of computer science that aims to create intelligent machines. It has become an essential part of the technology industry.

Research associated with artificial intelligence is highly technical and specialized. The core problems of artificial intelligence include programming computers for certain traits such as:

* Knowledge
* Reasoning
* Problem solving
* Perception
* Learning
* Planning
* Ability to manipulate and move objects

Knowledge engineering is a core part of AI research. Machines can often act and react like humans only if they have abundant information relating to the world. Artificial intelligence must have access to objects, categories, properties and relations between all of them to implement knowledge engineering. Initiating common sense, reasoning and problem-solving power in machines is a difficult and tedious task.

Machine learning is also a core part of AI. Learning without any kind of supervision requires an ability to identify patterns in streams of inputs, whereas learning with adequate supervision involves classification and numerical regressions. Classification determines the category an object belongs to and regression deals with obtaining a set of numerical input or output examples, thereby discovering functions enabling the generation of suitable outputs from respective inputs. Mathematical analysis of machine learning algorithms and their performance is a well-defined branch of theoretical computer science often referred to as computational learning theory.

Machine perception deals with the capability to use sensory inputs to deduce the different aspects of the world, while computer vision is the power to analyze visual inputs with a few sub-problems such as facial, object and gesture recognition.

Robotics is also a major field related to AI. Robots require intelligence to handle tasks such as object manipulation and navigation, along with sub-problems of localization, motion planning and mapping.

**“ZOMATO- Never have a bad meal”**

ZOMATO is India’s largest and most successful startup company where food industry meets technology and connects thousands of restaurant with one thread.

**Listed with more than 1.5 millions restaurants in 23 countries and still counting..**

**Over 2000 passionate and strong ZOMANS in 23countries representing 32 nationalities and become a more attractive Employer.**

ZOMATO is the brainchild of Mr. Deepinder Goyal and Pankaj Chadda who were colleagues in ‘**Bain And Company’**. The idea to start ZOMATO came from the demand for menu cards to order food among their colleagues. Mr. Deepinder Goyal and Pankaj Chadda initially started a database for food menu named “Foodiebay” in 2008 which soon had gone live with menus of 1200 restaurants in Delhi NCR (INDIA) by July 2008. Later the name was changed into ZOMATO (to avoid any legal implication as the name Foodiebay has Ebay annexed with it) in November 2010 which is now the largest restaurant detecting platform in India listed with more than 4200 restaurant across 12 cities in the country. And internationally it has lists of about 1.4 million restaurant across 10,000 cities in and at present it operates in 23 countries, including India, Australia and the United States. ZOMATO got its first funding of $1 million from Info Edge and it’s also ZOMATO’s largest shareholder. ZOMATO has over 2000 passionate Zomans strong across 23 countries, and this number is increasing day by day. ZOMATO is widely used by internet users who want information for dining and delivery options. ZOMATO is capturing customers by providing all the information a customer can need before going to a restaurant or ordering foods online. ZOMATO is following affiliate marketing model with website and mobile apps. Customers who are ordering foods online can pay through debit cards, credit cards and net banking. ZOMATO has also partnered with Paytm, PayPal, and Freecharge to enable online payments. For marketing purpose, ZOMATO’s key strategy is it’s friendly website with up-to-date information of the restaurants. It also has a global mobile application from where it gets more than 50% of its total traffic. ZOMATO has partnered with Delhivery and has invested in Pickingo and Grab to quick up its delivery process.

* 1. **Objectives of Research**

|  |  |  |  |
| --- | --- | --- | --- |
| **Key Activities**   * Provide local restaurants and hotel search services. * Collect data on food menus, contacts and relevant information to users. * Large database of restaurants across cities. * Presence across 23 countries. | **Value Proposition**   * One-stop shop for dines and offers restaurants a way to differentiate them. * Bridges the gap between customers and restaurants by efficient technology which reduces delivery time. * Rating based pricing model for foods. | **Customer Segments**   * Users who try to find local restaurants of various cuisines. * Restaurants who want their name to reach targeted audiences. * Customer preferring home delivery. * Database companies. * Market research companies. | **Key Resources**  **Key Partners**   * Uber Taxi * Paytm * Mobiwiki * Delhivery * Freecharge |
| **Revenue streams**   * Advertising * Ticket sales * Consulting | | **Channels**   * Mobile Application * Internet | **Customer relationship**  Online service build in mandatory rating system |

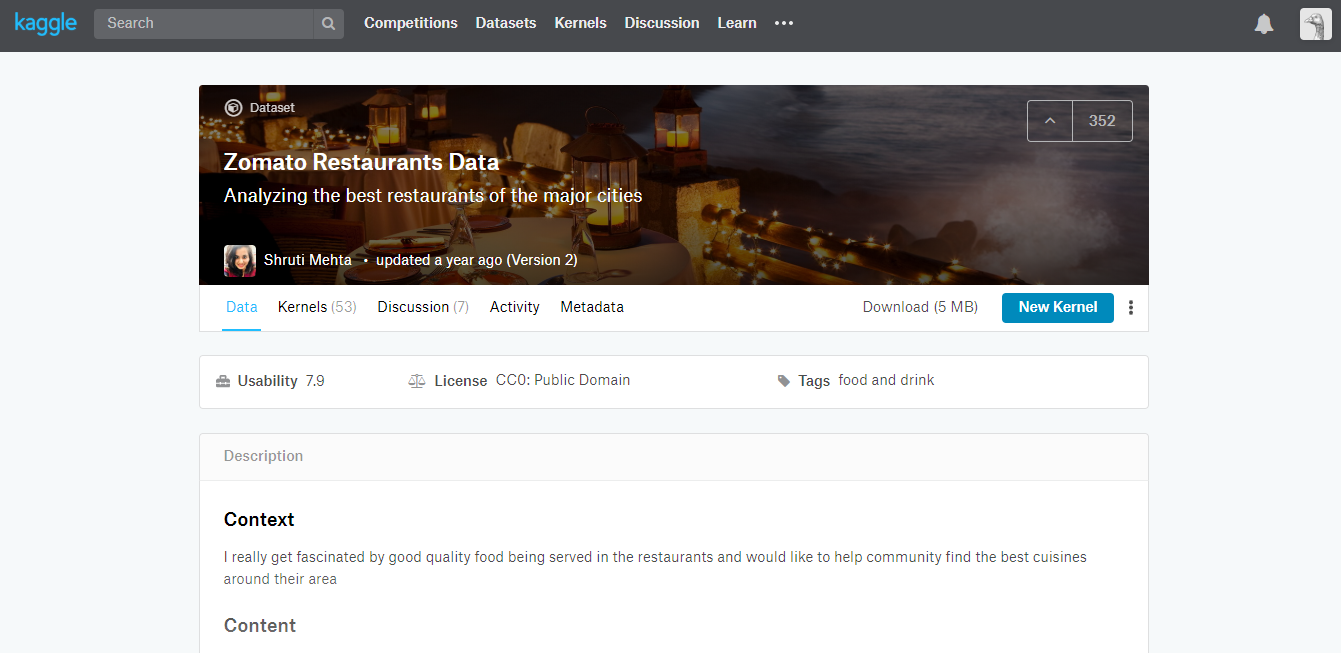
* 1. **PROBLEM STATEMENT:**

**To Predict the Reviews of the Restaurant given by Customers as Positive or Negative or Neutral on the basis of Collected Dataset of Zomato System.**

* 1. **INDUSTRY PROFILE**

ZOMATO provides information of all restaurants about their food menu, detailed analysis of restaurant, its branches and locations, service hours and reviews & comments from other user’s who have already experience of having food on respective restaurant. The business model of ZOMATO is affiliate marketing model which work with website and mobile apps. The website enables user to search for food in different ways. Such as by name of food, general search in a specific location, search for available restaurant open in particular time of a day and even they can find out if the restaurant has a bar, buffet ,live music, happy hours and other services. This is a great platform for people who love to eat and very choosy in terms of choice of food

1. **DATA COLLECTION :**

****

**Data Source:**

<https://www.kaggle.com/shrutimehta/zomato-restaurants-data/data#zomato.csv>

The data was duly collected from the above link and reverified for changed attributes by visiting different zomato APIs and checking for latest changes. The data set was modified accordingly and then used.

1. **Methodology**

**4.1** Importing Libraries

To perform majority of tasks in python it is very much important to import libraries that are used in python same as in C, C++ and java. There are different kinds of libraries that their respective functions in python and some of them are,

* + 1. **NumPy**

NumPy is a general-purpose array-processing package. It provides a high-performance multidimensional array object, and tools for working with these arrays.

It is the fundamental package for scientific computing with Python. It contains various features including these important ones:

* A powerful N-dimensional array object.
* Sophisticated (broadcasting) functions.
* Tools for integrating C/C++ and Fortran code.
* Useful linear algebra, Fourier transform, and random number capabilities.
  + 1. **Pandas**

Pandas is the most popular python library that is used for data analysis. It provides highly optimized performance with back-end source code is purely written in C or Python.

We can analyse data in pandas with:

* Series
* Data Frames
  + 1. **Matplotlib**

Matplotlib is a python library used to create 2D graphs and plots by using python scripts. It has a module named pyplot which makes things easy for plotting by providing feature to control line styles, font properties, formatting axes etc. It supports a very wide variety of graphs and plots namely - histogram, bar charts, power spectra, error charts etc. It is used along with NumPy to provide an environment that is an effective open source alternative for MATLAB. It can also be used with graphics toolkits like PyQt and wxPython.

* + 1. **Seaborn**

Seaborn is a library for making statistical graphics in Python. It is built on top of [matplotlib](https://matplotlib.org/) and closely integrated with [pandas](https://pandas.pydata.org/) data structures.

Here is some of the functionality that seaborn offers:

* A dataset-oriented API for examining [relationships](https://seaborn.pydata.org/examples/scatter_bubbles.html#scatter-bubbles) between [multiple variables](https://seaborn.pydata.org/examples/faceted_lineplot.html#faceted-lineplot).
* Specialized support for using categorical variables to show [observations](https://seaborn.pydata.org/examples/jitter_stripplot.html#jitter-stripplot) or [aggregate statistics](https://seaborn.pydata.org/examples/pointplot_anova.html#pointplot-anova).
* Options for visualizing [univariate](https://seaborn.pydata.org/examples/distplot_options.html#distplot-options) or [bivariate](https://seaborn.pydata.org/examples/joint_kde.html#joint-kde) distributions and for [comparing](https://seaborn.pydata.org/examples/horizontal_boxplot.html#horizontal-boxplot) them between subsets of data.
* Automatic estimation and plotting of [linear regression](https://seaborn.pydata.org/examples/anscombes_quartet.html#anscombes-quartet) models for different kinds [dependent](https://seaborn.pydata.org/examples/logistic_regression.html#logistic-regression) variables.
  + 1. **Sklearn**

Scikit-learn (formerly scikits.learn) is a [free software](https://en.wikipedia.org/wiki/Free_software) [machine learning](https://en.wikipedia.org/wiki/Machine_learning) [library](https://en.wikipedia.org/wiki/Library_(computing)) for the [Python](https://en.wikipedia.org/wiki/Python_(programming_language)) programming language. It features various [classification](https://en.wikipedia.org/wiki/Statistical_classification), [regression](https://en.wikipedia.org/wiki/Regression_analysis) and [clustering](https://en.wikipedia.org/wiki/Cluster_analysis) algorithms including [support vector machines](https://en.wikipedia.org/wiki/Support_vector_machine), [random forests](https://en.wikipedia.org/wiki/Random_forests), [gradient boosting](https://en.wikipedia.org/wiki/Gradient_boosting), [k-means](https://en.wikipedia.org/wiki/K-means_clustering) and [DBSCAN](https://en.wikipedia.org/wiki/DBSCAN), and is designed to interoperate with the Python numerical and scientific libraries [NumPy](https://en.wikipedia.org/wiki/NumPy) and [SciPy](https://en.wikipedia.org/wiki/SciPy).

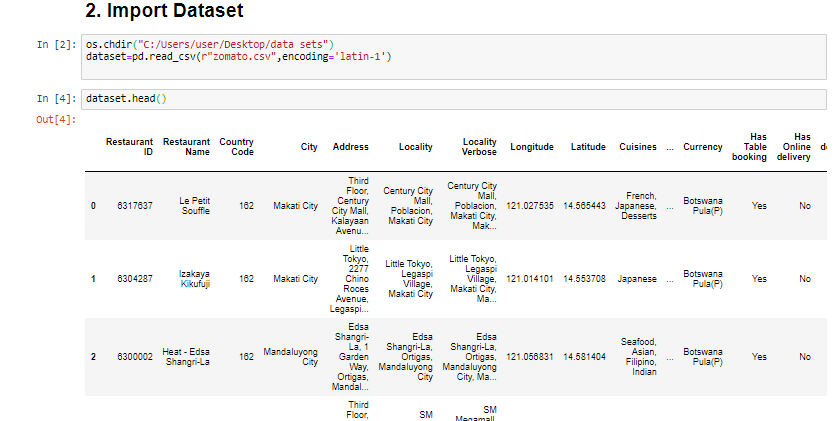
* 1. Import Data

To pull data from a CSV file, you must use the reader function to generate a reader object.

The reader function is designed to take each line of the file and make a list of all columns. Then, you just choose the column you want the variable data for.

It sounds a lot more complicated than it is. To prove it, let’s take a look at an example.

Next, we create the reader object, iterate the rows of the file, and then print them. Finally, we get Data.

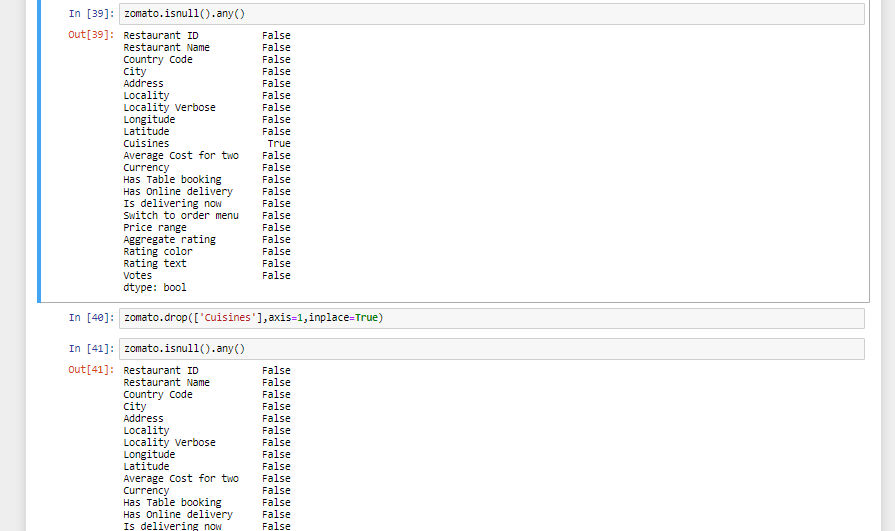


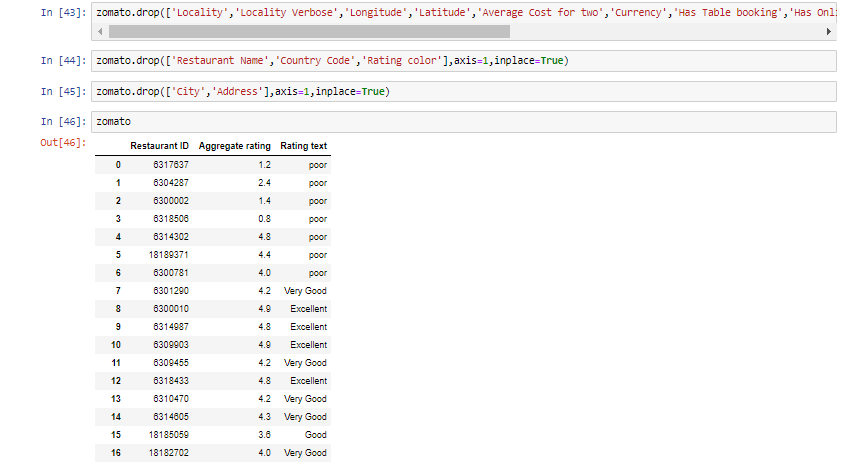
* 1. Exploratory Data Analysis

In [statistics](https://en.wikipedia.org/wiki/Statistics) and in Machine Learning, exploratory data analysis (EDA) is an approach to analysing [data sets](https://en.wikipedia.org/wiki/Data_set) to summarize their main characteristics, often with visual methods. A [statistical model](https://en.wikipedia.org/wiki/Statistical_model) can be used or not, but primarily EDA is for seeing what the data can tell us beyond the formal modelling or hypothesis testing task. Exploratory data analysis was promoted by [John Tukey](https://en.wikipedia.org/wiki/John_Tukey) to encourage statisticians to explore the data, and possibly formulate hypotheses that could lead to new data collection and experiments. EDA is different from [initial data analysis (IDA)](https://en.wikipedia.org/wiki/Data_analysis#Initial_data_analysis), which focuses more narrowly on checking assumptions required for model fitting and hypothesis testing, and handling missing values and making transformations of variables as needed. EDA encompasses IDA.

* 1. Cleaning of Data

The cleaning of data is very much important to increase all the parameters of the data. Cleaning up of data means the removal of null values from the data that makes it more robust to perform different types of operations.





* 1. Splitting of Data

When you’re working on a model and want to train it, you obviously have a dataset. But after training, we have to test the model on some test dataset. For this, you’ll a dataset which is different from the training set you used earlier. But it might not always be possible to have so much data during the development phase.

In such cases, the obviously solution is to split the dataset you have into two sets, one for training and the other for testing; and you do this before you start training your model.

But the question is, how do you split the data? You can’t possibly manually split the dataset into two. And you also have to make sure you split the data in a random manner. To help us with this task, the SciKit library provides a tool, called the Model Selection library. There’s a class in the library which is, aptly, named ‘[train\_test\_split](http://scikit-learn.org/stable/modules/generated/sklearn.model_selection.train_test_split.html).’ Using this we can easily split the dataset into the training and the testing datasets in various proportions.

There are a few parameters that we need to understand before we use the class:

**test\_size** — This parameter decides the size of the data that has to be split as the test dataset. This is given as a fraction. For example, if you pass 0.5 as the value, the dataset will be split 50% as the test dataset. If you’re specifying this parameter, you can ignore the next parameter.

**train\_size** — You have to specify this parameter only if you’re not specifying the test\_size. This is the same as test\_size, but instead you tell the class what percent of the dataset you want to split as the training set.

**random\_state** — Here you pass an integer, which will act as the seed for the random number generator during the split. Or, you can also pass an instance of the RandomState class, which will become the number generator. If you don’t pass anything, the RandomState instance used by np.random will be used instead.

* 1. Algorithm :

Step1:Import required libraries

Step 2: Import Dataset

Step3:Find and Remove NULL Values (If any).

Step4:Drop Unnecessary Parameters (If need).

Step5;Install VaderSentiment and Import SentimentIntensityAnalyzer

Step6: Using VanderSentiment Assign the scores to the text and label them.

Step7; Predict the system on basis of scores.

* 1. Evaluation

As the algorithm is successfully applied and we found the prediction values successfully. Now the question arises that to which extent the predicted values are correct. To Measure that we find the accuracy of that particular algorithm using accuracy score from sklearn and confirm with the accuracy and decide that it may happen or not. Basically, an accuracy of 80% and above are considered to be the best model. In the same way the sensitivity and specificity are also measured to determine some other specific factors of the data and the predicted values. Accuracy is done as follows,

1. **Findings and Suggestions :**

**The Project suggest that most of customers will follow and choose the best restaurants on basis of reviews and ratings .So,Reviews Plays a Crucial role in any model or system.**

1. **Conclusion :**

**The conclusion of the Project “ ZOMATO REVIEW SYSTEM” is Prediction of Customer reviews whether they are positive or negative on the basis of ratings.**

1. **References**

7.1 <https://www.kaggle.com/shrutimehta/zomato-restaurants-data/data#zomato.csv>

7.2 [www.wikipedia.com](http://www.wikipedia.com)

7.3 [www.geeksforgeeks.com](http://www.geeksforgeeks.com)

7.4 [www.tutorialspoint.com](http://www.tutorialspoint.com)

7.5 [www.stackoverflow.com](http://www.stackoverflow.com)

7.6 [www.python.org](http://www.python.org)