ZOMATO CUSTOMER REVIEW PREDICTION

1.1-Introduction

Artificial intelligence (AI) is the simulation of human intelligence processes by machines, especially computer systems. These processes include learning (the acquisition of information and rules for using the information), reasoning (using rules to reach approximate or definite conclusions) and self-correction. Particular applications of AI include expert systems, speech recognition and machine vision. AI can be categorized as either weak or strong. Weak AI, also known as narrow AI, is an AI system that is designed and trained for a particular task.

Virtual personal assistants, such as Apple's Siri, are a form of weak AI. Strong AI, also known as artificial general intelligence, is an AI system with generalized human cognitive abilities. When presented with an unfamiliar task, a strong AI system is able to find a solution without human intervention. Because hardware, software and staffing costs for AI can be expensive, many vendors are including AI components in their standard offerings, as well as access to Artificial Intelligence as a Service (AIaaS) platforms.

AI as a Service allows individuals and companies to experiment with AI for various business purposes and sample multiple platforms before making a commitment. Popular AI cloud offerings include Amazon AI services, IBM Watson Assistant, Microsoft Cognitive Services and Google AI services. While AI tools present a range of new functionality for businesses, the use of artificial intelligence raises ethical questions. This is because deep learning algorithms, which underpin many of the most advanced AI tools, are only as smart as the data they are given in training. Because a human selects what data should be used for training an AI program, the potential for human bias is inherent and must be monitored closely.

Some industry experts believe that the term artificial intelligence is too closely linked to popular culture, causing the general public to have unrealistic fears about artificial intelligence and improbable expectations about how it will change the workplace and life in general.

Researchers and marketers hope the label augmented intelligence, which has a more neutral connotation, will help people understand that AI will simply improve products and services, not replace the humans that use them.

1.1.1 Types of artificial intelligence

Type 1: Reactive machines

An example is Deep Blue, the IBM chess program that beat Garry Kasparov in the 1990s. Deep Blue can identify pieces on the chess board and make predictions, but it has no memory and cannot use past experiences to inform future ones. It analyzes possible moves -- its own and its opponent and chooses the most strategic move. Deep Blue and Google's AlphaGO were designed for narrow purposes and cannot easily be applied to another situation.

Type 2: Limited memory

These AI systems can use past experiences to inform future decisions. Some of the decision-making functions in self-driving cars are designed this way. Observations inform actions happening in the not-so-distant future, such as a car changing lanes. These observations are not stored permanently.

Type 3: Theory of mind

This psychology term refers to the understanding that others have their own beliefs, desires and intentions that impact the decisions they make. This kind of AI does not yet exist.

Type 4: Self-awareness

In this category, AI systems have a sense of self, have consciousness. Machines with self-awareness understand their current state and can use the information to infer what others are feeling. This type of AI does not exist yet.

1.1.2 Examples of AI technology

AI is incorporated into a variety of different types of technology. Here are seven examples.

Automation: What makes a system or process function automatically. For example, Robotic Process Automation (RPA) can be programmed to perform high-volume, repeatable tasks that humans normally performed. RPA is different from IT automation in that it can adapt to changing circumstances.

Machine learning: The science of getting a computer to act without programming. Deep learning is a subset of machine learning that, in very simple terms, can be thought of as the automation of predictive analytics. There are three types of machine learning algorithms:

- Supervised learning: Data sets are labeled so that patterns can be detected and used to label new data sets
- Unsupervised learning: Data sets aren't labeled and are sorted according to similarities or differences
- Reinforcement learning: Data sets aren't labeled but, after performing an action or several actions, the AI system is given feedback.

Machine vision: The science of allowing computers to see. This technology captures and analyzes visual information using a camera, analog-to-digital conversion and digital signal processing. It is often compared to human eyesight, but machine vision isn't bound by biology and can be programmed to see through walls, for example. It is used in a range of applications from signature identification to medical image analysis. Computer vision, which is focused on machine-based image processing, is often conflated with machine vision.

Natural language processing (NLP): The processing of human -- and not computer -- language by a computer program. One of the older and best known examples of NLP is spam detection, which looks at the subject line and the text of an email and decides if it's junk. Current

approaches to NLP are based on machine learning. NLP tasks include text translation, sentiment analysis and speech recognition.

Robotics: A field of engineering focused on the design and manufacturing of robots. Robots are often used to perform tasks that are difficult for humans to perform or perform consistently. They are used in assembly lines for car production or by NASA to move large objects in space. Researchers are also using machine learning to build robots that can interact in social settings.

Self-driving cars: These use a combination of computer vision, image recognition and deep learning to build automated skill at piloting a vehicle while staying in a given lane and avoiding unexpected obstructions, such as pedestrians.

1.1.3 AI applications

Artificial intelligence has made its way into a number of areas. Here are six examples.

AI in healthcare:

The biggest bets are on improving patient outcomes and reducing costs. Companies are applying machine learning to make better and faster diagnoses than humans. One of the best known healthcare technologies is IBM Watson. It understands natural language and is capable of responding to questions asked of it. The system mines patient data and other available data sources to form a hypothesis, which it then presents with a confidence scoring schema. Other AI applications include chatbots, a computer program used online to answer questions and assist customers, to help schedule follow-up appointments or aid patients through the billing process, and virtual health assistants that provide basic medical feedback.

AI in business:

Robotic process automation is being applied to highly repetitive tasks normally performed by humans. Machine learning algorithms are being integrated into analytics and CRM platforms to uncover information on how to better serve customers. Chatbots have been incorporated into websites to provide immediate service to customers. Automation of job positions has also become a talking point among academics and IT analysts.

AI in education:

AI can automate grading, giving educators more time. AI can assess students and adapt to their needs, helping them work at their own pace. AI tutors can provide additional support to students, ensuring they stay on track. AI could change where and how students learn, perhaps even replacing some teachers.

AI in finance:

AI in personal finance applications, such as Mint or Turbo Tax, is disrupting financial institutions. Applications such as these collect personal data and provide financial advice. Other programs, such as IBM Watson, have been applied to the process of buying a home. Today, software performs much of the trading on Wall Street.

AI in law:

The discovery process, sifting through of documents, in law is often overwhelming for humans. Automating this process is a more efficient use of time. Startups are also building question-and-answer computer assistants that can sift programmed-to-answer questions by examining the taxonomy and ontology associated with a database.

AI in manufacturing:

This is an area that has been at the forefront of incorporating robots into the workflow. Industrial robots used to perform single tasks and were separated from human workers, but as the technology advanced that changed.

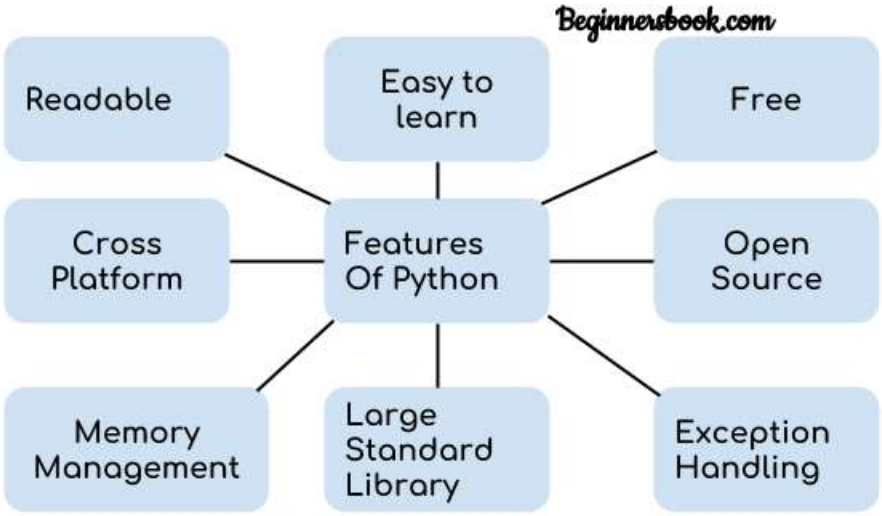
Python is a widely used general-purpose, high level programming language. It was initially designed by Guido van Rossum in 1991 and developed by Python Software Foundation. It was mainly developed for emphasis on code readability, and its syntax allows programmers to express concepts in fewer lines of code.

1.1.4 Introduction to Python Programming language

Python is developed by Guido van Rossum. Guido van Rossum started implementing Python in 1989. Python is a very simple programming language so even if you are new to programming, you can learn python without facing any issues. It is used for:

- web development (server-side)
- software development
- mathematics
- system scripting.

1.1.5 Features of Python programming language



1.**Readable:** Python is a very readable language.

2.**Easy to Learn:** Learning python is easy as this is a expressive and high level programming language, which means it is easy to understand the language and thus easy to learn.

3.**Cross-platform:** Python is available and can run on various operating systems such as Mac, Windows, Linux, Unix etc. This makes it a cross platform and portable language.

4.**Open Source:** Python is a open source programming language.

5.**Large standard library:** Python comes with a large standard library that has some handy codes and functions which we can use while writing code in Python.

6.**Free:** Python is free to download and use. This means you can download it for free and use it in your application. See: Open Source Python License. Python is an example of a FLOSS (Free/Libre Open Source Software), which means you can freely distribute copies of this software, read its source code and modify it.

7.**Supports exception handling:** If you are new, you may wonder what is an exception? An exception is an event that can occur during program exception and can disrupt the normal flow of program. Python supports exception handling which means we can write less error prone code and can test various scenarios that can cause an exception later on.

8.**Advanced features:** Supports generators and list comprehensions. We will cover these features later.

9.**Automatic memory management:** Python supports automatic memory management which means the memory is cleared and freed automatically. You do not have to bother clearing the memory.

1.1.6 What Can You Do with Python

You may be wondering what all are the applications of Python. The applications of Python are listed below:

1. Web development – Web framework like Django and Flask are based on Python. They help you write server side code which helps you manage database, write backend programming logic, mapping urls etc.

- Machine learning – There are many machine learning applications written in Python. Machine learning is a way to write a logic so that a machine can learn and solve a particular problem on its own. For example, products recommendation in websites like Amazon, Flipkart, eBay etc. is a machine learning algorithm that recognises user's interest. Face recognition and Voice recognition in your phone is another example of machine learning.

- Data Analysis – Data analysis and data visualisation in form of charts can also be developed using Python.

- Scripting – Scripting is writing small programs to automate simple tasks such as sending automated response emails etc. Such type of applications can also be written in Python programming language.

- Game development – You can develop games using Python.

- You can develop Embedded applications in Python.

- Desktop applications – You can develop desktop application in Python using library like TKinter or QT.

1.1.7 Natural Language Processing

Natural Language Processing (NLP) refers to AI method of communicating with an intelligent system using a natural language such as English.

Processing of Natural Language is required when you want an intelligent system like robot to perform as per your instructions, when you want to hear decision from a dialogue based clinical expert system, etc.

The field of NLP involves making computers to perform useful tasks with the natural languages humans use. The input and output of an NLP system can be −

- Speech
- Written Text

1.1.8 Components of NLP

There are two components of NLP as given −

Natural Language Understanding (NLU)

Understanding involves the following tasks −

- Mapping the given input in natural language into useful representations.
- Analyzing different aspects of the language.

Natural Language Generation (NLG)

It is the process of producing meaningful phrases and sentences in the form of natural language from some internal representation.

It involves −

- **Text planning** − It includes retrieving the relevant content from knowledge base.
- **Sentence planning** − It includes choosing required words, forming meaningful phrases, setting tone of the sentence.
- **Text Realization** − It is mapping sentence plan into sentence structure.

The NLU is harder than NLG.

1.1.9 Difficulties in NLU

NL has an extremely rich form and structure.

It is very ambiguous. There can be different levels of ambiguity

- **Lexical ambiguity** − It is at very primitive level such as word-level.
- For example, treating the word “board” as noun or verb?
- **Syntax Level ambiguity** − A sentence can be parsed in different ways.
- For example, “He lifted the beetle with red cap.” − Did he use cap to lift the beetle or he lifted a beetle that had red cap?
- **Referential ambiguity** − Referring to something using pronouns. For example, Rima went to Gauri. She said, “I am tired.” − Exactly who is tired?
- One input can mean different meanings.

- Many inputs can mean the same thing.

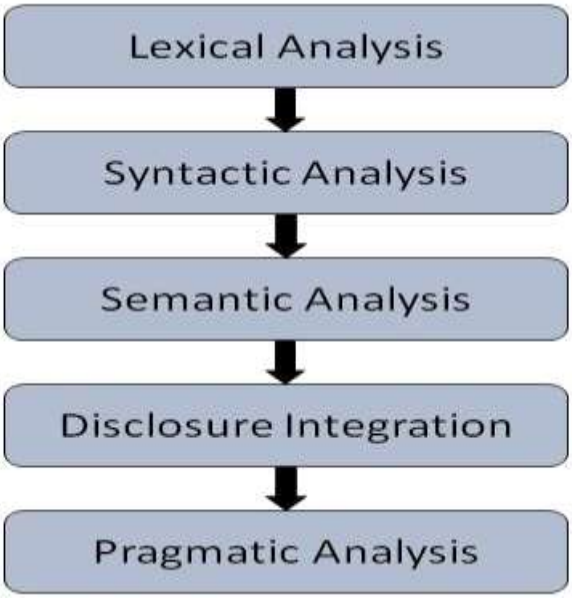
1.1.9 NLP Terminology

- **Phonology** − It is study of organizing sound systematically.
- **Morphology** − It is a study of construction of words from primitive meaningful units.
- **Morpheme** − It is primitive unit of meaning in a language.
- **Syntax** − It refers to arranging words to make a sentence. It also involves determining the structural role of words in the sentence and in phrases.
- **Semantics** − It is concerned with the meaning of words and how to combine words into meaningful phrases and sentences.
- **Pragmatics** − It deals with using and understanding sentences in different situations and how the interpretation of the sentence is affected.
- **Discourse** − It deals with how the immediately preceding sentence can affect the interpretation of the next sentence.
- **World Knowledge** − It includes the general knowledge about the world.

1.1.10 Steps in NLP

There are general five steps −

- **Lexical Analysis** − It involves identifying and analyzing the structure of words. Lexicon of a language means the collection of words and phrases in a language. Lexical analysis is dividing the whole chunk of txt into paragraphs, sentences, and words.
- **Syntactic Analysis (Parsing)** − It involves analysis of words in the sentence for grammar and arranging words in a manner that shows the relationship among the words. The sentence such as “The school goes to boy” is rejected by English syntactic analyzer.



- Semantic Analysis − It draws the exact meaning or the dictionary meaning from the text. The text is checked for meaningfulness. It is done by mapping syntactic structures and objects in the task domain. The semantic analyzer disregards sentence such as “hot ice-cream”.
- Discourse Integration − The meaning of any sentence depends upon the meaning of the sentence just before it. In addition, it also brings about the meaning of immediately succeeding sentence.
- Pragmatic Analysis − During this, what was said is re-interpreted on what it actually meant. It involves deriving those aspects of language which require real world knowledge.

1.1.11 Implementation Aspects of Syntactic Analysis

There are a number of algorithms researchers have developed for syntactic analysis, but we consider only the following simple methods −

- Context-Free Grammar
- Top-Down Parser

Let us see them in detail −

Context-Free Grammar

It is the grammar that consists rules with a single symbol on the left-hand side of the rewrite rules. Let us create grammar to parse a sentence −

“The bird pecks the grains”

Articles (DET) − a | an | the

Nouns − bird | birds | grain | grains

Noun Phrase (NP) − Article + Noun | Article + Adjective + Noun

= DET N | DET ADJ N

Verbs − pecks | pecking | pecked

Verb Phrase (VP) − NP V | V NP

Adjectives (ADJ) − beautiful | small | chirping

The parse tree breaks down the sentence into structured parts so that the computer can easily understand and process it. In order for the parsing algorithm to construct this parse tree, a set of rewrite rules, which describe what tree structures are legal, need to be constructed.

These rules say that a certain symbol may be expanded in the tree by a sequence of other symbols. According to first order logic rule, if there are two strings Noun Phrase (NP) and Verb Phrase (VP), then the string combined by NP followed by VP is a sentence. The rewrite rules for the sentence are as follows −

S → NP VP

NP → DET N | DET ADJ N

VP → V NP

Lexocon −

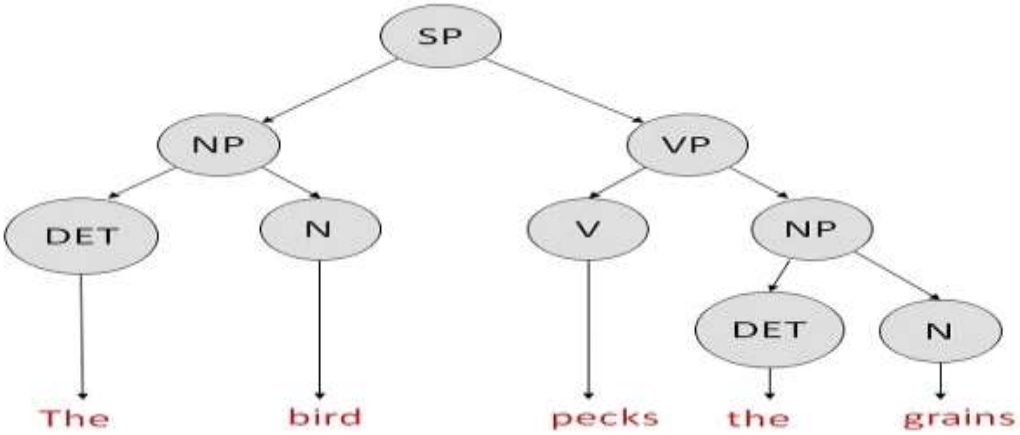
DET → a | the

ADJ → beautiful | perching

N → bird | birds | grain | grains

V → peck | pecks | pecking

The parse tree can be created as shown −



1.1.12 Merit

- The simplest style of grammar, therefore widely used one.
- Learn to use language with greater precision and elegance so that you and others can get what you want
- Think better, think more clearly and make better decisions

1.1.13 Demerit

- They are not highly precise. For example, “The grains peck the bird”, is a syntactically correct according to parser, but even if it makes no sense, parser takes it as a correct sentence.
- To bring out high precision, multiple sets of grammar need to be prepared. It may require a completely different sets of rules for parsing singular and plural variations, passive sentences, etc., which can lead to creation of huge set of rules that are unmanageable.

1.2 Objectives of Research

- The project demonstrates the efficiency of the Zomato.
- The user is free to give any kind of feedback which is ensured to be taken care of by the developer.
- It helps the developer to know how productive the app is in different places.
- It provides developer rough statistic of the reviews stated by the customers.
- The developer analyses the statistics and develops the app accordingly.

1.3 Problem Statement

The problem statment is ***Zomato Customer Review Prediction***. Here, the customer review is predicted using NLP technique.

In today’s world service sector contributes 64.80% in GDP. Zomato is one of the most popular applications that provide services to the user to discover restaurants.

The rise of digital technology is reshaping the industries. With the increased use of technology, the number of people engaging into the digital sector are rapidly increasing. Even Consumers are accustomed to shopping or even ordering online through apps or websites, with maximum convenience and transparency, expecting the same experience that they would get from the outlet itself.

To match up with the consumer's expectations apps are providing increased facilities and services to the customers. Being up to date with the customers' expectations helps firm retain customer's to a greater extent. With the predictions on the customer reviews the developer can make changes and act according to the feedbacks.

2. Review of literature

There are a couple of research papers published based on restaurant reviews and hotel reviews. It attempts to find classify the restaurant research. Many customers visit a restaurant based on food critics and reviews on websites such as Zomato.com. Restaurants strive at the initial stages of opening but their demand deteriorates after the initial hype.

Further, business of these restaurants are largely based on their reviews. What can the restaurant do to make their ratings better? Food taste is an obvious trigger to improve the ratings of a restaurant, but there are other factors that improve the ratings of a restaurant. Such as inclusion of cuisine, option for home delivery, presence of table etc. This paper aims a creating a prediction model for the reviews and analyze the trigger event that would improve the rating.

The model proposed in this paper is, extraction of review from social networking site using text processing, artificial neural network is used to classify the dataset as positive and negative.

In most of these research papers reviews are extracted from one website and classification model is created. This paper is an attempt to create a trigger model to improve restaurants based on Zomato dataset.

Digital services on the other hand are services that are anything that can be delivered through an information infrastructure such as the internet, in various forms i.e. applications, web pages, social media, etc. In the paper the major focus will be on the various apps that are available either by 3rdparty for deliverypurpose or by restaurants themselves for various purposes like delivery, pointing system, in house app ordering, etc.

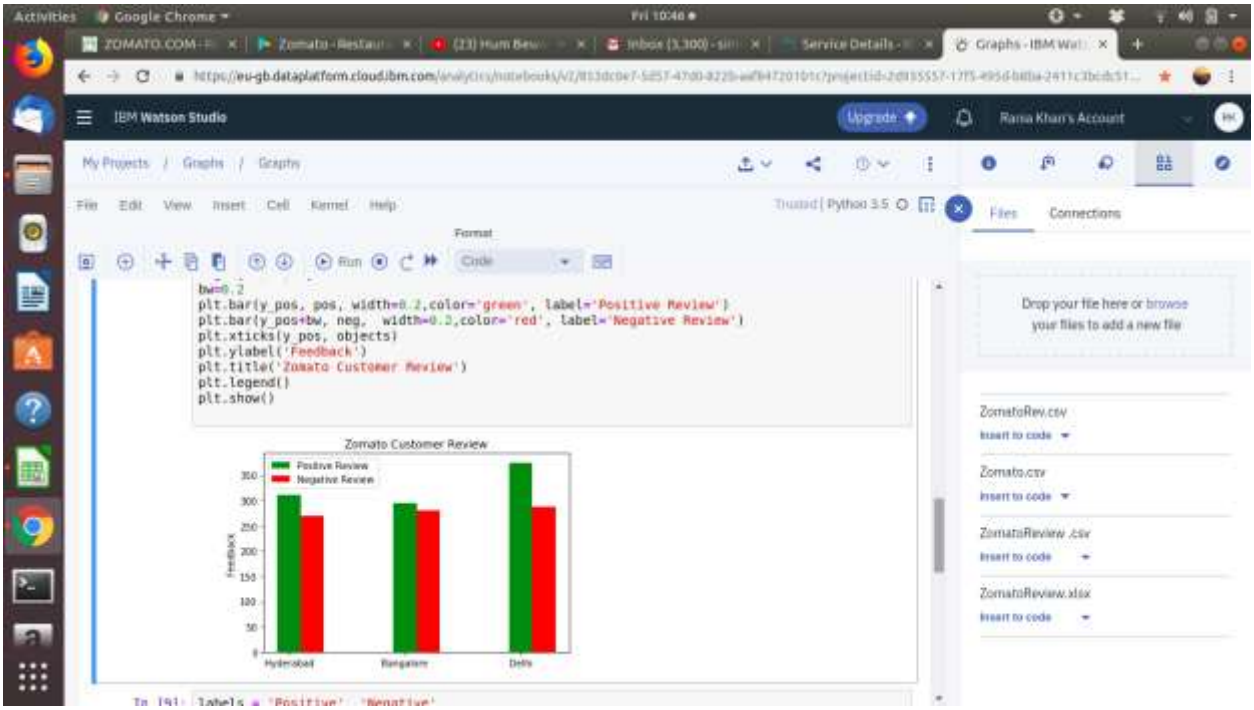
3. Data Collection

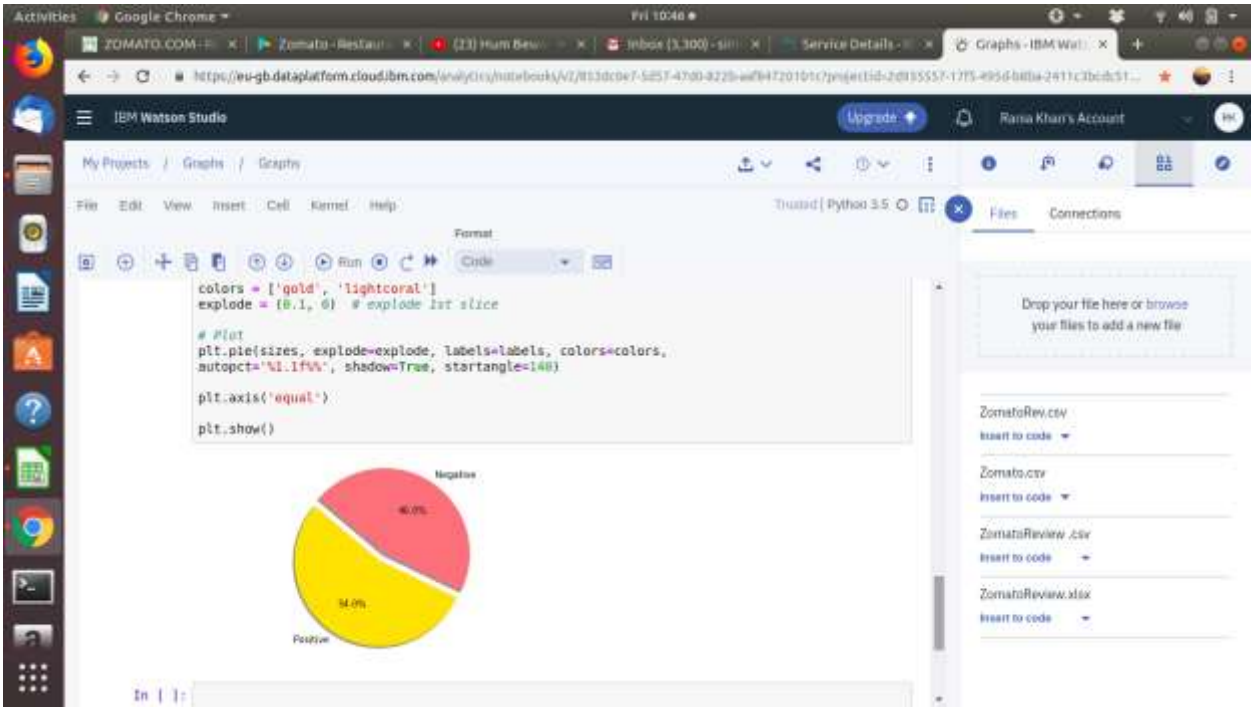
The data has been collected from primary and secondary sources. Primary data includes information collected based on attitude and perception of customers using zomato food delivery app . Secondary data included information from various websites like mouthshut.com, glassdoor.co.in and many more. The attributes present in the dataset are: city,liked,reviews.

4. Methodology

4.1 Exploratory Data Analysis

4.1.1 Figures and Tables





4.2 Data Modelling

In this project, the algorithm applied is NLP. Natural Language Processing is machine's ability to speak like a human.

Basically, Language->Alphabet->Words->Sentences.

NLP Terminology:

Syntax of Language: Vocabulary and Grammar

Text Mining : Converting unstructured data to structured data

Sentimental Analysis : It will segregate the category based on the sentiment of the user.

Chatbot : A chatbot (also known as a spy, conversational bot, chatterbot , interactive agent, conversational interface, Conversational AI, talkbot or artificial spy entity) is a computer program or an artificial intelligence which conducts a conversation via auditory or textual methods.

Speech Recognition : Speech recognition is the interdisciplinary subfield of computational linguistics that develops methodologies and technologies that enables the recognition and translation of spoken language into text by computers.

Machine Translation : Machine Translation, sometimes referred to by tge abbreviation MT(not to be confused with computer-aided translation, machine aided human translation (MAHT)or interactive translation)is a sub-field of computational linguistics that investigate the use of software to translate text or speech from one language to another.

Ambiguity : There are two types-

- Textutal: Confusion in texts due to use of homonyms.
- Ambiguous sentences :

Example: chicken is ready to eat

Here, in the above example the ambiguity is created as it gives two different meanings – either we are going to eat the chicken or the chicken is going to eat us.

NLU & NLG:

NLP is a processed using Natural Language Understanding(NLU) and Natural Language Generation (NLG)

Natural Language Understanding:

- Mapping input to useful representation
- Analyzing different aspects of the language

- Understanding the context of sentences

Natural Language Generation:

- Text planning
- Sentence planning
- Text realization

SOFTWARE USED:

Anaconda Navigator -

Anaconda Navigator is a desktop graphical user interface (GUI) included in Anaconda® distribution that allows you to launch applications and easily manage conda packages, environments and channels without using command-line commands.

Spyder -

Spyder, the Scientific Python Development Environment, is a free integrated development environment (IDE) that is included with Anaconda. It includes editing, interactive testing, debugging and introspection features. Spyder is also pre-installed in Anaconda Navigator, included in Anaconda.

LIBRARIES USED:

Pandas:

- This library is used carry out the whole process of data analysis and data modeling which was not possible in python priorly.
- Pandas allows us to focus much on research and less on programming . In this way it becomes very easy to learn. The bottom line is that it has increased our productivity.
- We use pandas to process time series data on our production servers.
- The simplicity and elegance of its API, and its high level of performance for high-volume datasets, made it a perfect choice for us**.**

NumPy:

- NumPy is a Python package which stands for 'Numerical Python'. It is a library consisting of multidimensional array objects and a collection of routines for processing of array.
- NumPy is the fundamental package for scientific computing with Python.

- Besides its obvious scientific uses, NumPy can also be used as an efficient multi-dimensional container of generic data.
- Arbitrary data-types can be defined. This allows NumPy to seamlessly and speedily integrate with a wide variety of databases.

Matplotlib:

- Matplotlib is an amazing visualization library in Python for 2D plots of arrays.
- Matplotlib is a multi-platform data visualization library built on NumPy arrays and designed to work with the broader SciPy stack.
- One of the greatest benefits of visualization is that it allows us visual access to huge amounts of data in easily digestible visuals.
- Matplotlib consists of several plots like line, bar, scatter, histogram etc.

Keras:

- Keras is a minimalist Python library for deep learning that can run on top of Theano or TensorFlow.
- It was developed based on four principles-
- Modularity: A model can be understood as a sequence or a graph alone. All the concerns of a deep learning model are discrete components that can be combined in arbitrary ways.
- Minimalism: The library provides just enough to achieve an outcome, no frills and maximizing readability.
- Extensibility: New components are intentionally easy to add and use within the framework, intended for researchers to trial and explore new ideas.
- Python: No separate model files with custom file formats. Everything is native Python.

Pickle:

- This library is used for serializing and de-serializing a Python object structure.
- Any object in python can be pickled so that it can be saved on disk. What pickle does is that it “serialises” the object first before writing it to file.
- Pickling is a way to convert a python object (list, dict, etc.) into a character stream.

- The idea is that this character stream contains all the information necessary to reconstruct the object in another python script.

Tkinter:

- The Tkinter module ("Tk interface") is the standard Python interface to the Tk GUI toolkit.
- Tkinter is not the only GUI Programming toolkit for Python. It is however the most commonly used one.
- The Tk interface is located in a binary module named _tkinter. This module contains the low-level interface to Tk, and should never be used directly by application programmers.
- It is usually a shared library (or DLL), but might in some cases be statically linked with the Python interpreter.

5. Findings and Suggestions

We attained the information of our dataset from various sites such as mouthshut.com and the Google Play. We applied the NLP algorithm in order to make the machine understand and generate content in the natural language which is chosen as English in the program, then we went on to create a user interface which gives the user an instant feedback based on their review. We analyzed the data to categorize the positive and negative reviews city wise.

We plotted two graphs, one depicting the number of positive and negative reviews per city and the other one depicting the overall percentage of negative and positive reviews.

We can expand the database to train the machine in a better way. We can furthermore use unsupervised learning to analyze the negative comments and work it. It should be made mandatory that each and every review given by the customer should be considered.

6. Conclusion

The solution for this problem statement is to create a dataset which consists of three things importantly. The three things that we have taken into consideration are the number of reviews given by the customers, places and last but not the least, the indication of the review whether it is positive or negative. Here,a python code based on NLP helps us to sort out the positive and negative reviews. Other specifications like color of the background, buttons, labels and many more can be added as per the requirement.

Basically, an interface between the customer and the zomato application holder is created directly or indirectly. Customers gives a review about the zomato application which is

categorized into positive or negative accordingly. The zomato customer review prediction makes it easier for the zomato application holder to improvise their serivce, makes more profit, add new facilities and have more number of customers with the help of these feedback predictions.

The UI is created both for the customer as well as the developer. The UI of the user has the window were he can give the input and the UI of the developer has the visualized part which includes graphs where he can analyze the statistics. The developer can actually know how the app is working and how people are reacting towards the service provided by the app. He can actually improvise the app by taking all the negative comments into consideration. He can also know the place where the app is used the most.

With positive reviews they can know how many of the customers are approaching and liking their services and with negative reviews they can improvise their services and know the drawbacks in their services.