

# **FLIPKART REVIEW SENTIMENT ANALYSIS USING PYTHON**

**Project submitted to**

**The University of Kerala**

**In partial fulfilment of the requirements for the  
Degree of Master of Science in Applied Statistics & Data Analytics**

**By**

**RESHMA R B**

**(Reg.no.85621615012)**

**Under the Supervision of**

**Dr.C. SATHEESH KUMAR**



**DEPARTMENT OF STATISTICS**

**UNIVERSITY OF KERALA**

**KARIAVATTOM CAMPUS THIRUVANANTHAPURAM**

**2021 - 2023**

**DEPARTMENT OF STATISTICS**  
**UNIVERSITY OF KERALA**



Dr.C. Satheesh Kumar  
Professor and Director  
University of Kerala

Kariavattom  
Thiruvananthapuram

---

**CERTIFICATE**

I hereby certify that the Project **“Flipkart Review Sentiment Analysis Using Python”** is a bonafide project work carried out by Ms. RESHMA R B, MSc Applied Statistics & Data Analytics student in the Department of Statistics, University of Kerala, Kariavattom campus, during 2021 – 2023 under my supervision and guidance, in partial fulfilment of the requirements for the MSc Degree in Applied Statistics & Data analytics of the university of Kerala.

Kariavattom  
September 2023

Dr.C Satheesh Kumar  
Professor and Director  
Supervised Guide

## **ACKNOWLEDGEMENT**

The success and final outcome of this project required a lot of guidance and assistance from many people and I am extremely privileged to have got this all along the completion of my project. All that I have done is only due to such supervision and assistance and I would not forget to thank them.

I wish to express my profound and deep sense of gratitude to Dr.C Satheesh Kumar, Professor and Director & Supervising guide, Department of Statistics, University of Kerala for providing me an opportunity to do this project, his guidance, supervision and constant encouragement throughout the study. His valuable and inspiring suggestions, constructive criticism and moral support helped me in the successful completion of this work. I am extremely thankful to him for providing such a nice support and guidance, although he had busy schedule managing the departmental affairs.

I am thankful to and fortunate enough to get constant encouragement, support and guidance from all Teaching staffs, non-teaching staffs, the Librarian and Research scholars of our department, my parents and all my friends who helped me in successfully completing my project work. I would also like to express my sincere thanks to all the respondents who extended their immense help by giving unstintingly their time and efforts, who helped me in achieving the objectives of the study.

Above all I thank the Almighty for His blessings showered on me throughout the course of this project work.

**RESHMA R B**

# DECLARATION

We hereby declare that the work entitled “FLIPKART REVIEW SENTIMENT ANALYSIS USING PYTHON” submitted to the Kerala University for the partial fulfilment of the Degree of **Applied Statistics & Data Analytics** is an original project work done by **Ms. RESHMA R B** under the guidance of Dr.C SATHEESH KUMAR, Professor and Director, **Department of Statistics, University of Kerala, Kariavattom.**

We further declare that this project report has not been submitted to any other universities institutions or Board for the award of any degree or diploma degree.

KARIAVATTOM

RESHMA R B

SEPTEMBER 2023

# FILPKART REVIEW SENTIMENT ANALYSIS USING PYTHON



MSc Applied Statistics and Data Analytics

# **ABSTRACT**

With the rise of online shopping, the requirement for product review categorization is very important and it can be done using sentiment analysis and with the help of NLP. Sentiment Analysis analyses an incoming message and determines if the underlying sentiment is positive, negative or neutral product. User's comments are useful information to estimate product quality. Customers may communicate their opinions and feelings more openly than ever before, understanding people's emotions is critical for business. It is really difficult for a human to look through each line and find the emotion that represents the user experience. With the advancement of technology, we can now analyse consumer feedback automatically, from survey replies to social media chats, allowing firms to listen to their customers and adjust products and services to match their demands. Natural Language Processing or NLP is a field of Artificial Intelligence that gives the machines the ability to read, understand and derive meaning from human languages. Through this analysis the company can get a full idea of customer feedback and can look after these particular areas. Increased business, fame and brand value profits make the company's customers more loyal. As Flipkart is one of the biggest ecommerce websites of India so we have used it for the sentiment analysis. In this work, propose a sentiment-based rating expectation technique to take care of this issue.

# CONTENTS

## **Chapter-1 INTRODUCTION** **9**

1.1	Motivation of the study	11
1.2	Objective of the study	13
1.3	Significance of the study	17
1.4	Dataset information	20
1.5	Review of Literature	21
1.6	Inference of Literature Review	24

## **Chapter-2 METHODOLOGY** **25**

2.1	Machine Learning	25
2.2	Python	34
2.3	Natural Language Processing (NLP)	38
2.4	Steps and Functionalities	45

2.5 Naïve Bayes .....	47
2.6 Sentiment Analysis .....	51

### **Chapter-3 DATA ANALYSIS**

3.1 Data Description .....	55
3.2 Representation of Data .....	56
3.3 Data Cleaning .....	58
3.4 Data Visualization .....	60
3.5 Key Findings .....	61

### **Chapter-4 CONCLUSION** .....

63

### **Chapter-5 REFERENCE** .....

65



# **Chapter-1**

## **INTRODUCTION**

As online marketplaces have been popular during the past decades, the online sellers and merchants ask their purchasers to share their opinions about the products they have bought. Everyday millions of reviews are generated all over the Internet about different products, services and places. This has made the Internet the most important source of getting ideas and opinions about a product or a service. However, as the number of reviews available for a product grows, it is becoming more difficult for a potential consumer to make a good decision on whether to buy the product. Different opinions about the same product on one hand and ambiguous reviews on the other hand makes customers more confused to get the right decision. Here the need for analysing these contents seems crucial for all e-commerce businesses. Sentiment analysis and classification is a computational study which attempts to address this problem by extracting subjective information from the given texts in natural language, such as opinions and sentiments. Different approaches have used to tackle this problem from natural language processing or NLP, text analysis, computational linguistics, and biometrics. In recent years, Machine learning methods have got popular in the semantic and review analysis for their simplicity and accuracy. Flipkart is one of the e-commerce giants that people are using every day for online purchases where they can read thousands of reviews dropped by other customers about their desired products. These reviews provide valuable opinions about a product such as its property, quality and recommendations which helps the purchasers to understand almost every detail of a product. This is not only beneficial for consumers

but also helps sellers who are manufacturing their own products to understand the consumers and their needs better. This project is considering the sentiment classification problem for online reviews using supervised approaches to determine the overall semantic of customer reviews by classifying them into positive and negative sentiment. The data used in this study is a set of beauty product reviews from Flipkart that is collected from Snap dataset.

Sentiment Analysis is key to determining the emotion of the reviews given by the customer. Sentiment analysis is mainly used in e-commerce platforms or any platform which requires customer opinion to make people express their experience of that product or a thing. Here, I will be demonstrating sentiment analysis on Flipkart Dataset.

Flipkart is one of the leading e-commerce companies/platforms based in India, which currently competes with Snapdeal, Myntra, Nykaa, and IndiaMART. Since the world is digitizing rapidly, e-commerce platforms are booming and making the best way possible to make and provide their services to the people.

The sentiment analysis of the product reviews is a way that can open ways to understand the performance of the products, which is beneficial for both future customers and the e-commerce platform. The e-commerce company can improve its product and customer services with the sentiment analysis technique. Sentiment analysis can also help the company better understand the product and plan further regarding products.

## 1.1 Motivation of the study

Sentiment analysis of Flipkart review using python can provide valuable insights for businesses, improve customer experiences, and serve as a compelling project for research and academic exploration. It bridges the gap between data science, business intelligence, and customer feedback analysis, making it relevant and impactful area of study. The motivation for studying Flipkart review sentiment analysis using python, or sentiment analysis of product reviews in general, can be multifaceted and can include several compelling reasons:

- **Business Insight:**

Sentiment analysis helps business gain valuable insights into customer opinions and preferences. Analysing Flipkart reviews can provide valuable feedback on their products and services, helping them make data-driven decisions for product improvements, marketing strategies, and customer service enhancements.

- **Competitive Analysis:**

Understanding customer sentiments can help business gain a competitive edge. Analysing Flipkart reviews can help companies compare their products and services with those of their competitors, identifying strengths and weaknesses.

- **Product Development:**

Sentiment analysis can guide product development efforts. By identifying recurring issues or feature request in reviews, companies

can prioritize product enhancements and new feature based on customer demand.

- **Marketing and Advertising:**

Positive sentiments in reviews can be leveraged in marketing campaigns. Companies can use positive customer feedback as testimonials or endorsements, attracting more customers.

- **Quality Assurance:**

Sentiment analysis can be part of quality assurance processes. By identifying negative sentiments early, companies can address product or service issues promptly, reducing customer dissatisfaction.

- **Customer Engagement:**

Responding to customer reviews, both positive and negative, can enhance customer engagement and loyalty. Companies can use sentiment analysis to prioritize which reviews to respond to and tailor their response accordingly.

- **Market Research:**

Analysing Flipkart reviews can be a cost-effective way to perform market research. Companies can gather information on consumer preferences, trends, and emerging needs without conducting expensive surveys or focus groups.

- **E-commerce Optimization:**

For E-commerce platforms like Flipkart, understanding sentiment can improve the use experience. Companies can use sentiment analysis to recommend products, personalize user experiences, and optimize search results.

- **Academic and Research Interests:**

Sentiment analysis is a prominent topic in natural language processing and machine learning. Studying Flipkart review sentiment analysis can serve as a practical and interesting project for individuals interested in these fields.

- **Social and Ethical Consideration:**

Studying the sentiment in reviews can also uncover ethical or social issues such as the spread of misinformation or the impact of fake reviews. Researchers may be motivated to address such issues.

## 1.2 Objective of the study

The specific objective may vary depending on the business's goals and the depth of analysis required. Ultimately, the objective is to use sentiment analysis to gain insights that can inform business strategies, improve customer satisfaction, and drive growth in the context of Flipkart's online marketplace.

The objective of conducting a sentiment analysis study on Flipkart reviews using python can be multifaceted and may include several specific goals. Here are some common objectives for such a study:

- **Assess Customer Satisfaction:**

Determine the overall satisfaction level of customers with products or services available on flipkart. This includes identifying areas where customers are most satisfied and areas that need improvement.

- **Identifying Key Product Insights:**

Uncover valuable insights into features, attributes, or aspects of products or service that customers appreciate the most and those that receive criticism.

- **Competitive Analysis:**

Compare products or services from different sellers on Flipkart to understand which ones are preferred by customers and why. Identify strengths and weakness of competitors.

- **Review Volume Analysis:**

Analyze then volume of reviews over time to identify trends, seasonality, or fluctuations in customer feedback.

- **Sentiment Trend Analysis:**

Track sentiment changes over time to identify whether customer sentiment is improving or declining, and correlate these changes with product updates or marketing efforts.

- **Feedback for Product Environment:**

Gather actionable feedback from negative reviews to identify specific areas for product or service improvement. This can inform the development or enhancement of offerings.

- **Marketing Insights:**

Use positive sentiment and feedback in marketing materials to attract potential customers. Highlight strengths and unique selling points in advertisements.

- **Customer Segmentation:**

Segment customer based on their sentiment. Understand the preferences and pain points of different customer groups and tailor marketing strategies accordingly.

- **Predictive Analytics:**

Build predictive models to forecast future sentiment trends and potential sales on historical sentiment data.

- **Business Decision Support:**

Provide decision-makers with data-driven insights that can guide strategic decisions related to inventory management, marketing budget allocation and customer engagement.

- **Customer Relationship Management:**

Develop Strategies for responding to reviews, both positive and negative, to engage with customers, address their concerns, and build trust.

- **Benchmarking:**

Benchmark your products or services against industry standards and competitors to understand your position in the market.

- **Compliance and Quality Control:**

Ensure compliance with quality standards and regulations by monitoring customer feedback for any product or service issues that may arise.

- **Brand Reputation Management:**

Monitor and manage the online reputation of the brand and action to protect or enhance it.

By analysing sentiments in reviews, we can understand what customers like or dislike about products, helping business make informed decision about product improvements, marketing strategies and inventory management.



## 1.3 Significance of the study

The significance of a sentiment analysis study on Flipkart reviews is its ability to provide actionable insights that can drive business growth, enhance customer satisfaction, and maintain or improve a company's competitive position in the market. It enables data-driven decision making and helps business adapt to changing market dynamics.

The significance of conducting a sentiment analysis study on Flipkart reviews using Python or any other method lies in its potential to provide valuable insights and benefits for various stakeholders. Here are the key aspects of its significance:

- **Customer Satisfaction and Experience Improvement:**

Understanding customer sentiment allows businesses to identify areas where customers are satisfied and where improvements are needed. This can lead to enhanced product/service quality and overall customer experience.

- **Product Development and Innovation:**

Analyzing reviews helps in identifying specific product features or attributes that customers appreciate or dislike. This information can

guide product development and innovation efforts to meet customer preferences.

- **Competitive Advantage:**

By comparing sentiment across products or services, businesses can identify their competitive advantages and disadvantages relative to competitors. This insight informs strategic decisions and helps maintain or gain market share.

- **Marketing and Sales Strategies:**

Positive sentiment can be leveraged in marketing campaigns to attract more customers. Addressing negative sentiment allows for tailored strategies to mitigate issues and improve sales.

- **Customer Engagement and Retention:**

Engaging with customers through reviews, addressing their concerns, and implementing their suggestions can build trust and loyalty, leading to increased customer retention.

- **Operational Efficiency:**

Sentiment analysis can highlight operational inefficiencies or issues in supply chains, customer service, or logistics. Addressing these issues can lead to cost savings and improved processes.

- **Data-Driven Decision-Making:**

Businesses can make informed, data-driven decisions rather than relying on intuition or assumptions. This reduces the risk of making incorrect choices that could harm the business.

- **Brand Reputation Management:**

Monitoring and responding to reviews can help protect and enhance the brand's online reputation. Quick responses to negative reviews can mitigate damage and demonstrate commitment to customer satisfaction.

- **Market Trends and Customer Preferences:**

Sentiment analysis can reveal evolving market trends and shifts in customer preferences. Staying attuned to these changes can help businesses adapt and remain competitive. I

- **Regulatory Compliance:**

For industries with regulatory requirements, monitoring sentiment can help identify potential compliance issues and facilitate proactive measures to address them.

- **Investor and Stakeholder Confidence:**

Positive sentiment and evidence of customer satisfaction can boost investor and stakeholder confidence, potentially leading to increased investments or partnerships.

- **Continuous Improvement:**

Ongoing sentiment analysis provides a feedback loop for continuous improvement. Businesses can track changes in sentiment over time and adjust strategies accordingly.

- **Customer Centric Approach:**

Demonstrating a commitment to understanding and responding to customer sentiment fosters a customer-centric approach, which is increasingly vital in today's competitive market.

- **Predictive Insights:**

Over time, historical sentiment data can be used to predict future trends and customer behavior, aiding in long-term planning and decision-making.

## 1.4 Dataset Information

The dataset I am using here for Flipkart reviews sentiment analysis is downloaded from Kaggle.

The dataset consists of reviews of the products which customers purchased from Flipkart. The reviews given are the experience about the product customer bought through Flipkart and given a rating to that product.

The dataset for Flipkart contains 3 columns:

**Product Name**: the name of the product with little details about it.

**Review**: review what the customer gives about that product.

**Rating**: ranges from 1 to 5 stars given by the customer.

## 1.5 Literature Review

Analyzing Flipkart reviews sentiment has been a popular topic in natural language processing (NLP) and sentiment analysis research. Here's a brief literature review summarizing some key studies and findings related to Flipkart review sentiment analysis:

### **1. "Sentiment Analysis of Flipkart Customer Reviews Using Machine Learning" (2019):**

- This study focused on classifying Flipkart customer reviews into positive, negative, or neutral sentiments using various machine learning algorithms.

The authors experimented with different features, including bag-of-words and TF-IDF, and compared the performance of classifiers like Support Vector Machines (SVM) and Naive Bayes.

They found that using TF-IDF features and SVM yielded promising results in accurately classifying sentiments.

## **2. "A Comparative Analysis of Sentiment Analysis Techniques on Flipkart Reviews" (2020):**

- This research compared the performance of multiple sentiment analysis techniques, including lexicon-based methods, machine learning, and deep learning, on Flipkart reviews.

The study found that deep learning models, such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs), outperformed traditional methods in sentiment classification tasks.

## **3. "Enhancing Sentiment Analysis of E-commerce Product Reviews Using Aspect-Based Sentiment Analysis on Flipkart Data" (2021):**

- This study delved into aspect-based sentiment analysis (ABSA) of Flipkart product reviews, aiming to identify sentiments towards specific aspects or features of products.

The authors employed techniques like dependency parsing and sentiment lexicons to extract aspect-based sentiments and observed that this approach provides more detailed insights than overall sentiment analysis.

#### **4. "Comparative Analysis of Sentiment Analysis Approaches on Flipkart Reviews" (2018)**

- This research compared traditional machine learning algorithms (e.g., Naïve Bayes and Random Forest) and deep learning models (eg. LSTM) for sentiment analysis on Flipkart reviews

The findings indicated that deep learning models generally outperformed traditional approaches, particularly in handling complex sentence structures and context.

#### **5. "Sentiment Analysis of Flipkart Product Reviews: A Study on User Experience" (2017):**

This study focused on analysing the sentiment of Flipkart product

reviews with a particular emphasis on the user experience aspect

It found that understanding user experiences from reviews can help businesses improve their products and services by identifying pain points and areas for enhancement.

These studies demonstrate the interest in sentiment analysis of Flipkart reviews and highlight the use of various techniques, from traditional machine learning to deep learning and aspect-based sentiment analysis, to gain valuable insights from user-generated content on e-commerce platforms like Flipkart. The research in this field continues to evolve with advancements in NLP and machine learning techniques.

## 1.6 Inference of Literature Review

Inferences from Literature review from the above literature review, it is inferred that sentiment analysis can be performed by using many algorithms but, Naïve Bayes Classifier and Logistic Regression gives optimum results. Out of the two, Logistic Regression proves to be better and works efficiently even on small datasets. For user trustworthiness, AHP is an effective approach in dealing with decision problems using a number of parameters to generate the ranking which is not possible in our case because of limited parameters. Hence, a custom scale would be used for user trustworthiness.



## Chapter-2

# METHODOLOGY

### 2.1 Machine Learning

Machine learning is a method of data analysis that automates analytical model building. It is a branch of artificial intelligence based on the idea that systems can learn from data, identify patterns and make decisions with minimal human intervention.

#### ✓ **How machine learning works**

UC Berkeley (link resides outside IBM) breaks out the learning system of a machine learning algorithm into three main parts.

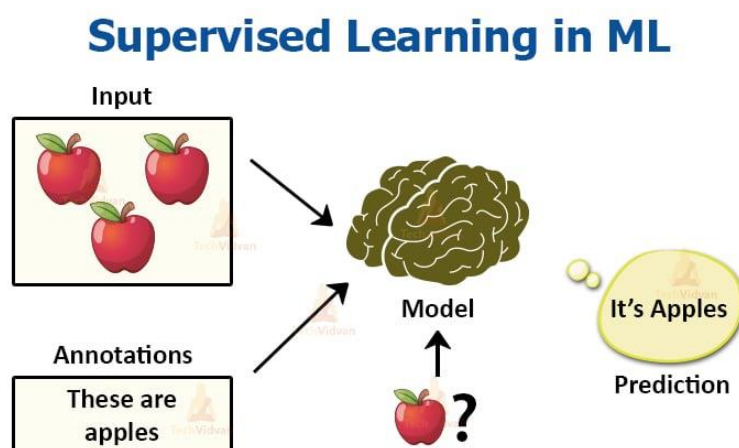
1. A Decision Process: In general, machine learning algorithms are used to make a prediction or classification. Based on some input data, which can be labelled or unlabelled, your algorithm will produce an estimate about a pattern in the data.
2. An Error Function: An error function serves to evaluate the prediction of the model. If there are known examples, an error function can make a comparison to assess the accuracy of the model.
3. A Model Optimization Process: If the model can fit better to the data points in the training set, then weights are adjusted to reduce the discrepancy between the known example and the model estimate. The algorithm will repeat this evaluate and optimize process, updating weights autonomously until a threshold of accuracy has been met.

## 🚦 Machine learning methods

Machine learning classifiers fall into three primary categories

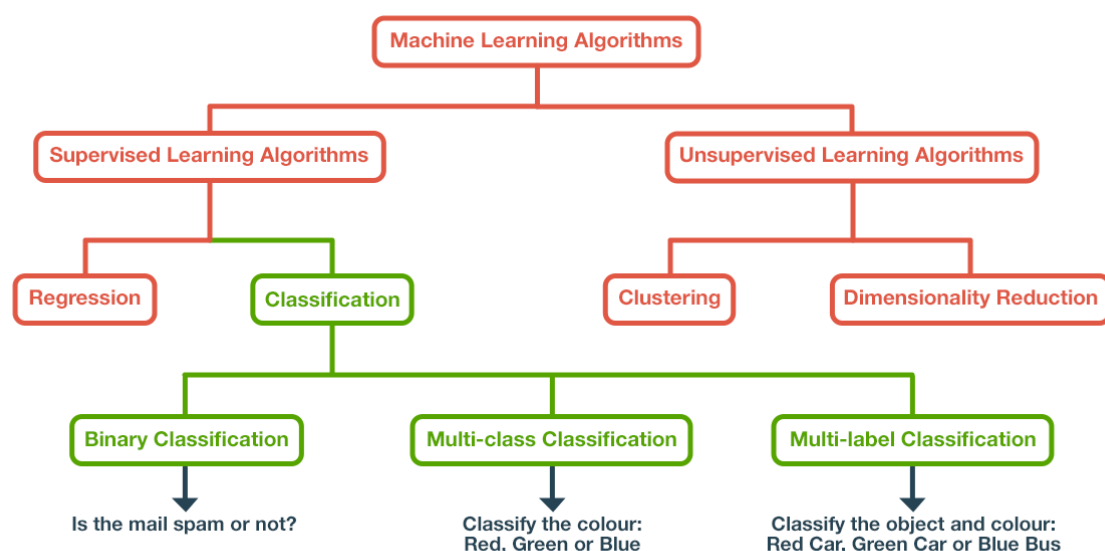
- **Supervised Machine Learning**

supervised learning, also known as supervised machine learning, is defined by its use of labelled datasets to train algorithms that to classify data or predict outcomes accurately. As input data is fed into the model, it adjusts its weights until the model has been fitted appropriately. This occurs as part of the cross-validation process to ensure that the model avoids overfitting or underfitting. Supervised learning helps organizations solve for a variety of real-world problems at scale, such as classifying spam in a separate folder from your inbox. Some methods used in supervised learning include neural networks, naïve bayes, linear regression, logistic regression, random forest, support vector machine (SVM), and more.



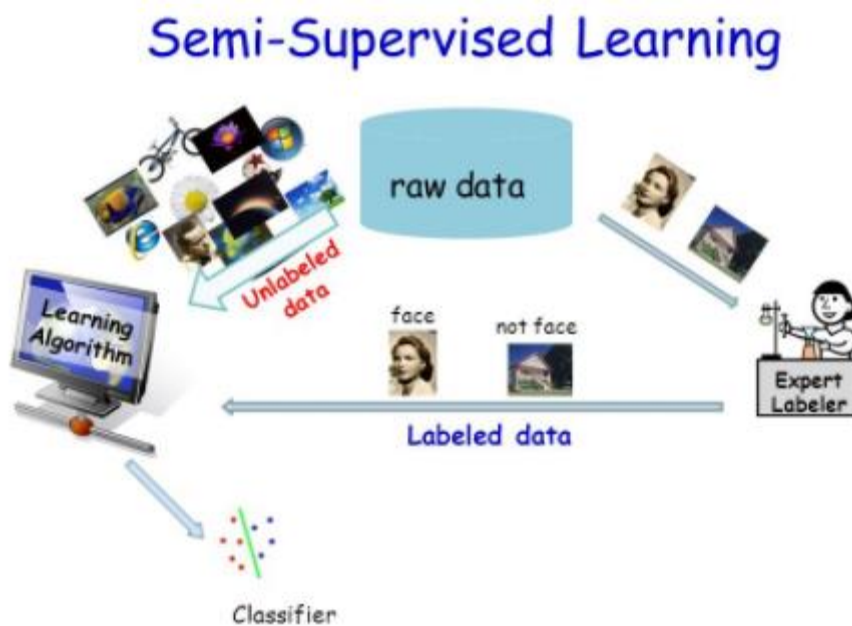
- **Unsupervised Machine Learning**

Unsupervised learning, also known as unsupervised machine learning, uses machine learning algorithms to analyze and cluster unlabelled datasets. These algorithms discover hidden patterns or data groupings without the need for human intervention. Its ability to discover similarities and differences in information make it the ideal solution for exploratory data analysis, cross-selling strategies, customer segmentation, image and pattern recognition. It's also used to reduce the number of features in a model through the process of dimensionality reduction; principal component analysis (PCA) and singular value decomposition (SVD) are two common approaches for this. Other algorithms used in unsupervised learning include neural networks, k-means clustering, probabilistic clustering methods, and more.



- **Semi-supervised Machine Learning**

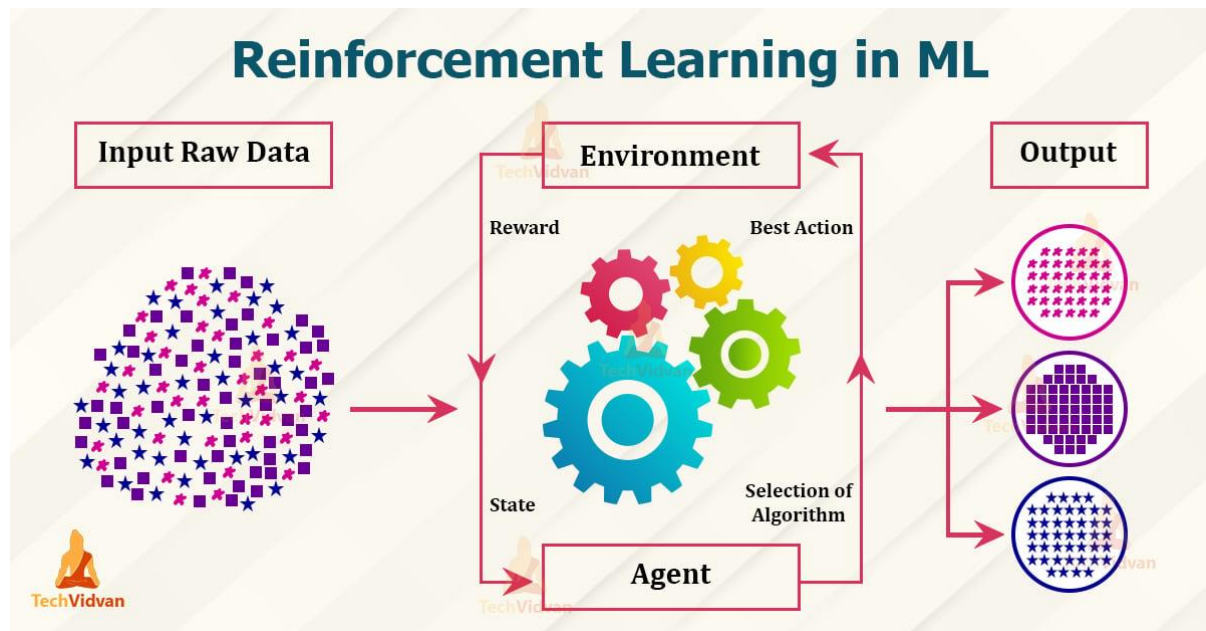
Semi-supervised learning offers a happy medium between supervised and unsupervised learning. During training, it uses a smaller labeled data set to guide classification and feature extraction from a larger, unlabeled data set. Semi-supervised learning can solve the problem of having not enough labeled data (or not being able to afford to label enough data) to train a supervised learning algorithm.



- **Reinforcement Machine Learning**

Reinforcement machine learning is a behavioral machine learning model that is similar to supervised learning, but the algorithm isn't trained using sample data. This model learns as it goes by using trial

and error. A sequence of successful outcomes will be reinforced to develop the best recommendation or policy for a given problem.



## Models

Performing machine learning involves creating a model, which is trained on some training data and then can process additional data to make predictions. Various types of models have been used and researched for machine learning systems.

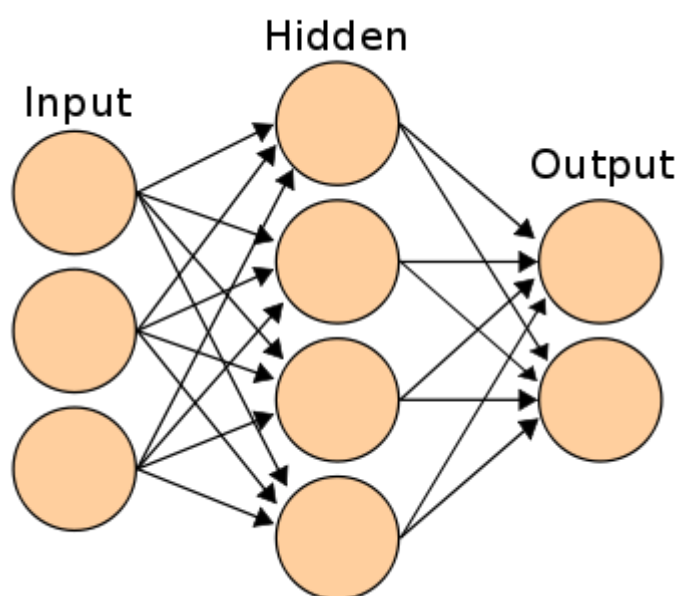
- **Artificial Neural Networks**

An artificial neural network is an interconnected group of nodes, akin to the vast network of neurons in a brain. Here, each circular node represents an artificial neuron and an arrow represents a connection from the output of one artificial neuron to the input of another. Artificial neural networks (ANNs), or connectionist systems, are computing systems vaguely inspired by the biological neural

networks that constitute animal brains. Such systems "learn" to perform tasks by considering examples, generally without being programmed with any task-specific rules.

An ANN is a model based on a collection of connected units or nodes called "artificial neurons", which loosely model the neurons in a biological brain.

ANN algorithm is depicted as follows:



Each connection, like the synapses in a biological brain, can transmit information, a "signal", from one artificial neuron to another. An artificial neuron that receives a signal can process it and then signal additional artificial neurons connected to it. In common ANN implementations, the signal at a connection between artificial neurons is a real number, and the output of each artificial neuron is computed by some non-linear function of the sum of its inputs. The connections between artificial neurons are called "edges". Artificial neurons and edges typically have a weight that adjusts as learning proceeds. The weight increases or

decreases the strength of the signal at a connection. Artificial neurons may have a threshold such that the signal is only sent if the aggregate signal crosses that threshold. Typically, artificial neurons are aggregated into layers. Different layers may perform different kinds of transformations on their inputs. Signals travel from the first layer (the input layer) to the last layer (the output layer), possibly after traversing the layers multiple times. The original goal of the ANN approach was to solve problems in the same way that a human brain would. However, over time, attention moved to performing specific tasks, leading to deviations from biology. Artificial neural networks have been used on a variety of tasks, including computer vision, speech recognition, machine translation, social network filtering, playing board and video games and medical diagnosis. Deep learning consists of multiple hidden layers in an artificial neural network. This approach tries to model the way the human brain processes light and sound into vision and hearing. Some successful applications of deep learning are computer vision and speech recognition.

- **Decision Trees**

Decision tree learning uses a decision tree as a predictive model to go from observations about an item (represented in the branches) to conclusions about the item's target value (represented in the leaves). It is one of the predictive modeling approaches used in statistics, data mining, and machine learning. Tree models where the target variable can take a discrete set of values are called classification trees; in these tree structures, leaves represent class labels and branches represent conjunctions of features that lead to those class labels. Decision trees where the target variable can take continuous values (typically real numbers) are called regression trees. In decision analysis, a decision tree can be used to visually and explicitly

represent decisions and decision making. In data mining, a decision tree describes data, but the resulting classification tree can be an input for decision making.

## Algorithms used in Decision Trees

1

### ID3

Iterative Dichotomiser 3, an extension of D3. Uses top-down approach to build a decision tree.

2

### C4.5

Uses C programming language. Succession of ID3.

3

### CART

Classification and Regression Tree. Generates future predictions based on already available values.

4

### CHAID

Chi-square automatic interaction detection. Based on adjusted significance testing.

5

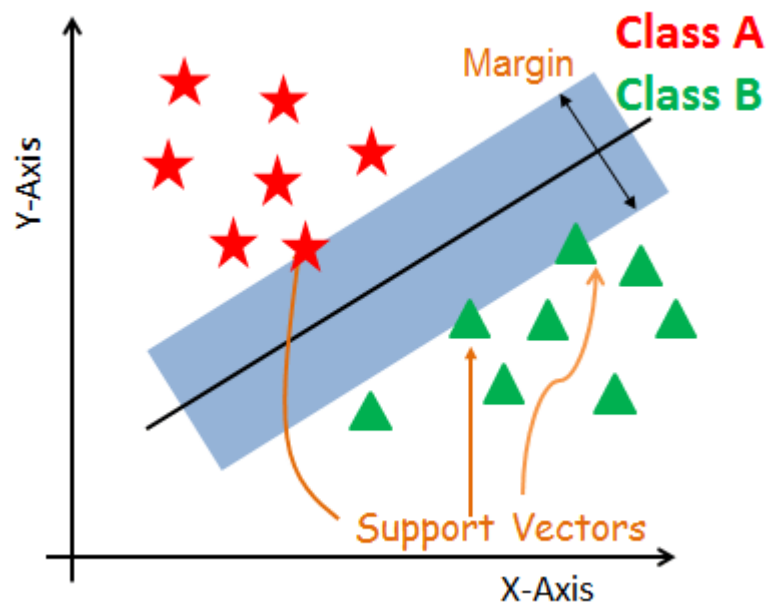
### MARS

Multivariate adaptive regression splines. Helps solve non-linear regression problems.



- **Support Vector Machine**

Support-vector machines (SVMs), also known as support-vector networks, are a set of related supervised learning methods used for classification and regression. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that predicts whether a new example falls into one category or the other. An SVM training algorithm is a non-probabilistic, binary, linear classifier, although methods such as Platt scaling exist to use SVM in a probabilistic classification setting. In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high dimensional feature access.



- **Regression Analysis**

Regression analysis encompasses a large variety of statistical methods to estimate the relationship between input variables and their associated features. Its most common form is linear regression, where a single line is drawn to best fit the given data according to a mathematical criterion such as ordinary least squares. The latter is often extended by regularization (mathematics) methods to mitigate overfitting and bias, as in ridge regression. When dealing with non-linear problems, go-to models include polynomial regression (for example, used for trendline fitting in Microsoft Excel), logistic regression (often used in statistical classification) or even kernel regression, which introduces nonlinearity by taking advantage of the kernel trick to implicitly map input variables to higher dimensional space.

## 2.2 Python

Python is a high-level, general purpose programming language. Guido Van Rossum began working on Python in the late 1980s as a successor to the ABC programming language and first released it in 1991 as Python 0.9.0. Its design offers filter, map, reduce functions, list comprehensions, dictionaries, sets etc.

Python's simplicity, readability, and extensive libraries have contributed to its popularity and widespread adoption in various industries and domains. It's an excellent choice for both beginners and experienced developers looking to build a wide range of applications efficiently. Here's a brief overview of Python:

### ➤ **Readability**

Python's syntax is designed to be clear and easy to read, making it an excellent choice for beginners and experienced programmers alike. It emphasizes code readability with its use of indentation (whitespace) to define code blocks.

### ➤ **Versatility**

Python is a versatile language used in a wide range of applications, from web development and scripting to data analysis, machine learning, scientific computing, automation, and more. It has a vast ecosystem of libraries and frameworks to support various tasks.

### ➤ **Interpreted Language**

Python is an interpreted language, which means you don't need to compile your code before running it. This makes development faster and more flexible.

### ➤ **Cross-Platform**

Python is available on multiple platforms, including Windows, macOS, and Linux, making it a cross-platform choice for developing applications.

### ➤ **Large Standard Library**

Python comes with a comprehensive standard library that provides modules and functions for a wide variety of tasks, such as file handling, networking, data manipulation, and more. This minimizes the need for writing custom code from scratch.

### ➤ **Community and Ecosystem**

Python has a large and active community of developers who contribute to its growth. It has a vast ecosystem of third-party libraries and frameworks, such as Django and Flask for web development, NumPy and pandas for data manipulation, TensorFlow and PyTorch for machine learning, and many more.

### ➤ **Dynamic Typing**

Python uses dynamic typing, which means you don't need to declare variable types explicitly. The interpreter infers the type based on the assigned value.

### ➤ **Object-Oriented**

Python is an object-oriented language, which means it supports the principles of object-oriented programming, including encapsulation, inheritance, and polymorphism.

### ➤ **Interoperability**

Python can easily integrate with other languages like C, C++, and Java, allowing you to leverage existing code and libraries.

### ➤ **Community Support**

Python has an active online community where developers can seek help, share knowledge, and collaborate on open-source projects.

### ➤ **Web Development**

Python is used for building web applications and web services with frameworks like Django, Flask, and FastAPI.

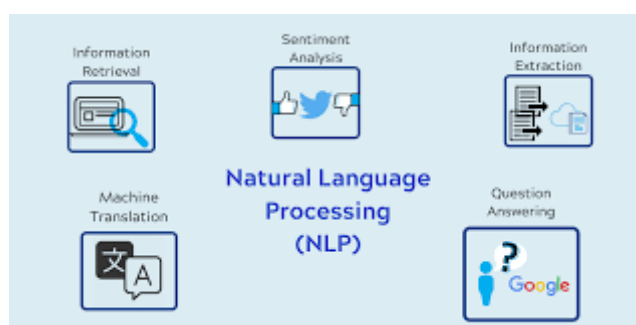
### ➤ **Data Science and Machine Learning**

Python has become the go-to language for data analysis and machine learning due to libraries like NumPy, pandas, scikit-learn, and deep learning frameworks like TensorFlow and PyTorch.

### ➤ **Automation and Scripting**

Python is often used for automating repetitive tasks and writing scripts for system administration, making it a valuable tool for DevOps

## 2.2 Natural Language Processing (NLP)



Natural language processing (NLP) is a subfield of linguistics, computer science, and artificial intelligence concerned with the interactions between computers and human language, in particular how to program computers to process and analyze large amounts of natural language data. The goal is a computer capable of "understanding" the contents of documents, including the contextual nuances of the language within them. The technology can then accurately extract information and insights contained in the documents as well as categorize and organize the documents themselves.

Challenges in natural language processing frequently involve speech recognition, natural language understanding, and natural language generation.

The following is a list of some of the most commonly researched tasks in natural language processing. Some of these tasks have direct real-world applications, while others more commonly serve as subtasks that are used to aid in solving larger tasks.

Though natural language processing tasks are closely intertwined, they can be subdivided into categories for convenience. A coarse division is given below.

- ❖ **Text and speech processing**
- ❖ **Optical character recognition (OCR)**

Given an image representing printed text, determine the corresponding text.

- ❖ **Speech recognition**

Given a sound clip of a person or people speaking, determine the textual representation of the speech.

This is the opposite of text to speech and is one of the extremely difficult problems colloquially termed "AI-complete" (see above). In natural speech there are hardly any pauses between successive words, and thus speech segmentation is a necessary subtask of speech recognition (see below). In most spoken languages, the sounds representing successive letters blend into each other in a process termed coarticulation, so the conversion of the analog signal to discrete characters can be a very difficult process. Also, given that words in the same language are spoken by people with different accents, the speech recognition software must be able to recognize the wide variety of input as being identical to each other in terms of its textual equivalent.

- ❖ **Speech segmentation**

Given a sound clip of a person or people speaking, separate it into words. A subtask of speech recognition and typically grouped with it.

- ❖ **Text-to-speech**

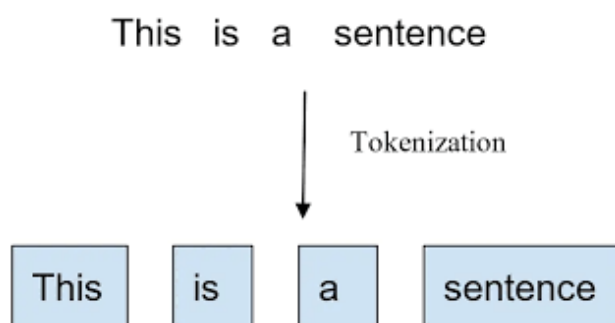
Given a text, transform those units and produce a spoken representation. Text-to-speech can be used to aid the visually impaired.

- ❖ **Word segmentation (Tokenization)**

Separate a chunk of continuous text into separate words. For a language like English, this is fairly trivial, since words are usually

separated by spaces. However, some written languages like Chinese, Japanese and Thai do not mark word boundaries in such a fashion, and in those languages text segmentation is a significant task requiring knowledge of the vocabulary and morphology of words in the language. Sometimes this process is also used in cases like bag of words (BOW) creation in data mining

For example:



### ❖ Morphological analysis

#### ❖ Lemmatization

The task of removing inflectional endings only and to return the base dictionary form of a word which is also known as a lemma.

Lemmatization is another technique for reducing words to their normalized form. But in this case, the transformation actually uses a dictionary to map words to their actual form.

#### ❖ Morphological segmentation

Separate words into individual morphemes and identify the class of the morphemes. The difficulty of this task depends greatly on the complexity of the morphology (i.e., the structure of words) of the language being considered. English has fairly simple morphology, especially inflectional morphology, and thus it is often possible to ignore this task entirely and simply model all possible forms of a word (e.g., "open, opens, opened, opening") as separate words. In



languages such as Turkish or Meitei, a highly agglutinated Indian language, however, such an approach is not possible, as each dictionary entry has thousands of possible word forms.

### ❖ **Part-of-speech tagging**

Given a sentence, determine the part of speech (POS) for each word. Many words, especially common ones, can serve as multiple parts of speech. For example, "book" can be a noun ("the book on the table") or verb ("to book a flight"); "set" can be a noun, verb or adjective; and "out" can be any of at least five different parts of speech.

### ❖ **Stemming**

The process of reducing inflected (or sometimes derived) words to a base form (e.g., "close" will be the root for "closed", "closing", "close", "closer" etc.). Stemming yields similar results as lemmatization, but does so on grounds of rules, not a dictionary.

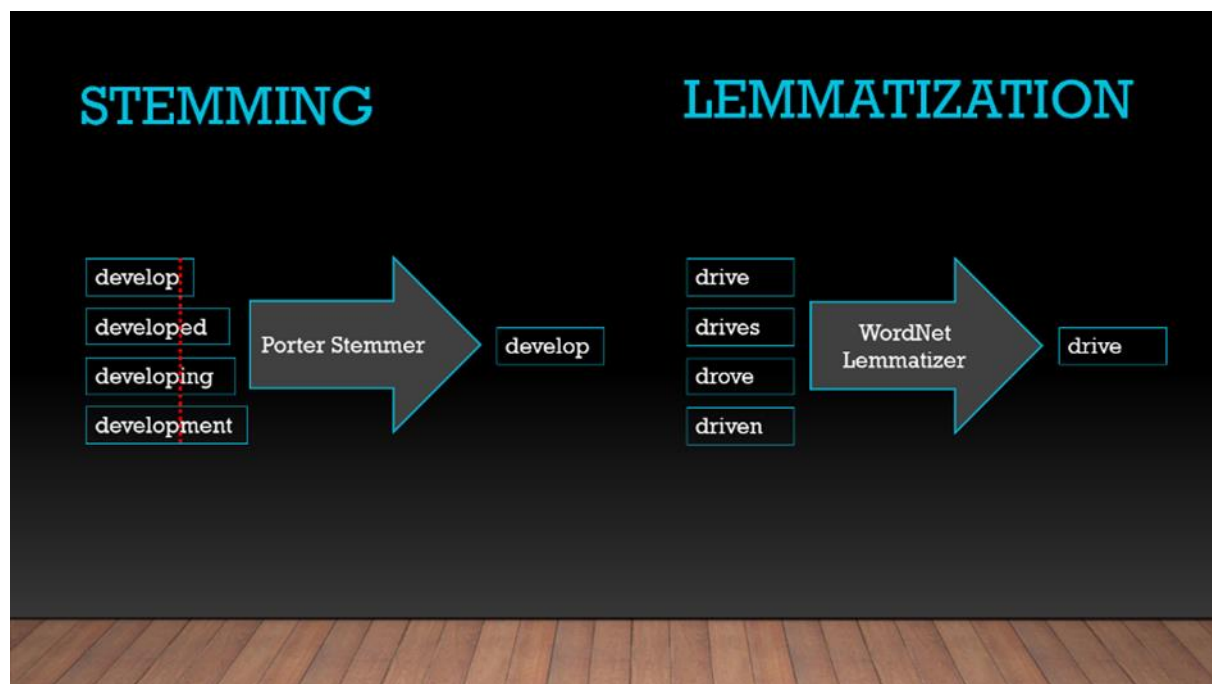
### ❖ **Syntactic analysis**

### ❖ **Grammar induction**

### ❖ **Generate a formal grammar that describes a language's syntax.**

### ❖ **Sentence breaking (also known as "sentence boundary disambiguation")**

Natural language processing strives to build machines that understand and respond to text or voice data—and respond with text or speech of their own—in much the same way humans do.



- **What is natural language processing?**

Natural language processing (NLP) refers to the branch of computer science—and more specifically, the branch of artificial intelligence or AI—concerned with giving computers the ability to understand text and spoken words in much the same way human beings can.

NLP combines computational linguistics—rule-based modeling of human language—with statistical, machine learning, and deep learning models. Together, these technologies enable computers to process human language in the form of text or voice data and to ‘understand’ its full meaning, complete with the speaker or writer’s intent and sentiment.

NLP drives computer programs that translate text from one language to another, respond to spoken commands, and summarize large volumes of text rapidly—even in real time. There’s a good chance you’ve interacted with NLP in the form of voice-operated GPS systems, digital assistants, speech-to-text dictation software,

customer service chatbots, and other consumer conveniences. But NLP also plays a growing role in enterprise solutions that help streamline business operations, increase employee productivity, and simplify mission-critical business processes.

### ❖ NLP Tasks

Human language is filled with ambiguities that make it incredibly difficult to write software that accurately determines the intended meaning of text or voice data. Homonyms, homophones, sarcasm, idioms, metaphors, grammar and usage exceptions, variations in sentence structure—these just a few of the irregularities of human language that take humans years to learn, but that programmers must teach natural language-driven applications to recognize and understand accurately from the start, if those applications are going to be useful.

Several NLP tasks break down human text and voice data in ways that help the computer make sense of what it's ingesting. Some of these tasks include the following:

Speech recognition, also called speech-to-text, is the task of reliably converting voice data into text data. Speech recognition is required for any application that follows voice commands or answers spoken questions. What makes speech recognition especially challenging is the way people talk—quickly, slurring words together, with varying emphasis and intonation, in different accents, and often using incorrect grammar. Part of speech tagging, also called grammatical tagging, is the process of determining the part of speech of a particular word or piece of text based on its use and context. Part of speech identifies ‘make’ as a verb in ‘I can make a paper plane,’ and as a noun in ‘What make of car do you own?’

Word sense disambiguation is the selection of the meaning of a word with multiple meanings through a process of semantic analysis that determine the word that makes the most sense in the given context.

For example, word sense disambiguation helps distinguish the meaning of the verb 'make' in 'make the grade' (achieve) vs. 'make a bet' (place). Named entity recognition, or NEM, identifies words or phrases as useful entities. NEM identifies 'Kentucky' as a location or 'Fred' as a man's name.

Co-reference resolution is the task of identifying if and when two words refer to the same entity. The most common example is determining the person or object to which a certain pronoun refers (e.g., 'she' = 'Mary'), but it can also involve identifying a metaphor or an idiom in the text (e.g., an instance in which 'bear' isn't an animal but a large hairy person). Sentiment analysis attempts to extract subjective qualities—attitudes, emotions, sarcasm, confusion, suspicion—from text.

Natural language generation is sometimes described as the opposite of speech recognition or speech-to-text; it's the task of putting structured information into human language. See the blog post “NLP vs. NLU vs. NLG: the differences between three natural language processing concepts” for a deeper look into how these concepts relate.

### ❖ **NLP tools and approaches**

#### ❖ **Python and the Natural Language Toolkit (NLTK)**

The Python programming language provides a wide range of tools and libraries for attacking specific NLP tasks. Many of these are found in the Natural Language Toolkit, or NLTK, an open-source collection of libraries, programs, and education resources for building NLP programs

The NLTK includes libraries for many of the NLP tasks listed above, plus libraries for subtasks, such as sentence parsing, word segmentation, stemming and lemmatization (methods of trimming words down to their roots), and tokenization (for breaking phrases, sentences, paragraphs and passages into tokens that help the

computer better understand the text). It also includes libraries for implementing capabilities such as semantic reasoning, the ability to reach logical conclusions based on facts extracted from text.

## 2.4 Steps and Functionalities

Choosing the right Python sentiment analysis library is crucial for accurate and efficient analysis of textual data. For organizations, sentiment analysis can help them understand customer sentiments toward their products or services. This information can be used to improve customer experience, target marketing efforts, and make informed business decisions.

Sentiment analysis can also be used in social media monitoring, political analysis, and market research. It can help governments and organizations gauge public opinion on policies, products, or events, and it can help researchers analyze and understand large amounts of textual data.

Here, the libraries used are given below:

### **Pandas**

Pandas is a popular Python library used primarily for data manipulation and analysis. It provides data structures and functions for working with structured data, making it an essential tool for data scientists, analysts, and anyone working with tabular data.

## **NumPy**

NumPy in short for "Numerical Python," is a fundamental Python library: for numerical and scientific computing. It provides support for working with large, multi-dimensional arrays and matrices, along with a collection of mathematical functions to operate on these arrays. NumPy is a critical library for data scientists, engineers, and researchers, and it serves several key purposes.

## **Matplotlib**

Matplotlib is a popular Python library for creating static, animated, and interactive visualizations in a wide range of formats. It provides a versatile and flexible framework for creating high-quality plots and charts, making it an essential tool for data visualization, scientific plotting, and presentation of data in various domains.

## **Seaborn**

Seaborn is a Python data visualization library based on Matplotlib that provides a high-level interface for creating informative and attractive statistical graphics. Seaborn is particularly useful for creating complex visualizations with minimal code and enhancing the aesthetics of your plots. It is built on top of Matplotlib and closely integrates with pandas DataFrames.

## **NLTK**

NLTK, which stands for Natural Language Toolkit, is a comprehensive Python library for natural language processing (NLP). NLTK provides

tools and resources to work with human language data, making it a valuable resource for researchers, linguists, and developers interested in tasks related to language understanding and text analysis.

## **WordCloud**

WordCloud is a Python library used for generating word clouds from text data. Word clouds are visual representations of text data, where words are displayed in varying sizes, with more frequent words appearing larger. Word clouds are often used for visualizing the most common words in a corpus of text and gaining insights into the prominent themes or topics within the text.

The WordCloud library in Python allows us to create customized word clouds with various options for coloring, font size, and layout. It is a popular choice for text analysis, data visualization, and generating aesthetically pleasing word clouds for presentations and reports.

## **Plotly**

Plotly is a Python library for creating interactive data visualizations. It allows us to create a wide range of interactive charts and graphs, including scatter plots, bar charts, line charts, heatmaps, and more. Plotly is often used in data analysis and data science projects to create visually appealing and interactive data visualizations for better data exploration and communication. It can be used in various Python environments, such as Jupyter Notebooks, standalone Python scripts, and web applications.

## Scikit-learn

Scikit-learn (also known as sklearn) is a popular Python library for machine learning and data science. It provides simple and efficient tools for data analysis and modeling, including a wide range of machine learning algorithms for tasks like classification, regression, clustering, dimensionality reduction, and more.

## 2.5 Naïve Bayes

The Naïve Bayes classifier is a supervised machine learning algorithm, which is used for classification tasks, like text classification. It is also part of a family of generative learning algorithms, meaning that it seeks to model the distribution of inputs of a given class or category.

- It is mainly used in *text classification* that includes a high-dimensional training dataset.
- Naïve Bayes Classifier is one of the simple and most effective Classification algorithms which helps in building the fast machine learning models that can make quick predictions.
- It is a probabilistic classifier, which means it predicts on the basis of the probability of an object.
- Some popular examples of Naïve Bayes Algorithm are **spam filtration, Sentimental analysis, and classifying articles.**



The Naïve Bayes algorithm is comprised of two words Naïve and Bayes, which can be described as:

- **Naïve:** It is called Naïve because it assumes that the occurrence of a certain feature is independent of the occurrence of other features. Such as if the fruit is identified on the bases of color, shape, and taste, then red, spherical, and sweet fruit is recognized as an apple. Hence each feature individually contributes to identify that it is an apple without depending on each other.
- **Bayes:** It is called Bayes because it depends on the principle of Bayes' Theorem.

## Bayes' Theorem:

- Bayes' theorem is also known as **Bayes' Rule** or **Bayes' law**, which is used to determine the probability of a hypothesis with prior knowledge. It depends on the conditional probability.
- The formula for Bayes' theorem is given as:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

**Where,**

**P(A|B) is Posterior probability:** Probability of hypothesis A on the observed event B.

**$P(B|A)$  is Likelihood probability:** Probability of the evidence given that the probability of a hypothesis is true.

**$P(A)$  is Prior Probability:** Probability of hypothesis before observing the evidence.

**$P(B)$  is Marginal Probability:** Probability of Evidence.

### Advantages of Naïve Bayes Classifier:

- Naïve Bayes is one of the fast and easy ML algorithms to predict a class of datasets.
- It can be used for Binary as well as Multi-class Classifications.
- It performs well in multi-class predictions as compared to the other Algorithms.
- It is the most popular choice for **text classification problems**.

### Disadvantages of Naïve Bayes Classifier:

- Naive Bayes assumes that all features are independent or unrelated, so it cannot learn the relationship between features.

### Applications of Naïve Bayes Classifier:

- It is used for **Credit Scoring**.
- It is used in **medical data classification**.
- It can be used in **real-time predictions** because Naïve Bayes Classifier is an eager learner.
- It is used in Text classification such as **Spam filtering** and **Sentiment analysis**

## Types of Naïve Bayes Model:

There are three types of Naive Bayes Model, which are given below:

- **Gaussian:** The Gaussian model assumes that features follow a normal distribution. This means if predictors take continuous values instead of discrete, then the model assumes that these values are sampled from the Gaussian distribution.
- **Multinomial:** The Multinomial Naïve Bayes classifier is used when the data is multinomial distributed. It is primarily used for document classification problems, it means a particular document belongs to which category such as Sports, politics, etc.  
The classifier uses the frequency of words for the predictors.
- **Bernoulli:** The Bernoulli classifier works similar to the Multinomial classifier, but the predictor variables are the independent Booleans variables. Such as if a particular word is present or not in a document. This model is also famous for document classification tasks.

## 2.6 Sentiment Analysis

Sentiment Analysis is a machine learning tool that analyses text for polarity, from positive to negative. By training machine learning tools with examples of emotions in text, machines automatically learn how to detect sentiment without human input.

To put it simply, machine learning allows computers to learn new tasks without being expressly programmed to perform them. Sentiment analysis models can be trained to read beyond mere definitions, to understand things like, context, sarcasm, and misapplied words.

For example:

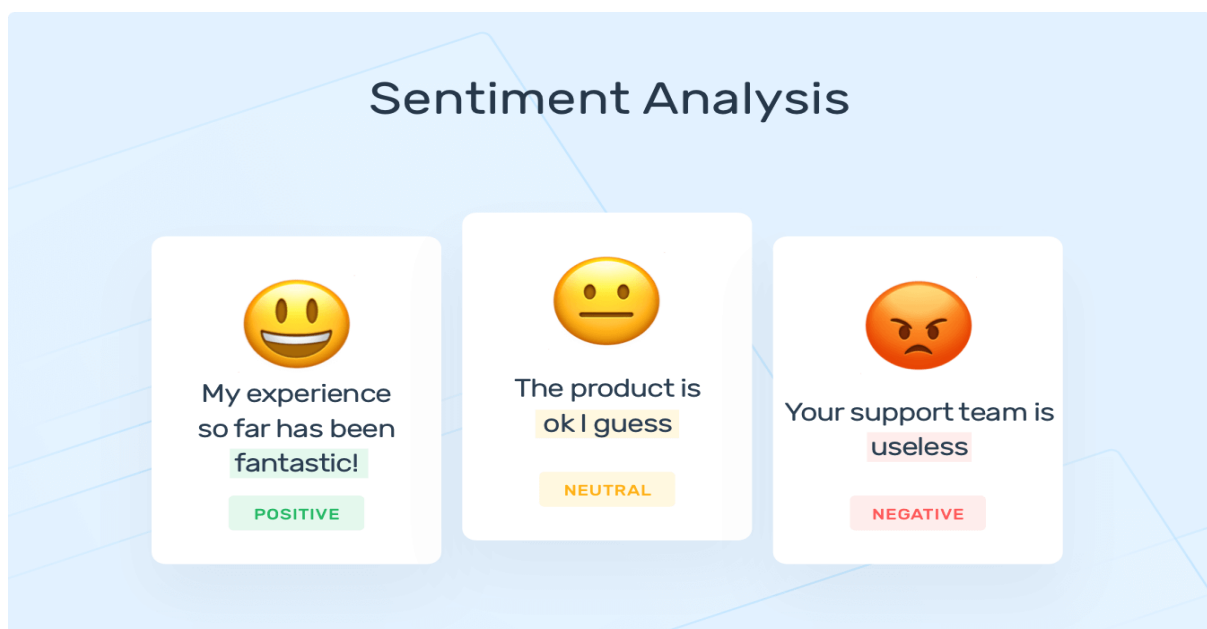
“Super user-friendly interface. Yeah right. An engineering degree would be helpful.”

Out of context, the words ‘super user-friendly’ and ‘helpful’ could be read as positive, but this is clearly a negative comment. Using sentiment analysis, computers can automatically process text data and understand it just as a human would, saving hundreds of employee hours.

Imagine using machine learning to process customer service tickets, categorize them in order of urgency, and automatically route them to the correct department or employee. Or, to analyze thousands of product reviews and social media posts to gauge brand sentiment.

Sentiment Analysis, or Opinion Mining, is a subfield of NLP (Natural Language Processing) that aims to extract attitudes, appraisals, opinions, and emotions from text. Inspired by the rapid migration of customer interactions to digital formats e.g., emails, chat rooms, social media posts, comments, reviews, and surveys, Sentiment Analysis has become an integral part of analytics organizations must perform to understand how they are positioned in the market. To be clear, Sentiment Analysis isn’t a novel concept. In fact, it has always been an important part of CRM (Customer Relationship Management) and Market Research — companies rely on knowing their customers better to evolve and innovate. The more recent rise is driven largely by the availability/accessibility of customer interaction records and well as improved computing capabilities to process these data. This advancement has really benefited consumers in meaningful ways. More than ever, organizations are listening to their constituents to improve. There are numerous approaches for Sentiment Analysis. In

this article, we'll explore three such approaches: 1) Naive Bayes, 2) Deep Learning LSTM, and 3) Pre-Trained Rule-Based VADER Models. We will focus on comparing simple out-of-the-box version of the models with the recognition that each approach can be tuned to improve performance. The intention is not to go into great details about how each methodology works but rather a conceptual study on how they compare to help determine when one should be preferred over another.

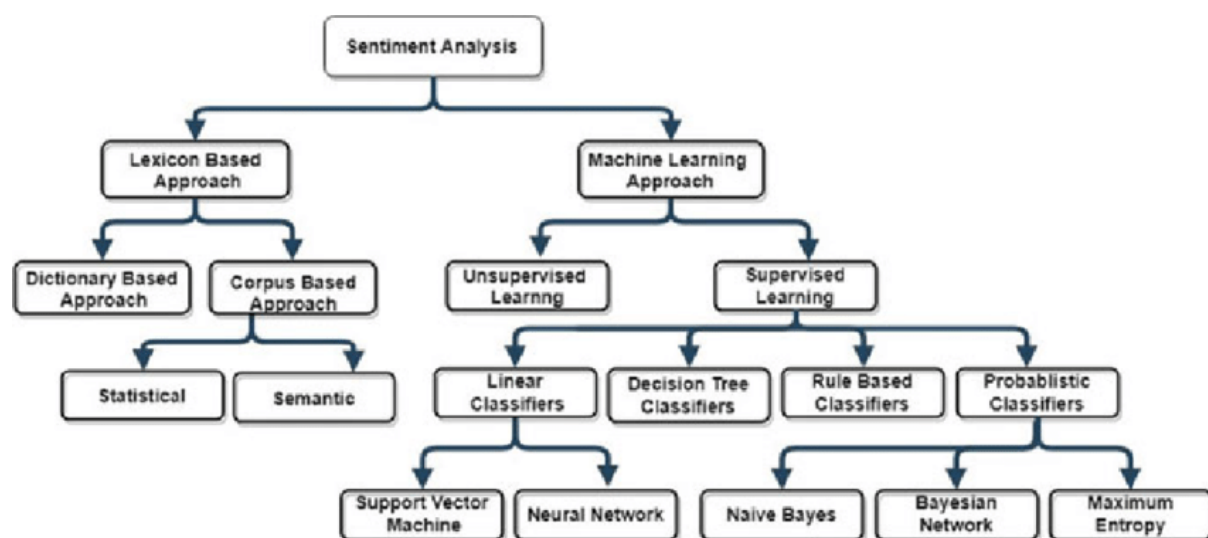


## Background on Sentiment Analysis

The objective of Sentiment Analysis ranges on the positive to negative spectrum. As with other NLP efforts, it is generally considered a classification problem, though it can be viewed as a regression problem when precision is important. Sentiment Analysis used to be accomplished by having a large labor force to read through and manually assess texts. This approach is costly and prone to human error. In an effort to automate this process, companies look to

advanced analytical methods for solving this problem. The challenge with Sentiment Analysis is that people express and interpret sentiment polarity and intensity differently. Furthermore, words and sentences can have multiple meanings based on the context (known as polysemy). While some of these issues can be mitigated, there is almost always a trade-off between speed and performance like any analytical tasks. We review three general methodologies, each with its own strengths and drawbacks:

### **There are two approaches to achieve Sentimental Analysis**



There are a number of techniques and complex algorithms used to command and train machines to perform sentiment analysis. There are pros and cons to each. But, used together, they can provide exceptional results. Below are some of the most used algorithms.

# Chapter-3

## DATA ANALYSIS

### 3.1 Data Description

The dataset I am using here for Flipkart reviews sentiment analysis is downloaded from Kaggle. Let's start this task by importing the necessary Python libraries and the dataset:

There are 2304 rows and 4 columns in this data. Here , given the first 5 rows of the dataset.

Unnamed: 0	Product_name	Review	Rating
0	0	Lenovo IdeaPad Gaming 3 Ryzen 5 Hexa Core 5600...	Best under 60k Great performanceI got it for a... 5
1	1	Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600...	Good perfomence... 5
2	2	Lenovo Ideapad Gaming 3 Ryzen 5 Hexa Core 5600...	Great performance but usually it has also that... 5
3	3	DELL Inspiron Athlon Dual Core 3050U - (4 GB/2...	My wife is so happy and best product 🙌👍 5
4	4	DELL Inspiron Athlon Dual Core 3050U - (4 GB/2...	Light weight laptop with new amazing features, 5

## 3.2 Representation of Data Analysis

The dataset consists of reviews of the products which customers purchased from Flipkart. The reviews given are the experience about the product customer bought through Flipkart and given a rating to that product.

The dataset for Flipkart contains 3 columns:

- Product\_name: the name of the product with little details about it.
- Review: review what the customer gives about that product.
- Rating: ranges from 1 to 5 stars given by the customer.

To represent and analyze the sentiment in Flipkart reviews, you can follow these steps:

### ▪ **Data Collection**

Collect the Flipkart reviews you want to analyze. You can use web scraping tools or APIs to gather the review data.

### ▪ **Text Preprocessing**

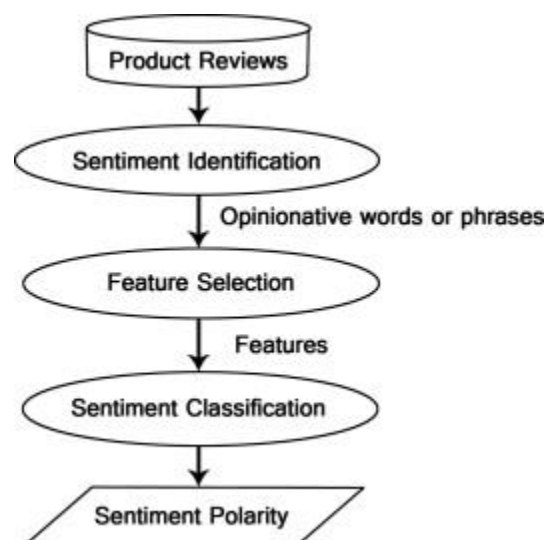
1. Converting words to lower/upper case
2. Removing special characters



3. Removing stopwords and high/low-frequency words
4. Stemming/lemmatization

## ▪ **Model Training**

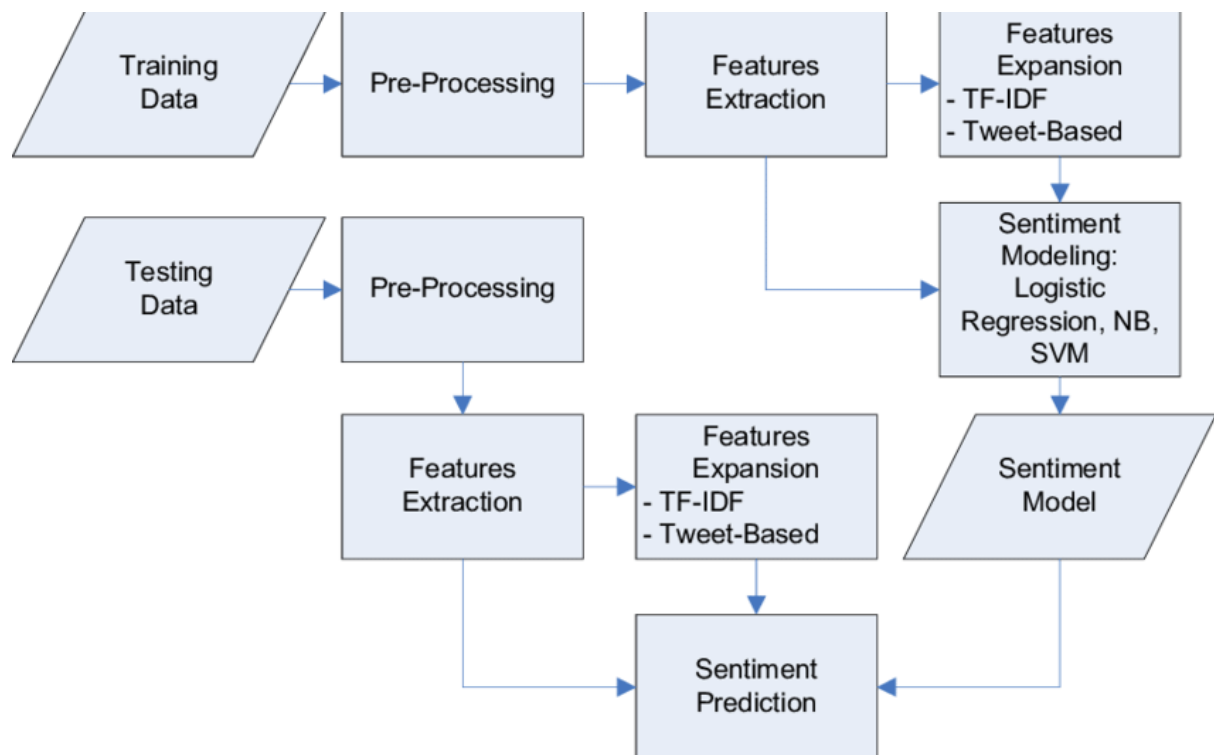
Fetching the datasets from the file and extracting all the corresponding words (feature words) like adjective, adverb and verb. Then datasets are labelled a respectively as “pos” for positive and “neg” for negative. Then performing frequency distribution over collected words and selecting 5000 words for training. Again, the shuffling of data is performed using random seed for better training. Here, the labeled datasets are divided into the percentile of 70-30% for training and testing, respectively. Training dataset to classification algorithms like Naïve Bayes classification algorithm.



## ▪ **Model Testing**

Here user can test and analysis the respective model by performing preprocessing over the input data. The preprocessing contains the removal of the symbol and number. Mapping to user input using saved

featured (based on training dataset). Then feed to saved model for prediction.



### 3.3 Data Cleaning

The dataset I am using here for Flipkart reviews sentiment analysis is downloaded from Kaggle. Let's start this task by importing the necessary Python libraries and the dataset:

The dataset contains only three columns. Let's have a look at whether any of these columns contains missing values or not:

```
print (data.isnull().sum())
```

So, the dataset does not have any null values. As this is the task of sentiment analysis of Flipkart reviews, I will clean and prepare

the column containing reviews before heading to sentiment analysis:

```
import nltk
```

```
import re
```

```
nltk.download('stopwords')
```

```
stemmer = nltk.SnowballStemmer("english")
```

```
from nltk.corpus import stopwords
```

```
import string
```

```
stopword=set (stopwords.words('english'))
```

```
def clean(text):
```

```
    text = str(text).lower ()
```

```
    text = re.sub ('\ [ . *? \]', "", text)
```

```
    text = re.sub ('https? ://\S+|www\.\S+', "", text)
```

```
    text = re.sub ('<. *?>+', "", text)
```

```
    text = re.sub('[%s]' % re.escape(string.punctuation), "", text)
```

```
    text = re.sub ('\n', "", text)
```

```
    text = re.sub ('\w*\d\w*', "", text)
```

```
    text = [word for word in text.split (' ') if word not in stopword]
```

```
    text="" “.join(text)
```

```
    text = [stemmer.stem(word) for word in text.split (' ')]
```

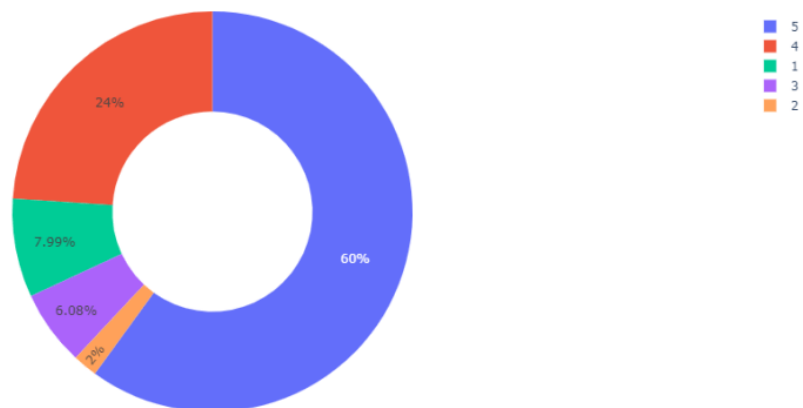
```
    text="" “.join(text)
```

```
    return text
```

```
data["Review"] = data["Review"]. apply(clean)
```

## 3.4 Data Visualization

The Rating column of the data contains the ratings given by every reviewer. So, let's have a look at how most of the people rate the products they buy from Flipkart:



So, 60% of the reviewers have given 5 out of 5 ratings to the products they buy from Flipkart. Now let's have a look at the kind of reviews people leave. For this, I will use a word cloud to visualize the most used words in the review's column:

Then the WordCloud visualization is given below.

Then the WordCloud visualization is given below:



### 3.5 Key Findings

Now let's see how most of the reviewers think about the products and services of Flipkart.

```
x = sum(data["Positive"])
```

```
y = sum(data["Negative"])
```

```
z = sum(data["Neutral"])
```

```
def sentiment_score (a, b, c):
```

if  $(a > b)$  and  $(a > c)$ :

```
print ("Positive 😊 ")
```

```
elif (b>a) and (b>c):  
    print ("Negative 😞 ")  
else:  
    print ("Neutral 😊 ")  
sentiment_score (x, y, z)
```

And now the finding is:

Neutral 😊

## Chapter-4

### CONCLUSION

Here, we offer a sentiment-based rating prediction and recommendation algorithm that may be used to predict product ratings based on user reviews. The purpose is to provide a feature-based impression of a large number of customer reviews of a product sold on the internet. To accomplish the rating prediction challenge, we combine sentiment similarity, interpersonal sentiment influence, and item reputation similarity into a single matrix factorization framework. We will examine complicated strategies for extracting opinion and product features, as well as novel classification models that can handle the organised names property in rating prediction and sentiment lexicons to apply fine-grained sentiment analysis in the future.

**Positive: 923.5529999999985**

**Negative: 96.77500000000013**

**Neutral: 1283.6880000000006**

So, most people give Neutral reviews, and a small proportion of people give Negative reviews. So, we can say that people are satisfied with Flipkart products and services. The Negative review score has a small proportion, and Positive reviews are midway. So, with the help of the above sentiment analysis implementation, we can conclude that most people are satisfied with the products and services they received from Flipkart.

This states that sentiment analysis is a good process to gather insights into each product, which ultimately benefit present and future customers and e-commerce companies. Sentiment analysis is important because, based on bad reviews, the

e-commerce company makes those products better or replaces those products with better and newer ones, which ultimately improves the overall customer service too.



# Chapter-5

## REFERENCE

- [1] Samaneh Moghaddam and Martin Ester, "Opinion Digger: An Unsupervised Opinion Miner from Unstructured Product Reviews", Proceedings of 19th ACM International Conference on Information and Knowledge Management, pp. 1825-1828, 2010.
- [2] Aurangzeb Khan, Baharum Baharudin and Khairullah Khan, "Sentiment Classification from Online Customer Reviews using Lexical Contextual Sentence Structure", Proceedings of International Conference on Software Engineering and Computer Systems, pp. 317-331, 2011
- [3] M. Hu and B. Liu, "Mining and Summarizing Customer Reviews", Proceedings of 10th ACM International Conference on Knowledge Discovery and Data Mining, pp. 166-177, 2005
- [4] A.M. Popescu and O. Etzioni, "Extracting Product Features and Opinions from Reviews", Proceedings of International Conference on Human Language Technology and Empirical Methods in Natural Language Processing, pp. 339-346, 2005
- [5] G. Vinodhini and R.M. Chandrasekaram, "Sentiment Analysis and opinion Mining: A Survey", International Journal of Advanced Research in Computer Science and Software Engineering, Vol. 2, No. 6, pp. 28-35, 2012.
- [6] A. Collomb, C. Costea, D. Joyeux, O. Hasan and L. Brunie, "A Study and Comparison of Sentiment Analysis Methods for Reputation Evaluation", Available at: <https://liris.cnrs.fr/Documents/Liris-6508.pdf>.

- [7] Mika V. Mantyla, Daniel Graziotin and Miikka Kuutila, "The Evolution of Sentiment Analysis-A Review of Research Topics", Computer Science Review, Vol. 27, No. 1, pp. 16-32, 2018.
- [8] F. Benamara, C. Cesarano and D. Reforgiato, "Sentiment Analysis: Adjectives and Adverbs are better than Adjectives Alone", Proceedings of International Conference on Weblogs and social media, pp. 1-7, 2006.
- [9] R.A. Hummel and S.W. Zucker, "On the Foundations of Relaxation Labeling Processes", Proceedings of International Conference on Computer Vision: Issues, Problems, Principles, and Paradigms, pp. 585-605, 1987.
- [10] Samaneh Moghaddam and Martin Ester, "ILDA: Interdependent LDA Model for Learning Latent Aspects and their Ratings from Online Product Reviews", Proceedings of 34<sup>th</sup> International ACM Conference on Research and Development in Information Retrieval, pp. 665- 674, 2011.
- [11] Jorge Carrillo De Albornoz, Laura Plaza, Pablo Gervas and Alberto Diaz, "A Joint Model of Feature Mining and Sentiment Analysis for Product Review Rating", Proceedings of International Conference on Advances in Information Retrieval, pp. 55-66, 2011.
- [12] Sentiment Analysis, Available at:  
<https://insightsatlas.com/sentiment-analysis>.
- [13] Andrew L. Maas, Raymond E. Daly, Peter T. Pham, Dan Huang, Andrew Y. Ng and Christopher Potts, "Learning Word Vectors for Sentiment Analysis", Proceedings of 49th Annual Meeting of the Association for Computational Linguistics, pp. 1-7, 2011.
- [14] Understanding Sentiment Analysis: What It Is and Why It's Used  
Understanding Sentiment Analysis: What It Is and Why It's Used,  
Available at: <https://www.brandwatch.com/blog/understanding-sentiment-analysis>.

**[15]** Sentiment Analysis Explained, Available at:  
<https://www.lexalytics.com/technology/sentiment-analysis>.