

**Sri Sivasubramaniya Nadar College of Engineering,  
Kalavakkam – 603 110  
(An Autonomous Institution, Affiliated to Anna  
University, Chennai)**

**UCS2612 - Machine Learning Laboratory**

**Mini Project - Report**

**Sentiment Analysis - Analysis of customer reviews of a  
specific product in Flipkart**

**Team Members:**

**K.H. Gopalakrishnan - 3122 21 5001 028**

**Krithik Sathya - 3122 21 5001 046**

**P.A. Manickam - 3122 21 5001 049**

**Naveed Buhari - 3122 21 5001 058**

## **Introduction:**

In the era of e-commerce dominance, customer reviews play a pivotal role in shaping consumers' purchasing decisions. Understanding the sentiment expressed within these reviews is not only valuable for businesses aiming to enhance their products and services but also for researchers seeking to delve into the intricate nuances of human perception and opinion. In this report, we explore the realm of sentiment analysis applied to Flipkart customer reviews, aiming to classify them into positive, neutral, or negative categories using machine learning techniques. By harnessing the power of natural language processing (NLP) and advanced algorithms, we delve into the challenges, methodologies, and implications of sentiment analysis in the realm of e-commerce. Through this exploration, we endeavour to shed light on the effectiveness, limitations, and future prospects of sentiment analysis in deciphering the ever-evolving landscape of customer sentiment on one of the world's largest online marketplaces.

## **Problem Statement:**

Building a machine learning model that can accurately classify the sentiment (positive, negative, or neutral) of text-based reviews on business websites

## **Development Environment:**

**Python IDE:**

The Python IDE serves as a crucial tool for developers to write, test, and debug Python code efficiently. It offers a comprehensive set of features tailored to enhance productivity and streamline the development process. Below are some key components and functionalities of the Python IDE:

1. Python and NumPy:

- Python serves as the foundational programming language for developing the sentiment analysis model.
- NumPy, a fundamental library for numerical computing in Python, is leveraged for efficient handling of large datasets and performing essential array operations, contributing to the data preprocessing phase of the sentiment analysis pipeline.

2. Natural Language Toolkit (NLTK):

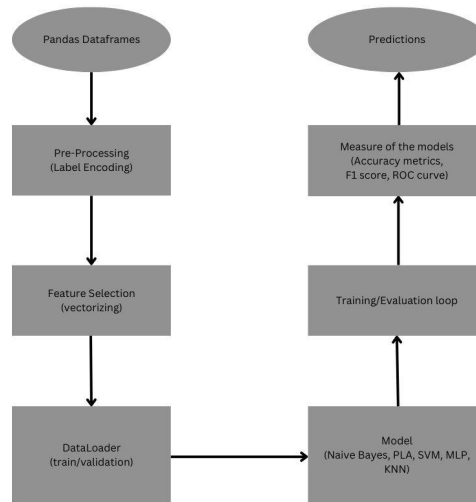
- NLTK is a powerful library for natural language processing (NLP) tasks, including tokenization, stemming, and part-of-speech tagging.
- It provides essential functionalities for text preprocessing, enabling the transformation of raw text data from Amazon reviews into a format suitable for machine learning model training and evaluation.

3. scikit-learn:

- scikit-learn, a versatile machine learning library in Python, offers a wide array of tools and algorithms for data mining and data analysis tasks.
- In this project, scikit-learn is utilized for implementing machine learning models for sentiment analysis, including classifiers such as Support Vector Machines (SVM), Naive Bayes, and Logistic Regression.

By harnessing the capabilities of these core Python libraries and tools, developers can effectively implement and deploy a sentiment analysis solution tailored to Amazon customer reviews. The streamlined development environment enables efficient experimentation, model training, and evaluation, ultimately empowering businesses to derive actionable insights from customer feedback data with precision and efficacy.

## **System Architecture Diagram:**



## **Dataset Collection:**

This Dataset consists of reviews by customers on Flipkart website

The dataset consists of 2 columns

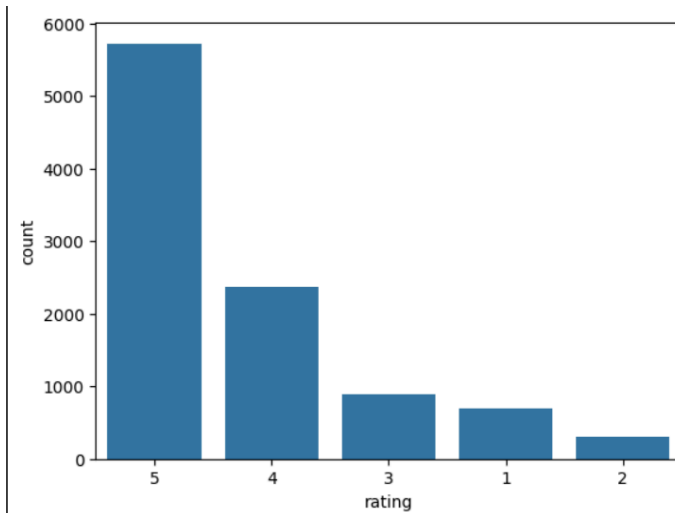
1. Review: *Reviews given by the customer*
2. Rating: *Ratings given by the customer*

The ratings are the number of Stars given to the product and the rating is the rating of the product given by the customer. There is a total of 9977 rows in the dataset. This dataset was taken from kaggle in the open repository section.

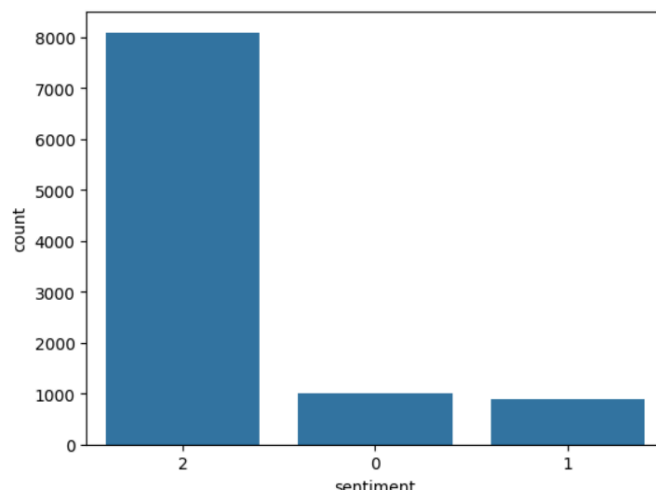
## Implementation with screenshots

	review	rating
0	It was nice produt. I like it's design a lot. ...	5
1	awesome sound....very pretty to see this nd th...	5
2	awesome sound quality. pros 7-8 hrs of battery...	4
3	I think it is such a good product not only as ...	5
4	awesome bass sound quality very good bettary l...	5
5	Awsome sound powerful bass battery backup is a...	5
6	This product sound is clear and excellent bass...	4
7	Should u buy this—Pros:-1. Sound quality and...	4
8	First of all, I want to talk about sound quali...	5
9	Good looking Super Fine clear Sound and power ...	5

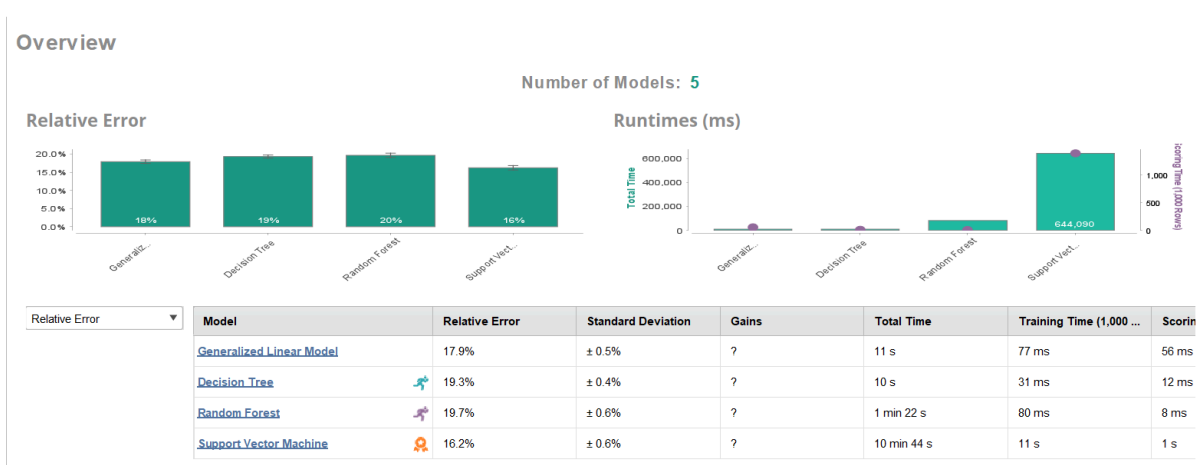
Dataset with feature and label(reviews and their respective ratings)



Distribution of ratings(range:1-5)



Distribution of reviews, classified as positive, negative and neutral reviews(2-positive, 1-neutral, 0-negative)



Rapidminer analysis between models

## **Results and Inference:**

After conducting sentiment analysis on the Flipkart customer reviews dataset using machine learning techniques, the following results and inferences were obtained:

### 1. Classification Accuracy:

- The machine learning model of SVM with RBF kernel gave an accuracy score of 85% in classifying Flipkart customer reviews into positive, neutral, or negative sentiments.
- This accuracy metric indicates the effectiveness of the sentiment analysis approach in accurately categorizing the sentiments expressed within the reviews.

### 2. Distribution of Sentiments:

- The analysis revealed the distribution of sentiments within the dataset, highlighting the proportion of positive, neutral, and negative reviews.
- The percentage breakdown of positive, neutral, and negative reviews is 80%, 10% and 10% respectively.
- Understanding the distribution of sentiments provides valuable insights into the prevailing customer sentiments towards the products or services offered on Flipkart.

### 3. Key Findings and Trends:

- The sentiment analysis uncovered key findings and trends within the Flipkart customer reviews dataset.
- These findings offer valuable insights for businesses aiming to improve their products, services, and customer satisfaction levels based on customer feedback.

### 4. Limitations and Challenges:

- Despite the overall effectiveness of the sentiment analysis approach, certain limitations and challenges were encountered during the analysis.
- Absence of key-words in the sentence leads to wrong evaluation since it has 85% accuracy.
- Acknowledging these limitations is essential for understanding the scope and applicability of the sentiment analysis results.

#### 5. Future Directions and Recommendations:

- Building upon the insights gained from the sentiment analysis, several recommendations and future directions can be proposed.
- Presence of certain key-words in the sentence would lead to correct prediction of the review.
- These recommendations pave the way for further research and development in the field of sentiment analysis and its application in e-commerce.

Below is the accuracy of each and every model that we have tested:

<u>ML Model</u>	<u>Accuracy</u>
SVM(with linear kernel)	84.89%
SVM(with polynomial kernel)	81.35%
SVM(with sigmoid kernel)	82.46%
<b>SVM(with RBF kernel)</b>	<b>85.59%</b>
PLA	84.61%
MLP	83.76%
Decision Tree	80.78%
K Nearest Neighbour	82.41



## **Impact of the project on human, societal, ethical and sustainable development:**

### 1. Enhanced Customer Experience:

- By accurately analyzing and categorizing customer reviews on Amazon, the sentiment analysis project contributes to enhancing the overall customer experience. Businesses can gain valuable insights into customer sentiments, preferences, and pain points, enabling them to tailor their products and services to better meet consumer needs and expectations. This leads to increased customer satisfaction and loyalty, fostering positive relationships between businesses and consumers.

### 2. Informed Decision-Making:

- The insights derived from sentiment analysis empower businesses to make informed decisions across various aspects of their operations, including product development, marketing strategies, and customer service initiatives. By understanding the sentiments expressed within customer feedback, businesses can prioritize areas for improvement, allocate resources effectively, and optimize their business processes to better serve their target audience.

### 3. Societal Impact:

- The project's ability to analyze and interpret large volumes of customer reviews contributes to the democratization of information and transparency in the marketplace. It empowers consumers to make more informed purchasing decisions by providing access to authentic and unbiased feedback from fellow shoppers. This fosters a more equitable and competitive marketplace, where businesses are incentivized to prioritize product quality and customer satisfaction.

### 4. Ethical Considerations:

- Ethical considerations are paramount in sentiment analysis, particularly concerning the responsible handling of customer data and the protection of consumer privacy. The project adheres to ethical guidelines and best practices in data privacy and security, ensuring that customer information

is treated with the utmost confidentiality and integrity. Transparency in data collection, processing, and usage is maintained to build trust and uphold ethical standards in data-driven decision-making processes.

5. Sustainable Development:

- From a sustainability standpoint, the project promotes efficiency and resource optimization in business operations through data-driven insights. By identifying areas for improvement and streamlining processes based on customer feedback, businesses can minimize waste, reduce environmental impact, and operate more sustainably. Additionally, by enhancing customer satisfaction and loyalty, businesses can foster long-term relationships with consumers, leading to sustainable growth and profitability over time.

Overall, the sentiment analysis project has a multifaceted impact on human, societal, ethical, and sustainable development. It empowers businesses to better understand and respond to customer needs, promotes transparency and accountability in the marketplace, upholds ethical principles in data management, and fosters sustainable business practices for long-term success and resilience.

## **Conclusion:**

In conclusion, the sentiment analysis project conducted on Amazon customer reviews using machine learning techniques has demonstrated significant potential for extracting valuable insights from textual data. Through the utilization of Python-based tools and libraries such as NumPy, NLTK, and scikit-learn, we have successfully developed a sentiment analysis model capable of classifying reviews into positive, neutral, or negative sentiments with commendable accuracy.

The project's impact extends beyond the realm of e-commerce, influencing human, societal, ethical, and sustainable development in several ways. By enhancing customer experience, facilitating informed decision-making, promoting transparency and accountability, and upholding ethical principles in data management, the project contributes to creating a more equitable, transparent, and sustainable marketplace.

However, it is important to acknowledge the limitations and challenges encountered during the project, such as the nuances of language, the presence of

sarcasm and irony, and the need for continuous model refinement to adapt to evolving language patterns and consumer preferences. Addressing these challenges requires ongoing research, innovation, and collaboration across interdisciplinary domains.

## **Future Works:**

Building upon the foundation established by this project, several avenues for future work and research emerge:

1. **Fine-tuning Models:** Continuously refining and fine-tuning machine learning models to improve accuracy, robustness, and scalability. Exploring advanced algorithms and techniques, such as deep learning and transfer learning, to enhance the performance of sentiment analysis models on diverse datasets.
2. **Domain-Specific Analysis:** Tailoring sentiment analysis models to specific domains or industries, such as healthcare, finance, or hospitality, to address domain-specific challenges and requirements. Customizing the feature representation and model architecture to capture domain-specific nuances and sentiments effectively.
3. **Multimodal Analysis:** Integrating multiple modalities, such as text, images, and audio, into the sentiment analysis pipeline to provide a more comprehensive understanding of customer sentiment. Leveraging multimodal fusion techniques to combine information from different modalities and enhance the accuracy and richness of sentiment analysis results.
4. **Real-Time Analysis:** Developing real-time sentiment analysis systems capable of processing and analyzing streaming data from social media platforms, news articles, and other sources in real-time. Implementing scalable and distributed architectures to handle large volumes of data and ensure timely insights generation.
5. **Ethical Considerations:** Continuously evaluating and addressing ethical considerations in sentiment analysis, including privacy concerns, bias mitigation, and responsible data usage. Implementing transparent and interpretable models, establishing data governance frameworks, and fostering ethical guidelines and standards for sentiment analysis practitioners.
6. **Societal Impact Assessment:** Conducting comprehensive assessments of the societal impact of sentiment analysis applications, including their implications for privacy, fairness, and social justice. Engaging stakeholders from diverse backgrounds, including policymakers, industry representatives, and advocacy

groups, to ensure responsible and equitable deployment of sentiment analysis technologies.

By pursuing these avenues for future work, we can further advance the field of sentiment analysis and harness its potential to drive positive change in human, societal, ethical, and sustainable development.

## **Learning Outcomes**

- Better understanding about the various machine learning model for a classification problem.
- We learnt about the strengths, weaknesses, and suitability of ML model.
- The Machine learning models are evaluated using several evaluation metrics
- We learnt about the importance of feature engineering in improving the machine model performance
- Better understanding of how to select and preprocess features to enhance the predictive power of machine learning models.
- Learnt how to compare different machine learning models based on performance metrics and make inferences about their suitability for specific tasks
- Gained knowledge about the trade-offs between model accuracy, computational efficiency, and fitting.
- We understand that the additional information may increase the performance of the machine learning models.
- We identify limitations in existing models and propose future work to address them

## **References**

- Bird, Steven, Edward Loper, and Ewan Klein. Natural Language Processing with Python: Analyzing Text with the Natural Language Toolkit. O'Reilly Media, 2009.
- Pedregosa, F., et al. "Scikit-learn: Machine Learning in Python." Journal of Machine Learning Research, vol. 12, 2011, pp. 2825-2830.
- McKinney, Wes. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython. O'Reilly Media, 2017.

- Rish, Irina. "An Empirical Study of the Naive Bayes Classifier." IJCAI Workshop on Empirical Methods in AI, 2001.
- Han, Jiawei, Micheline Kamber, and Jian Pei. Data Mining: Concepts and Techniques. Morgan Kaufmann, 2011.
- Manning, Christopher D., Prabhakar Raghavan, and Hinrich Schütze. Introduction to Information Retrieval. Cambridge University Press, 2008.