INTEGRATING NLP-BASED SENTIMENT ANALYSIS INTO SUPPLY CHAIN MANAGEMENT: A CASE STUDY OF INDIAN PRODUCTS ON AMAZON

A thesis submitted

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

MASTER OF TECHNOLOGY
IN
PRODUCTION ENGINEERING

Submitted by
KUSHAGRA VATS
2K19/PIE/05

Under the guidance of

Dr. N. YUVARAJ

DEPARTMENT OF MECHANICAL ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY (DTU)
BAWANA ROAD DELHI (INDIA)-110042

MAY, 2023

CERTIFICATE

This is to certify that the report entitled "**Integrating NLP-based Sentiment Analysis into Supply Chain Management: A Case Study of Indian Products on Amazon"** for the award of the degree **of Master of Technology** in **Production (PE)** is a record of bona fide work carried out by him under the supervision and guidance of undersigned. The thesis, in our opinion, is worthy of consideration for the award of a degree in accordance with Institute regulations.

This thesis's outcome has not been submitted to any other university or institute for the purpose of receiving a degree.

Forwarded by
Dr. N. YUVARAJ
Department of Mechanical Engineering
DELHI TECHNOLOGICAL
UNIVERSITY (DTU),
BAWANA ROAD DELHI-110042, INDIA.

DECLARATION

I, KUSHAGRA VATS, hereby certify that the work which is being presented in this thesis entitled "Integrating NLP-based Sentiment Analysis into Supply Chain Management: A Case Study of Indian Products on Amazon" the partial fulfillment of requirement for the award of degree of Master of Technology in Production Engineering submitted in the Department of Mechanical Engineering, Delhi Technological University, Delhi is an authentic record of my own work carried out during a period from July 2020 to June 2021, under the supervision of Dr. N. YUVARAJ, Department of Mechanical Engineering, Delhi Technological University, Delhi.

The matter presented in this thesis has not been submitted in any other University/Institute for the award of M. Tech Degree.

KUSHAGRA VATS
M. Tech. (Production Engineering)
Department of Mechanical Engineering
Roll No. 2k19/PIE /05

(Forwarded by)

Dr. N. YUVARAJ
Department of Mechanical Engineering
Delhi Technological University, Delhi

APPROVAL SHEET

This project report titled "**Integrating NLP-based Sentiment Analysis into Supply Chain Management: A Case Study of Indian Products on Amazon**" by **Kushagra Vats (Roll No: 2K19/PIE /05)** has been approved for the degree of M. Tech in Mechanical Engineering (Production Engineering) from Delhi Technological University, Delhi.

Examiners

____

____

____

Supervisor

____

____

____

Head of Department

____

Date: __________________

Place: __________________

ACKNOWLEDGEMENTS

It gives me great pleasure to deliver my dissertation paper, "**Integrating NLP-based Sentiment Analysis into Supply Chain Management: A Case Study of Indian Products on Amazon**". First and foremost, I owe a debt of gratitude to my mentor, Dr. N. YUVARAJ of the Mechanical Engineering Department, for his professional advice and unwavering encouragement throughout the whole process of the thesis. I consider myself fortunate to have the opportunity to work with them. It was highly thought stimulating not only to grasp the subject, but also to analyze the results obtained from the graphs. I am grateful for their compassion and generosity towards me, since it aided me in morally completing the assignment before beginning it.

I'd like to express my appreciation to Prof. S. K. Garg, Head, Mechanical Engineering Department, to allow me to complete this project.

Finally, and most importantly, I want to express my gratitude to my family members for their support, encouragement, and prayers during the past several months. They are the ones to whom I dedicate my work.

CONTENTS

List of figures and tables

ABSTRACT

A critical component of modern business operations is Supply chain management (SCM). It involves the effective coordination of activities from the procurement of raw materials to the delivery of finished products to customers. With the rapid technology advancements, the integration of Machine Learning (ML) and Artificial Intelligence (AI) techniques has emerged as a powerful tool in enhancing supply chain management practices. One particular area that has gained significant attention in recent years is sentiment analysis using Natural Language Processing (NLP).

Sentiment analysis (SA) or Opinion Mining (OM) is a special branch of NLP that focuses on extracting subjective information, sentiments and opinions from text data, such as customer reviews. By analyzing the sentiment expressed in these reviews, businesses can gain valuable insights into customer preferences, pain points and satisfaction levels. Applying sentiment analysis techniques to supply chain management allows organizations to better understand customer feedback and reviews, leading to informed decision-making and improved overall performance.

The goal of this study is to look into how NLP-based sentiment analysis could be used to enhance supply chain management, specifically through utilizing consumer feedback for Indian items. The study's goal is to examine the use of sentiment analysis to learn more about how customers feel about various elements of the supply chain, including product availability, delivery time, packaging, and customer service. The study attempts to identify patterns, trends, and customer feelings by analyzing a dataset of over 1000 customer reviews gathered from various categories, including clothing, hair and skin care goods, and technological devices on the Amazon platform.

The research's goal in conducting this case study is to draw attention to the useful uses of sentiment analysis in supply chain management. The study's results will help us comprehend sentiment analysis as a decision-making tool better and how it can be utilized to enhance different supply chain components. By giving businesses insights into client preferences and empowering them to make data-driven decisions to improve product offerings, inventory management, logistics optimisation, and overall customer pleasure, this research has the potential to be helpful to businesses.

We used a theoretical framework that is based on supply chain management, sentiment analysis, and NLP literature that has already been published. For the purpose of our study, we analyzed a corpus of research publications on supply chain management, sentiment analysis, and related subjects. Our study strategy included exploratory data analysis (EDA), the use of the VADER and RoBERTa models for sentiment analysis, and a qualitative analysis of the outcomes to pinpoint the most important conclusions and their ramifications.

The results imply that NLP-based sentiment analysis can offer useful supply chain management insights. Our research of the Amazon dataset showed that sentiment analysis may pinpoint a product's advantages and disadvantages, point out potential areas for development, and reveal consumer preferences and expectations. Additionally, the study found that sentiment analysis can offer helpful data for a variety of supply chain management functions, such as product design, manufacturing, inventory control, and customer service. By illustrating the importance of sentiment analysis in enhancing decision-making in several supply chain management sectors, the research adds to the body of knowledge on NLP-based sentiment analysis and supply chain management. The paper also emphasizes the need for additional investigation to examine sentiment analysis's potential in other supply chain management domains, such as supplier management and logistics.

The study reveals how sentiment analysis can help with decision-making and customer happiness, which has applications for supply chain managers. Additionally, our study offers a methodology that may be used in various contexts to undertake sentiment analysis in supply chain management. The study also emphasizes how critical it is to incorporate NLP-based sentiment analysis into supply chain management systems in order to monitor customer comments and reviews in real-time and create supply chains that are more responsive and centered on the needs of their customers.

As a result, our research shows how NLP-based sentiment analysis may enhance supply chain management. The significance of sentiment analysis in determining consumer preferences and expectations, enhancing product design and quality, and improving inventory management is highlighted in our work as a contribution to the literature. Because it offers a methodology for doing sentiment analysis and emphasizes the value of incorporating NLP-based sentiment analysis into supply chain management systems, our research also has applications for supply chain managers.

Keywords: NLP, EDA, SCM, RoBERTa, VADER, DL, ML, Sentiment Analysis, Opinion Mining

Introduction

Effective supply chain management has evolved as a critical component for businesses to achieve a competitive edge, maximize operational efficiency, and exceed customer expectations in today's fiercely competitive business environment. The importance of supply chain management has grown as businesses try to improve the efficiency of their operations and the delivery of goods and services. This thesis looks at the importance of supply chain management in the modern corporate environment and how supply chain analytics may be used to improve processes and boost overall company performance.

This study uses case studies from a variety of businesses (depending on the viewpoint of the consumer) to highlight the practical ramifications of supply chain management. These organizations include Ola, Sprint Mobile, an ed tech company, Amazon, and a service-based company. These businesses demonstrate the applicability and influence of supply chain management across several industries by representing a variety of sectors.

We may learn more about how businesses in many industries have effectively used supply chain management tactics to obtain a competitive advantage by looking at these case studies. For instance, Ola, a well-known supplier of ride-hailing services, has improved the efficiency of its supply chain management to guarantee quick and dependable transportation services. Sprint Mobile, a US-based telecom business that has since merged with T-Mobile, has successfully managed its supply chain to keep up with changing consumer needs and market trends. Supply chain management techniques have been used by an ed-tech company to effectively deliver online educational services. The supply chain processes have been revolutionized by e-commerce behemoth Amazon, enabling seamless product delivery and consumer satisfaction.

In this study, we explore the field of supply chain analytics, highlighting its importance for fostering advancements and supporting data-driven decision-making. Businesses can learn important lessons about demand forecasting, inventory management, logistics optimisation, and supplier performance monitoring by utilizing the power of analytics. Through the analysis of these case studies, we want to show how supply chain analytics may have a revolutionary effect on operational processes, reduce risks, and promote overall business performance through the analysis of these case studies.

This thesis aims to offer insightful information and practical consequences for businesses looking to enhance their supply chain management practices by examining the significance of supply chain management in various industries and the function of analytics in increasing supply chain operations. The results of this study can help businesses use supply chain analytics to make smart decisions, increase productivity, and gain a competitive edge in the complicated business environment of today.

It is impossible to exaggerate the value of supply chain analytics in enhancing international supply networks. Supply chains are significantly relied upon by businesses across a range of industries to provide clients with necessities like goods and services. However, issues like product shortages, delayed shipments, and defects can seriously harm customers' pleasure and even endanger lives, particularly when it comes to life-saving medications. As a result, there has never been a more pressing demand for qualified supply chain analytics professionals.

Supply chain analytics analyze and improve supply chain operations using statistical tools and techniques. Organizations can find chances for improvement and reduce possible risks in intricate global supply chains by leveraging the power of data and automation. Companies continuously attempt to improve their supply chain processes and realize greater time and cost savings by putting a strong emphasis on efficiency and resource optimisation.

Through analytics, we can recognise the crucial role that automation and data-driven decision-making play in solving the difficulties and complexity of supply chain management. We seek to demonstrate the transformative influence of supply chain analytics on operational efficiency, risk management, and overall supply chain performance by delving into real-world examples and case studies.

The report also highlights the significance of building expertise in supply chain analytics to suit the changing needs of the current business landscape and acknowledges the growing demand for qualified individuals in this discipline.

The main goal of supply chain management is to satisfy consumers by providing the appropriate goods at the appropriate time and location. However, supply chains must work to reduce waste and expenses related to materials, labour, and fuel in order to live up to investors' expectations. Businesses that successfully strike a balance between cost reduction and customer satisfaction see increases in revenues and profitability. However, establishing this delicate balance is difficult because every client transaction creates the possibility of mistakes like stockouts, delayed delivery, or defective goods.

When mistakes are made, supply chain managers frequently have to make crucial decisions, which creates uncertainty and raises questions regarding supplier selection, distribution centers, workforce capacity, and inventory management. In these circumstances, managers could decide hastily, leading to additional mistakes and monetary losses. Supply chain analytics can be useful in this situation. Supply chain analytics refers to a collection of instruments and techniques that make use of supply chain data to comprehend the past, pinpoint the causes, and foresee the future. Companies can establish effective strategies and make well-informed judgements based on data analysis rather than hunches.

The supply chain's wealth of data does, however, bring problems. The analytics team might not always have access to or be able to trust all data. When working with enormous amounts of data, prioritizing which issues to resolve or prevent becomes essential. Data management and prioritization are therefore necessary for successful supply chain analytics in order to concentrate on the most important problems and areas for development.

Thus, it is essential to understand the function and significance of applying AI/ML to supply chain analytics in order to satisfy the dual goals of customer pleasure and cost optimization. The study intends to show how supply chain analytics technologies assist businesses in understanding and mitigating errors, resulting in enhanced decision-making and performance. This research offers insights into the complications and potential solutions in efficiently utilizing supply chain analytics by looking at the issues related to data availability, dependability, and prioritization.

Finally, by emphasizing the importance of data-driven decision-making and the function of analytics in improving supply chain performance, this research adds to the increasing body of knowledge in supply chain management. The results of this study can assist managers and practitioners in establishing methods that improve supply chain efficiency overall, lower costs, and increase customer satisfaction.

According to the advice given in Google's machine learning handbook, it is suggested that whenever possible, one should choose to construct a simple rule-based system instead of using machine learning methods. In some circumstances, this strategy can be quite effective since it enables the development of precise and succinct rules that lead to the desired results. Without the use of complicated machine learning methods, the system can do jobs with precision and efficiency by utilizing predetermined rules and patterns.

For instance, the deployment of a rule-based system in the context of sentiment analysis for supply chain management can entail the creation of specific rules based on words, linguistic patterns, or phrases that are connected to good, negative, or neutral attitudes. However, the rapidly increasing volume of data poses a dilemma from a strategic standpoint.
Given the massive amount of data, it may not be feasible to rely entirely on a rule-based system for sentiment analysis in the long run. To support this claim, the following facts and evidence are provided:

Volume of data generated is rising dramatically as a result of the widespread use of digital platforms, social media, and e-commerce. By 2025, the amount of data generated globally is anticipated to surpass 175 zettabytes, according to Statista. It is impractical to manually construct and maintain sentiment analysis rules due to the vast amount of data.

- Data complexity and variety: Data from numerous sources, including social media, online reviews, and customer feedback, are available in a wide range of languages and forms. It becomes difficult and time-consuming to create rules that account for every conceivable event and complexity. By identifying patterns and drawing generalizations from data, machine learning models may manage such complexity.

- Language and expressions are always changing; new words, expressions, and slang are constantly being created. It is challenging to keep a comprehensive and current rule set current in light of these developments. On the other hand, machine learning models can adjust to changing language patterns by continuously learning from fresh data.
- Scalability and effectiveness: When working with huge datasets, rule-based systems may become computationally expensive and inefficient. The time and resources needed to handle and analyze the data using rule-based systems may become unworkable as data quantities increase. Machine learning algorithms have the advantage of scalability and efficiency since they can process data in parallel.
- Finding hidden patterns and insights: Rule-based systems can only take into account the rules that are openly stated, which may leave out more complex or subtle patterns in the data. Deeper insights and more precise predictions can be gained by utilizing hidden patterns and connections that may not be obvious through the use of predetermined criteria.

These facts demonstrate the drawbacks of using a rule-based system alone for sentiment analysis in the face of rising data complexity, volumes, as well as the requirement for scalability and adaptability. In the dynamic field of supply chain management, integrating machine learning or deep learning techniques can address these issues and provide stronger and more practical solutions for sentiment analysis.

Several models and methods have been created to perform accurate sentiment analysis, two of which are VADER (Valence Aware Dictionary and Sentiment Reasoner) and RoBERTa (Robustly Optimised BERT Pretraining Approach). These models have proven to be successful at capturing sentiment data, while using various methodologies and underlying architectural frameworks.

In 2014, Hutto and Gilbert [56] unveiled VADER, a rule-based sentiment analysis tool created exclusively for social media texts. To establish sentiment polarity, it combines lexical-based heuristics with a pre-built sentiment lexicon. In order to determine the sentiment intensity score for a particular text fragment, VADER assigns sentiment scores to individual words and then combines them. Intensifiers, negations, and contextual valence shifting are just a few of the variables that are taken into account by VADER, which offers a complex study of sentiment that goes beyond simple positive or negative categorization. VADER is a well-liked option for sentiment analysis in social media and short text domains due to its capacity to handle domain-specific terminology and comprehend sentiment in informal communications.

In contrast, RoBERTa, which Liu et al. [55]. unveiled in 2019, is a development in the area of contextualised language representation models. Based on the well-known BERT architecture (Bidirectional Encoder Representations from Transformers), RoBERTa improves pretraining by making adjustments and adding more training data. It uses a neural network built on transformers to learn contextualised word representations that capture the intricate semantic and syntactic details of a given text. RoBERTa achieves state-of-the-art performance on a variety of NLP tasks, including sentiment analysis, by utilising a large-scale corpus and a masked language modelling aim during pretraining. RoBERTa is able to grasp fine-grained sentiment nuances and achieve high accuracy in sentiment classification tasks thanks to its comprehension of sentence-level context and the complex relationships between words.

With a focus on accuracy, robustness, and flexibility, we compare and contrast the performance of VADER and RoBERTa in sentiment analysis tasks in this thesis. We want to provide light on the advantages, constraints, and potential uses of these models in various scenarios by evaluating them on benchmark datasets and carrying out actual tests.

Literature Review:

From the sourcing of raw materials to the delivery of items to customers, supply chain management is the strategic coordination and integration of all activities involved in the movement of commodities, services, information, and funds. It includes operations including purchasing, manufacturing, inventory control, shipping, warehousing, and customer support. Managing inventory levels to prevent stockouts or surpluses, coordinating suppliers to assure prompt delivery of goods, optimizing manufacturing schedules to meet demand, and organizing logistics to simplify distribution are a few examples of supply chain management. In today's complicated global marketplace, effective supply chain management provides cost reduction, increased productivity, improved customer satisfaction, and competitive advantage.
Therefore, it is impeccable to incorporate these crucial details to ace the management.

1) Importance of having supply chain goals: Setting up clear targets is essential before putting a supply chain analytics programme into action. These objectives might be based on current key performance indicators (KPIs) for the supply chain, such as inventory levels, quality evaluations, worker productivity, and on-time delivery.

2) Mapping stakeholders and identifying their needs: Understanding the needs of all involved parties, including suppliers, manufacturers, transportation firms, and customers, is crucial to creating better KPIs and enhancing the supply chain. Key success elements can be found by looking at their expectations and the supply chain's resources and outputs.

3) Supply chain analytics for risk mitigation: Analytics of the supply chain are essential for risk mitigation. It assists in keeping an eye on and foreseeing unforeseen hiccups like earthquakes, supplier failures, or unexpected rises in consumer demand. Organizations can create backup plans and automate decision-making procedures by employing analytics.

4) Cost-benefit analysis of supply chain analytics: A supply chain analytics program's implementation entails expenses for data collection, tool development, and human resources. It is crucial to think about whether the benefits of reaching the goals will outweigh these

expenses. A thorough cost-benefit analysis must be performed to guarantee a profitable return on investment.

The findings of Raut et al. [12] indicate that the biggest impediments in application of Data analytics in supply chain management of manufacturing companies are four constructs: a lack of top management support, a lack of financial support, a lack of skills, and a lack of processes or procedures.

S. N	**Barriers**	**12**	**11**	**10**	**9**	**8**	**7**	**6**	**5**	**4**	**3**	**2**
1	Poor data quality and lack of trust in data	A	A	A	A	O	V	O	O	X	A	V
2	Time-consuming activity	A	A	A	A	O	V	V	A	O	A	
3	Lack of sufficient resources	V	O	O	O	A	O	O	A	V		
4	Lack of security and privacy	A	A	O	A	O	V	V	O			
5	Lack of financial support	O	O	O	O	X	X	V				
6	Behavioural issues	V	O	O	V	A	A					
7	Return on investment (ROI) issues	A	O	O	A	A						
8	Lack of top management support	O	O	O	V							
9	Lack of skills	V	X	V								
10	Data scalability	V	O									
11	Lack of techniques or procedures	V										
12	Lack of data integration and management	---										

Fig. 1: Structural self-interaction of barriers [12]

The data is taken from experts. Three data analysts, two big data visualizers, two big data solution architects, three big data engineers, three big data researchers, two data warehouse managers, two data architects, two database managers, three business intelligence analysts, two data warehouse analysts, two data modelers, two database developers, two business system analysts, and two data miners made up the expert team, which was made up of 47 people. Here:

V- “Barrier i leads to barrier j.”

A- “Barrier j leads to barrier i.”

X- “Barrier i leads to barrier j and vice versa.”
O- “No relationship between the barriers.”

Traditional supply chain management tools:

Traditional SC analytics are categorized as predictive, prescriptive, descriptive, and diagnostic analytics. While diagnostic analytics aids in understanding why specific events occurred, descriptive analytics offers insights into what happened in the past. Prescriptive analytics prescribes actions based on complex computations, whereas predictive analytics uses models and data to predict future results.
Information like the percentage of late shipments or a rise in customer complaints can be found using descriptive analytics. On the other hand, diagnostic analytics assists in figuring out the causes of problems by taking into account elements like increased orders, maintenance needs, or staff availability.

Predictive analytics can foresee future events by analyzing trends and data, such as sales growth based on website traffic or the requirement for maintenance in the following month.
Prescriptive analytics uses data and sophisticated computations to recommend particular actions, such as buying new trucks to reduce late shipments and fleet costs over the long term. Based on the analysis, it might potentially suggest raising order sizes.

Supply chain management can profit from supply chain optimization simulations, which identify the best decisions for particular processes, in addition to traditional analytics tools, and cognitive analytics, which uses machine learning to quickly analyze large amounts of supply chain data.

The ability of an organization to successfully form and manage strategic alliances with outside partners, such as vendors, distributors, or service providers, is referred to as alliance management competency. These partnerships are established to take use of each party's skills, resources, and competencies in order to achieve mutual benefits and improve overall organizational performance.

The use of artificial intelligence (AI) in supply chain management has fundamentally changed how businesses process and make use of data in recent years. Large-scale supply chain data collection, processing, and analysis employing AI algorithms and techniques is referred to as AI-powered supply chain analytics capacity. It enables businesses to get insightful information, take well-informed decisions, and optimise key supply chain functions like demand forecasting, inventory control, logistics planning, and risk management.

The mediating impact of AI-powered supply chain analytics capabilities is relevant when analyzing the relationship between alliance management competence and organizational performance [14]. According to the mediating impact, the availability of AI-powered supply chain analytics capability serves as a connecting and strengthening mechanism between alliance management capability and organizational success.

Building out a **supply chain map** is a crucial step in deploying supply chain analytics into a management system. A typical supply chain map looks like fig. 1. Here, S1-S9 are suppliers.

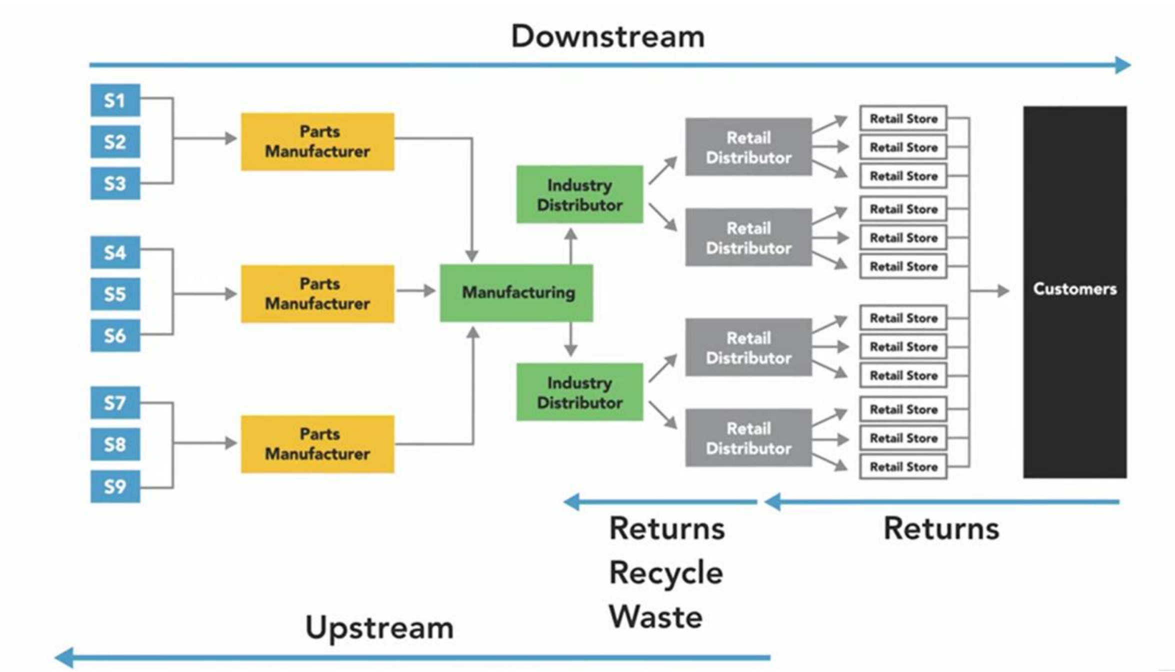

Fig. 2: Supply chain map

The first step in creating a supply chain map is to consider outlining stakeholders and their connections. merely being able to observe all parties involved, including all suppliers, clients, and intermediaries. So, it's clear how everyone is interconnected and how their actions may affect the placements and capabilities of the facilities. The flow of materials downstream, in the direction of the consumer, and the flow of returns and packaging upstream can also be included. We can map it using both the machinery and the people that operate in the supply chain. Since payments will be made to people, the flow of money can also be mapped. Then, using this information, we can map inventory levels, lead times, defect, damage, and theft rates. The datasets that are really being tracked must then be incorporated before we can move further with regulatory reporting. Yes, the supply chain map serves as a pointer to the data, but frequently, even though it makes us aware that the data is there, we find that the data actually belongs to the supplier or possibly one of the downstream partners, such as distributors, carriers, or retailers. We need access to real-time data in addition to historical data from previous years in order to get the greatest supply chain data. Consequently, it is essential to maintain a solid supply chain relationship with individuals who have access to the data. Between business partners and the industry in question, trust must be mutual. Particularly if the business has expanded through mergers and acquisitions. The company may have a number of ERP systems and other technologies that have been accumulated over many years. There are many problems a firm can run into with its own data. Various businesses are gathering data along the supply chain. Each stakeholder employs a unique set of hardware and software. Each is also a potential entrance point for a cyber assault, despite being a doorway to important data. Even if there is mutual trust between the organization in question and the supply chain, supply chain professionals are advised to be aware of potential attacks on the company's IT and supply chain infrastructure because software, cybersecurity, digital devices, and data storage are all expensive. As a result, allocation is and ought to be based on prioritizing ROI (return on investment). The whole cost of the data used to power the analytics programme must be taken into consideration by business leaders. Companies seek more data due to the size of the supply chain, which includes social media data, data from digital sensors, and even theoretical data from simulations. However, just like with books, just because we have access to every book in the library doesn't mean we can instantly find and read it. Determining the data required, locating it, gaining access to it, and then properly utilizing it can be quite difficult when we have TBs of supply chain data. The advantages and disadvantages of obtaining the best data should therefore be taken into account for the organization's smooth operation.

Following is the short description of the four traditional analytics procedure currently followed in the organizations:

Descriptive analytics: Supply chain professionals can now know what transpired thanks to these analytics. They help us understand the present with some situations as well. delivery and sales from yesterday. inventory levels throughout the previous seven days. These statistics are illustrative. Descriptive analytics, however, cannot predict the future or provide an adequate justification for past events. Retrospective analytics therefore include descriptive analytics. Supply chain managers would utilise descriptive analytics to track hundreds of different factors. For instance, data on on-time deliveries, lead times to various clients, personnel costs, gasoline costs, and inventory levels. Because there are so many of them, businesses choose the most noteworthy descriptive statistics as their KPIs, or key performance indicators. These KPIs can be used by managers to assess the performance of the supply chain. To present the findings of these analytics, simple statistics, sums, averages, standard deviations, percentages, and ratios are typically employed. Despite being simple, the data can be overpowering. To help with the transmission of these insights, businesses typically use graphs, charts, maps, and tables. To convey and evaluate descriptive analytics, enterprise resource planning tools, statistical software, and traditional spreadsheets are widely utilized. Dashboards are typically a stylish yet uncomplicated way to provide illuminating analytics. Additionally, stakeholders could receive digital notifications from automated systems when goals are achieved or, conversely, when dangerous levels are about to occur. Practically all supply chain experts often use descriptive analytics.Professionals in supply chain analytics who use descriptive analytics need to have a solid understanding of the supply chain, but just a basic grounding in statistics and little experience using technology tools. This is due to the fact that these frontline statistics are typically straightforward and provide frequent, sometimes real-time, updates on the state of the supply chain. Working with spreadsheets, producing aesthetically pleasing dashboards, maps, and graphs, and using data in reports and presentations are all necessary for understanding descriptive analytics and conveying to other team members the value of well-organized data.

Diagnostic analytics - These analytics explain to supply chain experts why that occurred. What transpired is revealed via descriptive analytics. Diagnostic analytics let us look into the reasons. Consequently, diagnostic analytics offer insight while descriptive analytics offer hindsight.

Diagnostic analytics would be used by supply chain management to conceivably look for underlying causes. Descriptive analytics, for instance, might have informed us that stockouts of our sports beverages increased. Our diagnostic analytics tools may provide information on current inventory levels, local weather conditions, variations in demand from the same week last year or from the same week this year, rates of damage and theft, and late supplier or store deliveries. All of these metrics have the potential to contribute to sports drink stockouts. A skilled supply chain analytics specialist could try to identify the root cause using analytics tools and their prior job experience, and then possibly try to create a strategy to prevent future stockouts. These analytics are frequently given following a statistical analysis. Data mining, root cause analysis, regression, sensitivity analysis, correlations, and time series analysis may all be included in this. Everyone is looking for proof that incident X led to outcome Y. Drill down analytics is another name for descriptive analytics. Dashboards for the supply chain frequently include descriptive analytics. A supply chain expert can click on the descriptive analytic or drill down to understand which metrics are frequently associated with certain outcomes when they notice something fascinating or troubling on the dashboard. The supply chain analytics expert who creates the dashboard will be in charge of discovering and then connecting a descriptive analytic to a prospective diagnostic analytic because these drill down linkages are not always intuitive. Supply chain detectives are diagnostic analytics experts. They search for the culprit whenever something goes wrong. Understanding the actual processes of purchasing, manufacturing, moving, and selling items enables you to connect the dots between various occurrences and results. Therefore, supply chain analytics experts who work with diagnostic analytics need to have a solid foundation in business statistics so they can develop correlations, as well as some dashboard technology abilities so they can help build interfaces. They also need to have a reasonable amount of supply chain understanding.

Predictive analytics: Supply chain experts can use these insights to predict what will happen next. In retrospect, descriptive analytics tools have shown us what occurred. diagnostic analysis, understanding why that occurred. To deliver insight, predictive analytics project into the future. We can only hope that predictive analytics will assist us in making wise judgements going forward and in avoiding bad choices in the past. For instance, descriptive analytics revealed that we experienced sports energy drink stockouts in 2019. Diagnostic analysis revealed a correlation between this and weather patterns five degrees above average and price rises for citric acid. Now, we can forecast the weather and the price of citric acid using predictive analytics models using both historical and real-time data. The teams in charge of demand and operations will keep an eye on these forecasts as they change and develop. Naturally, the hope is that better decision-making would lead to better business outcomes and more efficient use of resources. For decades, these methods are used that to forecast demand and price. However, one can create incredibly complex models when he/she has access to terabytes of data, including data on customers, sales, suppliers, logistics, the environment, unstructured social media, and your own machinery and equipment, all in real-time.The data can even be used in simulations, which can allow us to see a variety of possibilities, each with a distinct likelihood. Businesses can forecast when equipment will malfunction and which trucking routes would be the quickest. We can strengthen our reverse logistics crew by using social media to potentially detect impending returns. Predictive analytics are therefore significantly more difficult, whereas descriptive and diagnostic analytics required great supply chain expertise and undergraduate level statistics. We all know how difficult it is to anticipate the weather, sports results, volcanic eruptions, and the stock market. Even with all the variables and data available, there is still so much that humans do not fully comprehend. Therefore, experts in predictive analytics need advanced knowledge of data, statistics, and programming, even only Python and R. A data scientist is frequently required to create forecasting models and simulations. Companies are looking for data scientists that are prepared to apply machine learning systems as the amount of data, both historical and real-time data, rises. There aren't enough data scientists in the world today who are familiar with supply chain management.

Prescriptive analytics: What should be done is revealed by these analytics for supply chain specialists. What transpired is revealed via descriptive analytics. Hindsight. Analytics for diagnostics, why did that occur? Insight. The future was predicted through predictive analytics. Foresight. Analytics that are prescriptive go one step further. Based on your forecasts, they provide choices, courses of action, or tactics. They use the predicted scenarios from predictive analytics to foresee the consequences depending on various possibilities for making decisions. Descriptive analytics, for instance, reported stockouts of sports energy drinks. Diagnostic analysis linked this result to the weather and the cost of the materials. According to predictive analytics, there is a 60% likelihood that the temperature will be higher than average in the upcoming month. A 30% likelihood will be about typical. There's also a 10% risk that it won't be typical. Additional predictive analytics suggest a 40% likelihood that the price of citric acid will be extremely high. 40% probability that prices won't change. 20% probability that prices will decrease. The amount of information is overwhelming but also beneficial. Managers merely want to know how to maintain cheap prices, satisfy customers, and prevent stockouts. Prescriptive analytics techniques are useful in this situation. These tools simulate the many scenarios I have stated based on the forecasts. They examine the likelihood of each combination of circumstances using the various percentages. After that, the system will simulate each of the choices you're making for each scenario. Obtain more now? Only purchase more when stock levels are low? More items are kept in the warehouse? Deliver more goods to the shops now? ONe can work on a report detailing anticipated expenses and revenues, inventory levels, and the possibility of stockouts for every possible decision combination and scenario. Managers can then review the grid, talk about the findings, and make better decisions. Or they can let the system decide if they truly believe in the prescriptive analytics tools. Work with both organized and unstructured data, use optimisation techniques, and machine learning to enhance your models. These systems are really cutting-edge. Many businesses decide to employ prescriptive analytics techniques using industrial software. Others will create their own models and tools to exactly fit their supply chains.

Questions like what happened, why, what will happen next, and what should be done about it were answered by traditional supply chain analytics. These types of analytics technologies are capable of managing a much more real-time data from thousands of devices than cognitive supply chain analytics, which frequently aim to accomplish the same goals. So certainly, occasionally these are merely updated versions of your standard analytics. What exactly are cognitive supply chains, then? Cognitive supply chains attempt to resemble people. thinking, communicating, and making decisions. mimicking people? Why not just leave it to people, you could ask? Well, cognitive systems are capable of processing large amounts of structured and unstructured input. Yes, the foundation of cognitive supply chain is frequently seen in IoT and supply chain optimisation technologies. As you can see, cognitive supply chains make use of IoT infrastructure, IoT data, and all of the outputs from advanced supply chain optimisation to apply machine learning, a subfield of artificial intelligence, to uncover deeper patterns and knowledge. It is hoped that cognitive supply chains will be able to make discoveries more quickly than people and perhaps things that people might miss. So, certainly, cognitive supply chain systems can process data more quickly, use the internet of things to instantly drive optimized decisions down the supply chain, and then learn from the outcomes, a never-ending cycle of instruction and practise. Though it doesn't actually, it sounds futuristic. Although it's difficult, many firms with extensive worldwide supply chains are already making investments in the creation of cognitive supply chains. For any organization, creating a cognitive supply chain is overwhelming. Because of this, it takes a special team of people, each with their own supply chain and analytics abilities.

Natural Language Processing (NLP)

Organizations can gain useful insights from unstructured text data with the development of Natural Language Processing (NLP) tools. Computers can now comprehend, interpret, and produce human language thanks to a variety of technologies and approaches known as natural language processing, or NLP. Businesses can use NLP to automate processes like document classification and entity extraction, extract pertinent information from massive amounts of text, and acquire a deeper knowledge of client feelings. Additionally, NLP is essential for sentiment analysis, topic modeling, and text summarization, enabling organizations to base choices on textual data on data-driven analysis. The incorporation of NLP into analytics frameworks improves the total analytical capabilities, enabling companies to use language as a competitive advantage and a growth engine.

Recent years have seen a substantial increase in interest in the subject of natural language processing (NLP), notably in regard to the sentiment analysis of online reviews. It is important to carefully consider consumer happiness and feedback.

A branch of linguistics, computer science, and artificial intelligence called "natural language processing" studies how computers and human language interact, with a focus on how to programme computers to process and analyze massive amounts of natural language data.

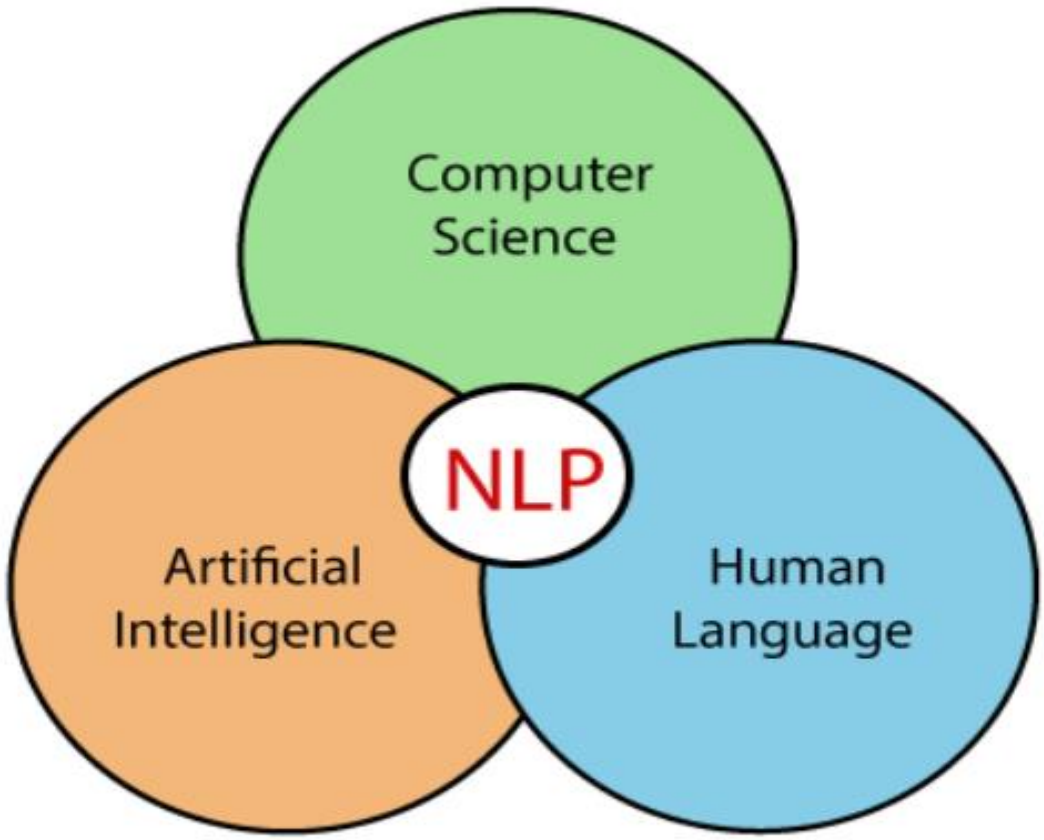

Fig. 3: NLP Venn diagram

Natural language is distinguished from formal and constructed languages such as those used to study logic or program computers.

The different approaches to perform NLP tasks are as follows:

1) Rule-based (heuristic) approach
2) Machine learning based approach
3) Deep learning approach

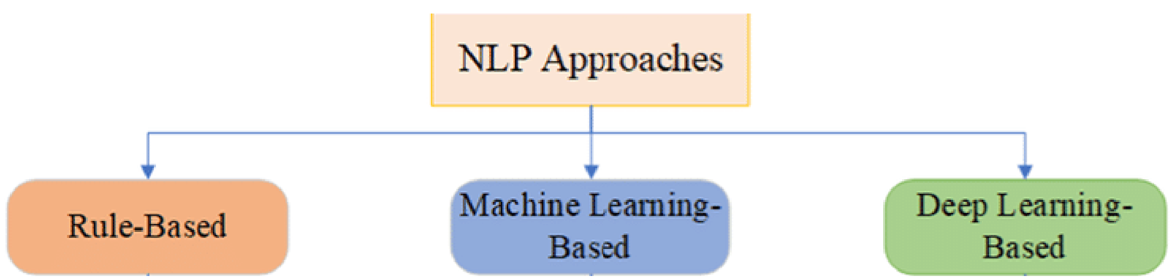

Fig 4: Different approaches used for NLP

Rule-based (heuristic) approach: The Natural Language Processing (NLP) heuristic or rule-based approach entails creating a collection of predetermined rules or heuristics to analyse and interpret linguistic data. These laws frequently depend on linguistic patterns, grammatical structures, semantic connections, or expertise of a certain subject area. Here is a description of the NLP heuristic approach's methodology:

Rule Creation:

Based on their expertise and comprehension of the language and the particular task at hand, linguistic and domain specialists develop a set of rules.

Language theories, grammatical rules, syntactic constructions, semantic links, and other pertinent knowledge sources can all be used to generate these rules.

Regular expressions, pattern matching, if-then statements, or certain linguistic restrictions are frequently used to represent rules.

Preprocessing:

To extract linguistic properties that help with rule application, the input text is placed through preprocessing procedures such tokenization, part-of-speech tagging, and parsing.

Tokenization entails separating the text into tokens, or single words.

Each token is given a grammatical label (such as a noun, verb, or adjective) by part-of-speech tagging.

Parsing examines the grammatical links between words in sentences, such as the subject-verb-object relationship.

Rule Application:

The preprocessed text is subjected to the rules in order to extract pertinent data or carry out particular linguistic tasks.

A particular language pattern, syntactic structure, or semantic link is often the emphasis of each rule.

In order to recognise particular language constructs or entities, the rules may use pattern matching, syntactic parsing, or semantic analysis.

Rule Prioritization and Ordering:

The priority of certain linguistic elements or patterns is established by the sequence in which the rules are implemented.

Rules are frequently applied in descending order of specificity, with more detailed rules taking precedence over more broad ones.

The resolution of competing rules in an acceptable manner and the precise capture of the relevant linguistic qualities are both guaranteed by rule prioritisation.

Iterative Refinement:

The heuristic technique frequently entails an iterative process of fine-tuning and modifying rules in response to input, assessment, and domain-specific requirements.

To gauge the correctness and efficiency of the rule-based system, gold standard datasets or expert opinions are compared to it.

To enhance the functionality of the system, the rules may be altered, enlarged, or fine-tuned based on the evaluation results.

Limitations:

The effectiveness and scope of the predefined rules are crucial to the heuristic approach.

A thorough system of rules can take a lot of time and work to create and maintain.

It can be difficult to deal with linguistic ambiguity and exceptions because some rules might not cover every variant or edge case.

The heuristic method could not generalise effectively to linguistic patterns or domains that are new or uncharted.

The heuristic method has been extensively employed in NLP applications like named entity identification, information extraction, text categorization, and rule-based machine translation despite these drawbacks. It offers a clear and understandable technique for examining and processing language input, giving linguistic and subject-matter specialists more control over how NLP systems behave.

Machine learning based approach:

Natural Language Processing (NLP) uses machine learning to automatically understand patterns, relationships, and representations of language by training models on massive volumes of text data. Following that, these models can be applied to a variety of NLP tasks, including sentiment analysis, text classification, named entity identification, machine translation, and more. The methodology for the machine learning-based approach to NLP is described below.:

Data Preparation:

The collection and preprocessing of the text data is the initial phase. To do this, you must perform activities like tokenization (dividing text into words or other smaller units), remove stopwords (frequent words with little meaning), and deal with text normalisation (such as handling contractions and changing words to lowercase).

Feature Extraction:

Text data needs to be translated into numerical features because machine learning algorithms demand numerical inputs. In NLP, common techniques for feature extraction include:

Bag-of-Words: Represents text as a group of distinctive words and the frequency of each word in a document.

TF-IDF (Term Frequency-Inverse Document Frequency): gives words weights based on their rarity across the corpus and their frequency in a document..

Word Embeddings: captures the links between words' semantic properties by representing words as dense vectors in a continuous space. The word embedding methods Word2Vec, GloVe, and FastText are all well-liked.

Model Training:

Various machine learning algorithms can be used to train NLP models, including:

Naive Bayes:

Definition: Based on Bayes' theorem, Naive Bayes is a probabilistic algorithm. Given the class, it is assumed that the characteristics are conditionally independent.

Advantages:

efficient in terms of computation and capable of handling big feature spaces.

demonstrates good performance with little training data.

works well for text categorization tasks like sentiment analysis and spam detection.

Cons: Makes assumptions about feature independence that may not be true in practical situations.

If it comes across a feature in the test set that wasn't present in the training set, it can experience the "zero probability problem".

Use Case: Document classification, sentiment analysis, and email spam detection.

Support Vector Machines (SVM):

Definition: SVM is a binary classification algorithm that, by maximising the margin between classes, locates a hyperplane in a high-dimensional space to divide them.

Advantages:

able to handle huge feature spaces and effective in high-dimensional settings.

strong generalisation abilities even with little data.

works well for document classification and text classification, including sentiment analysis.

Disadvantages:

big datasets may be computationally expensive.

It can be difficult to select the right kernel function and tune hyperparameters.

Use Case: Sentiment analysis, named entity identification, and text classification.

Decision Trees:

Definition:With the help of decision trees, which are hierarchical structures, predictions can be made by dividing data based on specific traits.

Advantages:

simple to understand and offer information on the significance of a characteristic.

can manage data that is both numerical and categorical.

Can capture non-linear relationships and handle feature interactions well.

Disadvantages:

overfitting is a risk, especially with complicated trees.
sensitive to slight data differences and may produce various trees.
Use Case: Sentiment analysis, subject identification, and text classification.

Random Forest:

Definition: An ensemble model called Random Forest mixes various decision trees to produce predictions.
Advantages:
by averaging the predictions from various decision trees, reduces overfitting.
ably handles a variety of input features.
estimates the value of the characteristic.
Disadvantages:
expensive computationally, especially for big datasets.
less comprehensible than distinct decision trees.
Use Case: Text classification, sentiment analysis, and named entity recognition.

Logistic Regression:

Definition: A statistical model which is utilised for binary or multiple-class classification tasks.
Advantages:
efficient and effective with data that can be separated linearly.
provides class membership probabilities.
works well for sentiment analysis and text classification.
Disadvantages:
assumes that characteristics and the target variable's log-odds have a linear relationship.
Complex non-linear interactions might not be captured.
Sentiment analysis, document classification, and named entity recognition are use cases.

Recurrent Neural Networks (RNN):

Definition:RNNs are neural networks with recurrent connections that are made to analyse sequential data. They can recognise relationships and patterns in sequential data thanks to a "memory" they possess.

Benefits: Captures long-term dependencies and effectively handles sequential data.

allows for variable-length inputs.

Suitable for activities including named entity identification, sentiment analysis, language modelling, and machine translation.

Disadvantages:

suffers from the "vanishing gradient" problem, which makes it challenging to detect long-term relationships since gradients become less pronounced as they spread through time.

costly in terms of computation, and it might have trouble with excessively long sequences.

Use cases include named entity identification, sentiment analysis, language modelling, and machine translation.

Convolutional Neural Networks (CNN):

Definition: CNNs are neural networks created specifically to interpret grid-like input, including images or word sequences. To extract regional information from the input data, they employ convolutional layers.

Advantages: Captures regional trends and spatial linkages well.

They are computationally effective due to parameter sharing.

Good for jobs like text classification, sentiment analysis, and document classification.

Cons: Sequential data is less successful at capturing long-range dependencies.

require fixed-size inputs, making it difficult to handle variable-length sequences.

Text categorization, sentiment analysis, document categorization, and image captioning are examples of use cases.

Transformer Models:

Definition: Transformer models include a self-attention mechanism at its core that enables them to recognise contextual linkages among words in a sequence. They have attracted a lot of attention in NLP because they are good at capturing long-distance dependencies.

Advantages: Even for lengthy sequences, effectively captures contextual information.

They are useful for large-scale training since they are parallelizable and computationally efficient.

performance at the cutting edge for a variety of NLP tasks, including sentiment analysis, text generation, and machine translation.

Cons:

Needs a substantial amount of training data to work at its best.

costly in terms of computation during training due to the quantity of parameters.

Use Cases:

Text production, sentiment analysis, and question-answering by machines.

Transformer model advancements have produced important designs such as RoBERTa (Robustly Optimised BERT), GPT (Generative Pre-trained Transformer), and BERT (Bidirectional Encoder Representations from Transformers). The outcomes of several NLP tasks using these models have greatly surpassed the state-of-the-art.

These are just a few examples of machine learning algorithms used in NLP.

Deep learning-based approach:

Neural networks with numerous layers are used in deep learning NLP techniques to process and comprehend textual data. These models develop hierarchical representations of language, which enables them to identify complex links and patterns in the data. Let's explore some commonly used deep learning algorithms in NLP:

Recurrent Neural Networks (RNN):

RNNs are made to handle sequential data by preserving a state in internal memory. Information can endure over multiple time steps because they work with a series of inputs and have connections that form a directed cycle.

RNNs are effective at performing tasks like sentiment analysis, machine translation, and language modelling.

Standard RNNs, on the other hand, experience the "vanishing gradient" problem, where gradients become smaller as they spread over time, making it challenging to detect long-term dependencies.

Long Short-Term Memory (LSTM):

A kind of RNN that lessens the vanishing gradient issue is the LSTM. To regulate the information flow inside the network, they add specialised memory cells and gating systems.

Long-term dependencies can be captured by LSTMs, which are now often employed for sentiment analysis, text generation, and machine translation.

Gated Recurrent Unit (GRU):

Another RNN architectural option that deals with the vanishing gradient issue is GRUs. They use LSTM-like gating methods, but with fewer parameters.

GRUs have been used in applications including machine translation, question answering, and dialogue systems because they are computationally effective.

Convolutional Neural Networks (CNN):

CNNs are frequently related to computer vision; nevertheless, they have also demonstrated success in NLP applications involving grid-like data, such as text represented as a list of words.

In order to identify regional patterns and dependencies within the input text, CNNs in NLP commonly use 1D convolutions.

CNNs excel at tasks like named entity identification, sentiment analysis, and text categorization.

Transformer:

Since their release, transformers have had a big impact on NLP. They use a self-attention mechanism instead of repetition or convolution to identify contextual relationships among words in a sequence.

In tasks including machine translation, document summarising, question answering, and text production, transformers have displayed impressive performance.

BERT (Bidirectional Encoder Representations from Transformers), GPT (Generative Pre-trained Transformer), and RoBERTa (Robustly Optimised BERT) are notable transformer models.

These are but a few instances of how deep learning algorithms are applied in NLP. Each algorithm has its own advantages and disadvantages, making them appropriate for various tasks and datasets. When selecting an algorithm, it is crucial to take into account elements including the nature of the data, processing capabilities, and particular task needs. Additionally, continuing research keeps introducing new algorithms and improving existing ones, pushing the limits of deep learning for NLP.

Model Evaluation and Optimization:

The model must be evaluated after it has been trained using assessment metrics like accuracy, precision, recall, F1-score, or area under the ROC curve.
The model can be further adjusted by changing the hyperparameters (settings that affect model performance) in order to preserve robustness and prevent overfitting.

Deployment and Testing:

After model evaluation and optimisation, the trained model can be used to make predictions on new, unanticipated data.
The model is put to the test using real-world scenarios to help identify any performance issues, modify the parameters, and improve the model's accuracy over time.
It's important to keep in mind that new algorithms and techniques are always being developed in the field of NLP. As the field of NLP continues to grow swiftly, researchers and practitioners are actively looking into novel architectures, pre-training techniques, and transfer learning procedures.

For the sentiment analysis of Amazon reviews the following 2 algorithms are used for the research.
VADER(Valence Aware Dictionary and sEntiment Reasoner): VADER is a sentiment analysis method built on machine learning. It employs a pre-established sentiment lexicon, which consists of words with corresponding sentiment scores. It examines the text using rules and heuristics, taking into account the modifiers, punctuation, and the polarity and strength of the words.
Benefits: Simple to use and apply.

It works well with informal content, such as posts on social media, and is particularly useful for sentiment analysis.
both the phrase and document levels, delivers sentiment and intensity scores.

Limitation to the sentiment lexicon, which might not include all slang or industry-specific terminology, is a drawback.
may have trouble with sarcasm, negation, and situational emotions.
relies on heuristics and might not be able to grasp sentiment's intricacies.

RoBERTa(Robustly Optimized BERT):

Algorithm: A deep learning model called RoBERTa is built on the Transformer architecture. It is a BERT (Bidirectional Encoder Representations from Transformers) model version that has been enhanced for extensive pre-training on a variety of data sources.
Advantages: Recognises the relationships between words in context and is linguistically sophisticated.
excels at identifying distant dependencies and does well on a variety of NLP tasks.
can handle slang and vocabulary particular to a given domain through pre-training on a big, varied dataset.

Demerits:
uses a lot of computational power for both inference and training.

need lots of labelled data to be fine-tuned, especially for certain domains.
comparatively poor interpretability to rule-based models like VADER.

Comparison:
RoBERTa is a deep learning model, whereas VADER is a rule-based approach.
RoBERTa learns representations through extensive pre-training, while VADER uses a predefined sentiment lexicon.

VADER is simpler to use and appropriate for informal text, although it might have trouble expressing attitudes that depend on the context. Although RoBERTa excels at a number of tasks, it needs additional computer power and labelled data.
RoBERTa's sentiment scores are learned from data, while VADER's sentiment scores are rule-based.
Even though VADER may be easier to use for rapid sentiment analysis, RoBERTa offers greater adaptability and versatility with finer customization.

Both models can be applied to the sentiment analysis of Amazon reviews for Indian products. VADER has a simple implementation and excels at handling casual text. However, RoBERTa needs additional processing power and labelled data to be able to capture the context and subtleties of sentiment.
There are various steps involved in adding an NLP sentiment analysis code to an Amazon review's existing comment area. A step-by-step manual to assist the deployment procedure is provided below:

Step 1: Prepare the Development Environment
Create a development environment and add the required software and libraries. Typically, this consists of Python, a code editor, and the necessary tools for sentiment analysis and NLP, like NLTK or spaCy.
Step 2: Collect and Preprocess Data
The comment section can be used to collect a representative sample of Amazon reviews.
Preprocess the data by eliminating extraneous characters, changing all text to lowercase, and taking care of any demands made by your sentiment analysis model.

Step 3: Train a Sentiment Analysis Model

Choose a sentiment analysis algorithm or model, such as a classic machine learning technique like Naive Bayes or a pre-trained neural network model like BERT.

Create training and testing sets from your data.

Utilise the training data to train the sentiment analysis model.

Using the testing results, assess the model's performance and make any necessary adjustments.

Step 4: Implement the Sentiment Analysis Code

Create new code or alter current code to include the sentiment analysis model.

Connect the code to the Amazon reviews platform's comment section.

Make that the sentiment analysis code can manage user input, analyse the text, and deliver results in real-time.

Step 5: Test and Debug

To ensure the sentiment analysis code's accuracy and dependability, test it on a variety of example comments.

Fix any problems or mistakes that may appear when testing.

If necessary, make code modifications to boost performance.

Step 6: Deploy the Code

Deploy the sentiment analysis code in your working environment.

Easily incorporate it into the existing Amazon review comments area.

Make sure that the platform's usability and user experience are not affected by the deployment procedure.

The intricacy of the sentiment analysis model, the scope of the deployment, and the necessary infrastructure are only a few of the variables that affect cost estimation. Development time, server or cloud hosting fees, as well as potential continuing maintenance and support, are possible expenses. It is advised to speak with the technical and financial teams at your company to obtain a more precise cost estimate that takes into account your unique requirements.

Please be aware that there may be extra factors to take into account when integrating a sentiment analysis code into a live platform like Amazon reviews, such as adhering to Amazon's terms and conditions, privacy rules, and API usage guidelines. During the deployment process, be sure to read and abide by all pertinent policies and rules.

A blockchain-based mechanism to gather customer reviews of vaccines is suggested by Hu et al. [43]. Vaccine manufacturers can learn more about consumer attitudes and demands by utilising text sentiment analysis algorithms, which enables them to recommend high-quality vaccines that are suited to specific tastes. Their system promotes competition among vaccine enterprises to enhance vaccine quality and address consumer concerns. The efficiency of machine learning methods, notably the trainable-BILSTM model, in precisely categorising sentiment is demonstrated in a case study utilising a dataset of online medicine reviews. The suggested system not only enables ongoing vaccination quality improvement, but also offers a useful resource for consumers and manufacturers to use when making decisions that would promote public health. A promising way to guarantee safe and effective vaccines while raising customer happiness is to use NLP techniques when analysing vaccine evaluations.[43]

Table 1 describes the different models used for sentiment analysis by different researchers on several datasets based on the use case.

Authors	Dataset	Method	Classification
Mahadevaswamy et al.[1]	Amazon reviews	Bi-LSTM	Positive/negative
Jose Luis Arroyo-Barrigiiete[2]	Lovecraft’s fiction	BoW	Surprise/disgust/joy/ anger/ anticipation/sadness/trust/fear

Chlapanis et al.[3]	CMU-MOSEI dataset	Adapted Multimodel BERT (AMB)	Positive/negative/ neutral/anger/disgust/ fear/happy/sad/ Surprise/emotion intensity
Singh et al.[4]	Mid-sized car reviews (CarWale)	Aspect level SA with bias corrected least squared dummy variable	Positive/negative/ neutral
Mello et al.[5]	Multilingual news articles covering London 2012 and 2016 Olympics	SenticNET, SentiStrength, BERT and Vader	Positive/negative/ neutral
Kaur et al.[6]	Consumer reviews (SemEval-2014 restaurant reviews, Sentiment140, STS-Gold)	LSTM	Positive/negative/ neutral
Savci et al.[7]	Customer reviews and complaints from e-commerce website in Turkish, English and Arabic language	Turkish - Bert-based Multilingual model, Convert-based- turkish- cased, Distilbert-base-turkish-cased English – Bert-base-multilingual-case,	Positive/negative

		DistilRoberta-base-model, Albert-base-v2 Arabic- Bert-base-multilingual-cased, Xml-roberta, Distilbert-base-multilingual-cased, RNN, CNN, Random forest, Multinomial Naïve Bayes, SVM	
Hernández et al.[8]	Twitter tweets (Kaggle)	Formal Concept Analysis (FCA)	Positive/negative
Balahur et al.[9]	ISEAR database	EmotiNet KB	Disgust/fear/ shame/anger/ Sadness/joy/guilt
Altamimi et al.[10]	Airline tweets	FastText, Linear SVC	Positive/negative/ neutral
Patel et al.[11]	Customer feedback and reviews for airline services (Kaggle)	Random forest, BERT, LSTM, RoBERTa, Electra	Positive/negative/ neutral

Table 1: List of approaches proposed for sentiment-analysis

Methodology

Data retrieval:

Amazon customer reviews and raw web-scraped data for Indian products have been retrieved from Kaggle. Typically, for web scraping python is considered due to easy data handling and procedure. Most popular libraries for the task are BeautifulSoup and Requests. The Requests library is used to send an HTTP GET request to the Amazon webpage that contains the products with their reviews. A required header, such as User-Agent is used to mimic a web browser. The HTML response is then captured which is returned by the request. The response will contain the contents of the webpage, namely, the reviews, product name, asin, and the score of the review, which are present in the particular class the type of data is in.

After getting the response library like BeautifulSoup is used to parse the HTML response and then to extract specific components, like review titles, review text, name of the reviewer, ratings, etc. Following this, locating and extracting the review data is done after inspecting the HTML structure of the page and identifying the attributes and tags that contain the data. Many such functions are available in the library to locate these elements. Following this, the data is generally stored in the csv (comma-separated values) format for further analysis.

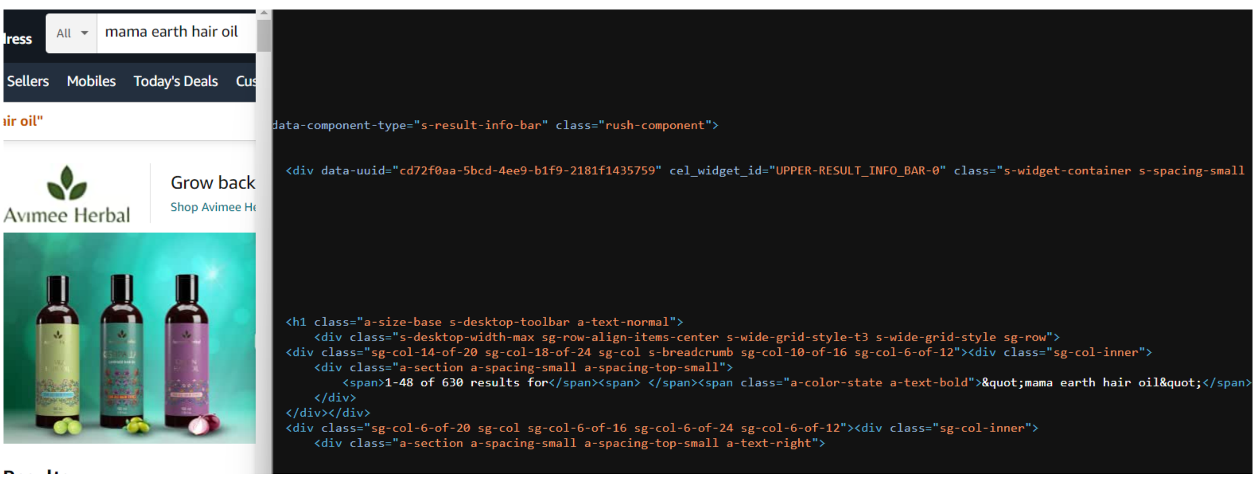

Fig. 5: An example of HTML script for a webpage from which data is parsed

The data is parsed using Requests and BeautifulSoup library.

The final data has the following fields:

asin : Amazon Standard Identification Number - a unique id for each product on Amazon name : name of the product

date : date on which review was posted

rating : rating given to the product (out of 5)

review: review given to the product

There are 1133 reviews for various Indian Products pertaining to categories like hair and skin care products, clothes, electronic gadgets, etc from companies like Godrej, Titan, Maaza, Paper, Indiana, Coca, Natural, Maggi, Glucon-D, Amul, Patanjali, Dettol, Savlon, Cinthol, Britannia, NutriChoice, Streax, Himalaya, Society, Tata, Fastrack, Reflex, Mysore.

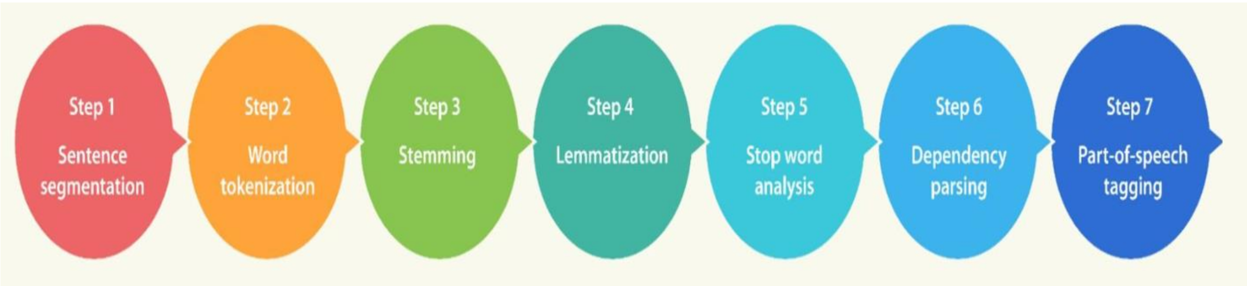

Fig. 6: NLP pipeline

Several procedures are routinely taken to clean up the text and get it ready for analysis when prepping Amazon review data for NLP. The preprocessing processes are broken down in more detail below:

Text Cleaning:
Remove any non-textual components from the text that don't add to its meaning, such as HTML tags, URLs, special characters, and punctuation.
To maintain consistency, change the text's case to lowercase or uppercase.

Tokenization:
Separate the text into tokens or individual words. This stage enables additional word-level analysis. Whitespace separation or more sophisticated methods like word or sentence tokenizers can be used for tokenization.

Stopword Removal:
Stop using words like "the," "is," and "and," which are frequently used but don't add any significance to the text.
Stopword lists are available for many different languages and can be tailored based on the particular needs of the study.

Stemming and Lemmatization:
Consolidate word variations by reducing them to their base or root form.
Stemming uses straightforward linguistic guidelines to eliminate prefixes or suffixes.
On the other hand, lemmatization uses linguistic information to transform words into their canonical form while taking into account elements like part of speech.

Handling Abbreviations and Acronyms:
For consistent analysis, expand or normalise acronyms to their full forms.
Making a dictionary of popular acronyms and their equivalent expansions can be part of this phase.

Handling Contractions:
To guarantee consistent word representation, expand contractions like "can't" to "cannot" and "I'll" to "I will".
This can entail employing a contraction mapping dictionary, just like with abbreviations.

Removing Duplicates:
Find any duplicate reviews or sections of reviews that are duplicates, and eliminate them.
Duplicate data might cause biases and affect analytical conclusions.

Handling Noisy Data:

Any noisy or incorrect data points that can affect analysis should be addressed.

Spell checking, fixing frequent typos, or addressing certain data concerns specific to the Amazon reviews dataset may all be part of this.

Vectorization:

Create numerical representations of the text that have been analysed that are appropriate for machine learning techniques.

The term frequency-inverse document frequency (TF-IDF) and word embeddings like Word2Vec and GloVe are examples of common approaches.

It's vital to keep in mind that the precise preprocessing procedures and sequence may change depending on the particular needs of the study and features of the dataset used for the Amazon reviews. For proper cleaning and preparation for NLP jobs, it is also advised to properly investigate and comprehend the dataset before preprocessing.

Handling emojis:

Emojis in Amazon reviews must be handled appropriately during preprocessing in order to take into consideration their potential impact on sentiment analysis.

Emojis should be recognised and processed so that sentiment analysis can take into account their meaning.

Emojis frequently express emotion and provide meaning for texts.

There are a few possible approaches to handle emojis:

Substitute textual equivalents for emojis: Transform each emoji into its corresponding text equivalent, such as ":)" for a love sign or "😄" for a smiling face.

Emojis to sentiment scores conversion: Depending on the sentiments they are frequently connected with, give each emoji a sentiment score. Give a happy face emoji a good score, and a sad face emoji a negative one, as an example.

Remove emojis: Emojis can be completely eliminated if they don't significantly add to the sentiment analysis work or aren't necessary for the analysis.

Emoji handling techniques are determined by the nature of the sentiment analysis task and the importance of emojis in expressing emotion in the Amazon reviews dataset. Emojis might contribute significant information or introduce noise, depending on the use case, thus it's important to take these factors into account when analysing sentiment.

The code above imports the following libraries:

pandas (as pd): Used for data manipulation and analysis.
Numpy (as np): Used for arrays and numerical operations.
matplotlib.A plotting library for making visualisations, pyplot module (as plt)is calling different methods for plotting.
A data visualisation package that integrates nicely with matplotlib and offers extra plotting features is called seaborn (as sns).

The line plt.style.use('ggplot') sets the plotting style to 'ggplot', which is a popular style for creating visually appealing plots. Shape of dataset is (1132,5). The nltk library for word tokenization has a punkt resource. This resource includes pre-trained models for tokenization. The tokenize function takes the review as a string and split it up into tokens or words (in this case). Tokenization can also be done with a pair of words. Each word is stored in the ‘List’ data structure.
The ‘averaged_perceptron_tagger’ resource of nltk library is used for part-of-speech (POS) tagging on the list of tokens.
Here, is an example of how POS tagging looks like:
[('This', 'DT'), ('is', 'VBZ'), ('an', 'DT'), ('example', 'NN'), ('sentence', 'NN'), ('.', '.')]
Here, each token is paired with it’s corresponding POS tag in a tuple.
Here, DT - Determiner
NN - Noun
VBZ - Verb

A task in natural language processing (NLP) called named entity recognition (NER) tries to recognise and categorise named items in text into specified categories. Named entities are particular pieces of information that can be categorised. Examples include names of people, businesses, places, events, seasons, currencies, and more.

The objective of NER is to automatically recognise, extract, and categorise these named entities from unstructured text into the appropriate categories or types. As a result, applications like information extraction, question answering, sentiment analysis, and others are made possible by NER, which also aids in understanding the structure and meaning of the text.
For example, in the sentence "I opted Mama earth hair oil thinking that the product will be delivered in Hyderabad (my hometown), but the product delivered in Surat." NER would recognize "Mama earth" as an organization and "Hyderabad", "Surat" as a location.

Tokenizing the text into individual words or tokens, giving part-of-speech tags to the tokens, and then utilising machine learning or rule-based techniques to identify and categorise the named entities are a few examples of the various stages that commonly go into NER systems.
Overall, NER is widely employed in numerous NLP applications to improve language interpretation and analysis and plays a significant role in extracting important information from text.
To perform NER using nltk, the resource 'maxent_ne_chunker' has been used. By passing the tagged tokens to nltk.chunk.ne_chunk() performs NER. This function locates named entities (such as names of people, businesses, or locations) and chunks them. The respective entity types are listed next to the specified entities, which are surrounded in brackets (e.g., ORGANISATION, GPE for geographical/political entities).

After downloading resource called "vader_lexicon," which is a vocabulary created especially for sentiment analysis, the SentimentIntensityAnalyzer class is imported, which is provided by nltk as a pre-trained sentiment analysis model. Tqdm is used to create the progress bars for progress of the loop/iteration tracking. A sia object is assigned to create an instance of 'SentimentIntensityAnalyzer' class. By giving each word a sentiment score and calculating the overall sentiment score for a specific sentence or document, this object is used to analyse the sentiment of the text. Following this polarity_scores() method provide the overall sentiment of the document.
For example: running this line of code
sia.polarity_scores('I am a loyal customer.')
will execute and will output the sentiment scores as follows:
{'neg': 0.0, 'neu': 0.238, 'pos': 0.762, 'compound': 0.4588}

The four keys of a dictionary are used to express the sentiment scores.

‘neg’: Negative emotion score, with a range of 0.0 to 1.0. It is 0.0 in this instance, demonstrating a lack of unfavourable feeling.

The score of neutral sentiment, which ranges from 0.0 to 1.0, is "neu." It is 0.238 in this instance, indicating a moderate degree of neutrality.

‘pos’: The score of the positive sentiment, which ranges from 0.0 to 1.0. It is 0.762 in this instance, which denotes a higher level of positive mood.

"Compound": The sentiment score that results from adding the three distinct sentiment scores together. It is 0.4588 in this instance, which indicates a largely favorable attitude.

After creating a loop to iterate through each review and creating these scores for each one of them we have a dataset that looks like the fig 7.

	asin	neg	neu	pos	compound	name	date	rating	review
0	B07W7CTLD1	0.0	0.748	0.252	0.9704	Mamaearth-Onion-Growth-Control-Redensyl	43625	1	I bought this hair oil after viewing so many g...
1	B07W7CTLD1	0.0	0.748	0.252	0.9704	Mamaearth-Onion-Growth-Control-Redensyl	14-08-2019	5	Used This Mama Earth Newly Launched Onion Oil ...
2	B07W7CTLD1	0.0	0.748	0.252	0.9704	Mamaearth-Onion-Growth-Control-Redensyl	19-10-2019	1	So bad product...My hair falling increase too ...
3	B07W7CTLD1	0.0	0.748	0.252	0.9704	Mamaearth-Onion-Growth-Control-Redensyl	16-09-2019	1	Product just smells similar to navarathna hair...
4	B07W7CTLD1	0.0	0.748	0.252	0.9704	Mamaearth-Onion-Growth-Control-Redensyl	18-08-2019	5	I have been trying different onion oil for my ...
...	...	...	...	...	...	...	...	...	...
1127	B07MVHJ6CH	0.0	0.680	0.320	0.8720	Mysore-Sandal-Soaps-Pack-Bars	43713	5	Fantastic
1128	B07MVHJ6CH	0.0	0.680	0.320	0.8720	Mysore-Sandal-Soaps-Pack-Bars	43833	5	Long lasting freshness throughout the day.
1129	B07MVHJ6CH	0.0	0.680	0.320	0.8720	Mysore-Sandal-Soaps-Pack-Bars	24-10-2019	5	My preferred soap
1130	B07MVHJ6CH	0.0	0.680	0.320	0.8720	Mysore-Sandal-Soaps-Pack-Bars	21-06-2019	4	Super Product
1131	B07MVHJ6CH	0.0	0.680	0.320	0.8720	Mysore-Sandal-Soaps-Pack-Bars	43897	5	Best soothing, cooling fragrance for hot summe...

1132 rows × 9 columns

Fig. 7: Resulting dataset after adding fields for neg, neu and pos scores

The program starts by importing the necessary libraries for data manipulation and analysis, such as transformers.

It creates a DataFrame called vaders by converting the res dictionary into a DataFrame and transposing it. The columns of the DataFrame correspond to different sentiment scores.

The DataFrame vaders is merged with another DataFrame called df using a left join operation.

A bar plot is created using seaborn (sns.barplot) to show the relationship between the Amazon rating and the compound sentiment score.

Subplots are created (fig, axs = plt.subplots(1,3, figsize=(12,3))) to display three bar plots: positive sentiment score, neutral sentiment score, and negative sentiment score, all grouped by the Amazon rating.

Titles are assigned to the subplots to indicate their content.

The plots are tightly laid out and displayed using plt.tight_layout() and plt.show().

The program installs the transformers library using the command pip install transformers.

It imports the necessary modules from transformers, including AutoTokenizer, AutoModelForSequenceClassification, softmax, and torch.

A pretrained Roberta model, specifically the "cardiffnlp/twitter-roberta-base-sentiment" model, is loaded using AutoTokenizer and AutoModelForSequenceClassification.

The program prints the value of the variable example.

It applies the VADER sentiment analysis on the example using sia.polarity_scores().

The program runs the example through the Roberta model by encoding the text with the tokenizer, passing it to the model, and obtaining the scores. The scores are then passed through the softmax function to obtain probabilities.

The scores from the Roberta model are stored in a dictionary called scores_dict.

The program defines a function called polarity_scores_roberta that performs sentiment analysis using the Roberta model. It follows a similar process as above and returns the scores in a dictionary format.

The program reimports the required modules from transformers to use the FlaxAutoModelForSequenceClassification instead of AutoModelForSequenceClassification.

It prints the value of the variable example.

It applies the VADER sentiment analysis on the example using sia.polarity_scores().

The program runs the example through the Roberta model using the updated modules. The scores are obtained and passed through the softmax function.

The scores from the Roberta model are stored in the scores_dict dictionary.

The polarity_scores_roberta function is defined again with the updated modules.

It iterates over each row in the DataFrame df using df.iterrows().

For each row, it retrieves the text and ID values.

It applies VADER sentiment analysis on the text using sia.polarity_scores() and stores the result in vader_result.

It renames the keys in the vader_result dictionary to include the prefix 'vader_' and stores the result in vader_result_rename.

It performs sentiment analysis using the Roberta model on the text using the polarity_scores_roberta function and stores the result in roberta_result.

The vader_result and roberta_result dictionaries are merged into a single dictionary called both.

The both(name of the dictionary) dictionary is stored in the res dictionary with the ID as the key.

If any errors occur during the iteration, a message is printed indicating the ID for which the process failed.

The res dictionary is converted into a DataFrame called results_df.

The index of the DataFrame is reset, and the column name 'index' is renamed to 'asin'.

The DataFrame results_df is merged with the original DataFrame df using a left join operation.

A pairplot is created using seaborn (sns.pairplot) to compare sentiment scores between VADER and Roberta models. The variables used for comparison are 'neg', 'neu', 'pos', 'roberta_neg', 'roberta_neu', and 'roberta_pos'. The plot is colored by the 'rating' column.

The program displays the review with the highest positivity score among 1-star ratings and highest negativity score among 5-star ratings using the Roberta model and VADER and thus the data results are validated.

Results and analysis

Exploratory data analysis (EDA)

The EDA of the dataset shows that we have 122 unique products in our dataset.

After splitting company names from the product names, I found that the words **PATANJALI** and **Patanjali** are the same and same applies to **MYSORE** and **Mysore** but these occur as uniqueness takes into consideration the case-sensitive nature of the strings.

The dataset has a total of 24 unique Indian brands namely, 'Mamaearth', 'Godrej', 'Titan', 'Maaza', 'Paper', 'Indiana', 'Coca', 'Natural', 'Maggi', 'Glucon', 'Amul', 'Patanjali', 'Dettol', 'Savlon', 'Cinthol', 'Britannia', 'NutriChoice', 'Streax', 'Himalaya', 'Society', 'Tata', 'Fastrack','Reflex', 'Mysore'.

The top five brands with the least number of reviews can be seen in fig. The least number of reviews are of Indiana and NutriChoice.

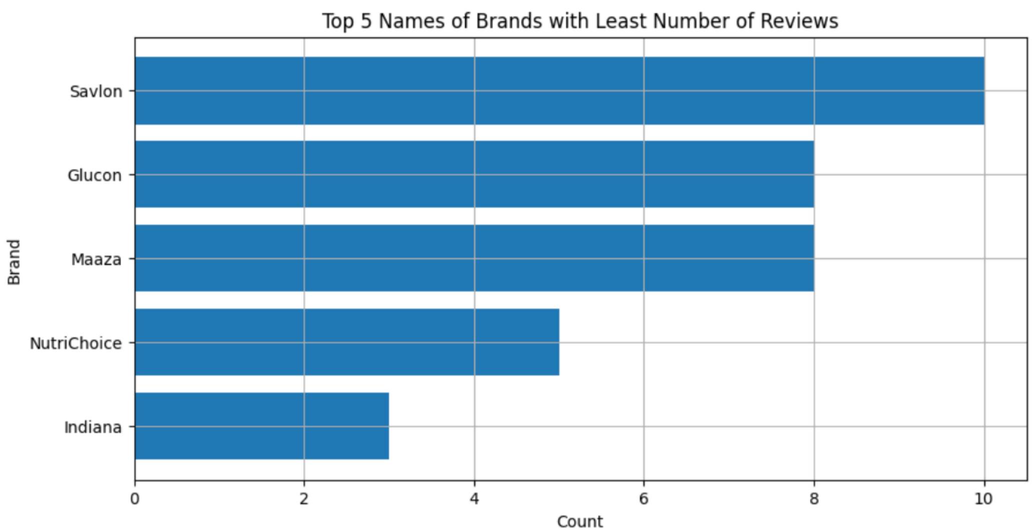

Fig. 8: Top 5 names of brands with least number of reviews

The top five brands with the most number of reviews can be seen in fig. The most number of reviews are of Godrej and Titan.

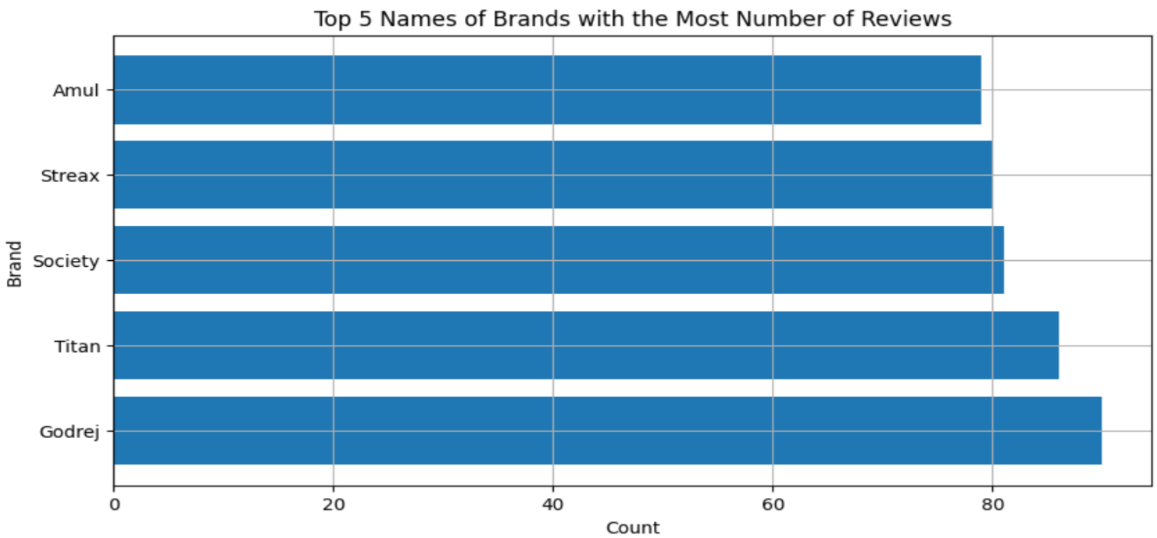

Fig. 9: Top 5 names of brands with most number of reviews

The dataset shows that we have polarized and biased review dataset. The 5 (highest) and 1(lowest) score reviews are more in numbers than any other scores.

The wordcloud of all reviews looks like fig 10.

Fig. 10: Wordcloud (all reviews)

The proportion of score using ‘textblob’ as sentiment analyser is shown in fig.

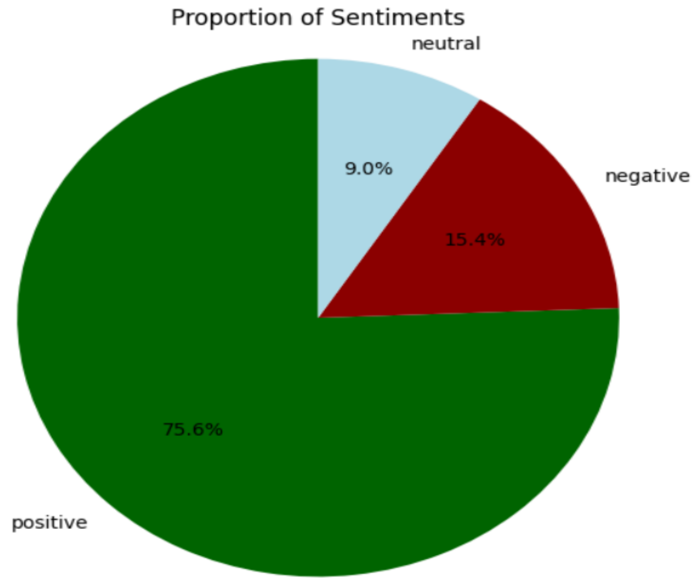

Fig. 11: Sentiment proportion

Here, we can see that the number of positive characterized reviews using textblob is much greater than negative and neutral.

The positive sentiment wordcloud is shown in fig 12.

Fig. 12: Wordcloud (positive reviews)

The negative sentiment wordcloud is shown in the fig 13.

Fig. 13: Wordcloud (negative reviews)

The neutral wordcloud is shown in fig 14.

Fig. 14: Wordcloud (neutral words)

Fig. 15: Review barplot

The barplot for vaders dataframe is shown in fig 15.

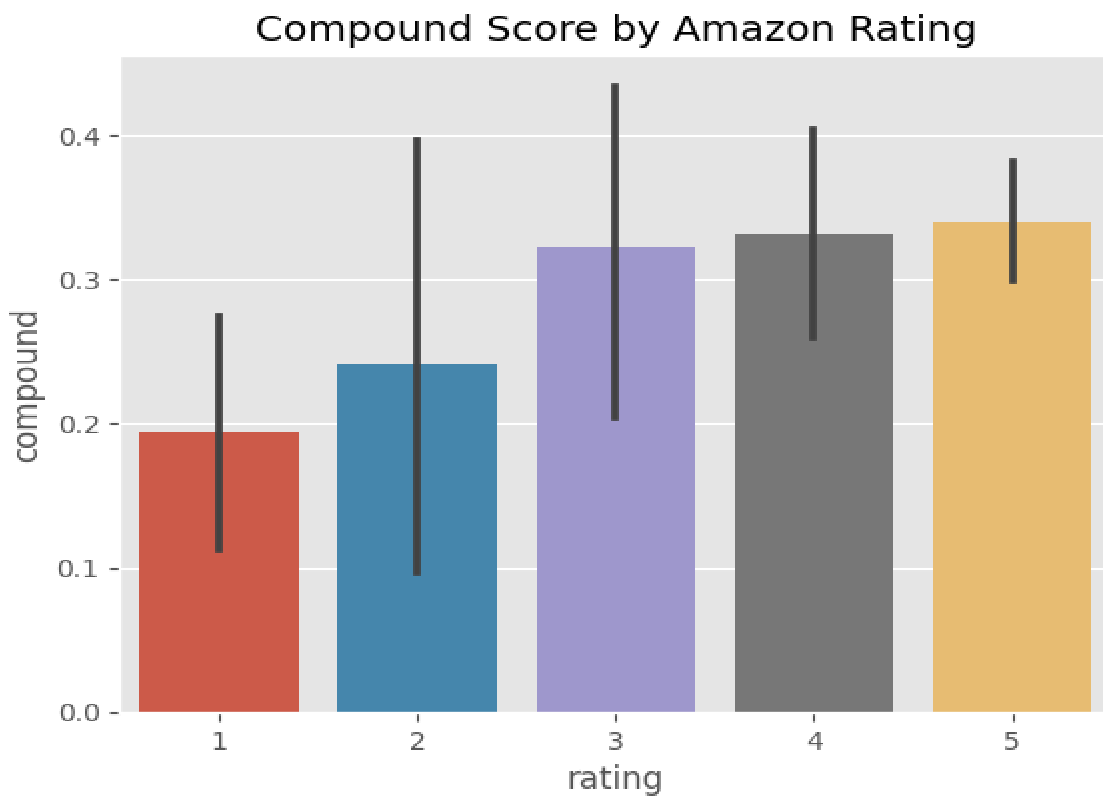

Fig. 16: Compound score by amazon rating

The barplot() function from seaborn library is used. Here, the ‘x’ parameter is ‘rating’ which indicates the x-axis that represents the Amazon rating.

The ‘y’ parameter is ‘compound’ depicting that it is compound score by the Amazon rating.

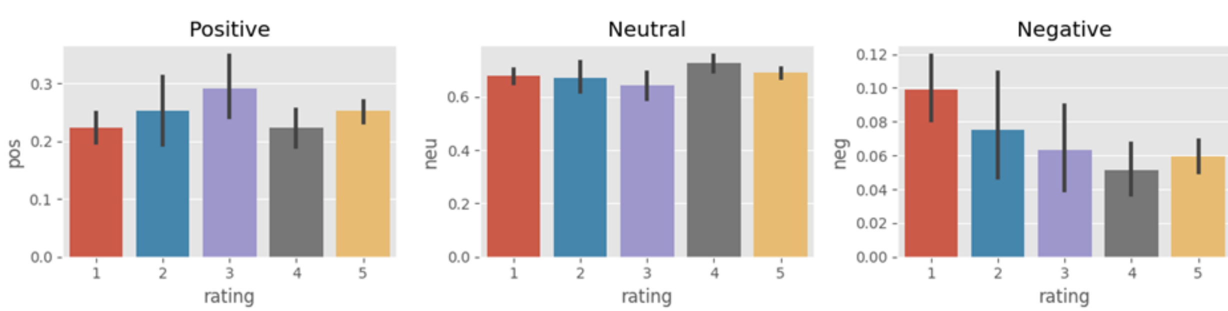

Fig. 17: Positive, neutral and negative score barplots

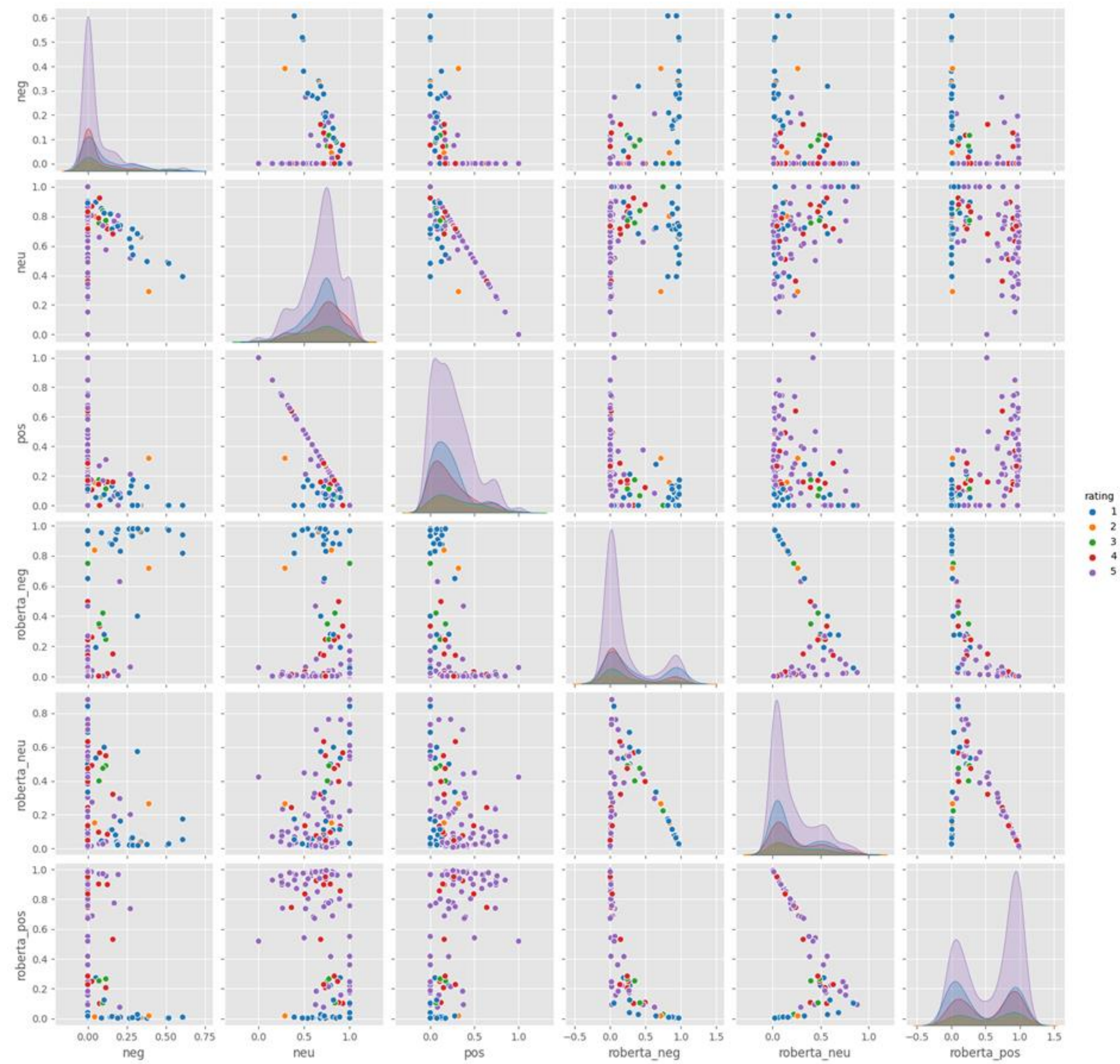

Fig. 18: RoBERTa and Vader model variables pairplots

Our analysis revealed several key findings. First, we found that the sentiment of the reviews varied significantly across different product categories. For example, hair and skin care products received more positive reviews than electronic gadgets, which had a more neutral sentiment overall.

Second, we found that the sentiment analysis results varied significantly between the Vader and Roberta models. While the Vader model provided a more straightforward and rule-based approach to sentiment analysis, the Roberta model provided a more nuanced and context-specific analysis of the text.

Finally, our study demonstrated the potential applications of sentiment analysis in supply chain management. By analyzing customer reviews, companies can gain insights into customer preferences, identify areas for improvement, and make informed decisions about product development, pricing, and marketing.

Case study and Industry examples:

1) Leveraging NLP for Improving Supply Chain Management in the Ride-Hailing Industry: A Case Study of Ola

In handling customer calls and resolving numerous problems relating to Ola's services, customer support representatives experience difficulties that are highlighted in this case study.

With an average response time of 3 minutes per call, a typical customer care agent deals with 90 customer calls every day. The main duties include attending to client complaints, resolving problems, and offering the best solutions based on the company's policies. The nature of customer inquiries covers a wide range of topics, including driver-related issues (e.g., refusal of pickup, dropping off more than 500 metres from the destination, unsafe driving, bad mood), technical problems (e.g., faulty odometer, vehicle breakdown midway), service-related issues (e.g., another driver showing up for pickup, driver refusing to cancel the ride), billing-related questions (e.g., incorrect cancellation charges, bill breakup), and a variety of other issues. Additionally, there are cases involving accidents that also require prompt resolution.

It is essential to look into technology developments that might speed up the process and improve customer satisfaction in order to handle the difficulties encountered in managing customer calls properly. The use of Natural Language Processing (NLP) methods to enhance Ola's supply chain management is one such area of focus.

NLP has important potential applications in Ola's supply chain management. Several areas of the supply chain can be optimised by utilising the huge amounts of customer call data that can be analysed using NLP approaches, including:

Sentiment analysis: SA enables Ola to see trends and obtain a better grasp of customer satisfaction levels by analysing the sentiment that customers convey throughout their calls. Sentiment analysis can be used to pinpoint critical supply chain improvement areas including driver behaviour, service quality, and billing accuracy.

Issue Categorization and Resolution: By using NLP models, customer concerns and questions can be automatically classified, allowing for effective routing to the appropriate department or customer support agent. In order to facilitate quicker and more reliable issue resolution, NLP can also help in generating ideas or automated responses based on predefined resolution parameters.

Customer Feedback Analysis: NLP methods can be used to analyse customer feedback gathered through a variety of sources, like calls, surveys, and social media. This research can offer insightful information about areas that could use improvement, empowering Ola to make data-driven decisions to optimise its supply chain operations.

Demand Forecasting and Planning: NLP can help with demand forecasting and planning by handling client calls and extracting pertinent data. Ola can better predict and address future supply chain difficulties by analysing client preferences, complaints, and suggestions. This will increase operational effectiveness and boost customer happiness.

Real-time Monitoring and Alerts: NLP algorithms have the ability to track consumer calls in real-time and automatically indicate urgent problems. This proactive approach can reduce service interruptions, stop possible escalations, and guarantee prompt redress for consumer complaints.

Knowledge Base Development: Ola may create a comprehensive knowledge base with commonly asked questions, standard operating procedures, and best practices for customer support employees by utilizing NLP approaches.

Review segregation: This review segregation procedure may offer insightful information on the particular specializations or areas of expertise of each customer service employee. The system may intelligently assign cases to agents who are best suited to address particular customer difficulties by taking into account their backgrounds, abilities, and experience. A representative with experience in solving technical issues might be given the case, for instance, if a customer review relates to a technical problem. A person with experience in billing and financial issues can be given the case if a review finds billing inconsistencies.

2) Improving Supply Chain Management of Sprint Mobile Communications (now merged with T-mobile) through Natural Language Processing

The case study highlights the difficulties that customer care agents encounter when handling discussions with customers and resolving various problems pertaining to Sprint's services.

The researcher claimed that they assist Sprint clients via chat while simultaneously resolving two different customer issues. The issues with customers cover a wide range of subjects, such as number activation, credit score-based payment plans, bill splitting inquiries, optimal family plan recommendations, offering various plans based on income ranges (for example, offering special military plans to qualified customers), educating customers about new product offerings, and fixing signal problems, among others.

It is crucial to look at technological improvements that can expedite procedures and boost customer satisfaction in order to manage customer service operations and improve supply chain management. There are several ways that NLP could be used in Sprint's supply chain management. Sprint can enhance the following things by utilising NLP techniques:

Chat Analysis and Categorization: NLP can be used for conversation analysis and categorization, classifying client chats into distinct subjects or problems. Sprint can effectively route customer chats to the proper customer care agents with the necessary skills by automatically categorising customer enquiries into predetermined categories, such as activation, payment, invoicing, plans, and technical assistance.

Sentiment Analysis: Sprint can analyse client conversation sentiment using NLP approaches, allowing it to measure customer satisfaction levels and pinpoint areas that need work. Sprint can better understand consumer pain spots, spot re-occurring problems, and prioritise efforts to address them by analysing sentiment.

Knowledge Base Development: By examining customer interactions and identifying pertinent data, NLP can make it easier to create an extensive knowledge library. Using this knowledge base as a resource for commonly asked questions, how-to manuals, and best practises will enable customer service agents to deliver precise and reliable support.

Automated Responses and Suggestions: By examining customer interactions and identifying pertinent data, NLP can make it easier to create an extensive knowledge library. Using this knowledge base as a resource for commonly asked questions, how-to manuals, and best practises will enable customer service agents to deliver precise and reliable support.

Customer Insights and Demand Forecasting: NLP can offer useful insights into client preferences, wants, and problem issues by examining customer interactions. These revelations could guide Sprint's decision-making, leading to more effective tactics for demand forecasting, inventory control, and product development.

Continuous Improvement and Training: NLP approaches can be used to find common problems and places where customer care agents might need more guidance or training. By identifying knowledge gaps through customer conversation analysis, Sprint is able to tailor training initiatives and raise the level of expertise throughout its customer care workforce.

Sprint can boost customer happiness, optimise customer service processes, and manage its supply chain more effectively by utilising NLP. Sprint can optimise its resources, shorten response times, and provide individualised and effective customer experiences through chat analysis, sentiment analysis, knowledge base building, automated responses, customer insights, and continuous improvement projects.

3) Leveraging Natural Language Processing for Enhancing Supply Chain Management in an EdTech Company

Let's consider the utilization of NLP techniques to improve supply chain management in an edtech company.

a) Doubt Solving: Students in the online classes could ask questions and get clarifications using the chatbox function. It can take a while to individually evaluate and address each query, though. The requirement to delete any offensive content added to the moderators' workload.

b) Repetitive Questions: Students frequently have the same questions in reaction to the same video, necessitating repeated efforts to answer them. If the system could recognise and present previously answered questions, this inefficiency might be overcome.

c) Subject Relevance: It was crucial to make sure that questions were sent to the appropriate subject team because the edtech company offers courses in a variety of disciplines. Manual triage of doubts can be time-consuming and error-prone.

d) Chatbox Spam:Some students took advantage of the chatbox feature by spamming unrelated stuff, which interfered with others' ability to learn.

e) Personalized Course Recommendations: Based on the information submitted by students, the company hoped to provide personalised course recommendations; however, this process required manual analysis and could use automation.

NLP Solutions:

To address these challenges, the edtech company can leverage NLP techniques as follows:

a) Automated Doubt Solving: The business can create a system that automatically processes and replies to textual doubts by utilising NLP algorithms. Implementing a system to detect pornographic content can also speed up the moderation process.

b) Question Similarity Analysis: NLP models can be used to examine how frequently students ask questions that are similar to one another. This would allow the system to display questions that have already been answered, cutting down on repetition and giving teachers more time back.

c) Subject Classification: NLP can help automatically classify doubts according to their subject. This avoids the need for manual triage by directing pertinent questions straight to the appropriate subject team.

d) Chatbox Spam Detection: To provide a conducive learning environment, NLP algorithms can be used to recognise spam communications and take the relevant action, such as imposing temporary or permanent limitations.

e) Personalized Course Recommendations: In order to create personalised course recommendations automatically, the system can employ NLP to analyse the data entered by students. The recommendation process would be more effective and efficient as a result.

Supply Chain Management Benefits:

The supply chain management of the edtech company may experience a number of advantages, such as:

a) Increased Efficiency: By automating the time spent on repetitive operations, such as question similarity analysis and subject classification, the company may free up teachers' time to concentrate more on providing high-quality instruction.

b) Improved Student Experience: The application of NLP-powered systems will result in quicker doubt clarification, a decrease in pointless inquiries, and subject-based responses that are pertinent. In the end, this will improve students' overall learning experiences.

c) Optimal Resource Allocation: The edtech organisation may allocate resources more efficiently and make sure that subject specialists are used by automating the doubt triaging and course suggestion processes.

.4) Leveraging Natural Language Processing for Enhanced Supply Chain Management in Amazon (an e-Commerce website):

Let's explore the application of Natural Language Processing (NLP) techniques to improve supply chain management in Amazon. This case study is based on 1000+ web-scraped Amazon reviews dataset of Indian products of companies such as Godrej, Titan, Maaza, Paper, Indiana, Natural, Maggi, Glucon-D, Amul, Patanjali, Dettol, Savlon, Cinthol, Britannia, NutriChoice, Streax, Himalaya, Society, Tata, Fastrack, Reflex, Mysore, etc. on the website.

Challenges Identified:

This case study is based on the research after analyzing customer reviews of Indian products on the platform. The findings have significant implications for our field esp. in case of industrial engineering.:

a) Customer Sentiment Analysis: For the purpose of identifying areas for improvement and addressing potential supply chain problems, it is essential to comprehend consumer sentiment. Large-scale manual customer review analysis is time-consuming and prone to mistakes.

b) Product Quality Assessment: It's crucial for preserving consumer happiness to evaluate product quality based on client feedback. Traditional methods for evaluating a product's quality based on customer reviews might not be scalable or effective.

c) Supply Chain Performance Evaluation: Without automated analysis tools, it might be difficult to glean information regarding supply chain performance from customer reviews, such as delivery delays, packaging problems, or broken goods.

NLP Solutions:

The researcher suggested integrating NLP techniques into Amazon's supply chain management system to address these issues:

a) Sentiment Analysis: NLP algorithms can be used to analyse the sentiments expressed in customer reviews and classify them as neutral, positive, or negative. This automated process can discover possible supply chain improvement opportunities and offer insightful information about consumer perceptions.

b) Aspect-Based Sentiment Analysis: Aspect-based sentiment analysis, which enables the identification of particular features of the product or supply chain that customers mention in their reviews, can be performed using NLP approaches. When used to guide focused adjustments, this study can provide a detailed insight of client happiness or discontent.

c) Quality Assessment: NLP models can be taught to evaluate the quality of products based on consumer reviews. The system may give quantitative and qualitative assessments of product quality by finding patterns and keywords linked to favourable or unfavourable product experiences. This enables proactive measures to be made to solve any shortcomings.

d) Supply Chain Performance Evaluation: NLP algorithms can extract pertinent data from customer reviews to assess the effectiveness of various supply chain components, including packaging, delivery time, and customer service. Through the identification of bottlenecks and potential improvement areas, the performance of the supply chain as a whole can be improved.

Supply Chain Management Benefits:

Using NLP technologies in Amazon's supply chain management has the following major advantages:

a) Real-time monitoring of customer sentiment and feedback is possible thanks to sentiment analysis enabled by NLP. As a result, Amazon has access to crucial information about consumer preferences and can take quick action to resolve any problems and raise customer happiness.

b) Proactive Issue Resolution: Amazon can pinpoint specific supply chain points that need attention using aspect-based sentiment analysis. This proactive strategy enables focused enhancements, improving client experiences and lowering the possibility of disruptions.

c) Data-driven Decision Making: The supply chain management process benefits from data-driven decision making by utilising NLP tools to analyse customer evaluations. Amazon can improve operations, inventory management, and supplier relationships by spotting patterns and trends.

d) Continuous Improvement: Including NLP in supply chain management ensures a constant customer feedback loop. This feedback enables Amazon to refine and improve its operations, leading to an iterative process of supply chain optimisation.

Conclusion

The integration of NLP-based sentiment analysis into supply chain management has been examined in this thesis, with a focus on Indian products sold on the Amazon platform.com. It has been shown that NLP can actually play a vital role in enhancing supply chain management in several sectors through a thorough research of four different businesses.

The results of this study have shown how NLP-based sentiment analysis might improve supply chain procedures. Companies may better understand how consumers feel about their products by gleaning insightful information from customer reviews and feedback. They can use this knowledge to make informed choices about product creation, inventory management, and demand forecasting.

The capability to proactively identify and address customer problems and difficulties is one of the main advantages of incorporating NLP-based sentiment analysis into supply chain management. Companies can rapidly spot any issues with the product or supply chain by examining the opinions stated in reviews. This makes it possible for them to act immediately to fix problems, so raising customer happiness and enhancing supply chain efficiency.

Additionally, NLP-based sentiment analysis can contribute significantly to demand forecasting. Companies can learn about shifting consumer preferences and new trends by examining customer sentiment and opinions. This information can help with precise demand forecasts, enabling improved inventory management and lowering expenses related to stockouts or surplus inventory.

The assessment and selection of suppliers is another area where NLP may improve supply chain management. Companies can learn more about the reputation and calibre of their suppliers by examining consumer attitudes towards particular products and brands. Making informed decisions on supplier alliances using this information will result in a reliable and effective supply chain network.

It is crucial to remember that a successful supply chain management implementation of NLP-based sentiment analysis requires the right technical setup and knowledge. To successfully extract valuable insights from massive amounts of textual data, businesses must invest in cutting-edge NLP tools, machine learning algorithms, and data analytics skills.

Despite the fact that this study concentrated on Indian products sold on Amazon, its conclusions apply to supply chain management across a range of sectors and places. Businesses in a variety of industries might benefit from integrating NLP-based sentiment analysis to obtain a competitive advantage and improve customer satisfaction.

As a result, supply chain management has a major opportunity to enhance operational effectiveness, customer happiness, and decision-making by integrating NLP-based sentiment analysis. Companies may improve demand forecasts, optimise supplier relationships, and take a proactive approach to supply chain difficulties by utilising the power of NLP. NLP-based sentiment analysis is emerging as a viable tool for success as firms continue to negotiate the dynamic and competitive world of supply chain management.

References

[1] U. B. Mahadevaswamy and P. Swathi, "Sentiment Analysis using Bidirectional LSTM Network Sentiment Analysis using Bidirectional LSTM Network," *Procedia Comput. Sci.*, vol. 218, pp. 45–56, 2023, doi: 10.1016/j.procs.2022.12.400.

[2] J. L. Arroyo-Barrigüete, "Sentiment analysis of Lovecraft's fiction writings," *Heliyon*, vol. 9, no. 1, p. e12673, 2023, doi: 10.1016/j.heliyon.2022.e12673.

[3] O. S. Chlapanis, G. Paraskevopoulos, and A. Potamianos, "Adapted Multimodal BERT with Layer-wise Fusion for Sentiment Analysis," 2022. Available: http://arxiv.org/abs/2212.00678.

[4] A. Singh, M. Jenamani, J. J. Thakkar, and N. P. Rana, "Quantifying the effect of eWOM embedded consumer perceptions on sales: An integrated aspect-level sentiment analysis and panel data modeling approach," *J. Bus. Res.*, vol. 138, no. September 2020, pp. 52–64, 2022, doi: 10.1016/j.jbusres.2021.08.060.

[5] C. Mello, G. S. Cheema, and G. Thakkar, "Combining sentiment analysis classifiers to explore multilingual news articles covering London 2012 and Rio 2016 Olympics," *Int. J. Digit. Humanit.*, 2022, doi: 10.1007/s42803-022-00052-9.

[6] G. Kaur and A. Sharma, "A deep learning-based model using hybrid feature extraction approach for consumer sentiment analysis," *J. Big Data*, vol. 10, no. 1, 2023, doi: 10.1186/s40537-022-00680-6.

[7] P. Savci and B. Das, "Prediction of the customers' interests using sentiment analysis in e-commerce data for comparison of Arabic, English, and Turkish languages," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 35, no. 3, pp. 227–237, 2023, doi: 10.1016/j.jksuci.2023.02.017.

[8] M. Ojeda-Hernández, D. López-Rodríguez, and Á. Mora, "Lexicon-based sentiment analysis in texts using Formal Concept Analysis," *Int. J. Approx. Reason.*, vol. 155, pp. 104–112, 2023, doi: 10.1016/j.ijar.2023.02.001.

[9] A. Balahur, J. M. Hermida, and A. Montoyo, “Detecting Implicit Expressions of Sentiment in Text Based on Commonsense Knowledge,” *Proc. 2nd Work. Comput. Approaches to Subj. Sentim. Anal. WASSA 2011*, pp. 53–60, 2011, [Online]. Available: http://www.aclweb.org/anthology/W11-1707.

[10] E. Altamimi and E. Altamimi, “Text Analysis of Airline Tweets by,” 2022.

[11] A. PATEL, P. OZA, and S. AGRAWAL, “Sentiment Analysis of Customer Feedback and Reviews for Airline Services using Language Representation Model,” *Procedia Comput. Sci.*, vol. 218, pp. 2459–2467, 2023, doi: 10.1016/j.procs.2023.01.221.

[12] R. D. Raut, V. S. Yadav, N. Cheikhrouhou, V. S. Narwane, and B. E. Narkhede, “Big data analytics: Implementation challenges in Indian manufacturing supply chains,” *Comput. Ind.*, vol. 125, p. 103368, 2021, doi: 10.1016/j.compind.2020.103368.

[13] S. Luthra, D. Garg, and A. Haleem, “An analysis of interactions among critical success factors to implement green supply chain management towards sustainability: An Indian perspective,” *Resour. Policy*, vol. 46, pp. 37–50, 2015, doi: 10.1016/j.resourpol.2014.12.006.

[14] R. Dubey, D. J. Bryde, C. Blome, D. Roubaud, and M. Giannakis, “Facilitating artificial intelligence powered supply chain analytics through alliance management during the pandemic crises in the B2B context,” *Ind. Mark. Manag.*, vol. 96, no. February, pp. 135–146, 2021, doi: 10.1016/j.indmarman.2021.05.003.

[15] R. Kumar, R. K. Singh, and R. Shankar, “Critical success factors for implementation of supply chain management in Indian small and medium enterprises and their impact on performance,” *IIMB Manag. Rev.*, vol. 27, no. 2, pp. 92–104, 2015, doi: 10.1016/j.iimb.2015.03.001.

[16] Y. Wang, J. Guo, C. Yuan, and B. Li, “Sentiment Analysis of Twitter Data,” *Appl. Sci.*, vol. 12, no. 22, pp. 1–14, 2022, doi: 10.3390/app122211775.

[17] N. V. Babu and E. G. M. Kanaga, “Sentiment Analysis in Social Media Data for Depression Detection Using Artificial Intelligence: A Review,” *SN Comput. Sci.*, vol. 3, no. 1, pp. 1–20, 2022, doi: 10.1007/s42979-021-00958-1.

[18] K. Muduli, K. Govindan, A. Barve, D. Kannan, and Y. Geng, “Role of behavioural factors in green supply chain management implementation in Indian mining industries,” *Resour. Conserv. Recycl.*, vol. 76, pp. 50–60, 2013, doi: 10.1016/j.resconrec.2013.03.006.

[19] S. Luthra, D. Garg, and A. Haleem, “The impacts of critical success factors for implementing green supply chain management towards sustainability: An empirical investigation of Indian automobile industry,” *J. Clean. Prod.*, vol. 121, pp. 142–158, 2016, doi: 10.1016/j.jclepro.2016.01.095.

[20] K. Govindan, K. Muduli, K. Devika, and A. Barve, “Investigation of the influential strength of factors on adoption of green supply chain management practices: An Indian mining scenario,” *Resour. Conserv. Recycl.*, vol. 107, pp. 185–194, 2016, doi: 10.1016/j.resconrec.2015.05.022.

[21] P. T. Narasimha, P. R. Jena, and R. Majhi, “Impact of COVID-19 on the Indian seaport transportation and maritime supply chain,” *Transp. Policy*, vol. 110, no. May, pp. 191–203, 2021, doi: 10.1016/j.tranpol.2021.05.011.

[22] S. Luthra, S. Luthra, and A. Haleem, “Hurdles in Implementing Sustainable Supply Chain Management: An Analysis of Indian Automobile Sector,” *Procedia - Soc. Behav. Sci.*, vol. 189, pp. 175–183, 2015, doi: 10.1016/j.sbspro.2015.03.212.

[23] M. Winter, S. Dopler, J. M. Müller, and A. Zeisler, “Information sharing and multi-tier supply chain management of SMEs in the context of Industry 4.0,” *Procedia Comput. Sci.*, vol. 217, no. 2022, pp. 1378–1385, 2023, doi: 10.1016/j.procs.2022.12.336.

[24] H. Lin, J. Lin, and F. Wang, “An innovative machine learning model for supply chain management,” *J. Innov. Knowl.*, vol. 7, no. 4, 2022, doi: 10.1016/j.jik.2022.100276.

[25] A. Desai and S. Rai, “Knowledge Management for Downstream Supply Chain Management of Indian Public Sector Oil Companies,” *Procedia Comput. Sci.*, vol. 79, pp. 1021–1028, 2016, doi: 10.1016/j.procs.2016.03.129.

[26] K. Soomro, M. N. M. Bhutta, Z. Khan, and M. A. Tahir, “Smart city big data analytics: An advanced review,” *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 9, no. 5, pp. 1–52, 2019, doi: 10.1002/widm.1319.

[27] K. Schouten and F. Frasincar, “Survey on Aspect-Level Sentiment Analysis,” *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 3, pp. 813–830, 2016, doi: 10.1109/TKDE.2015.2485209.

[28] S. Mohan and R. Sunitha, “Survey On Aspect Based Sentiment Analysis Using Machine Learning Techniques,” *Eur. J. Mol. Clin. Med.*, vol. 07, no. 10, 2020.

[29] D. Mathivathanan, D. Kannan, and A. N. Haq, “Sustainable supply chain management practices in Indian automotive industry: A multi-stakeholder view,” *Resour. Conserv. Recycl.*, vol. 128, pp. 284–305, 2018, doi: 10.1016/j.resconrec.2017.01.003.

[30] A. Gupta, G. Prakash, and J. Jadeja, “Supply Chain in the Public Procurement Environment: Some Reflections from the Indian Railways,” *Procedia - Soc. Behav. Sci.*, vol. 189, pp. 292–302, 2015, doi: 10.1016/j.sbspro.2015.03.225.

[31] D. Das, *The impact of Sustainable Supply Chain Management practices on firm performance: Lessons from Indian organizations*, vol. 203. Elsevier Ltd, 2018.

[32] R. Rajesh, “Sustainable supply chains in the Indian context: An integrative decision-making model,” *Technol. Soc.*, vol. 61, p. 101230, 2020, doi: 10.1016/j.techsoc.2020.101230.

[33] A. Mojumder and A. Singh, “An exploratory study of the adaptation of green supply chain management in construction industry: The case of Indian Construction Companies,” *J. Clean. Prod.*, vol. 295, p. 126400, 2021, doi: 10.1016/j.jclepro.2021.126400.

[34] F. Alshuwaier, A. Areshey, and J. Poon, “Applications and Enhancement of Document-Based Sentiment Analysis in Deep learning Methods: Systematic Literature Review,” *Intell. Syst. with Appl.*, vol. 15, no. April, p. 200090, 2022, doi: 10.1016/j.iswa.2022.200090.

[35] A. P and M. M. Patil, “A Review on Data Analytics for Supply Chain Management: A Case study,” *Int. J. Inf. Eng. Electron. Bus.*, vol. 10, no. 5, pp. 30–39, 2018, doi: 10.5815/ijieeb.2018.05.05.

[36] J. Wu, T. Zhu, X. Zheng, and C. Wang, “Multi-Modal Sentiment Analysis Based on Interactive Attention Mechanism,” *Appl. Sci.*, vol. 12, no. 16, 2022, doi:

10.3390/app12168174.

[37] L. Zhang and L. Corporation, “RAAF Richmond,” *A + U-Architecture Urban.*, no. 487, p. 121, 2011.

[38] W. Etaiwi, D. Suleiman, and A. Awajan, “Deep Learning Based Techniques for Sentiment Analysis: A Survey,” *Inform.*, vol. 45, no. 7, pp. 89–95, 2021, doi: 10.31449/inf.v45i7.3674.

[39] M. Birjali, M. Kasri, and A. Beni-Hssane, “A comprehensive survey on sentiment analysis: Approaches, challenges and trends,” *Knowledge-Based Syst.*, vol. 226, p. 107134, 2021, doi: 10.1016/j.knosys.2021.107134.

[40] Q. A. Xu, V. Chang, and C. Jayne, “A systematic review of social media-based sentiment analysis: Emerging trends and challenges,” *Decis. Anal. J.*, vol. 3, no. June, p. 100073, 2022, doi: 10.1016/j.dajour.2022.100073.

[41] M. Suaib, “a Survey on Deep Learning Techniques,” *Strad Res.*, vol. 7, no. 8, pp. 1–4, 2020, doi: 10.37896/sr7.8/037.

[42] K. Govindan, M. Kaliyan, D. Kannan, and A. N. Haq, “Barriers analysis for green supply chain management implementation in Indian industries using analytic hierarchy process,” *Int. J. Prod. Econ.*, vol. 147, no. PART B, pp. 555–568, 2014, doi: 10.1016/j.ijpe.2013.08.018.

[43] H. Hu, J. Xu, M. Liu, and M. K. Lim, “Vaccine supply chain management: An intelligent system utilizing blockchain, IoT and machine learning,” *J. Bus. Res.*, vol. 156, no. November 2022, p. 113480, 2023, doi: 10.1016/j.jbusres.2022.113480.

[44] L. Liu, W. Song, and Y. Liu, “Leveraging digital capabilities toward a circular economy: Reinforcing sustainable supply chain management with Industry 4.0 technologies,” *Comput. Ind. Eng.*, vol. 178, no. April 2022, p. 109113, 2023, doi: 10.1016/J.CIE.2023.109113.

[45] T. Shaik, X. Tao, C. Dann, H. Xie, Y. Li, and L. Galligan, “Sentiment analysis and opinion mining on educational data: A survey,” *Nat. Lang. Process. J.*, vol. 2, no. October

2022, p. 100003, 2023, doi: 10.1016/j.nlp.2022.100003.

[46] Z. Li *et al.*, “Multimodal Sentiment Analysis Based on Interactive Transformer and Soft Mapping,” *Wirel. Commun. Mob. Comput.*, vol. 2022, 2022, doi: 10.1155/2022/6243347.

[47] C. Chen, X. Hu, H. Zhang, and Z. Shou, “Fine grained sentiment analysis based on Bert,” *J. Phys. Conf. Ser.*, vol. 1651, no. 1, 2020, doi: 10.1088/1742-6596/1651/1/012016.

[48] R. R. Menon and V. Ravi, “Analysis of enablers of sustainable supply chain management in electronics industries: The Indian context,” *Clean. Eng. Technol.*, vol. 5, p. 100302, 2021, doi: 10.1016/j.clet.2021.100302.

[49] J. Hartmann, M. Heitmann, C. Siebert, and C. Schamp, “More than a Feeling: Accuracy and Application of Sentiment Analysis,” *Int. J. Res. Mark.*, no. xxxx, pp. 1–13, 2022, doi: 10.1016/j.ijresmar.2022.05.005.

[50] R. Bhaskaran *et al.*, “Intelligent Machine Learning with Metaheuristics Based Sentiment Analysis and Classification,” *Comput. Syst. Sci. Eng.*, vol. 44, no. 1, pp. 235–247, 2022, doi: 10.32604/csse.2023.024399.

[51] D. Suhartono, K. Purwandari, N. H. Jeremy, S. Philip, P. Arisaputra, and I. H. Parmonangan, “Deep neural networks and weighted word embeddings for sentiment analysis of drug product reviews,” *Procedia Comput. Sci.*, vol. 216, no. 2022, pp. 664–671, 2023, doi: 10.1016/j.procs.2022.12.182.

[52] A. Chamekh, M. Mahfoudh, and G. Forestier, “Sentiment Analysis Based on Deep Learning in E-Commerce,” *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 13369 LNAI, pp. 498–507, 2022, doi: 10.1007/978-3-031-10986-7_40.

[53] Z. A. Diekson, M. R. B. Prakoso, M. S. Q. Putra, M. S. A. F. Syaputra, S. Achmad, and R. Sutoyo, “Sentiment analysis for customer review: Case study of Traveloka,” *Procedia Comput. Sci.*, vol. 216, no. 2022, pp. 682–690, 2023, doi: 10.1016/j.procs.2022.12.184.

[54] P. Baby and K. B, “Sentimental Analysis and Deep Learning : A Survey,” *Int. J. Sci. Res. Sci. Eng. Technol.*, pp. 212–220, 2020, doi: 10.32628/ijsrset207135.

[55] Y. Liu *et al.*, "RoBERTa: A Robustly Optimized BERT Pretraining Approach," no. 1, 2019.

[56] Hutto, C., & Gilbert, E. (2014). VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*, *8*(1), 216-225. https://doi.org/10.1609/icwsm.v8i1.14550