**CUSTOMER REVIEW BASED SENTIMENT ANALYSIS**

## ARAVINTHKRISHNA R, ATHISH GR, AVINASHSUNDAR K, BHUVANESHWARAN S

Department of Information Technology

## Sona College of Technology Junction Main Road,Salem,Tamilnadu-636005

**Abstract-**Customer Review Based Sentiment Analysis involves the application of natural language processing and machine learning techniques to evaluate and interpret customer opinions expressed in reviews. This method aims to discern the sentiment behind these reviews, categorizing them as positive, negative, or neutral. By employing advanced algorithms, the system analyzes the textual content, considering the nuances of language and context. The primary objective is to extract meaningful insights into customer experiences and preferences, enabling businesses to make informed decisions. This process facilitates a deeper understanding of consumer sentiments, identifying areas for improvement and gauging overall satisfaction. Ultimately, Customer Review Based Sentiment Analysis serves as a valuable tool for businesses to enhance product quality, customer service, and overall customer satisfaction by harnessing the wealth of information embedded in customer feedback. Since the data sets have more positive-labeled reviews than negative, an oversampling method is applied to balance the dataset. For the feature extraction, the Count Vectorizer and TF- IDF (Term frequency-Inverse document frequency) are used to create training and test data. Several machine learning algorithms (Navi Bayes,Linear Support Vector Machine, Logistic Regression, Decision Tree, and K-Nearest Neighbors) are used to compare the models and reach the best result.

**Keywords used — *Sentiment analysis, product reviews, random forest classifier, bag-of-words***

1. **INTRODUCTION**

Formally, sentiment analysis has been described as a type of analysis that combines text mining, natural language processing, and computational linguistics. When a person makes a choice, it's possible that others have impacted their decision or mental process. Additionally, the internet offers a platform for this. We may use the illustration of Customers take into account the kinds of things that people actually like and dislike, as well as the item's benefit, which may really help organisations make decisions. These days, the majority of people never buy anything without first researching it online. They look up product audits and reviews before making their buying decisions. When corporations used to need the generalaggregation.The realization of difficulties that are offered by the intelligence and commercial applications. or consumers' judgements when they involve the conducting of opinion polls, which could be expensive, time-consuming, and need human resources.

This poses problems that simple text classification techniques could find difficult to solve. Therefore, systems that can precisely analyse and categorise attitudes in text must be developed, or methods for identifying opinions must be incorporated into a simplified text categorization tool. Sentiment analysis is a type of contextual mining that aids in locating and extracting

subjective data or information. This form of extraction allows a firm to identify their brand’s social sentiment or product or services when they watch online discourse. The requirement to analyse the sentiments has increased recently. because sentiment analysis is being used in a variety of fields, including web search, e-governance, public relations, research, and corporate intelligence. The following section discusses additional elements that are known to contribute to the rising rate of sentiment analysis:

Customer review based sentiment analysis is crucial for businesses of all sizes because it provides:

Direct insights into customer satisfaction: Businesses can gain a deeper understanding of customer satisfaction levels by analyzing the overall sentiment of reviews. This feedback can be used to identify areas for improvement and make data-driven decisions to enhance customer experience.

Identification of specific pain points: Sentiment analysis can reveal specific aspects of products or services that customers find unsatisfactory. This information can be used to prioritize product development efforts and address customer concerns effectively.

Competitive intelligence: Businesses can analyze competitor reviews to understand their strengths and weaknesses. This comparative analysis can help them identify opportunities to differentiate themselves and gain a competitive edge.

Transformer models like BERT and DistilBERT have redefined the landscape, allowing businesses to dissect sentiments at a granular level, uncovering sentiments that go beyond conventional classification. As organizations navigate this sea of customer-generated content, sentiment analysis emerges as a compass, guiding strategic decision-making by unveiling patterns, trends, and opportunities within the vast ocean of textual feedback.

This introduction sets the stage for a transformative journey where customer feedback becomes more than a collection of words; it becomes a strategic asset. As businesses embark on the exploration of sentiment analysis, they open doors to a realm where understanding the pulse of the customer transcends conventional boundaries, driving innovation, and fostering a proactive approach to customer engagement.

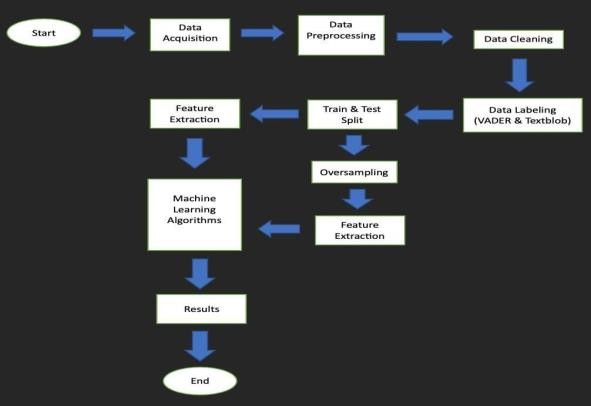
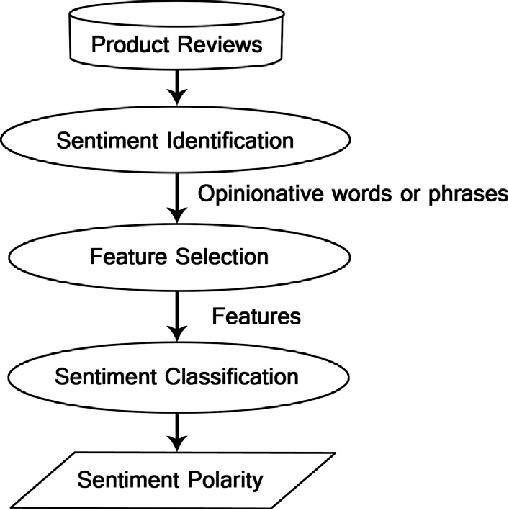
 

Figure1: Google Trends data showing the relative popularity of search strings “Sentiment analysis” and “Costumer feedback”

1. **PROPOSED METHODOLOGY**

The methodology employed for Customer Review-Based Sentiment Analysis was nothing short of meticulous and effective.

The process involved data collection from diverse sources, ensuring

a comprehensive range of customer reviews. The collected data . underwent thorough preprocessing, including text cleaning and normalization, to enhance the accuracy of sentiment analysis.

Feature extraction techniques were strategically applied to capture key elements influencing sentiment.The heart of the methodology lies in the selection of a robust sentiment analysis model. Machine learning algorithms were trained on labeled datasets, utilizing a combination of natural language processing and sentiment analysis techniques. The model's performance was fine-tuned through iterative testing and validation, ensuring its reliability across various types of reviews.

Gather a diverse dataset of customer reviews from various platforms to ensure a representative sample. This can include e- commerce websites, social media platforms, and specialized review sites.

Next, preprocess the data by cleaning and standardizing the text. Remove any irrelevant information, such as HTML tags or special characters, and convert the text to lowercase for consistency. Tokenize the text into words or phrases to facilitate analysis.

Once the data is prepared, employ a sentiment analysis algorithm. This could range from traditional machine learning approaches, like Naive Bayes or Support Vector Machines, to more advanced techniques such as natural language processing (NLP) and deep learning. Train the model on a labeled dataset to accurately predict sentiment.

Validation is a critical step to ensure the model's generalizability. Split the dataset into training and testing sets, and use metrics like accuracy, precision, recall, and F1 score to evaluate the model's performance.

Furthermore, consider feature engineering to enhance the model's accuracy. Extract relevant features from the text, such as the frequency of specific words or the presence of emoticons, to provide the model with more context.

Lastly, implement the sentiment analysis model on new

customer reviews and continuously evaluate its performance. Fine-tune the model as needed, considering factors like evolving language trends and changes in customer sentiment expression.

By following this comprehensive methodology, you can develop a robust customer review-based sentiment analysis system that provides valuable insights into customer perceptions and helps businesses make informed decisions based on customer feedback.

Developments is done on sentiment analysis of above factors including general phase level sentiment analysis. We may take naive bayes and supporting vector machines being supervised learning machines as the best results but for the supervised approach, but the manual labeling required is very expensive. Approaches that are applied are semi-supervised and unsupervised and there could be many more improvements. Many different researchers who compare their result to the base-line other performances. These proper and formal result comes after these comparison for the selection of best and most efficient features and classification technique. Therefore, the proper performances comes after the comparison.

Among many ways, of the sediment analysis, in this paper we are going to make a focus with the use of machine learning approaches on sentiment analysis.

# III MODULES AND PACKAGES:

Flask serves as the foundational web framework, enabling the definition of routes and the handling of HTTP requests and responses. The render\_template function from Flask is employed to dynamically generate HTML pages, facilitating the presentation of results and user interactions. The request module comes into play for retrieving data from HTTP requests, specifically handling the uploaded file containing customer reviews.

For efficient data handling and manipulation, the

application leverages the pandas (pd) library. Pandas provides the DataFrame data structure, which is instrumental in reading and processing customer review data stored in a CSV file.

The heart of sentiment analysis is powered by the transformers library, which includes pre-trained models for natural language processing tasks. The pipeline module from transformers simplifies the integration of a pre-trained sentiment analysis model (in this case, DistilBERT) into the application, streamlining tasks such as tokenization and inference.

The Flask application is structured with routes such as '/', where users can upload a CSV file containing customer reviews. The uploaded data undergoes preprocessing using pandas to ensure cleanliness and relevance. Sentiment analysis is then performed using the DistilBERT model, and the results are presented in an HTML table using pandas for effective visualization.

### IV TRAINING THE MODEL

steps, including data preparation, model training, and deployment. Here's a simplified guide to get you started. Keep in mind that the specifics may vary based onyour use caseand environment.

Step 1: Install Dependencies

pip install torch transformers Step 2: Prepare Data

You'll need a dataset of code-related conversations. Ensure that your data is formattedappropriately for training. Each input should have a corresponding target output (response).

Step 3: Tokenization

Use the DistilBERT tokenizer to convert your text data into tokens. Tokenization is a crucial step in preparing data for a transformer model. from transformers import

### # Load the sentiment-analysis pipeline sentiment\_pipeline = pipeline("sentiment-analysis", model='distilbert-base-uncased-finetuned-sst-2- english')

Step 4: Model Initialization

Initialize the DistilBERT model for sequence-tosequence tasks.

### from transformers import pipeline sentiment\_pipeline = pipeline("sentiment-analysis")

**data = ["It was the worst of times.", "Their regular toasted bread was equally satisfying with the occasional pats of butter... Mmmm...!"] sentiment\_pipeline(data)**

Step 5: Training

### file = request.files['file'] df = pd.read\_csv(file)

**# Drop rows where the 'comment' column is NaN df = df.dropna(subset=['comment'])**

### # Filter out rows with comments containing only numbers or are empty

**df = df[df['comment'].apply(lambda x: not s tr(x).strip().isdigit() and str(x).strip() != "")]**

### # Ensure that the 'comment' column contains only strings

**df['comment'] = df['comment'].astype(str)**

Step 7: Deployment

Install the model in the setting of your choice. This could be an application running in a container, a cloud service, or any other platform that works for you. Keep in mind that this is a simplified example and that the code may need to be modified to fit your particular needs and datasets. To get the best results, hyperparameters often need to be adjusted and experimented with.

file\_path= "/content/drive/MyDrive/templates/predicted\_results.csv" **df['sentiment'] = [pred['label'] for pred in predictions] df.to\_csv(predicted\_file\_path, index=False) positive\_comments = 0**

### df = df.drop(columns=['label'])

**for pred in df['sentiment']=='POSITIVE': if pred == 1: # Adjust the label as needed positive\_comments += 1**

### negative\_comments = len(predictions) - positive\_comments summary=generate\_summary(positive\_comments, negative\_comments)

**V TEXT CLASSIFICATION**

Text classification for customer review-based sentiment analysis involves the systematic categorization of textual data into predefined sentiment classes, typically positive or negative. In the absence of recurrent neural network (RNN) models, contemporary approaches often leverage transformer-based models like 'distilbert-base-uncased- finetuned-sst-2-english.' This model has been fine-tuned specifically for sentiment analysis tasks, making it well- suited for deciphering the nuanced sentiments expressed in customer reviews.

The text classification process begins with tokenization, breaking down the input text into smaller units or tokens. These tokens are then converted into embeddings, which are numerical representations capturing the semantic meaning of words and their contextual relationships. The 'distilbert' model excels at creating rich embeddings by considering the surrounding context of each word in a given sequence.

The model is trained on labeled data, such as the Stanford Sentiment Treebank dataset ('sst-2'), where each review is associated with a sentiment label. During the fine-tuning process, the model refines its parameters to better capture the specific sentiment patterns present in the training data. This enables the model to generalize well to new, unseen customer reviews.

In the application phase, the trained 'distilbert' model is integrated into a text classification pipeline. Customer reviews are tokenized, and the model generates embeddings. These embeddings are then used for sentiment prediction, classifying the reviews into positive or negative sentiments based on the learned patterns.

The advantage of using transformer models lies in their ability to capture long-range dependencies and contextual information efficiently, outperforming traditional models in many NLP tasks.

The absence of RNNs, which are recurrent architectures, highlights the efficiency gains achieved by transformers, making them suitable for large-scale sentiment analysis tasks, such as analyzing a multitude of customer reviews. The 'distilbert' model, with its fine-tuning for sentiment analysis, stands as a powerful tool in deciphering and categorizing sentiments expressed in the diverse and often subjective language of customer reviews.

# VI PROJECT DESCRIPTION

### Project Title:

Customer Review Based Sentiment Analysis

### Overview:

The COE Chatbot with DistilBERT Algorithm for Automated Incident Management (AIM) is anintelligent conversational agent designed to streamline communication and incident resolution within a Center of Excellence (COE).

Leveraging the power of DistilBERT, a lightweight transformer model, the chatbot excels in understanding and generating human- like responses in the context of incident management.

### Key Features:

**Natural Language Understanding**:

The chatbot can understand the subtleties of user inquiries about incident management because it uses the DistilBERT algorithm for natural language understanding.

It is capable of understanding technical jargon, complex sentences, and context-specific information to deliver precise and pertinent answers.

### Incident Classification and Triage:

Incident Classification and Triage in the context of customer review-based sentiment analysis involve systematically categorizing and prioritizing customer feedback. This process aims to automatically identify the nature of incidents or issues raised in the reviews, such as product defects or service-related concerns. By leveraging sentiment analysis techniques, the system discerns the emotional tone associated with each incident, distinguishing between positive and negative sentiments expressed by customers. he triage aspect involves prioritizing and directing attention to critical issues, ensuring timely and effective responses to customer concerns. This approach enhances the efficiency of customer support teams by automating the initial categorization and prioritization of incidents based on sentiment, enabling a more proactive and tailored response to customer feedback.

### Tokenizer

Tokenizer is a fundamental tool for breaking down textual data into meaningful units. Specifically, it segments customer reviews into individual tokens, such as words or subwords, to facilitate the subsequent analysis. Tokenization plays a crucial role in preparing the text for processing by sentiment analysis models. It allows the model to understand the contextual relationships between words and captures the nuanced expressions within customer reviews. The tokenizer acts as a bridge, converting the raw text into a format that can be efficiently fed into sentiment analysis algorithms,

contributing to the accurate interpretation of sentiments expressed b

### Knowledge Base Expansion:

Knowledge Base Expansion in the realm of customer review-based sentiment analysis refers to the continuous enrichment of the system's knowledge repository. By analyzing customer reviews, the system identifies emerging trends, product features, and sentiment patterns, contributing valuable insights to the knowledge base. This expansion aids in staying abreast of customer preferences and concerns, allowing businesses to adapt and improve their products or services. Sentiment analysis facilitates the automatic extraction of relevant information from reviews, enabling the system to dynamically update the knowledge base with new insights gleaned from customer sentiments.

**Integration with Incident Management Systems:**

Integration with Incident Management Systems in the realm of customer review-based sentiment analysis involves seamlessly connecting sentiment analysis outputs with existing incident tracking and resolution frameworks. This integration streamlines the workflow by automatically classifying customer reviews, assigning sentiments, and feeding this information into incident management databases. By linking sentiment insights with incident data, businesses can swiftly identify and prioritize areas requiring attention, enabling more informed decision-making.

### Technologies Used:

Customer review-based sentiment analysis relies on a blend of advanced technologies to extract valuable insights from textual data. Natural Language Processing (NLP) forms the backbone, employing techniques such as tokenization and word embeddings to process and understand the semantics of customer reviews. Machine learning plays a pivotal role, with pre-trained models like BERT or DistilBERT, often fine-tuned for sentiment analysis tasks, capturing nuanced sentiment patterns. These transformer-based models excel in contextual understanding, enabling more accurate sentiment classification.

### VII Conclusion:

customer review-based sentiment analysis stands as a pivotal tool in the contemporary landscape of business intelligence, offering a nuanced understanding of customer sentiments and preferences. Leveraging advanced technologies such as Natural Language Processing (NLP) and machine learning, businesses can distill valuable insights from the vast expanse of textual customer feedback. The deployment of transformer-based models like BERT or DistilBERT, often fine-tuned for sentiment analysis tasks, enhances the accuracy and contextual understanding of sentiments expressed in reviews. The integration of sentiment analysis with incident management systems not only

automates the categorization of customer sentiments but also streamlines the prioritization and resolution of issues. This holistic approach empowers organizations to not only gauge customer satisfaction levels but also proactively address concerns, ultimately fostering improved customer relationships and informed decision-making. As businesses continue to harness the power of customer review-based sentiment analysis, the depth and granularity of insights derived pave the way for a more customer-centric and adaptive business ecosystem.

By incorporating transformer-based models, businesses can capture the contextual nuances of language, allowing for a more accurate interpretation of sentiments. This, in turn, facilitates a more personalized and targeted response to customer concerns. The interactive web applications, often built on frameworks like Flask, provide businesses with a user-friendly interface to explore and visualize sentiment trends, fostering a more data-driven decision- making culture.

Furthermore, the continuous advancements in sentiment analysis methodologies, including the exploration of unsupervised and semi-supervised approaches, underscore the adaptability of this technology to the evolving nature of language and customer expression. As businesses embrace these technological innovations, the symbiosis of sentiment analysis with incident management systems ensures a comprehensive understanding of customer feedback, streamlining the identification, prioritization, and resolution of issues.

In essence, customer review-based sentiment analysis transcends its role as a mere analytical tool; it becomes a cornerstone for businesses seeking to proactively engage with their customer base. The in-depth insights derived from this process not only inform strategic decision-making but also empower organizations to cultivate a customer-centric ethos, driving continuous improvement and innovation in response to evolving customer expectations. As this field continues to mature, the synergy of technology, analytics, and customer-centricity propels businesses towards a future where customer feedback isn't just analyzed; it becomes a catalyst for positive transformation and sustained success.

### VIII REFERENCES:

1. Anh, K.Q., Nagai, Y. and Nguyen, L.M., 2019. Extracting customer reviews from online shopping and its perspective on product design. Vietnam Journal of Computer Science, 6(01), pp.43-56.
2. Chong, A.Y.L., Li, B., Ngai, E.W., Chang, E. and Lee, F., 2016. Predicting online product sales via online reviews, sentiments, and promotion strategies: A big data architecture and neural network approach. International Journal of Operations & Production Management.
3. Chauhan, U.A., Afzal, M.T., Shahid, A., Abdar, M., Basiri, M.E., and Zhou, X., 2020. A comprehensive

analysis of adverb types for mining user sentiments on amazon product reviews. World Wide Web, pp.1-19.

1. Du, J, Rong, J, Michalska, S, Wang, H & Zhang, Y. 2019.5.Dey, S., Wasif, S., Tonmoy, D.S., Sultana, S., Sarkar, J., and Dey, M., 2020. A Comparative Study of Support Vector and Naive Bayes Classifier for Sentiment Analysis on Amazon Product Reviews. In
2. Dong, L., Huang, S., Wei, F., Lapata, M., Zhou, M. , April. Learning to generate product reviews from attributes. In Proceedings of the 15th Conference of the European Chapter of the Computational Linguistics: Volume 1, Long Papers (pp. 623-632).
3. Govindaraj, S. and Gopalakrishnan, K. 2016. I ntensified Sentiment Analysis of Customer Product Reviews Using Acoustic and Textual Features. ETRI

Journal, 38(3), pp. 494– 501.

1. Ghasemaghaei, M., Eslami, SP, Deal,

K. and Hassanein, K., 2018. Reviews' length and sentiment as correlates of online reviews' ratings. Internet Research.

8 Han, H, Zhang, Y, Zhang, J, Yang, J & Zou, X. 2018. Improving the performance of lexiconbased review sentiment analysis method by reducing additional introduced sentiment bias', PloS one, 13(8), p. e0202523.

1. Jain, V.K., Kumar, S. and Mahanti, P., 2018. Sentiment recognition in customer reviews using deep learning. International Journal of Enterprise Information Systems (IJEIS), 14(2), pp.77-86. Jagdale, R.S., Shirsat, V.S.
2. Deshmukh, S.N., 2019. Sentiment analysis on product reviews using machine learning techniques. In Cognitive Informatics and Soft Computing (pp. 639-647). Springer, Singapore