



UNLOCKING BUSINESS INSIGHTS: STATISTICAL ANALYSIS OF SENTIMENT DATA FOR STRATEGIC MANAGEMENT

¹Mr. Abhishek Ganguly, ²Mr. Jeewesh Mishra, ³Ms. Madhumita Ghosh, ⁴Ms. Mansi Tiwari, ⁵Mr. Ritik Sharma

¹Student, ²Student, ³Student, ⁴Student, ⁵Student

'Institute of Management Studies, Ghaziabad, India'

Abstract : In an era where data-driven decisions are paramount, the project titled "Unlocking Business Insights: Statistical Analysis of Sentiment Data for Strategic Management" seeks to harness the power of sentiment analysis to provide actionable insights for strategic decision-making in modern enterprises. The project centers on the analysis of sentiment data derived from various sources, including customer feedback, social media interactions, and employee surveys. Through rigorous statistical techniques, this research aims to uncover hidden patterns, trends, and Correlations within the sentiment data that can be leveraged by management teams to inform critical business strategies. Analyzing their relationship between sentiment and key performance indicators (KPIs) such as customer satisfaction, employee engagement, and Brand perception and developing predictive models to forecast sentiment changes and their impact on business outcomes as well as recommending data-driven strategies for enhancing customer relations, optimizing marketing efforts, are some aspects that every organization wants to thrive in.

Index Terms -: Sentiment analysis, Strategic decision making, Key Performance indicators, Employee engagement, Social media Interaction, Brand perception

I. INTRODUCTION

Sentiment analysis, also known as opinion mining, is a natural language processing (NLP) technique that involves the use of computational algorithms to determine the sentiment or emotional tone expressed in a piece of text. The primary goal of sentiment analysis is to Automatically classify text as being positive, negative, neutral, or sometimes even more fine-grained emotions like happy, sad, angry, etc. Here are the key components and aspects of sentiment analysis: 1. Text Input: Sentiment analysis typically takes text as input, although it can also be applied to other forms of data, such as social media posts, reviews, comments, or transcripts of spoken language. 2. Sentiment Classification: The core task of sentiment analysis is to classify the sentiment expressed in the text. This classification is often done on a scale, such as: • Positive • Negative • Neutral • Mixed sentiments (when both positive and negative sentiments are present) • Fine-grained sentiment labels (e.g., very positive, somewhat positive, very negative, etc.) 3. NLP Techniques: Sentiment analysis relies on NLP techniques to preprocess and analyze text. This includes tasks like tokenization (breaking text into words or phrases), removing stop words (common words that don't carry much meaning), and applying machine learning algorithms or rule-based methods to classify sentiment.

4. Applications: • Business: Companies use sentiment analysis to gauge customer opinions and feedback, monitor brand perception, and make data-driven decisions for marketing and customer service. • Social Media Monitoring: It is used to track public sentiment on social media platforms, helping organizations understand trends and public opinion. • Product Reviews: Sentiment analysis is applied to analyze product or service reviews to determine customer satisfaction and identify areas for improvement. • Financial Markets: In finance, sentiment analysis can be used to analyze news articles and social media to predict market movements. • Politics: It's used in political campaigns to gauge public sentiment and understand voters' opinions. • Customer Support: Sentiment analysis can assist in routing customer inquiries to the right support agents based on the sentiment of the query. 5. Challenges: Sentiment analysis is a challenging task due to linguistic nuances, context, sarcasm, and cultural differences. Developing accurate sentiment analysis model often requires large labeled datasets and sophisticated algorithms. Sentiment analysis plays a crucial role in various industries and applications by providing valuable insights into public sentiment, customer feedback, and the overall emotional tone of text data, helping organizations make informed decisions and improve their products and services.

Sentiment analysis holds significant importance from a management perspective across various industries and sectors. Here are several key reasons why sentiment analysis is valuable for management: Customer Insights: Product Feedback: Sentiment analysis allows businesses to monitor and analyze customer feedback and reviews. It helps identify product strengths and weaknesses, enabling improvements and product development in line with customer preferences. Market Research: Management can use

sentiment analysis to gain insights into market trends, consumer preferences, and emerging issues, aiding in strategic planning and decision-making. Brand Reputation Management: Real-time Monitoring: Sentiment analysis tools can provide real-time monitoring of brand mentions and sentiment on social media and other online platforms. This enables rapid response to potential reputation threats and the ability to capitalize on positive sentiment. Customer Service Enhancement: Issue Identification: Sentiment analysis helps identify customer issues and concerns quickly. Management can prioritize addressing these concerns, improving customer satisfaction, and reducing churn. Efficiency: Efficiently routing customer inquiries based on sentiment allows companies to provide tailored support, reducing response times and improving customer experiences. Marketing and Advertising: Campaign Assessment: Sentiment analysis helps evaluate the effectiveness of marketing campaigns by measuring consumer reactions and sentiment changes. Management can adjust strategies based on this feedback. Competitor Analysis: Companies can use sentiment analysis to gauge the sentiment around their competitors, identifying opportunities to gain a competitive edge. Financial Markets : Trading Decisions: Sentiment analysis can be applied in financial markets to gauge market sentiment from news, social media, and financial reports. This data can in form of investment decisions and risk management. Product Development: Feature Prioritization: Sentiment analysis can guide product development by revealing which Features or improvements are most desired by customers. Early Warning System: It can act as an early warning system for identifying potential issues with new products or services, allowing for proactive adjustments. Employee Satisfaction and Feedback: Employee Engagement: Sentiment analysis can be applied to internal communication channels to gauge employee satisfaction, identify concerns, and improve overall employee engagement. HR Decision Support: HR management can use sentiment analysis to make informed decisions related to employee well-being, training needs, and workplace improvements. Political and Social Impact: Public Opinion: Sentiment analysis helps gauge public sentiment regarding political candidates, policies, and social issues. This is valuable for political campaigns, public relations, and policy-making. Risk Management: Risk Assessment: In industries like insurance and finance, sentiment analysis can assist in assessing risks by monitoring sentiment-related indicators that might impact financial stability or market behavior.

II. Need of the Study:

Strategic Planning: Data-Driven Decision-Making: Sentiment analysis provides data-driven insights that can inform strategic planning, helping organizations adapt to changing market conditions and consumer sentiments. In summary, sentiment analysis empowers management by providing valuable insights into customer opinions, market trends, brand reputation, and employee satisfaction. These insights enable organizations to make data-driven decisions, enhance customer experiences, mitigate risks, and stay competitive in today's dynamic business environment. In sentiment analysis, the concept of "polarity" refers to the emotional or tonal orientation of a piece of text, indicating whether the sentiment expressed is positive, negative, neutral, or somewhere in between. Polarity is a fundamental aspect of sentiment analysis and helps determine the overall sentiment or emotional tone of the text. Here are the primary polarity categories in sentiment analysis: Positive Polarity: Text with positive polarity expresses favorable or optimistic sentiments. It often includes words, phrases, or expressions that convey happiness, satisfaction, approval, or enthusiasm. Examples of positive polarity sentences: "I love this product; it's amazing!" "The weather is perfect today." Negative Polarity: Text with negative polarity conveys unfavorable or pessimistic sentiments. It typically contains words, phrases, or expressions associated with dissatisfaction, disappointment, criticism, or unhappiness. Examples of negative polarity sentences: "The service at the restaurant was terrible." "I'm very disappointed with this purchase." Neutral Polarity: Text with neutral polarity lacks strong positive or negative sentiment. It often consists of Factual statements or descriptions that do not express emotion or opinion. Examples of neutral polarity sentences: "The sky is blue." "The book contains 300 pages." Mixed or Ambiguous Polarity: Some text can have mixed polarity, where both positive and negative sentiments are present within the same text, making it challenging to categorize as purely positive or negative. Example of mixed polarity sentence: "I like the product's features, but the customer service needs improvement." In practice, sentiment analysis algorithms use various techniques, including machine learning, lexicon-based analysis, and rule-based approaches, to classify text into one of these polarity categories. Lexicon-based methods rely on sentiment lexicons or dictionaries containing words and their associated sentiment scores (e.g., a positive word like "happy" might have a positive score, while "sad" would have a negative score). Machine learning approaches, on the other hand, learn from labeled training data to predict polarity based on contextual patterns and features within the text. Polarity analysis is essential for understanding how individuals perceive and express emotions or opinions in textual data, making it a valuable tool for businesses, social media monitoring, customer feedback analysis, and various other applications where assessing Sentiment is crucial

III. Literature Review

A survey on sentiment analysis methods, applications, and challenges

By Mayur Wankhade Annavarapu Chandra Sekhara Rao Chaitanya Kulkarni

An extensive description of the process for accomplishing this assignment and the uses of sentiment analysis are covered in this article. After that, it assesses, contrasts, and explores the methods employed in order to acquire a thorough comprehension of their benefits and drawbacks. In order to determine future directions, the sentiment analysis challenges are finally reviewed.

A Study of Sentiment Analysis: Concepts, Techniques, and Challenges By Ameen Abdullah Qaid Aqlan, B. Manjula and R. Lakshman Naik

This study's primary goal is to present a comprehensive understanding of SA approaches, including their classification and application methodologies. Additionally, it provides a succinct synopsis of big data methodologies and their application in the field of SA. Since the usage of Big Data (Hadoop) in the process of gathering data and reviewing it from social networks for analysis has significantly advanced in the last few years.

Sentiment Analysis of Twitter Data: A Survey of Techniques by Vishal A. Kharde, S.S. Sonawane

In this study, we present a survey and comparative analysis of current opinion mining methodologies, including lexicon-based approaches and machine learning, along with assessment criteria. We provide study on twitter data streams using a variety of machine learning algorithms, including Naive Bayes, Max Entropy, and Support Vector Machine.

IV. RESEARCH METHODOLOGY

To extract actionable insights from sentiment data, enabling businesses to improve customer engagement, satisfaction, and loyalty through informed strategies and initiatives. / To evaluate the effectiveness of marketing campaigns by analyzing sentiment data, allowing adjustments for better targeting.

We start with hypothesis generation:

Null hypothesis- Sentiment polarity does not depend upon qualification of people as well as the Chat bot used
 Alternative hypothesis- Sentiment polarity depends upon qualification of people as well as the Chat bot used
 The following analyses on specific data subsets to gain insights: The dataset which we have taken is from Kaggle. It provides data on the statements by the bot, the polarity of the statement, the age group of the people who judged it and their age.

Source	polarity	Age	division	Education	polarity	Statements
Flipkart	0.5	24	positive	Graduation	0.5	I love this product!
Yelp	0.3	24	positive	Graduation	0.3	The service was terrible.
IMBD	0.2	24	positive	Graduation	0.2	This movie is amazing!
Amazon reviews	0.4	24	positive	Graduation	0.4	I'm so disappointed with their customer support.
Amazon reviews	0.2	24	positive	Graduation	0.2	The quality of this product is subpar.
Amazon reviews	0.1	24	positive	Master	0.1	Just had the best meal of my life!
Amazon reviews	0.4	24	positive	Graduation	0.4	This book made me feel inspired. Highly recommended!", Positive, Goodreads, 2023-07-02 12:34:56, bookworm789, London, 0.88
Goodreads	0.5	24	positive	Graduation	0.5	The customer service at this store is top-notch
Swiggy	0	24	neutral	Graduation	0	
Swiggy	-0.1	24	negative	Graduation	-0.1	The food at this restaurant was awful. Never going back again!
Swiggy	0	24	neutral	Graduation	0	Great food
Swiggy	0.8	24	positive	Basic	0.8	Great food
Zomato	0	24	neutral	Graduation	0	Nice
Zomato	0.1	24	positive	Master	0.1	Loved the food
Zomato	0.3	24	positive	Graduation	0.3	Loved the food
Zomato	0.5	24	positive	Graduation	0.5	Loved the food
Zomato	0.5	24	positive	Graduation	0.5	Loved the food
Zomato	0.6	24	positive	Graduation	0.6	Loved the food
Zomato	0.4	24	positive	Master	0.4	Loved the food
Zomato	0	24	neutral	Graduation	0	
Twitter	0.2	24	positive	Graduation	0.2	his playlist is my go-to for workouts. Energizing and motivating!
Spotify	0.3	24	positive	Graduation	0.3	I can't stop listening to this song. It's my new favorite!
Spotify	-0.1	24	negative	Graduation	-0.1	Awful way to extort money
Twitter	0	24	negative	Graduation	0	Awful way to extort money
Byjus	0.1	24	positive	Graduation	0.1	I had a great chat with their customer support. Helpful and friendly.
Byjus	0	24	neutral	Graduation	0	I had a great chat with their customer support. Helpful and friendly.
Byjus	0	23	neutral	Graduation	0	I had a great chat with their customer support. Helpful and friendly.
Byjus	0	23	neutral	Graduation	0	I had a great chat with their customer support. Helpful and friendly.

Data preprocessing is a critical step in the data analysis and machine learning pipeline. It involves cleaning, transforming, and organizing raw data into a format that is suitable for analysis or modeling. Proper data preprocessing can significantly improve the quality of our results and the performance of machine learning models. The two data pre-processing techniques we are going to use are as follows: - Label encoding is a technique for converting categorical data into numeric form by assigning a unique integer label to each category. It's a straightforward way to represent categorical data in a format that can be used for various machine learning algorithms, including regression analysis.

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Education	Label encoded	Qualification encoded																	
		1																	
Graduation		2																	
Graduation		3																	
Graduation		2																	
Graduation		2																	
Graduation		2																	
Master		0																	
Graduation		2																	
Graduation		2																	
Graduation		2																	
Graduation		2																	
Basic		1																	
Graduation		2																	
Master		3																	
Graduation		2																	
Graduation		2																	
Graduation		2																	
Graduation		2																	
Graduation		2																	
Graduation		2																	
Graduation		2																	
Graduation		2																	
Graduation		2																	
Graduation		2																	
Graduation		2																	
Graduation		2																	
Graduation		2																	
Graduation		2																	

Figure 2: A snapshot of after-label encoding dataset One-hot encoding is a technique used to convert categorical data into a binary format where each category is represented as a binary (0 or 1) variable. This is particularly useful for categorical data where there is no ordinal relationship between categories After the various preprocessing procedures have been done, the analysis can be performed on the dataset. Statistical Analysis that we are going to perform are –

Correlation -

A statistical measure that describes the degree to which two or more variables are related or associated with each other. It quantifies the strength and direction of the relationship between variables. Here the two variables as stated before, are polarity and the qualification.

Qualification encoded	polarity
2	0.5
2	0.3
2	0.2
2	0.4
2	0.2
2	0.1
3	0.4
2	0.5
2	0
2	-0.1
2	0
1	0.8
2	0
3	0.1
2	0.3
2	0.5
2	0.5
2	0.6
3	0.4
2	0
2	0.2
2	0.3
2	-0.1
2	0
2	0.1
2	0
2	0
2	0
2	0

Figure 3: A snapshot of the variables for correlation

Regression

A set of statistical methods used for modeling and analyzing the relationship between a dependent variable (also called the response or outcome variable) and one or more independent variables (also called predictors or features). The main goal of regression analysis is to understand how changes in the independent variables are associated with changes in the dependent variable. It is widely used for prediction, forecasting, and understanding the nature of relationships in data.

Source	Source code	polarity
Flipkart	1	0.5
Yelp	2	0.3
IMBD	3	0.2
Amazon reviews	4	0.4
Amazon reviews	4	0.2
Amazon reviews	4	0.1
Amazon reviews	4	0.4
Goodreads	4	0.5
Swiggy	5	0
Swiggy	5	-0.1
Swiggy	5	0
Swiggy	5	0.8
Zomato	6	0
Zomato	6	0.1
Zomato	6	0.3
Zomato	6	0.5
Zomato	6	0.5
Zomato	6	0.6
Zomato	6	0.4
Zomato	6	0
Alibaba	7	0.2
Spotify	8	0.3
Spotify	8	-0.1
Spotify	8	0
Byjus	9	0.1
Byjus	9	0
Byjus	9	0
Byjus	9	0

Fig 4: A snapshot of encoded source names alongside polarity

V. RESULTS AND OBSERVATION

The results of the test are as follows:

	Qualification encoded	polarity
Qualification encoded	1	
polarity	-0.13963744	1

Fig 5: Result of correlation

The correlation between Qualification encoded and Polarity show a weak negative relationship. While there is a tendency for sentiment polarity to be more negative as qualification encoded values increase, the correlation is not strong, suggesting that other factors may also be influencing sentiment polarity. Correlation measures linear relationships, and the strength and direction of

relationships can vary widely based on the data and context. A weak correlation does not imply causation, and other factors not considered in the correlation analysis may also influence sentiment polarity.

SUMMARY OUTPUT	
<i>Regression Statistics</i>	
Multiple R	0.443562118
R Square	0.196747352
Adjusted R Square	0.107497058
Standard Error	3.13329417

Fig 6: Results of Regression

The Multiple R value suggests a moderate positive relationship between the independent variables (X) and the dependent variable (Y). The R Square value indicates that the independent variables explain about 19.67% of the variability in the dependent variable. The Adjusted R Square, while lower, is a more conservative estimate of explained variability after considering model complexity. The Standard Error represents the model's prediction accuracy, with smaller values indicating better accuracy.

VI. CONCLUSION

Based on the results from above, we can conclude that

- Even though the correlation was weak but it did suggest that higher a person qualification, lower the sentiment score that is, more critical the people are.
- As for the relation between the chat bots and polarity score, The positive correlation between the Polarity Score and the Chat bot variable suggests that, on average, as the Polarity Score increases (indicating more positive sentiment), the Chat bot variable tends to increase as well. This might imply that higher Polarity Scores are associated with better or more active Chat bot interactions. However, it's important to note that the Polarity Score explains only a relatively small portion (about 10.75%) of the variability in the Chat bot variable. This suggests that while there is a correlation, other factors not included in the model also influence the Chat bot variable.
- The Standard Error of 3.1333 indicates the average prediction error, which means that the model's predictions may not be very precise. Other unmeasured factors likely contribute to the variability in the Chat bot variable.

VII. REFERENCES

- [1] A Study of Sentiment Analysis: Concepts, Techniques, and Challenges, Ameen Abdullah Qaid Aqlan, B. Manjula and R. Lakshman Naik, Research Gate (January 2019)
- [2] Wankhade, M., Rao, A.C.S. & Kulkarni, C : A survey on sentiment analysis methods, applications, and challenges (February 2022)
- [3] Sentiment analysis using product review data, Xing Fang and Justin Zhan, Journal of Big Data, (2015)