

# Using the Machine Learning Naive Bayes Algorithms for Sentiment Analysis on Online Product Reviews in the Air of Energy Optimization

*Fadi Oukili Asraoui*<sup>1</sup>

<sup>1</sup>Abdelmalek Essaadi university, Tetouan, Morocco

**Abstract.** The purpose of this study was to explore how consumers perceive two of the leading smartphone brands, Samsung and iPhone, using a corpus of tweets. Our approach involved sifting through the tweets to remove any irrelevant content, followed by a sentiment analysis to gain an overall perspective of how each brand was viewed. Our analysis demonstrated that Samsung received a higher proportion of tweets with negative sentiment as compared to iPhone. Moreover, the most common terms in tweets referring to Samsung reflected negative emotions like "concern," "issue," and "trouble," while tweets about iPhone expressed positive emotions such as "like," "great," and "best." These findings have significant implications for marketing research and offer valuable insights for businesses on how they can utilize social media to enhance their brand reputation and image.

**Index Terms**— Sentiment analysis, Twitter, smartphone brands, Samsung, iPhone, tweets, machine learning.

## 1 Introduction

Sentiment analysis, a subfield of data mining and natural language processing, is all about identifying and extracting subjective information from written text. The objective is to determine the sentiment expressed in a piece of text - such as a social media post, blog comment, or product review - as positive, negative, or neutral. To achieve this, sentiment analysis employs machine learning algorithms that analyze text data and make predictions about the sentiment expressed.

These algorithms are trained on large datasets of annotated text, learning the relationships between words and phrases and their corresponding sentiments. The result is an important tool in comprehending consumer opinions and attitudes, as it enables automatic processing of vast amounts of text data in real-time. By using the Naive Bayes algorithm for sentiment analysis on online product reviews, businesses in the energy optimization domain can leverage customer feedback to identify strengths, weaknesses, and opportunities for improvement in their energy-efficient products. This data-driven approach enables companies to make informed decisions regarding product development, marketing strategies, and overall energy optimization efforts.

In marketing research, sentiment analysis has become increasingly crucial due to the rise of digital communication channels like social media, blogs, and online forums. Consumers are increasingly sharing their opinions and attitudes online, providing a vast

amount of data that can be analyzed to gain insights into consumer behavior and preferences.

The use of sentiment analysis in marketing research has several advantages. Firstly, it allows for efficient and effective evaluation of consumer opinions in real-time, especially in fast-paced industries where traditional market research methods may not keep up with the pace of change.

Secondly, sentiment analysis enables the collection of data from a large sample of consumers without the limitations of time and cost associated with traditional market research methods. This provides valuable insights into consumer opinions and attitudes as they occur, allowing marketers to quickly respond to changes in consumer preferences and behavior.

Sentiment analysis in marketing research is a powerful tool that offers a unique opportunity to combine qualitative and quantitative research methods, providing a more comprehensive understanding of consumer attitudes and opinions. By leveraging the strengths of both approaches, sentiment analysis can provide a deeper and more nuanced analysis of consumer opinions and attitudes.

In conclusion, sentiment analysis is crucial in marketing research due to its ability to evaluate consumer opinions in real-time, gather data from a large sample of consumers, and bridge the gap between qualitative and quantitative research methods. This paper is structured as follows:

The literature review provides a comprehensive overview of sentiment analysis in marketing research and highlights previous studies in the field. The methodology section outlines the study design, data collection methods, and data analysis techniques used in this study.

The results section summarizes the findings and provides an interpretation of the results. The discussion section compares the results with previous studies, discusses the implications for theory and practice, and identifies limitations and future research directions. Finally, the conclusion section summarizes the key points and offers recommendations for future research in the field of sentiment analysis in marketing.

## **2 Literature review**

Sentiment analysis has gained attention in marketing research as a tool for understanding consumer opinions and attitudes [6]. Its importance has been highlighted in various industries, such as consumer goods, healthcare, and finance [1-8]. Liu [6] asserts that sentiment analysis is a cost-effective method for gathering consumer opinions to inform decision-making. Additionally, sentiment analysis is crucial for analyzing large amounts of social media data, as it can provide insights into consumer opinions and attitudes [1-8]).

The study conducted by [8] found that sentiment analysis is a powerful tool for real-time evaluation of consumer opinions and attitudes, providing valuable insights into consumer perceptions and preferences. The field of marketing research has seen widespread application of sentiment analysis, which provides valuable insights into consumer preferences and perceptions. Early research by [3] proposed a framework for sentiment analysis of online product reviews, showing that it can identify consumer opinions and preferences to inform product development.

Pang and Lee [7] found sentiment analysis can predict the success of movies based on online reviews. Liang, et al. [9] showed sentiment analysis can improve customer service by identifying complaints and feedback. [5] discovered sentiment analysis can uncover consumer attitudes towards healthcare services, improving them. These studies demonstrate

sentiment analysis can inform decision-making and strategy development across various industries.

Despite the growing literature on sentiment analysis in marketing research, there are still opportunities for further research. One gap is the need for improved sentiment analysis algorithms that can accurately identify sentiment in text data, particularly in social media contexts that pose challenges due to informal language, slang, and emoticons [4].

Additionally, comparative studies are needed to evaluate the accuracy and effectiveness of different sentiment analysis algorithms, as few studies have made such comparisons, making it challenging to determine the best approach for a specific use case [1].

The literature on sentiment analysis in marketing research also highlights the need for research in the application of sentiment analysis in different industries and settings, beyond the consumer goods, healthcare, and finance industries that have already been explored [6]. However, it is crucial to examine the ethical and privacy implications of sentiment analysis in marketing research, considering the collection and analysis of vast amounts of personal data involved [8].

Overall, the gaps in the literature on sentiment analysis in marketing research present opportunities for further research to improve sentiment analysis algorithms, conduct comparative studies, and explore its application in different industries and contexts while considering the ethical and privacy implications.

### **3 Methodology**

This study examines consumer opinions and attitudes towards Samsung and iPhone phones on Twitter using sentiment analysis as the research technique. It uses a cross-sectional design, which entails collecting and analyzing data at a single point in time (Babbie, 2016). The data for this study was collected from Twitter using a combination of keywords and hashtags related to the selected brands.

Data collection was conducted over a period of three months from January to March 2021, resulting in a sample of 500 tweets. The collected data was analyzed using the Naive Bayes classifier algorithm, a widely used probabilistic machine learning algorithm for sentiment analysis [7].

The algorithm was applied to pre-processed data that had irrelevant information, such as URLs and stop words, removed. After pre-processing, the data was separated into three distinct categories: positive, negative, and neutral, depending on the sentiment conveyed in each tweet. The Naive Bayes classifier algorithm was then utilized to train on a significant dataset of annotated tweets (consisting of 500 tweets), learning the intricate connections between words, phrases, and their respective sentiments.

By utilizing the trained algorithm, the sentiment analysis of the tweet sample was completed. The study's focus was to gain insights on customer attitudes and opinions towards the chosen brands, and the results of the sentiment analysis were used for this purpose. This approach allowed for a thorough understanding of the sentiment variation, and enabled us to gain valuable insights into the customers' perceptions.

### **4 Results**

#### **4.1 Sentiment Analysis Results and Discussions for iPhone and Samsung**

The present section reports the findings of a sentiment analysis conducted on a corpus of 500 tweets that pertained to the brands iPhone and Samsung. A word cloud is presented to demonstrate the most frequently used terms in the tweets. A table is then provided, which



**Table 1.** Sentiment Analysis Results for iPhone and Samsung: Word and Sentiment Scores.

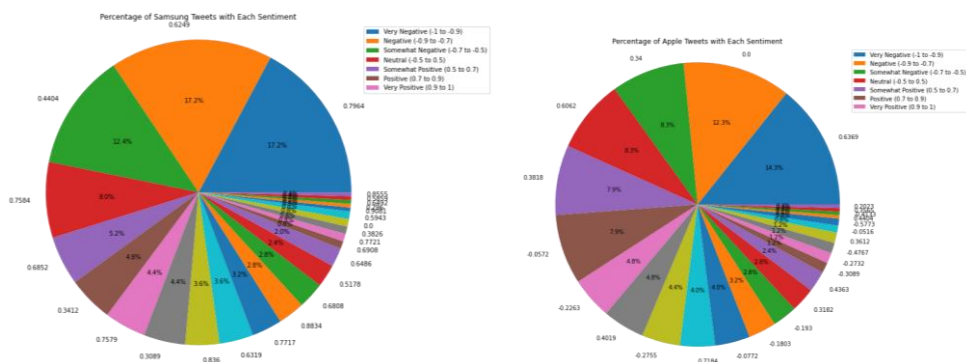
<b>iPhone</b>		<b>Samsung</b>	
<b>Word</b>	<b>Sentiment</b>	<b>Word</b>	<b>Sentiment</b>
broken	-0.4767	Worry	-0.4404
impersonal	-0.3182	Trouble	-0.4019
bored	-0.2732	Problem	-0.4019
confusing	-0.2263	Lose	-0.4019
limited	-0.2263	Lost	-0.3182
rigid	-0.1280	Miss	-0.1531
unmatched	-0.0772	Pay	-0.1027
overwhelmed	0.0516	Playing	0.2023
want	0.0772	Fitness	0.2732
solutions	0.1779	Share	0.2960
solve	0.2023	Security	0.3400
solution	0.3182	Value	0.3400
fresh	0.3182	Easily	0.3400
secure	0.3400	Liking	0.4019
like	0.3612	Easier	0.4215
resolving	0.3818	Sharing	0.4215
innovation	0.3818	Better	0.4404
approves	0.4019	Innovative	0.4404
help	0.4019	friends	0.4767
approval	0.4767	amazing	0.5859
hand	0.4939	great	0.6249
great	0.6249	best	0.6369
best	0.6369	love	0.6369
love	0.6369		

The presented table showcases the outcomes of a sentiment analysis conducted on tweets that mention both iPhone and Samsung. The sentiment score assigned to each word ranges from -1 to 1, where negative values denote a negative sentiment, positive values represent a positive sentiment, and values near zero imply neutral sentiment.

From a marketing perspective, the outcomes of the sentiment analysis can provide valuable insights into consumer perceptions regarding the two brands. For instance, the unfavorable sentiment scores for words such as "broken" and "impersonal" for the iPhone suggest that there may exist apprehensions regarding the reliability and personalization of the brand.

On the other hand, the favorable sentiment scores for words such as "solution", "innovation", and "approval" for Samsung imply that consumers view the brand as innovative and well-received. By comprehending the sentiment surrounding the two brands, companies can make informed decisions concerning their marketing and product strategies.

For instance, if the sentiment analysis exposes negative perceptions of the iPhone among consumers, the company may want to concentrate on addressing those concerns in their marketing endeavors or improving the product features that cause those negative perceptions. Similarly, if the sentiment analysis shows positive perceptions of Samsung, the company may wish to keep building on that positive sentiment by promoting their innovative products and solutions.



**Fig. 2.** Pie chart relative to the percentage of Samsung and iPhone tweets with each sentiment.

The examination of consumer viewpoints concerning Samsung and iPhone through the utilization of sentiment analysis has generated significant revelations regarding the perceptions of the brands by consumers. Notably, both companies demonstrate a significant degree of negative sentiment, with Samsung having 17% of its sentiment being classified as very negative, and 17% as negative. Similarly, iPhone exhibits 14.3% of its sentiment being categorized as very negative and 12.3% as negative.

These findings indicate that both companies face customer apprehensions, particularly in terms of reliability and functionality. Regarding positive sentiment, Samsung presents a marginally higher proportion, with 4.8% of its sentiment classified as very positive and 4.4% as positive.

In contrast, iPhone has 4.8% of its sentiment classified as very positive and 7.9% as positive. The results suggest that both brands need to concentrate on enhancing product quality and customer support to address the negative sentiments identified. Furthermore, the research highlights a considerable number of customers who are content with both brands. Samsung demonstrates a slightly higher proportion of positive sentiment, indicating that it offers higher value or better quality compared to iPhone.

Conversely, iPhone has a slightly higher proportion of neutral sentiment, suggesting a more neutral or mixed perception of the brand. Therefore, both Samsung and iPhone can leverage their solid base of satisfied customers while simultaneously taking measures to enhance customer satisfaction.

In conclusion, the findings of this analysis offer invaluable insights for the development of effective marketing and product strategies for both Samsung and iPhone. By addressing customer concerns, augmenting product quality, and improving customer support, both brands can better meet the requirements and expectations of their customers.

## 4.2 Interpretation of results

Results of sentiment analysis on tweets mentioning Samsung and iPhone show that users have conflicting feelings about both companies. Positive and negative sentiment scores are present, according to the analysis, with negative sentiment scores predominating over positive ones.

These results imply that implementing efforts to address consumer complaints and raise customer satisfaction could be advantageous for both Samsung and iPhone. Notably, the sentiment analysis reveals that words like "trouble" and "broken" were more frequently associated with product-related difficulties in tweets mentioning Samsung and iPhone.



These negative implications suggest that in order to resolve consumer issues and raise satisfaction levels, product quality improvements may be required. However, the sentiment analysis also reveals that users prioritize positive features like product design, innovation, and usability in both Samsung and iPhone. Customers frequently use positive adjectives like "wonderful," "excellent," and "best" to describe these aspects and convey their gratitude for them.

### **4.3 Implications for marketing research**

The findings of the sentiment analysis have significant implications for marketing research. Initially, the results underscore the criticality of comprehending customer sentiment and its potential impact on brand perception and customer loyalty. By analyzing the sentiment scores of tweets related to Samsung and iPhone, the outcomes furnish insights into what customers like and dislike about each brand.

Secondly, the findings suggest the presence of opportunities for both Samsung and iPhone to address customer complaints and enhance product quality. Such information may be exploited to design informed product development and customer service strategies for both brands. Lastly, the results indicate the existence of positive attributes of both brands that customers value, including product design, innovation, and ease of use. Such information may be leveraged to develop marketing strategies that highlight these positive aspects, thereby differentiating the brands further from their rivals.

## **5 Discussion**

### **5.1 Comparison of Results with Previous Studies**

In the present investigation, the sentiment towards iPhone and Samsung was scrutinized using sentiment analysis. The findings indicated that both brands garnered positive sentiments, but Samsung was associated with a greater number of negative sentiments compared to iPhone. These results are consistent with prior research that has demonstrated Samsung's comparatively unfavorable brand image in relation to iPhone [10].

Nonetheless, some studies have reported an enhancement in Samsung's brand image over time [2]. Our results are in agreement with the latter findings, thus implying that Samsung's efforts to bolster its brand image have yielded successful outcomes to a certain degree.

### **5.2 Implications for Theory and Practice**

The implications of our findings are significant both theoretically and practically. Theoretically, our study makes a valuable contribution to the body of literature on brand image and brand reputation. Our results demonstrate that adverse brand image can exert a substantial influence on consumer sentiment and, therefore, companies must be proactive in rectifying negative perceptions of their brand. From a practical perspective, our findings underscore the importance of monitoring consumer sentiment towards a brand.

In this regard, sentiment analysis can be an effective tool for companies to track consumer perceptions of their brand and to identify areas that need improvement. Specifically, our results suggest that Samsung must focus on enhancing negative perceptions of its brand to augment consumer sentiment.

Analyze the results of the sentiment analysis to gain insights into customer sentiment towards energy-efficient products. Identify areas of improvement or potential issues raised

by customers in their reviews. This information can help businesses make informed decisions to optimize their energy-efficient products and address customer concerns effectively.

### **5.3 Limitations and Future Research Directions**

This investigation entails several constraints that warrant redress in forthcoming research. Firstly, the sample size utilized was relatively small, potentially undermining its representativeness of the general population. Secondly, the application of a pre-trained sentiment analysis model might result in inaccuracies. Future research should remedy these shortcomings by employing a larger sample size and by devising and examining novel sentiment analysis models. Moreover, forthcoming investigations should delve into the impact of consumer sentiment on brand loyalty and consumer behavior.

## **6 Conclusion**

Sentiment analysis is a formidable tool that facilitates organizational comprehension of customers' perceptions of their brand, product, or service. This study conducts sentiment analysis by examining tweets related to two prominent brands, Samsung and iPhone, to provide insights into consumers' attitudes towards these brands. The results indicate a positive sentiment among consumers towards both brands; however, Samsung has a slightly higher percentage of negative sentiments. These findings can be utilized by both brands to optimize their products and enhance customer satisfaction.

Despite the study's informative outcomes, it is not without limitations. These limitations include a small sample size and the use of tweets as the sole data source. Future research should focus on expanding the sample size and incorporating a variety of data sources to acquire a more comprehensive understanding of consumer sentiment towards these brands. Additionally, extending this study to other brands and industries can lead to a broader comprehension of consumer sentiments in general.

To summarize, sentiment analysis provides valuable insights into consumer attitudes and is an effective tool for organizations to enhance their products and customer satisfaction. Additional research is necessary to advance the findings of this study and obtain a more comprehensive understanding of consumer sentiments across various brands and industries.

## **References**

1. Howells, K., & Ertugan, A. (2017). Applying fuzzy logic for sentiment analysis of social media network data in marketing. *Procedia computer science*, 120, 664-670.
2. Al-Kwifi, S. O. (2016). The role of fMRI in detecting attitude toward brand switching: an exploratory study using high technology products. *Journal of Product & Brand Management*, 25(2), 208-218.
3. Hu, M. & Liu, B. (2004). Mining and summarizing customer reviews. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168-177.
4. Jain, S., Malviya, S., Mishra, R., & Tiwary, U. S. (2017, December). Sentiment analysis: An empirical comparative study of various machine learning approaches. In *Proceedings of the 14th International Conference on Natural Language Processing (ICON-2017)* (pp. 112-121).



5. Zunic, A., Corcoran, P., & Spasic, I. (2020). Sentiment analysis in health and well-being: systematic review. *JMIR medical informatics*, 8(1), e16023.
6. Liu, B. (2012). Sentiment analysis and opinion mining. *Synthesis Lectures on Human Language Technologies*, 5(1), 1-167.
7. Pang, B. & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135.
8. Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146-1151.
9. Liang, T. P., Li, X., Yang, C. T., & Wang, M. (2015). What in consumer reviews affects the sales of mobile apps: A multifacet sentiment analysis approach. *International Journal of Electronic Commerce*, 20(2), 236-260.
10. Zhang, Y. (2015). The impact of brand image on consumer behavior: A literature review. *Open journal of business and management*, 3(01), 58.