ISSN: 2688-9323

User Opinion Analysis of Huawei Consumer Electronic Products in the Dominican Republic

Pablo Gabriel Santana Gómez

Chongqing University of Post and Telecommunications, China

Abstract

Recently, social media is part of our daily lives. It has given many people the power of sharing what they want (pictures, videos, moments, opinions, etc.) whenever they want. The number of users using social media had increased drastically and is expected to keep growing in the following years. This phenomenon generates tons of valuable and analyzable data (user preferences, tendencies, user opinions about certain products, trending topics, etc.) that can help companies to reach their goals by trancing their consumer's behavior. This research was conducted to determine the customer perception of Huawei consumer electronic products in the Dominican Republic's market, mining user comments from Huawei's official account on Instagram and applying the Naïve Bayes and Decision Tree classification model as well as using VADER sentiment analyzer to conduct a Sentimental Analysis using Natural Language processing tools.

Keywords

Sentiment Analysis; Natural Language Processing; Text Classification; Consumer Electronic Products.

1. Introduction

Natural Language Processing (NLP) is the automatic computational processing of human languages [2]. One of its main areas is Sentiment Analysis which has the purpose of extracting the polarity of a given string of text, to distinguish between, positive, negative and neutral opinions. [3]. Even though sentiment analysis is one of the most popular research topics related to NLP, there is no evidence of such studies conducted in the Dominican Republic (DR) using these computational tools. However, the constant economic growth that DR is experiencing [4] has made this country attractive for many local and international firms that seek to take advantage of the low manufacturing costs and growing market. The increasing social media usage by the Dominican population, which by 2020 already had 6.4 million active users [5], encourages these firms to explore these popular social platforms to study the consumer opinions about their products and services, in order to take well-oriented decisions.

Among all companies in DR, this research focuses on the multinational Huawei, which has an increasing popularity in the Dominican Republic due to its high-quality electronic gadgets that can be acquired at a reasonable price, even though recent controversies have affected its reputation and operations.

There is no formal research conducted in DR aiming to categorize the sentiment of comments on social media. Therefore, this study will reveal what are the strengths and opportunities of Huawei in this market, mining the opinions about Huawei on social platforms to know what the people are talking about this company, to check if the sentiment is the same for both men and women (H1), cellphone features (H2), and to confirm if the sentiment score varies with time (H3), considering the recent events affecting this company's reputation.

2. Literature Review

Sentiment analysis or opinion mining is defined as the is the field of study that analyzes people's emotions, sentiments, or opinions towards entities and their attributes [6]. The expression sentiment analysis appeared for the first time in a research paper by Nasukawa and Yi [7] but the research of opinions sentiments appeared in several earlier works such as Das and Chen, 2001 [8] Pang et al.,2002 [9] and Wiebe, 2000. [10]

Sentiment analysis has many applications. With the increasing use of social media, some organizations are using this data to help them make better decisions and give direction to their strategic planning. Some researchers had also explained how sentiment analysis can be used in different fields. In Liu et al. 2007 [11] sentiment analysis was used to predict sales, McGlohon et al. 2010 [12] purchases reviews helped to rank products and suppliers, and Tumasjan et al. 2010 [13] used Twitter data to predict elections results, and in Bollen et al. 2011 [14] also Twitter data is used but in this case, to predict stock market behavior. Some other works have been related to social media comment sentiment extraction, that present the advantages of social platforms as an environment for opinion mining [15], the use of machine learning methods to extract the sentiment from Instagram comments [16] , for later the accuracy of these models [17]. Another study shows the opinions collected during the begging of the commercial war between the United States and China [18].

3. Research Methodology

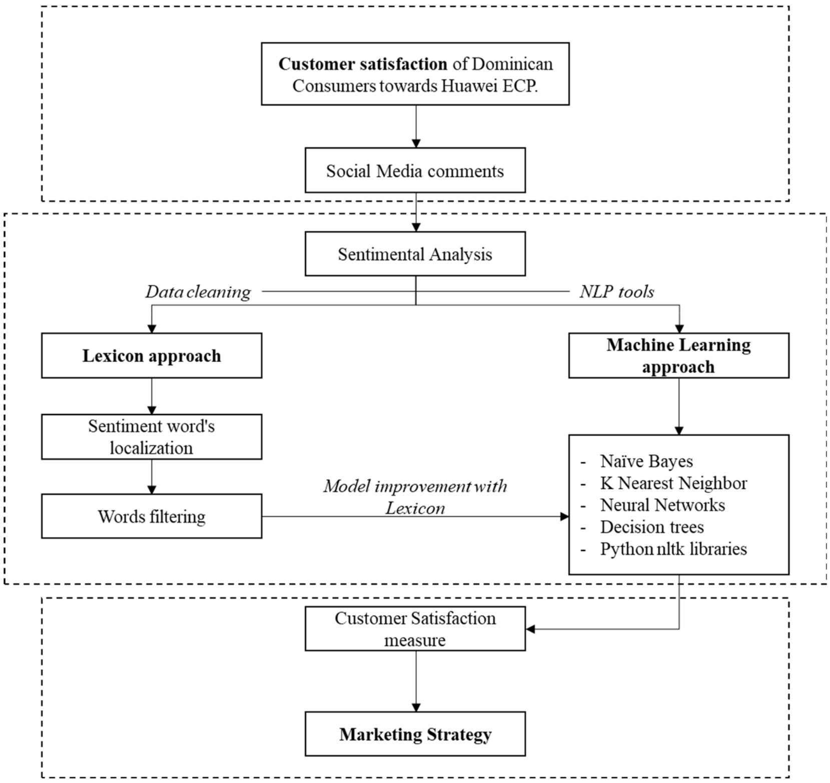

Figure 1. Research Process

This research will follow a quantitative methodology, using statistical inference for analyzing the data collected in the study. Web scraping techniques will be used for collecting the data from social media, taking the JSON code from each web page, and cleaning it with Python. The judgmental sampling procedure will be used to select the sample units from different comments on different electronic consumer products of Huawei in the Dominican Republic. All Instagram comments from de account @huaweimobilerd will be taken as the population of the

ISSN: 2688-9323

study, comments related to the four different product categories selected (cellphones, earbuds, smartwatches and laptops). Comments from 2019 to 2021 will be considered for this research.

Data processing and interpretation using Natural Language Processing classification methods and Python programing Language. Using a Python interpreter, this research aims to apply two classification algorithms (Naïve Bayes and Decision trees) and one Python library from the Natural Language Processing Toolkit to get the sentiment polarity from the comments.

4. Data Collection Technique

Since automatic web-scrapping from social media platforms is penalized. The safest and most accurate way to obtain user comments from Instagram is extracting the JSON code from every single post, scrolling the comments section down manually. The comments will be extracted directly from the posts related to Huawei on its official Instagram account @huaweimobilerd. Besides the comment itself, the gender of the user as well as the date on which the comment was posted, will also be extracted, using the NLTK names list.

After the data is extracted, invalid data types must be deleted, like duplicated values and empty records. Also, in the text string itself, some common characters in social media communication must be removed as can be appreciated on Table 1.

Table 1. Values to be extracted

Data type	Example
Link	http://abcde/efg
Emoji	😆
Punctuation	?!.,;”
White Space	“ ”
Digits	123456
Tags	#Sales#

Once we extract the comment and user information, the sentiment polarity (positive, negative, neutral) can be analyzed using one Python library for Natural Language Processing-Sentiment Analysis called VADER, and classification algorithms like Naïve Bayes and Decision Trees.

4.1. Naïve Bayes Text Classification Model for Sentiment Analysis

Naïve Bayes text classification model implies that every feature gets a say in determining which label should be assigned to a given input value [19]. For assigning a label to a specific input value, this model calculates the probability for each label, which is determined considering the occurrence of each label in the training set. Several features have to be considered for this model. In the end, the input value is tagged with the one that has the highest probability.

$$P\ (sentiment|features) = P\ (features,\ sentiment)/P(features)$$
$$P(features) = \Sigma label \in labels\ P\ (features,\ sentiment)$$
$$P\ (features,\ sentiment) = P(sentiment) \times P\ (features|\ sentiment)$$
$$P\ (features,\ sentiment) = P(sentiment) \times \Pi f \in features\quad P\ (f|\ sentiment)$$

4.2. Decision Trees Text Classification Model for Sentiment Analysis

This classification model is a flowchart that allocates labels for input values. Contains some decision nodes that analyze features and leaf nodes that tag labels. The first node is called the

ISSN: 2688-9323

root node and is where the first decision is made [19], the first feature is checked for allowing it later to get to the next node. It continues this process until arriving at the final node.

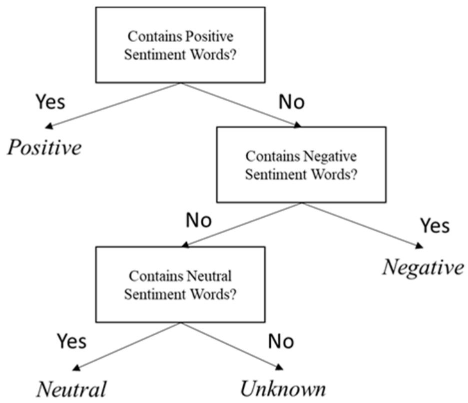

Figure 2. Decision Tree Model

4.3. VADER Python Library for Sentiment Analysis

VADER (Valence Aware Dictionary for Sentiment Reasoning) is a classification model for text sentiment analysis that shows both polarity (positive/negative) and intensity (strength) of the opinion. It comes with the Natural Language Processing Toolkit of Python (NLTK) package module uses dictionaries that conduct the lexical features to emotion intensities known as sentiment scores. The sentiment score of a text can be obtained by summing up the intensity of each word in the text.

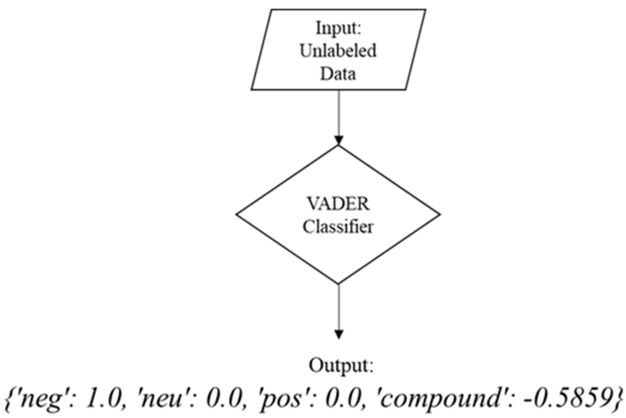

Figure 3. VADER Model

5. Results and Discussion

5.1. Results

We collected nearly 10,000 comments from Instagram and saved them on an Excel file. For generating the sentiment polarity for each comment this research found out that the best alternative was the Python module VADER for Natural Language Processing because this one offers the best features in terms of user-friendliness, accuracy, and efficiency [18]. Besides, some classification models, Naïve Bayes and Decision Tree both upgraded with non-standard Spanish words used frequently in the Dominican Republic. The results are expressed in the Figure 1 andd Figure 2.

ISSN: 2688-9323

Table 2. Comments by Classification Method

Classifier	Polarity	Qty. of comments
Naïve Bayes	Positive	2530
	Neutral	5727
	Negative	1743
Decision Tree	Positive	2269
	Neutral	6340
	Negative	1391
VADER	Positive	2041
	Neutral	6575
	Negative	1384

Table 3. Accuracy by Classification Method

Classifier	Accuracy (%)
Naïve Bayes	86.07
Decision Tree	65.5
VADER	77.8

After putting the data into a probability distribution graph, a normal distribution with mean nearlyzero can be appreciated on Figure 4, therefore the Central Limit Theorem can be appied. In addition, all the tests and statistical tools will be carried by the variables showed on Table 4.

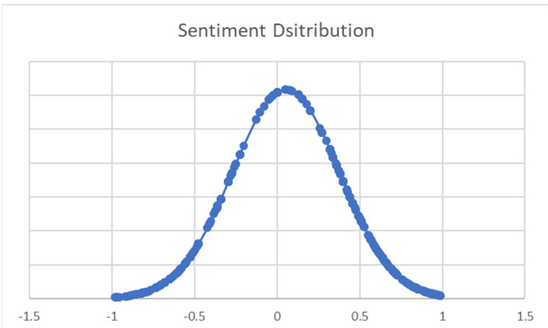

Figure 4. Sentiment Score Distribution

Table 4. Selected Variables

Variable	Values
User gender	Male
	Female
Products	Cellphones
	Earbuds
	Smartwatch
	Laptops
Features	Price
	Battery
	Performance

ISSN: 2688-9323

To confirm if the overall sentiment has variations after the implementation of the sanctions and since the variances of the population σ_{before} and σ_{after} are unknown, a Z-score test was conducted applying the concepts of the Central Limit Theorem, considering the probability of rejecting the null hypothesis while it is true of α = 0.05. For this kind of test the Z_{score} is calculated as follows:

$$Z_{score} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{\sqrt{\left(\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}\right)}}$$

For the chosen value of α, the level of significance is $\pm$ 1.96 and the obtained Z_{score} is -0.74 which gives a p_{value} of 0.2296. As can be appreciated the p_{value} is greater than the established significance level (p_{value}= 0.2296 > α = 0.05) therefore there is not enough evidence to reject the null hypothesis (H3). The previous test was made taking all the data from the comments independently of the controlled variables like type of product or gender. Therefore, we will conduct the same test for each variable, considering the period of time before and after the restrictions as two different samples. The results are shown in the following table:

Table 5. Z-test

Product	Z-stat	p-value	Result
Cellphone	0.33	0.73	Fail to Reject Null
Smart Watch	-0.29	0.77	Fail to Reject Null
Earbuds	-2.67	0.01	Reject Null
Laptop	1.54	0.12	Fail to Reject Null

As can be appreciated, for a significance level of 0.05, only the Earbuds have a considerable variation after the implementation of the sanctions. Doing the same test for the gender variable que obtain a Z_{score} of -4.62 with a p_{value} almost equal to zero, which implies that the sentiment score does vary among genders. An ANOVA for product features was also conducted to appreciate where the variation exists among this variable. Where can be appreciated that there is a considerable variation among the groups.

Table 6. ANOVA

Source of Variation	SS	df	MS	F	P-value	F crit
Between Groups	0.80	2	0.40	3.82	0.02	3
Within Groups	208.61	9998	0.10			
Total	209.41	10000				

Linear regression with sentiment and gender, before and after sanctions.

Table 7. Linear Regression

	Coefficients	Standard Error	t Stat	P-value
Intercept	0.10	0.016	6.42	0.00
Gender	-0.07	0	-4.70	0.00
Sanctions	-0.011	0.016	-0.688	0.492

ISSN: 2688-9323

6. Conclusions and Future Work

This study aims to determine the user perception of Huawei consumer electronic products in the Dominican Republic by specifically conducting a Sentimental Analysis on Instagram comments. Based on the final results written in the previous chapter, is concluded that men represent most comments with 61.08% of all comments. The most accurate model was the Naïve Bayes with an 86.07 score. The most commented product was the cellphone with 52.56% of all comments, from which 23.57% were negative and 28.99% were positive. As for the features, most of the comments were about performance with 76.02% of all comments, of which 42.59% were positive. Overall, the sentiment classification for the 10,000 comments was as follows: positive 20.41%, negative 13.84%, and neutral 65.75%.

In addition, there is not enough evidence to confirm that the restrictions represent a major impact on the sentiment score, according to the z-test, except for the product “earbuds” that showed a slight variation. Some advice for future research are: 1) Use another classification algorithm for conducting the sentiment analysis like Neural Network, known for having a high accuracy; 2) Conduct the sentiment analysis for Huawei Consumer Electronic Products in a different time range; 3) Add more Dominican slang words to the model. Furthermore, the implementation of a modified classification algorithm might provide better accuracy for the algorithm and sentiment score.

References

[1] Statista, "Statista.com, " 2021. [Online]. Available: https:// www.statista. com/ statistics/ 278414/ number- of-worldwide-social-network-users/. [Zugriff am 11 January 2022].

[2] Y. Goldberg, Neural Network Methods in Natural Language Processing, 2017.

[3] K. J. a. M. Morzy, "Opinion Mining and Social Networks: A Promising Match, " 2011.

[4] W. Bank, databank.worldbank.org, 2021. [Online]. Available: https:// databank. worldbank. org/ source / world-development-indicators. [Zugriff am 11 January 2022].

[5] DATAREPORTAL, datareportal.com, 2020. [Online]. Available: https:// datareportal. com/ reports / digital-2020-dominican-republic. [Zugriff am 11 January 2020].

[6] B. Liu, Sentiment Analysis and Opinion Mining, Toronto: Morgan & Claypool, 2012.

[7] T. a. J. Y. Nasukawa, "Sentiment analysis: Capturing favorability using natural language processing., " 2003.

[8] S. a. M. C. Das, "Yahoo! for Amazon: Extracting market sentiment from stock message boards" 2001.

[9] B. L. L. a. S. V. Pang, "Thumbs up?: sentiment classification using machine learning techniques." 2002.

[10] J. Wiebe, Learning subjective adjectives from corpora., 2000.

[11] J. j. Y. C. C.-Y. L. Y. H. a. M. Z. Liu, Low-quality product review detection in opinion summarization, 2007.

[12] M. N. G. a. Z. R. McGlohon, "Star quality: Aggregating reviews to rank products and merchants." 2010.

[13] A. T. O. S. P. G. S. a. I. M. W. Tumasjan, Predicting elections with Twitter: What 140 characters reveal about political sentiment., 2010.

[14] J. H. M. a. X.-J. Z. Bollen, "Twitter mood predicts the stock market., " Journal of Computational Science, 2011.

[15] K. J. a. M. Morzy, Opinion Mining and Social Networks:a Promising Match, 2011.

[16] A. L. D. Y. R. Hanif Sudira, "Instagram Sentiment Analysis with Naive Bayes and KNN: Exploring Customer Satisfaction of Digital Payment Services in Indonesia," 2019.

[17] M. Ahmad, "Sentiment Analysis System of Indonesian Tweets using Lexicon and Naïve Bayes Approach, " 2019.

[18] N. H. K. a. H. Aye, "Visualization of Twitter Sentiment during the Period of US Banned Huawei," 2020.

ISSN: 2688-9323

[19] E. K. E. L. Steven Bird, Natural Language Procesing with Python, O'Reilly Media, 2009.