





“Applying advanced sentiment analysis for strategic marketing insights: A case study of BBVA using machine learning techniques”

AUTHORS	Luis Miguel Garay Gallastegui  Ricardo Reier Forradellas  Sergio Luis Náñez Alonso 
ARTICLE INFO	Luis Miguel Garay Gallastegui, Ricardo Reier Forradellas and Sergio Luis Náñez Alonso (2024). Applying advanced sentiment analysis for strategic marketing insights: A case study of BBVA using machine learning techniques. <i>Innovative Marketing</i> , 20(2), 100-115. doi: 10.21511/im.20(2).2024.09
DOI	http://dx.doi.org/10.21511/im.20(2).2024.09
RELEASED ON	Wednesday, 17 April 2024
RECEIVED ON	Friday, 01 September 2023
ACCEPTED ON	Wednesday, 03 April 2024
LICENSE	 This work is licensed under a Creative Commons Attribution 4.0 International License
JOURNAL	"Innovative Marketing "
ISSN PRINT	1814-2427
ISSN ONLINE	1816-6326
PUBLISHER	LLC “Consulting Publishing Company “Business Perspectives”
FOUNDER	LLC “Consulting Publishing Company “Business Perspectives”



NUMBER OF REFERENCES

54



NUMBER OF FIGURES

10



NUMBER OF TABLES

4

© The author(s) 2024. This publication is an open access article.



BUSINESS PERSPECTIVES



LLC "CPC "Business Perspectives"
Hryhorii Skovoroda lane, 10,
Sumy, 40022, Ukraine
www.businessperspectives.org

Received on: 1st of September, 2023

Accepted on: 3rd of April, 2024

Published on: 17th of April, 2024

© Luis Miguel Garay Gallastegui,
Ricardo Reier Forradellas, Sergio Luis
Náñez Alonso, 2024

Luis Miguel Garay Gallastegui, Ph.D.,
Dekis Research Group, Social and Legal
Sciences Faculty, Catholic University of
Ávila, Spain. (Corresponding author)

Ricardo Reier Forradellas, Ph.D., Dekis
Research Group, Social and Legal
Sciences Faculty, Catholic University of
Ávila, Spain.

Sergio Luis Náñez Alonso, Ph.D.,
Professor, Dekis Research Group, Social
and Legal Sciences Faculty, Catholic
University of Ávila, Spain.



This is an Open Access article,
distributed under the terms of the
[Creative Commons Attribution 4.0
International license](https://creativecommons.org/licenses/by/4.0/), which permits
unrestricted re-use, distribution, and
reproduction in any medium, provided
the original work is properly cited.

Conflict of interest statement:

Author(s) reported no conflict of interest

Luis Miguel Garay Gallastegui (Spain), Ricardo Reier Forradellas (Spain),
Sergio Luis Náñez Alonso (Spain)

APPLYING ADVANCED SENTIMENT ANALYSIS FOR STRATEGIC MARKETING INSIGHTS: A CASE STUDY OF BBVA USING MACHINE LEARNING TECHNIQUES

Abstract

In the digital era, understanding public sentiment toward brands on social media is essential for crafting effective marketing strategies. This study applies sentiment analysis on Banco Bilbao Vizcaya Argentaria (BBVA) tweets using advanced machine learning techniques, particularly the eXtreme Gradient Boosting (XGBoost) algorithm, which showed remarkable precision (91.2%) in sentiment classification. This process involved a systematic approach to data collection, cleaning, and preprocessing. The precision of XGBoost highlights its effectiveness in analyzing social media conversations about banking. Additionally, this paper achieved improvements in neutral tweet classification, with accuracy rates at 87-88% and a reduced misclassification rate, enhancing the analysis reliability. The findings not only uncover general sentiments toward BBVA but also provide insight into how these sentiments shift in response to marketing activities and global events. This gives marketers a valuable tool for real-time assessment of campaign effectiveness and brand perception. Ultimately, employing the XGBoost algorithm for sentiment analysis offers BBVA a strategic advantage in understanding and engaging its online audience, demonstrating the significant benefits of using sophisticated machine learning in banking. The study emphasizes the crucial role of data-driven sentiment analysis in developing informed business strategies and improving customer relationships in the banking industry's competitive landscape.

Keywords

customer sentiment, marketing analytics, data-driven strategies, artificial intelligence

JEL Classification

M31, G21

INTRODUCTION

In the era of digital interconnectedness, the banking sector represented by institutions like Banco Bilbao Vizcaya Argentaria (BBVA) stands at a crossroads of challenges and opportunities. This landscape is especially pronounced in leveraging customer sentiments expressed across social media platforms, highlighting a crucial shift in the approach to analyzing customer feedback and market trends for strategic marketing decisions. The integration of banking with digital communication technologies signals a transformation in how insights into customer preferences and market dynamics are harnessed to drive marketing strategies and business development.

The central challenge arises from the voluminous and unstructured nature of social media data, which traditional manual analysis techniques cannot efficiently or accurately process. This presents a significant scientific inquiry: How can banks like BBVA systematically convert this expansive, unstructured data into structured, actionable insights that inform marketing strategies? The solution lies within the

realms of data science and machine learning, where cutting-edge algorithms hold the potential to revolutionize sentiment analysis. This technological advancement allows banks to quickly respond to customer feedback, anticipate market trends, and refine their marketing and service offerings.

However, the integration of machine learning for sentiment analysis within the banking industry is accompanied by intricate challenges. A primary concern is the identification and fine-tuning of algorithms capable of precisely classifying sentiments, with a particular focus on detecting negative feedback to enable timely and proactive marketing interventions. The capacity of these algorithms to manage large volumes of data and their sensitivity to the subtleties of human emotion and expression on social media are pivotal in determining their utility. Thus, the adoption of these technologies is not merely a matter of operational enhancement but a strategic imperative that can provide banks with a competitive edge in marketing intelligence and customer relationship management.

1. LITERATURE REVIEW

The exploration of social networks transcends mere modern digital phenomena, tracing back to the 1930s with the pioneering concept of sociometry by Jacob Levy Moreno. Moreno (1953) not only laid the foundations for the sociometric movement but also introduced sociometry as a measurable and experimental method to analyze human relationships. This innovative approach, rooted in the principles of interrelation and spontaneous creativity (Moreno, 1955), underscored the complexity and interdependence of individuals within groups, setting a precedent for subsequent social network analyses.

The surge in interest in social network theory during the 1970s brought to light seminal theories such as Milgram's six degrees of separation (Milgram, 1967) and Granovetter's importance of weak ties in social relationships (Granovetter, 1973), which expanded one's understanding of social connectivity. This period also saw scholars like Wasserman and Kass (1995) and Borgatti et al. (2009) developing detailed methodologies to describe social networks as sets of actors connected by various types of relationships, thereby emphasizing the importance of nodes and the diversity of interactions within networks.

Boyd and Ellison (2007) characterized the transition to the digital era, marked by the emergence of virtual social networks, notably through platforms like Facebook. The initial modern virtual social network, Sixdegrees (Williams, 2019), highlighted a significant evolution in how individuals connect and communicate. Concurrently,

network centrality explored by Zhang and Luo (2017), along with degree centrality (Bródka et al., 2011), closeness centrality (Okamoto et al., 2008), and betweenness centrality (Stephenson & Zelen, 1989), provided deep insights into the structural dynamics of networks.

The use of social networks has also evolved as users gain experience in their use. The "stories" format in social networks, analyzed by Veissi (2017) and Rossmann (2018), represents an adaptation to new ways of sharing content, highlighting the need for detailed analysis in Big Data environments. Anshari et al. (2019) discuss the growing competition in paid advertising and the need for intelligent audience segmentation. Che and Ip (2017) noted how social networks have become influential marketplaces where users search for and buy products influenced by brands.

Consequently, social networks often mirror situations occurring in the physical world and serve as platforms for expression and relationship, triggering human motivations, such as the need for belonging, recognition, and identity construction. Brandes et al. (2013), García-Galera et al. (2014), and Nieto (2013) highlight this dynamic. They discussed the significant social impact and the growing importance of social networks in the coordination of economic society and as relationship channels within organizations.

The evolution of messaging platforms, as pointed out by Nouwens et al. (2017) and Gill et al. (2019), underscores the transition of users from public spaces to private ones, demanding personalization based on detailed information processed through

Big Data environments (Sudha & Sheena, 2020). Hootsuite Inc. (2022) predicts future trends in social networks focused on rebuilding user trust, adapting content to specific business objectives, and decoding the purchase funnel. Xu et al. (2020) emphasize the need to humanize brands to establish deeper connections with users, while Isaak and Hanna (2018) and Dickerson et al. (2014) discuss security and transparency challenges on platforms like Facebook and Twitter.

Big Data in social networks, highlighted by Bonnet and Westerman (2021), Kumar et al. (2020), and Hariri et al. (2019), plays a crucial role in today's digital era. It involves analyzing information generated on social platforms to guide corporate decision-making. IAB Spain (2021) reports that active social media users in Spain represent 85% of the online population, with a dominant 97% usage rate of mobile devices. These statistics underline the importance of analyzing data generated on these platforms, offering insights into the complex social dynamics unfolding within these digital realms. Here, 81% use these networks for entertainment, 72% for user interaction, and 66% for staying informed.

The application of Big Data methodologies to social media data analysis offers a range of valuable insights. Micu et al. (2017) and Lee and Kim (2020) stressed that audience analysis involves deciphering customer behaviors to enhance strategies across the pre-sales, sales, and post-sales spectrum. Conversation analysis, according to Sun et al. (2022), focuses on recognizing important topics for users, such as objects, individuals, and concepts, with sentiment in user posts often reflecting their emotional dispositions, typically categorized as positive, neutral, or negative. Furthermore, campaign analysis, both online and offline (Kim & Ellison, 2021; Piatak & Mikkelsen, 2021), shows how offline campaigns resonate on social platforms through mentions, feedback, and overall sentiment. Reputation management is another crucial aspect where monitoring public sentiment about brands, celebrities, political entities, or specific events is essential, especially during crises, to effectively manage narratives (Auter & Fine, 2018). Additionally, the analysis of virtual communities, as studied by Gutiérrez-Batista et al. (2016) and Molu et al. (2014), involves statistical evaluations

based on mentions to understand discourses within large online communities. Even central banks have applied the technique of sentiment analysis to test the acceptance or rejection of their digital currencies (CBDC) (Ozili & Nañez Alonso, 2024).

The banking sector's engagement with social media exemplifies the practical application of social network analysis in the industry. The American Bankers Association (2023) indicates a strong recognition among banks of the importance of social media for business strategy. This acknowledgment is mirrored in the marketing domain, where a shift toward allocating a larger portion of the marketing budget to social media is evident (Gartner, 2023). The emphasis on understanding consumer needs and interactions with brands on social media platforms (The Financial Brand, 2022) highlights the need for sophisticated data analysis to inform marketing strategies and customer engagement.

The emotion analysis within textual content is conducted on data sets that have been categorized in terms of positive, negative, or neutral sentiments. This process is executed by deploying an array of classification algorithms. Prior to classification, it is crucial to prepare the text through text mining techniques, which encompass the removal of symbols and punctuation, reducing words to their stems, and discarding stop words (Witten et al., 2011). These steps facilitate the creation of a list of significant terms. Subsequently, term frequency and inverse document frequency methods are employed to establish a vector space model. Upon obtaining this model, sentiment analysis is pursued through a classification procedure.

In their sentiment analysis investigation, Singh et al. (2017) examined four cutting-edge machine learning classifiers, including Naïve Bayes (NB), J48, BFTree, and OneR, aiming to optimize the process. They discovered that NB was expedient in learning, and OneR offered promising outcomes, with an accuracy of 91.3% in precision. Huq et al. (2017) employed machine learning algorithms Support Vector Machines (SVM) and k-NN on Twitter data, achieving accuracy rates ranging between 58.39% and 79.99% after extracting features using n-grams. Amolik et al. (2016) executed sentiment analysis by classifying tweets with a Feature Vector and classifiers such as NB and SVM.

Although NB exhibited higher precision in comparison to SVM, the latter proved to be more effective in terms of overall accuracy. Başarslan and Kayaalp (2020) focused on social media posts of users about health topic to categorize sentiments as positive, negative, and neutral, using the NB, SVM, and Artificial Neural Networks algorithms.

This comprehensive exploration of social networks from their inception as a sociometric study to the contemporary utilization of Big Data and machine learning reveals a significant evolution in how human connections and interactions are understood and analyzed. Through decades of scholarly investigation, the field has expanded from analyzing the direct interpersonal connections to understanding the vast, complex networks facilitated by digital technology.

Despite the substantial advancements in social network analysis, particularly through the lens of Big Data and machine learning algorithms, a gap remains in the practical application of these advanced techniques in specific sectors, such as banking. The literature underscores a growing need for nuanced tools capable of parsing the intricate web of digital interactions for actionable insights, especially in understanding customer sentiment and behavior on social media platforms (Lee & Kim, 2020). This gap points toward the potential impact of leveraging sophisticated machine learning approaches, like XGBoost, to not only enhance sentiment analysis but also to refine customer engagement strategies and decision-making processes in the banking industry.

This review sets the stage for the current study's objective: to rigorously evaluate the efficacy of specific machine learning algorithms, including XGBoost, in the sentiment analysis of BBVA's social media interactions. It is hypothesized that the machine learning algorithm XGBoost will outperform traditional models such as SVM, Random Forest, and Neural Networks in the task of sentiment classification within the banking sector's social media. This superiority is predicated on XGBoost's efficiency in handling the complexities and voluminous data typical of social network analysis, thereby providing more accurate actionable insights, and significantly contributing to the enhancement of customer relations and market engagement.

2. METHODS

A data-driven approach is used to collect and store the data, which is done using the Twitter API to gather tweets related to Banco Bilbao Vizcaya Argentaria (BBVA) and efficiently stored in a local PostgreSQL database. This process is automated by scheduled scripts. Subsequently, data processing and cleaning are moved onto, where official BBVA accounts' tweets are filtered out by a secondary script and duplicates are removed to ensure data precision. Data analysis and feature engineering are then proceeded to, where a subset of tweets is carefully examined, and a supervised dataset is curated. Through language analysis, linguistic diversity within the dataset is uncovered. Sophisticated feature engineering techniques are applied to refine the data, enhancing the precision of model training. The journey continues into feeling identification and rating, where the power of Natural Language Processing (NLP) is harnessed (Reier Forradellas et al., 2020). This includes sentiment analysis using the Google Cloud Natural Language (CNL) API, where emoticons, capital letter usage, and influential words are considered to gain a holistic view of tweet sentiments. To validate sentiment analysis, precision is rigorously assessed by comparing manual classifications with sentiments derived from the CNL. Additionally, feature selection is undertaken, employing correlation matrices to identify relationships between variables and eliminate redundancy. Finally, the application of Machine Learning models is delved into, evaluating various supervised learning algorithms, including SVM, XGBoost, Random Forest, and Neural Networks (Reier Forradellas et al., 2020). This evaluation assists in choosing the most suitable algorithm for the specific task. As part of the refinement process, Parameter Tuning is also engaged in, systematically optimizing model configurations by analyzing confusion matrices. This comprehensive approach ensures the robustness and precision of the data analysis pipeline.

Access to data is facilitated through the utilization of the Twitter API. For this experimental pilot, the free version of the API has been chosen to create a universally accessible environment suitable for a wide range of organizations and professionals. This version aligns perfectly with the requirements, accommodating both the daily tweet volume from BBVA's account and the 30-day analysis timeframe.

For the performance of data retrieval, the R library, `rtweet`, has been employed, a choice made based on its superior tweet download capacity compared to alternative libraries like `twitterR`. The script “`rTweetScript.R`” has been meticulously crafted for this purpose. An authentication process using the appropriate key and token is undergone by this script, a connection is established, and tweets are retrieved using the `search_tweets` function. This pilot study searched for the term “BBVA” and fetched 5,000 tweets per call. The retrieved data were temporarily stored in a list, and the decision to exclude retweets (`include_rts = FALSE`) to mitigate redundancy was made.

One of the key project objectives is the automation of the entire process. This pilot study adopted a novel approach using the `cronR` task scheduler instead of conventional scheduling tools to initiate the primary script daily. The primary script is triggered by this task scheduler, which subsequently activates a chain of auxiliary scripts. This streamlined approach allows for a focus on the crucial tasks of data processing and machine learning.

Following data retrieval, data processing is the subsequent step. Initially, tweets are temporarily housed in a list structure. Subsequently, they are persistently stored in a local PostgreSQL database for further in-depth analysis. After the database creation, a connection to it is established using the DBEaver management tool. The final phase of data storage involves the development of an R script that leverages the `RPostgreSQL` library to manage database operations.

The functionality of the “`rTweetScript.R`” script is extended by incorporating the capabilities of the `RPostgreSQL` library (R Studio Team, 2020). After a database connection is successfully established using a pre-configured driver, the tweet list is systematically saved to the “`original_tweets`” table using the `dbWriteTable` function. To ensure the preservation of data integrity and prevent overwriting during subsequent data downloads, the `append` parameter is judiciously employed. After the execution of “`rTweetScript.R`,” the stored tweets can be conveniently reviewed using the DBEaver tool by users (R Studio Team, 2020). The process of automatically storing tweets in the database is seamlessly managed by the task scheduler.

In the experimental pilot, tweets from individuals, whether they are BBVA customers or not, that mention BBVA are analyzed. As a result, a secondary script, denoted as “`firstCleaning.R`,” interlinked with the primary script, “`rTweetScript.R`,” has been devised (R Studio Team, 2020). This structural configuration unveils the tapestry of interconnected tools and processes orchestrating the experimental pilot, with tasks synergistically functioning with a high degree of automation.

The primary objective within the “`firstCleaning.R`” script is the filtration of tweets emanating from official BBVA accounts (R Studio Team, 2020). These designated accounts are documented within a dedicated table, aptly named “`bbva_accounts`.” For the duration of this pilot initiative, this catalog is vigilantly maintained through manual oversight. Subsequently, in the data cleansing phase, the systematic eradication of duplicate tweets is given precedence. Given the automated nature of data retrieval, conducted in daily increments with the acquisition of the latest 5,000 tweets, the potential for overlapping tweets from consecutive days looms large. Consequently, the upholding of the pristine integrity of the database by methodically purging duplicate records becomes imperative. This critical verification process is seamlessly integrated into the “`firstCleaning.R`” script (R Studio Team, 2020).

Ultimately, this methodological framework, characterized by its automated task scheduling and meticulous data management, facilitates a nuanced analysis of BBVA-associated tweets. This not only unveils customer sentiments but also provides a strategic edge in marketing analysis, enabling BBVA to engage more effectively with its digital audience and refine its marketing strategies in the competitive banking industry landscape.

This exploratory pilot study uses a robust approach grounded in machine learning and text mining disciplines for analyzing sentiment in tweets about BBVA. This methodology facilitates the categorization of text by the sentiment’s polarity (Niu et al., 2005), reflecting agreement or disagreement on a topic (Balahur et al., 2009), distinguishing between favorable and unfavorable news (Ku et al., 2005), and delineating pros and cons (Kim & Hovy, 2006). It encompasses the stages of

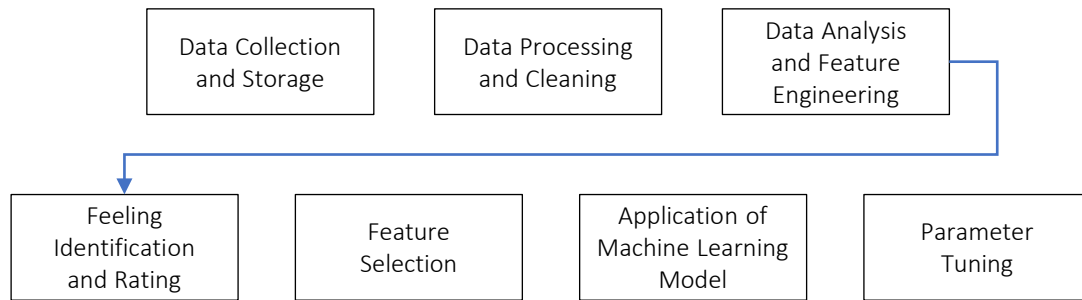


Figure 1. Sentiment analysis process

data collection, storage, processing, analysis, feature engineering, sentiment identification, model application, and parameter tuning. A comprehensive overview of each step in the methodology facilitates a clear understanding of the sentiment analysis process (Figure 1). Each step provides a clear insight into the sentiment analysis process, thereby enhancing marketing strategies through a deeper understanding of public opinion on social media.

Upon completing the foundational phase with R, the project transitions into a Python-executed phase, broadening the scope to include data analysis, feature definition, value extraction, and the construction of a machine learning-based model (Chang, 2018). This transition marks a shift toward a more nuanced analysis, necessary for crafting targeted marketing strategies.

Initially, the model operates without supervision, relying solely on input data devoid of labels. Faced with the choice of continuing with an unsupervised model or manually labeling a dataset for a more structured approach, this study opts for the latter, ensuring a refined analysis conducive to strategic marketing insights.

From an extensive collection of over 10,000 tweets, a subset of 1,000 tweets was analyzed and labeled, forming a supervised dataset. This process not only categorizes sentiments ranging from negative to positive but also allows for the exploration of the linguistic diversity within the tweets, a factor vital for understanding market sentiment nuances.

The data analysis delves into the intricate details of the 60+ fields present in each record. Preliminary exploration showed that 87.3% of the captured tweets are in Spanish, 5.1% in English, with the re-

maining 7.5% spanning other languages. The daily influx of tweets is found to fluctuate between 250 and 3,000, averaging around 1,500 tweets.

After the variables for analysis are determined and the labeled training dataset of 1,000 records is established, the methodology extends to feature engineering, where new variables are derived from the data, thereby increasing the precision of model training. This step is critical for identifying key sentiment drivers, which in turn, informs marketing strategies.

The primary goal of the study is discerned to be the sentiment of each tweet. To achieve this, text processing techniques based on Natural Language Processing (NLP), which enable the analysis, interpretation, and comprehension of the content of each tweet, are employed. For this purpose, the Google Cloud Natural Language (CNL) API is utilized. However, the choice of this API is accompanied by an understanding of the challenges inherent in determining sentiment using NLP tools. Many libraries, particularly free ones, are tailored for a specific language, usually English, because their underlying sentiment databases are language specific. While they may be accurate for one language, their effectiveness is diminished for others. This presents a significant challenge for the experimental pilot, as there arises a need to translate almost 94.9% of the tweets. Literal translations are often found to distort the original essence of a tweet.

Before finalizing the choice on the Google API, experiments with several libraries were conducted, but none were found to be satisfactory. NLTK (Natural Language Toolkit), while comprehensive and supporting multiple languages, primarily focuses on English, and its precision drops sig-

nificantly for other languages. TextBlob operates slower than NLTK and faces similar language-related challenges. The ‘Spanish-sentiment-analysis’ library, though designed for Spanish, lacks the extensive data repositories of its predecessors, resulting in less accurate outcomes.

Given these challenges, the decision for the experimental pilot was made to focus on analyzing tweets in both English and Spanish, which make up 92.4% of the total tweets. The procedure involves loading tweets from the database, tokenizing, eliminating stop words, and determining sentiment.

The methodological journey concludes with an update to the database with newly acquired sentiment data, ensuring that marketing strategies are informed by the most current public sentiment. This comprehensive approach, from data collection through to sentiment analysis, underpins the strategic marketing insights derived from the study, highlighting the potential of machine learning techniques to transform social media data into actionable marketing intelligence.

3. RESULTS

Table 1 illustrates the comparison between the manual classification of the 1,000 training tweets and the sentiment derived using the Cloud Natural Language Google API (CNL). To categorize the sentiment of the tweets, equidistant segments are established. Specifically, values less than -0.3 are indicated to denote a negative sentiment, values

between -0.3 and 0.3 are suggested to represent a neutral sentiment, and values greater than 0.3 are signified to imply a positive sentiment. The values of the ordinate axis are indicated to represent the number of tweets. Figures 2, 3, and 4 show the histograms obtained using the feeling obtained from CNL. Figure 5 shows the distribution of the classification result in a confusion matrix.

Table 1. Distribution of tweets based on the manual classification and automatic classification (CNL) of the feeling of the tweets

Feeling	Manual Classification	CNL
Positive	68	221
Neutral	559	475
Negative	373	304

Tweets that have been correctly classified (True Positive/Negative) are positioned on the diagonal, with their respective values being: $T_{positive} = 40$, $T_{neutral} = 313$, and $T_{negative} = 163$. Any other classifications are either False Positives or False Negatives. The current accuracy of the model is calculated as $Accuracy (Acc) = (T_{positive} + T_{neutral} + T_{negative}) / Total$. This gives an accuracy value of 0.516.

The sentiment analysis provided by the Google API aligns with the manual classification just over half the time, a mere 50%, which is only slightly better than random guessing. This indicates a need for further refinement of the model. To enhance the model’s accuracy, the following additional considerations are considered to derive new features:

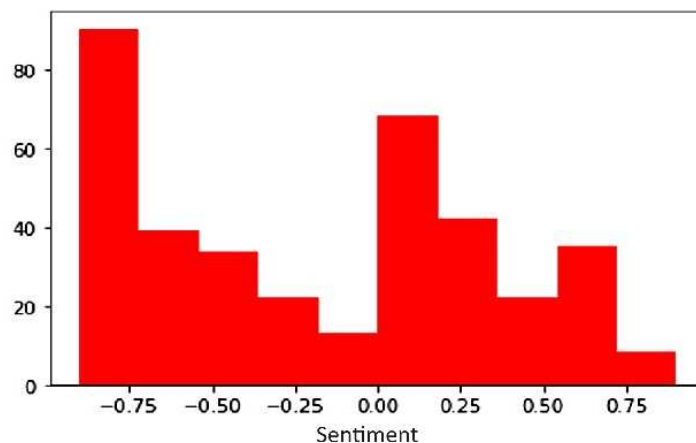


Figure 2. Negative tweet feeling histogram

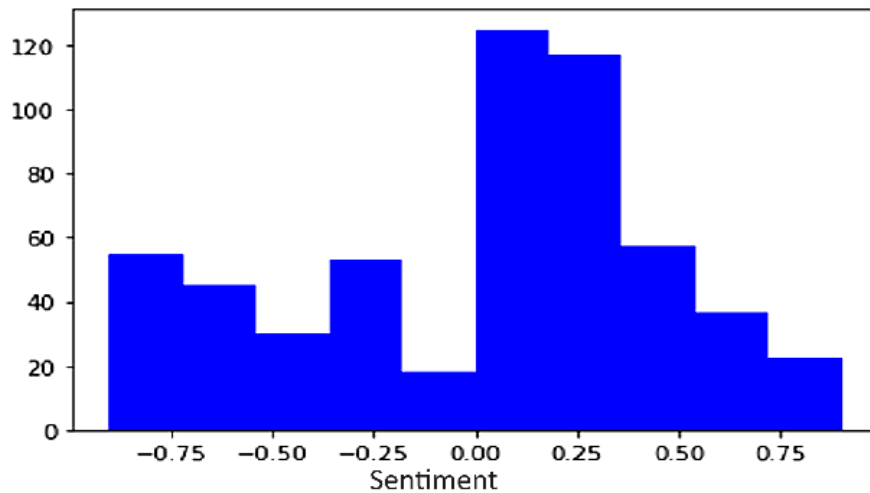


Figure 3. Neutral tweet feeling histogram

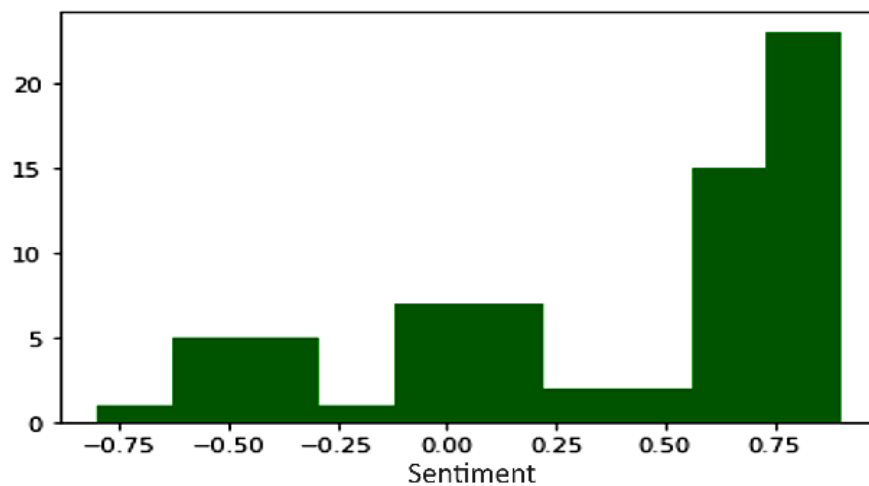


Figure 4. Positive tweet feeling histogram

1. Tweet sentiment concentrations: As depicted in Figure 2, there are two distinct clusters of negative tweets. One is close to -1, and the other is around 0. The latter cluster is attributed to tweets mentioning potential “Corporate Reputation” issues. While these are not direct customer complaints, they convey negativity concerning potential reputational damage. Figures 3 and 4 display expected distributions, with positive sentiments nearing the value of 1 and neutral sentiments centered around 0.
2. BBVA’s sponsorship: It is crucial to note that Banco Bilbao Vizcaya Argentaria (BBVA) was the official sponsor of the Professional Soccer League in Spain. In the manual classification, tweets related to this were labeled as neutral. However, the sentiment derived can swing between positive and negative, influenced by match summaries, statistics, sports news, and virtual soccer leagues.
3. Emoticon interpretation: Emoticons often convey sentiments or opinions. The Python Demoji library is employed to analyze texts, returning the emoticons they contain and their respective descriptions (Brad, 2019). All stored tweets are processed using this library, and the results are cross-referenced with pre-defined lists of positive and negative emoticons.
4. Capital letter usage: The use of capital letters can indicate strong emotions such as anger, rage, or extreme joy. The model calculates the percentage of capital letters in a tweet to consider its potential sentiment.

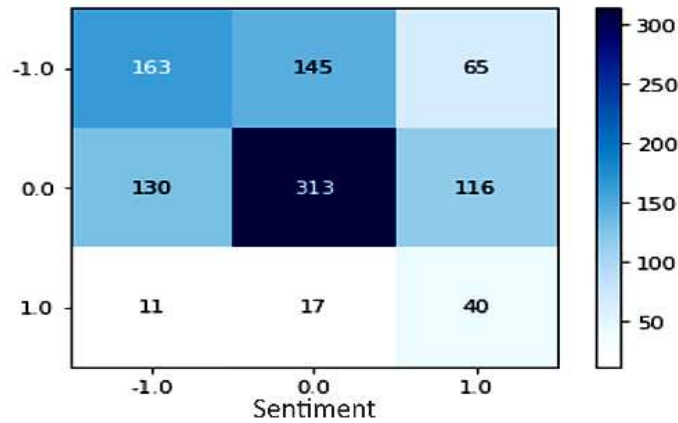


Figure 5. Confusion matrix classification vs feeling

5. Influential words: Certain words carry strong positive or negative connotations, significantly influencing tweet classification. Two additional lists of such words are created. Additionally, a linguistic stemming process is applied, especially important for the Spanish language. This process extracts the stem or root of each word in a tweet, representing its declinable forms. This simplification aids in achieving a more precise subsequent classification, considering singular forms for nouns, masculine singular for adjectives, and infinitives for verbs.

As illustrated in Figure 6, these refinements enable the precise classification of close to 3,000 tweets as either positive or negative. This advancement underscores the importance of continuous model optimization and the integration of diverse features to achieve more accurate sentiment analysis results.

Before accepting all the variables of the first selection as good, it is necessary to verify that there is a relationship between them. If there were correlation between variables, it would mean redundant data and, therefore, they could be dispensed with without losing a large amount of information.

In the study, the utilization of a correlation matrix, constructed with the Pandas library and visualized using the Seaborn library, is crucial in elucidating the relationships between variables through correlation coefficients. A coefficient of 1 is always displayed on the matrix's diagonal, which is expected since any variable will invariably have a perfect correlation with itself. Coefficients that are close to 0 are signified to denote little to no linear relationship between the paired variables. In contrast, coefficients nearing 1 are indicated to suggest a strong positive linear relationship, implying that as one variable increases, the other is

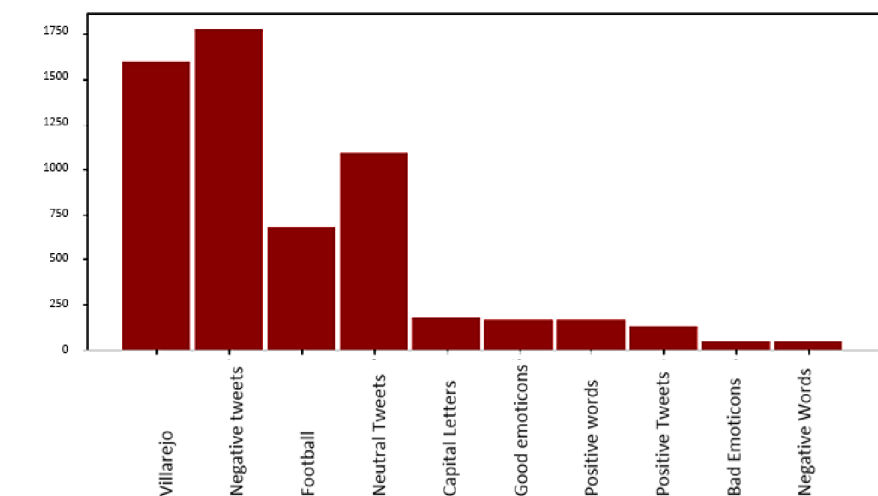


Figure 6. Histogram over 3,000 tweets with new properties

tended to increase as well. Conversely, coefficients approaching -1 are signified to denote a strong negative linear relationship, meaning that as one variable increases, the other is likely to decrease. When a direct or inverse correlation is observed between two variables, one of them should be discarded to avoid redundancy and ensure the clarity of the analysis. In the matrix, the absence of extreme values is indicated, a sign that the variables are not correlated and therefore, they will be able to be used in the machine learning algorithm. This selection of features allowed reducing the training time, simplifying the model, and reducing the possibility of overfitting.

The model under consideration is based on principles of supervised learning, utilizing a dataset that has been manually labeled for training. Initially, 1,000 labeled tweets were available, but this number has been expanded to 3,000. Of this total, 60% are allocated to the training set, and the remaining 40% to the testing set. Any additional tweets serve as a validation set, providing fresh data that the model has not encountered, allowing for an assessment of classifier performance to be made.

To ensure representativeness and avoid biases, a proportionate distribution of tweets across various categories, such as emoticons (102 tweets), corporate reputation (1,598 tweets), football (667 tweets), and others, should be had by both the training and testing sets. A Python script is developed to enumerate tweets from each category and distribute them: 60% to the training set and 40% to the testing set.

Several supervised learning algorithms suitable for classification tasks, including SVM, XGBoost, Random Forest, and Neural Networks, are evaluated. A primary focus is placed on analyzing their confusion matrices to identify the highest precision in classifying negative tweets, as these are crucial for immediate interventions. Neutral tweets, while less actionable, and positive tweets, which might be used for retweets or promotional campaigns, are also assessed.

The confusion matrices are derived in Figures 7, 8, 9, and 10, respectively, by applying the SVM, XGBoost, Random Forest, and neural network algorithms.

The study evaluated the performance of various machine learning algorithms in analyzing sentiments expressed in tweets related to BBVA. The algorithms under scrutiny included Support Vector Machines (SVM), XGBoost, Random Forest, and Neural Networks, with a particular focus on their precision in sentiment classification. Among these, the XGBoost algorithm emerged as the standout, demonstrating a remarkable precision rate of 91.2%. This high precision is not only indicative of the effectiveness of machine learning techniques in parsing the complexities of large-scale social media data but also underscores the nuanced capability of these algorithms to accurately classify sentiments across a spectrum of contexts, including negative and neutral tweets, with precision rates of 94-95% and 87-88%, respectively.

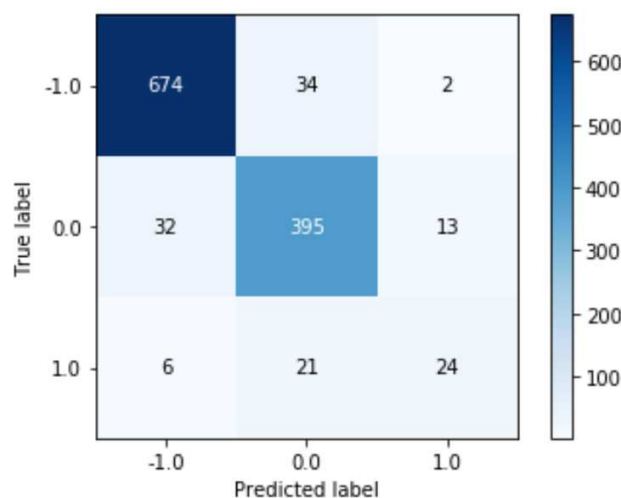


Figure 7. SVM confusion matrix

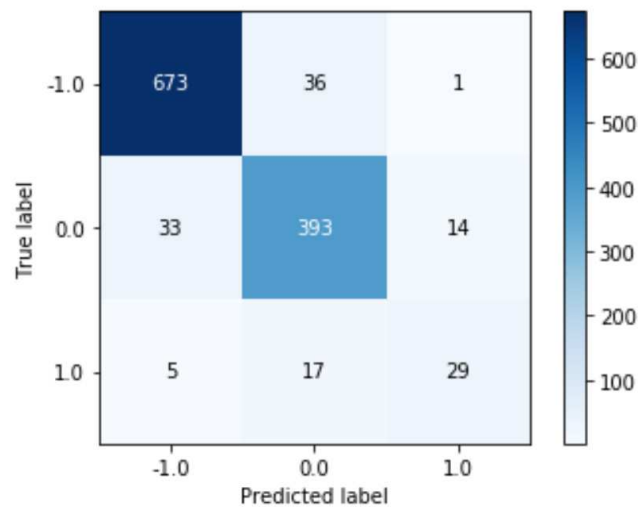


Figure 8. XGBoost confusion matrix

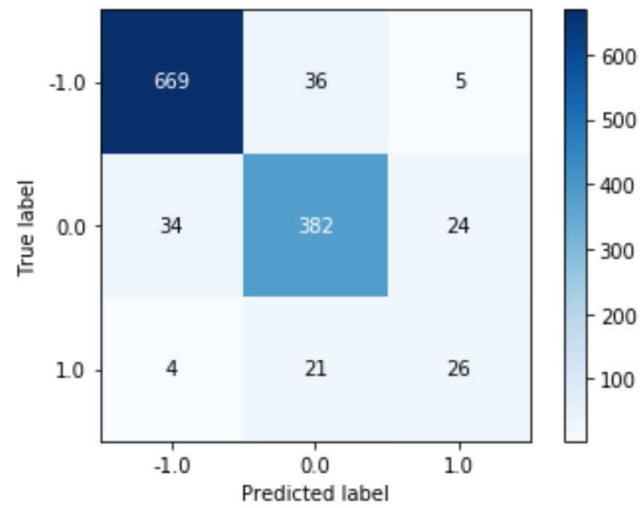


Figure 9. Random Forest confusion matrix

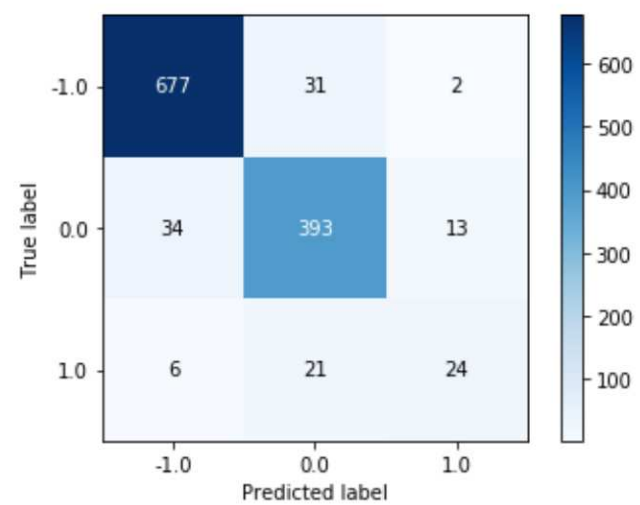


Figure 10. Confusion matrix neural networks

Table 2 compares the precision of each algorithm, revealing a closely contested field, yet highlighting XGBoost’s slight edge.

Table 2. Comparison of precision between algorithms

Algorithm	Precision
SVM	91%
XGBOOST	91.2%
RANDOM FOREST	89.7%
NEURAL NETWORKS	91.1%

Given the narrow precision margins, particular attention was directed toward the algorithms’ ability to predict negative sentiments accurately, necessitating immediate intervention for customer service improvement.

Further analysis, detailed in Table 3, delves into the misclassification rates and accuracies for positive, neutral, and negative tweets.

Table 3 highlights a significant disparity in the classification of positive tweets, underscoring the critical importance of selecting an algorithm that maximizes precision across all sentiment categories. XGBoost’s superior performance, particularly in accurately predicting negative and positive sentiments, positions it as the optimal algorithm for this study’s objectives.

Table 3. Comparison of results between algorithms

Data	SVM	XGBOOST	RANDOM FOREST	NEURAL NETWORKS
% MisClassifiedPOS	53%	43%	49%	53%
% MisClassifiedNEU	10%	10%	13%	11%
% Mis ClassifiedNEG	5%	5%	6%	5%
Accuracy Positives	62%	66%	47%	62%
Accuracy Neutrals	88%	88%	87%	88%
Accuracy Negatives	95%	95%	95%	94%

Table 4. Comparison improvement of results of the XGBoost algorithm

Data	Result XGBOOST	Improvement over second best	Improvement over last
Precision	91.2%	+0.1%	+1.4%
% MisClassifiedPOS	43%	+6%	+10%
% MisClassifiedNEU	10%	0%	+3%
% Mis ClassifiedNEG	5%	0%	+1%
Accuracy Positives	66%	+4%	+19%
Accuracy Neutrals	88%	0%	+1%
Accuracy Negatives	95%	0%	+1%

To quantify the enhancements offered by XGBoost over its competitors, Table 4 provides a comparative overview of the performance enhancements of the XGBoost algorithm relative to the other algorithms.

The findings unequivocally support the hypotheses posited at its outset. Specifically, the hypothesis that the XGBoost algorithm would demonstrate superior accuracy in sentiment classification on BBVA’s social media platforms, compared to other machine learning algorithms, is confirmed with a precision rate of 91.2%. This not only validates the hypothesis but also emphasizes the critical role of advanced machine learning techniques, like XGBoost, in enhancing sentiment analysis, particularly in the banking sector where real-time customer feedback is invaluable.

Moreover, the results affirm the assumption that employing advanced machine learning techniques for sentiment analysis can significantly enhance strategic decision-making and customer engagement in the banking industry. The high accuracy in negative sentiment classification, achieved by the XGBoost algorithm, illustrates its potential to enable immediate and informed responses to customer feedback, thereby fostering a more dynamic and responsive customer engagement framework for BBVA.

4. DISCUSSION

This study not only aligns with the high precision benchmarks set by the OneR Classifier as reported by Singh et al. (2017) but also surpasses the performance of other algorithms like Naïve Bayes, J-48, and BFTree in their analysis. The precision of the XGBoost algorithm, standing at 91.2%, indicates a significant improvement over the results obtained by Huq et al. (2017) using SVM and k-NN on Twitter data, which achieved accuracy rates between 58.39% and 79.99%. Similarly, while Amolik et al. (2016) observed that Naïve Bayes (NB) had better precision compared to SVM, the findings of this study suggest that the advanced machine learning techniques, especially XGBoost and neural networks, offer more reliable alternatives for sentiment analysis. This observation is in concordance with the conclusions drawn by Başarslan and Kayaalp (2020), further establishing the utility of these algorithms in extracting meaningful insights from social media data.

The superior performance of the XGBoost algorithm can be attributed to its ability to handle large-scale data efficiently, its robustness against overfitting, and its adaptability to various types of data, which are critical attributes for analyzing the vast and diverse datasets typical of social media Aarshay (2016). Neural networks, with their deep learning capabilities, also show promising results, particularly in deciphering the complexities of human language expressed through tweets. This effectiveness is reflective of the continuous advancements in machine learning technologies and their increasing applicability in real-world scenarios, like sentiment analysis on social media platforms.

The continuous evolution of social networks and their significant impact on a company's brand im-

age underscore the importance of this study's focus on Twitter. However, as the social media landscape expands to include platforms like Instagram and TikTok, the dynamics of sentiment expression and interpretation may vary. The textual content on Twitter, which is more concise due to character limits, offers a different context for sentiment analysis compared to the visual and video content predominant on Instagram and TikTok. This variation suggests that the algorithms' performance might differ when applied to these other platforms, pointing to the necessity for future research to adapt and test machine learning techniques across diverse social media environments.

Looking ahead, the exploration of sentiment analysis on emerging social media platforms represents a promising avenue for research. Additionally, the integration of multimodal data analysis, which includes text, images, and videos, could offer a more holistic understanding of social media sentiments. Another potential direction involves the development of algorithms that can automatically adjust to the nuances of different languages and dialects, enhancing the accuracy of sentiment analysis in a global context. This approach is particularly relevant for global brands like BBVA, which engage with a diverse customer base across various cultural and linguistic backgrounds.

This study contributes to the growing body of knowledge on the application of machine learning algorithms in sentiment analysis, demonstrating the potential of advanced techniques like XGBoost and neural networks. The findings underscore the necessity for ongoing research to keep pace with the rapid evolution of social media and the complex dynamics of user interactions, ensuring that sentiment analysis remains a powerful tool for understanding and engaging with the digital audience.

CONCLUSION

The aim of this paper was to rigorously evaluate and compare the effectiveness of specific machine learning algorithms for sentiment analysis on Banco Bilbao Vizcaya Argentaria (BBVA)'s social media platforms, with a particular focus on assessing the precision and practical utility of these methods. This investigation has successfully demonstrated the superior effectiveness and precision of machine learning algorithms in conducting sentiment analysis, highlighting their invaluable role in strategic decision-making processes related to customer relationship management and market analysis.

This study's results offer compelling evidence of the efficacy of machine learning algorithms, especially XGBoost, in accurately analyzing and classifying sentiments expressed in social media content. This insight is instrumental for BBVA and similar financial institutions aiming to leverage social media analytics for improved customer interaction and market responsiveness, marking a significant stride toward the integration of advanced data analytics in strategic business operations.

Among the algorithms tested, the XGBoost algorithm stood out for its exceptional precision, achieving a 91.2% success rate in sentiment classification. This finding not only highlights the capability of machine learning technologies to provide detailed and accurate sentiment analyses on social media platforms but also emphasizes their strategic importance for companies like BBVA. By leveraging these analyses, BBVA can gain significant advantages in understanding and engaging with its digital audience, thereby enhancing business strategies, and fostering improved customer relationships.

In conclusion, this study affirms the pivotal role of advanced machine learning techniques in the realm of social media sentiment analysis, offering profound insights for financial institutions in their pursuit of strategic engagement and market responsiveness. As such, machine learning stands as a cornerstone technology that enables businesses to navigate the complexities of digital communication landscapes effectively, ensuring that they remain attuned to the evolving dynamics of customer sentiment and market trends.

AUTHOR CONTRIBUTIONS

Conceptualization: Luis Miguel Garay Gallastegui, Ricardo Reier Forradellas, Sergio Luis Nández Alonso.

Data curation: Luis Miguel Garay Gallastegui.

Formal analysis: Luis Miguel Garay Gallastegui, Ricardo Reier Forradellas, Sergio Luis Nández Alonso.

Investigation: Luis Miguel Garay Gallastegui, Ricardo Reier Forradellas, Sergio Luis Nández Alonso.

Methodology: Luis Miguel Garay Gallastegui, Ricardo Reier Forradellas, Sergio Luis Nández Alonso.

Software: Luis Miguel Garay Gallastegui.

Supervision: Luis Miguel Garay Gallastegui, Ricardo Reier Forradellas, Sergio Luis Nández Alonso.

Validation: Luis Miguel Garay Gallastegui, Ricardo Reier Forradellas, Sergio Luis Nández Alonso.

Writing – original draft: Luis Miguel Garay Gallastegui, Ricardo Reier Forradellas, Sergio Luis Nández Alonso.

Writing – review & editing: Luis Miguel Garay Gallastegui, Ricardo Reier Forradellas, Sergio Luis Nández Alonso.

REFERENCES

1. Aarshay, J. (2016). *Complete guide to parameter tuning in XGBoost with codes in Python*. GitHub. Retrieved April 3, 2020, from <https://github.com/analyticsvidhya/Complete-Guide-to-Parameter-Tuning-in-XGBoost-with-codes-in-Python/blob/master/Complete%20Guide%20to%20Parameter%20Tuning%20in%20XGBoost%20with%20codes%20in%20Python.ipynb>
2. American Bankers Association. (2023). *Banks report active social media engagement*. Retrieved from <https://www.aba.com/news-research/analysis-guides/social-media-in-banking-2023-report>
3. Amolik, A., Jivane, N., Bhandari, M., & Venkatesan, M. (2016). Twitter sentiment analysis of movie reviews using machine learning techniques. *International Journal of Engineering and Technology*, 7(6), 1-7. Retrieved from https://www.researchgate.net/publication/291837156_Twitter_Sentiment_Analysis_of_Movie_Reviews_using_Machine_Learning_Techniques
4. Anshari, M., Almunawar, M. N., Lim, S. A., & Al-Mudimigh, A. (2019). Customer relationship management and big data enabled: Personalization & customization of services. *Applied Computing and Informatics*, 15(2), 94-101.
5. Auter, Z., & Fine, J. (2018). Social media campaigning: Mobilization and fundraising on Facebook. *Social Science Quarterly*, 99(1), 185-200. <https://doi.org/10.1111/ssqu.12391>
6. Balahur, A., Kozareva, Z., & Montoyo, A. (2009). Determining the

- polarity and source of opinions expressed in political debates. In A. Gelbukh (Ed.), *Computational linguistics and intelligent text processing* (pp. 468-480). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-642-00382-0_38
7. Başarslan, M. S., & Kayaalp, F. (2020). Sentiment analysis with machine learning methods on social media. *Advances in Distributed Computing and Artificial Intelligence Journal*, 9(3), 5-15. <https://doi.org/10.14201/AD-CAIJ202093515>
 8. Bonnet, D., & Westerman, G. (2021). The new elements of digital transformation. *MIT Sloan Management Review*, 62, 82-89. Retrieved from <https://sloanreview.mit.edu/article/the-new-elements-of-digital-transformation/>
 9. Borgatti, S. P., Mehra, A., Brass, D. J., & Labianca, G. (2009). Network analysis in the social sciences. *Science*, 323(5916), 892-895. <https://doi.org/10.1126/science.1165821>
 10. Boyd, D. M., & Ellison, N. B. (2007). Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 210-230. <https://doi.org/10.1111/j.1083-6101.2007.00393.x>
 11. Brad, S. (2019). *Demoji. Project description*. Retrieved March 15, 2022, from <https://pypi.org/project/demoji/>
 12. Brandes, U., Freeman, L. C., & Wagner, D. (2013). Social networks. In R. Tamassia (Ed.), *Handbook of graph drawing and visualization* (pp. 805-839). London: Chapman & Hall. Retrieved from <https://d-nb.info/1112610073/34>
 13. Bródka, P., Skibicki, K., Kazienko, P., & Musiał, K. (2011). A degree centrality in multi-layered social network. *2011 International conference on computational aspects of social networks (CASoN)* (pp. 237-242). IEEE. Retrieved from <https://ieeexplore.ieee.org/document/6085951>
 14. Chang, V. (2018). A proposed social network analysis platform for big data analytics. *Technological Forecasting and Social Change*, 130, 57-68. <https://doi.org/10.1016/j.techfore.2017.11.002>
 15. Che, X., & Ip, B. (2017). *Social networks in China*. Chandos Publishing.
 16. Dickerson, J. P., Kagan, V., & Subrahmanian, V. S. (2014). Using sentiment to detect bots on Twitter: Are humans more opinionated than bots? *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)* (pp. 620-627). IEEE. Retrieved from <https://ieeexplore.ieee.org/document/6921650>
 17. García-Galera, M. C., Del-Hoyo-Hurtado, M., & Fernández-Muñoz, C. (2014). Engaged youth in the Internet: The role of social networks in social active participation. *Comunicar: Revista Científica Iberoamericana de Comunicación y Educación*, 43(2), 35-43. Retrieved from <http://eprints.rclis.org/23377/1/c4303en.pdf>
 18. Gartner. (2023). *The state of marketing budget and strategy 2022. Insights from Gartner's annual CMO spend survey*. Retrieved from <https://www.gartner.co.uk/en/marketing/research/annual-cmo-spend-survey-research>
 19. Gill, S. S., Tuli, S., Xu, M., Singh, I., Singh, K. V., Lindsay, D., Tuli, S., Smirnova, D., Singh, M., Jain, U., Pervaiz, H., Sehgal, B., Kaila, S. S., Misra, S., Aslanpour, M. S., Mehta, H., Stankovski, V., & Garraghan, P. (2019). Transformative effects of IoT, Blockchain and Artificial Intelligence on cloud computing: Evolution, vision, trends and open challenges. *Internet of Things*, 8, Article 100118. <https://doi.org/10.1016/j.iot.2019.100118>
 20. Granovetter, M. S. (1973). The strength of weak ties. *American Journal of Sociology*, 78(6), 1360-1380. Retrieved from <https://snap.stanford.edu/class/cs224w-readings/granovetter73weakties.pdf>
 21. Gutiérrez-Batista, K., Campaña, J., Martínez-Folgoso, S., Vila, M., & Martín-Bautista, M. (2016). About the effects of sentiments on topic detection in social networks. *International Journal of Design & Nature and Ecodynamics*, 11(3), 387-395. <https://doi.org/10.2495/dne-v11-n3-387-395>
 22. Hariri, R. H., Fredericks, E. M., & Bowers, K. M. (2019). Uncertainty in big data analytics: Survey, opportunities, and challenges. *Journal of Big Data*, 6(1), Article 44. <https://doi.org/10.1186/s40537-019-0206-3>
 23. Hootsuite Inc. (2022). *Social media trends 2022*. Retrieved March 23, 2022, from <https://blog.hootsuite.com/es/informe-digital-estadisticas-de-redes-sociales/>
 24. Huq, M. R., Ali, A., & Rahman, A. (2017). Sentiment analysis on Twitter data using KNN and SVM. *(IJACSA) International Journal of Advanced Computer Science and Applications*, 8(6), 19-25. Retrieved from <https://thesai.org/Publications/ViewPaper?Volume=8&Issue=6&Code=IJACSA&SerialNo=3>
 25. IAB Spain. (2021). *Estudio de redes sociales 2021 [Social media study 2021]*. (In Spanish). Retrieved from <https://iabspain.es/estudio/estudio-de-redes-sociales-2021/>
 26. Isaak, J., & Hanna, M. J. (2018). User data privacy: Facebook, Cambridge Analytica, and privacy protection. *Computer*, 51(8), 56-59. <https://doi.org/10.1109/MC.2018.3191268>
 27. Kim, D., & Ellison, N. (2022). From observation on social media to offline political participation: The social media affordances approach. *New Media & Society*, 24(12), 2614-2634. <https://doi.org/10.1177/1461444821998346>
 28. Kim, S. M., & Hovy, E. (2006). Automatic identification of pro and con reasons in online reviews. *Proceedings of the COLING/ACL Main Conference Poster Sessions* (pp. 483-490). <http://dx.doi.org/10.3115/1273073.1273136>
 29. Ku, L.W., Lee, L.Y., Wu, T.H., & Chen, H. H. (2005). Major topic detection and its application to opinion summarization. *Proceedings of the ACM Special Interest Group on Information Retrieval (SIGIR)* (pp. 627-628). Salvador, Brazil. <http://dx.doi.org/10.1145/1076034.1076161>

30. Kumar, A., Sangwan, S. R., & Nayyar, A. (2020). Multimedia social big data: Mining. In *Multimedia big data computing for IoT applications* (pp. 289-321). Singapore: Springer. http://dx.doi.org/10.1007/978-981-13-8759-3_11
31. Lee, Y., & Kim, K. H. (2021). Enhancing employee advocacy on social media: The value of internal relationship management approach. *Corporate Communications: An International Journal*, 26(2), 311-327. <https://doi.org/10.1108/CCIJ-05-2020-0088>
32. Micu, A., Micu, A. E., Geru, M., & Lixăndroi, R. C. (2017). Analyzing user sentiment in social media: Implications for online marketing strategy. *Psychology & Marketing*, 34(12), 1094-1100. <https://doi.org/10.1002/mar.21049>
33. Milgram, S. (1967). The small-world problem. *Psychology Today*, 1(1), 61-67. Retrieved from <http://snap.stanford.edu/class/cs224w-readings/milgram67smallworld.pdf>
34. Molu, F., Findik, N., & Dalci, M. (2014). Enhancing user experience in financial services. *International Journal of E-Services and Mobile Applications*, 6(2), 12-22. <https://doi.org/10.4018/ijesma.2014040102>
35. Moreno, J. L. (1953). *Who shall survive? Foundations of sociometry, group psychotherapy and socio-drama* (2nd ed.). Beacon House. Retrieved from <https://psycnet.apa.org/record/1954-04178-000>
36. Moreno, J. L. (1955). The sociometric school and the science of man. *Sociometry*, 18(4), 15-35. <https://doi.org/10.2307/2785839>
37. Nieto, B. G. (2013). Nuevos espacios comunicativos para las organizaciones: Las redes sociales [New communicative spaces for the organizations: The social networks]. *Historia y Comunicación Social*, 18, 731-741. (In Spanish). Retrieved from <https://pdfs.semanticscholar.org/d4bd/acea5fc283d1e71803534976334f-0936dae7.pdf>
38. Niu, Y., Zhu, X., Li, J., & Hirst, G. (2005). Analysis of polarity information in medical text. *Proceedings of the American Medical Informatics Association 2005 Annual Symposium* (pp. 570-574). Retrieved from <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1560818/>
39. Nouwens, M., Griggio, C. F., & Mackay, W. E. (2017). "WhatsApp is for family; Messenger is for friends." Communication places in app ecosystems. *Proceedings of the 2017 CHI conference on human factors in computing systems* (pp. 727-735). <https://doi.org/10.1145/3025453.3025484>
40. Okamoto, K., Chen, W., & Li, X. Y. (2008). Ranking of closeness centrality for large-scale social networks. In *International workshop on frontiers in algorithmics* (pp. 186-195). Berlin, Heidelberg: Springer. https://doi.org/10.1007/978-3-540-69311-6_21
41. Ozili, P. K., & Nañez Alonso, S. L. (2024). Central bank digital currency adoption challenges, solutions, and a sentiment analysis. *Journal of Central Banking Theory and Practice*, 13(1), 133-165. <https://doi.org/10.2478/jcbtp-2024-0007>
42. Piatak, J., & Mikkelsen, I. (2021). Does social media engagement translate to civic engagement of-fline? *Nonprofit and Voluntary Sector Quarterly*, 50(5), 1079-1101. <https://doi.org/10.1177/0899764021999444>
43. Rossmann, A. (2018). Digital maturity: Conceptualization and measurement model. *Thirty Ninth International Conference on Information Systems*. San Francisco, CA, USA. Retrieved from https://www.researchgate.net/publication/345760193_Digital_Maturity_Conceptualization_and_Measurement_Model
44. Singh, J., Singh, G., & Singh, R. (2017). Optimization of sentiment analysis using machine learning classifiers. *Human-centric Computing and Information Sciences*, 7, Article 32. <https://doi.org/10.1186/s13673-017-0116-3>
45. Stephenson, K., & Zelen, M. (1989). Rethinking centrality: Methods and examples. *Social Networks*, 11(1), 1-37. [https://doi.org/10.1016/0378-8733\(89\)90016-6](https://doi.org/10.1016/0378-8733(89)90016-6)
46. Sudha, M., & Sheena, K. (2020). Impact of influencers in consumer decision process: The fashion industry. *Interdisciplinary Journal on Law, Social Sciences and Humanities*, 1(2), 1-13. <http://dx.doi.org/10.19184/ijl.v1i1.19146>
47. Sun, K., Wang, H., & Zhang, J. (2022). The impact factors of social media users' forwarding behavior of Covid-19 vaccine topic: Based on empirical analysis of Chinese Weibo users. *Frontiers in Public Health*, 10. <https://doi.org/10.3389/fpubh.2022.871722>
48. The Financial Brand. (2022). *Social Media Trends for Q1 2022: Top Banks Losing Facebook Followers*. Retrieved from <https://thefinancialbrand.com/news/social-media-banking/social-media-trends-bank-facebook-likes-up-as-credit-unions-struggle-154079/>
49. Veissi, I. (2017). *Influencer marketing on Instagram* (Master's Thesis). Haaga-Helia University of Applied Sciences. Retrieved from <https://www.theseus.fi/handle/10024/135448>
50. Wasserman, L., & Kass, R. E. (1995). A reference Bayesian test for nested hypotheses and its relationship to the Schwarz criterion. *Journal of the American Statistical Association*, 90(431), 928-934. <https://doi.org/10.2307/2291327>
51. Williams, J. R. (2019). The use of online social networking sites to nurture and cultivate bonding social capital: A systematic review of the literature from 1997 to 2018. *New Media & Society*, 21(11-12), 2710-2729. <https://doi.org/10.1177/1461444819858749>
52. Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data mining: Practical machine learning tools and techniques*. Elsevier.
53. Xu, J., Hu, Zh., Zou, Zh., Zou, J., Hu, X., Liu, L., & Zheng, L. (2020). Design of smart unstaffed retail shop based on IoT and artificial intelligence. *IEEE Access*, 8, 147728-147737. <https://doi.org/10.1109/ACCESS.2020.3014047>
54. Zhang, J., & Luo, Y. (2017). Degree centrality, betweenness centrality, and closeness centrality in social network. *IProceedings of the 2017 2nd International Conference on Modelling, Simulation and Applied Mathematics (MSAM2017)* (pp. 300-303). Atlantis Press. Retrieved from <https://www.atlantis-press.com/proceedings/msam-17/25874733>