International Journal of Economics and Financial Issues

ISSN: 2146-4138

available at http: www.econjournals.com

International Journal of Economics and Financial Issues, 2025, 15(2), 344-354.

Behavioral Economics and Stock Market Sentiments in Investment Decisions in Mexico: Web Scraping, Natural Language Processing, and Pearson Correlation of Scores

Sandra Yolotzin Zúñiga-Cedillo, Ana Lorena Jiménez-Preciado, Salvador Cruz-Aké, Francisco Venegas-Martínez*

Instituto Politécnico Nacional, Escuela Superior de Economía, México. *Email: fvenegas1111@yahoo.com.mx

Received: 28 October 2024 **Accepted:** 30 January 2025 **DOI:** https://doi.org/10.32479/ijefi.18142

ABSTRACT

This paper analyzes market sentiment using digital news to understand the influence of cognitive biases (behavioral economics) on the investment decisions of economic agents. Web scraping techniques were employed to gather news about "América Movil" stock, the company with the highest market capitalization value. Subsequently, Natural Language Processing (NLP) tools are used to determine sentiment polarity scores, Pearson correlation sentiment scores are carried out. The data consists of digital news articles related to América Movil and its corresponding historical stock price data, providing a basis for sentiment analysis and trend comparison. The findings expose a consistent trend between sentiment polarity scores and stock price movements. Moreover, economic and political factors significantly influence sentiment serving as early indicators of stock price behavior. The model has practical implications for behavioral economics, demonstrating how news profoundly influences investment decisions and shapes the behavior of the Mexican market. This research uniquely combines sentiment analysis of digital news with stock price data to highlight the impact of cognitive biases on investment decisions in the Mexican market.

Keywords: Behavioral Economics, Sentiment Analysis, Web Scraping, Cognitive Biases, Natural Language Processing, Stock Prices
JEL Classifications: G14, D03, C81

1. INTRODUCTION

Mainstream economists consider all economic agents rational when making financial decisions (Muth, 1961; Lucas and Prescott, 1971; Lucas, 1978). However, the Behavioral Economics (BE) approach challenged this paradigm. They recognize the importance of psychology in the economic agent's decision-making process. Behaviorists showed that individuals do not make pure risk-neutral financial decisions; those decisions are influenced by emotional, cognitive, and social factors (Kahneman and Tversky, 1984).

Under high uncertainty, an investor may base their decision on fear generated by negative news without conducting additional research. A similar process can occur during a euphoric rally about an economic asset. Because of that, it is crucial to understand how investors process information and how cognitive biases influence their investment decisions (Fernández et al., 2017; Shleifer and Summers,1990; Shiller, 1984).

Cognitive biases are mental shortcuts that speed up decision-making by using available information. The challenge of measuring the effect of cognitive bias on financial market behavior remains a challenge. To address this issue, researchers such as Liu (2015), Zhao et al. (2016), Agarwal and Natarajan (2023), and Shapiro et al. (2022) have turned to sentiment analysis to extract the emotions and feelings of economic agents. Sentiment analysis proposes identifying and classifying opinions or sentiments into positive, negative, or neutral categories.

Today, some online news platforms are considered reliable information sources, and access to their information and content determines financial decision-making. With their real-time updates and broad reach, these platforms play a crucial role in shaping investor sentiment. Thus, they can be a good source of information for sentiment analysis as long as they are structured and accessible texts that facilitate data collection and processing for sentiment analysis.

On the other hand, the explosion of artificial intelligence tools offers methodologies that efficiently search and analyze large volumes of information to measure the "market sentiment," its effect on real financial markets, and cognitive biases. Several researchers have employed this approach to examine events of great economic relevance. One example is the study by Costola et al. (2023) that investigate how the news affected investor behavior from January to June 2020 during the COVID-19 pandemic in the financial markets. This unprecedented event experienced periods of rapid information dissemination, making it difficult for investors to process. Another prominent example is the work of Sun et al. (2021) that proposes the theory that the cryptocurrency market exhibits high volatility due to changes in investor sentiment. Both studies applied machine learning techniques to analyze the influence of investor sentiment on market behavior.

This study analyzes market sentiment using digital news from the Mexican company América Movil, SAB de CV, AMXB.MX, to understand cognitive biases on investment decisions. América Movil has the largest market capitalization in México. It is a telecommunications company currently in the S&P/BMV IPC index and is the company with the largest market capitalization in Mexico. This paper hypothesizes that cognitive biases influence investment decisions reflected in market behavior.

This research uses a Natural Processing Language (NLP)-based model. It is built to identify words and perform sentiment analysis in digital news texts. It will determine whether words carry positive or negative sentiments. This approach will allow an understanding of how these biases influence investor behavior, analyzing the volatility of stocks in the market.

The paper mixed methodology combines qualitative and quantitative variables. Qualitative variables include emotional tone (positive, negative, or neutral), opinions' polarity, and text subjectivity. On the other hand, quantitative variables involve the frequency of specific keywords, the length of the text, and the use of punctuation marks. The news texts were collected from January 2023 to April 2024.

The paper is organized as follows: section 1 discusses on some theoretical issues about behavioral economics, namely cognitive biases and measuring market sentiment; section 2 studies the behavior of Mexican market sentiment indices over time and how they influence the price of financial instruments; section 3 develops the model for measuring investor bias based on news sentiment analysis; finally, section 4 gives conclusions and recommendations.

2. COGNITIVE BIASES IN INVESTMENT DECISIONS AND MARKET SENTIMENT

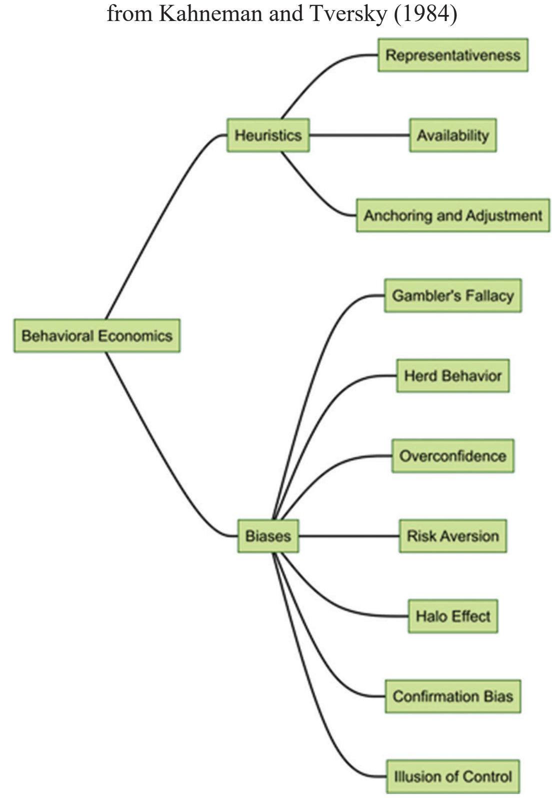
All investments face the yield–risk trade-off. Financial theory is built on such facts, but most assessments depend heavily on risk-neutral and entirely rational assumptions. In the mainstream financial theory, rational investment depends on the market information. However, economic agents do not understand or process the market information similarly.

BE points out that cognitive and heuristic biases influence decisions, that is, previous experiences and context can alter decisions and how economic agents set trading algorithms, leading to partially rational decisions that do not always maximize their long-run profits. Behavioral economists call this non-long-run maximizing actions "cognitive bias."

The underlying ideas behind the apparently sub-optimal decision-making process are the biases and heuristics that explain how humans behave in risk situations. Making parallelism with mainstream economics, the risk aversion curvature, the cost of recent and reliable information and the available and cheaper experience change the short-run utility function, creating suboptimal decisions.

Thus, the investors' decision-making process can be understood knowing which biases and heuristics affect them. Although there is no universally accepted classification of biases and heuristics, this research will try to resume those defined by the leading experts in behavioral economics. Figure 1 shows the heuristics and how they may be reflected in investors' decisions.

Figure 1: Heuristics and biases. Own elaboration with information from Kahneman and Tversky (1984)



The BE has two prominent linkages. First, heuristics are split into representativeness and anchoring. Representativeness occurs when an investor assumes that a specific asset is a well-performing stock, and he holds it based on its past performance and availability. When the same investor hears reliable negative news about the company, he/she will sell the assets without further analysis. On the other hand, anchoring and adjustment occur when the investor bases his expectations for future earnings on the stocks past performance without reviewing the present conditions.

The second ramification referred to in BE is bias. The most common bias is the gambler's fallacy. The fallacy occurs when an investor holds onto a declining stock, believing it is due for an upturn. Another bias is herd behavior, when an investor buys a trending stock without evaluating the company's fundamentals. In this case, the heuristic behind the behavior is that "so many people cannot be wrong" and that "somebody already did the analysis." These biases are a consequence of the information's cost. The overconfidence is another kind of bias. It occurs when an investor makes risky trades or trades beyond a risk measurement because he trusts his "superior" market knowledge. The counterpart for the overconfidence bias is the risk aversion bias. It occurs when an investor avoids potentially profitable opportunities because he greatly fears losses; this fear is represented in mainstream economics as a highly concave utility function. The last biases refer to the halo effect, the confirmation bias, and the illusion of control. The first bias occurs when an investor overvalues a stock based on the company's brand name alone. The second bias is the confirmation bias. It occurs when investors only seek or incorporate information that supports their existing views on a stock. Finally, the illusion of control means an investor takes an excessive risk, believing he can control market outcomes.

Recent advancements in behavioral economics highlighted the role of sentiment analysis in understanding investor decision-making processes and market dynamics. This methodological approach extracts opinions and emotions from textual data, providing insights into the cognitive biases that influence investment decisions. Then, it will be conceptualized that "market sentiment" as the aggregate expectations of investors, which substantially impact asset pricing and market movements.

The literature identifies three primary methodologies for quantifying market sentiment:

a) Market-based measures, such as Close-end Fund Discount (CEFD) and Initial Public Offerings (IPO) activity.
b) Survey-based measures, including the Investor's Intelligence Survey and the American Association of Individual Investors (AAII) survey.
c) Text and media-based sentiment analysis.

The latter can capture short and long-term emotional trends using diverse data sources, including social media platforms, financial blogs, and news outlets. Recent empirical studies statistically demonstrated significant correlations between sentiment scores and financial market behavior, including traditional equity markets and emerging crypto assets.

Measuring sentiment in financial markets posed challenges for investors, leading them to rely on proxy indicators for their evaluations. However, with the advent of Big Data, sentiment analysis is performed by Web Scraping and NLP; the later classify and categorize market sentiments. In this sense, Dhuria (2015) mentions several approaches to NLP, including Subjective Lexicon, N-gram models, and Machine Learning, as suitable techniques to analyze text and classify it into positive, negative, or neutral sentiments.

On the other hand, researchers like Shapiro et al. (2022) employed subjective lexicons and machine-learning techniques to evaluate news sentiment. Costola et al. (2023) utilized the BERT (Bidirectional Encoder Representations from Transformers) model to generate sentiment scores during the COVID-19 pandemic. Sun et al. (2021) developed a sentiment index for the Bitcoin market based on Chinese news, and Fazlija and Harder (2022) applied pre-trained language models to predict movements in the S&P 500.

Despite its potential, NLP has shortcomings, Agarwal and Natarajan (2023) noted that sentiment analysis based on digital news faces significant limitations, primarily related to the technique and source used. Likewise, Blasco et al. (2012) emphasize the importance of adequately selecting data frequency, while Shapiro et al. (2022) mention the challenge of constructing extensive and labeled training datasets. Integrating these advanced methodologies in financial research offers promising avenues for understanding market dynamics, albeit with the caveat of acknowledging and addressing inherent limitations in data collection and analysis techniques.

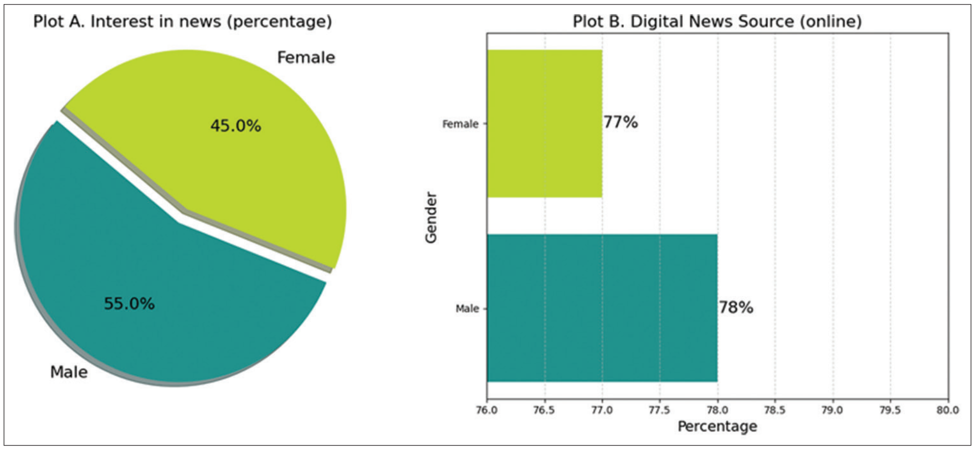
In the previous analysis has been discussed the paradigm shift from traditional economic rationality to behavioral economics, highlighting the impact of cognitive biases on investment decisions using digital news data. In the present research, machine learning and web scraping techniques will be used for collecting and classifying news information into positive, negative, or neutral categories as a methodology to comprehend market perceptions and their influence on price movements.

3. RELEVANCE OF NEWS FOR MARKET SENTIMENT FORMATION

Reuters Institute (2023) conducted extensive research to analyze news consumption patterns across various countries. This research primarily utilizes online questionnaires undertaken in collaboration with YouGov (a private online survey firm, originated in United Kingdom and owned by Black Rock). This study carefully designs and constructs representative samples of key demographic variables such as age, gender, region, education, and, when relevant, political preferences. The collected data are weighed to match census or specific industry or market standards to ensure the statistical validity and generalization of the findings.

Reuters Institute key findings provide valuable insights into news consumption. Plot A in Figure 2 shows that 48% of respondents are interested in news content, displaying a notable gender

Figure 2: News interest. Own elaboration with data from Reuters Institute (2023)



disparity 53% of male respondents indicated a high interest in news, compared to 43% of female respondents. Likewise, plot B reveals a similar high use of online platforms and websites as their primary news source. This trend is particularly pronounced among male respondents.

Figure 3 delineates specific online platforms preferred by each demographic group and illustrates the distribution of primary news sources across major social media platforms. Facebook is the predominant platform, accounting for 30.8% of news consumption. YouTube is the second most significant source, with 23.1% of users relying on it for news content. Despite being primarily a messaging application, WhatsApp ranks third, with 19.2% of users accessing news through it. Instagram and Twitter hold 15.4% and 11.5% of news consumption, respectively.

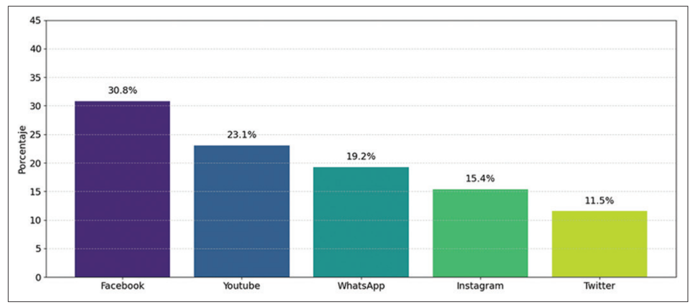
For Mexico, the Reuters Institute (2023) reports a gradual decline in the prominence of television and print media as primary news sources (Figure 4 plot B). Conversely, social media platforms gained traction across various age groups, facilitated by Internet penetration in Mexican households (Figure 4 plot B).

As news sources, Mexican consumers show confidence in social media platforms. YouTube and TikTok are the fastest-growing sources. Figure 5 shows that Facebook is the dominant platform, here 36.4% of users rely on it as their primary news source. YouTube is the second platform accounting for 24.2% of users using it as a news provider.

WhatsApp, a messaging application, ranks third, with 18.2% of users accessing news through peer-to-peer and group messages on this social media platform. Twitter and Instagram are the fourth and fifth sources, representing 12.1% and 9.1% of news consumption, respectively.

The consumers' distribution pattern reflects the evolving nature of news consumption in Mexico, with social media platforms playing a crucial role. The data indicates a preference for platforms that offer diverse content formats, from text and images to videos and instant messaging. Because of their extensive reach and importance among economic agents, news sources are valuable for analyzing

Figure 3: Main platforms as a source of news. Own elaboration with data from Reuters Institute (2023)



market sentiment. The fast spread of information through these channels and their capacity to influence perceptions and investment decisions render an invaluable understanding of market behavior and financial trends.

4. METHODOLOGY FOR MARKET SENTIMENT ANALYSIS

Next the sentiment analysis of news related to public companies to determine if cognitive biases influence the behavior of economic agents will be now analyzed. To do this, companies' historical stock prices and text news from several sources will be used to understand if the news appearance is timely coordinated with the share price movements, validating the cognitive biases' potential influence on investment decisions.

Figure 6 illustrates the model for analyzing news sentiment and historical price behavior. It describes the web scraping massive data collection (publication date, title, and content) from public digital news texts obtained in the websites' HTML codes. Once the texts are collected, the will be preprocessed with the Natural Language Toolkit (NLTK), A Python based platform capable of analyzing human language data. This process involves tokenization (separating words and eliminating irrelevant text as stopwords, lemmatization, and grammatical signs).

Finally, in order to obtain the sentiment scores for the remaining words to measure their polarity (negative, positive, or neutral

Figure 4: Internet access and main source of news in Mexico. Own elaboration with data from Reuters Institute (2023)

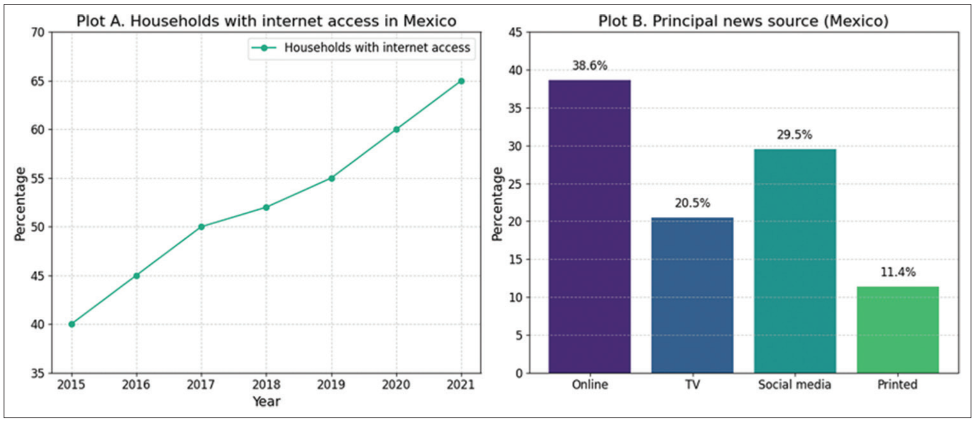
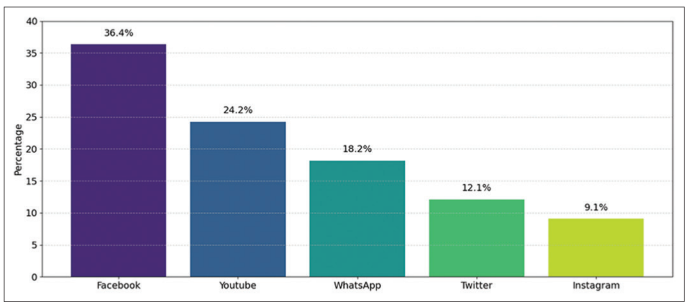


Figure 5: Main social media source of news in Mexico. Own elaboration with data from Reuters Institute (2023)



sentiments) using the "VADER" lexical dictionary. Finally the historical prices using the Yahoo finance platform will be extracted.

To carry out the proposed analysis, data from América Móvil will be used. This stock is included in the S&P IPC BMV (Mexican Stock Index) and the third most significant company (9.99% index participation). Given its relevance and liquidity in the Mexican stock market, this company has been chosen for this study. Two data sets are built, first, one of relevant news and the other of price data from January 2023 to April 2024:

A. Google News (https://news.google.com)
B. Yahoo Finance (https://finance.yahoo.com)
C. Finviz (https://finviz.com)

Information is also obtained from other platforms, such as blogs and newspaper websites. Specifically on Mexican news websites:

i. El Universal (https://www.eluniversal.com.mx/)
ii. Expansion (https://expansion.mx/)
iii. El Financiero (https://www.elfinanciero.com.mx/)
iv. Reforma (https://www.reforma.com/)
v. Forbes Mexico (https://www.forbes.com.mx/)

All the sites above are publicly accessible. Therefore, web scraping will be used only as a tool to facilitate news downloading, limiting it to collecting publicly available information without getting involved in obtaining private or restricted data.

The web scraping process involved converting unstructured data from different news websites, typically HTML, by extracting key details such as URL, title, content, and publication date. HTML sets the information for easy storage and analysis using tags to structure web pages. The tags vary across websites, leading to differences in the scraping process for information extraction from each page.

After collecting essential information from various news sites, a database was generated with 211 news links related to the "América Móvil" search, written in Spanish and English between January 2023 and April 2024. Figure 7 shows the number of news items collected by month and year.

The language needed to be standardized to count words; this ensures accurate analysis of the collected news texts. Since the primary libraries and dictionaries for sentiment analysis are in English, in this case, it is used the "=GOOGLETRANSLATE" function in Google Sheets to translate the texts from Spanish to English. All the words from the articles were collected. Figure 8 shows the most frequent one.

In Figure 9, several words stand out, such as "América" and "Móvil," understandably, since they refer to the company's name. In addition, terms such as "Slim" are mentioned, referring to the owner of América Móvil, as well as other terms related to the sector to which the company belongs, such as "Services" and "Telecommunications." These terms reveal the key aspects surrounding the company and its business context.

Figure 10 illustrates the text length for each studied date; it helps identify patterns, trends, or significant dates (too much information) within the study, such as February and March 2024, when large volume of news is collected, indicating a noteworthy event for América Móvil.

At the end of the data collection period, 211 news texts are gathered that, on average, have approximately 275 words, with a standard deviation of 241, indicating a significant variation in word count between texts. The length of the texts varied from a minimum of 5 words to a maximum of 1105 words, with 50\% of the texts having a length between 86 and 390 words. The variety in the size of the texts is considerable, ranging from short to longer ones. Table 1 provides an overview of how words are distributed within the data set collected.

Figure 6: Development of the model algorithm. Own elaboration

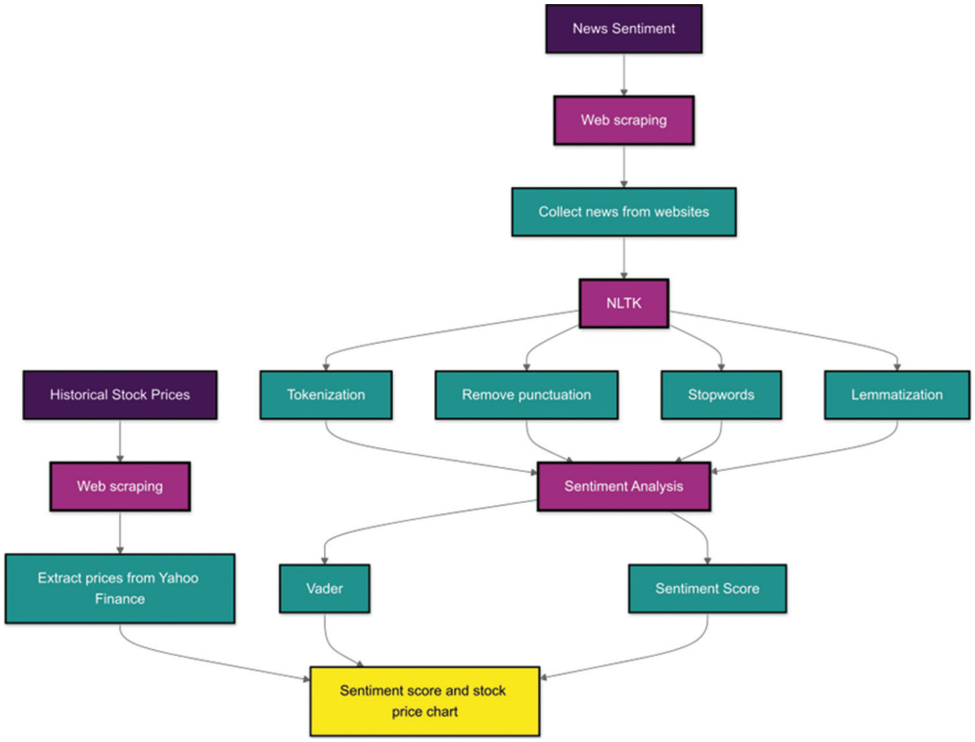


Figure 7: News collected from América Móvil. Own elaboration

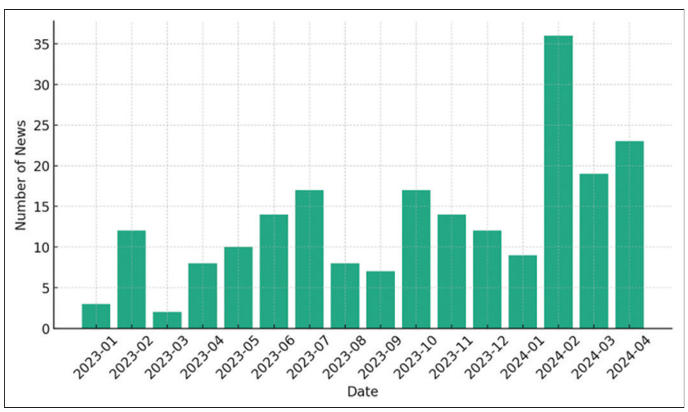


Figure 8: Most common words in the collected text. Own elaboration

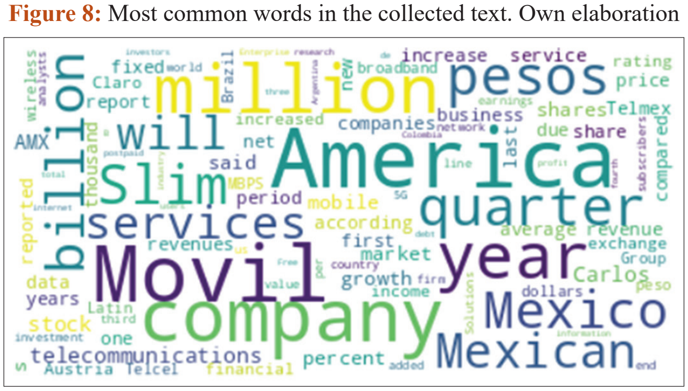
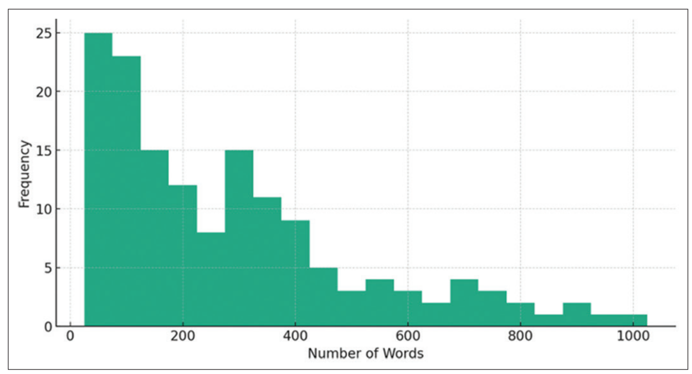


Figure 9: Frequency of words (clean). Own elaboration

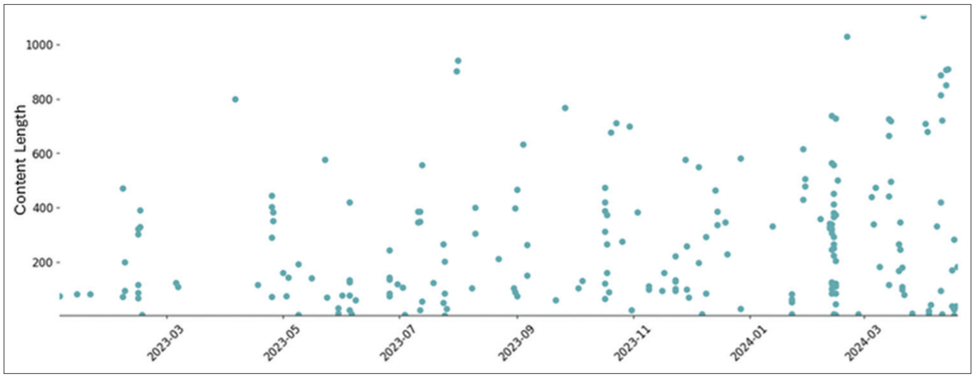


5. EMPIRICAL RESULTS

Once the database has been collected and its behavior and characteristics studied, it is necessary to preprocess the text before analyzing it. This process involves several different steps and libraries.

a. Eliminate stopwords: This is a crucial step in text preprocessing, as common words contribute little meaning to the analysis. Removing them reduces the number of general words and increases the focus on keywords, improving word classification and the accuracy of classification models. Various studies (Rakholia and Saini, 2016; Sarica et al., 2021) emphasize the importance of this step. This study used a list of 179 English stopwords from the Python library "NLTK," designed for working with human language data, and offered resources for tasks like classification, tokenization, stemming, and parsing.
b. Remove punctuation and numbers.
c. Lemmatize: This procedure involves morphological normalization that converts each word into its base form or lemma using dictionaries and morphological analysis. This method allows words to be reduced to their common root,

Figure 10: Content length by date. Own elaboration



which facilitates the comparison and analysis of the text. For example, the words "investing," "invested," and "invests" would all be normalized to their base form, "invest." This lemmatization process simplifies the text by reducing different word forms to a common root, making it easier to compare and analyze the text across countless contexts.

d. Vectorize: Once the text has been preprocessed, it needs to be converted into a format that machine learning models can understand. Vectorization transforms the text into numerical representations, often using methods like Bag of Words (BoW) or Term Frequency-Inverse Document Frequency (TF-IDF). These methods convert words or phrases into vectors, allowing algorithms to analyze the text numerically. For instance, TF-IDF assigns weight to each word based on its importance in the text, making it easier to distinguish between common and significant terms during analysis.

This analysis also employs the *CountVectorizer* function from Scikit-learn, which utilizes the BoW method. *CountVectorizer* converts a collection of text documents into an array of token counts, where each column represents a vocabulary word, and each row corresponds to a document. The values in the array indicate the number of times each word appears in the document. BoW performs the following process:

i. Tokenization: CountVectorizer splits text into individual words or tokens.
ii. Vocabulary building: A vocabulary of all unique words appearing in the corpus (collection of documents) is created.
iii. Word Counter: CountVectorizer counts the number of times each word appears in the vocabulary for each document.

The result is a sparse matrix, where each row represents a document, and each column represents a word in vocabulary. The values indicate the number of times that word appears in the corresponding document.

After being reprocessed, the text reveals significant changes in several statistical metrics. Table 2 shows the reduction in average text length, which explains the decrease in standard deviation. In addition, the maximum length decreased, suggesting the elimination of redundant or non-essential content.

Figure 11 shows a word cloud that contrasts with the original version of the text, highlighting the most relevant and significant

Figure 11: Most common words in the preprocessed text. Own elaboration

Table 1: Descriptive statistics of text compilation

Count	Mean	Standard Deviation	Max	Min
211 news	74 words	241 words	1106 words	85 words

Table 2: Descriptive statistics of clean text compilation

Count	Mean	Standard Deviation	Max	Min
211 news	156 words	137 words	722 words	4 words

Source: Own elaboration

words. It also shows a greater concentration of ideas and expressions directly associated with the company offering a more precise view of the critical issues addressed in the text.

To quantitatively analyze news sentiment and its relationship with market behavior, a mathematical model based on natural language processing techniques and time series analysis is developed. This model allows us to represent the textual content of news in a vector space and quantify its sentiment, thus facilitating comparison with América Móvil stock price movements.

Let D = $\{d_1, d2,\ldots,d_n\}$ be the set of news documents. Each document d_i is represented as a vector in the feature space $d_i = (w_{i1}, w_{i2},\ldots,w_{im})$ where w_{ij} is the weight of term j in document i, and m is the total number of unique terms in the corpus.

The vector representation of news documents is performed using the TF-IDF method. As mentioned before, TF-IDF weighs the importance of each term based on its frequency in the document and its rarity in the total corpus (Salton and Buckley, 1988). This technique enables us to capture the semantic essence of each news item numerically. The TF-IDF weight is calculated as:

$$w_{ij} = tf_{ij} \cdot \log\left(\frac{N}{df_j}\right) \tag{1}$$

where tf_{ij} is the frequency of term j in document i, N, is the total number of documents, and df_j is the number of documents containing term j.

For sentiment analysis, the VADER (Valence Aware Dictionary and Sentiment Reasoner) model is applied, which has proven particularly effective in analyzing short texts and financial content (Hutto and Gilbert, 2014). It uses qualitative and quantitative methods to produce a sentiment lexicon used in social networks (Elbagir and Yang, 2019). VADER's function receives a text and returns a dictionary of scores in four categories: negative, neutral, positive, and composite, which is calculated by normalizing negative, neutral, and positive scores. VADER assigns a compound sentiment score to each document d_i:

$$s_i = \frac{s_i^+ + s_i^- + s_i^n}{\sqrt{\left(s_i^+\right)^2 + \left(s_i^-\right)^2 + \left(s_i^n\right)^2}} \tag{2}$$

where s_i^+, s_i^-, and s_i^n are the positive, negative, and neutral scores respectively, normalized between -1 and 1; i.e., if the Compound is negative, it means that the news has a negative sentiment. For example:

A. News 1 = Compound: −0.5
B. News 2 = Compound: −0.1

Finally, to evaluate the relationship between news sentiment and market behavior, it is computed the Pearson correlation coefficient ρ between the time series of sentiment scores $S = \{s_1, s_2, \ldots, s_T\}$ and the time series of logarithmic returns of América Móvil stock $R = \{r_1, r_2, \ldots, r_T\}$:

$$\rho = \frac{\sum_{t=1}^{T}\left(s_t - \bar{s}\right)\left(r_t - \bar{r}\right)}{\sqrt{\sum_{t=1}^{T}\left(s_t - \bar{s}\right)^2}\sqrt{\sum_{t=1}^{T}\left(r_t - \bar{r}\right)^2}} \tag{3}$$

where $\bar{s}$ and $\bar{r}$ are the means of S and R respectively. In this case, both news items have negative sentiment. However, news item 1 is more negative than news item 2. The same principle applies to positive sentiment, but the Compound is greater than 0. On the other hand, if the Compound is equal to 1, then the sentiment is neutral. Figure 12 shows América Móvil Compound from January 2023 to April 2024. Most news has a neutral sentiment, with scores close to 1.

Figure 13 shows the trend of the Compound score created by Vader for América Móvil from January 2023 to April 2024. Throughout the study period, scores close to one predominate, indicating mainly neutral to positive sentiments. However, in September 2023 and March 2024, it is observed very negative sentiment scores.

6. GENERAL DISCUSSION

To compare América Móvil market sentiment with its market performance, it is gathered the historical share price of América Móvil, "AMX," from the Yahoo Finance platform. Figure 14 highlights two key aspects: the closing price of AMX, which shows a positive trend in the share price until July 2023, followed by a decline that reached its lowest point in November of the same year. Afterward, the stock exhibited signs of recovery and continued upward until April 2024. The Compound sentiment score also reveals a similar trend, with some atypical behaviors.

These behaviors are particularly notable during the share price decline in early July 2023 and the subsequent recovery in November 2023. During these periods, sentiment scores displayed increased volatility, indicating that economic agents reacted to changes in news polarity.

Figure 15 shows that a better trend visualization is obtained when establishing a threshold for the positive sentiment score, eliminating those exceeding 50%. It is important to point out that in October 2023, the sentiment was mainly positive, while the share price was not yet increasing, suggesting that it takes time to reflect market sentiment fully. Similarly, in March 2024, the share price increased, but the sentiment score was mainly negative, and the sentiment affected the price days later.

Examining the most frequently used words during periods of stock market booms and busts may help to clarify the links between cognitive biases and fluctuations in stock prices. Next, the most frequently used words during stock booms and busts periods are examined to clarify the links between cognitive biases and fluctuations in stock prices. The most common words in September 2023, during the price's significant lows, were "complaint," "failure," and "service". It is important to note that not all words directly relate to company performance; economic and political factors also play a role. Terms such as "peso," "Slim," "disagreement," and "regulation" appear Figure 16 in this analysis. Those words highlight the importance of considering diverse aspects that can influence agents' sentiments.

When focusing on the lowest scores for that month (those below 0.5), several words show a negative association, such as "complaint, disagreement," failure, Figure 17.

Figure 18 shows the most common words during March 2024. They have a positive sentiment, even when the stock's price was not growing.

A closer look at the highest scores (above 0.5) for March 2024 shows that words such as "income," "quarter," "revenue," "increase," etc., stand out Figure 19.

Figure 12: Frequency score compound América Móvil. Own elaboration with VADER

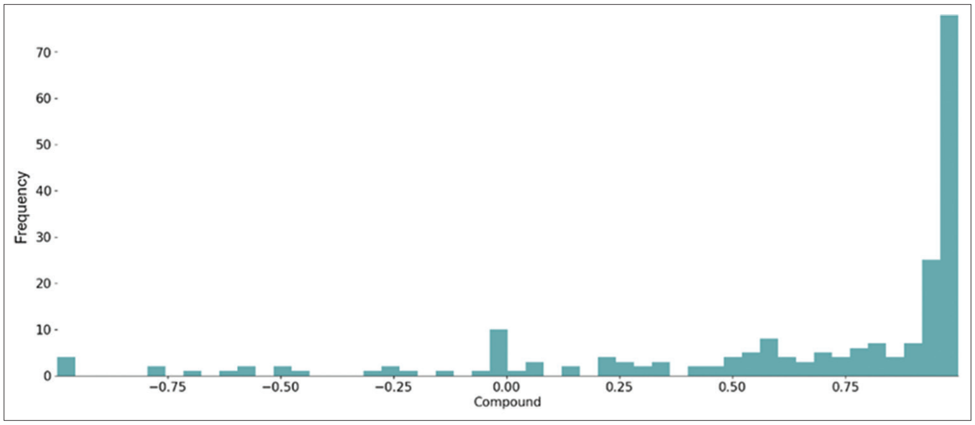


Figure 13: Score compound trend América Móvil. Own elaboration

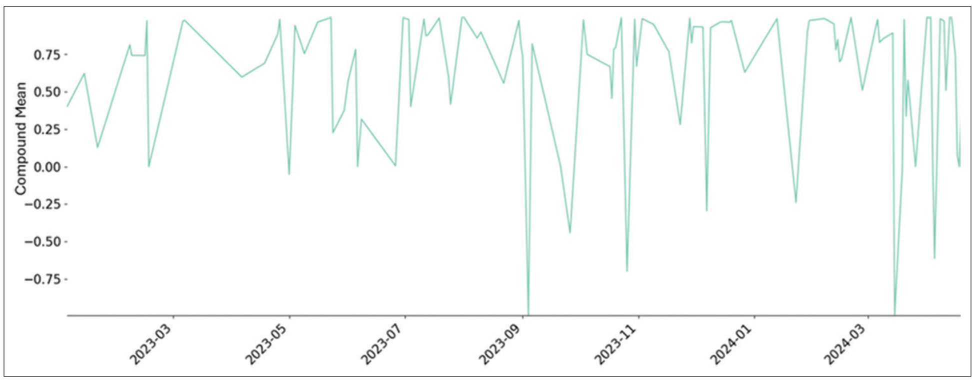


Figure 14: Sentiment score and price history América Móvil. Own elaboration with Python

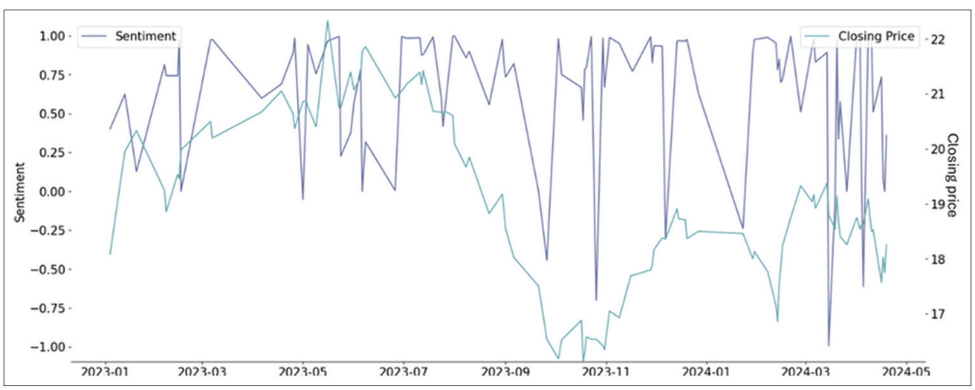


Figure 15: Sentiment score and price history América Móvil with threshold. Own elaboration with Python

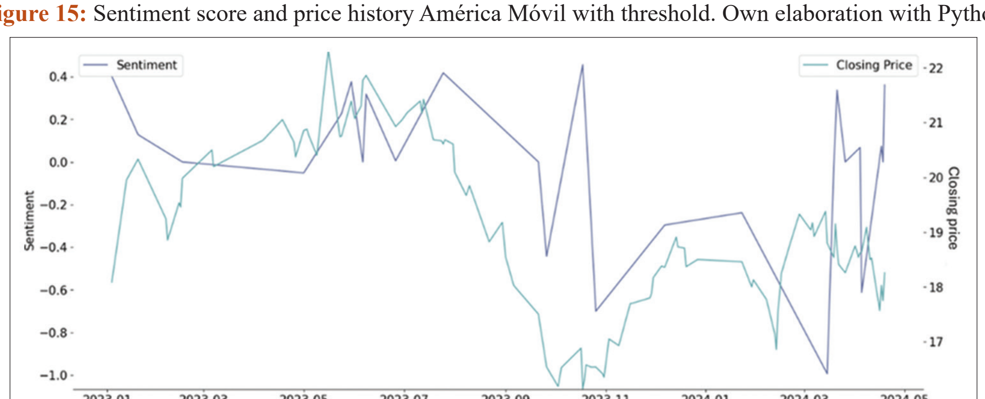
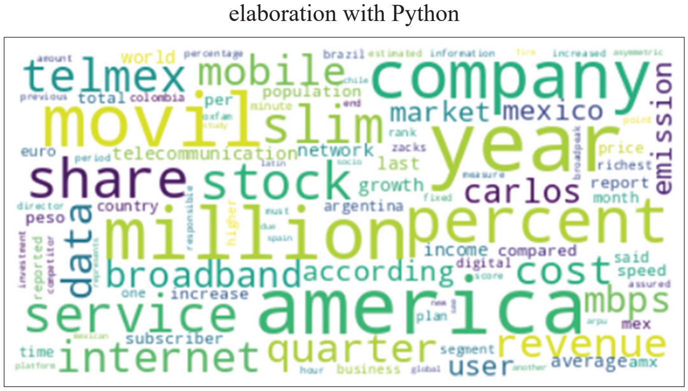


Figure 16: Most common words in September 2023. Own elaboration with Python

Figure 17: Words with score <0.5 in September 2023. Own elaboration with Python

Figure 18: Most common words in March 2024. Own elaboration with Python

Figure 19: Words with score >0.5 in March 2024. Own elaboration with Python



The proposed analysis demonstrates that negative sentiment is associated with a decline in the stock price, while positive sentiment correlates with an increase. These findings highlight a clear relationship between the cognitive biases of market agents and the stock's behavior, reinforcing the impact of sentiment on market dynamics.

7. CONCLUSION

Understanding the influence of emotions, feelings, and biases on economic agents' investment decisions helps us measure behavioral economics sentiments. This study contributes to the field by transforming abstract cognitive aspects into quantifiable data, providing a more comprehensive understanding of the impact of cognitive biases. In this research, a model to examine cognitive biases was developed to analyze the news sentiment from different websites and compare it with the América Móvil stock prices from January 2023 to April 2024.This company has a prominence in the S&P IPC BMV index.

This paper used web scraping tools to extract large volumes of text. Subsequently, it is preprocessed by removing irrelevant words, spaces, and stopwords, followed by vectorization using the BoW technique. The "sentiment scores" using the VADER library in Python are computed assigning positive, negative, or neutral sentiment values to specific words. Subsequently, historical stock prices from Yahoo Finance are collected to compare sentiment trends with stock prices' movements over the study period.

The sentiment analysis revealed that words associated with positive and negative sentiments helped to identify the events to which economic agents react. For instance, negative sentiment was linked to reports of service failures or reduced revenues, indicating that agents' reactions are not solely based on a company's performance but are also influenced by broader economic and political contexts.

The present investigation has found that stock prices sometimes exhibit a lagged response to investor sentiment, which can be interpreted as an early indicator of future stock behavior. Furthermore, the predominance of neutral sentiment in most news articles underscores the necessity of a large dataset to extract more pronounced sentiment polarities.

The growing accessibility of digital media, such as social networks, platforms, and websites, offers new opportunities to analyze their impact on economic agents' behavior. In this context, studying cognitive biases within Mexican society allows for a deeper understanding of the factors influencing investment decisions and the diverse reactions to online information.

Future research could expand the study of cognitive biases across different sectors and political issues. Enhancing the database with a more extensive collection of news and online texts would refine sentiment analysis across various dates. Employing a Spanish sentiment library or developing a specialized dictionary could improve sentiment analysis without translating. Another promising direction for research is predicting stock behavior based on sentiment scores.

REFERENCES

Agarwal, D., Natarajan, B. (2023), Tracking and handling behavioral biases in active learning frameworks. Information Sciences, 641, 119117.

Blasco, N., Corredor, P., Ferreruela, S. (2012), Market sentiment: A key factor of investors' imitative behaviour. Accounting and Finance, 52(3), 663-689.

Costola, M., Hinz, O., Nofer, M., Pelizzon, L. (2023), Machine learning sentiment analysis, covid-19 news and stock market reactions. Research in International Business and Finance, 64, 101881.

Dhuria, S. (2015), Sentiment analysis: An approach in natural language processing for data extraction. International Journal of New Innovations in Engineering and Technology, 2(4), 27-31.

Elbagir, S., Yang, J. (2019), Twitter sentiment analysis using natural language toolkit and vader sentiment. In: Proceedings of the International Multiconference of Engineers and Computer Scientists. Vol. 22. Hong Kong: IMECS. p13-15.

Fazlija, B., Harder, P. (2022), Using financial news sentiment for stock price direction prediction. Mathematics, 10(13), 2156.

Fernández, M., De Guevara, R.L., Madrid, R.M. (2017), Las finanzas conductuales en la toma de decisiones. Fides et Ratio-Revista de Difusión Cultural y Científica de la Universidad La Salle en Bolivia, 13(13), 127-144.

Hutto, C.J., Gilbert, E. (2014), VADER: A parsimonious rule-based model for sentiment analysis of social media text. In: Proceedings of the Eighth International AAAI Conference on Weblogs and Social Media. p216-225.

Kahneman, D., Tversky, A. (1984), Choices, values, and frames. American Psychologist, 39(4), 341-350.

Liu, B. (2015), Sentiment Analysis: Mining Opinions, Sentiments, and Emotions. 1st ed. Cambridge: Cambridge University Press.

Lucas, R.E. Jr. (1978), Asset prices in an exchange economy. Econometrica, 46(6), 1429-1445.

Lucas, R.E. Jr., Prescott, E.C. (1971), Investment under uncertainty. Econometrica, 39(5), 659-681.

Muth, J.F. (1961), Rational expectations and the theory of price movements. Econometrica, 29(3), 315-335.

Rakholia, R.M., Saini J.R. (2016), Lexical Classes Based Stop Words Categorization for Gujarati Language. In: 2nd International Conference on Advances in Computing, Communication, and Automation (ICACCA) (Fall), Bareilly, India. p1-5.

Reuters Institute. (2023), Digital News Report. Technical Report. United Kingdom: Reuters Institute for the Study of Journalism.

Salton, G., Buckley, C. (1988), Term-weighting approaches in automatic text retrieval. Information Processing and Management, 24(5), 513-523.

Sarica, S., Song, B., Luo J., Wood, K.L. (2021), Idea generation with technology semantic network. Artificial intelligence for engineering design. Analysis and Manufacturing, 35(3), 265-283.

Shapiro, A.H., Sudhof, M., Wilson, D.J. (2022), Measuring news sentiment. Journal of Econometrics, 228(2), 221-243.

Shiller, R.J. (1984), Stock prices and social dynamics. Brookings Papers on Economic Activity, 1984(2), 457-510.

Shleifer, A., Summers, L.H. (1990), The noise trader approach to finance. Journal of Economic Perspectives, 4(2), 19-33.

Sun, Y., Kong, X., Chen, T., Su, H., Zeng, X., Shen, Y. (2021), Measuring investor sentiment of cryptocurrency market-using textual analytics on chain node. Procedia Computer Science, 187, 542-548.

Zhao, J., Liu, K., Xu, L. (2016), Sentiment analysis: Mining opinions, sentiments, and emotions. Computational Linguistics, 42(3), 595-598.