



MASTER'S DEGREE IN DATA SCIENCE AND BUSINESS
ANALYTICS

SENTIMENT-DRIVEN AUTOMATED TRADING: A COMPARISON WITH TRADITIONAL METHODS

Author: Jordi Agut Sáiz

Supervised by: Juan Manuel Moreno Lamparero

- Tarragona August 2025 -

ABSTRACT

In recent years, the rapid evolution of algorithmic trading has been fuelled by the integration of advanced data science techniques, particularly Natural Language Processing (NLP). This project explores the development of an intelligent algorithmic trading system that uses real-time sentiment analysis from social media and news sources. The primary objective is to enhance trading decision-making by capturing market sentiment, a critical factor influencing retail-driven assets such as cryptocurrencies. By employing NLP techniques, including sentiment classification, this system can analyse the emotional tone of text data and quantify its impact on market prices.

This thesis evaluates the performance of a Natural Language Processing (NLP) sentiment analysis strategy compared with a naïve BUY & HOLD approach across both stocks and cryptocurrencies. The results indicate that while large-cap equities show limited added value from NLP due to market efficiency and institutional dominance, the strategy is more effective in volatile and less efficient markets. In particular, cryptocurrencies and small-cap stocks benefited from more active trading signals, often outperforming passive investing. Nevertheless, performance was not uniformly consistent, as trade timing and model robustness played a critical role. Overall, the study highlights the potential of NLP as a complementary tool for investment decision-making, while emphasizing the importance of further refinement and the integration of hybrid approaches.

Keywords: Sentiment analysis, Natural language processing, Algorithmic trading, Financial irrationality, API, Cryptocurrencies.

INDEX

1.	INTRODUCTION	1
2.	OBJECTIVES OF THE RESEARCH	3
3.	TRADITIONAL MARKET-EFFICIENCY PARADIGMS.....	4
3.1	EFFICIENT MARKET HYPOTHESIS	4
3.2	IRRATIONAL MARKET CRASHES.....	4
4	BEHAVIOURAL FINANCE.....	7
4.1	EVIDENCE ON FINANCIAL IRRATIONALITY	8
4.2	ROLE OF SOCIAL MEDIA	9
4.2.1	CONFIRMATION BIAS ON SOCIAL PLATFORMS	10
4.2.2	HERDING BEHAVIOR IN ONLINE INVESTOR COMMUNITIES.....	11
5.	DATA AND METHODOLOGY	12
5.1	DATA SOURCES	13
5.1.1	APIs.....	13
5.2	SENTIMENT ANALYSIS PIPELINE	15
5.2.1	BERT	17
5.2.2	FINE-TUNED MODELS.....	18
5.3	DECISION MAKING LOGIC	21
5.4	TRADING SIGNAL.....	23
6.	RESULTS.....	26
6.1	EVALUATION METHODOLOGY.....	26
6.2	STOCKS.....	27
6.3	CRYPTOCURRENCIES.....	29
7.	LIMITATIONS	32
8.	CONCLUSIONS	33
9.	REFERENCES.....	34

10. ANNEX	36
10.2 FIGURES	36
10.3 TABLES	40
10.4 CODE	40

1. INTRODUCTION

Financial markets have long been considered rational environments where asset prices are determined by fundamental factors such as earnings, interest rates, and economic indicators. Behavioural finance, which emerged in the late 1970s, has transformed this paradigm by demonstrating that investors' emotions, cognitive biases, and collective psychology significantly influence markets.

Researchers such as Robert J. Shiller and Richard L. Peterson have demonstrated that individuals frequently behave in manners that cause market prices to diverge from their intrinsic or “rational” values.

Daniel Kahneman and Amos Tversky developed Prospect Theory to elucidate the often irrational thought processes of humans regarding gains and losses. One of the fields where this theory could be cross-checked empirically is in finance and investing, where investors deal every day with these potential situations where they must confront benefits or losses depending on the action chosen. The Disposition Effect, Herd Mentality, and Overconfidence Bias are significant theories illustrating the influence of psychological factors on the market. These insights are particularly significant in the contemporary landscape, where social media and online news serve as primary information sources and can rapidly alter public sentiment regarding the market.

The emerging research domain and significant transformations in algorithmic trading—facilitated by advanced data science techniques and novel data sources—present numerous opportunities for investigation and alter investor behaviour. If these cognitive biases are reflected in the market, new quantitative approaches provide traders the capability of interpreting and capitalising on these shifts in sentiment and an opportunity to generate profit.

News articles, social media, and various online platforms have emerged as significant sources of sentiment data influencing the stock market. Cryptocurrencies and other retail-driven assets demonstrate the influence of public sentiment on market dynamics. Significant occurrences, such as policy declarations from the Trump administration or Twitter's influence on cryptocurrency valuations, have demonstrated the rapidity with which market sentiment can shift, resulting in abrupt price fluctuations.

This project integrates the two worlds of finance and data science with their subsequent ramifications, natural language processing (more specifically sentiment analysis), and the use of APIs or automation techniques to exploit this new paradigm in trading.

The primary objective is to develop an advanced algorithmic trading system that will be based purely on real-time market sentiment analysis derived from news and social media. The system will employ NLP techniques such as entity recognition and sentiment analysis to identify and evaluate the emotional tone of textual data. This will enable an immediate response to fluctuations in public sentiment regarding the market.

2. OBJECTIVES OF THE RESEARCH

The main goal of this project is to assess the impact and exploit the behavioural-finance phenomena, such as confirmation bias and herding behaviour, and measure how they affect trading outcomes in a variety of asset classes, such as cryptocurrencies, and stocks.

A second goal is to create and put into place a fully automated data-processing pipeline that takes in real-time social media signals (like sentiment and tweet volume metrics) and uses the predictions to execute buy and sell orders in markets. As a tool for people interested in this subject who don't know much about coding, the finished code is meant to be as straightforward and easy to use as possible, so the user could have the ability to adjust some parameters of interest in a superficial way, only writing or deleting some variables without generating new scripts.

Additionally, one of the programming objectives is to enable querying different APIs using their respective credentials and tokens, ensuring that the system can be replicated for various users without issues.

With respect to the models of sentiment analysis, increase as much as possible the accuracy of the predictions for this specific context, financial tweets: short text, probably containing special characters and noise, such as emojis, with information concentrated in few words; tokenising and fine-tuning the model for these inputs will be one of the biggest steps because all the rest of the pipeline will depend on that.

Finally, measure the added value of integrating behavioural finance insights into algorithmic trading. The project will compare this social media-enhanced trading system to traditional fundamental and quantitative strategies, assessing improvements in performance metrics (and % return strategies implemented).

3. TRADITIONAL MARKET-EFFICIENCY PARADIGMS

All the dominant theories in finance historically have operated under the premise that investors are completely rational beings who maximise a well-behaved utility function. From the first expected utility theories, Bernoulli (1738) or Von Neumann & Morgenstern (1944), ideas imply that, when faced with risky prospects, investors always choose the “rationally” optimal lottery: the one delivering the greatest weighted average of utilities.

3.1 EFFICIENT MARKET HYPOTHESIS

Thirty years later, Fama (1970) developed the Efficient Markets Hypothesis (EMH). This theory maintains that if investors are rational and process information without bias, then asset prices fully reflect all available information. There are three types of EMH:

- **Weak form:** Technical analysis cannot consistently outperform the market because prices reflect all historical trading data.
- **Semi-strong form:** Prices take into account all information that is available to the public, so neither public news nor fundamental analysis can produce consistent α .
- **Strong form:** All information, both public and private, is reflected in prices, meaning that even insiders cannot consistently outperform.

Then how to explain the anomalies? Under EMH, any observed market “anomalies” must be random noise or quickly arbitrated away, since rational investors would exploit and thus eliminate them.

Stock markets quickly reveal that investors were not as logical as previously thought. Shiller (1981) challenged this hypothesis by showing that stock prices fluctuate far more than can be justified by subsequent dividends. Moreover, the empirical outcomes in the indices demonstrate that the EMH was not holding.

3.2 IRRATIONAL MARKET CRASHES

Oil Crisis & Bear market 1973-74

- The OPEC oil embargo in October 1973 resulted in a surge in crude prices from approximately \$3 to over \$12 per barrel within six months, initiating a bear market that persisted from January 1973 to December 1974 (see Figure 1). The recession and rising inflation were intensified by the quadrupling of energy prices, causing significant investor fear. The Dow Jones Industrial Average

declined over 45% during the 694 trading days of the downturn, rendering it one of the most severe bear markets since the Great Depression.

- Investors possessed minimal fundamental information in this stagflationary context characterised by concurrent high inflation and recession. Herding behaviour, or the propensity to follow the herd rather than one's own judgement, became more noticeable as markets liquidated. Fearful of being "the last one out," both retail investors and portfolio managers collectively liquidated their positions or redeemed funds.

The collective selling intensified price declines, creating a feedback loop in which falling prices incited additional selling, further exacerbating the reduction in prices.

Black Monday Crash 1987

- The U.S. Dow Jones Industrial Average experienced the biggest single-day percentage decline in history on October 19, 1987, when it abruptly fell 22.6%, shocking global equity markets (see Figure 2). Importantly, the collapse was driven by psychological contagion, as no significant new fundamental news—economic data, earnings reports, or geopolitical events—could account for it.

Widespread alarm was stoked by media coverage of the crash. Investors gave up on any presence of fundamental analysis after observing peers and institutions hurrying to sell. This widespread panic demonstrated how emotion can override reasoned judgement under extreme stress by pushing volume to all-time highs and prices even lower.

Dot-Com Bubble 2000

- Convinced that the internet would disrupt conventional business models, investors poured enormous sums of money into any company with a ".com" suffix during the latter half of the 1990s. Stories of network effects and first-mover advantages drove this euphoria, while cash flows, earnings, and revenues were mostly disregarded. The Nasdaq Composite increased by almost 580% between January 1995 and March 2000, from about 750 to over 5,100.

Sentiment suddenly changed when the promised profits did not materialize. The Nasdaq reached a peak of roughly 5,132 points on March 10, 2000, but by

October 2002, it had fallen to about 1,140, a 78% decline that erased almost all the gains made during the bubble (see Annex, Figure 14). Fear spread because of media reports of plummeting stock prices and well-publicized bankruptcies (such as Pets.com and Webvan). Observing their peers leave in large numbers, investors gave up on long-term analysis and joined the rush out of "overvalued" names, according to Number Analytics.



Figure 1: DOW (Adj. for inflation) 1972-1986. Source: <https://www.macrotrends.net>

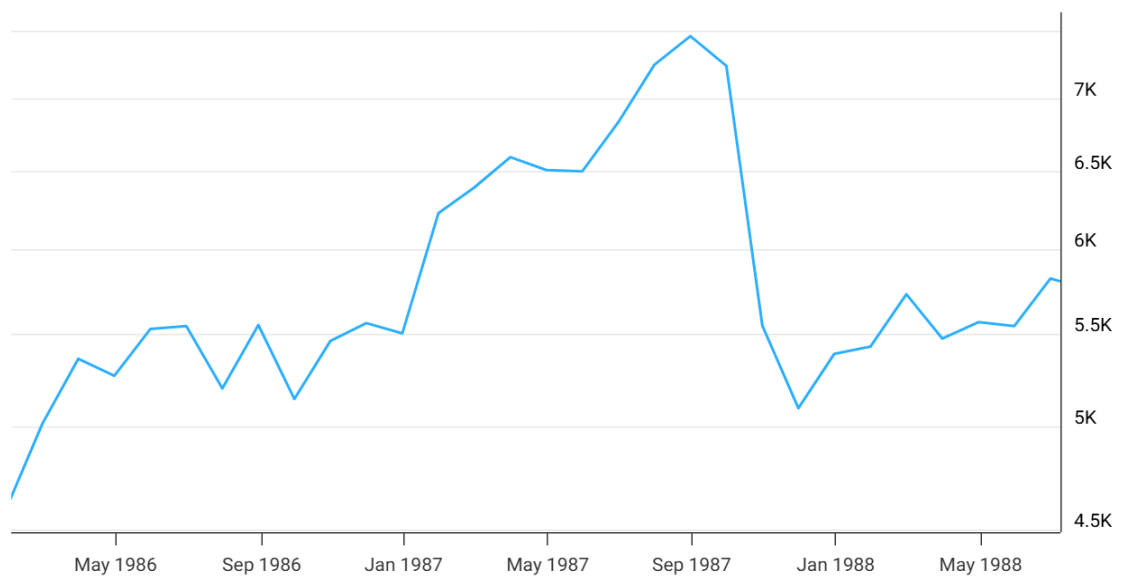


Figure 2: DOW (Adj. for inflation) Feb 1986- Jun 1988. Source: <https://www.macrotrends.net>

4 BEHAVIOURAL FINANCE

Following Investopedia definition, behavioural finance, a subfield of behavioural economics, is a theory proposes that psychological influences and biases affect the financial behaviours of investors and financial practitioners. Moreover, influences and biases can explain all types of market anomalies, including those in the stock market, such as severe rises or falls in stock price.

Behavioural finance posits that financial participants are psychologically influenced and exhibit moderate self-control rather than being entirely rational and self-disciplined. An investor's psychological condition often fluctuates in accordance with their overall health status. This influences their decision-making and the logical approach they adopt towards all real-world problems, including financial matters.

The Efficient Market Hypothesis theory was based on the premise of rationality. This belief of rationality was somewhere a key shortcoming of EMH, resulting into market anomalies. Behavioural Finance aims to investigate these anomalies or irregularities by clarifying what, why and how of investment from the perspective of a human being.

Information and attributes of market agents have a great influence on investors' investment decisions and market outcomes. As a result, when making investment decisions, investors act irrationally making suboptimal decisions.

From this theory arises important concepts in finance, such as the Market Psychology, the study of collective sentiment and behaviour of market participants that drives price movements, often influenced by emotions such as fear and greed, not just fundamentals. Fundamentals drive stock performance, but market psychology can override the fundamentals, pushing a stock's price in an unexpected direction (Investopedia).

CBOE Volatility Index (VIX), measures expected market volatility using portfolio options on the S&P 500. By quantifying how fast price changes, we can estimate the particular degree of fear among market participants on the market. Historically, values greater than 30 are generally linked to large volatility periods, resulting from high uncertainty, risk and investors' fear. Below 20, generally corresponds to stable periods. From the examples in Chapter 3.2, the correlation is clear (see Figure 3).

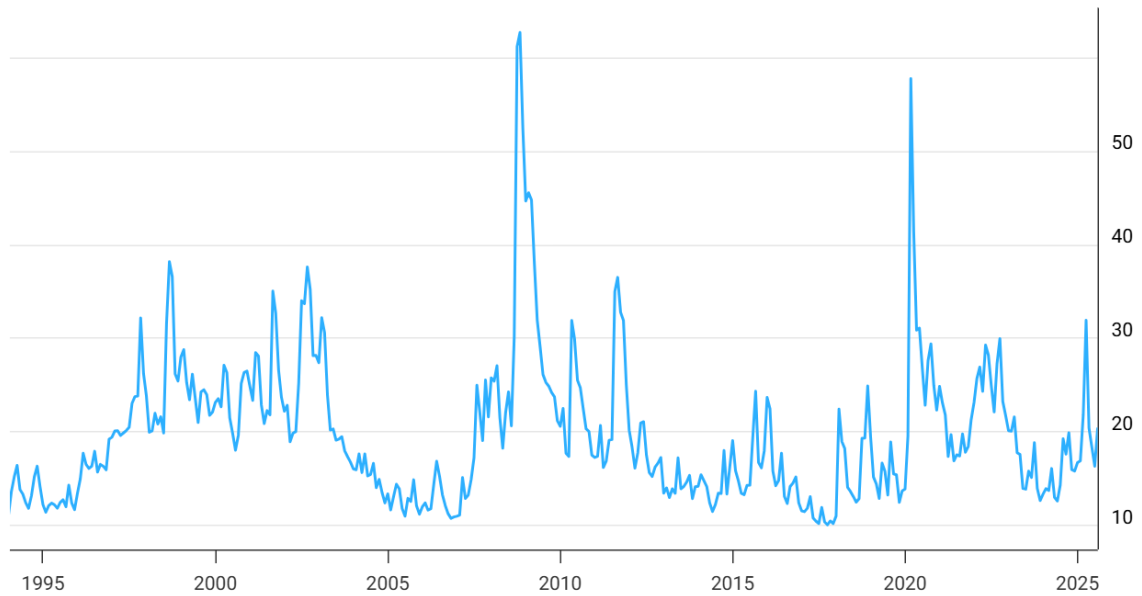


Figure 3: VIX Index 1995-2025. Source: <https://www.macrotrends.net>

On the basis of these psychological underpinnings, a number of irrational behaviours that affect investor decision-making have been discovered. These include overconfidence, which causes investors to overestimate their knowledge or aptitude for predicting market movements, and herd behaviour, in which people follow the herd rather than doing their own independent research. Investors frequently hang onto losing positions for too long due to loss aversion, which is the propensity to fear losses more than to value comparable gains. Other prevalent biases include anchoring, which occurs when decisions are unduly influenced by arbitrary reference points like past stock prices, and confirmation bias, which occurs when people look for information that supports their preconceived notions. Additionally, investors may be prone to familiarity bias, which favours well-known assets over potentially superior unknown ones, and recency bias, which places an excessive amount of weight on recent occurrences. When taken separately or in combination, these biases contribute to the explanation of irrational investment patterns and ongoing market inefficiencies.

4.1 EVIDENCE ON FINANCIAL IRRATIONALITY

Several studies analysed and quantified the effect of this irrationalities, Understanding the role of social media sentiment in identifying irrational herding behaviour in the stock market (Li, Chen, et al. 2023), examined the role of social media sentiment in identifying irrational herding behaviour in the stock market. Selecting the Chinese stock market, where majority

are retail investors, gathered the sentiment of 227,353 microblog text messages through deep learning techniques and constructed the identification method of herd behaviour.

Results show that social media sentiment has a significant impact on irrational herding behaviour in the stock market. The findings of the study complement the theory of investor actions and can aid in investors' trading decisions and financial regulators' policy recommendations. Moreover, determined the impact of the social media disaggregating it in official or unofficial accounts. The first ones have differential impact on herding behaviour by being more neutral and less likely to influence these irrational behaviours, while unofficial accounts reduced the limited rationality of investors. The more follower's unofficial social media accounts have, the more likely they are to influence irrational herding behaviour.

However, this "follower effect" has a certain threshold value, and when the threshold value is exceeded, despite gaining more followers, unofficial social media accounts' posting sentiment no longer influence irrational herding behaviour.

4.2 ROLE OF SOCIAL MEDIA

Social media platforms have fundamentally reshaped retail investing by speeding the spread of market commentary, amplifying peer influence, and creating powerful feedback loops between sentiment and price. On X (formerly Twitter), each user's "For You" feed is driven by an algorithm that prioritizes content similar to what they've liked, retweeted, or spent time viewing. An investor who clicks on bullish analyses will see more of the same voices, while contrarian or critical perspectives remain hidden unless they go viral. This personalization deepens confirmation bias: optimistic narratives are rewarded with engagement and further amplification, while dissenting views struggle for visibility. During the 2024 Bitcoin halving cycle, for example, tweets hyping "halving hype" consistently dominated high-engagement feeds, whereas detailed critiques of miner economics and network security received scant attention.

Reddit's topic-focused subreddits operate on a simple but potent upvote/downvote system that transforms popular sentiment into top-ranked posts. In communities like r/WallStreetBets, posts that reinforce the prevailing narrative accumulate upvotes, drawing more eyeballs and more endorsements in a self-reinforcing spiral. Meanwhile, comments that question or temper the consensus are quickly downvoted into obscurity. Anonymity on the platform further reduces the social risk of joining the majority view, fostering herd behavior over independent analysis. This dynamic was clearly illustrated during the January 2021 GameStop squeeze: a handful of high-karma posts alleging that hedge funds had "shorted

GameStop to oblivion” spurred a wave of buying that drove the stock up over 1,600 percent in just a few weeks despite no underlying improvement in the company’s fundamentals.

More recently, TikTok has emerged as a venue for bite-sized trading tips and viral “hot stock” recommendations. Thirty- to sixty-second videos combine trending audio, filter effects, and enthusiastic delivery to turn complex financial topics into entertainment. Influencers with hundreds of thousands of followers can spark double-digit price moves in low-liquidity or small-cap stocks overnight, often without any new fundamental catalyst. During Q2 2024, multiple microcap “meme stocks” rallied sharply following TikTok endorsements, then reversed just as quickly once the social momentum subsided. Unlike longer-form Twitter threads or Reddit discussions—where users can post attachments and debate in threaded comments—TikTok’s interface offers little room for immediate pushback or detailed scrutiny, further accelerating momentum trading.

These social-media-driven phenomena have not gone unnoticed by academics. Antweiler and Frank (2004) found that message-board volume and tone added statistically significant predictive power to next-day stock returns, while Bollen et al. (2011) showed that aggregated Twitter mood indicators correlate closely with major-index fluctuations. More recently, Greyling and Rossouw (2025) combined high-frequency trade data with minute-by-minute sentiment analysis of X, reporting that this hybrid model improved short-term volatility forecasts by up to 12 percent over traditional econometric approaches. Yet, most studies treat “the market” as a single entity, without differentiating among asset classes or trading venues.

The present research addresses that gap by examining how algorithmic curation on X produces heterogeneous effects across equities, cryptocurrencies, and derivatives.

In sum, social media influences retail investors through personalized content curation, community-driven voting mechanisms, and entertaining short-form formats that reward speed and virality over depth of analysis. These platforms create filter bubbles and groupthink, leading to momentum-driven trading spikes that often diverge sharply from fundamental valuations. By dissecting these dynamics across multiple asset classes, this thesis will illuminate the nuanced ways that digital social networks shape modern financial markets.

4.2.1 CONFIRMATION BIAS ON SOCIAL PLATFORMS

Confirmation bias refers to the tendency of individuals to seek out, interpret, and remember information in a way that confirms their pre-existing beliefs. On social media, this bias is magnified by algorithmic personalization: platforms curate content based on users’ past

interactions, so investors who engage with bullish analyses receive ever more optimistic posts, while sceptical or contrarian viewpoints are systematically under-exposed. This creates self-reinforcing “echo chambers” in which investors receive a highly skewed view of market conditions.

Empirical studies confirm that these dynamics affect financial decision-making. Pelster and Romero Gonzalez (2016) analyse a large social-trading platform and find that retail investors disproportionately overweight information consistent with their existing positions, leading to poorer portfolio performance and reduced diversification. Likewise, Ostendorff et al. (2023) demonstrate that confirmation bias in digital media not only promotes group polarization but also exacerbates trading errors when investors rely on socially-filtered news feeds

4.2.2 HERDING BEHAVIOR IN ONLINE INVESTOR COMMUNITIES

Herding behaviour occurs when individuals mimic the actions of a larger group, often under the assumption that the crowd’s behaviour reflects superior information. In online forums and social platforms, rapid dissemination of sentiment—measured via tweet volume or upvote counts—can trigger cascades of synchronized trading, as users interpret widespread activity as a signal to follow suit

Frontiers in Physics (2022) employed machine-learning text analysis of social-media posts to show that abnormal retail posting activity is significantly associated with subsequent clustering of buy and sell orders, evidencing genuine herding rather than isolated noise. Complementing this, Aut’s SSRN paper (2023) finds that spikes in positive social-media sentiment led to measurable increases in herding metrics—such as return autocorrelation and variance ratios—indicating that online chatter directly undermines market informational efficiency.

With all the past evidence, the study aims to take advantage of it, and if assumptions hold have a profitable strategy by exploding the bias of the financial markets.

5. DATA AND METHODOLOGY

The developing of this project has been conceived in empirical-practical approach taking advantage of the real time processing queries of the APIs used to provide to those interested a useful tool to use for their purposes.

The conceptual idea of the project is represented in Figure 4, where user is querying the twitter accounts and their respective tweets, then this input is ingested in the desired model (user can select the pretrained one it wants from Hugging Face or train one itself with the scripts from the code). Once the model analysed the sentiment of the text and search for the company name, the criteria of investment (sell, buy or no action) is activated and then transmitted to the broker via API. This project could be replicated by installing the public repository: <https://github.com/jordiagut10/financial-nlp>

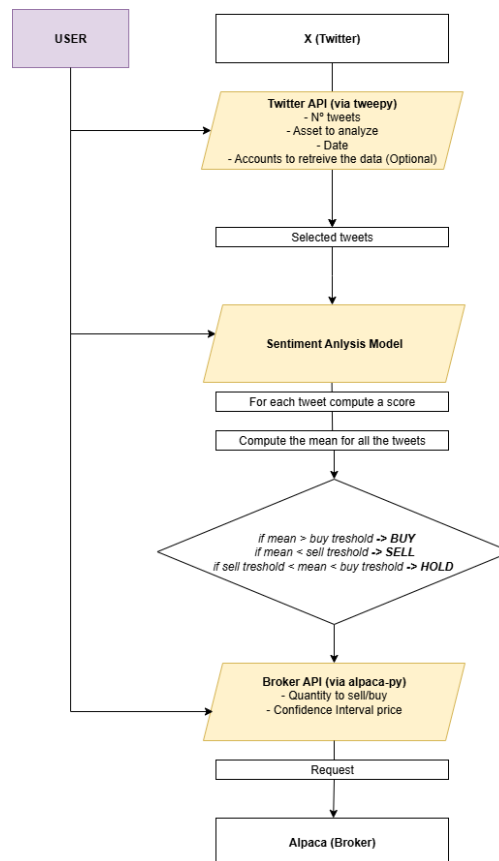


Figure 4. Conceptual framework of the pipeline.
Source: Own elaboration

- **Data Ingestion:** Tweets are streamed via the Twitter API, filtered by target company handles and keywords to capture relevant market commentary

- **Pre-Processing:** Raw tweet text is normalized—converted to lowercase, URLs and special characters removed, tokenized to the specific NLP model selected.
- **Sentiment Analysis:** Each tokenized tweet is scored by a domain-adapted NLP model, outputting negative, neutral, or positive sentiment labels with confidence scores.
- **Decision Logic:** Sentiment labels are aggregated and then mapped to investment actions: “negative” triggers a sell signal, “positive” triggers a buy signal, and “neutral” results in no action. Confidence thresholds (parameter user can modify) are applied to suppress low-certainty signals.
- **Trade Execution:** Final buy/sell orders are sent to the broker’s REST API with built-in safeguards on order size, rate limits, and error handling to ensure reliable execution.

This sequence guarantees end-to-end processing latency under one second, enabling automated, sentiment-driven trading decisions in live market conditions.

5.1 DATA SOURCES

The research was conducted using financial data from the social network X (Twitter). The analysis will focus on prominent stocks such as Apple (AAPL), and Amazon (AMZN), as well as smaller low-capitalization stocks like Aurora Cannabis (ACB) and Digi (DGII). Respect cryptocurrencies similar methodology has been used, including famous assets such as Bitcoin (BTC/USD) and Ethereum (ETH/USD), alongside mid-tier coins like Polkadot (DOT/USD) and ChainLink (LINK/USD).

In order to maximize the potential irrationalities influential personalities of the financial environment had been selected: ZeroHedge (1.6M followers), MichelleLaRosa (150k followers), JPMorganAM (200K followers), VitalikButerin (3M followers), TheBlock (300k followers) and Cointelegraph (2M followers). In case that these influencers do not have any tweet for the specific assets, last tweets from the market are queried.

5.1.1 APIs

The first and last step of the pipelines are produced via APIs (Application Programming Interface). An API is a set of rules that allows different software systems to communicate with each other. It defines how requests for data or services should be made, and how responses are delivered. APIs are commonly used to enable apps to access external services, in the scope of this study, used to gather information from an external provider and to interact

with “actions” in the broker. In top of these two APIs, code is developed to make easier the implementation of the overall pipeline.

5.1.1.1 X API

Twitter has is own and free (with some limitations) API, with this any user it is capable to interact with the application and read, post or search for specific tweets, users or trends. This last part is what it had been used in the analysis. Via tweepy, a Python wrapper for this API where simplifies the process of using it turning complex API request into easy-to use Python methods.

5.1.1.2 ALPACA API

Alpaca is a brokerage platform that exposes a powerful, REST-and-WebSocket-based API designed for algorithmic trading. For this study, the paper account (play money) had been used.

Historical bar data, real-time trade and quote feeds, account information, and order management all live under one API roof for maximum developer efficiency. With their own Python library alpaca-py, implementation and interaction with the broker is scalable and maintainable.

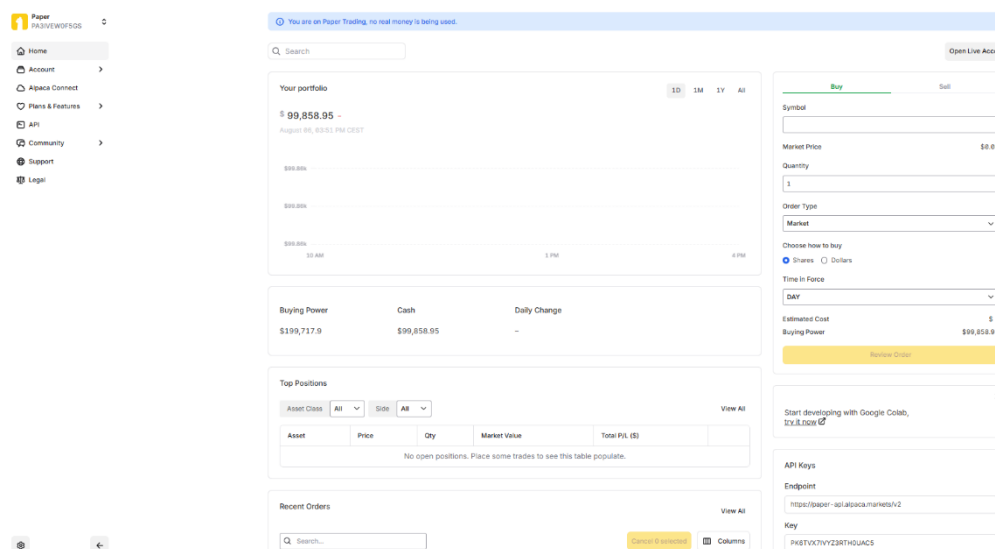


Figure 5: Alpaca main page. Source: <https://app.alpaca.markets/dashboard/overview>

```

from alpaca.trading.client import TradingClient

trading_client = TradingClient(BROKER_TOKEN, BROKER_SECRET)
print("Account ID: ", trading_client.get_account().account_number, "\nPortfolio value: ", trading_client.get_account().portfolio_value, "$")

✓ 0.2s
Account ID: PA31VEW0F56S
Portfolio value: 99858.95 $

```

Figure 6: Tweepy API code snippet. Source: Own elaboration

5.2 SENTIMENT ANALYSIS PIPELINE

Natural language processing is a field that combines computer science, artificial intelligence and language studies. It helps computers to understand process and create human language. It includes two main types of analysis: syntactical analysis and semantical analysis. Syntactical and semantical analysis are its two primary categories. By parsing the word syntax and using preprogrammed grammatical rules, syntactical analysis determines the meaning of a word, phrase, or sentence. Semantical analysis interprets the meaning of the words within the sentence structure by using the syntactic output.

Sentiment analysis is NLP technique used to classify the text based on the mood or mentality expressed in the text, which can be positive negative, or neutral.

Conventional computational methods employed by NLP to comprehend and replicate human language are as follows:

Text preprocessing: Prepares raw text into a format that machines can more easily understand. Here is we find the important concept of tokenization (see Annex 8.2, Figure 15): which involves splitting the text into smaller units like words sentences or phrases. Other transformations such as converting all the characters to lowercase, the stop words removal or stemming (reduce words to their root form. Eg: Running -> Run) are frequently used. After this preprocessing text is standardized and able to ingest to ML models.

Feature Extraction: Process of converting text into numbers that machines can analyze and interpret. Different methods are used like, Bag of Words and Term Frequency – Inverse Document Frequency (TF-IDF) which quantify the presence and importance of words in a document. More advances techniques (Word2Vec or GloVe) include word embedding, which represent words as a vectors capturing semantic relationships between words.

$$\mathbf{TF-IDF}(\mathbf{w}, \mathbf{d}) = \mathbf{TF}(\mathbf{w}, \mathbf{d}) \times \log\left(\frac{N}{\mathbf{DF}(\mathbf{w})}\right)$$

Equation 1: Term Frequency – Inverse Document Frequency

TF = Term frequency in the document

DF = Number of documents containing the word

N = Total number of documents

Document doesn't have to be a long article or file. A document can be phrase, sentence, paragraph, or text, depending on the level of granularity.

Text analysis: Step where model extract and interpret meaningful information from text data. This process includes tasks such as part-of speech (POS) tagging which identifies grammatical roles of words and named entity recognition (NER), which detects specific entities like names, locations and dates. Dependency parsing analyses grammatical relationships between words to understand sentence structure, while sentiment analysis determines the emotional tone of the text, assessing whether it is positive, negative or neutral. Topic modelling identifies underlying themes or topics within a text or across a corpus of documents. Natural language understanding (NLU) is a subset of NLP that focuses on analysing the meaning behind sentences. NLU enables software to find similar meanings in different sentences or to process words that have different meanings. Through these techniques, NLP text analysis transforms unstructured text into insights.

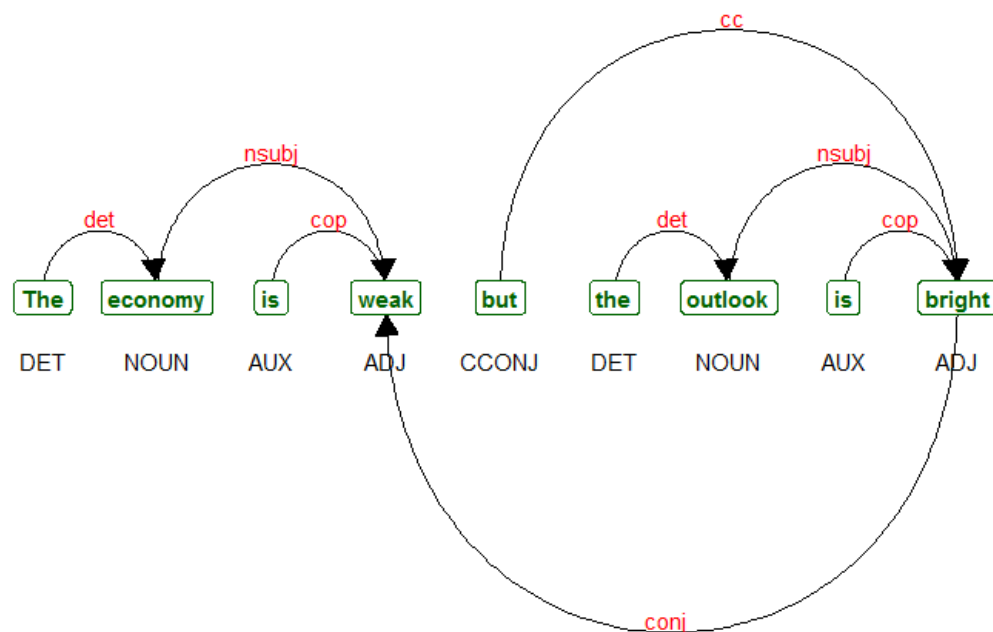


Figure 7: Illustration of dependency parsing. Source: <https://www.r-bloggers.com/2019/07/dependency-parsing-with-udpipe/>

Processed data is utilized to train machine learning models that identify patterns and relationships within the data. Throughout training, the model modifies its parameters to reduce errors and enhance its performance. After training, the model can be applied to new, unseen data to produce outputs or make predictions. The efficacy of NLP modelling is

improved through assessment, validation, and optimization to augment precision and pertinence in practical applications.

There are different types of NLP models:

- *Sequence-to-Sequence*: Based on RNN, used for machine translation
- *Autoregressive models*: Trained specifically to predict the next word in a sequence (GPT, Llama)
- *Transformer models*: They use tokenization of language and self-attention (captures dependences and relationships) to compute the relation of different language parts to another. BERT is one of the main models within this category and the one selected as a foundation for the research.

5.2.1 BERT

Bidirectional Encoder Representations from Transformers is a technique for NLP developed by Google AI Language at end of 2018. BERT was the language model that made the whole concept of pre-training and fine-tuning flow very popular. It brought two core innovations to language modelling: borrowed the transformer architecture from machine translation, which does a better job of modelling long-term dependencies than RNN-based ones. And it introduced the Masked Language Modelling (MLM) task, where a random 15% of all tokens are masked and the model predicts them, enabling true bi-directionality. At difference with other models, what BERT is doing is trying to predict onwards and backwards a hidden word, improving the quality of the output.

BERT's ability to fine-tune the model to a particular domain is one of its key advantages. This means that after being pretrained on large, general-purpose text databases, BERT can be trained on a smaller, domain-specific dataset to adapt to specialised tasks like financial sentiment analysis. Because the pretrained model already captures rich language representations, this fine-tuning procedure is effective and requires comparatively little labelled data. Additionally, it is simple to add task-specific heads and carry out fine-tuning with a few lines of code thanks to tools like Hugging Face's Transformers library.

5.2.2 FINE-TUNED MODELS

This research utilized three models as a potential candidates for the sentiment analysis purposes, all based on the FinBERT model, which in turn is derived from the original BERT model.

FinBERT (ProsusAI/finbert) is an adaptation of the NLP model designed specifically for sentiment analysis, trained on 5,000 sentences to ensure robust generalisation within the financial domain.

FinTwitBERT (StephanAkkerman/FinTwitBERT-sentiment) is the refined iteration of the previously fine-tuned model, further tailored to financial data, specifically designed for concise and precise tweet-style communication. Utilised 12 million tweets sourced from various sources, including a variety of tweets related to cryptocurrency and stock-specific datasets.

Finally, another derivation of **FinBERT (yiyanghjust/finbert-tone)** has been analysed and trained on financial communication texts, positing that it could more effectively interpret financial tweets from official accounts.

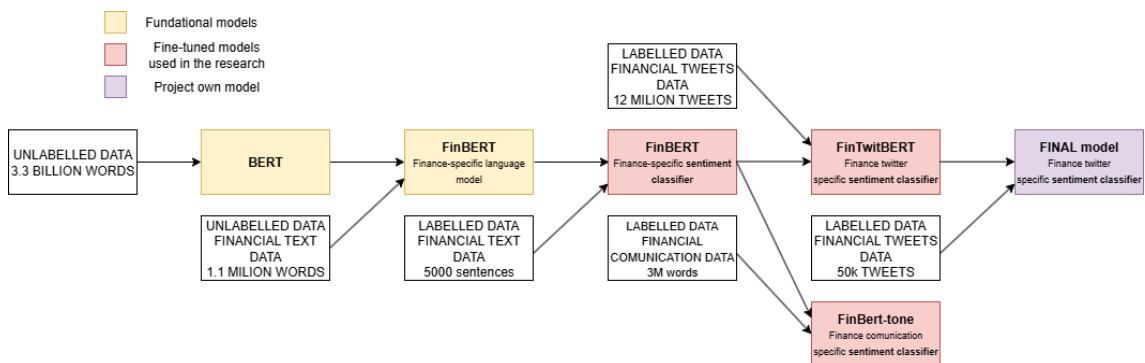


Figure 8: Visual representation of the BERT fine-tuning. Source: Own elaboration

All three models were evaluated using unseen data (n= 10.000), yielding the following results:

ProsusAI/finbert

The confusion matrix [\(see Annex 8.2, Table 10\)](#) shows the model defaults to neutral predictions regardless of the true class. Of 3,215 true negatives, only 311 were identified correctly, while 2,579 were labeled neutral and 325 positive. For 2,196 neutrals, just over half (1,141) were correct; 920 were flagged positive and 135 negative. Positives fared worst—only 265 of 4,589 were detected, with 3,324 labeled neutral or negative. This “safe” bias reflects training on long-form financial prose, not suitable to Twitter’s slang and brevity. Total accuracy: **17.17%**.

StephanAkkerman/FinTwitBERT_sentiment

Its confusion matrix ([see Annex 8.2, Table 11](#)) shows strong, balanced performance. Of 3,215 negatives, 2,467 were correct, with 307 neutral and 441 positive errors. Neutrals (2,196) were almost all correct (1,994), with only 145 positive and 57 negative errors. Of 4,589 positives, 3,904 were detected, with 685 misclassified. High and symmetric true positive/negative rates confirm that pretraining and fine-tuning on millions of tweets enable nuanced understanding of informal, emoji-rich language. Total accuracy: **83.65%**.

yiyanghkust/finbert_tone

The confusion matrix ([see Annex 8.2, Table 12](#)) shows mid-range performance. Of 3,215 negatives, 2,800 were correct, with 415 mislabeled. Neutral detection was weak—only 552 of 2,196 were correct, with 1,499 called negative and 145 positive. Positives (4,589) saw 1,276 correct identifications, with 3,204 labeled negative and 109 neutral. This bias toward negative likely stems from formal “tone” pretraining. Total accuracy: **46.28%**.

5.2.2.1 FINAL NLP MODEL

ProsusAI/finbert misclassifies most tweets as neutral, detecting ~10% of true negatives and <6% of true positives. FinTwitBERT sentiment delivers balanced, high accuracy; 75% for negatives, ~90% for neutrals, and ~85% for positives—with minimal confusion. Yiyanghkust/finbert tone captures most negatives but underperforms on neutrals and positives (<33% for both). Only the tweet-trained FinTwitBERT achieves robust, evenly distributed accuracy, underscoring the importance of domain-specific pretraining or targeted fine-tuning for conversational tweet sentiment.

With a Macro-F1 score of 0.83, the NLP model demonstrated strong overall performance. Positive (F1: 0.86) and negative (F1: 0.83) sentiments perform well at the class level, whereas neutral sentiment lags somewhat (F1: 0.81). Predictions that were neutral (0.73) had lower precision than those that were negative (0.89) and positive (0.87). Neutral cases had the highest recall (0.91), followed by positive cases (0.85) and negative cases (0.77), suggesting that the model is more likely to correctly identify neutral cases at the expense of precision. The model's performance is balanced overall, though neutral-class precision could be increased.

Class	Precision	Recall	F1
Negative	0.893517	0.767341	0.825636
Neutral	0.725355	0.908015	0.806471
Positive	0.869488	0.850730	0.860007

Table 1: Macro F1 results FinTiwT Bert. Source: Own elaboration

Accuracy in the sentiment analysis is a key step in the overall pipeline; consequently, a new model has been fine-tuned in this research with a data source containing 38,100 tweets of financial data (<https://huggingface.co/datasets/TimKoornstra/financial-tweets-sentiment>) using as a base model the best NLP performer model of the 5.2.2 section.

After performing an optuna hyperparameter optimization with a subsample of 5,000 tweets and fitting the model (now with 10,000 rows) in the best combination, order to save computational time (see [Anenx 8.3 src/nlp_finertuning.ipynb](#) for more details). The F1 metric increased by 3.2 p.p posterior of applying the fine-tuning.

Additionally, in the pipeline the model of interest it is not restricted to be the best one out of the three analyzed, it can be freely selected. For the NLP pipeline, it is selected FinTwitBert due to capacity constraints after training the own model.

Metric	Value
loss	0.367494
accuracy	0.8675
precision	0.866367
recall	0.859574
f1	0.862778
Runtime (sec)	114.5431

Table 2: Results after fine-tuning. Source: Own elaboration

5.3 DECISION MAKING LOGIC

To transform sentiment into concrete trading actions, a well-defined criterion that the Broker API can interpret is required. Since the analysis involves multiple tweets filtered by a specific keyword, such as AAPL or Bitcoin, the output of the sentiment model takes the form of a dataframe. Each row of the dataframe contains the labelling "positive," "negative," or "neutral" and its corresponding score, which reflects the probability assigned by the model to that category.

The dataframe taken globally could be seen as a current sentiment regarding the asset for the overall market, then used as a proxy for the herding behaviour it could give rise to. Since each row of this dataframe is the tweet of an individual, we can aggregate all of them and then obtain the general view of the market. If most of the tweets are labelled as “positive”, investors expect a bullish asset, and vice versa with “negative”. If most of them are “neutral” or we have an equal amount between “positive” and “negative”, it is probable that herding behaviour will not occur and consequently there will be no changes in the asset price for the high-frequency analysis.

Tweet	Label	Score
"Apple just reported record quarterly earnings, beating analyst expectations. \$AAPL soaring!"	Positive	0.95
"Strong iPhone sales are driving Apple's revenue growth. Investors are happy. \$AAPL"	Positive	0.9
"Apple's services segment is growing faster than expected. Bullish outlook for \$AAPL."	Positive	0.92
"Apple just announced a stock buyback program. Shareholders love it! \$AAPL"	Positive	0.88
"Analysts upgraded Apple, citing strong fundamentals and growth potential. \$AAPL"	Positive	0.85
"Apple's supply chain issues could hurt next quarter's earnings. \$AAPL might dip."	Negative	0.8
"Apple's stock price moved slightly today after the earnings report. \$AAPL"	Neutral	0.5

Table 3: Sample table for 10 AAPL tweets. Source: Own elaboration

In table 3, we could see an example if we queried 10 tweets of AAPL, taken as a representative of the overall investor feelings. The scores could be aggregated, obtaining 71% of them think AAPL stock is doing great and consequently a good potential investment making the price rise by the assumptions of behavioural finance explained above.

5.3.1 TRADING SIGNAL GENERATION

In this study it has been developed a function, *generate_trading_signal()*, that take the overall sentiment analysis table and converts it into trading actions; “BUY”, “SELL” or “HOLD”. It can be applied in two ways.

First, the average sentiment is computed by converting qualitative sentiment labels (positive, neutral, negative) into quantitative sentiment values by assigning confidence scores. If the label is positive the number remains equal, if negative it is converted into a negative value and if it is neutral is assigned as a 0. These values are aggregated across the dataset to compute the mean sentiment.

If the mean exceeds the user-specified *buy_threshold*, the function return “BUY”. If it falls below *sell_threshold* (by default the negative of buy threshold) it returns “SELL” and if the mean sentiment is between these two numbers, we will have “HOLD”.

The second method overcomes the scale issue of the first one: in a polarized market, the average sentiment can be near zero, a lot of neutral (hence 0s) and a mix of positive and negative that will cancel themselves (as one is the negative of the other), making difficult to assign an arbitrary number for the two parameters presented above.

Here, the function computes proportion of “positives” and “negatives” of the overall dataframe. If the positive or negative proportion exceeds *ratio_threshold*, (between 0 and 1, threshold to trigger any action, specified by the user) it returns “BUY” or “SELL” respectively. Otherwise, it returns “HOLD.”

Both criteria operate simultaneously within a single decision rule: the function returns an action as soon as any condition is satisfied, resolving conflicts by a fixed priority (mean-based before ratio-based, and BUY before SELL); if none are satisfied, the outcome is HOLD.

BUY if avg_sentiment > buy_threshold
else BUY if pos_ratio >= ratio_threshold
else SELL if avg_sentiment < sell_threshold
else SELL if neg_ratio >= ratio_threshold
otherwise HOLD

In table 3, if user decided; $ratio_threshold = 0.7$ and $buy_threshold = 0.3$, action flown to the broker API will be “BUY” as the first condition is satisfied.

```

ACTION = generate_trading_signal(df=res, buy_threshold=0.3, sell_threshold=-0.3,ratio_threshold=0.7)
ACTION

✓ 0.0s
Action: BUY – Average sentiment (0.56) is above the buy threshold (0.3).
(np.float64(0.5643692016601562), 'BUY')

```

Figure 9: Code snippet of the decision logic implemented. Source: Own elaboration

5.4 TRADING SIGNAL

In order to operationalize the trading strategy derived from sentiment analysis and market data, a dedicated function was developed to generate and execute trading signals. The function is implemented through the *AlpacaTrader* class, which integrates directly with the Alpaca API to handle both stock and cryptocurrency assets. This implementation allows the system to translate abstract sentiment indicators into concrete buy, sell, or hold decisions, while simultaneously incorporating safeguards such as user-defined thresholds and timeframes. By combining historical data retrieval with threshold-based limit order placement, the function ensures that trades are only executed when market conditions align with the predefined strategy, thereby bridging the gap between theoretical modelling and actionable trading execution.

A threshold percentage specified by the user is applied to the reference mean price, which is established by the trading logic (output of Section 5.3) using recent historical data. A price corridor that is deemed acceptable for trades is defined by this threshold. For example, sell orders are validated when the price does not drop below the floor, while buy orders are only activated when the most recent market price does not surpass the computed ceiling. In this sense, the system includes a safety feature that stops orders from being carried out under adverse or extremely unstable circumstances.

This is key since the analysis will be based on high-frequency trading rather than long-time investments to take advantage of the irrationality. If the price is high (if buying) or low (if selling), the arbitrage caused by the market irrationality disappears, and then the strategy will not be profitable.

Orders are placed as limit orders through Alpaca's trading client when the market conditions match these restrictions. The design also accommodates the customs of various markets: cryptocurrency orders stay open until specifically cancelled, whereas stock orders default to a one-day validity. This eliminates the need for distinct implementations and guarantees flexibility across asset classes.

It is important to note that the function is recommended to be used primarily when markets are open. Attempting to query historical data or place orders outside trading hours—particularly for equities—may result in incomplete data retrieval or execution errors (see Appendix, Figure 16) as the Alpaca API cannot access price bars for closed markets. In practice, this limitation highlights the importance of aligning the signal generation process with live trading sessions.

All things considered, by carefully connecting forecasts to market execution, this trading signal function operationalises the forecasting model. The implementation places a strong emphasis on reproducibility because the logic is organised according to a systematic and rule-based structure, and transparency because every decision is recorded along with market data and order details.

```
trader.execute_action(  
    symbol= TICKER,  
    asset_type=ASSET_TYPE,  
    action=ACTION[1],  
    qty= 0.2,  
    threshold_pct=0.1,  
    timeframe_unit="Hour",  
    limit=20,  
)  
trader.trading_client.get_orders()  
✓ 0.2s  
  
Latest price for ETH/USD is 4228.66 (mean 4162.44)  
No action taken, outside your price threshold, try to adjust the limit or market already moved.
```

Figure 10: Code snippet showing an attempt to buy a stock using the AlpacaTrader class. Source: Own elaboration

To illustrate this more clearly, in the example above (Figure 10), we attempt to execute a trading action on Ethereum, as indicated by the ticker printed in the first output line. The action itself is determined by the label defined in Section 5.3.1; in this case, the system

attempts to buy a fractional quantity of 0.2 units. The function checks whether the current price of the asset deviates by more than 1% (*threshold_pct*) from the mean of the last 20 hours. In this example, the calculated mean price is 4162, while the current price is 4228, representing a deviation of 1.5%. As a result, the function is stopped, and no order is executed. This mechanism is particularly useful because excessive price movements may already reflect speculative hype, reducing the possibility of capturing profitable opportunities.

For the sake of demonstration, these parameters were chosen; however, given that the scope of the analysis is more focused on high-frequency trading, the recommended configuration would involve setting the *timeframe_unit* in minutes and using a *limit* between 10 and 60.

6. RESULTS

6.1 EVALUATION METHODOLOGY

The empirical assessment of the trading strategy aims to validate the fundamental premise of this study: that investor behavior in financial markets is not fully rational, and that sentiment indicators derived from public discourse can consequently offer predictive capabilities that surpass conventional methodologies. To confirm this, it compared the sentiment-based strategy with a buy-and-hold benchmark, representing the performance of a rational, passive investor who refrains from active trading decisions. If the sentiment-driven approach yields enhanced profitability and improved risk-adjusted results, it would demonstrate that market irrationality can be systematically leveraged through sentiment analysis. In other words, if the proposed method works better, it will mean that emotional and behavioral factors that are part of investor sentiment are shown in asset prices in ways that a purely passive strategy can't.

Sentiment-based trading strategies work best for short-term or high-frequency time frames. This is because news or social media activity can quickly change how investors feel about something, which can then quickly change market prices. Behavioral responses, including fear, overconfidence, or speculative hype, frequently induce transient deviations from fundamental value, which are progressively rectified as markets assimilate new information. So, the ability of sentiment signals to predict what will happen quickly fades over time, which makes them best for trading on the same day or in the short term.

In this situation, taking advantage of inefficiencies caused by emotions requires acting within minutes or hours instead of days or weeks, because the market usually fixes overreactions quickly. Sentiment analysis can help with medium-term strategies, but the best proof of investor irrationality is likely to be found in high-frequency settings, where emotional and behavioral responses are most directly reflected in asset prices before rational adjustments happen.

Then for all 8 assets analysed: AAPL, AMZN, ACB, DGII, BTC/USD, ETH/USD, DOT/USD and LINK/USD, the same process will be followed.

On day x , the pipeline is activated, querying tweets, transforming sentiment, and executing an action through the broker. Simultaneously, for the buy and hold strategy, based on data from [Yahoo Finance](#), the action most frequently recommended by analysts is also executed (see Figure 11). On day $x + i$, with i representing an arbitrary number in the range (1, 7, 14), the pipeline is re-activated, and the returns from both approaches are subsequently compared.

For cryptocurrencies, where analyst recommendations are not available, the strategy consists of purchasing the asset on day x and applying the same evaluation procedure described above.

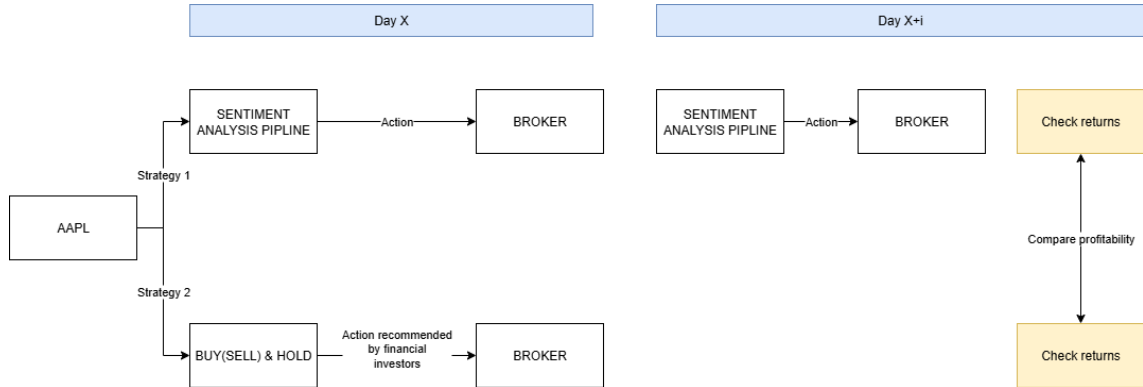


Figure 11: Schema investing strategies. Source: Own elaboration

In the special case where both on day x and on day $x + i$ the sentiment analysis pipeline generates consecutive buy signals; the comparison of returns is conducted at the end of the day in order to allow the price to fluctuate before evaluation.

The days of analysis of this project were 7th of August 2025 at 15:00 pm (opening hour of American stock exchange at Spanish time) for the initial price. Consequently, for the test days we had 8th of August, 14th of August and 21st of August.

6.2 STOCKS

For large stocks with high capitalization backed by extensive analysis and institutional investors behind, the impact of irrationality it is found to be limited, at least during “normal” periods without general market crisis. For the two stocks analysed, NLP pipeline suggested a buying action only once out of eight possible opportunities, specifically for AAPL at the first day. On the test dates where pipeline is also activated, HOLD was the action taken for both. This indicates that the quantitative method suggested a BUY & HOLD (B&H) strategy for AAPL while preferred to do not invest in AMZN during the period of analysed.

For both actions, Yahoo analysts suggested a BUY, then for the naïve strategy used to test against the NLP pipeline, we performed a BUY & HOLD. We obtained the following returns:

	07/08	08/08	14/08	21/08	1D Return (%)	1W Return (%)	2W Return (%)
AAPL	219.54	222.01	232.05	225.33	1.13%	5.70%	2.64%
AMZN	225.54	222.01	230.05	221.13	-1.57%	2.00%	-1.96%
ACB	5.762	5.863	7.446	6.594	1.75%	29.23%	14.44%
DGII	31.16	32.4	33.95	32.81	3.98%	8.95%	5.30%

Table 4: Returns for the BUY & HOLD strategy. Source: Own elaboration

	07/08	08/08	14/08	21/08	1D Return (%)	1W Return (%)	2W Return (%)
AAPL	BUY->219.54	HOLD	HOLD	HOLD	1.13%	5.70%	2.64%
AMZN	HOLD	HOLD	HOLD	HOLD	0.00%	0.00%	0.00%
ACB	BUY->5.762	BUY->5.863	HOLD	HOLD	0.87%	29.40%	14.60%
DGII	HOLD	HOLD	BUY-> 33.95	HOLD	0.00%	0.00%	-3.35%

Table 5: Actions and returns of NLP pipeline. Source: Own elaboration

When comparing the two baselines, we observe identical returns for AAPL (since the same strategy was applied) and better results for AMZN in the 1-day and 2-week horizons. However, we miss the potential gain from the 1-week horizon, during which the stock showed a positive return (see Table 6).

	1D Return (%)	1W Return (%)	2W Return (%)
AAPL	0 p.p	0 p.p	0 p.p
AMZN	1.57 p.p	-2 p.p	1.96 p.p
ACB	-0.88 p.p	0.17 p.p	0.16 p.p
DGII	-3.98 p.p	-8.95 p.p	-8.65 p.p

Table 6: Return NLP - return Buy (Sell) & Hold. Source: Own elaboration

These results could suggest that in big companies where huge capital is invested, and where public and private financial institutions converge, herding behaviour is less likely to occur. This is because the individual investors represent only a small percentage of shareholders and hence limited ability to move upwards or downwards the price.

In contrast, a different picture emerges for small capitalization stocks. We find out that NLP model is more “active” taking actions, but results are less conclusive: for one stock, ACB, results were

positive, while for the other do not. This could be due to some of the limitations of the study discussed in Section 7.

For ACB we have better returns compared to a simple B&H since we have in our portfolio 2 stocks, bought at day 1 and 2, since the price went up both actions are weighting more. In the other side, for DGII, the timing of the call unfavourable and the performance was clearly worse as the model bought at a relatively high price (see Figure 12).

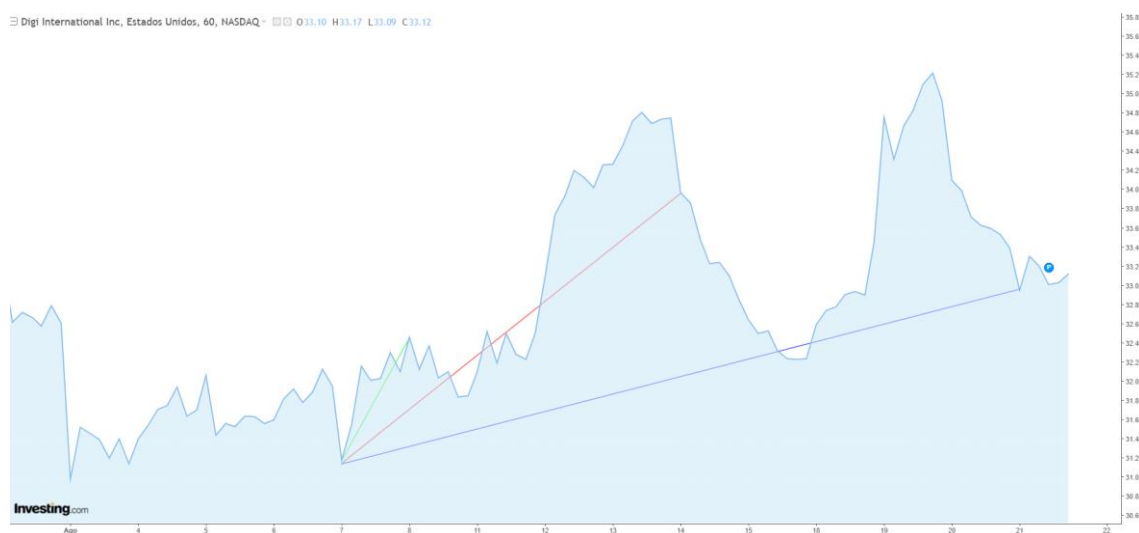


Figure 12: Price evolution for different time spans. Green line 1D, Red line 1W, Blue line 2W. Source: <https://es.investing.com/>

6.3 CRYPTOCURRENCIES

The case of the crypto actives is totally different to the stock ones, here the NLP model clearly outperformed a simple strategy of B&H for small and big assets. The lack of fundamental analysis, the endogenous behaviour of the currency or the large following of the assets in X could be factors that benefit this approach and favour irrationalities.

	07/08	08/08	14/08	21/08	1D Return (%)	1W Return (%)	2W Return (%)
BTC	115136	116653	121668	113520	1.32%	5.67%	-1.40%
ETH	3726.31	3901.35	4724.41	4187.26	4.70%	26.79%	12.37%
DOT	3.682	3.911	4.196	3.83	6.22%	13.96%	4.02%
LINK	16.74	19.2	23.36	25.64	14.70%	39.55%	53.17%

Table 7: Returns for the BUY & HOLD strategy. Source: Own elaboration

For the NLP strategy the calls were mostly accurate regardless the time span of the analysis or the respective price. For cryptos, NLP generated 7 trading actions (compared with the 4 of the stocks) doubling the broker requests.

For BTC, the strategy entered a one-week position, buying and later selling the asset, and achieved a 4.2% return. This took advantage of volatility whereas the B&H strategy did not intercept losing -1.4% after 2 weeks.

For Ethereum, the return was the same, with the difference that in the last period the asset was sold, leaving the cash in the account ready for reinvestment.

For smaller cryptocurrencies, in the case of DOT, NLP adopted a short strategy by selling during the second and third periods. Although the 1-day and 1-week returns temporarily showed negative values, after two weeks this strategy outperformed by 1.58 percentage points.

Finally for LINK, the NLP strategy followed the same behaviour of the B&H strategy so in this case, sentiment analysis did not provide any additional benefit.

	07/08	08/08	14/08	21/08	1D Return (%)	1W Return (%)	2W Return (%)
BTC	HOLD	BUY->115563	SELL->121668	HOLD	0.00%	4.20%	4.20%
ETH	BUY->3726.31	HOLD	HOLD	SELL->4187.26	4.70%	26.79%	12.37%
DOT	HOLD	SELL->3.911	SELL->4.196	HOLD	0.00%	-6.79%	5.60%
LINK	BUY->16.74	HOLD	HOLD	HOLD	14.7%	39.55%	53.17%

Table 8: Actions and returns of NLP pipeline. Source: Own elaboration

	1D Return (%)	1W Return (%)	2W Return (%)
BTC	0 p.p	-1.47 p.p	5.6 p.p
ETH	0 p.p	0 p.p	0 p.p
DOT	-6.22 p.p	-20.75 p.p	1.58 p.p
LINK	0 p.p	0 p.p	0 p.p

Table 9: Return NLP - return Buy (Sell) & Hold. Source: Own elaboration

Hence, it is possible to conclude that NLP strategies are profitable in the long time compared vs B&H. After 2 weeks of investing neither of the strategy followed by each asset returned less benefits than the naïve strategy, remarking the possible underlying effect of the irrationality that characterizes this type of asset.



Figure 13: Price evolution for ETH. Source: <https://es.investing.com/>

7. LIMITATIONS

Although the results of this study are promising, several limitations should be acknowledged, as they may introduce bias into the analysis.

First, the limitations of the X API in the number of tweets available for querying could significantly affect the results. In principle, the more tweets accessible to the NLP model, the better the ability to capture and generalize market sentiment. However, access was restricted to 100 tweets per month, which is likely insufficient. With few queries used in this research, the dataset may contain noise and fail to fully represent overall market dynamics.

Second, this study pre-selected specific stocks for analysis. While this choice allowed for focused testing, it also excluded potentially valuable opportunities in lesser known, more volatile assets, where the NLP model might have demonstrated even greater benefits. A more robust approach could involve integrating entity recognition with sentiment analysis to dynamically identify investment opportunities rather than relying solely on a predetermined set of assets.

Third, the NLP pipeline relied on a pre-trained model (FinTwitBERT) to analyze tweets. While this model provided adequate performance, employing a domain-adapted pre-trained model could likely improve accuracy significantly. The main reason for not training or fine-tuning a custom model in this study was infrastructure constraints, since tokenizers and related model components require substantial storage and computational capacity.

8. CONCLUSIONS

This study has evaluated the performance of an NLP-based sentiment analysis strategy compared with a naïve BUY & HOLD approach across both stocks and cryptocurrencies. The results suggest that NLP-driven decisions can, under certain conditions, improve investment outcomes, although the benefits vary depending on the asset class and market characteristics.

For large-cap stocks, where institutional investors and extensive market coverage reduce the influence of retail sentiment, the NLP strategy produced fewer trading signals and its performance was largely aligned with BUY & HOLD. This suggests that in highly liquid and efficient markets, sentiment extracted from text data has limited ability to systematically outperform passive strategies.

In contrast, for small-cap stocks, the NLP pipeline generated more trading actions and, in some cases, delivered higher returns. Nevertheless, the evidence was mixed: while positive results were observed for ACB, the model's timing was less effective for DGII, leading to underperformance. This indicates that sentiment-based approaches may capture inefficiencies in smaller markets, but their reliability remains subject to limitations such as timing risk and data sensitivity.

For cryptocurrencies, the NLP model proved to be more active and effective, producing a higher number of signals compared with stocks. The results for BTC and Ethereum demonstrate the model's ability to exploit short-term volatility, outperforming BUY & HOLD in scenarios where passive strategies incurred losses. For DOT, the adoption of a short-selling strategy delivered superior results after two weeks, despite initial temporary underperformance. However, for LINK, the strategy behaved identically to BUY & HOLD, showing no added value from sentiment analysis.

Overall, the findings highlight three key points. First, NLP-based strategies show greater potential in volatile and less efficient markets, such as cryptocurrencies and small-cap stocks. Second, while the model successfully identified profitable entry and exit points in several cases, it also exhibited limitations, particularly when trades were executed at unfavourable moments. Third, the analysis underscores the importance of complementing sentiment-driven models with additional signals or filters to enhance robustness and mitigate risks.

In conclusion, sentiment analysis via NLP can represent a valuable tool for investment decision-making, especially in markets where retail sentiment and volatility play a larger role. However, its effectiveness is neither universal nor guaranteed. Future research should focus on expanding the dataset, refining the model to better capture market-specific dynamics, and exploring hybrid strategies that combine NLP with other quantitative techniques.

9. REFERENCES

- Akkerman, S., & Koornstra, T.** (2023). *FinTwitBERT: A specialized language model for financial tweets* [GitHub repository]. GitHub. <https://github.com/TimKoornstra/FinTwitBERT>
- Alpaca Markets.** (2022). *Getting started with Alpaca-py: the official Python SDK* [Documentation]. Alpaca Markets. https://alpaca.markets/sdks/python/getting_started.html
- Antweiler, W., & Frank, M. Z.** (2004). *Is all that talk just noise? The information content of internet stock message boards.* *The Journal of Finance*, 59(3), 1259–1294. <https://doi.org/10.1111/j.1540-6261.2004.00662.x>
- Barbosa, A.** (2024, October 29). *From Tweets to Trades: The Risks of Social Media in Investing.* CFA Institute Enterprising Investor. <https://blogs.cfainstitute.org/investor/2024/10/29/from-tweets-to-trades-the-hidden-risks-of-social-media-in-investing/>
- Bernoulli, D.** (1738). *Specimen theoriae novae de mensura sortis. Commentarii Academiae Scientiarum Imperialis Petropolitanae*, 5, 175–192.
- Bollen, J., Mao, H., & Zeng, X.** (2011, March). *Twitter mood predicts the stock market.* *Journal of Computational Science*, 2(1), 1–8. <https://doi.org/10.1016/j.jocs.2010.12.007>
- Ciampaglia, G. L., Menczer, F., & The Conversation US.** (2018, June 21). *Biases make people vulnerable to misinformation spread by social media.* *Scientific American*. <https://www.scientificamerican.com/article/biases-make-people-vulnerable-to-misinformation-spread-by-social-media/>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K.** (2018, October 11). *BERT: Pre-training of deep bidirectional transformers for language understanding* (arXiv 1810.04805). arXiv. <https://doi.org/10.48550/arXiv.1810.04805>
- Fama, E. F.** (1970). *Efficient capital markets: A review of theory and empirical work.* *Journal of Finance*, 25(2), 383–417.
- Genc, Z.** (2020, July 31). *FinBERT: Financial Sentiment Analysis with BERT.* Prosus AI Tech Blog (Medium). <https://medium.com/prosus-ai-tech-blog/finbert-financial-sentiment-analysis-with-bert-b277a3607101>
- Greyling, T., & Rossouw, S.** (2025, March 26). *Twitter sentiment and stock market movements: The predictive power of social media.* CEPR's VoxEU. <https://cepr.org/voxeu/columns/twitter-sentiment-and-stock-market-movements-predictive-power-social-media>
- IBM.** (2024, August 11). *What is Natural Language Processing?* IBM Think. <https://www.ibm.com/think/topics/natural-language-processing>
- Investopedia.** (n.d.). *Behavioral finance: Biases, emotions and financial behavior.*

- Investopedia. <https://www.investopedia.com/terms/b/behavioralfinance.asp>
- Investopedia.** (2006, May). *Market psychology: What is it and predictions*. Investopedia. <https://www.investopedia.com/terms/m/marketpsychology.asp>
- Investopedia.** (2024, August). *CBOE Volatility Index (VIX): What does it measure in investing?* Investopedia. <https://www.investopedia.com/terms/v/vix.asp>
- Khalid, S.** (2019, November 3). *BERT Explained: A Complete Guide with Theory and Tutorial*. Medium. <https://medium.com/@samia.khalid/bert-explained-a-complete-guide-with-theory-and-tutorial-3ac9eb8fa7c>
- Modgil, S., Singh, R. K., Gupta, S., & Dennehy, D.** (2024). *A confirmation bias view on social media induced polarisation during COVID-19*. *Information Systems Frontiers*, 26(2), 417–441. <https://doi.org/10.1007/s10796-021-10222-9>
- Mossin, J.** (1966). *Equilibrium in a capital asset market*. *Econometrica*, 34(4), 768–783.
- Pelster, M., & Romero-Gonzalez, G.** (2016). *Social media interactions and biases in investment decisions* [Discussion paper]. Centre for Economic Policy Research.
- Rogers, J.** (2024, May 16). *These behavioral trends drove the GameStop and AMC meme-stock rally*. *MarketWatch*. <https://www.marketwatch.com/story/these-behavioral-trends-drove-the-gamestop-and-amc-meme-stock-rally-3e7ed0ab>
- Tweepy developers.** (2025). *Tweepy 4.14.0 API documentation*. Tweepy. <https://docs.tweepy.org/en/stable/api.html>
- Von Neumann, J., & Morgenstern, O.** (1944). *Theory of games and economic behavior*. Princeton University Press.
- X Developer Platform.** (2025). *Products – X Developer Portal* [Webpage]. X Developer Platform. <https://developer.x.com/en/portal/products>
- Yahoo Finance.** (n.d.). *Yahoo Finance*. <https://finance.yahoo.com/> – Leading finance portal offering real-time data, news, analytics, and tools for investors.
- Yang, X., Fernández-Pérez, A., & Indriawan, I.** (2023). *Social media sentiment, investor herding and informational efficiency*. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.5274226>
- Yoon, J., & Oh, G.** (2022). *Investor herding behavior in social media sentiment*. *Frontiers in Physics*, 10, 1023071. <https://doi.org/10.3389/fphy.2022.1023071>

10. ANNEX

10.2 FIGURES

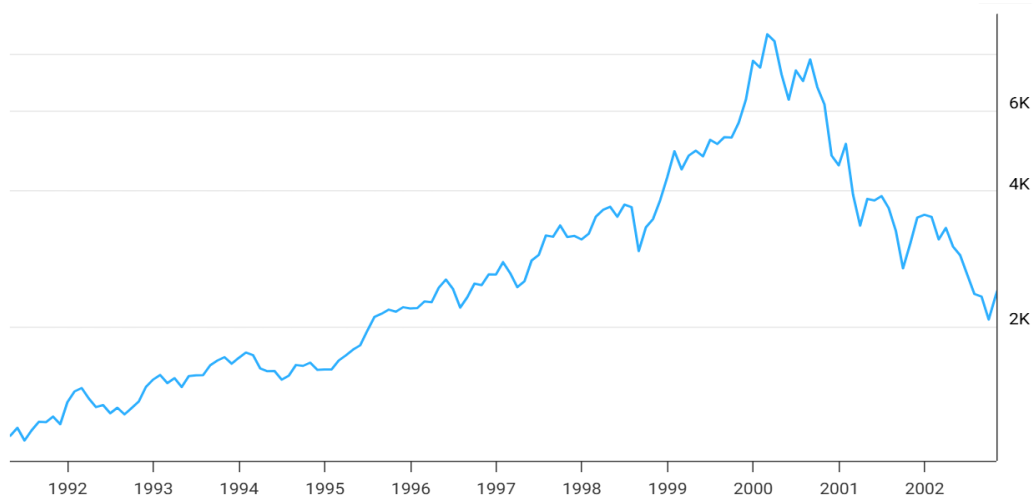


Figure 14: NASDAQ Composite (Adj. for inflation) 1991-2002. Source: <https://www.macrotrends.net>

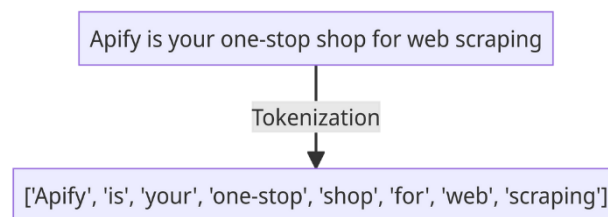


Figure 15: Example of tokenization. Source: <https://blog.apify.com/nlp-techniques/>

```
trader.trading_client.get_clock()

✓ 0.5s

{
  'is_open': False,
  'next_close': datetime.datetime(2025, 8, 20, 16, 0, tzinfo=TzInfo(-04:00)),
  'next_open': datetime.datetime(2025, 8, 20, 9, 30, tzinfo=TzInfo(-04:00)),
  'timestamp': datetime.datetime(2025, 8, 20, 4, 27, 53, 314306, tzinfo=TzInfo(-04:00))}

trader.execute_action(
    symbol= TICKER,
    asset_type=ASSET_TYPE,
    action=ACTION[1],
    qty= 0.2,
    threshold_pct=0.0001,
    timeframe_unit="Min",
    limit=10,
)
trader.trading_client.get_orders()

✗ 0.2s

Error fetching data for AAPL: {"message": "subscription does not permit querying recent SIP data"}
, market closed
```

Figure 16: Code snippet, example of buying an action with the market closed. Source: Own elaboration

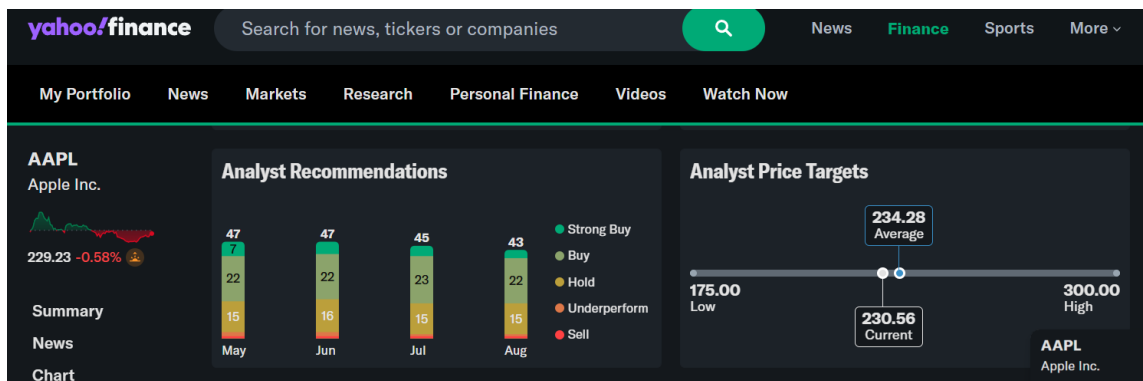


Figure 17: Analysis AAPL. Source: <https://finance.yahoo.com/>

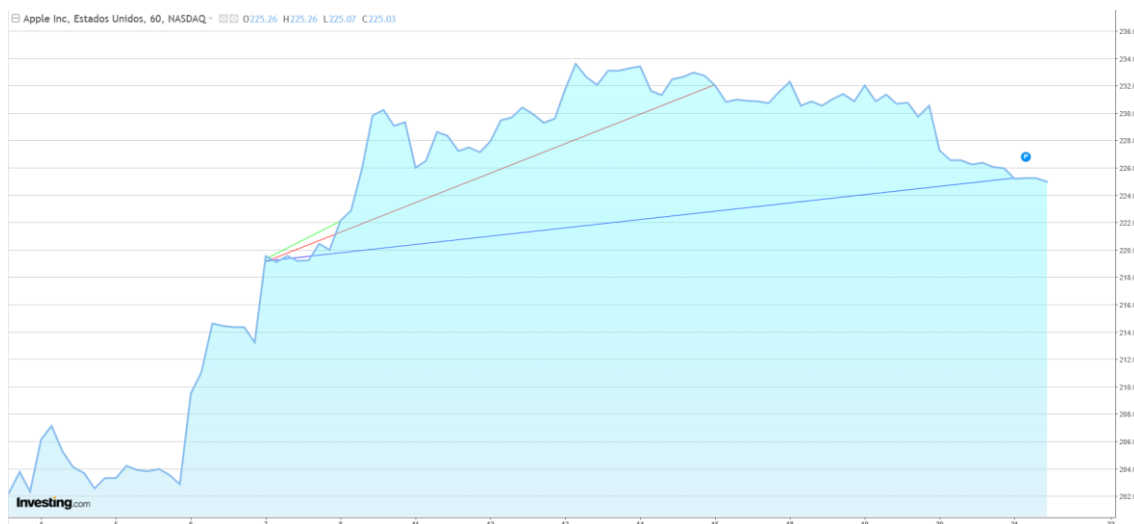


Figure 18: Price evolution for AAPL. Source: <https://es.investing.com/>



Figure 19: Price evolution for Amazon. Source: <https://es.investing.com/>

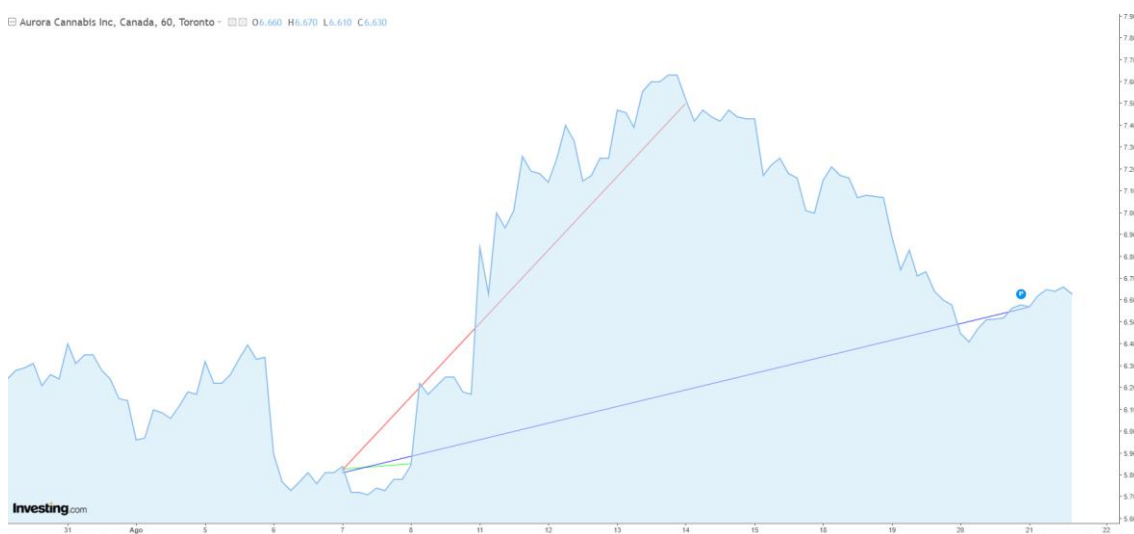


Figure 20: Price evolution for ACB. Source: <https://es.investing.com/>



Figure 21: Price evolution for Bitcoin. Source: <https://es.investing.com/>

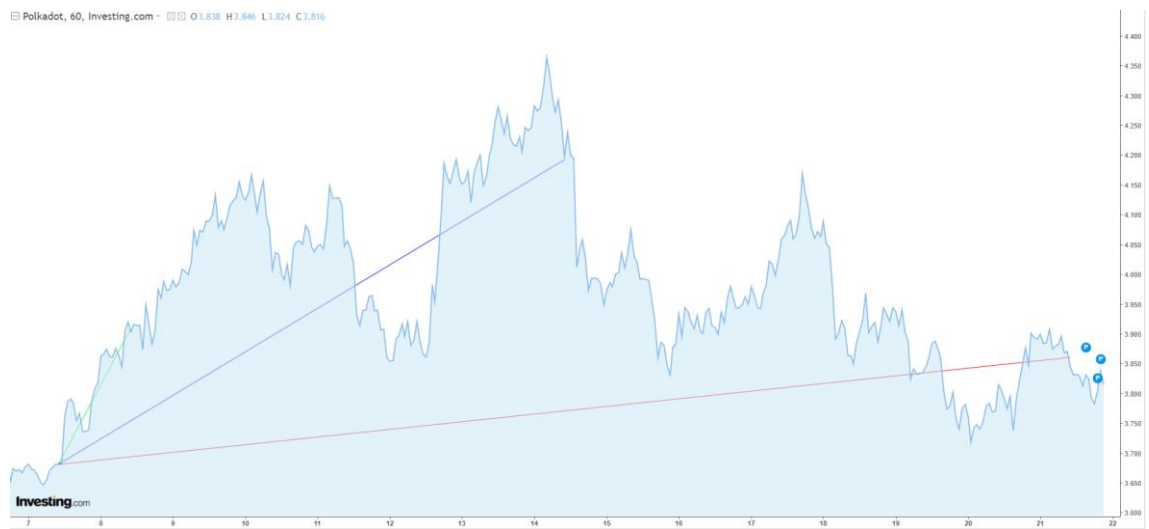


Figure 22: Price evolution for DOT. Source: <https://es.investing.com/>

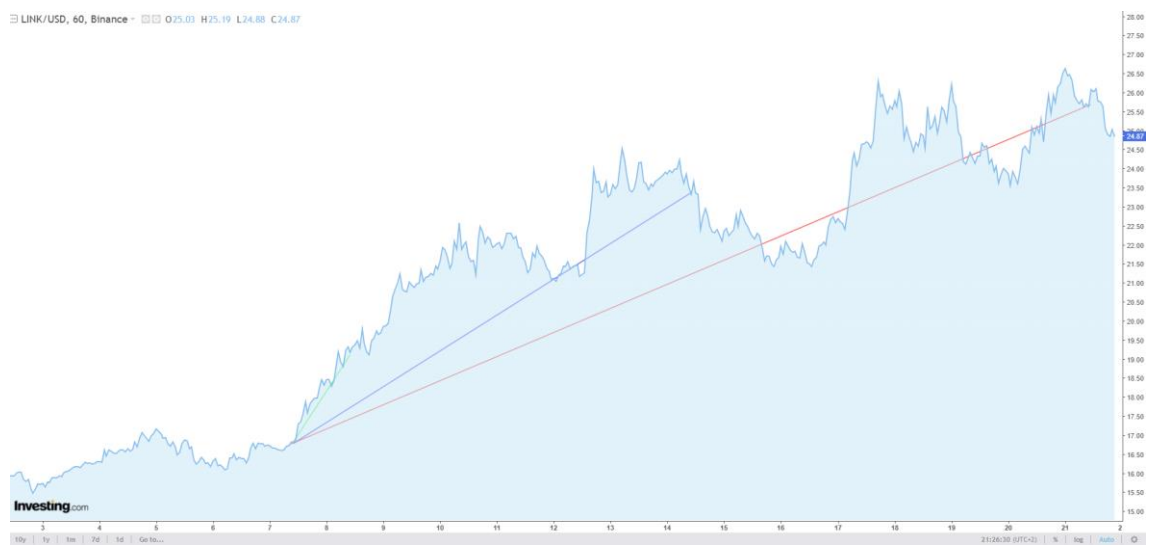


Figure 23: Price evolution for LINK. Source: <https://es.investing.com/>

10.3 TABLES

Actual\Predicted	Negative	Neutral	Positive
Negative	311	2579	325
Neutral	135	1141	920
Positive	1217	3107	265

Table 10: Confusion matrix model ProsusAI/finbert. Source: Own elaboration

Actual\Predicted	Negative	Neutral	Positive
Negative	2467	307	441
Neutral	57	1994	145
Positive	237	448	3904

Table 11: Confusion matrix model StephanAkkerman/FinTwitBERT-sentiment. Source: Own elaboration

Actual\Predicted	Negative	Neutral	Positive
Negative	2800	157	258
Neutral	1499	552	145
Positive	3204	109	1276

Table 12: Confusion matrix model yiyanghkust/finbert-tone. Source: Own elaboration

10.4 CODE

Project repository: [joas1847/financial-nlp](https://github.com/joas1847/financial-nlp)

The notebook that runs the overall pipeline is *demo.ipynb*. In the first cells, user must provide the necessary tokens (via .env or by entering them directly in the Jupyter notebook) to make the API connections, as well as the the variables of interest such as the ASSET to query or ACCOUNTS to retrieve tweets.

The pipeline_scripts/ folder, it is where the helper modlues (steps of the NLP pipeline) are coded.

In *nlp_finetunning.ipynb*, you can find the performance analysis of the NLP models used in this study as well the code to train the own fine tuned NLP model .

The data used to develop these processes are stored in src/data/.

Moreover, user can build a Conda environment in their own computer for this specific project by following the commands in the GitHub Readme.md. If some error arises in the installation it is encouraged to install only the necessary libraries in the terminal.