

## UNIVERSIDADE CATÓLICA PORTUGUESA

# The Effects of *Twitter* Sentiment on Renewable Energy Stock's Returns

A Portuguese Study about EDP Renováveis Stocks

Diogo Fragata Barros

Católica Porto Business School June 2023



## UNIVERSIDADE CATÓLICA PORTUGUESA

# The Effects of *Twitter* Sentiment on Renewable Energy Stock's Returns

A Portuguese Study about EDP Renováveis Stocks

Master thesis presented to

Católica Porto Business School

Submitted as partial requirement for the conferral of master's in Finance

by

Diogo Fragata Barros

Under orientation of

PhD. Mário Pedro Leite de Almeida Ferreira

Católica Porto Business School June 2023

## Acknowledgments

I would like to thank firstly to my supervisor PhD. Mário Ferreira for all the guidance and support he gave me during the elaboration of this thesis. Also, I would like to thank my master colleagues Bruno Moreira, Levi Leandro, and Paulo Dias, which were my usual work group during the last two years and helped me achieve better comprehension and grades during this Master. I would similarly like to appreciate the student's union I was part of during the first year of the master, especially my department, which helped me develop my soft skills and understanding of the real world of work we live in.

To my bachelor's friends group Alexandra Monteiro, Carolina Barreto, Mariana Morais, and Filipe Simões, who are the main reason I miss my bachelor years so much. For my hometown friends Pedro Mesquita, Hrithik Harsadrai, Duarte Araújo, Afonso Morais, Nuno Afonso, Pedro Monteiro, João Rodrigues, Diogo Viana, João Ferreira, Guilherme Lima, Inês Araújo, Miguel Amaral, José Pequeno, and Luísa Amorim thank you for always being there for me. For my university partners Pedro Caldas, Diogo Santos, Tomás Baguinho, Adolfo Magalhães, Guilherme Ricardo, Diogo Pinheiro, Diogo Simões, Francisco Ribeiro, Francisco Mendes, Diogo Alves, Matilde Novo, Margarida Varejão, Leonor Lopes, Catarina Nunes, Inês Costa, Tiago Rebelo, and all the other I might be forgetting, thank you for making these years the best of my life.

Finally yet importantly, the biggest thank you goes to my family. To my grandparents Eunice and Manuel Fragata for being my motivation. To my uncle, aunt and cousins, for helping me being a better person. To my godfather, João Morais, for being the best father ever, my godmother, Monica Morais, for leading me in the right path, and their children, Mariana, Miguel e Martim, for being like brothers and sister to me. To my mother, Teresa Fragata and Luis, thank you for never given up on me even on my darkest times and making me who I am today. To my sister, Carolina Barros, for being my role model, the person I most look up to and my biggest proud. Finally, to my father, Luís Barros, the reason I wake up every day trying to be a better person, the person I want the most to be proud of me, even not being here I know you support me wherever you are. I miss you. And I love you all. Thank you.

#### Resumo

A racionalidade dos investidores no processo de decisão de investimento tem sido tópico de discussão nas últimas décadas devido ao conflito entre duas linhas de pensamento diferentes. Várias anomalias que não iam de encontro com a hipótese do mercado eficiente deram origem a uma nova escola de pensamento à racionalidade dos investidores chamada de finanças comportamentais. Análise de sentimentos é um dos ramos desta nova linha de pensamento que estuda a influência das emoções dos investidores em diferentes variáveis económicas. Não existe consenso entre académicos se estas emoções conseguem enviesar as decisões de investimento ou não. O objetivo desta tese é observar se o sentimento presente em tweets consegue fazer prever os retornos das ações de uma empresa de energias renováveis do mercado português. Este estudo analisa a segunda maior empresa portuguesa por capitalizações, a EDP Renováveis (EDPR), no período temporal entre o dia 1 de junho de 2021 e o dia 1 de julho de 2022, e não encontrou evidência com significância de uma relação entre o estado de espírito do Twitter e os retornos das ações da EDP Renováveis. As razões que justificam estes resultados podem ser o facto da EDPR pertencer a um mercado muito pequeno e concentrado como o português, indo de encontro com a evidência empírica, assim como a composição dos proprietários das ações da empresa ter uma percentagem muito reduzida de investidores individuais, que são o tipo de investidor mais facilmente influenciado por heurísticas presentes nos tweets. Este resultado tem implicações para o desenvolvimento da teoria de análise do sentimento, dando mais detalhes da influência deste em mercados mais pequenos e concentrados, no ramo das energias Renováveis, no período de tempo do início da guerra entre a Ucrânia e a Rússia e a recuperação financeira mundial pós-Covid-19.

Palavras-Chave: Finanças Comportamentais, Sentimento, Análise Sentimental, Twitter, Energias Renováveis, Retornos.

## **Abstract**

Investors' rationality in the decision-making process has been topic of discussion in the last decades due to conflicts between schools of thought. Several anomalies in the Efficient Market Hypothesis (EMH) led to a new line of thought in the matter of rationality called behavior finance. Sentiment analysis is one branch of this new school of thought who studies investors' emotions influence on economic variables. There is no consensus between academics if these emotions can make the investment decision biased or not. The aim of this paper is to observe if the prevailing sentiment in tweets can predict the stock returns for a renewable energy company of the Portuguese market. This study looks at the second biggest company by capitalizations of the Portuguese market, EDP *Renováveis* (EDPR), in the period from the June 1<sup>st</sup> 2021, to June 1<sup>st</sup> 2022, and finds no significant evidence of a relationship between Twitter mood and EDP Renováveis stock returns. The reasons for this result might be explained by EDPR belonging to a very small and concentrated market, corroborating the existing theory, as well as the stakeholder composition of the company only having a very small percentage of individual investors, being this kind of investors the most influenced by biases and heuristics present in the tweets. These findings have implications for the development of the sentiment analysis theory, giving more details of the influence of sentiment in smaller and concentrated market, in the renewable energy branch, and in the period of the beginning of the war between Ukraine and Russia and the worldwide economic recovery from the Covid-19 pandemic.

Keywords: Behavioral Finance, Sentiment, Sentiment Analysis, Twitter, Renewable Energies, Returns.

## Index

Acknowledgments	i
Resumo	iii
Abstract	v
Graph Index	x
Table Index	xii
1. Introduction	2
1.1. General Background	
1.2. Research Gaps	
1.3. Research Question	
1.4. Originality	
1.5. Contribution to Knowledge	
1.6. Outline of the Following Chapters	
2. Literature Review	
2.1. Definition of Sentiment in Finance	
2.2. Modern Finance Theories	
2.2.1. Efficient Market Hypothesis	
2.2.1.1. Efficient Market Hypothesis Critics	
2.2.3. Sentiment Analysis	
2.2.3.1. Context	
2.2.3.2. Theoretical Analysis	
2.2.3.3. Impact on Financial Markets	
2.3. Empirical Evidence of Sentiment Analysis on Stock Returns	17
2.4. Context	1
2.4.1. <i>Twitter</i>	1
2.4.2. Macroeconomic Context Between 2021-2022	2
2.4.3. Portugal	2
2.4.4. <i>EDP</i> and <i>EDPR</i>	5
2.5. Hypothesis	7
3. Data and Methodology	
3.1. Method	10

	3.2. Data Collection	. 10
	3.3. Variables	. 11
	3.3.1. Dependent Variable	12
	3.3.2. Independent Variable	12
	3.4. Main Methodology	. 13
	3.5. Software	. 14
4.	Empirical Results	. 17
	4.1. Descriptive Statistic and Data Characteristics	. 17
	4.2. Main Results	. 23
	4.3. Discussion	. 25
5.	Conclusion	. 30
	5.1. Main Conclusions	. 30
	5.2. Implications for Management	. 31
	5.3. Limitations	. 32
	5.4. Future Research	. 32
В	ibliography	. 35
A	ppendix	. 49

# **Graph Index**

Graph 1 - Euronext Lisbon Market Capitalizations. Source: CEIC DATA (2023a)
3
Graph 2 - Number of Tweets Made by Day of the Week. Source: Twitter Search
API
Graph 3 - Number of Tweets Made by Month. Source: Twitter Search API 49
Graph 4 - Number of Tweets Made by Language. Source: <i>Twitter</i> Search API . 50
Graph 5 - Min, Max, and Average Daily Returns of EDPR Stocks. Source:
https://finance.yahoo.com/50
Graph 6 - Daily Returns of EDPR Stock. Source: https://finance.yahoo.com/.51
Graph 7 - Net Income of <i>EDPR</i> 51
Graph 8 - ROE and ROA of <i>EDPR</i> 52
Graph 9 - Debt-To-Equity and Equity Ratio of <i>EDPR</i>
Graph 10 - Dividend Yield of EDPR53

## **Table Index**

Table 1 – Anomalies in Efficient Market Hypothesis. Self-Made
Table 2 - Empirical Evidence on the Relationship Between Social Media
Sentiment and Stock Market Variables. Self-Made (Continuation)
Table 3 - Biggest Companies me Portugal by Market Capitalization. Source:
(Papadopoulos, 2023)4
Table 4 - Tweets' Polarity by Month. Source: Umigon
Table 5 - Min, Max, and Average Monthly Returns of EDPR Stocks. Source:
https://finance.yahoo.com/
Table 6 - Descriptive Statistics of the variables from $01/06/2021$ to $01/06/2022$
Table 7 - Descriptive Statistics of the EDPR historical daily returns from
01/061/2013 to 31/05/2021
Table 8 - Pearson Correlation
Table 9 - Pearson's Correlation between MOOD and CHANGE
Table 10 - Regression Statistics between CHANGE, VTOT, VPOS and VNEG. 23
Table 11 - Multiple Regression Model between CHANGE, VTOT, VPOS and
VNEG
Table 12 - Ridge Regression between CHANGE, VTOT, VPOS and VNEG 24
Table 13 - Regression Statistics Between MOOD and CHANGE
Table 14 - Regression Model Between CHANGE and MOOD24
Table 15 - Polynomial Regression Statistics between MOOD and CHANGE 25
Table 16 - Polynomial Regression Model between MOOD and CHANGE 25

#### 1. Introduction

## 1.1. General Background

Fama (1970) developed the Efficient Market Hypothesis (EMH), which was considered a universal assumption in all financial fields by many researchers, in which stock prices can't be predicted because future prices are random, markets are efficient, and investors possess all the necessary information to make financial decisions. However, the EMH was not sufficient to explain the financial crisis, market collapses, and speculative bubbles that followed it (Sharma & Kumar, 2019). In the 1980s, behavioral finance began to emerge as a response to the limitations of modern financial models in explaining financial markets. Proponents of behavioral finance argue that the modern finance theory does not accurately represent real-world scenarios because individual behavior differs from rational theory, and individuals are subject to psychological biases (Le Bon, 1896).

Behavioral finance encompasses a variety of research topics, including biases, heuristics, investor psychology, market anomalies, decision-making behavior, prospect theory, herding, and investor sentiment (Barberis & Thaler, 2003). The concept of investor sentiment was first introduced in Graham & Dodd's (1934) book, which was published shortly after the 1929 market crash. They suggested that market prices could be influenced by investors' emotional behavior. However, it was not until the 1980s that investor sentiment gained more recognition and was studied more frequently by behavioral researchers who believed that investors were influenced by their emotions and sentiments in the decision-making process (De Bondt & Thaler, 1985; Shleifer & Summers, 1990).

Investor sentiment is a branch of behavioral finance that is constantly evolving with new authors and research gaps. Notable authors such as Barberis & Thaler (1985), and Shleifer & Summers (1990), mentioned in the previous paragraph, have proposed theoretical frameworks to understand better the role of sentiment in the markets and to relate sentiment to market anomalies and inefficiencies. More recent but equally notable authors include Malcolm Baker and Kent Daniel, who have published several papers on the role of investor sentiment in financial markets and have contributed to the evolution of this branch (Baker & Wurgler, 2006; Daniel et al., 2002). Nowadays, sentiment analysis continues to be a hot

topic in finance research, particularly with the rise of social media (Stieglitz & Dang-Xuan, 2013) and cryptocurrencies (Aste, 2019). Inside the social platforms, there are more relevance into some particular media like *Twitter*, for being a microblog with a low character limitation in which posts tend to be less ambiguous and subjective. These are relatively new areas of research that have expanded beyond the generic analysis of investor sentiment and stock returns, to include sentiment analysis over other platforms and its influence on more recent financial markets or its effects on corporate financial decisions (Gryglewicz, 2011) or risk management (C. Chen & Hafner, 2019).

EDP Renovíveis is a subsidiary of the EDP Group, created as a spin-off from that group in 2008, and operates in the field of renewable energies. It is the third-largest renewable energy company in the world and the second-largest wind energy operator, with operations in Europe, North America, and South America. Even though both enterprises operate independently from each other, the two companies are connected because EDP Group own a significant stake of EDPR and it is its main shareholder, having influence over EDPR strategic decision and operation, even sharing some resources, staff, and know-how. EDP Renováveis is a company that seems relevant for sentiment analysis because it belongs to a small and concentrated market, for being the second biggest Portuguese company by market capitalizations, and for being one of the biggest renewable energy company of the world (Yahoo! Finance, 2023).

## 1.2. Research Gaps

While the focus of investor sentiment research is primarily on the stock market, there is a lack of research on other asset classes such as bonds, commodities, utilities, or real estate. It remains unclear if these asset classes follow the same patterns as the stock market or if they have unique characteristics that alter the relationship between sentiment and the market (Huang et al., 2014). Additionally, most research examines the impact of sentiment on the stock market, but not how it influences corporate decision-making. This information is crucial for companies to understand better how to respond to particular investor sentiment (W. Chen, 2013). Finally, the majority of research has been conducted in the United States, overlooking the potential influence of cultural differences on how investors act based on their sentiments (Tran & Tran, 2023).

When studying the sentiment present in social media, there is no consensus in which is the most effective sentiment analysis method. Each method can produce different results, analysis, and accuracy levels, leading the results to be different from each other, even though the data is the same (Volkova et al., 2013). Moreover, a major challenge in social media sentiment analysis is identifying the polarity accurately. Analyzing tweets individually can be time-consuming and costly, so most studies rely on automated sentiment analysis tools. However, even with the use of artificial intelligence, there can be discrepancies in interpreting the sentiment of a tweet, and experts can disagree on whether a tweet has a buy or sell signal or a positive, negative, or neutral sentiment. Additionally, individuals themselves can be inconsistent with their own sentiments, further complicating the accuracy of sentiment analysis (Younis, 2015).

Studies on *Twitter* sentiment analysis for financial purposes rely on analyzing web text messages. However, despite the high quality of available databases, the correlation between the stock price time series and web time series is limited due to the complexity of their relationship (Ranco et al., 2015). Furthermore, the majority of studies in this field have focused on large markets, indexes, or well-known companies. The US market, including NASDAQ, DJIA, and S&P-500, has been the most extensively studied (Fudholi et al., 2022; Mendoza-Urdiales et al., 2022). However, there is a lack of research on smaller or less well-known markets. In addition, the research is limited to English tweets, ignoring non-English tweets which may contain valuable information about the sentiment as well (Barrière & Balahur, 2020).

### 1.3. Research Question

The primary objective of this dissertation is to examine whether *Twitter* sentiment can predict the *EDP Renováveis* stock returns by analyzing tweets from June 1st, 2021, to June 1st, 2022. This is a relevant question considering the previous research gaps for various reasons. First, even though the research is about a western European country, Portugal is a country with just a few research, leading to new discoveries in terms of cultural differences dealing with investor sentiment. Second, the polarity of the tweets will be both analyzed by a machine and a human, thanks to the few data available, leading to a better analysis of the

information and a more accurate study. Also, the relationship of the different data sets will be normalized to avoid the simple measuring correlation leading to more precise results. Moreover, it is important to have more studies about peripheral market like the Portuguese because most research are around main markets. Finally, the study about *EDPR* give us better comprehension of the effects of *Twitter* on renewable energy companies, as well as firms from small and concentrated markets.

## 1.4. Originality

Previous literature on forecasting stocks through *Twitter* sentiment has mostly focused on American indexes such as Bollen et al. (2011) and Mao et al. (2012), or large markets such as the American industrial industry (Teti et al., 2019). When considering only the renewable energy industry, almost every investigation has centered on companies or indexes from the United States (Herrera et al., 2022; Song et al., 2019). This paper is, to the best of my knowledge, the first to consider a renewable energy company from a small market like Portugal and the first to examine a small market with data from the actual war period, specifically in the context of the Ukraine-Russia conflict.

## 1.5. Contribution to Knowledge

This dissertation is another empirical study that offers important insights into the relationship between *Twitter* sentiment and stock returns in the Portuguese renewable energy market. By studying the effects of investor sentiment on stock returns using *Twitter* as a proxy, this research can provide a better understanding of how human behavior influences the stock market. Furthermore, this thesis can provide valuable insights into how not only *Twitter* sentiment, but also general sentiment, can affect the renewable energy sector, as well as the energy sector in general.

The findings of this study can be particularly valuable for CEOs, who can use the results to assess the impact of customer satisfaction on stock returns and consider using social media to keep shareholders informed. For investors, this work can provide guidance on how to create trading strategies based on the polarity of tweet sentiments. Finally, market regulators may benefit from this research by gaining a better understanding of the potential impact of *Twitter* on stock returns and developing new policies to prevent market bias.

## 1.6. Outline of the Following Chapters

The following of the paper is structured into four chapters. Chapter 2 provides a comprehensive review of the relevant literature in the field of sentiment analysis and behavioral finance. Chapter 3 outlines the methodology employed in this study. Chapter 4 presents the empirical results. Finally, chapter 5 concludes the study by summarizing the main findings.

## 2. Literature Review

#### 2.1. Definition of Sentiment in Finance

The term "sentiment" first emerged in finance in the 1980s when behavioralists began testing whether stock prices could be mispriced (Baker & Wurgler, 2007). These scientists examined words and sentences, revealing signs of sentiment within text.

Behavioral researchers have generally agreed on the definition of sentiment. For instance, Kaplanski & Levy (2010) describe sentiment as a misperception that can lead to the mispricing of an asset. D'Arms & Jacobson (2000) define sentiment as "any occurrent, object-directed, affect-laden mental state" (p. 3). They distinguish mood from sentiment because a mood is an affective state that lacks an object, whereas sentiment does not. Pang & Lee (2008) definition of sentiment also includes other related terms. For example, they describe opinion as a conclusive thought that can be debated and may differ among individuals. Meanwhile, sentiment, according to the authors, is a "settled opinion reflective of one's feelings" (p. 5).

However, the sentiment definition we have is still too vague. In order to achieve a more detailed classification, we turned to Kearney & Liu (2014) exhaustive study on sentiment in finance. They divide sentiment as a subject of behavioral finance research into two distinct types: investor sentiment, which refers to beliefs about future investments that are not justified, and text-based sentiment. Yu & Hatzivassiloglou (2003) defined textual sentiment as a written statement, whether direct or indirect, that conveys the Holder's opinion, whether positive, negative, or neutral, about the claim related to the topic, which aligns with Kearney and Liu's definition. Textual sentiment can be further divided into three segments: corporation-expressed sentiment, media-expressed sentiment, and internet-expressed sentiment. Corporation-expressed sentiment comes from official releases by insiders about the company, while media-expressed sentiment is the polarity expressed in news stories, commentaries, or analyst reports. Internet-expressed sentiment refers to feelings expressed in internet posts.

Every definition of sentiment described before aligns with other scholars' characterizations of sentiment as an opinion that carries either a positive or a negative connotation (Hu & Liu, 2004; Melville et al., 2009). Therefore, the general

definition of sentiment is easily and accurately defined by a wide range of behavioralists. Thus, as we delve more specific classifications, we begin to see variations from the general definition. For instance, investor sentiment and textual sentiment main difference is that while investor sentiment analyzes the subjective behavioral characteristics of investors, textual sentiment can also include the subjective behavior of investors as well as more objective reflections. Similarly, the distinction between corporation, media, and internet sentiment is based on the source and location of the sentiment.

In our study, even though Pang & Lee (2008) definition is precise, it is very generic, so we will adopt Kearney & Liu (2014) definition of sentiment as we are specifically investigating the textual internet-expressed sentiment present in tweets, which are analyzed from textual posts by various individuals on the internet. In addition, the data we analyze are less accurate, clear, formally written, and noisier which usually do not happen in corporate disclosures or media articles.

#### 2.2. Modern Finance Theories

### 2.2.1. Efficient Market Hypothesis

The Efficient Market Hypothesis (EMH) emerged in the 1960s, with Samuelson (1965) being one of the first to formalize the hypothesis, and Fama (1965) being the first to use the term "efficient markets." The martingale model and the random walk hypothesis were considered implications of the EMH, marking the beginning of the EMH literature (A. W. Lo, 2007).

Fama (1970) published the definitive Efficient Market Hypothesis. He defined an efficient market as "A market in which prices always 'fully reflect' available information" (p. 383). According to Lo (2005), the market incorporates all available information into the market price "fully, accurately, and instantaneously". This implies that when new information arrives, it is quickly incorporated into the stock market without delay.

Fama (1970) not only presented the Efficient Market Hypothesis but also proposed three different forms of market efficiency. The weak form of efficiency suggests that stock prices are unpredictably random, making technical analysis ineffective, all current information is reflected in stock prices, and past information has no relationship with current market prices. The semi-strong

form suggests that prices reflect both historical and publicly available information, meaning that an investor cannot outperform the market based on public information alone. Lastly, the strong form proposes that stock prices reflect all possible information, including insider information, making it impossible for investors to earn excess profits by trading on inside information.

#### 2.2.1.1. Efficient Market Hypothesis Critics

At the beginning of the 21st century, the Efficient Market Theory lost its universality as economists began to recognize the impact of psychological and behavioral factors on stock-price determination. They argued that past stock price patterns and valuation metrics could be used to predict stock prices. Furthermore, several events have occurred since the 1960s that challenge this theory, despite the considerable evidence supporting the Efficient Market Hypothesis. Therefore, there are a few but significant anomalies that are inconsistent with the EMH.

Table 1 shows that the anomalies in the EMH might be rare but significant. If the market behaved in the way Fama described, these anomalies would have faster corrections. However, most of these, like seasonal patterns, size effects and over and underreactions are spread out across markets, industries and even through time. Shiller (2000) observes that EMH illustrates the ideal world, but it does not describe the actual financial markets, has the anomalies do not corroborate their theory. These questions lead to the rising of Behavioral Finance.

#### **DESCRIPTION**

VALUE LINE ENIGMA	Low price-earnings stocks (value stocks) have higher returns than high price-earnings stocks (growth stocks) (Copeland & Mayers, 1982; Nicholson, 1960).
VOLATILITY OF ORANGE JUICE FUTURE PRICES	Majority of the produced frozen orange juice comes from Florida. The weather is a main factor for the orange juice crop, so anticipating the weather in central Florida can help us predict the price of orange juice (Roll, 1984).
MARKET CRASH 1987	"Black Monday", investors panic drove the market to a breakdown (French, 1988).
INTERNET BUBBLE 1990S	Irrational valuation of Internet and related high-tech companies (Shiller, 2000).
FINANCIAL CRISIS 2008	Predatory lending and taking excessive risk increased the systematic risk, leading to a burst in the United Stated market that spread globally (Chernomas & Hudson, 2017).
SIZE EFFECT	Small companies tend to outperform big companies in terms of returns (Banz, 1981; Keim, 1983; Roll, 1983)
SEASONAL PATTERNS	Patterns that occur regularly and are widely known, such as the "January Effect" (Haugen & Lakonishok, 1988), weekend seasonality (French, 1980), patterns around holidays (Ariel, 1990), and around the turn of the month (Lakonishok & Smidt, 1988).
OVER AND UNDERREACTION	Investors sell stocks that are losing value or buy stocks that are increasing in value, causing prices to exceed their actual market value or underreact to new information, especially with regard to future earnings announcements and short-term momentum (Lo & MacKinlay, 1999; Malkiel, 2003).

Table 1 – Anomalies in Efficient Market Hypothesis. Self-Made

#### 2.2.2. Behavioral Finance

The modern theory of finance, specifically the Efficient Market Hypothesis and its concept of the "random walk" and the rational investor, has been criticized for numerous failures. As a result, a new branch of finance has emerged: Behavioral Finance. This field originated in the 1980s through various academic journals, although its roots can be traced back much further. For instance, Selden (1912) applied the psychology of stockholders to the stock market, while le Bon (1896) analyzed how crowd characteristics can unconsciously influence individual decisions. The key difference between economics and psychology at the time was that economics treated human behavior as rational, while psychology considered

both rational and irrational aspects of behavior. They recognized that people have motivations and goals to achieve, and that they use reasons (either well or poorly) to respond to those motivations (Simon, 1986).

Behavioral Finance seeks to explain the emotional processes that occur in the decision-making process of individuals, groups, and entities. Humans tend to overestimate their own predictive abilities for success and often fail to learn from their past errors. Langer (1975) argues that decision-makers are irrational and overestimate their ability to control events, which explains why individuals often take higher risks. Behaviorists also contend that the decision-making process is influenced by psychological factors and emotions, which suggests that market behavior is not synonymous with the fundamental value of an asset (Peterson, 2016). Furthermore, prices do not always follow a random walk, and therefore can be predictable (Bollen, Mao, & Zeng, 2011), which contradicts the EMH. This led to the development of a theory that not only analyzes investor decisions but also the reasoning behind these decisions and what influences the investor's subjective representation of the decision model.

The key models of behavioral finance that explain the non-rational decision making of investors focus on biases and heuristics. These are mental shortcuts or rules of thumb that allow us to process information and make decisions quickly and efficiently (Busenitz & Barney, 1997). However, these biases and heuristics can also lead to errors in judgment and decision making. They are considered one of the reasons for market irrationality.

Cognitive biases can be divided into three categories. The first category relates to biases influenced by investor emotions, such as overconfidence, overreaction, confirmation bias (the tendency to justify previous ideas with preconceptions), hindsight bias (the belief that one predicted an event after it has occurred), and regret. The second category relates to the way in which investors interpret information differently depending on how it is presented, leading to different conclusions, also known as heuristics. This includes the Prospect Theory (Kahneman & Tversky, 1979), representativeness (making judgments based on how something matches our stereotype), anchoring (estimate based on an initial point of reference), framing (interpretation changes depending on how the problem is presented), availability (rely on easily recalled information when making decisions), and affect (when emotions influence the judgment). The final category of biases is influenced by interactions between investors and entities,

such as herding (the bandwagon effect), contagion effect, and informational cascade (an investor ignores private information and follows the common market belief).

#### 2.2.3. Sentiment Analysis

#### 2.2.3.1. Context

As behavioral finance gained popularity and the internet expanded, a new field of research emerged in the early 21st century known as sentiment analysis. This approach, also referred to as subjectivity analysis, opinion mining, and appraisal extraction (Pang & Lee, 2008), employs text and data mining tools, computational linguistics, and language processing techniques to examine how individuals perceive and evaluate data. Such data can consist of words, phrases, sentences, or even entire documents. The emergence can be attributed to the advancements in computational linguistics and natural language processing (NLP) techniques during the late 20th and early 21st centuries. However, analyzing and classifying opinions involves more than just determining polarity, but also determining whether the text is objective or subjective, who holds the opinion, what it is about, the position of the holder, the target of the opinion, and whether the author's intended opinion is accurately conveyed (Liu, 2006; Yu & Hatzivassiloglou, 2003).

The investor sentiment paradigm experienced a significant shift in the late 2000s and early 2010s, with the introduction of social media platforms such as *Twitter*, Facebook, and MySpace. Initially, sentiment analysis on social media had its limitations, such as manual monitoring of social media feeds, the use of basic text analysis tools, small datasets, and manual sentiment analysis techniques (Paltoglou & Thelwall, 2012). Sentiment analysis in finance was first employed to forecast stock prices (Tetlock, 2007). Currently, sentiment analysis is employed in diverse financial areas such as investment decision-making (Hasselgren et al., 2023), risk management (Cao & Chen, 2022), and fraud detection (Goel & Uzuner, 2016).

#### 2.2.3.2. Theoretical Analysis

To Russel (2003) emotions are constructed psychological experiences that arrive from the combination of cognitive appraisals, contextual factors, and conceptual knowledge to a non-specific feeling state that people experience. The polarity of the sentiment depends on the valence and the arousal form the specific emotion location in a two-dimensional circular space. High valence and arousal represent a positive sentiment, while a low valence and arousal are associated with negative sentiments. Additionally, the human mind associates a non-specific feeling with specific concepts or objects in the environment, leading her to associate a specific emotion to that concept over time.

Damásio (1995) neuroscientific case study suggest that sentiment is not separate from reason, guiding our decision-making process helping us make fast and intuitive judgments based on the emotional valence associated with certain choices or actions based on past experiences, helping us correct cognitive bias. Barrett (2017) goes further explaining the emotion is constructed by the brain combining sensory input, past experiences, and contextual information to generate emotional experiences. To her, the brain generates predictions about incoming sensory information based on past experiences and bodily sensations and updates them based on feedback received. Cultural and social contexts influence the way individuals interpret and experience emotions due to differences in conceptual frameworks, learned associations, and interpretation of bodily sensations.

The emotions can be influenced by a wide range of factors. One such factor is the occurrence of events that affect the economy or specific groups. The polarity of these events (whether positive or negative) can influence investor sentiment and their subsequent decisions, especially among individual investors who rely on publicly available and cost-effective information, unlike institutional investors (Tetlock, 2007). The quality and quantity of news reported can also influence the trading behavior of both individual and institutional investors. (Wu & Lin, 2017). In addition, market trends, such as bear and bull markets (Hanna et al., 2020) can also influence investor sentiment.

Another set of factors are past experiences like encounters, memories, and associated events, influencing the perception and response of an individual to various situations, cultural and social norms, values, beliefs, and practices which

provide framework for understanding and expressing emotions, as well as personality traits like neuroticism, extraversion, and agreeableness, which can affect the intensity, frequency, and duration of emotional experiences (Barrett, 2017). Moreover, assessing the personal significance, relevance, and meaning of a particular situation can influence the resulting emotional response. Social interaction, support, acceptance, and quality of relationships can also shape sentiments, as well as biological factors like genetics, hormones, and neurological processes. Besides, the physical environment such as the presence of nature, light and noise can evoke emotions from relaxation to stress. Additionally, media and social media can shape and influence attitudes, values, and emotional responses (Mesquita & Fridja, 1992).

Investors can also be influenced by biases and heuristics, which may lead them to conform to the actions of others, disregard information that contradicts their beliefs, or hold onto their beliefs even when presented with evidence to the contrary. Confirmation bias can occur as users engage with others who share their views and ignore those who contradict them affecting sentiment by leading individuals to selectively focus on and interpret information that aligns with their existing sentiments (Nickerson, 1998). Additionally, the availability heuristic, which is the tendency to rely on easily accessible information, can lead to some ideas being over-represented, relying on their readily available emotional experiences or vivid memories to assess the intensity or prevalence of a particular sentiment (Tversky & Kahneman, 1973). In-group bias can also influence sentiment, as individuals tend to prefer opinions from those in their own community, leading to a more positive sentiment towards insiders and a negative one towards outsiders.

Furthermore, people may feel pressured to follow the opinions of their community or the people they follow, rather than considering different points of view (Maitland & Sammartino, 2015). The affect heuristic, where opinions are formulated based on emotions rather than facts, can also influence sentiment analysis, especially in strongly worded posts, biasing individuals to perceive and express sentiments that align with their current emotional state. Moreover, clickbait titles can persuade individuals, and anchoring can lead people to form their opinion based on the first headline or tweet they see. Also, herding behavior of the investor follow the crowd and investor overconfidence where they believe

they are more skilled than what they are can influence investor sentiment and the decision-making process (Gigerenzer, 1991).

Likewise, emotional contagion, where individual's emotions are influenced by others' emotions, can influence the sentiment the individual adopt for being exposed to the emotion of others (Barsade, 2002). Implicit biases, which are unconscious attitudes, directly affect sentiment by automatically shape emotion responses that may be inconsistent with an individual's consciously held beliefs or values (Greenwald et al., 1998). Stereotyping can also shape sentiments towards individuals or groups based on preexisting biases, formatting a positive or negative emotional reaction (Devine, 1989), and representative heuristics where people judge an event by a familiar prototype can create certain sentiments to express their sentiments or interpret the sentiments of others (Kahneman & Tversky, 1979). Finally, the way information is framed, the framing heuristic, can influence judgments and decisions, shaping how individuals perceive their emotional experiences, leading to different sentiments based on the framing context (Tversky & Kahneman, 1981).

#### 2.2.3.3. Impact on Financial Markets

Sentiment analysis has a significant impact on trading strategies as it provides investors with valuable insights into the current mood of the market. Sentiment analysis can also serve as a useful tool for developing trading strategies. For instance, investors can choose to follow the trend based on the polarity of the sentiment, either by following the crowd or going against the trend by buying or short selling stocks, respectively (Hurst et al., 2017). Additionally, contrarian investing, which involves buying past losers and selling past winners or buying after prices go down and selling after prices go up, is another popular strategy based on sentiment analysis. This strategy is based on the belief that sentiment will eventually shift, enabling investors to avoid market corrections or take advantage of rebounding stock prices (Kim & Park, 2015).

Investor sentiment serves as a tool for risk management as it provides valuable insights about target organizations or executives, which can help industries manage risk and make strategic or long-term decisions. Aggressive customers, advocacy groups, and company insiders can create risks related to business ethics and corporate social responsibility. Sentiment analysis can assist in mitigating

these risks by organizing data and understanding public perception, vulnerabilities, and how to adjust security postures (Kunze et al., 2020). By using sentiment analysis, industry leaders can stay ahead of potential issues and make more informed decisions to avoid negative impacts on their organizations.

Inside social media and *Twitter* itself, sentiment expressed in posts can help individuals to shape their overall perception of the market. The sentiment indicators derived from the social media data can be used as a tool for developing trading strategies (Bollen, Mao, & Zeng, 2011). Moreover, positive sentiment might increase market optimism leading the prices do rise and creating a bullish market sentiment. On the other hand, negative sentiment might have the opposite effect, causing market pessimism, driving prices down, making a selling pressure on investors and creating a bearish market sentiment. So, social media sentiment can influence market volatility as well as market immediate reaction, amplifying the impact of certain events or news on stock prices (Haritha & Rishad, 2020). Besides, monitoring sentiment trends in social media can assist in risk management strategies by identifying potential risks or opportunities (Demek et al., 2018).

## 2.3. Empirical Evidence of Sentiment Analysis on Stock Returns

There is evidence of *Twitter* sentiment polarity affecting a variety of markets, such as the cryptocurrency market, the commodity market, and the stock market, which is the focus of this dissertation. In the Table 2 presented below, there is stated several studies' information, like the authors, the market / industry and country of analysis, utilized method and results. Overall, there is a wide variety of empirical methods used, mixed markets, industries, and results. However, the chosen industry is usually from a big country like the USA or an important market or index like the DAX Index or the Tunidex, some of the most influent Indexes inside the European Union and Africa.

A small number of studies have found no relationship between social media sentiment and stock returns. Nofer & Hinz (2015) found no significant relationship using an OLS regression between Social Mood Index and share returns in the following 4 trading days after an event in Germany, explaining it because investors incorporate the mood level in their models rapidly. When changing the SMI with a Weighted Social Mood Index only volatility have a

significant relationship with the mood. Reboredo & Ugolini (2018) discovered that *Twitter* sentiment has no useful information in forecasting prices, volatility, or trading volumes for renewable energy companies using a VAR model. Hamraoui & Boubaker (2022) found that sentiment polarity is not useful for predicting stock performance, because four companies pass the Granger causality test, two financial companies, one electronic and one food chain company. However, the number of tweets have great significance in price volatility. Nisar & Yeung (2018), using a multiple regression model, found no relationship between the number and polarity of tweets, the closing prices, the daily returns, and the FTSE Index. They found that the polarity of tweets about the major English stocks don't have influence on predicting the daily returns, the multiple regression model is not a good fit for the hypothesis, and that the independent variables of polarity and mood are not appropriate to predict stock returns.

In these four papers, only one of them analyzes the USA market, all the other examine smaller markets (Germany, England, and Tunisia). However, the one that explores the USA market does not investigate the companies from the big indexes like Nasdaq, DIJA or S&P, but a specific industry, which is the renewable energy. The companies that the Index contain are not big enterprises like the other papers who found a positive relationship between the public mood and returns analyze. There is a pattern in these papers of considering relatively smaller companies or indexes than the ones who have different results. Also, most these studies focus on long-term effects, like Nofer & Hinz (2015), where he did not found relationship in the following 4 days after event, as other factors play a more significant role than the *Twitter* sentiment. Also, the methodology of these studies tend to be less complex, like OLS regression (Nisar & Yeung, 2018; Nofer & Hinz, 2015) or VAR models (Hamraoui & Boubaker, 2022; Reboredo & Ugolini, 2018), but the sample data and the sentiment analysis techniques tend to be more refined, suggesting that the Twitter sentiment may be noisier and unreliable, which leads to challenges in identifying patterns.

In contrast, most studies have found a significant relationship between social media and stock returns. Bollen, Mao, & Zeng (2011) created a public mood time series that has shown predictive power on the DJIA closing values, but not because they analyzed the general positive vs. negative sentiment of the tweets, but instead the calmness of the public mood. Oliveira et al. (2017) also found

predictive power on returns of the S&P 500 Index by analyzing *Twitter* mood, plus that microblogging sentiment is useful for predicting the returns of lower market capitalization portfolios, and the returns of some industries like the energy and high technology ones.

The methodological approach, the industry analyzed, the specific country, as well as other variables, can influence the results. For example, Ranco et al. (2015) found significant evidence of relationship between stock returns and Twitter sentiment while analyzing DIJA using a similar VAR model used by Reboredo & Ugolini but a different index. Similarly, Nofer & Hinz (2015) found no relationship while using an OLS model but when chasing to a VAR model found that polarity is significant in predicting market statistics. Another example is Ho et al. (2017) paper that justify their findings of extreme values having a crucial role in forecasting stock returns because of the chosen DLM-SUR modeling approach. Both Sprenger & Welpe (2010) and Sul et al. (2017) have similar findings in why there is a relationship between the sentiment and returns. While Sprenger & Welpe (2010) concludes that there is important information on *Twitter* that is still not incorporated in the stock market, Sul et al. (2017) complete this conclusion by finding that users with less retweets and followers have the strongest effects on future returns, because the information they provide about the stock is acquired by fewer people than the users with lots of followers and retweets. Gu & Kurov (2020) corroborates these papers by concluding that Twitter sentiment has a strongest predictive power on the firms with less analyst's coverage. All these papers analyze the USA largest listed companies and use a Panel Data regression approach, which may justify why they found the same results.

Every paper analyzed in the table below which gave us a significant relationship between *Twitter* sentiment and stock returns have three main similarities. First, the data sample, where every company is from USA and all of them are from very well-known and non-concentrated indexes, being the companies' large ones. Secondly, almost every study, except for (Sul et al., 2017), examine the immediate or next-day relationship, the short-term effect. Finally, these studies analyze most of the times the effects of *Twitter* sentiment during specific events, like earning announcements or product launches, investigating the sentiment surrounding these events (Bollen, Mao, & Zeng, 2011).

Despite the mixed and often complex findings of these studies, it is clear that sentiment analysis can provide valuable insights into stock market dynamics. The results suggest that social media sentiment can have predictive power for stock returns in certain contexts, making it a useful tool for predicting them.

AUTHORS	Market/Industry	Country	Method	Results
NOFER & HINZ (2015)	Dax Index (30 major German Companies)	Germany	OLS (Ordinary Least Squares) Regression	• No significant relationship found between Social Mood Index and share returns in the following 4 days after an event.
				DAX intraday returns are positively influenced by increases in Follower- Weighted Social Mood Index
REBOREDO & UGOLINI (2018)	WilderHill Clean Energy Index (Renewable Energy companies)	USA	VAR Model	• <i>Twitter</i> sentiment has no useful information in forecasting prices, volatility, or trading volumes for renewable energy companies.
HAMRAOUI &	Tunindex (22 companies,	Tunisia	VAR Model	Sentiment polarity is not useful for predicting stock performance.
BOUBAKER (2022)	mainly from the Tunisian financial sector)			Number of tweets is useful in predicting price volatility
BOLLEN, MAO, & ZENG (2011)	DJIA (30 large publicly owned companies listed on NYSE and NASDAQ)	USA	Time series	Public mood time series has predictive power on the DJIA closing values.
HO ET AL. (2017)	DJIA (30 large publicly	USA	DLM Time series and	Extreme values in the data are crucial for forecasting stock returns.
	owned companies listed on NYSE and NASDAQ)		SUR Model	<ul> <li>Relationship between social media sentiment and stock returns is time- varying and cross-correlated.</li> </ul>
SOUZA ET AL. (2015)	5 big US companies	USA	Auto-regressive Model	Twitter moves in the market in respect to the excess log-returns of the stocks
SPRENGER & WELPE	S&P 100 (largest 100	USA	Panel regression and	Microblogs contain information not yet incorporated in the stock prices.
(2010)	companies listed in the U.S. stock exchanges)		Time-sequencing regressions	Increased bullishness of microblogs is associated with high returns.
				Message volume explains trading volume, but not returns or volatility.
RANCO ET AL. (2015)	DJIA (30 large publicly owned companies listed on NYSE and NASDAQ)	USA	VAR Model	• Significant evidence of relationship between stock returns and <i>Twitter</i> sentiment.

RAO & SRIVASTAVA (2012)	DJIA, NASDAQ-100 and 13 other tech U.S. companies	USA	OLS (Ordinary Least Squares) Regression	•	Public mood carries strong cause-effect relationship with stock returns.
SUL ET AL. (2017)	S&P 500 (largest 500 companies listed in the U.S. stock exchanges)	USA	Panel Data Regression	•	Users with less followers are significant in predicting stock returns in one, ten and twenty trading days.  Users with less followers and with no retweets has the strongest effects in future returns.
OLIVEIRA ET AL. (2017)	DJIA, S&P 500, Russel 2000, NASDAQ 100	USA	Multiple Regression, Neutral Network, Support Vector Machine, Random Forest, Ensemble Averaging	•	Microblogging data is useful to predict stock market variables like stock returns, volatility, trading volume and indexes.
BEHRENDT & SCHMIDT (2018)	DJIA (30 large publicly owned companies listed on NYSE and NASDAQ)	USA	Panel HAR Model	•	Twitter sentiment have statistically significant feedback effects of return volatility as well as Twitter count and vice versa.
GU & KUROV (2020)	Russell 3000 (3000 Largest traded companies in the U.S.)	USA	Cross-sectional Panel Regression	•	Daily <i>Twitter</i> sentiment contains information that is useful for predicting next day stock returns.
RUAN ET AL. (2018)	Firms with largest number of tweets	USA	Linear Regression	•	Users' sentiment helps making better predictions of abnormal stock returns

Table 2 - Empirical Evidence on the Relationship Between Social Media Sentiment and Stock Market Variables. Self-Made (Continuation)

## 2.4. Context

### 2.4.1. Twitter

Twitter was launched in October 2006 as an SMS-based communication platform that allows users to post short messages about a wide range of topics. Its most popular uses include sharing news and current events in real-time, expressing personal thoughts and opinions, promoting and marketing, sharing educational and informative content, providing entertainment and humor, networking, and collaborating with others (Bartov et al., 2018).

Twitter has experienced tremendous growth over the years and has become one of the most popular social media platforms worldwide. As of April 2023, the platform has approximately 373 million monthly active users, making it the 15th biggest social media platform globally. The platform's users tend to be younger, highly educated, with higher incomes, interested in news and current events, and comfortable with technology and social media, in comparison to other social media platforms (Datareportal, 2023b). In addition, there are a wide range of types of active traders using *Twitter*, from retail to institutional investors, financial professionals, news organizations and journalists.

Twitter has a strong relationship with finance has it becomes a noticeable platform for financial information, news, updates, and analysis dissemination. Investors can acquire real-time information on the stock market, as it has been a channel where it is possible to provide rapidly real-time financial news and updates. This includes monitoring trends and breaking news, following analysts, news outlets, industry experts, and thought leaders to remain up to date on upcoming events and announcements, tracking sentiment and buzz around specific companies or stocks to gain insights on investment directions, engaging with other investors and traders to share ideas, as well as interacting with companies directly. Twitter financial data can also be used for sentiment analysis for a better investment decision-making process, for market influence and herding by influential financial figures, for customer service, support and engagement, for market transparency and democratization, and for networking and collaboration (Twitter, 2023). Up to date, there is no specific regulations pertaining to Twitter as a platform, even though the content shared may fall

under specific jurisdiction depending on the nature of the information being shared. *Twitter* is the third most used social media by financial advisors for business purposes, being used by 52% of them, staying behind LinkedIn (72%) and Facebook (62%) (Putnam Investments, 2023). Each social media is preferred for different reasons, and *Twitter* business development initiatives like thought leadership, as well as the short posts with limited characters and the large and diverse user base make it a good platform for financial use.

### 2.4.2. Macroeconomic Context between 2021-2022

By the end of 2021, inflation had surpassed expectations, primarily due to the significant increase in fossil fuel prices, almost doubling energy costs. In addition, food prices had risen. The Covid-19 pandemic had caused disruptions in production, leading to global supply distribution problems, which, coupled with high demand for goods, clogged ports, and landside constraints, contributed to the rise in inflation (International Monetary Fund, 2022b)

On the morning of February 24, 2022, the conflict between Ukraine and Russia began, which is considered the most significant armed conflict since World War II. As a result of the conflict, there were negative effects on energy and food supplies, as well as on economic indicators (Polyzos, 2022).

In the first half of 2022, the GDP experienced a contraction for the first time since 2020. In European countries, inflation continued to rise, triggering global financial conditions. Consumer prices rose faster than expected, with the euro area experiencing an 8.6% increase (the highest since the creation of the monetary union). Wage growth did not keep up with higher food and energy prices, resulting in a decline in household purchasing power. Central banks began to raise interest rates, particularly in developing economies, leading to a rise in borrowing costs and a decline in equity prices (International Monetary Fund, 2022a).

## 2.4.3. Portugal

Twitter ranks as the 8th most popular social media platform in Portugal, with 1.9 million users. The platform is commonly used by individuals, celebrities, companies, media outlets, influencers, activists, and non-profit organizations.

Additionally, traders, financial institutions, and stakeholders in Portugal use *Twitter* to stay updated with the latest news and developments in the stock market (Datareportal, 2023a).

There is a strong rising of individual investment in Portugal, particularly among middle-class families, as a means of building wealth. This trend has been reinforced over time by the promotion of stock market investment by banks and financial advisors, as well as the availability of more educational resources among younger generations. Additionally, technological advancements such as the rise of online trading platforms and robo-advisors have made it easier for individual investors to enter the stock market. However, financial literacy levels in Portugal are lower than the OECD average among adults aged 18-64, despite recent improvements (*CMVM*, 2023a).

The Lisbon Euronext is the Portuguese stock exchange market. The Portuguese market is small and concentrated by just a few companies, which have the majority of the trading volumes and capitalization. Also, this bigger companies have less liquidity risk. However, this risk increases when we are getting closer to the end of trading day (Pereira da Silva, 2016). The Portuguese Market capitalization between 2021 and 2022 decreased from 37,867% of GDP to 34,346% of GDP (CEIC DATA, 2023b), having its lower value of about 86k USD bn in June 2021 and its highest value of about 97k USD bn in October 2021 has it is showed in graph 2 (CEIC DATA, 2023a).



Graph 1 - Euronext Lisbon Market Capitalizations. Source: CEIC DATA (2023a)

The PSI-20 is the top stock market index in Portugal and is part of the Euronext Lisbon stock exchange. It monitors the performance of the most frequently traded listed stocks, covering a wide range of sectors. This index used to analyze the top 20 Portuguese companies, but nowadays only 18 companies are considered, showing the high concentration of the market. Among the companies included in this index are *EDP* and *EDP Renováveis*, which are, in 2023, the most valuable Portuguese companies by market capitalizations as it is showed in the table 3 below.

FIRMS	SECTOR	MARKET CAPITALIZATION (BILLION USD)
EDP RENOVÁVEIS, S.A.	Energy	21,8
EDP – ENERGIAS DE PORTUGAL, S.A.	Energy	20,8
JERÓNIMO MARTINS, SGPS, S.A.	Consumer Retail	13,8
GALP ENERGIA, SGPS, S.A.	Energy	8,8
BANCO COMERCIAL PORTUGUÊS, S.A.	Financial	3
THE NAVIGATOR COMPANY	Paper and Forest Production	2,4
NOS, SGPS, S.A.	Telecommunications	2,3
SONAE – SGPS, SA	Consumer Retail, Telecommunications, etc	2
REN – REDES ENERGÉTICAS NACIONAIS, SGPS, S.A.	Energy	1,9
CORTICEIRA AMORIM, SGPS, S.A.	Paper and Forest Production	1,4
SEMAPA	Industrial	1,1
ALTRI, SGPS, S.A.	Paper and Forest Production	1

Table 3 - Biggest Companies me Portugal by Market Capitalization. Source: (Papadopoulos, 2023)

The Portuguese securities market regulatory authority is *CMVM*, who is responsible for supervising and regulating, ensure fair and transparency, protect and promote investors' confidence, authorize, and oversight entities operations, maintain the integrity and efficiency of the market, cooperate with other regulatory authorities, and coordinate the supervision of the security market. Some of the key regulations of CVM are the regulation on listed companies to disclose all the information to the market, regulate and prevent market abuse, like manipulation and inside trading, regulation on public offers, listing, investment funds, corporate governance, take overs, mergers, and market participants (*CMVM*, 2023b).

### 2.4.4. EDP and EDPR

EDP is a publicly traded utility company in Portugal that is involved in the generation, distribution, and sale of electricity and gas. It is an important driver of economic growth and job creation in Portugal and in the other countries it operates in. On the other hand, EDP Renováveis is a subsidiary of EDP that specializes in developing, constructing, and operating wind, solar, and hydro renewable energy projects around the world. It is one of the biggest renewable energy companies worldwide, and it is considered the major wind energy producer of the world, also having an exponential growth in the solar energy production. Portugal ranks as the 11th most dependent on imported energy EU member mainly because of the lack of fossil energy sources. However, the Portuguese renewable energy sector continues to be a global leader (International Trade Administration, 2023).

Investors in *EDP Renováveis* tend to have a long-term investment horizon and a global perspective, as the company operates in various countries. They are interested in sustainability and are attracted by the growth potential of the company. These investors typically have diversified portfolios and are willing to take risks, as renewable energy stocks are influenced by a wide range of factors such as specific country legislation (e.g. windfall tax), global legislation (e.g. embargoes), policy uncertainty, development of new technologies, and availability of capital-intensive renewable energy projects. Therefore, investing in *EDP Renováveis* requires a certain level of risk tolerance (Reboredo & Ugolini,

2018). Only 4% of *EDPR* stakeholders are individual investors (*EDP Renováveis*, 2023a).

EDP Renováveis is influenced by various specific regulations and legislation established by the Portuguese government. The RTSE which set out the tariff regulation for the energetic Portuguese market, the RCC which defines the commercial relationship between electricity market participants, and DL 34/2011 which regulates the renewable energy sector in Portugal by promoting renewable sources, through the feed-in tariff system and power purchase agreements are the main regulations EDPR have to deal with by being established in the renewable energy market. The Portuguese government and the European Union offer grants, subsidies, and tax incentives opportunities for renewable energy projects in which EDPR may benefit from, as well as research and development fundings. Also, the Portuguese government set renewable energy targets and standards, requiring a certain percentage of energy consumed or generated to be from renewable sources creating opportunities for *EDPR* to act. Similarly, the EU has an Emission Trading System where limits the gas emissions for certain industries in which EDPR can benefit indirectly because it is not subject to the same restrictions as fossil power companies.

Analyzing the financial performance of the company in the last 10 years, EDPR is a very solid firm. In terms of profit, the net income of the company has grown exponentially yearly (Graph 7 of the Appendix). The Return on Assets and the Return on Equity are also improving over the years, with an average of 0,017754123 and 0,04048048, respectively, on the last 10 years, suggesting the company generates a modest level of profitability relative to its assets and equity (Graph 8 of the Appendix). In terms of financial autonomy, the company Debtto-Equity and Equity Ratio indicate the company has potentially a higher level of financial risk for relying heavily on debt financing to fund its operations and growth (Graph 9 of the Appendix), justifying why the investors tend to be longterm investors. EDPR also pays consistently dividends, increasing them over the years as showed by Graph 10 of the Appendix where the dividend yield keeps growing over the last 10 years, with a 6% average. These dividend yields can be attractive to income-oriented investors seeking regular income from their investments, being higher than the average yield of the renewable energy industry (EDP Renováveis, 2023b).

## 2.5. Hypothesis

 $H_1$ : EDP Renováveis stock returns can't be predicted by analyzing Twitter Sentiment between the  $1^{st}$  of June 2021 and the  $1^{st}$  of June 2022.

Emotions are constructed psychological experiences that arrive from the combination of cognitive appraisals, contextual factors, and conceptual knowledge to a non-specific feeling state that people experience (Russel, 2003). Investors have motivations and goals which they want to achieve. To reach their objective they use reasons to respond to those motivations, but these reasons might not always be rational (Simon, 1986). Therefore, investors, in their decision-making process, tend to have emotional processes, making them irrational, to fail to learn from their past errors, and to overestimate their own abilities for success and to control events, which explains why individuals often take higher risks while investing (Langer, 1975). This might justify why investors invest in *EDP Renováveis* stock instead of other companies, because *EDPR* is a renewable energy company influenced by factors like country and global legislation, policy uncertainty development of new technologies, and availability of capital-intensive renewable energy projects, which might require a certain level of risk tolerance (Reboredo & Ugolini, 2018).

Twitter is used by almost 20% of the Portuguese population, its only available for people over 13 years old, and is used mainly by people with higher education level and higher financial literacy (Datareportal, 2023a) which makes it the best social media to extract the user's sentiment. Also, *Twitter* has a wide range of active financial players in Portugal, who use it as a source of knowledge to the decision-making process, making it a reliable source of information to post and analyze investors opinions and insights (Bartov et al., 2018).

Analyzing the empirical evidence around renewable energy companies, there is no significance found between stock returns and *Twitter* mood. Reboredo & Ugolini (2018) found no correlation between an index constituted only by renewable energy companies just like *EDPR*. Hamraoui & Boubaker (2022), in their study, only found relationships between financial, food supply and electronic companies, founding no relationship between *Twitter* sentiment and stock returns of energetic enterprises. In addition, Nofer & Hinz (2015) found no significant relationship between those variables while studying a concentrated

and small market like the Portuguese market in which *EDPR* is inserted. Besides, *EDPR* is a stable company, in which the investors tend to have a long-term investment horizon, not being easily influenced by situational changes, rumors, or news. Thus, *EDPR* is a company with a structured dividend payment and with a dividend yield higher than the average of the renewable energy industry, suggesting it is a relatively safe company to invest, and being an attractive firm for investors seeking income from their investments in a longer-term horizon than the investors who seek fast income, being those investors the most influenced by *Twitter* posts (*EDP Renováveis*, 2023b). Also, *EDPR* percentage of individual investor's share is only 4%, and institutional investors have the remaining 96%. Individual investors, who rely on publicly available and cost-effective information, like the *Twitter* information, do not have much impact on the stock prices changes for having such a small share (Tetlock, 2007). On the other hand, institutional players are not usually influenced by bias and heuristics, meaning they do not usually follow *Twitter* sentiment.

# 3. Data and Methodology

## 3.1. Method

This thesis employs a quantitative approach instead of a qualitative approach. This method can collect data from large samples, allowing for generalizations of findings, making it more time and cost-efficient than the other approach because the data is analyzed quickly and with fewer resources (Creswell, 2014). *Twitter* analysis generates a vast volume of data that would be challenging to analyze manually every tweet, making sentiment analysis software an efficient tool. Additionally, in a fast-moving social media platform like *Twitter*, speed is crucial when analyzing data. Therefore, quantitative methods can analyze volumes of data more quickly than qualitative methods. In conclusion, a quantitative approach provides numerical measures that can be helpful in comparing tweet polarities over time and across different groups, aiding in the search for trends and patterns (Wolstenholme, 1999).

### 3.2. Data Collection

In this thesis, secondary data, which refers to data collected by other sources instead of directly from the primary source, is utilized for various reasons. Firstly, secondary data on *Twitter* sentiment is readily available and provides a wealth of diverse opinions and viewpoints. Secondly, collecting secondary data is cost-effective compared to collecting primary data. Additionally, secondary data is reliable and consistent, and it can be compared over time and space to uncover trends and patterns (Johnston, 2014).

Daily time series data was collected from both *Twitter* and the market for the period spanning from June 1, 2021, to June 1, 2022. *Twitter* data was manually collected using the *Twitter* Search API, with a search query focused on the *EDP Renováveis* cash-tag (\$*EDPR* and \$*EDPR*.LS). As noted by Zhang et al. (2011), both original tweets and retweets were included in the data collection process, as retweets can be indicative of a post's popularity and influence. The collected tweets were then filtered and cleaned to remove spam, irrelevant content, and posts from fake accounts. Additionally, URLs, cash-tags, user mentions, and repeated letters were removed from the tweets (Rane & Kumar, 2018). Finally,

the sentiment polarity of each tweet was determined, and a time series was created to track changes in sentiment over time.

The market data was collected from Yahoo! Finance on *EDP Renováveis* stock, which included the volume of stocks traded per day, opening price, and closing price, and PSI-20 Index closing prices. The objective was to extract the time series of daily returns. Yahoo! Finance is a widely used and reputable source of financial data (Awan et al., 2021; Batra & Daudpota, 2018). Although it is generally accurate, some stocks may have a few data errors. However, market data is not available on weekends or holidays when the market is closed. To address this, I applied the concave function proposed by Mittal & Goel (2009). This function estimates the given day with no information as the average of the last day with available data and the next available day with data. While *Twitter* data was available for all days, this method was necessary to fill in the missing market data.

Time series was chosen over panel data mainly because the hypothesis only analyzes one firm. With a Time Series approach, we can analyze trends, seasonality, autocorrelation and other time-dependent patters specific from *EDPR*, because this method allows to capture and model the temporal dynamics and patters within the firm's data

## 3.3. Variables

The multiple regression used to obtain the results included the following independent variables related to *EDPR*: daily volume of tweets (VTOT), daily volume of positive tweets (VPOS), daily volume of negative tweets (VNEG), daily average mood (MOOD) and weighted social mood index (WSMI). The dependent variables were the daily returns (CHANGES) of *EDPR* stocks.

In order to examine the correlation between CLOSE, VTWEETS, VPOS, and VNEG, I applied a Z-score standardization process to normalize the variables. This was done to enable comparisons between the variables despite having different units of measurement, as recommended by authors such as Bollen, Mao, & Zeng (2011) and Jain et al. (2005). The formula for the z-score is:

$$Z = \frac{(X+\mu)}{\sigma},$$

Z is the z-score, X is the original value of the variable,  $\mu$  is the mean of the variable, and  $\sigma$  is the standard deviation of the variable.

While there are other normalization methods such as Min-Max Normalization and Decimal Scale Normalization (Henderi et al., 2021; Saranya & Manikandan, 2013), these methods may distort the distribution of the data if there are extreme outliers. In contrast, the z-score normalization maintains the distributional characteristics of the data, while also removing the influence of extreme values. This allows for a more accurate comparison and interpretation of the data.

## 3.3.1. Dependent Variable

The dependent variable in the regressions that will be made is the daily returns. The daily returns represent the percentage change in the value of the asset over a single trading day. In order to calculate daily returns, I utilized the logarithmic return method (Andersen et al., 2003), using the following equation:

$$CHANGE_t = \ln\left(\frac{CLOSE_t}{CLOSE_{t-1}}\right)$$

While there are other methods for calculating daily returns, such as simple daily returns or weighted daily returns (Busse, 1999), the logarithmic return method tends to be more symmetrical when dealing with stock returns, which are often skewed in their distribution. This provides a more realistic representation of volatile investments, such as stock performance over time (Hudson & Gregoriou, 2015).

## 3.3.2. Independent Variable

The independent variables are VTOT, VPOS, VNEG and MOOD. VTOT, VPOS and VNEG are, respectively the total daily tweets made about *EDPR*, the total daily tweets that have a positive polarity, and the total daily tweets with a negative polarity. VTOT is used because it can provide a hint of the overall activity surrounding *EDPR*, while VPOS and VNEG can reflect the emotional tone of discussion around the company, as well as provide insights about the market sentiment (Nisar & Yeung, 2018).

The daily average mood is the share of positive mood on all word occurrences. To determinate the daily average mood, I will utilize a sentiment score based on the formula proposed by Nisar & Yeung (2018) where the only assumption is that neutral sentiment tweets are considered as 0 so may be ignored when calculating the average, even though this assumption of having no implication is not always true:

$$MOOD_t = rac{VPOS_t - VNEG_t}{VTOT_t}$$
 , where t is a given t day.

There are other approaches like the one Nofer & Hinz (2015) had, where they don't consider the total tweets but only does which sentiment is positive, grief, hopelessness, tiredness, or anger. However, to scrutinize in more depth the tweeter sentiment it can cause more difficulties for the software to analyze which mood the tweets represent, because this categorization is much more subjective and can lead to divergent analysis. The approach of the Weighted Social Mood Index, an extension to the daily mood average, proposed by Nofer & Hinz (2015) was tested but, due to the low quantity of tweets posted about *EDPR*, the results of the WSMI were the same as the daily mood average.

# 3.4. Main Methodology

The main methodology used in this thesis is a Time Series Regression.

I will apply an OLS multiple regression model that includes CHANGE, VTOT, VPOS and VNEG variables and OLS simple regression model between CHANGE and MOOD.

$$CHANGE = \alpha + \beta_1 VTOT + \beta_2 VPOS + \beta_3 VNEG + \varepsilon_t$$
$$CHANGE = \alpha + \beta_1 MOOD + \varepsilon_t$$

The purpose of using as the regression technique an OLS multiple regression is to establish the relationship between the independent variables and the dependent variable. This model enables us to estimate the magnitude and direction of the association between the variables, while accounting for the influence of other variables that may also affect the dependent variable. By doing

so, we can identify which variables significantly impact stock returns, to what degree, and gain a comprehensive understanding of the link between *Twitter* sentiment and stock returns (Brooks, 2014).

Alternative approaches, such as Granger Causality analysis, could be used instead of regression, as Ranco et al. (2015) did. However, an OLS multiple regression offers advantages as it allows for the examination of several independent variables simultaneously, capturing the potential effects of other factors that may be influencing stock prices. Additionally, the r coefficient helps to identify the relative significance of each independent variable in predicting daily stock returns (Dumouchel & Duncan, 1983). Moreover, the OLS multiple regression results are simple and more straight-forward to interpret and compare. Also, an OLS regression allows for the inclusion of multiple independent variables, enabling the analysis of the combined effects of different factors on stock returns and provides standard errors, p-values, and confidence intervals for the estimated coefficients, enabling statistical inference (Fumio Hayashi, 2000).

However, OLS regression relies on some assumptions, like linearity, independence of errors, homoscedasticity (constant variance), and absence of multicollinearity, in which if these assumptions are violated it can lead to biased or inefficient coefficient estimates and misleading conclusions. Therefore, to avoid reaching distorted results, in case some assumptions are violated there will be proceeded some calibration methods. If the linearity is violated, a polynomial regression will be used as calibration method because it allows for fitting a curve to the data by adding polynomial terms to the regression equation (James et al., 2021a). On the other hand, if there is multicollinearity between variables, a ridge regression will be used as a regularization technique to help mitigate overfitting and stabilize the coefficient estimates by introducing a penalty term to the regression equation (James et al., 2021b).

### 3.5. Software

Two different software programs will be used to analyze the data. For analyzing *Twitter* sentiment, an online tool named Umigon will be employed. Using a lexicon-based sentiment classifier software is preferable over manually examining each tweet, not only due to time and cost savings but also to avoid the

potential for biases affecting data interpretation. However, Umigon may have some imprecision in detecting negative sentiment, thus, after the initial software analysis, a manual review will be conducted to ensure accuracy (Ribeiro et al., 2016). R software was chosen as the statistical software for its ability to handle the data efficiently and for allowing flexible and customizable regression modeling options. While other programs such as Stata or Gretl are available, R is a powerful tool that can yield similar results (Mizumoto & Plonsky, 2016).

# 4. Empirical Results

## 4.1. Descriptive Statistic and Data Characteristics

When observing the *Twitter* data, we can notice that only 101 tweets of 211 tweets gathered by the *Twitter* Search API were considered suitable for analysis. This happens because 110 tweets were considered as Spam or Irrelevant, mainly because of one *Twitter* user who was a trader bot who posted the same information more than one time. The monthly and weekly distribution of the tweets is showed in Graph 3 and 4 located on the Appendix. Also, when looking for tweets with the cash-tag \$*EDPR*, most of the tweets were written in French, while when looking for \$*EDPR*.LS the preferred language was English (more data in Graph 5 of the Appendix). As showed in the Table 4 below, more than half the tweets presented a positive sentiment. Moreover, we labeled some tweets as Mixed because the polarity was both positive and negative. However, these tweets will be considered as neutral in the remaining of this thesis.

While analyzing the average monthly data (Table 4 below and Graph 6 in the appendix), investors might think this instrument has almost no volatility. However, when evaluating the daily returns (Graph 7 in the appendix) and the Maximum and Minimum daily returns values (Table 4 bellow and Graph 6 in the Appendix) it is possible to some major spikes. These three spikes are justified by events that happened in the country (24th July 2021 the Portugal transmission systems were disconnected from the synchronous are Continental Europe), inside the company (3rd November 2021 *EDPR* acquires Sunseap), or internationally (24th February 2022 the war between Ukraine and Russia started).

	MIXED	NEGATIVE	NEUTRAL	POSITIVE	TOTAL
June	0	3	4	0	7
July	0	1	1	6	8
August	0	0	0	2	2
September	0	0	2	13	15
October	1	1	1	9	12
November	0	2	2	2	6
December	0	0	4	5	9
January	1	0	4	3	8
February	0	1	3	3	7
March	0	1	4	6	11
April	0	0	0	5	5
May	0	0	2	9	11
Total	2	9	27	63	101

Table 4 - Tweets' Polarity by Month. Source: Umigon

DAILY RETURNS	MINIMUM	MAXIMUM	AVERAGE
JUNE	-0,032412716	0,035000999	0,000429227
July	-0,070082913	0,043701044	0,000798949
August	-0,025416715	0,047281534	0,003779764
September	-0,042101964	0,037636475	-0,001669291
October	-0,027825076	0,039599669	0,004399575
November	-0,061648941	0,037740328	-0,002729246
December	-0,041243004	0,028222218	-0,00104348
January	-0,047180954	0,04926109	-0,005233881
February	-0,038507379	0,100083403	0,005499897
March	-0,041134446	0,038761093	0,00232049
April	-0,023693878	0,04542482	-0,001691742
May	-0,040463721	0,043733205	0,000938614
Total	-0,070082913	0,100083403	0,000462812

Table 5 - Min, Max, and Average Monthly Returns of EDPR Stocks. Source: https://finance.yahoo.com/

The descriptive statistics of *EDPR* daily returns and daily average mood are displayed in the table 6 below:

VARIABLES	MEAN	MEDIAN	STANDARD DEVIATION	KURTOSIS	SKEWNESS	MINIMUM	MAXIMUM
EDPR	0.04984831	0.03171811	1	2.95064756	0.24426341	-3.7493	5.41478571
MOOD	0.70083592	0.35041796	1	1.80238689	1.12176437	-2.1574753	2.8583112
VTOT	0.87502908	0.43751454	1	4.15266172	2.19147919	0.43751454	5.06629154
VPOS	0.81779848	0.40889924	1	5.10324319	2.37944507	0.40889924	5.30228359
VNEG	0.26682044	0.13341022	1	76.7218908	8.31662291	0.13341022	12.3070929

Table 6 - Descriptive Statistics of the variables from 01/06/2021 to 01/06/2022

The EDPR daily returns is non-normally distributed and exhibits an average around 0,05 and a daily volatility of 1, indicating that the daily returns have a moderate amount of volatility or dispersion around the mean. Also, it indicates a slightly right-skewed distribution, suggesting the distribution has a longer right tail and is skewed towards the right, meaning that may have more frequent small positive values and fewer extreme negative returns pulling the mean and the median towards the right side of the distribution. This value can be justified by specific sector, company and market conditions, dynamics, or events between 2021 and 2022, like the quarterly earnings announcements, the marker persistent upward movements and recovery from the Covid-19 pandemic, the beginning of the war between Russia and Ukraine which led to embargos on the Russian oil, acquisitions, and others. The kurtosis value of 2.95064756 suggests that the dataset has a peak that is moderately higher or sharper compared to a normal distribution, indicating that the data may have a slightly more concentrated distribution around the mean compared to a normal distribution, suggesting the presence of les outliers or extreme values in the dataset. These outliers can be seen in graph 7 of the Appendix where there are not usually extreme values in the daily returns, triggered by news events and market shocks (3 of them named above) which cause a volatility clustering phenomenon. However, market inefficiencies like behavioral biases and the fact energetic companies are very volatile thanks to the influence of rare events, like the war and regulations changes, in their stock returns, can lead to more frequent price movements. Thus, the distribution is not highly non-normal since the kurtosis value is close to 3.

VARIABLES	MEAN	MEDIAN	STANDARD DEVIATION	KURTOSIS	SKEWNESS	MINIMUM	MAXIMUM
EDPR HISTORICAL VALUES	0.09349261	0.07865289	1	5.94569614	-0.366515	-7.8286627	6.19322131

Table 7 - Descriptive Statistics of the EDPR historical daily returns from 01/061/2013 to 31/05/2021

By analyzing the *EDPR* historical daily returns data from 2013 until the June 2021, we can observe that our sample period has a lower average daily returns and kurtosis and a higher skewness (Table 7 above). This suggests that, on average, the returns generated during the sample period were lower daily compared to the historical norm, indicating a deviation in the daily performance between the sample period and the historical data. It can imply that the sample period may have experienced relatively lower daily profitability or performance compared to the historical period, so the selected period is having lower daily returns than the average of the last 10 years. Also, the higher kurtosis in the historical data suggests it may exhibit more significant deviations in terms of extreme values, volatility, or non-linear patterns, and the skewness in sample period may exhibit a relatively more balanced distribution, while the historical data may have a more pronounced skewness in one direction.

The Mood Index, VPOS, VNEG and VTOT also displays a volatility of 1, means of 0,70, 0,82, 0,27, and 0,86, respectively. The kurtosis values indicate that the distribution have higher tails than a normal distribution and is leptokurtic shaped, except for the VTOT variable in which the kurtosis is Platykurtic distribution, meaning that the dataset has lighter tails and a flatter peak compared to a normal distribution, having fewer extreme values and being more dispersed. The skewedness is skewed towards the right and has a longer right tales, meaning there is a greater frequency of higher values and that the variables are non-normally distributed. The MOOD variable is skewed positively due to the tendency for users to express more positive sentiment, as is showed in the table 4 above, where 63 of 101 tweets are positive being only 9 negatives. In addition, there are more significant positive events associated with *EDPR* like earning announcements, successful campaigns like the acquisition, or more favorable economic indicators after the Covid-19 pandemic, which can generate a higher proportion of positive tweets. The heavy tailed kurtosis can be explained

by the viral nature of *Twitter* in which it can amplify sentiments to more extreme ones only by sharing or retweeting them.

When examining the impact of *Twitter* sentiment on stock returns, researchers employ Pearson's Correlation to determine if a linear relationship exists between these two variables, and whether *Twitter* sentiment can be considered a dependable predictor of changes in stock returns (Hamraoui & Boubaker, 2022; Jahjah & Rajab, 2020). I will utilize Pearson's Correlation to examine the relationship between the variables CHANGE and VTOT, VPOS, and VNEG to determine whether positivity and negativity are linearly related to daily stock returns, and which sentiment has a greater influence.

	CHANGE	VTOT	VPOS	VNEG
CHANGE				
PEARSON CORRELATION	1	0,054311284	0,041805797	0,03883488
SIG. (2-TAILED)		0,956717037	0,966676499	0,969043384
VTOT				
PEARSON CORRELATION	0,054311284	1	0,925126553	0,328484899
SIG. (2-TAILED)	0,956717037		0,355514677	0,742734467
VPOS				
PEARSON CORRELATION	0,041805797	0,925126553	1	-0,054701204
SIG. (2-TAILED)	0,966676499	0,355514677		0,956406603
VNEG				
PEARSON CORRELATION	0,03883488	0,328484899	-0,054701204	1
SIG. (2-TAILED)	0,969043384	0,742734467	0,956406603	

Table 8 - Pearson Correlation

Table 8 shows a positive correlation between VTOTAL, VPOS and VNEG in relation to CHANGE implying that as the discussion grows, so does the stock returns, regardless of sentiment., indicating a positive linear relationship. However, the correlations observed (0,054, 0,042, and 0,039) are almost null, this relationship is very weak, meaning that VTOTAL, VPOS and VNEG might have a slight tendency to have a minor influence on daily returns or even no influence. Additionally, the significance values (p-value) show that the observed correlations are not statistically significant because the values (0,957, 0,967, and 0,969) are greater than 0,05, meaning that there is no meaningful association

between these variables. This relationship is further tested using a multiple regression.

It is important to note VPOS and VTOT correlation value (0,925126553) is very high, indicating multicollinearity. This can lead to difficulties in the understanding of the unique effects of these variables on the CHANGE variable, and provide redundant, overlapping, or overfitting information, as well as reduce the robustness of the statistical test estimations. This is explained by a much higher proportions of the total tweets being considered positive tweets, so the positive number of tweets does not provide much additional information other than the data provided by the total number of tweets.

	CHANGE
MOOD	
PEARSON CORRELATION	0,01802876
SIG. (2-TAILED)	0,98562582

Table 9 - Pearson's Correlation between MOOD and CHANGE

On the other hand, the Pearson's Correlation coefficient between CHANGE and MOOD of 0,01802876 (Table 9) is very close to zero, indicating a very weak positive correlation or no linear relationship between the variables and that the variable MOOD has a very low explanatory power in predicting daily returns. The p-value is very high (0,98562582) suggesting no statistically significant evidence of existing correlation between the variables.

## 4.2. Main Results

### **REGRESSION STATISTICS**

R SQUARE	Adjusted R Square	F-statistic	P-value	<b>Durbin-Watson</b>	P-value
0,00344380	-0,002062	0,6255	0,5355813	1,8724	0,7499

Table 10 - Regression Statistics between CHANGE, VTOT, VPOS and VNEG

	COEFFICIENTS	STANDARD ERROR	T STAT	P-VALUE
INTERCEPT	-0,005502521	0,052863217	-0,10409	0,91716
VTOT	0,102608949	0,1307254028	0,785	0,43301
VPOS	-0,058546995	0,1381984140	-0,42364	0,67208
VNEG	NA	NA	NA	NA

Table 11 - Multiple Regression Model between CHANGE, VTOT, VPOS and VNEG

After proceeding to resolve the multiple regression between CHANDE, VTOT, VPOS, and VNEG we achieve the results presented on table 9 and 10. The low values of R Square and Adjusted R Square show the weak or negligible relationship between the independent and the dependent variables and that daily *EDPR* returns can be driven by other variables not included in the model. Also, that the relationship between the variables might not be strictly linear which causes the low value of the R Square. The F-statistic value and its corresponding p-value indicate that the overall model is not a good fit for the problem and is not statistically significant. So, the chosen independent variables don't capture adequately the factors influencing the *EDPR* daily returns. The Durbin-Watson values suggest a presence of positive autocorrelation in the residuals, which is not uncommon for time series data, specifically daily returns, which usually exhibit market trends and influence of previous days returns. However, a positive autocorrelation violates the assumption of independent errors, which can lead to inadequate coefficient estimates an unreliable statistical inference.

In table 11, we can see this inadequate coefficient estimate happening in the VNEG variable. Duo to singularities, most probably because of multicollinearity, the values for this variable were not calculated. To circumvent this problem, I tried three different approaches. First, I changed from an OLS regression to a

Generalized Linear Model, which gave the same results. Second, I changed the OLS regression to a polynomial regression, to try to dodge the nonlinearity problem, which gave the same results. Finally, used a ridge regression, which is a regularization technique, to mitigate the problems related to multicollinearity, in which it gave us similar results as before. As seen in the table 12 below, the R Square and Adjusted R Square value are still very low meaning the weakness of the relationship between the variables which is corroborated by the very low coefficients of the Intercept, VTOT, VPOS, and VNEG, which might not be the best variables to explain the changes in *EDPR* daily returns.

### **COEFFICIENTS**

INTERCEPT	-0,00113317
VTOT	0,02113093
VPOS	0,01487079
VNEG	0,02153037
R SQUARED	0,002079324
ADJUSTED R SQUARE	-0,006213646

Table 12 - Ridge Regression between CHANGE, VTOT, VPOS and VNEG

### **REGRESSION STATISTICS**

R SQUARE	Adjusted R Square	F-statistic	P-value	<b>Durbin-Watson</b>	P-value
0,00032503614	-0,002428889	0,118026481	0,731383760	1,8911	0,1465
Table 13 - Reg	ression Statistics Between MO	OD and CHANGE			

	COEFFICIENTS	STANDARD ERROR	T STAT	P-VALUE
INTERCEPT	0	0,0524059	0	1
MOOD	0,01802875863	0,052477857	0,343549823	0,731383759

Table 14 - Regression Model Between CHANGE and MOOD

For the second equation between CHANGE and MOOD I obtained the results in the tables 13 and 14 above. The results are very similar with the results of the first regression: very low values of R Square and Adjusted R square meaning MOOD have a weak or a null correlation with CHANGE (also shown in the coefficients in table 14); the F-statistic p-value show the model is not a good fit for the problem being MOOD not relevant to predict the daily stock returns; and the Durbin-Watson is below 2 and shows a positive autocorrelation violating the assumption of independent errors. These results show that the variables are not strictly linear, and, because of that, this model is not the best fit. To avoid the

nonlinearity of the variables I proceeded to change the OLS regression for a Polynomial regression. The results are the one's shown in the table 15 and 16 below.

#### **REGRESSION STATISTICS**

R SQUARE	Adjusted R Square	F-statistic	P-value	Durbin-Watson	P-value
0,003566	-0,00194	0,6477	0,5239	1,8707	0,1062
Table 15 - Polynomial Regression Statistics between MOOD and CHANGE					

	COEFFICIENTS	STANDARD ERROR	T STAT	P-VALUE
INTERCEPT	-0,03586	0,06195	-0,579	0,563
MOOD	-0,02209	0,06418	-0,344	0,731
(MOOD)^2	0,03596	0,03314	1,085	0,279

Table 16 - Polynomial Regression Model between MOOD and CHANGE

However, these results show that even calibrating the regression to avoid the nonlinearity problem, the new model is still not a good fit, not globally relevant, the variable relationship is still very weak or even zero, and there is still a positive autocorrelation problem.

### 4.3. Discussion

According to the results, H1 is validated because there was not statistically significance found between daily returns and the *Twitter* sentiment variables. Between 1st of June 2021 and 1st of June 2022 it is not possible to predict *EDPR* stock returns by analyzing the daily *Twitter* Sentiment.

Emotions are constructed psychological experiences that arrive from the combination of cognitive appraisals, contextual factors, and conceptual knowledge to a non-specific feeling state that people experience (Russel, 2003). However, sentiment is not separate from reason, and guides our decision-making process helping us make fast and intuitive judgments (Damásio, 1995). The brain generates predictions about incoming sensory information based on past experiences and bodily sensations and updates them based on feedback received, helping us correct cognitive bias (Barrett, 2017). The investors bias correction differs from individual to individual, depending on which past experiences, cultural and social norms, values, beliefs, practices, personality traits they have

which can affect the intensity, frequency, and duration of emotional experiences (Barrett, 2017). Even though media and social media can shape and influence attitudes, values, and emotional responses (Mesquita & Fridja, 1992). This is not verified in this study, being one of the possible reasons the specific context of the Portuguese investors and *EDPR* investors, which tend to have a long-term investment horizon, a global perspective, are interested in sustainability, are attracted by the growth potential of the company, and are willing to take risks, as renewable energy stocks like *EDPR* are influenced by a wide range of factors.

The long-term investment horizon of *EDPR* investors can help explain these findings. Renewable energy stock investors, in general, tend to be risk-seekers due to the inherent risks associated with investing in a company dependent on external factors such as legislation. Investors in these firms often hold their stocks in portfolios for extended periods of time. Additionally, the consistent payment of dividends by *EDPR*, along with its dividend yield, attracts investors with a focus on long-term investments.

On the other hand, *Twitter* is a platform commonly used for short-term interactions, which appeals to investors seeking quick profits. However, this type of investor is not typical for *EDPR*, as institutional players make up around 96% of the company's shareholder base(*EDP Renováveis*, 2023a). Consequently, *EDPR* is considered a robust company with steady net income, operating in a solid energy industry (*EDP Renováveis*, 2023b). The recent conflict between Ukraine and Russia has further strengthened the energy market, causing energy prices to rise. This has positively impacted *EDPR*, as evidenced by the significant increase in daily returns when the war began. This highlights the resilience and solidity of the company in such circumstances (Yahoo! Finance, 2023).

These conclusions corroborate the findings of Reboredo & Ugolini (2018) where they found no useful information in *Twitter* sentiment to forecast prices of renewable energy companies, such as *EDP Renováveis*. Though, goes against the findings of Oliveira et al. (2017) who found predictive power of returns from energy companies by analyzing *Twitter* sentiment. However, Oliveira et al. (2017) study analyzes companies from big and non-concentrated indexes and markets, which is the contrary of the *EDPR* context, who belongs to a concentrated and small market like the Portuguese, being one of the reasons for H1 not being rejected. Actually, studies that examine small and concentrated markets or indexes tendency is to not find relationship between *Twitter* sentiment and stock

returns. My findings validate Nisar & Yeung (2018) and Nofer & Hinz (2015) results of no significant relationship found while using very similar regression techniques, which are multiple and OLS regressions, and variables like the Mood indexes and total number of positive and negative tweets, while analyzing markets similar with the Portuguese market.

On one hand, Hamraoui & Boubaker (2022), analyzing a market similar to the Portuguese one, very concentrated and with smaller companies, found predictive power of sentiment in stock returns in some companies, but none of them where energy enterprises, being every energetic firm considered to have no relationship when investigating the correlation between *Twitter* sentiment and their returns, corroborating our findings. On the other hand, *EDPR* is a company with few analysts' coverage and with posts with very little retweets, which according to Gu & Kurov (2020) and Sul et al. (2017) are the firms with strongest predictive power, going against my observation of *Twitter* mood having no forecasting ability for predicting daily returns between 2021 and 2022 for *EDPR* stocks.

Furthermore, between the period studied there were several events that could had led the decision-making process for more biased judgements. Quarterly earnings announcements, the Covid-19 pandemic which led to economic uncertainty as well as the economic recovery from this virus, the armed conflict between Ukraine and Russia which had direct impact on the energy sector in Europe with the embargo imposed in the Russian oil and petrol, limiting the Portuguese energy supply and rising the energetic prices, and the disconnection from the European power grid because of structural problems, resulting in national energetic distribution problems, are some of the main events between this period. The theory suggests investors are influenced by events that impact the company or them directly, biasing their sentiment and their decisions (Tetlock, 2007). In this case, the Twitter sentiment did not have influence even when there was a known or "unknown" event on stock returns. This can be justified because events have much more impact on investors who rely on publicly available and cost-effective information, meaning individual investors. EDPR stockholder structure is composed only by 4% of individual investors, being the remaining percentage composed by institutional investors between corporations and funds, justifying why events, even though the theory suggests so, did not have influence in the relationship between Twitter mood and stock

returns, because there is a very small ownership by individual investors (*EDP Renováveis*, 2023a).

Moreover, theory indicates the existence of bias and heuristics which can influence trading strategies or the decision-making process. In our specific Twitter case, my findings suggests that there is no influence of the Twitter sentiment for investors to follow the trends (Hurst et al., 2017), to have herding behavior and follow the crowd (Gigerenzer, 1991), for investors to fell pressure to follow the opinions of their community or the people they follow on Twitter that post about EDPR (Maitland & Sammartino, 2015), or to be influenced by clickbait titles, the availability of the *Twitter* information, or the anchoring of the posts, which could led for the decision -making process to not be as accurate as investors think (Gurdgiev & O'Loughlin, 2020). Even though the study does not analyze these biases directly, not finding relationship between the Twitter mood and stock returns suggest that the one's related to Twitter posts about EDPR does not have much influence on the investors which rely on Twitter information. Once again, these can be justified for the very small percentage of individual investors existent on the stockholder structure of EDPR, because these investors are the most influenced by these biases and heuristics for relying in more easy access information and for having less financial literacy than the institutional investors.

## 5. Conclusion

## 5.1. Main Conclusions

The goal of this thesis was to research the possibility of *EDP Renováveis* stock returns being predicted by analyzing *Twitter* sentiment. The methodology applied was taken from Nisar & Yeung (2018) and the daily returns of *EDPR* were employed to the sample period of 1st of June 2021 to 1st of June 2022. The results suggests that neither the 101 tweet's polarity about *EDPR* collected neither the social mood index built from that data have impact in the variations of its daily returns. The independent variables chosen revealed limited explanatory power in predicting daily returns and the regression model had a lack of meaningful explanatory power.

By focusing on the company *EDPR*, this study provides valuable insight into the specific dynamics of sentiment and stock returns inside the Portuguese market, a small and non-influent market, and the sector of renewable energies. The results do not support the behavioral finance theory suggesting the investors of the Portuguese market may not be influenced by sentiment, events, biases, emotions, and heuristics, as other empirical evidence suggest.

This study adds to the growing body of research that casts questions on the effectiveness of sentiment analysis in predicting stock market trends. It aligns with previous literature that have suggested the limited practicality of using *Twitter* sentiment to forecast stock prices and performances. By confirming these findings, this study strengthens the existing evidence and enhances our understanding of the role of sentiment in behavioral finance.

Moreover, the findings of this study carry implications for the wider scientific community engaged in behavioral finance research. They challenge established theories and models by raising doubts about the assumed impact of *Twitter* sentiment on the dynamics of the renewable energy stock market. This study encourages researchers to reconsider their assumptions and explore alternative explanations for investor behavior and market outcomes. It also stimulates further investigation into the complex interplay of psychological, economic, and social factors that shape the decision-making processes of investors.

## 5.2. Implications for Management

With the constant growing of sentiment analysis, there are some advice that can be made to market players in order for them to be able to manage risks better by considering the results of this thesis.

For investors and traders, they should recognize that, at least in a small market like the Portuguese or at a renewable energy company like *EDPR*, there is limited predictive power in forecasting the stock market through *Twitter* sentiment. Only relying on sentiment analysis technics to make investment decision should never be an option. However, incorporating other fundamental and technical analysis tools alongside the sentiment analysis can be a beneficial way to make a better financial decision. Also, investors should stay updated on new research around this topic to gain a deeper and better understanding of the market dynamics.

On the other hand, financial analysts and advisors should acknowledge these results and understand the shortcomings of sentiment analysis as a stand-alone tool to forecast the market. When providing clients with their tips and predictions, they should take in consideration that they need to provide more comprehensive analysis which include other relevant factors. Thus, they should highlight the importance of a diversified investment portfolio not only based in sentiment-driven strategies, educate their costumers about the limitations these strategies have, and encourage a balanced approach to investment decision-making.

Market regulators like the *CMVM*, whose one objective is to promote the transparency of the market, should disclosure the importance of being aware in using sentiment analysis strategies for investment decisions, as well as its limitations and potential risks. Additionally, the can take into account the findings and consider implementing guidelines or regulations to ensure a responsible and accurate use of sentiment-based strategies in the decision-making process.

For the data providers or even the social media platforms which are constantly growing popularity for research, they should enhance the accuracy of data analysis by refining algorithms or incorporating contextual information into the information provided. Besides, the creation of disclaimers and explanations of the limitation and biases sentiment analysis can have to prevent misinterpretation by market participants should be added. Collaborating with

researchers to provide more insight of the actual benefits and reliability of sentiment analysis in financial contexts might also be a favorable option.

### 5.3. Limitations

This thesis has several limitations that can affect the empirical evidence. Firstly, the sample size of *Twitter* data is relatively small. This happens because the method of tweets search was through cash tags (\$EDPR and \$EDPR.LS). There might be many other tweets related to *EDPR*, but which were not considered due to not having the cash tags in them. So, the sample collected may not fully represent the sentiments and behaviors of the investor population. Also, the presence of a lot of spam, irrelevant and neutral tweets can restrict the generalizability of these findings. Additionally, the sentiment analysis might not be accurate because the algorithm of Umigon has its limitations and biases while interpreting sentiments. It is also important to acknowledge that there is a nonlinear relationship between variables, so the chosen model might not be the most appropriate one to obtain the best results.

Furthermore, this study only has in consideration sentiment variables ignoring other contextual factors that, if analyzed simultaneously with these sentiment variables, might confound the relationship between sentiment and stock returns, like macroeconomics indicators or geopolitical events. This study also is focused on a specific company, in a specific market, on a specific time frame, some the findings ca nott be directly applicable to other companies, markets, financial contexts, or even the same company in another timeframe, being the unique characteristics of the study context considered.

### 5.4. Future Research

For future research, a larger and more diverse sample size from multiple companies from the Portuguese market, multiple renewable energy companies or multiple time periods can provide more comprehensive understandings and enhance the generalizability. Moreover, more longitudinal analysis can help capture trends, patters, and potential shifts in the influence of the sentiment in the market.

There should be investigation of this particular context but with different techniques, algorithms, tools, data sources, like nonlinear models, cross-market and cross-country analysis, to compare and evaluate the effectiveness of these differences and help identify the most accurate and reliable approach. Also, the use of other sentiment indicators, variables and market conditions can moderate the relationship between sentiment and stock returns.

## **Bibliography**

- 1. Andersen, T. G., Bollerslev, T., Diebold, F. X., & Labys, P. (2003). Modeling and Forecasting Realized Volatility. *Econometrica*, 71(2), 579–625. https://doi.org/https://doi.org/10.1111/1468-0262.00418
- 2. Ariel, R. A. (1990). High Stock Returns before Holidays: Existence and Evidence on Possible Causes. *The Journal of Finance*, 45(5), 1611–1626.
- 3. Aste, T. (2019). Cryptocurrency market structure: connecting emotions and economics. *Digital Finance*, 1(1), 5–21. https://doi.org/10.1007/s42521-019-00008-9
- 4. Awan, M. J., Rahim, M. S. M., Nobanee, H., Munawar, A., Yasin, A., & Zain, A. M. (2021). Social Media and Stock Market Prediction: A Big Data Approach. *Computers, Materials and Continua, 67*(2), 2569–2583. https://doi.org/10.32604/cmc.2021.014253
- 5. Baker, M., & Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. *The Journal of Finance*, *61*(4), 1645–1680. https://doi.org/https://doi.org/10.1111/j.1540-6261.2006.00885.x
- 6. Baker, M., & Wurgler, J. (2007). Investor Sentiment in the Stock Market. *Journal of Economic Perspective*, 21(2), 129–151.
- 7. Banz, R. W. (1981). The Relationship Between Return and Market Value of Common Stocks. *Journal of Financial Economics*, 9(1), 3–18.
- 8. Barberis, N., & Thaler, R. H. (2003). A Survey of Behavioral Finance. In G. Constantinides, R. Stulz, & M. Harris (Eds.), *Handbook of the Economics of Finance* (1st ed., Vol. 1B, pp. 1053–1128). Elsevier.
- 9. Barrett, L. (2017). *How Emotions Are Made: The Secret Life of the Brain.* Houghton Mifflin Harcourt.
- 10. Barrière, V., & Balahur, A. (2020). Improving Sentiment Analysis over non-English Tweets using Multilingual Transformers and Automatic Translation for Data-Augmentation. *Proceedings of the 28th International Conference on Computational Linguistics*, abs/2010.03486, 266–271.
- 11. Barsade, S. G. (2002). The Ripple Effect: Emotional Contagion and its Influence on Group Behavior. *Administrative Science Quarterly*, 47(4), 644–675. https://doi.org/10.2307/3094912

- 12. Bartov, E., Faurel, L., & Mohanram, P. S. (2018). Can *Twitter* help predict firm-level earnings and stock returns? *Accounting Review*, 93(3), 25–57. https://doi.org/10.2308/accr-51865
- 13. Batra, R., & Daudpota, S. M. (2018). Integrating StockTwits with sentiment analysis for better prediction of stock price movement. 2018

  International Conference on Computing, Mathematics and Engineering Technologies (ICoMET), 1–5.

  https://doi.org/10.1109/ICOMET.2018.8346382
- 14. Behrendt, S., & Schmidt, A. (2018). The *Twitter* myth revisited: Intraday investor sentiment, *Twitter* activity and individual-level stock return volatility. *Journal of Banking & Finance*, 96(1), 355–367. https://doi.org/https://doi.org/10.1016/j.jbankfin.2018.09.016
- 15. Bollen, J., Mao, H., & Pepe, A. (2011). Modeling Public Mood and Emotion: *Twitter* Sentiment and Socio-Economic Phenomena. *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, 450–453. www.aaai.org
- 16. Bollen, J., Mao, H., & Zeng, X. (2011). *Twitter* mood predicts the stock market. *Journal of Computational Science*, 2(1), 1–8. https://doi.org/10.1016/j.jocs.2010.12.007
- 17. Brooks, C. (2014). Modelling Volatility and Correlation. In *Introductory Econometrics for Finance* (1st ed., Vol. 2). Cambridge University Press.
- 18. Busenitz, L. W., & Barney, J. B. (1997). Differences between entrepreneurs and managers in large organizations: Biases and heuristics in strategic decision-making. *Journal of Business Venturing*, 12(1), 9–30. https://doi.org/https://doi.org/10.1016/S0883-9026(96)00003-1
- 19. Busse, J. A. (1999). Volatility Timing in Mutual Funds: Evidence from Daily Returns. *The Review of Financial Studies*, 12(5), 1009–1041. https://doi.org/10.1093/rfs/12.5.1009
- 20. Cao, C. X., & Chen, C. (2022). Political sentiment and stock crash risk. *The Journal of Risk Finance*, 23(2), 139–154. https://doi.org/10.1108/JRF-11-2021-0186
- 21. CEIC DATA. (2023a, May 15). *Portugal Capitalização de Mercado*. Https://Www.Ceicdata.Com/Pt/Indicator/Portugal/Market-Capitalization.

- 22. CEIC DATA. (2023b, May 15). Portugal Market Capitalization: % of GDP. Https://Www.Ceicdata.Com/En/Indicator/Portugal/Market-Capitalization--Nominal-Gdp.
- 23. Chen, C., & Hafner, C. (2019). Sentiment-Induced Bubbles in the Cryptocurrency Market. *Journal of Risk and Financial Management*, 12(2), 53. https://doi.org/10.3390/jrfm12020053
- 24. Chen, W. (2013). Can Corporate Governance Mitigate the Adverse Impact of Investor Sentiment on Corporate Investment Decisions? Evidence from Taiwan. *Asian Journal of Finance and Accounting*, 5(2), 101–126.
- 25. Chernomas, R., & Hudson, I. (2017). Finance, Crisis and the Efficient Market Hypothesis. In *The Profit Doctrine* (pp. 125–146). Pluto Press. https://doi.org/10.2307/j.ctt1jktsbd.12
- 26. CMVM. (2023a, May 15). O Perfil do Investidor Particular Português. Https://Www.CMVM.Pt/Pt/EstatisticasEstudosEPublicacoes/Estudos/Pages/O%20Perfil%20do%20Investidor%20Particular%20Portugu%C3%AAs.Aspx?V=#3.
- 27. CMVM. (2023b, May 15). Who Are We? Https://Www.CMVM.Pt/En/The\_CMVM/Overview/Pages/Who-We-Are.Aspx.
- 28. Copeland, T. E., & Mayers, D. (1982). The value line enigma (1965–1978):

  A case study of performance evaluation issues. *Journal of Financial Economics*, 10(3), 289–321.

  https://doi.org/https://doi.org/10.1016/0304-405X(82)90004-6
- 29. Creswell, J. W. (2014). Research Design: Qualitative, Quantitative, and Mixed Methods Approaches (3rd ed.). Sage.
- 30. Damásio, A. (1995). *Descartes' Error: Emotion, Reason, and the Human Brain.* Harper Perennial.
- 31. Daniel, K. D., Hirshleifer, D., & Teoh, S. H. (2002). Investor Psychology in Capital Markets: Evidence and Policy Implications. *Journal of Monetary Economics*, 49(1), 139–209.
- 32. D'Arms, J., & Jacobson, D. (2000). Sentiment and value. *Ethics*, 110(4), 722–748. https://doi.org/10.1086/233371
- 33. Datareportal. (2023a, May 15). *Digital* 2023: *Portugal*. Https://Datareportal.Com/Reports/Digital-2023-Portugal.

- 34. Datareportal. (2023b, May 15). *Twitter Users, Stats, Data & Trends*. Https://Datareportal.Com/Essential-*Twitter*-Stats.
- 35. De Bondt, W. F. M., & Thaler, R. (1985). Does the Stock Market Overreact? *The Journal of Finance*, 40(3), 793–805.
- 36. Demek, K. C., Raschke, R. L., Janvrin, D. J., & Dilla, W. N. (2018). Do organizations use a formalized risk management process to address social media risk? *International Journal of Accounting Information Systems*, 28(1), 31–44.
  - https://doi.org/https://doi.org/10.1016/j.accinf.2017.12.004
- 37. Devine, P. (1989). Stereotypes and Prejudice: Their Automatic and Controlled Components. *Journal of Personality and Social Psychology*, 56(1), 5–18. https://doi.org/10.1037//0022-3514.56.1.5
- 38. Dumouchel, W. H., & Duncan, G. J. (1983). Using Sample Survey Weights in Multiple Regression Analyses of Stratified Samples. *Journal of the American Statistical Association*, 78(383), 535–543. https://doi.org/10.1080/01621459.1983.10478006
- 39. *EDP Renováveis*. (2023a, June 15). *Shareholder Structure*. Https://Www.*EDPR*.Com/En/Investors/Shareholder-Structure.
- 40. *EDP Renováveis*. (2023b, July 5). *Reports & Presentations*. Https://Www.*EDPR*.Com/En/Investors/Investor-Information/Reports-and-Presentations.
- 41. Fama, E. F. (1965). Random Walks in Stock Market Prices. *Financial Analysts Journal*, 21(5), 55–59.
- 42. Fama, E. F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *The Journal of Finance*, 25(2), 383–417.
- 43. French, K. (1980). Stock Returns and the Weekend Effect. *Journal of Financial Economics*, 8(1), 55–69.
- 44. French, K. (1988). Crash Testing the Efficient Market Hypothesis. In S. Fischer (Ed.), *NBER Macroeconomics Annual* (Vol. 3, pp. 277–286). M.I.T. Press.
- 45. Fudholi, D. H., Nayoan, R. A. N., & Rani, S. (2022). Stock Prediction Based on *Twitter* Sentiment Extraction Using BiLSTM-Attention. *Indonesian Journal of Electrical Engineering and Informatics*, 10(1), 187–198. https://doi.org/10.52549/ijeei.10i1.3011

- 46. Fumio Hayashi. (2000). Finite-Sample Properties of OLS. In *Econometrics* (1st ed., pp. 3–87). Princeton University Press.
- 47. Gigerenzer, G. (1991). How to Make Cognitive Illusions Disappear: Beyond "Heuristics and Biases." *European Review of Social Psychology*, 2(1), 83–115. https://doi.org/10.1080/14792779143000033
- 48. Goel, S., & Uzuner, O. (2016). Do Sentiments Matter in Fraud Detection? Estimating Semantic Orientation of Annual Reports. *Intelligent Systems in Accounting, Finance and Management,* 23(3), 215–239. https://doi.org/https://doi.org/10.1002/isaf.1392
- 49. Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, 74(6), 1464–1480.
- 50. Gryglewicz, S. (2011). A theory of corporate financial decisions with liquidity and solvency concerns. *Journal of Financial Economics*, 99(2), 365–384. https://doi.org/https://doi.org/10.1016/j.jfineco.2010.09.010
- 51. Gu, C., & Kurov, A. (2020). Informational role of social media: Evidence from *Twitter* sentiment. *Journal of Banking & Finance*, 121(1), 1–44. https://doi.org/https://doi.org/10.1016/j.jbankfin.2020.105969
- 52. Gurdgiev, C., & O'Loughlin, D. (2020). Herding and anchoring in cryptocurrency markets: Investor reaction to fear and uncertainty. *Journal of Behavioral and Experimental Finance*, 25(1), 1–10. https://doi.org/https://doi.org/10.1016/j.jbef.2020.100271
- 53. Hamraoui, I., & Boubaker, A. (2022). Impact of *Twitter* sentiment on stock price returns. *Social Network Analysis and Mining*, 12(1), 28. https://doi.org/10.1007/s13278-021-00856-7
- 54. Hanna, A. J., Turner, J. D., & Walker, C. B. (2020). News media and investor sentiment during bull and bear markets. *The European Journal of Finance*, 26(14), 1377–1395. https://doi.org/10.1080/1351847X.2020.1743734
- 55. Haritha, P. H., & Rishad, A. (2020). An empirical examination of investor sentiment and stock market volatility: evidence from India. *Financial Innovation*, 6(1), 34. https://doi.org/10.1186/s40854-020-00198-x
- 56. Hasselgren, B., Chrysoulas, C., Pitropakis, N., & Buchanan, W. J. (2023). Using Social Media: Sentiment Analysis to Make Investment Decisions. *Future Internet*, *15*(1), 1–23. https://doi.org/10.3390/fi15010005

- 57. Haugen, R. A., & Lakonishok, J. (1988). *Incredible January Effect: The Stock Market's Unsolved Mystery* (1st ed., Vol. 1). Dow Jones -Irwin.
- 58. Henderi, Wahyuningsih, T., & Rahwanto, E. (2021). Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer. *International Journal of Informatics and Information System*, 4(1), 13–20.
- 59. Herrera, G. P., Constantino, M., Su, J. J., & Naranpanawa, A. (2022). Renewable energy stocks forecast using *Twitter* investor sentiment and deep learning. *Energy Economics*, 114. https://doi.org/10.1016/j.eneco.2022.106285
- 60. Ho, C.-S., Damien, P., Gu, B., & Konana, P. (2017). The time-varying nature of social media sentiments in modeling stock returns. *Decision Support Systems*, 101, 69–81. https://doi.org/https://doi.org/10.1016/j.dss.2017.06.001
- 61. Hu, M., & Liu, B. (2004). Mining and Summarizing Customer Reviews. *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 168–177. https://doi.org/10.1145/1014052.1014073
- 62. Huang, C., Yang, X., Yang, X., & Sheng, H. (2014). An Empirical Study of the Effect of Investor Sentiment on Returns of Different Industries. *Mathematical Problems in Engineering*, 2014, 1–11. https://doi.org/10.1155/2014/545723
- 63. Hudson, R. S., & Gregoriou, A. (2015). Calculating and comparing security returns is harder than you think: A comparison between logarithmic and simple returns. *International Review of Financial Analysis*, 38, 151–162. https://doi.org/https://doi.org/10.1016/j.irfa.2014.10.008
- 64. Hurst, B., Ooi, Y. H., & Pedersen, L. H. (2017). A Century of Evidence on Trend-Following Investing. *The Journal of Portfolio Management*, 44(1), 15–29.
- 65. International Monetary Fund. (2022a). World Economic Outlook Update.
- 66. International Monetary Fund. (2022b). World Economic Outlook Update January 2022.
- 67. International Trade Administration. (2023, May 15). *Portugal Country Commercial Guide*. Https://Www.Trade.Gov/Country-Commercial-Guides/Portugal-Energy.

- 68. Jahjah, F., & Rajab, M. (2020). Impact of *Twitter* Sentiment Related to Bitcoin on Stock Price Return. *Journal of Engineering*, 26(6), 60–71.
- 69. Jain, A., Nandakumar, K., & Ross, A. (2005). Score normalization in multimodal biometric systems. *Pattern Recognition*, *38*(12), 2270–2285. https://doi.org/https://doi.org/10.1016/j.patcog.2005.01.012
- 70. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021a). Moving Beyond Linearity. In *An Introduction to Statistical Learning with Applications in R* (2nd ed., pp. 289–326). Springer.
- 71. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2021b). Linear Model Selection and Regularization. In *An Introduction to Statistical Learning with Applications in R* (2nd ed., pp. 225–288). Springer.
- 72. Johnston, M. (2014). Secondary Data Analysis: A Method of Which the Time has Come. *Qualitative and Quantitative Methods in Libraries*, 3(1), 619–626.
- 73. Kahneman, D., & Tversky, A. (1979). "Prospect Theory: An Analysis of Decision under Risk." *Econometrica: The Econometric Society*, 47(2), 263–292. https://about.jstor.org/terms
- 74. Kaplanski, G., & Levy, H. (2010). Sentiment and stock prices: The case of aviation disasters. *Journal of Financial Economics*, 95(2), 174–201. https://doi.org/10.1016/j.jfineco.2009.10.002
- 75. Kearney, C., & Liu, S. (2014). Textual sentiment in finance: A survey of methods and models. *International Review of Financial Analysis*, 33, 171–185. https://doi.org/10.1016/j.irfa.2014.02.006
- 76. Keim, D. B. (1983). Size-related anomalies and stock return seasonality: Further empirical evidence. *Journal of Financial Economics*, 12(1), 13–32. https://doi.org/https://doi.org/10.1016/0304-405X(83)90025-9
- 77. Kim, M., & Park, J. (2015). Individual Investor Sentiment and Stock Returns: Evidence from the Korean Stock Market. *Emerging Markets Finance and Trade*, 51(sup5), S1–S20. https://doi.org/10.1080/1540496X.2015.1062305
- 78. Lakonishok, J., & Smidt, S. (1988). Are Seasonal Anomalies Real? A Ninety-Year Perspective. *The Review of Financial Studies*, 1(4), 403–425. https://about.jstor.org/terms
- 79. Langer, J. (1975). The Illusion of Control. *Journal of Personality and Social Psychology*, 32(2), 311–328.

- 80. le Bon, G. (1896). *The Crown: A Study Of The Popular Mind*. The Macmillan Co.
- 81. Liu, B. (2006). Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (2nd ed.). Springer.
- 82. Lo, A., & MacKinlay, A. C. (1999). *A Non-Rnadom Walk Down Wall Street*. Princeton University Press.
- 83. Lo, A. W. (2005). Reconciling Efficient Markets with Behavioral Finance: The Adaptive Markets Hypothesis. *The Journal of Investment Consulting*, 7(2), 21–44.
- 84. Lo, A. W. (2007). Efficient Market Hypothesis. In *The New Palgrave: A Dictionary Of Economics* (2nd ed.). Palgrave Macmillan. http://ssrn.com/abstract=991509
- 85. Maitland, E., & Sammartino, A. (2015). Decision making and uncertainty: The role of heuristics and experience in assessing a politically hazardous environment. *Strategic Management Journal*, 36(10), 1554–1578. https://doi.org/https://doi.org/10.1002/smj.2297
- 86. Mao, Y., Wei, W., Wang, B., & Liu, B. (2012). Correlating S&P 500 stocks with Twitter data. Proceedings of the 1st ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research, HotSocial 2012, 69–72. https://doi.org/10.1145/2392622.2392634
- 87. Melville, P., Gryc, W., & Lawrence, R. (2009). Sentiment Analysis of Blogs by Combining Lexical Knowledge with Text Classification. *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '09*, 1275–1284. https://doi.org/10.1145/1557019.1557156
- 88. Mendoza-Urdiales, R. A., Núñez-Mora, J. A., Santillán-Salgado, R. J., & Valencia-Herrera, H. (2022). *Twitter* Sentiment Analysis and Influence on Stock Performance Using Transfer Entropy and EGARCH Methods. *Entropy*, 24(7). https://doi.org/10.3390/e24070874
- 89. Mesquita, B., & Fridja, N. H. (1992). Cultural variations in emotions: a review. *Psychological Bulletin*, 112(2), 179–204.
- 90. Mittal A, Goel A. Stock Prediction Using Twitter Sentiment Analysis; 2009. Available from: https://pdfs.semanticscholar.org/4ecc/55e1c3ff1cee41f21e5b0a3b22c58 d04c9d6.pdf. Accessed April 21, 2023

- 91. Mizumoto, A., & Plonsky, L. (2016). R as a Lingua Franca: Advantages of Using R for Quantitative Research in Applied Linguistics. *Applied Linguistics*, 37(2), 284–291. https://doi.org/10.1093/applin/amv025
- 92. Nicholson, S. F. (1960). Price-Earnings Ratios. *Financial Analysts Journal*, 16(4), 43–45. https://doi.org/10.2469/faj.v16.n4.43
- 93. Nickerson, R. S. (1998). Confirmation Bias: A Ubiquitous Phenomenon in Many Guises. *Review of General Psychology*, 2(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175
- 94. Nisar, T. M., & Yeung, M. (2018). *Twitter* as a tool for forecasting stock market movements: A short-window event study. *Journal of Finance and Data Science*, 4(2), 101–119. https://doi.org/10.1016/j.jfds.2017.11.002
- 95. Nofer, M., & Hinz, O. (2015). Using *Twitter* to Predict the Stock Market: Where is the Mood Effect? *Business and Information Systems Engineering*, 57(4), 229–242. https://doi.org/10.1007/s12599-015-0390-4
- 96. Oliveira, N., Cortez, P., & Areal, N. (2017). The impact of microblogging data for stock market prediction: Using *Twitter* to predict returns, volatility, trading volume and survey sentiment indices. *Expert Systems with Applications*, 73(1), 125–144. https://doi.org/https://doi.org/10.1016/j.eswa.2016.12.036
- 97. Paltoglou, G., & Thelwall, M. A. (2012). *Twitter*, MySpace, Digg: Unsupervised Sentiment Analysis in Social Media. *ACM Transactions on Intelligent Systems and Technology*, 3(4), 1–19.
- 98. Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(2), 1–135.
- 99. Papadopoulos, A. (2023, May 15). *The Biggest Companies in Portugal by Market Capitalization (March 28, 2023)*. Https://Ceoworld.Biz/2023/03/28/the-Biggest-Companies-in-Portugal-by-Market-Capitalization-March-28-2023/.
- 100. Pereira da Silva, P. (2016). A Anatomia do Mercado Português: Uma Análise da Microestrutura e Liquidez da Euronext Lisbon.
- 101. Peterson, R. (2016). *Trading on Sentiment: The Power of Minds Over Markets* (1st ed.). Wiley Finance.
- 102. Polyzos, E. (2022). Escalating Tension and the War in Ukraine: Evidence Using Impulse Response Functions on Economic Indicators and Twitter Sentiment. https://ssrn.com/abstract=4058364

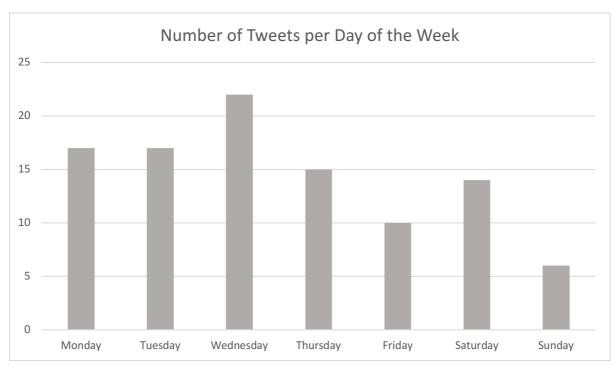
- 103. Putnam Investments. (2023, July 5). *Use of Social Media Nearly Universal Among Financial Advisors, According to New Putnam Investments Study*. Https://Www.Putnam.Com/Newsroom/Press-Release/970-Use-of-Social-Media-Nearly-Universal-among-Financial-Advisors-According-to-New-Putnam-Investments-Study/.
- 104. Ranco, G., Aleksovski, D., Caldarelli, G., Grčar, M., & Mozetič, I. (2015). The effects of *Twitter* sentiment on stock price returns. *PLoS ONE*, 10(9). https://doi.org/10.1371/journal.pone.0138441
- 105. Rane, A., & Kumar, A. (2018). Sentiment Classification System of *Twitter* Data for US Airline Service Analysis. 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), 01, 769–773. https://doi.org/10.1109/COMPSAC.2018.00114
- 106. Rao, T., & Srivastava, S. (2012). Analyzing Stock Market Movements Using *Twitter* Sentiment Analysis. *International Conference on Advances in Social Networks Analysis and Mining*, 119–123.
- 107. Reboredo, J. C., & Ugolini, A. (2018). The impact of *Twitter* sentiment on renewable energy stocks. *Energy Economics*, 76, 153–169. https://doi.org/10.1016/j.eneco.2018.10.014
- 108. Ribeiro, F. N., Araújo, M., Gonçalves, P., André Gonçalves, M., & Benevenuto, F. (2016). SentiBench a benchmark comparison of state-of-the-practice sentiment analysis methods. *EPJ Data Science*, *5*(1), 23. https://doi.org/10.1140/epjds/s13688-016-0085-1
- 109. Roll, R. (1983). Vas ist das? The turn-of-the-year effect and the return premia of small firms. *The Journal of Portfolio Management*, 9(2), 18–28.
- 110. Roll, R. (1984). Orange Juice and Weather. *The American Economic Review*, 74(5), 861–880. http://www.jstor.org/stable/549
- 111. Ruan, Y., Durresi, A., & Alfantoukh, L. (2018). Using *Twitter* trust network for stock market analysis. *Knowledge-Based Systems*, 145(1), 207–218. https://doi.org/https://doi.org/10.1016/j.knosys.2018.01.016
- 112. Russel, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review*, 110(1), 145–172.
- 113. Samuelson, P. (1965). Proof That Properly Anticipated Prices Fluctuate Randomly. *Industrial Management Review*, 6(2), 41–49.

- 114. Saranya, C., & Manikandan, G. (2013). A Study on Normalization Techniques for Privacy Preserving Data Mining. *International Journal of Engineering and Technology*, 5(3), 2701–2704.
- 115. Selden, G. (1912). *Psychology of the Stock Market* (1st ed.). Ticker Publishing Company.
- 116. Sharma, A., & Kumar, A. (2019). A review paper on behavioral finance: study of emerging trends. *Qualitative Research in Financial Markets*, 12(2), 137–157. https://doi.org/10.1108/QRFM-06-2017-0050
- 117. Shiller, R. (2000). Irrational Exuberance. *The American Journal of Economics and Sociology*, 59(3), 537–540.
- 118. Shleifer, A. V. D., & Summers, L. H. l. (1990). The Noise Trader Approach to Finance. *Journal of Economic Perspectives*, 4(2), 19–33.
- 119. Simon, H. A. (1986). Rationality in Psychology and Economics. *The Journal of Business*, 59(4), 209–224. https://www.jstor.org/stable/2352757
- 120. Song, Y., Ji, Q., Du, Y. J., & Geng, J. B. (2019). The dynamic dependence of fossil energy, investor sentiment and renewable energy stock markets. *Energy Economics*, 84. https://doi.org/10.1016/j.eneco.2019.104564
- 121. Souza, T., Kolchyna, O., Treleaven, P., & Aste, T. (2015). Twitter Sentiment Analysis Applied to Finance: A Case Study in the Retail Industry.
- 122. Sprenger, T., & Welpe, I. (2010). Tweets and Trades: The Information Content of Stock Microblogs.
- 123. Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior. *Journal of Management Information Systems*, 29(4), 217–247. http://www.jstor.org/stable/43590107
- 124. Sul, H. K., Dennis, A. R., & Yuan, L. (Ivy). (2017). Trading on *Twitter*: Using Social Media Sentiment to Predict Stock Returns. *Decision Sciences*, 48(3), 454–488. https://doi.org/https://doi.org/10.1111/deci.12229
- 125. Teti, E., Dallocchio, M., & Aniasi, A. (2019). The relationship between *Twitter* and stock prices. Evidence from the US technology industry. *Technological Forecasting and Social Change*, 149. https://doi.org/10.1016/j.techfore.2019.119747

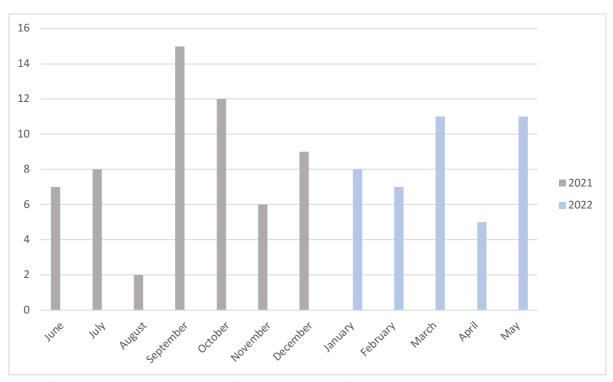
- 126. Tetlock, P. C. (2007). Giving Content to Investor Sentiment: The Role of Media in the Stock Market. *The Journal of Finance*, *62*(3), 1139–1168. http://www.jstor.org/stable/4622297
- 127. Tran, P. H. H., & Tran, T. H. (2023). Stock markets' reaction to COVID-19: the joint impact of uncertainty avoidance culture and government response–evidence from emerging countries. *Review of Behavioral Finance*, 15(1), 55–64. https://doi.org/10.1108/RBF-08-2021-0146
- 128. Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232. https://doi.org/https://doi.org/10.1016/0010-0285(73)90033-9
- 129. Tversky, A., & Kahneman, D. (1981). The Framing of Decisions and the Psychology of Choice. *Science*, 211(4481), 453–458. https://doi.org/10.1126/science.7455683
- 130. Twitter. (2023, May 15). A wealth of conversation: How Twitter connects those interested in finance and investment. Https://Marketing.Twitter.Com/En\_gb/Insights/How-Twitter-Connects-Those-Interested-in-Finance-and-Investment-.
- 131. Volkova, S., Wilson, T., & Yarowsky, D. (2013). Exploring demographic language variations to improve multilingual sentiment analysis in social media. *Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, 1815–1827.
- 132. Wolstenholme, E. F. (1999). Qualitative vs quantitative modelling: the evolving balance. *Journal of the Operational Research Society*, 50(4), 422–428. https://doi.org/10.1057/palgrave.jors.2600700
- 133. Wu, C.-H., & Lin, C.-J. (2017). The impact of media coverage on investor trading behavior and stock returns. *Pacific-Basin Finance Journal*, 43(1), 151–172. https://doi.org/https://doi.org/10.1016/j.pacfin.2017.04.001
- 134. Yahoo! Finance. (2023, May 2). *PSI-20, EDP and EDPR Historical Data*. Https://Finance.Yahoo.Com/.
- 135. Younis, E. M. G. (2015). Sentiment Analysis and Text Mining for Social Media Microblogs using Open Source Tools: An Empirical Study. *International Journal of Computer Applications*, 112(5), 44–48.

- 136. Yu, H., & Hatzivassiloglou, V. (2003). Towards Answering Opinion Questions: Separating Facts from Opinions and Identifying the Polarity of Opinion Sentences. *Proceedings of the 2003 Conference on Empirical Methods in Natural Language Processing*, 129–136. https://doi.org/10.3115/1119355.1119372
- 137. Zhang, X., Fuehres, H., & Gloor, P. A. (2011). Predicting Stock Market Indicators Through *Twitter* "I hope it is not as bad as I fear." *Procedia Social and Behavioral Sciences*, 26, 55–62. https://doi.org/https://doi.org/10.1016/j.sbspro.2011.10.562

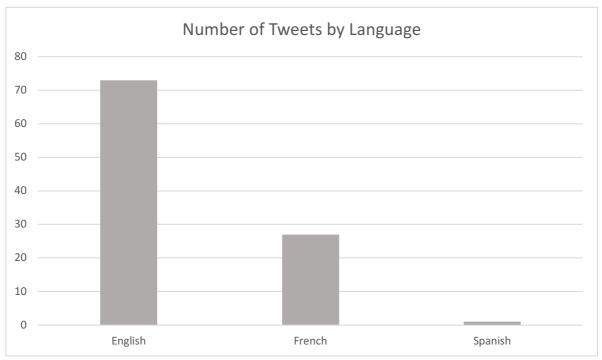
## Appendix



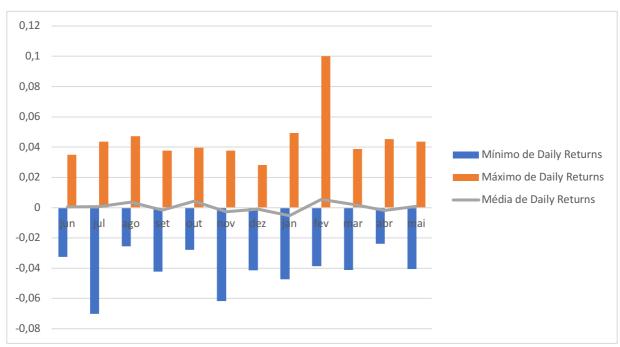
Graph 2 - Number of Tweets Made by Day of the Week. Source: Twitter Search API



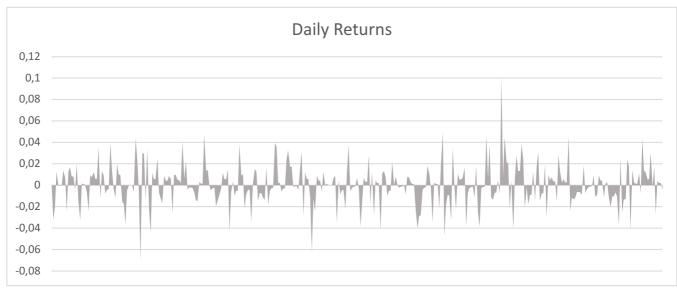
Graph 3 - Number of Tweets Made by Month. Source:  $\mathit{Twitter}$  Search API



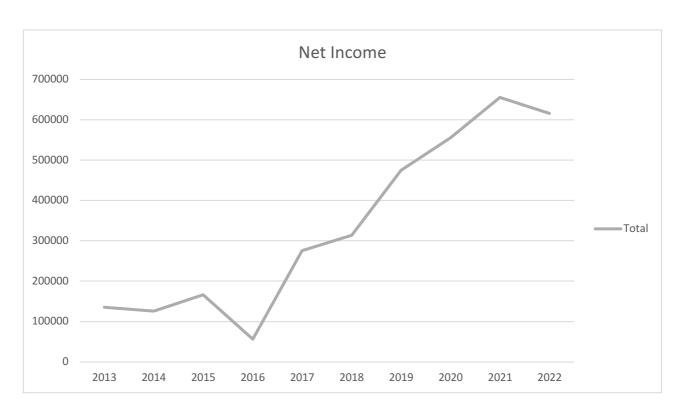
Graph 4 - Number of Tweets Made by Language. Source: Twitter Search API



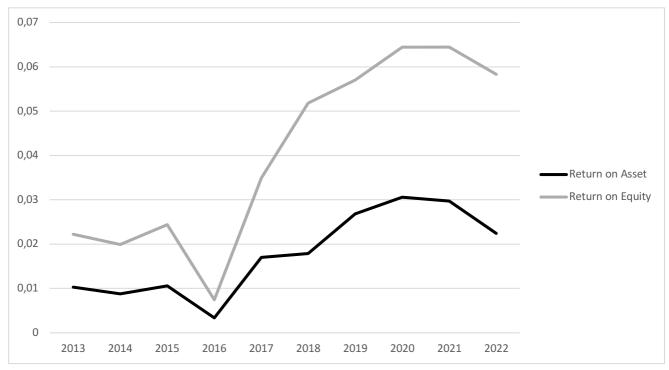
Graph 5 - Min, Max, and Average Daily Returns of EDPR Stocks. Source: https://finance.yahoo.com/



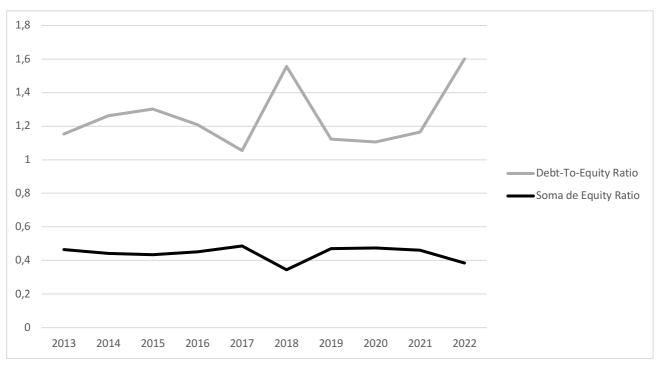
Graph 6 - Daily Returns of EDPR Stock. Source: https://finance.yahoo.com/



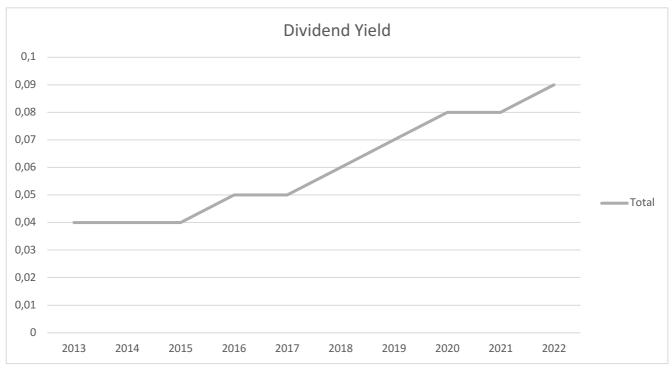
Graph 7 - Net Income of EDPR



Graph 8 - ROE and ROA of EDPR



Graph 9 - Debt-To-Equity and Equity Ratio of EDPR



Graph 10 - Dividend Yield of EDPR