Analyzing Moroccan Stock Market using Machine Learning and Sentiment Analysis

Hind Bourezk, Amine Raji, Nawfal Acha, Hafid Barka
EM2TI, National Institute of Posts and Telecommunications
Rabat, Morocco
h.bourezk@gmail.com; amine.raji.perso@gmail.com; nawfal.acha@gmail.com; hafid.barka@gmail.com

***Abstract*—Behavioral Finance studies demonstrated over the last decade that stock market can be driven by emotions for market participants. On the other hand, online sentiment tracking over social media network and news showed promising results in predicting financial markets. Hence, measuring investor sentiment has become a key research issue in financial predictions. In this paper, we present our methodology for collecting, analyzing and inferring sentiments from several information sources regarding Casablanca Stock Exchange Market. With this data we apply sentiment analysis and machine learning algorithms to infer the relationship between the general public view regarding a stock and its evolution within the stock market.**

***Index Terms*—Investor sentiment, Big data, Sentiment analysis, Behavioral finance, Machine learning, Stock market news.**

I. Introduction

Stock market prediction has been an interesting field of research for the last decades. The Efficient Market Hypothesis (EMH) asserts that financial market valuations includes all existing, new, and even hidden information, since investors act as rational agents who seek to maximize profits [8]. However, behavioral finance states that financial markets can be driven by emotions and claims the important role of behavioral and emotional factors, including social mood in financial decision-making [18].

The past decade has seen an explosion in alternate sources of information available to stock market participants. With the advent of the Internet, and especially social media, individual investors increasingly rely on each other while making investement decisions.

In this paper, we test the hypothesis based on the premise of behavioral economics, that the emotions and moods affect the decison making process of investors. Hence, leading to a direct correlation between "public sentiment" and "market sentiment".

We present our methodology to collect, analyze and process sentiments from financial data extracted from different online sources of news publishing about the studied market (Casablanca Stock Exchange). Computed sentiments are then correlated with the evolution of the stocks to expose at what extent the available news can be leveraged to predict the market evolution.

978-1-7281-4979-0/20/$31.00

We perform sentiment analysis on pubicly available financial news to detect the public mood. We use moroccan indices values to predict future stock movements from Casablanca Stock Exchange (CSE).

This paper is organized as follows: Section 1 presents background notions used in this paper. Section 2 summarize related works that address the same problem. Section 3 presents data sources used in this study and sentiment analysis methodology. Section 4 presents and discusses our first results. Last section concludes with main challenges and a discussion of future works.

II. Background concepts

A. News media and Investor sentiment

An important part of behavioral finance is the relation between investor sentiment and stock market returns [26]. The important role of investor sentiment in financial markets has received notable attention from economists lately [13], [19]. The empirical literature examined the effects of sentiments extracted from online networks on the stock market and show that both investor sentiment and investor behavior have significant impacts on stock returns [1], [28], [32] Many researches [2], [32], [33] argue that sentiment measures do have predictive power with respect to equity returns.

[24] argues that news plays a crucial role in buying and selling decisions among traders, who constantly react to new information and show that the news media are important players in creating market sentiment and similar thinking as it spreads ideas and, thus, can significantly contribute to herding behavior and influence price movement on financial markets.

[21] examine the impact of news media sentiment on financial market returns and volatility in the long-term by taking the hypothesis that the way the media formulate news to the public engenders different perceptions and,hence, induces different investor behavior.

Moreover, [23] found that investors respond to earnings news differently according to sentiment and that the stock price reaction to positive earnings surprises is significantly greater for firms with high sentiment.

In the other hand, [17] investigate the dependence patterns in 24 European equity markets by examining whether global financial crisis and sovereign debt crisis trigger contagion. Their results reveal heterogeneity in the time-varying dependence and across markets.

[10] shows that short-term stock price movements can be predicted using financial news articles and found that definite predictive power for the stock price movement in the interval starting 20 minutes before and ending 20 minutes after news articles become publicly available.

B. Sentiment analysis

Sentiment analysis or opinion mining is one of the most important points in big data research [6]. It describes various computational techniques focused to discover, extract and distil the human emotions, feelings or opinions from textual information within the web content towards the certain entities [9]. Broadly, there is two types of methods for sentiment analysis: machine learning method and lexicon-based methods [14] as shown in figure 1.

Machine learning methods can be further divided into supervised and unsupervised approaches but it often rely on supervised classification approaches, where sentiment detection is framed as a binary (positive or negative). This approach requires labeled data to train classifiers [12].

For supervised approaches, we need two sets of annotated data, one each for training and testing [20]. A training set is used by an automatic classifier to learn the differentiating characteristics of documents, and a test set is used to validate the performance of the automatic classifier [29]. Some of the most applied classifiers for supervised learning are Decision Tree (DT), SVM, Neural Network (NN), Naïve Bayes, and Maximum Entropy (ME).

An advantage of machine learning is its ability to be defined in detail for specific contexts. On the contrary, a disadvantage is its low applicability to new data due to costly text tagging or unavailability of sufficient examples from which the machine learning program can learn [11], [12].

For exemple, [3] claim that sentiment analysis can be cast as a classification problem where the task is to classify messages into two categories depending on whether they convey positive or negative feelings. They discuss the challenges that Twitter data streams pose, focusing on classification problems, and then consider these streams for opinion mining and sentiment analysis.

On the other hand, lexical based approaches use a predefined list of words, lexicon or dictionary where each word is associated with a specific sentiment. Dictionary based approach will use an existing dictionary, which is a collection of opinion words along with their positive or negative sentiment strength. The lexical methods vary according to the context in which they were created [12].

One advantage of this approach is that pre-tagging is not necessary, but it is a shortcoming when specific context needs to be analyzed because a unique dictionary can be unavailable [12], [15].

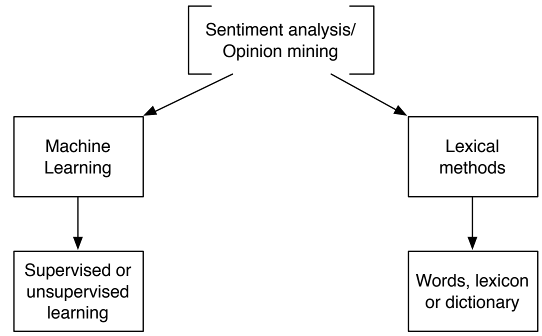

Fig. 1. Sentiment classification process

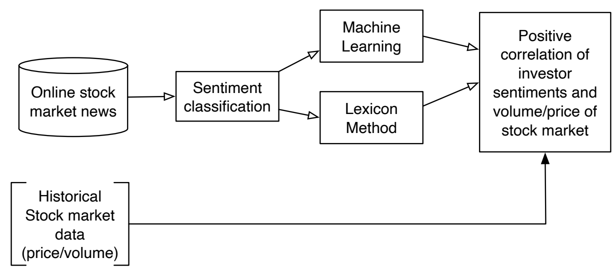

Fig. 2. Data collection and analysis process

Mining is used to help people to extract valuable information from large amount of data. Sentiment analysis focuses on the analysis and understanding of the emotions from the text patterns. It identifies the opinion or attitude that a person has towards a topic or an object and it seeks to identify the viewpoint underlying a text span [16].

The process of sentiment analysis consists of several steps illustrated in figure 2. Firstly, big data has to be collected, secondly web content should be transformed to text content only, thirdly data transformation from a qualitative into a quantitative nature based on the machine learning or lexical method [6].

III. RELATED WORKS

A vast body of empirical literature examined the effects of sentiments sensed online on the stock market and indicate that both investor sentiment and investor behavior have significant impacts on stock returns [1], [28]. Others argue that sentiment measures do have predictive power with respect to equity returns [2], [32], [33].

[21] construct a monthly media sentiment indicator by taking the ratio of the number of newspaper articles that contain predetermined negative words to the number of newspaper articles that contain predetermined positive words in the headline and/or the lead paragraph. The results indicate that pessimistic news media sentiment is positively related to global market volatility and negatively related to global market returns 12 to 24 months in advance.

However, [25] studies how the simultaneous use of financial news articles with different levels of relevance to the target stock can give an advantage in financial news-based forecasting. They employ the Global Industry Classification Standard (GICS) to divid stocks by industries and sectors and for assigning their corresponding news articles to five categories and found that the simultaneous usage of five news categories improves the prediction performance in comparison with methods based on a lower number of news categories.

Given a stock price time series, for each time interval [10] classify price movement as ”up,” ”down,” or (approximately) ”unchanged” relative to the volatility of the stock and the change in a relevant index. Each article in a training set of news articles is therefore labeled ”up,” ”down,” or ”unchanged” according to the movement of the associated stock in a time interval of the publication of the article. They trained a naïve Bayesian text classifier to predict which movement class an article belongs to.

Moreover [22] examines a predictive machine learning approach for financial news articles analysis using several different textual representations by investigating 9,211 financial news articles and 10 millions stock quotes covering the S&P 500 stocks during a five week period. They found that a Proper Noun scheme performs better than the de facto standard of Bag of Words in all three metrics.

In the other hand, Twitter has become an important tool for businesses and individuals to communicate and share information [5]. Twitter sentiment analysis is not an easy task as a tweet can contain a significant amount of information in very compressed form, and simultaneously carry positive and negative feelings [3]. Authors in [31] used a novel investor sentiment proxy extracted from Twitter, they investigate whether investor sentiment has predictive power for stock returns in 10 international stock markets by using Granger causality tests. Also, authors in [4] considered a collection of public tweets and analysed the text content of daily Twitter feeds by two mood tracking tools, namely ”OpinionFinder” (OF) that measures positive vs. negative mood and *Google-Profile of Mood States (GPOMS)* that measures mood in terms of 6 dimensions (Calm, Alert, Sure, Vital, Kind, and Happy).

Instead of predicting stock price movements [27] examined the relation between Facebook data and investors decision making in stock markets with a unique data based on investors transactions on Nokia. Investors were divided into five groups according to their sector codes and they use *Social Data Analytics Tool (SODATO)* to collect daily numbers of posts from Nokia's Facebook wall.

Other works [7] measured investor sentiments with *Financial and Economic Attitudes Revealed by Search (FEARS)* by aggregating the volume of Internet search queries and conducted different tests to show how FEARS predicts market returns and daily volatility. On the other hand, authors in [30] argue that higher investors attention is related to higher market efficiency using Google queries on broad market indices like the Dow Jones Industrial Average(DJIA), S&P 500, NASDAQ and other indices.

From this short literature review, we can argue that several studies use different data sources (news, tweets, Facebook comments, search engine queries) to infer investor sentiments then find that there is/isn't correlations between investors sentiments and market behavior.

In this paper, we choose a different approach to study the general public perspective over different stocks by analyzing the large body of content published over several financial news websites. Such news websites are frequently consulted by Moroccan individual investors on a daily basis to get informed about different stocks. The next section presents the data collection used in this research and how we computed investors sentiments from it.

IV. Data processing and sentiment analysis

This section lists data collection and analysis techniques used as well as the tools used to extract investors sentiments from raw internet data.

In this research, we used two main datasets:
1. Publicly available raw financial news from their online websites.
2. Moroccan stock market indices extracted from the site of ”La bourse de Casablanca” (Casablanca Stock exchange).

A. Data preprocessing

Financial News Articles: News articles are considered as raw materials that need to be preprocessed before aligning them with the historical stock prices data. Given a set of news articles, the aim is to extract the useful information in them. For example, we are interested in the information that will trigger the stock price changes. One of the key challenges in data preprocessing is to know how to extract useful information from the raw data. Some researchers align the whole contents of news articles to the stock prices movements. We believe that only certain terms from the news articles may have high predictability to trigger the stock price changes.

The data that was obtained from above mentioned sources had been pre-processed to make it appropriate for reliable analysis. We used a python script to collect published entries related to Moroccan stock market using six different online news websites:

- leconomiste.com (450 entry)
- media24.ma (325 entry)
- boursenews.ma (441 entry)
- lematin.ma (624 entry)
- lesecho.ma (177 entry)
- leboursier.ma (195 entry)

Historical morrocan stock market Prices: We used daily data of morrocan stocks prices. The data was obtained from June 2019 to October 2019 and includes the open, close, high and low values for a given day.

B. Data mining

Sentiment classification of stock market news involves identifying positive and negative news articles, and is an emerging technique for making stock trend predictions which can facilitate investor decision making. In this paper, we propose two types or methods of sentiment classification applied to online stock market news articles to predict the movements of stock prices in the Casablanca Stock Exchange (BVC).

After extracting the data, we conduct sentiment analysis each entry of text using machine learning techniques. Python happens to be one of the best programming language choices when it comes to machine learning and textual analytics as it is easy to learn, is open source, and is effective in catering to machine learning requirements like processing large data sets and performing mathematical computations. Natural Language ToolKit (NLTK) is one of the popular packages in Python that can aid in sentiment analysis. For this, we used two approaches:

Analyzing sentiments using lexicon: NLTK comes with an inbuilt sentiment analyser module *nltk.sentiment.vader* that can analyse a piece of text and classify the sentences under positive, negative and neutral polarity of sentiments. Because we are dealing with text entries in french not English, we used *Stanford Log-linear Part-Of-Speech Tagger* that support french language, we can produce polarity tags for each sentence or our data. The advantage of such method is that the analysis is fully automated, however, it falls short regarding its accuracy due to the variations and figures of speech used to qualify different stocks (figurative descriptions). When running the sentiment intensity analyzer, sentiment proportions for individual sentences are obtained as shown in figure 3. Under each sentence, our program measures the sentiment based on the vocabulary used to qualify each subject (text in red font) and produce an estimated sentiment using (neg, neu, pos and compound) metrics.

Using Naive Bayes Classifier: Apart from Vader, one can create one's own classification model using Naïve s Bayes Classifier. In the machine learning context, Naïve s Bayes Classifier is a probabilistic classifier based on Bayes' theorem that constructs a classification model out of training data. This classifier learns to classify the reviews to positive or negative using the supervised learning mechanism. The learning process starts by feeding in sample data that aids the classifier to construct a model to classify these reviews. The drawback is the we need some training data to train our model which can be tedious and time consuming. In our example, we trained our classified using a training dataset of about 200 entries, each tagged with one of sentiment tags: "positive", "negative", "neutral" (figure 4 show an excerpt of data used to train the model.

V. Discussion and first results

Figure 5 plots the first results obtained from analyzing sentiments over more than 2200 articles related to an excerpt of five stocks of Casablanca Stock Market. We only retained sentences that list keywords related to such stocks using their respective stock ids.

We investigated the causative relation between public mood as measured from a large scale collection of financial news and the moroccan stock indices values. Our results show that firstly public mood can indeed be captured from the large-scale news mood by means of simple natural language processing techniques, as indicated by the responses.

We can observe the frequency of negative and positive terms extracted from Moroccan News related to Stock market using six different online news websites for a period of one month.

As it is shown in the figure, the BMCI Stock show the most negative text content within 2200 entries, while the best positives opinions was for M2M stocks.

Those results can be used to predict the variations of stock prices in the Casablanca stock exchange and is a better opportunity for institutional investors for managing their stock portfolios.

VI. Conclusion and future works

In this paper, we performed sentiment analysis on public data related to Moroccan stock market collected used automated scripts from several online sources. Two different techniques of sentiment analysis were tested to evaluate which will provide the best accurate sentiment evaluation. Based on our experimentations, we retained the supervised machine learning algorithm of Naive Bayes because of the subtleties and diversity of the french language used to qualify stocks.
In future work, we will collect larger data set and perform several correlation analysis between sentiments and the evolution of the stocks over the same period of time.

This study systematically explores the interactions between media content and stock market We investigated the causative relation between public mood as measured from a large scale collection of news and the stock values of (CSE).

Our results show that firstly public mood can indeed be captured from newspapers by means of simple natural language processing techniques. Secondly, among the observed polarity, the negative mood had huge impact on the fluctuations of stock market indices.

Finally, its worth mentioning that our analysis doesn't take into account many factors. Firstly, our dataset doesn't really map the real public sentiment, it only considers major financial and economic newspapers. It's possible to obtain a higher correlation if the actual mood is studied. It maybe hypothesized that people's mood indeed affect their investment decisions, hence the correlation. But in that case,there's no direct correlation between the people who invest in stocks and who read newspapers more frequently, though there certainly is an indirect correlation investment decisions of people may be affected by the moods of people around them,ie. the general public sentiment. All these remain as areas of future research.

```
Les principaux indices de la Place ont terminé en territoire négatif vendredi.
neg: 0.802, neu: 0.198, pos: 0.0, compound: -0.86,
Le Masi a dévissé de 1,51% à 11.087,32 points sous la pression des secteurs
Banques, Assurances, Immobilier, Bâtiment et matériaux de construction, Electricité
et Télécommunications.
neg: 0.710, neu: 0.02, pos: 0.0, compound: -0.64,
```

Fig. 3. Sentiment tagging example

```
train = [(" Le Masi accentue ses pertes .", "neg"),
    (" Léger rebond des volumes grâce à Maroc Telecom.", "pos"),
    (" Les secteurs télécoms, BTP et Banques font dévisser le Masi", "neg"),
    (" Le Masi lâche 0,25%.", "neg"),
    (" Les secteurs vedettes soutiennent le Masi.", "pos"),
    (" La belle dynamique se poursuit", "pos"),
    (" La Bourse a retrouvé des couleurs, pourvu que ça dure", "pos"),
    . . . |
]
```

Fig. 4. Excerpt of data used to train the classifier

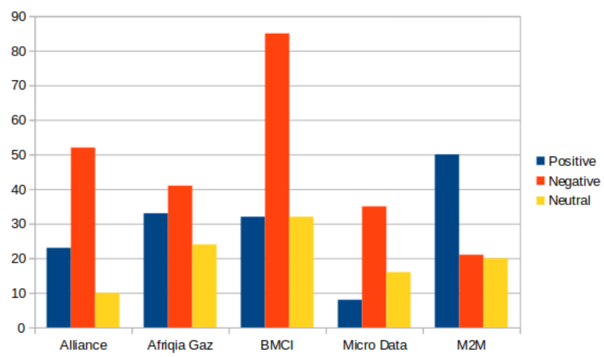

Fig. 5. Excerpt of computed sentiments for 5 stocks

References

[1] Baker, M., Wurgler, J.: Investor sentiment in the stock market. Journal of Economic Perspectives (December) (2006)

[2] Baker, M., Wurgler, J.: Investor sentiment in the stock market. Journal of Economic Perspectives 21(2), 129–151 (Spring 2007)

[3] Bifet, A., Frank, E.: Sentiment knowledge discovery in twitter streaming data. International conference on discovery science, Springer (2010)

[4] Bollen, J., Mao, H., Zeng, X.: Twitter mood predicts the stock market. Journal of Computational Science 2 (October 2010)

[5] Brown, E.: Will twitter make you a better investor? a look at sentiment, user reputation and their effect on the stock market. Twitter Sentiment for Investing Decision Support (January 2012)

[6] Bukovina, J.: Social media big data and capital markets—an overview. Journal of Behavioral and Experimental Finance 11, 18–26 (2016)

[7] Da, Z., Engelberg, J., Gao, P.: The sum of all fears investor sentiment and asset prices. Review of Financial Studies (October 2014)

[8] Fama, E.: Efficient capital markets: A review of theory and empirical work. Journal of Finance 25, 2,, 383–417 (1970)

[9] Fang, X., Zhan, J.: Sentiment analysis using product review data. Journal of Big Data a SpringerOpen Journal (2015)

[10] Gidófalvi, G.: Using news articles to predict stock price movements (Juin 2001)

[11] Godsay, M.: The process of sentiment analysis: A study. International Journal of Computer Applications (0975–8887) 126(7) (September 2015)

[12] Goncalves, P.: Comparing and combining sentiment analysis methods. In: In: Proceedings of the First ACM Conference on Online Social Networks. No. 27-38 (2013)

[13] Karabulut, Y.: Can facebook predict stock market activity? In: AFA 2013 San Diego Meetings Paper (2013)

[14] Medhat, W., Hassan, A., Korashy, H.: Sentiment analysis algorithms and applications: a survey,. Electrical Engineering 5(4), 1093–1113 (DEcember 2014)

[15] Michelle Annett, G.K.: A comparison of sentiment analysis techniques: Polarizing movie blogs. In: In: 21st Conference on Advances in Artificial Intelligence. pp. 25–35 (2008)

[16] Mouthami, K., Nirmala Devi, K., Murali Bhaskaran, V.: Sentiment analysis and classification based on textual reviews. International Conference on Information Communication and Embedded Systems (ICICES) (2013)

[17] Nitoi, M., Pochea, M.: Time-varying dependence in european equity markets: A contagion and investor sentiment driven analysis. Economic Modelling (2019)

[18] Prechter, R.: The wave principle of human social behavior and the new science of socionomics. New Classics Library (1999)

[19] Qadan, M., Nama, H.: Investor sentiment and the price of oil. Energy Economics 69, 42–58 (January 2018)

[20] Ravi, K., Ravi, V.: A survey on opinion mining and sentiment analysis: Tasks, approachesand applications. Knowledge-Based Systems pp. 14–46 (2015)

[21] Roman, K., Elizaveta, M.: News media sentiment and investor behavior. CFS Working Paper Series 492 (2014)

[22] Schumaker, R., Chen, H.: Textual analysis of stock market prediction using breaking financial news: The azfintext system. 12th Americas Conference on Information Systems (AMCIS) (2009)

[23] Seoka, S., Cho, H., Ryu, D.: Firm-specific investor sentiment and the stock market response to earnings news. The North American Journal of Economics and Finance 48, 221–240 (April 2019)

[24] Shiller, R.: From efficient markets theory to behavioral finance. Journal of Economic Perspectives 17(1), 83–104 (2003)

[25] Shynkevich, Y., McGinnity, T., Coleman, S., Belatreche, A.: Predicting stock price movements based on different categories of news articles. IEEE Symposium Series on Computational Intelligence (2015)

[26] Siganos, A., Vagenas-Nanosa, E., Verwijmeren, P.: Facebook's daily sentiment and international stock markets. Journal of Economic Behavior and Organization pp. 730–743 (2014)

[27] Siikanena, M., Baltakysa, K., Kanniainena, J., Vatrapub, R., Mukkamalab, R., Hussain, A.: Facebook drives behavior of passive households in stock markets. Finance Research Letters (May 2018)

[28] Tetlock, P.: Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance LXII(3) (June 2007)

[29] Vinodhini, G., Chandrasekaran, R.: Sentiment analysis and opinion mining: A survey. International Journal of Advanced Research in Computer Science and Software Engineering 2(6) (June 2012)

[30] Vozlyublennaia, N.: Investor attention, index performance, and return predictability. Journal of Banking and Finance (2014)

[31] Wanhai, Y., Yawei, G., Cheng, P.: Twitter's daily happiness sentiment and the predictability of stock returns. Finance Research Letters (2017)

[32] Yang, C., Zhou.L: Individual stock crowded trades, individual stock investor sentiment and excess returns. North American Journal of Economics and Finance 38, 39–53 (2016)

[33] Zhong, X., Da, Z., Wen, J.: Investor sentiment and its nonlinear effect on stock returns-new evidence from the chinese stock market based on panel quantile regression model. Economic Modellingo (50), 266–274 (2015)