Ca' Foscari
University
of Venice

Second Cycle Degree programme

in Accounting and Finance
LM-77 (Management)

Final Thesis

Extracting Information from Social Media to Track Financial Markets

Supervisor
Prof. Diana Barro

Graduand
Mirco Guidi
Matriculation Number: 854793

Academic Year
2015 / 2016

Table of Contents

INTRODUCTION

The internet has become an unlimited source of information about almost any topic. Its spreading in every aspect of our lives, allow us to access data in real-time, everywhere we are. A particular implementation of the internet has developed in the 21st century: social media. With this form of communication, users can broadcast their thoughts to a global audience at almost no price. For the first time in human history, it is possible to know people's thoughts, emotions, and moods about any topic. All this information can be collected in real time and with very small costs.

For these reasons, researchers have begun to use social media data to analyze real-world phenomena, like brand popularity [De Vries 12], elections [Lampos 12] and others. Our focus is on the financial markets, and how the analysis of data from Twitter and other comparable data sources, such as Google Trends, can help us to track and predict them. Lately, many studies have been focused on this area of interest, like [Counts 11], [Preis 13], [Zheludev 15], however there are still many questions regarding the potential predictive power of social media information. Researchers do not agree with each other on the fact that this kind of information can anticipate financial markets. Moreover, there is still uncertainty about the quantity and quality of information we can extract from social media.

This new area of research is still in its infancy. It presents computational challenges due to the size of the dataset considered and it is often open to skepticism. Much work is still needed, especially regarding how textual information from social media can be transformed and used as predictive indicators, and if they contain indeed any predictive power.

Of particular interest could be the possibility of building predictive investment strategies based on indicators generated using signals from social media. Nevertheless, this does not represent the goal of this study. Rather we want to focus on the issues that arise when we try to extract information from social media data and try to investigate the extent to which these data can contains information useful to track financial markets' trend. To do this we will conduct two analysis. The first one employs data from Twitter and aims to investigate if the mood

that emerges from this social network about Apple Inc. is somehow correlated with the stock indicators. The second one focuses on search volume of financial terms and how this information can be useful to track the performances of financial indexes such as S&P 500 and FTSE 100.

This thesis is organized as follows. Chapter 1 will introduce the concept of information extraction, analyzing the main features of this activity. It contains a section that deals with a particular type of information extraction, sentiment analysis. We will use this technique in our Twitter-based analysis. Chapter 2 will deal with information extraction from social media and contains a part of literature review about tracking financial markets using social media data. Chapter 3 contains the practical part of this work: the Twitter-based analysis and the search volume-based analysis. First, we will introduce the methodology we used to extract the data we need, then we will explain how we have conducted our analysis: data set-up, correlation tests, and results. Chapter 4 provides conclusions, discussing difficulties and limits we met during this study and directions for future research.

1 INFORMATION EXTRACTION

During the last decades, the world witnessed a rapid growth of textual information available in digital form and delivered through networks, in particular on the Internet. The great majority of this information is made by unstructured data, transmitted through online news, corporate and government reports, legal and court acts, medical alerts, social media communication, etc. so it is hard to search in. For this reason, the need for effective and efficient techniques for analyzing these data started to grow. Thus, first Information Extraction technologies have emerged, with the goal of discovering valuable and relevant knowledge from these free-text data and make them accessible in the form of structured information.

According to [Piskorski 13] "the task of Information Extraction (IE) is to identify a predefined set of concepts in a specific domain, ignoring other irrelevant information, where a domain consist of a corpus of texts together with a clearly specified information need". IE is about getting structured information from unstructured text. Natural language has some implicit internal structures; researchers leverage those structures to build Information Extraction systems that convert unstructured data into specific values.

Table 1 shows a simplified example of values and features extracted from this fragment of an online news article: *"Three bombs have exploded in north-eastern Nigeria, killing 25 people and wounding 12 in an attack carried out by an Islamic sect. Authorities said the bombs exploded on Sunday afternoon in the city of Maiduguri".*

Feature	Value
Type	*Crisis*
Subtype	*Bombing*
Location	*Maiduguri*
Dead-Count	*25*
Injured-Count	*12*
Perpetrator	*Islamic sect*
Weapons	*Bomb*
Time	*Sunday Afternoon*

Table 1: *Example of automatically extracted information from a news article on a terrorist attack [Piskorski 13].*

The process of extraction of this structured information requires some sort of predefined structures, like nouns denoting a person or a group, geographical references and expression related to numbers. In addition, most of the time, some domain-specific knowledge is required, in order to aggregate the partial information into a more structured form. The complexity of Information Extraction task is mostly due to the ambiguity of natural language: we can use many different ways to express the same fact [Huttunen 02] and a great part of significant information may be implicit, which means we need a lot of background knowledge about the domain to perform an effective Information Extraction from natural language.

To better understand IE and its applications, it may be useful to compare it with Full Text Understanding and Information Retrieval, since they are similar to each other.

1.1 Full Text Understanding

Full Text Understanding implies that a computer fully understands natural language, taking into consideration all possible interpretations and variations in meaning. Information Extraction instead, is a more limited task [Grishman 97] and its scope is narrower than the scope of Full Text Understanding. Nonetheless, IE applications are still very useful if the user is not interested in the full text but just in specific features. As a consequence, IE needs less sophisticated analysis tools and less knowledge engineering to operate. In general, it is a more simple issue than Full Text Understanding, since IE can ignore much of the information text [Tianhao 02].

1.2 Information Retrieval

According to [Smeaton 98], Information Retrieval (IR) is "the task of finding documents, usually texts, which are relevant to a user's information need". One of the best IR system for the web is Google. Just like Google, the output of an IR system is a subset of documents that are relevant to particular queries, usually based on key-word search. In general, Information Extraction aims to extract pre-defined features from documents while Information Retrieval extracts documents themselves. As [Adam 01] states *"Both IR and IE are difficult because they must overcome the ambiguities inherent in language"*. However, IE is more complex since it needs a certain degree of knowledge about the domain, in order to establish relationships between features.

Full Text Understanding, Information Extraction and Information Retrieval are all part of Textual Information Access. All of them aims to understand textual documents in some manner, but the degree of understanding needed to pursue each task is quite different [Tianhao 02]. The complexity and the level of details required, increase from IR to IE to Full Text Understanding.

1.3 Data Sources

Structured data. It refers to those data that can be stored in a database, in tables with rows and columns. The role of the words of this kind of texts is easily predictable and mapping these data in pre-defined fields is not a hard task. They represent the easiest way to manage information. That is why the focus of Information Extraction research is not on structured data but on unstructured and semi-structured ones [Tianhao 02].

Semi-structured data. It refers to information that cannot fit in a relational database because the format of the structure is imprecise, but that do have some kind of organizational properties anyway. According to [Eikvil 99] *"semi-structured text is often ungrammatical and does not follow a rigid format. It sometimes does not contain full sentences either"*. These features make semi-structured data not suitable for methods that work with structured and narrative text. An example of unstructured data is a Word Document with metadata tags. These tags represent the document's content and make it easier to found the document when people search these words. However, the document still lacks the structured organization of a database.

Unstructured data. It refers to information that does not present a pre-defined structure and it is not organized in a pre-defined manner. They represent around 80% of all data and often include text and multimedia content [Ronk 14]. Unstructured text or narrative text is text without explicit formatting, examples can be newspapers, online articles, emails, business reports, social media messages and posts, website content etc. According to [Eikvil 99] one of the main goals of Information Extraction research is to create systems able to catch relevant features from free natural language text. In order to do that IE system needs to apply rules to the inputs to extract meaningful features and their values. These rules can be hand coded or developed from annotated training sets. The last approach allows the IE system to work in a sort of automatic mode, however in this way performances are poor due to the fact that narrative text is usually very complex. *"Unrestricted natural language understanding is a long way from being solved"* [Eikvil 99], however IE

systems on narrative texts can still provide relevant outcomes because the specified features can have regular structures. The analytical section of this research will take into consideration this kind of data.

1.4 Learning Methods

Several learning methods can be employed in Information Extraction; they have been developed starting from the late 80's, reaching more efficiency and a higher degree of automation [Piskorski 13].

Knowledge Engineering. All the early IE system were developed using this methodology [Appelt 99]. With this approach, human experts build rules for the systems by hand, using their knowledge about the application domain. It requires a lot of time to build and maintain such a system and a crucial role in the performances of it is played by the skills of the experts. In this case, the machine does not learn anything from data, it is only a tool to implement what humans expert knows. A major limit of this method it is lack of modularity, systems are usually monolingual and not easily adaptable to new scenarios [Piskorski 13]. Nevertheless, most of the best performing Information Extraction systems are Knowledge Engineering based.

Supervised Learning. The first step towards automation has been taken developing a new method called Supervised Learning. The focus of this method is to shift the human effort away from knowledge engineering and towards annotating training data, to use as an input for machine learning algorithms [Piskorski 13]. While Knowledge engineering requires efforts from both the system developer and the domain expert, Supervised Learning only involves the domain expert's knowledge. This method allows to convert training data into patterns without any human intervention; it represents an improvement compared to the first approach, however the needs for a large amount of manually prepared data, still makes it very time-consuming for human experts.

Semi-Supervised Learning. In order to reduce human expertise involvement, researchers introduced another method called Semi-Supervised Learning. [Basu 02] describes it in this way:

> *"Using Semi-Supervised Learning, a system learns from a mixture of labelled (annotated) and unlabelled data. In many application there is a small labelled data set together with a huge unlabelled data set. It is not good to use only the small labelled data set to train the system because it is well known that when the ratio of the number of training samples to the number of features measurements is small, the training result is not accurate. Therefore, the system needs to combine labelled and unlabelled data during training to improve performance. […] the system extracts patterns from the annotated data, and labels the unannotated data automatically using the patterns. As a result, all data are labelled for the training."*

Semi-Supervised Learning can actually save human efforts and times granting performances as good as the ones with Supervised Learning method.

Active Learning. Active Learning is another method that aims to reduce the human intervention in form of annotations for data. The idea, as described by [Piskorski 13] is that the annotator initially tags a small amount of samples, then based on this set, the learner actively decides which other examples need to be tagged by the human expert in order to maximize gains. The active learner will choose those examples in which its uncertainty is greatest, since it will benefit more if the annotator marks up samples that are very different among themselves, instead of providing redundant information. As described by [Cohn 94] this approach will *"help users select suitable features in order to minimize the number of examples the user must annotate".*

The main difference between Active and Semi-Supervised Learning is that the first approach uses a system to help human experts choose the unannotated text to label, while following the latter approach they manually select the samples to tag, or a small set of annotated text has already been determined.

Bootstrapping or Unsupervised Learning. Like the Active Learning methods, the system developer provides an initial set of annotated samples, but in Bootstrapping the learning process goes on without further supervision [Piskorski

13]. This kind of approach is also called Unsupervised Learning, even though a certain supervision is required, at the beginning - to determine the initial examples - and in the final phase - to check the results and remove "bad" candidates. [Paliouras 99] describes the learning method according to this approach as follows:

> *"The basic approach of Unsupervised Learning includes the following steps. First of all, an Unsupervised Learning system is seeded with a couple of labelled facts or patterns. Next, the system searches a large unannotated corpus for new candidate patterns based on the seeds. After the new patterns are found, the system can use them to uncover additional facts. The system then adds the facts to the seed set. After that, the system is retrained based on the new extended seed set. This process is repeated until no more patterns can be found."*

1.5 Evaluation

Starting from an input text or a collection of texts, it is easy to assess precisely the output of an IE system. The most frequently used evaluation metrics are precision and recall. They measure the system's effectiveness from the user's point of view, assessing if the system generates all the appropriate output (recall) and only the appropriate output (precision) [Piskorski 13]. In order to define them formally, we introduce the following notion, according to [Tianhao 02]:

- TP (true positive) indicates the number of feature values in both target and selected sets;
- TN (true negative) is the number of feature values that are not in the selected set nor in the target set;
- FP (false positive) represents the wrongly chosen feature values in the selected set;
- FN (false negative) are the feature values in the target set that were incorrectly not selected.

Figure 1 from [Manning 99] shows a diagram to make it clearer.

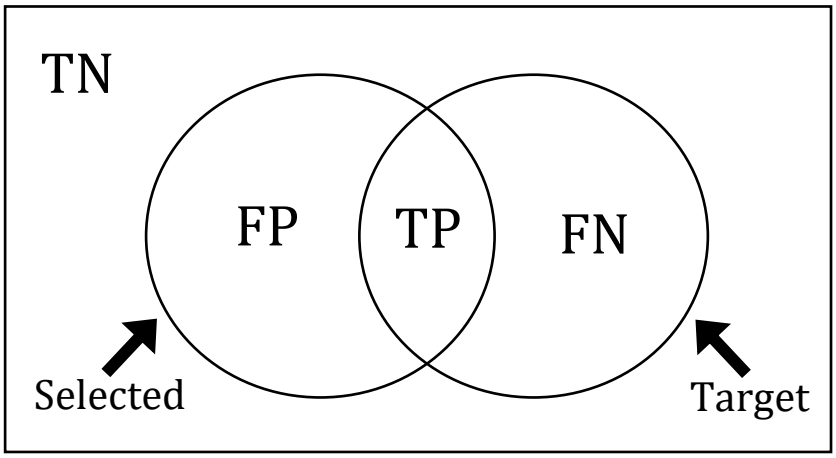


Figure 1: *A diagram of TP, TN, FP and FN [Manning 99].*

Precision (P) and recall (R) are defined as follows:

$$P = \frac{TP}{TP + FP} \qquad R = \frac{TP}{TP + FN}$$

The values of precision and recall are always between 0 and 1. Higher values represent better results. Precision and recall are inversely related: selecting an entire collection of documents will give you a recall of 100%, but a very low precision. On the other hand, choosing only a few relevant documents will give you high precision and low recall. Usually, a precision-recall curve is used to plot the trade-off.

Figure 2 shows an example of precision-recall curve.

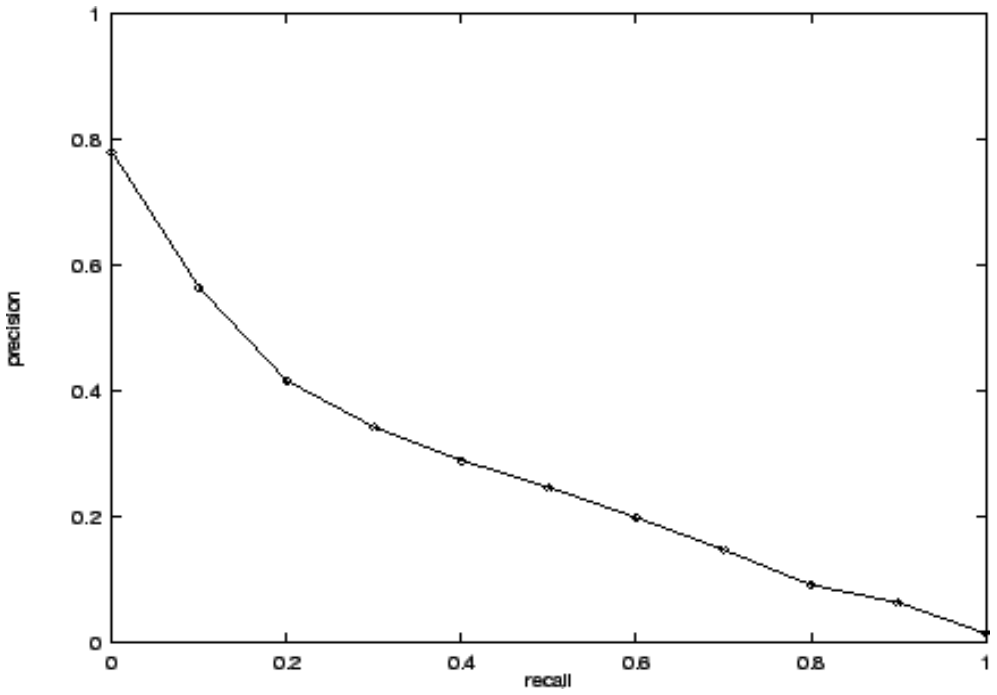


Figure 2: *An example of precision-recall curve [Tianhao 02].*

In order to obtain a more refined and harmonic indicator of IE systems' effectiveness, it is possible to combine precision and recall using the F-measure [Rijsbergen 79].

$$F = \frac{(\beta^2 + 1) * P * R}{(\beta^2 * P) + R}$$

β is a non-negative value, used to adjust their relative weights. When β equals 1, precision and recall have the same weight; if β > 1 recall is more important than precision, on the other hand for lower values of β precision is more important than recall. *"Using the F-measure, it is easy to compare the performance of systems with different precision and recall"* [Tianhao 02].

1.6 Challenges for the Future

According to [Piskorski 13], there are two main directions towards which IE researches will head in the future.

In the past, the majority of researches focused on Information Extraction in English. These days, the growing amount of data available in other languages, generate a shift of the focus on non-English IE systems and language independent IE techniques. Until fews years ago, relatively little work has been done on non-English IE techniques, mainly because of various linguistic phenomena that are not present in English and that contribute to make Information Extraction tasks harder. [Piskorski 13] indicates some of these linguistic "downsides":

- Lack of whitespace, which makes hard to define words' boundaries (eg. Chinese);
- Productive compounding, that complicates morphological analysis (eg. German);
- Complex proper name declension (typical of Slavic languages);
- Zero anaphora, whose resolution is fundamental in the context of IE extraction (typical of Japanese, Romance, and Slavic languages);
- Free word order, that makes harder to extract relevant relations (typical of Germanic, Slavic and Romance languages).

This list is only partial and there are many other factors that make non-English IE systems more complicated, nevertheless it is possible to observe considerable progress in the recent years.

Another focus that distinguishes current researches from the classical IE, is the extraction of information from multiple sources [Piskorski 13]. This is due to the huge amount of information we can access today on the Internet: most facts or stories in fact, do not exist only in an isolated text but are told by many online sources, maybe even in different languages. This redundancy can be exploited to validate facts, and to follow the evolution of the story through time, with

elaborations and modifications. An example of approaching to Information Extraction from multiple sources is pointed out by [Downey 05].

1.7 A Subjective Form of Information Extraction: Sentiment Analysis

After a general overview of Information Extraction, it is useful for the aim of this research to introduce what can be considered a subjective form of IE: Sentiment Analysis. Sentiment Analysis deals with subjective elements, defined by [Wiebe 04] as "linguistic expressions of private states[1] in context". These can be words, sentences, paragraphs or even whole documents, but it is usually recognized that sentiment lies in smaller linguistic units [Pang 08].

According to [Liu 06] sentiment can be either implicit, when the text implies an opinion, or explicit, when the opinion is expressed directly. This last type is the one easier to be analyzed, therefore most of the work done so far uses this kind of sentiment. Sentiment has polarity, which is generally distinguished between positive or negative, but it can also be expressed as a range. It is important not to confuse polarity with strength: *"One may feel strongly about a product being OK, not particularly good or bad; or weakly about a product being very good, because perhaps one owned it for too short time to form a strong opinion"* [Mejova 09]. Another important feature of sentiment is its target. It can be anything, a concept, an object, a person and can be easy to identify (like for product reviews) or more complicated [Mejova 09].

Sentiment Analysis is a complex process that usually involves five different steps, as indicated by [D'Andrea 15]. These steps are:

1. *Data collection*. The first step consists of collecting user-generated data from blogs, social media, etc. These data are unstructured and often disorganized.

[1] As described by [Quirk 10], *private state* is "something that is not open to objective observation or verification".

2. *Text preparation*. The extracted data need to be cleaned before analysis. Non-textual contents and irrelevant contents are discarded.
3. *Sentiment detection*. The extracted text is analyzed: sentences with subjective expression are maintained, those with objective information, like facts, are eliminated.
4. *Sentiment classification.* The selected subjective messages are ranked according to different scales (e.g. positive/negative/neutral, single score, multiple points, etc.).
5. *Presentation of the output.* After the unstructured text has been converted into meaningful information the results need to be displayed in spreadsheets, charts, graphs, etc.

1.7.1 Goals of Sentiment Analysis

Sentiment Analysis embraces several tasks that are usually combined to produce an overall knowledge about the opinion found in the text. These tasks, as pointed out by [Mejova 09], are:

- *Opinion or sentiment detection*, which is the classification of texts as objective or subjective. Usually, this selection is based on the presence of adjectives in sentences.
- *Polarity classification,* that deals with classifying an opinionated text according to two opposite sentiment polarities or locate a position on the continuum between these two polarities [Pang 08]. In the first case, there will be either a positive or a negative classification. In the second case, classification will use a multi-point scale, like the five-star system review for movies. Defining polarity is not always an easy task, sometimes it is straightforward, like in the case of product reviews, in other cases, like news articles, may be more complex (an article can report a "bad" news without actually containing any subjective terms) [Mejova 09].
- *Features extraction,* that refers to the identification of features in the text. Given an object or a topic, features are defined by [Liu 06] as components

or attributes of that item. Breaking down the discussion into features ensures a more elaborate sentiment analysis and a more detailed report about the results.

- *Discovery of the opinion's target*, which difficulty varies on the base of the domain of analysis. For example, it is easy to assume that product reviews deal with the specified product; it is more complex to identify the objects of general writing, like websites, blogs, social networks etc. Sometimes it is possible to find more than one target in a sentiment sentence, like in the case of comparative sentences. These sentences can be identified because they usually contain comparative adjective and adverbs to order objects according to preferences. For example *"more", "less", "better"*, words ending with *–er*, etc. [Liu 06].

Tasks mentioned above can be applied at three main different levels [D'Andrea 15].

Document-level Sentiment Analysis. Classifying a document's opinion is considered the simplest form of sentiment analysis [Feldman 13]. It is required to consider the whole document as a basic information unit and to assume that it contains an opinion on the main object expressed by the text.

Sentence-level Sentiment Analysis. A document may presents different opinions, even about the same entities. In a case like this, a document-level classification is not appropriate, so to have a more fine-grained view of the different opinions in the text we must switch to a sentence-level classification. With this approach, it is possible to classify the sentiment of each sentence. It is typical to use the output of one level as the input of the higher one [Turney 03]. For example, applying sentiment analysis to sentences and then use the outcome to evaluate the document.

Aspect-level Sentiment Analysis. In many cases, people discuss about objects and topics that have many aspects (attributes), and they have different opinions about each aspect [Feldman 13]. This often happens for product reviews. For example, of a smartphone people can judge screen, battery, performances, size, weight, camera etc. Some of these aspects can be reviewed positively while others can be classified as negative. Classification at aspect level does not look at language construction

(sentences, paragraphs, etc.) but focuses on the opinion about specific aspects of entities mentioned within the document. Usually these aspects are explicit, however sometimes are not clearly mentioned in the text but need to be inferred by the sentiment expression that mentions them implicitly [Feldman 13].

1.7.2 Classification of Sentiment Analysis Approaches

As pointed out by [Maynard 11] there are three sentiment analysis approaches, with specific techniques and features and different advantages and limitations: *machine learning approach, lexicon-based approach, hybrid approach*.

Machine learning approach. It uses two sets of documents: a training set and a test set. The first one is used by an automatic classifier to learn the different characteristics of a document while the second set is needed to check the performance of the classification. [D'Andrea 15] identifies three features of machine learning approach:

- Terms presence and their frequency.
- Part of speech information, used for determining the sense of the document.
- Negations, that has the potential to reverse sentiment words and phrases.

The main advantage of this approach is the ability to adapt to specific purposes and contexts, creating ad-hoc trained models. The main downside deals with the high costs of training new data to employ this method in contexts where training data are not already available [D'Andrea 15].

Lexicon based approach. It uses sentiment dictionaries containing predefined lists of words, where each word is linked to a specific sentiment. The analysis consists of matching the words from the dictionary with the data in order to determine polarity [D'Andrea 15]. It is possible to identify three techniques to build a sentiment lexicon:

- Manual construction. It implies building an ad-hoc lexicon related to the context of the analysis. It is a difficult and time-consuming task.

- Corpus-based method. A corpus is a collection of thousands of different texts including novels, academic books and papers, newspapers, blogs, recorded conversation etc. It usually ensures high-accuracy classification.
- Dictionary-based method. This methodology is the most common. It consists of collecting a predefined set of positive and negative words and then grow this set by searching for their synonyms and antonyms. For general sentiment tasks, some prebuilt dictionaries are publicly available such as the LIWC[2] - Linguistic Inquiry and Word Count. Prebuilt dictionaries are easy to use and have been applied to many different environments, however they are often context-dependent and can potentially lead to serious mistakes in research [Rice 13].

An advantage of lexicon based approaches is that general knowledge lexicons have broad term coverage, however the number of words is not infinite, and this could be a problem when working in very dynamic environment. Moreover, with this approach words tend to have a fixed sentiment orientation and score without considering how these words are used in the document [D'Andrea 15].

Hybrid approach. It combines machine learning and lexicon based approach, ensuring an improvement of sentiment classification performance. The main advantages, as affirmed by [D'Andrea 15] are represented by the lexicon/learning symbiosis and the lesser sensitivity to changes of context. The only limitation of this approach is the noisy reviews, that often consider irrelevant words.

[2] http://www.liwc.net/

1.7.3 Sentiment Analysis Applications

There are different application fields for sentiment analysis. [D'Andrea 15] identifies the main as:

- Business;
- Politics;
- Public actions;
- Finance.

On the business field, sentiment analysis has been used in the area of reviews of consumer product and services. Many websites provide automated summaries of reviews about products, using systems like SumView, a web-based platform developed by [Wang 13] that summarizes product reviews and customer opinions. It incorporates reviews from Amazon.com, automatic features extraction and a text field where users can search for desired characteristics of products. In the business domain, sentiment analysis is also used to monitor brand reputation. For example, tools like Tweetfeel[3] allow to perform real-time analysis of tweets that refers to a certain product. Online advertising also employs sentiment analysis, like in the case of Blogger Centric Contextual Advertising, which, as highlighted in his study by [Fan 11], refers to the "assignment of personal ads to any blog page, chosen according to bloggers' interests" [Fan 11]. Online commerce represents another application of sentiment analysis in the business context. It is well known that consumers value others' opinions about products, restaurants, stores, etc. and that is why service like Google introduced star ratings. In this regard, a study by [Kang 12] developed a senti-lexicon that optimize the sentiment analysis of restaurant reviews.

With respect to the politic domain, sentiment analysis has been used to track how voters feel about different issues and proposals of politics. It turned out to be a very practical tool for campaign managers. Examples of studies on sentiment analysis applied to politics can be found in [Wang 12] that describes a system for

[3] http://www.tweetfeel.com

real-time analysis of public sentiment towards presidential candidates in the 2012 U.S. elections, or the more recent research discussed by [Sheth 16], in which Cognovi Labs set up a real-time monitoring of the Brexit campaign.

The public action implementation process represents another field in which sentiment analysis is highly employed, in particular in the context of tracking real-world events. Examples are the monitoring of people's opinions about pending policies and government regulations [Cardie 06] and the analysis of critical information about earthquakes' locations, magnitude [Sakaki 16] or other rapidly evolving situations.

Regarding the finance domain, implementing sentiment analysis using the huge quantity of financial information spread by articles, blogs, and tweets, makes it possible to develop sentiment indicators suitable to predict financial markets. In [Schumaker 12], the authors paired a financial news articles prediction tool, the *Arizona Financial Text* system (AZFinText), with a sentiment analysis classifier. Through this research, they found out that financial news articles have a direct influence on the price of commodities and shares. Another example is contained in [Feldman 13], that presents the Stock Sonar[4], a service developed by Digital Trowel, that collecting, reading and analyzing information from different online sources, is able to show graphically the daily positive and negative sentiment about each stock of companies traded in the US market

Nowadays there is a common aspect that associate all these domains, social media. The opportunity to apply sentiment classification approaches and tools to social media represents an interesting challenge for the future. This innovative form of communication is producing a huge quantity of data, available in real time at almost no cost. However handling these data places many challenges. In the following chapter, we will discuss information extraction from social media, analyzing where and how this kind of data can be employed. We will focus in particular on financial markets.

[4] www.thestocksonar.com

2 INFORMATION EXTRACTION FROM SOCIAL MEDIA

Social media drastically transformed the way people express themselves and communicate. At this very moment, huge amounts of information are being transferred through social media, blogs, and micro-blogging platforms like Facebook and Twitter. These services allow people to communicate, comment and chat about topics of any kind, like current events, politics, products, etc. This trend spread especially in the last 10-15 years, supported by low-barrier cross-platform and global proliferation of mobile devices [Piskorsi 13]. One of the main features that distinguishes social media from conventional sources of information, such as online news, is the fact that in many situations they can provide more up-to-date information, like in the context of natural disasters or live events. The research community showed interest in this field and it started performing some work on social media content, including attempts to Information Extraction from social media. This kind of extraction is more complex than classic Information Extraction. Some of the issues that make this task more challenging, as pointed out by [Piskorsi 13] are:

- Texts are usually very short, e.g. Twitter limits messages to 140 characters;
- Texts are informal and noisy, they often include misspellings, non-standard abbreviations, lack punctuation and capitalization, and do not always contain grammatically correct sentences;
- High uncertainty about the reliability of the sources, that makes information transmitted through social media not always trustworthy, compared to news media.

Since applying standard tools for IE on social media content results in poor performances, taking into consideration the above-mentioned features, a new line of research in Information Extraction has been developed, one with a focus on techniques to extract information from informal and noisy texts. This field is still in its early stages and focuses mostly on analysis in English. Future researches will concentrate on adoption of classical IE techniques to extract information from social

media, and systems to gather and aggregate information extracted from conventional sources with the ones from social media, for example *"using Twitter to enhance events descriptions extracted from classic online news with situational updates"* [Piskorsi 13].

2.1 Tracking Real World Phenomena with Social Media

Real-time data from social media can be used to track real world phenomena across many fields, not only related to financial markets. They have the ability to report public opinion without lags and filters associated with the production of data on real-world phenomena by governments and other agencies [Dodds 11].

Social data have been used to track, forecast and make analysis about many real world events. [Pologreen 08] and [Ginsberg 09] examined the relationship between queries for influenza on Yahoo and Google search engines and actual influenza occurrence, creating a model that can use online searches to detect influenza epidemics in places with many web users. [De Vries 12], analyzing 355 brand posts published on fan pages of 11 international companies, determines how the activities of social media marketing can influence brand popularity. [Carrière-Swallow 13], analyzing the automotive industry in Chile, showed that data about Google search queries can be used to improve the prediction of consumer behavior patterns in emerging markets. Researches about economic variables have been carried on by [Baker 11] that built a job search index based on Google search data, in order to test unemployment rate; and by [Vosten 14] that introduced an indicator for private consumption based on data extracted from Google Trends. [Lampsos 13] estimated patterns of mood variation using sentiments derived from words in Twitter messages, with the goal of tracking moods of people. Social data have been used to track voting intentions during political races, like [Lampos 12] that implemented a sentiment analysis data extracted from Twitter with the goal to predict the percentage of voting intention polls during the General Election of 2010 in the UK. Recently Cognovi Labs set up a real-time monitoring of the Brexit

campaign. Running Twitter chatters through their tools, they have been able to predict the "leave" outcome of the vote six hours before the news got out [Sheth 16]. This result is even more impressive if we think that the more experienced opinion polls have asserted until the very end that "remain" was leading the way.

An important distinction must be made between social data used to predict the future or to track the present. The latter, called "nowcasting", focuses on using social data to track real-world phenomena in real-time [Cristianini 12] and makes sense when applied to events like influenza outbreaks, voting intentions, people's happiness, etc. On the contrary, "nowcasting" financial markets is pointless: real-time data in this field are readily available, therefore the focus is on social media information that can lead financial markets in the future [Zheludev 15].

2.2 Analyzing Social Media Data to Track Financial Markets

As [Zheludev 15] points out there are two schools of thought about the best methods for tracking financial markets using information extracted from social media. The first one focuses on the volume of data, analyzing information from social media messages [Mao 12], search engine queries [Preis 13] [Challet 13] [Bordino 12], Wikipedia articles' views [Moat 13] and financial news [Alanyali 13], in order to predict changes in the returns of stocks and indexes. These methods ignore the evaluation of the content of data, considering just their volume. The second methodology focuses on the quantitative evaluation of the content of social media messages; it is possible to find examples of this line of research in [Pepe 11], [Zhang 11], [Bollen 11] and others. To apply these kinds of techniques the researchers must perform a quantitative analysis of the meaning of social texts from a big number of people, since as showed by [Saavedra 11] an estimation applied to a large group offers stronger results than the viewpoint of just a few individuals. This approach is referred to as Sentiment Analysis: the use of natural language processing and text analysis to identify and extract subjective information from unstructured texts (*see section 1.8*). It allows a deeper analysis of social media power to lead financial markets, "over and above what is possible with solely message-volume based

analysis" [Zheludev 15]. Nevertheless, the value of these techniques in predicting financial markets has not been fully uncovered and that is why this research will focus more on this second approach.

For a theoretical discussion and analysis about how investor sentiment affects stock market and how to quantify this effect, refer to [Baker 07].

[Thelwall 10] examines the application of sentiment analysis to unstructured and informal text, with a focus on internet-sourced data. This research proves that the majority of the current sentiment analysis tools are not suitable, because they are not specifically developed to analyze colloquial and unstructured texts like Tweets and internet-sourced messages. To overcome this issue [Thelwall 10] presents an innovative system called SentiStrength[5], created specifically to rank colloquial texts from the internet. It works on the principle of dictionary matching and it is based on the work of [Pennebaker 07] that developed LIWC[6] – Linguistic Inquiry and Word Count, a multi-mood dictionary-term matching software. There are other approaches available, like the "part of speech tagging" method [Mitkov 05], but usually they can't ensure high-quality results because they are not designed to analyze informal, short and noisy texts, which often ignore standard spelling and grammar rules [Brill 92].

The implementation of tools for Sentiment Analysis has demonstrated that the Internet, and more in particular the analysis of discussions on Twitter [Bollen 11] and the observation of web-search data [Choi 12], are a valuable source of information, useful to track the mood of users about social, political and economic events. As underpinned by [Nofsinger 05], people's mood is strongly linked to the performance of stock indexes and vice-versa. Proofs of this can be found in [Zhang 11], that using a sample of 1% of all global Tweets to measure collective hope and fear, showed that the emotional tweet percentage significantly negatively correlated with Dow Jones, NASDAQ and S&P 500, but significantly positively correlated with the VIX (Chicago Board Options Exchange Volatility Index). As one of the first

[5] http://sentistrength.wlv.ac.uk/
[6] http://www.liwc.net/

research in this field [Zhang 11] used a “part of speech” sentiment classification, that, as mentioned before, it is not specifically developed for analyzing unstructured messages. A more appropriate methodology has been used by [Bollen11], that investigated the same field by employing a multi-mood approach. With the help of a tool named GPOMS – Google-Profile of Mood States, this work categorizes text in six dimensions of emotion (calm, alert, sure, vital, kind, happy) and shows that a sample of global Tweets can predict the direction of the Dow Jones Industrial Average index with a level of accuracy of 86.7%. This study has inspired many other researches. For example, it has been emulated by [Mittal 12], that confirmed the results by [Bollen 11] proposing also a portfolio management strategy based on the predicted values.

The researches mentioned above focused on the issue of predicting the return of stock indices, while [Ruiz 12] considered instead the problem of predicting specific stocks. This research underlines a significant correlation between Twitter messages on companies and market trading volumes for some publicly traded stocks. Others achieved similar outcomes, like [Mao 12] that, using a linear regression model, showed that the daily number of Tweets mentioning S&P 500 stocks significantly correlate with S&P 500 daily closing prices, daily price changes and absolute daily price changes. Similar results have also been found making analogous analysis using Google Trends [Preis 13] [Challet 13], Yahoo Search Engine [Bordino 12] and Wikipedia [Moat 13]. An interesting result is displayed by [Ranco 15], that analyzing 30 stock companies from the DJIA index, found out a significant correlation between Twitter sentiment and abnormal returns in particular when tweet volume reaches its peaks. Other attempts to use sentiment analysis to track the prices of specific stocks, instead of indices, can be found in [Oliveira 13], which provides an analysis of the content of Twitter data for determining future performances of specific stocks. The analysis shows weak evidence of return predictability using sentiment indicators, but meaningful correlation between Twitter posting volume and trading volume as well as volatility. The work applies a simple regression analysis to nine stocks, considering a range of time of 32 days, leaving room for further in-depth analysis.

Twitter is not the only source of social media data that could be used to track financial markets. Queries from internet search engines can have predictive power as well, as shown by [Bordino 12] that found correlation between daily trading volumes of stocks traded in NASDAQ100 and daily volumes of queries related to the same stocks. A similar result is presented by [Preis 10] that found clear evidence of a correlation between weekly transaction volumes of S&P 500 companies and weekly search volume of corresponding company names. On the base of these evidences, [Preis 13] even developed a profit-making trading strategy based on the analysis of economic terms using Google Trends. However, according to [Zheludev 15]: "These strategies were structured based on an ex-post facto identification of the search terms which would result in the highest profits retroactively". For this reason, this cannot be fully considered a demonstration that data from social media can predict financial markets, in fact a study by [Challet 13] demonstrates that the volume of internet searches relating to non-finance words contains the same predict power as finance-related searches, if considered with the same ex-post facto methodology.

Furthermore, there are researches that show how search volume itself can be a mood indicator for financial markets. The results presented by [Da 15] show that the more people search about economic negative terms, such as *"recession" "unemployment" "bankruptcy",* the more pessimistic people feel about the economic situation. Similar results can be found in the work of [Mao 11] that shows how the search queries of 19 negative economic words present a strong correlation with unemployment rate, consumer confidence, and investor sentiment. This research displays also evidences about how it is possible to help the prediction of the unemployment rate using certain search queries related to the matter.

3 DATA COLLECTION AND ANALYSIS

Our study aims to investigate if information extracted from social media represent a useful source of data to understand and track financial markets. In this work, following [Counts 11], two different analysis will be carried out. The first one takes into consideration Twitter and wants to investigate whether the mood that emerges from this social network about Apple Inc. is somehow correlated with its stock indicators. The second one focuses on search volume of financial terms and how this information can be useful to track the performances of financial indexes such as S&P 500 and FTSE 100.

This chapter will first outline the methods behind the collection of data from Twitter, Google Trend, and the stock market. Afterward, it will present the methodologies used to evaluate the extent to which the collected data can be useful to track financial markets.

3.1 Twitter Data Collection

The reasons why we decide to select Twitter as a source of data are mainly two. First, this platform represents one of the most popular microblogging services, second, its API – Application Programming Interface, allows to collect a large number of messages. In fact, it is not possible to find aggregated datasets of historic tweets, but since the License Agreement between Twitter and its users makes most tweets public, we had the chance to collect and store them locally.

Before discussing the methodology used to extract and collect tweets it is useful to briefly assess the demographics of users of this social media, in order to understand if these users can represent an accurate picture of society. Twitter has more than 300 million monthly active users, who send more than 500 million tweets every day [Newberry 16]. 79% of these accounts are located outside the United States, with the top countries been Brazil, Japan, Mexico, and India. Considering the U.S.A., roughly one-quarter of online adults (24%) uses Twitter with almost no bias regard to gender. *Table 2* summarizes the results of a Pew Research study by

[Greenwood 16] and shows how it is more likely to find younger Americans on Twitter, with a percentage of 36% of online adults aged 18-29. Then the usage starts to slide down a little bit as the age grows, and once users hit age 65, they start leaving the social network with only a 10% of senior using this service. *Table 2* shows also how Twitter is more popular among the highly educated: 29% of online adults with college degrees use Twitter, 20% have high school diplomas or less. Talking about income demographics, Twitter users are quite evenly split across the income brackets. As underlined by [Zheludev 15], the racial demographics of Twitter users are often representative of society, while differences exist depending on geographical locations within the USA. For example, Hispanic users are underrepresented in the South-West of the United States; African-American users are underrepresented in the South and Caucasian users are overrepresented in major cities.

24% of online adults (21% of all Americans) use Twitter

% of online adults who use Twitter

All online adults	24%
Men	24
Women	25
18-29	36
30-49	23
50-64	21
65+	10
High school degree or less	20
Some college	25
College+	29
Less than $30K/year	23
$30K-$49,999	18
$50K-$74,999	28
$75,000+	30

Note: Race/ethnicity breaks not shown due to sample size.
Source: Survey conducted March 7-April 4, 2016.
"Social Media Update 2016"

PEW RESEARCH CENTER

Table 2: *Twitter demographics [Greenwood 16].*

In order to carry out our Twitter-based analysis, we decided to focus the attention on Apple Inc.'s stock because it is one of the most tweeted companies [Hentschel 14]. In Twitter community, users usually refer to a company's stock using the stock symbol preceded by a dollar sign (*cashtag*), for example $AAPL for Apple Inc. We used this method to collect all the tweets related to Apple's stock because the word "apple" is a common word and using it as keyword would result in a large amount of misleading tweets. In order to collect tweets, we used *"Twitter Archiver*[7]*"* an add-on for Google Sheets developed by Amit Agarwal. Twitter Archiver allows saving tweets for any search keyword or hashtag, creating a Google Spreadsheet. After entering a search query or a hashtag, all the matching tweets are automatically saved in the Google Sheet every hour. It is possible to write simple queries, use Boolean search and even include advanced Twitter search operators to create more complex queries.

With this web app, we built a database of 46,444 tweets about Apple's stock, collecting tweets containing the cashtag $AAPL from December 12th 2016 to January 12th 2017. The analysis of this data will be presented in section 3.4.

3.2 Search Volume Data Collection

In order to extract search queries volume about some financial terms, we decided to use *"Google Trends*[8]*"*, a Google service that provides search volume data for particular search terms from January 2004 to the present. This tool can show how often a particular word is entered relative to the total search volume, giving also the chance to restrain the analysis to specific regions and languages. Google Trends results are displayed in a main graph, which shows the popularity of words over time, and in a second one that breaks down results by countries, regions, and cities. Data in these graphs can be exported into a .csv file and opened in Excel and other spreadsheets applications.

[7] https://chrome.google.com/webstore/detail/twitterarchiver/pkanpfekacaojdncfgbjadedbggbbphi
[8] https://www.google.com/trends/

For our analysis, following the work by [Counts 11], we started downloading the weekly search volume data for a set of financial keywords in English, including "*stock market", "FTSE 100", "S&P 500", "financial market", "bullish", "bearish"*, without any geographical restraints. Then we expanded the search to those terms that are suggested as top relevant from Google Trends. We end up with a list of 20 financial search terms, as shown in *Table 3*, for which we downloaded search frequency, resulting in a time series from January 1st 2013 to January 1st 2016.

Bearish, Bullish, Finance News, Finance, Financial Market, Financial News, FTSE 100, Market Value, S&P 500, Stock Decline, Stock Exchange, Stock Fall, Stock Market Crash, Stock Market News, Stock Market Today, Stock Market, Stock Price, Stock to Buy, Stock Value, Stock.

Table 3: *List of search terms.*

3.3 Stock Market Data Collection

For our Twitter-based analysis, we collected financial data from Bloomberg. We focused on Apple Inc.'s stock, gathering daily values of opening and closing price, high and low, and trading volume from December 12th 2016 to January 12th 2017. With information about the closing price, we computed the stock return for the same period of time.

The opening price represents the price at which a security first trades during the opening of an exchange, while the closing price is the price of the last transaction of a security during a day's trading session. High and low are respectively the highest and lowest price at which a stock trades during a trading day. Trading volume represents the quantity of securities traded in each day during a trading session. Return is the gain or loss of a security in a determined period. It is the percentage change in the investment value. We computed it with the adjusted

close price, which is the official closing price adjusted for capital actions and dividends. Return on a given day t (R_t) is computed as:

$$R_t = Ln\left(\frac{P_t}{P_{t-1}}\right)$$

Where P_t is the adjusted close price at day t and P_{t-1} is the adjusted close price at the previous day.

For the search volume-based analysis, we focused on the performances of S&P 500 and FTSE 100, gathering information about closing price, trading volume, and volatility. We collected weekly[9] data from Bloomberg starting from January 1st 2013 until January 1st 2017. With information about the closing price, we computed the return for the same period of time.

Volatility is a measure of risk associated with a particular security or investment. It can be defined using several different approaches.

- Historical volatility. It represents daily changes in stock prices and it is computed as the standard deviation of the change in price of a stock relative to its historic price, over a period of time.
- Implied volatility. It represents the estimated volatility of a security's price and it is computed using stock options' prices.
- Intraday volatility. It shows the price movement of a stock during a trading day.

We decided to use implied volatility, in line with [Ederington 02] that consider it an appropriate estimator of volatility. Thus, to assess S&P 500's volatility we considered VIX – CBOE[10] Volatility Index, while VFTSE – FTSE 100 Volatility Index, reflects the expected volatility of FTSE 100. These indices measure the market's expectation of volatility contained in option prices.

[9] We used last day values.
[10] Chicago Board Options Exchange.

As we pointed out, our two analysis consider different time frequency. The information collected for the Twitter-based analysis are on daily base. This allows us to have a better representation of the immediacy that characterizes this social platform. The Search volume-based analysis utilizes weekly data, because Google Trend provides information with this time frequency and because this analysis takes into consideration a longer time frame.

3.4 Analyzing the Sample: Twitter-based Analysis

The following section will present our Twitter-based analysis. First we will discuss about tools and method needed to conduct a sentiment analysis on the sample of tweets we collected, underlying features and limits of this activity. Then we will examine the results of a correlation analysis between tweets' sentiment, volume and stock indicators about Apple Inc.

Figure 3 displays the time series of Apple Inc.'s closing price during the period considered.

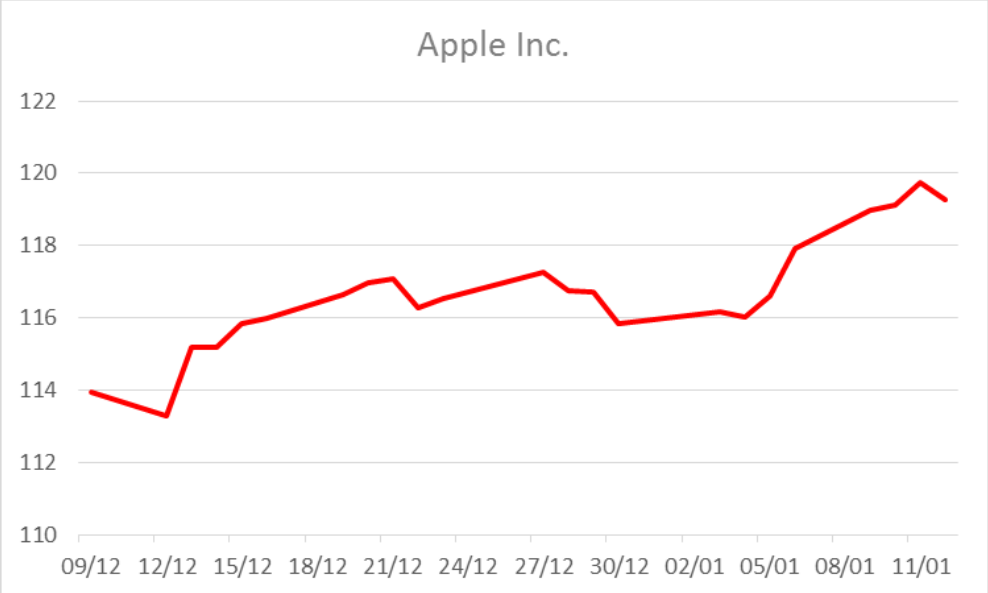


Figure 3: *Time series of Apple Inc. closing price [Personal elaboration].*

3.4.1 Choosing a Sentiment Classifier

As presented in the literature review (*see section 2.2*), previous works on the analysis of the correlation between tweets and financial markets have shown a significant relationship that links the two data sets together. In order to investigate the mood that emerges from tweets' contents we needed to perform a sentiment analysis, thus we started looking for a classifier able to fulfill the following three criteria, as pointed out by [Zheludev 15]:

- Accuracy. The tool must be designed to function properly in the context it is used, in this case social media. This means it has to consider the nature of the messages, for example incorrect spelling and formatting.
- Convenience of use in dealing with big amounts of data. This is necessary since the amount of tweets considered in the analysis is quite large and practical applicability is required.
- Transparency of internal operations. The sentiment classifier's rules of operation must be clear and accessible.

Based on the work of [Zheludev 15], we took into consideration four different tools. Now we will briefly discuss their features with reference to the aforementioned criteria.

AlchemyAPI[11]*.* It is a commercial language analysis system that offers access to a series of tools for natural language processing, including a machine learning sentiment classifier. This tool, as shown by [Batool 2013], is not suitable for the analysis of unstructured texts. In fact, it has been shown to ignore misspelling of common words instead of considering them. AlchemyAPI's sentiment classifier ensure practical applicability, however it is a commercial system (part of IBM's Watson Developer Cloud) so it is not free to use. Furthermore, the methodology used to determine the sentiment is not disclosed by the company and this undermines transparency.

[11] http://www.alchemyapi.com/

LIWC[12] *(Linguistic Inquiry and Word Count).* It is a dictionary-term matching sentiment classifier based on a corpus of 9,906 words. This software's strength is its accuracy in ranking formal English texts, its corpus and all attributions of words to various emotions have been checked by groups of human judges [Pennebaker 07]. However, this system is not able to rank informal texts, for example it does not recognize common misspelling of common English words and it does not take into consideration the negative effects of negations (e.g. *"not good"*). Practical applicability presents limits, due to the method chosen to feed in data, which requires users to insert text manually into a graphical user interface. Classification methodology is fully transparent and available to users.

Sentiment140[13]*.* This is a machine learning-based sentiment classifier that allows ranking tweets based on their polarity (positive, negative, neutral). It was made specifically to rank tweets so it is suitable to understand and classify unstructured texts. Practical Applicability is not assured since its free version requires users to manually insert data. Technical reports about approaches and methodologies of the classification are available to users on the classifier's website.

SentiStrength[14]. This lexicon based sentiment classifier is specially designed for ranking short informal texts in English. For this reason, it works properly in the context of social media, ensuring accuracy in dealing with the colloquial and often grammatically incorrect nature of the messages. However, since it is heavily based on LIWC – Linguistic Inquiry and Word Count, it is able to rank also formal English texts. SentiStrength can deal with a large amount of data (it can process about 16,000 tweets per second[15]) and its interface ensure practical applicability. Furthermore, it is free for academic research. Its classification methodologies are completely transparent and can be accessed by the user on the webpage of the software.

[12] http://liwc.wpengine.com/
[13] http://www.sentiment140.com/
[14] http://sentistrength.wlv.ac.uk/
[15] As stated on the webpage of the project.

Since SentiStrength is the only classifier that meets all three criteria, we chose it as sentiment analysis tool for this research. In the following section, we will examine more in depth the approach that stands at the base of its classification process.

3.4.2 SentiStrength

The heart of SentiStrength is a lexicon of 2,310 sentiment words from LIWC (Linguistic Inquiry and Word Count) program and the General Inquirer list of sentiment terms, plus some additions made during testing [Thelwall 16]. SentiStrength considers a scale from +1 (not positive) to +5 (extremely positive) to rank positive sentiments, while negative ones are ranked on a scale from -1 (not negative) to -5 (extremely negative). Each word in the lexicon has a positive or a negative score within one of these two scales. These scores were human assigned based on the analysis of 2,600 comments from the social network MySpace, then updated through further testing.

[Thelwall 16] offers a clear explanation of how SentiStrength operates. Every time it reads a text, it splits it into words, separating emoticons and punctuation. Each word is checked with the lexicon for matching any of the sentiment terms. When it found a match, the associated sentiment score is memorized. The overall score for a sentence is the highest positive and negative score of its composing words. If there are multiple sentences, the scores of the whole text correspond to the maximum positive and negative scores of the single sentences. Every analyzed phrase will display two scores, a positive and a negative one. The reason why SentiStrength uses this dual scale is that psychological research reports that humans can feel positive and negative emotions simultaneously [Norman 11]. For example, the sentence: *"Susan is pretty but her sister is boring. I hate her."* would be classified as: "*Susan is pretty [3] but her sister is boring [-2]. <sentence: 3,-2> I hate [-4] her. <sentence: 1,-4>".* The number in square brackets indicates sentiment strength of the preceding words, while angle brackets contain sentence scores. The overall classification for

the text is given by the highest positive score and the lowest negative score, which are (3,-4).

SentiStrength can also report binary (positive/negative), trinary (positive/negative/neutral) and single scale (-4 to +4), however the free version available for academic research, only allows to rank tweets using the double score approach.

As I mentioned before SentiStrength's dual scale can go from 1 to 5 for positivity and from -1 to -5 for negativity. An important aspect to notice is that a score of 1 does not indicate positivity, like a score of -1 does not imply negativity. For this reason the sentence: *"I bought a new laptop and a pair of shoes"* would be classified as: "*I bought a new laptop and a pair of shoes. <sentence: 1,-1>".* Since no sentiment words have been matched with the lexicon, the classifier returns a score of (1,-1), which indicates the lack of any sentiment, either positive or negative. We define cases like this as neutral. We will discuss them more in depth in section 3.4.5.

SentiStrength includes also a list of words that increase or decrease the strength of emotion of following words by 1 or 2 points, like *"very", "extremely"*, *"little"* etc. In addition, the algorithm contains words that reverse the polarity of subsequent emotion words, like negation with *"not"* - e.g. *"very happy"* scores (3,-1) while *"not very happy"* scores (1,-2). Since emoticons are of common use in social media, the classifier also embodies a list of popular emoticons that affect scores in a positive or negative sense [Gonzalez 2016].

As we mentioned, SentiStrength is a lexicon based sentiment analysis classifier. Compared to machine learning based classifiers, tools like SentiStrength usually ensure more accuracy in ranking sentiments from social web texts. This topic is discussed by [Thelwall 12] who reports an experiment that shows how SentiStrength outperforms other machine learning competitor tools in ranking sentiments on social media. First, a panel of human judges classified a set of 1,041 MySpace comments. Afterward, SentiStrength's output about these texts has been compared to the outcome of other machine learning classification algorithms. The results, based on the ranking produced by the human judges, showed that

SentiStrength's ability to determine sentiments was significantly above the best result from those machine learning tools. This could be explained by the fact that lexical sentiment analysis *"is less likely to pick up indirect indicators of sentiment that will generate spurious sentiment patterns"* [Thelwall 16]. For example, a machine learning classifier might consider names of unpopular politicians as negative sentiment words because they tend to occur in negative texts. This can potentially jeopardize the accuracy of any derived sentiment analysis about that topic. [Thelwall 16]. The exceptions where lexicon based classifiers perform less well than machine learning competitors are texts with sarcasm or irony.

For our research, we submitted the database of tweets about Apple Inc.'s stock to SentiStrength software. The output of the evaluation is a ranked database, in which every tweet has a positive sentiment score (how positive that message is) and a negative sentiment score (how negative that message is). *Figure 4* presents an extract from the tweet database ranked by SentiStrength.

Date	Tweet	Positive	Negative	EmotionRationale
12/12/2016 06:59	@Sassy_SPY Hmmm AAPL has always be generous to me with free stuff	2	-1	@Sassy_SPY[0] Hmmm/Hm[0][+0.6 MultipleLetters] AAPL/APL[0] has[0] always[0] be[0] generous[1] to[0] me[0] with[0] free[0] stuff[0] [[Sentence=-1,2=word max, 1-5]][[[2,-1 max of sentences]]]
12/12/2016 06:59	AAPL call strike 115 last 0.33 this week (4 days left).	1	-2	AAPL/APL[0] call[0] strike[-1] 115[0] last[0] 0[0] .33[0] this[0] week[0] (4[0] days[0] left[0] [[Sentence=-2,1=word max, 1-5]][[[1,-2 max of sentences]]]
12/12/2016 07:00	FIVE MINUTE TRADES ARE WORKING - https://t.co/moCrpqc4V7 SPY IWM QQQ UVXY VXX FB NFLX PCLN BBRY USO GDX NUGT JNUG AAPL	2	-1	FIVE[0] MINUTE[0] TRADES[0] ARE[0] WORKING[0] https[0] ://t[0] [[Sentence=-1,1=word max, 1-5]] co/moCrpqc4V7[0] SPY[0] IWM[0] QQQ/Q[0][+0.6 MultipleLetters] UVXY[0] VXX/VX[0] FB[0] NFLX[0] PCLN[0] BBRY[0] USO[0] GDX[0] NUGT[0] JNUG[0] AAPL/APL[0] [[Sentence=-1,2=word max, 1-5]][[[2,-1 max of sentences]]]
12/12/2016 07:00	Apple Stock Price: 113.00 #apple AAPL	1	-1	Apple[0] Stock[0] Price[0] 113[0] .00[0] #apple[0] AAPL/APL[0] [[Sentence=-1,1=word max, 1-5]][[[1,-1 max of sentences]]]
12/12/2016 07:00	My column on how to pick good momentum, growth, small-cap and value #stocks https://t.co/ImwBmvZBVU AMZN COST https://t.co/zIIz3biEAt	2	-1	My[0] column[0] on[0] how[0] to[0] pick[0] good[1] momentum[0] growth[0] small[0] cap[0] and[0] value[1] #stocks[0] https[0] ://t[0] [[Sentence=-1,2=word max, 1-5]] co/ImwBmvZBVU[0] AMZN[0] COST[0] https[0] ://t[0] [[Sentence=-1,1=word max, 1-5]] co/zIIz3biEAt[0] [[Sentence=-1,1=word max, 1-5]][[[2,-1 max of sentences]]]
12/12/2016 07:01	FB Support @ weekly 10MA. Must retake 119.16 (21MA on 60 min) by end of 11:30EST. Still in uptrend move. No panic. AAPL NFLX GOOGL AMD	2	-3	FB[0] Support[1] @[0] weekly[0] 10MA[0] [[Sentence=-1,2=word max, 1-5]] Must[0] retake[0] 119[0] .16[0] (21MA[0] on[0] 60[0] min[0] by[0] end[0] of[0] 11[0] :30EST[0] [[Sentence=-1,1=word max, 1-5]] Still[0] in[0] uptrend[0] move[0] [[Sentence=-1,1=word max, 1-5]] No[0] panic[-2] [[Sentence=-3,1=word max, 1-5]] AAPL/APL[0] NFLX[0] GOOGL[0] AMD[0] [[Sentence=-1,1=word max, 1-5]][[[2,-3 max of sentences]]]
12/12/2016 07:01	RT @OptionTrader101: AAPL call strike 115 last 0.33 this week (4 days left).	1	-2	RT[0] @OptionTrader101[0] AAPL/APL[0] call[0] strike[-1] 115[0] last[0] 0[0] .33[0] this[0] week[0] (4[0] days[0] left[0] [[Sentence=-2,1=word max, 1-5]][[[1,-2 max of sentences]]]
12/12/2016 07:01	GOOG AAPL validj: Nice chart set-ups SFUN CNIT NQ RENN https://t.co/8U3njQRF3u	2	-1	GOOG[0] AAPL/APL[0] validj[0] Nice[1] chart[0] set[0] ups[0] SFUN[0] CNIT[0] NQ[0] RENN/REN[0] https[0] ://t[0] [[Sentence=-1,2=word max, 1-5]] co/8U3njQRF3u[0] [[Sentence=-1,1=word max, 1-5]][[[2,-1 max of sentences]]]
12/12/2016 07:01	PortfolioBuzz: Track trending assets in 1 watchlist AAPL GBPUSD LMT USDCAD USDYEN https://t.co/bwHofTxdC8 https://t.co/kNoWRS5uz4	1	-1	PortfolioBuzz/PortfolioBuz[0] Track[0] trending[0] assets[0] in[0] 1[0] watchlist[0] AAPL/APL[0] GBPUSD[0] LMT[0] USDCAD[0] USDYEN[0] https[0] ://t[0] [[Sentence=-1,1=word max, 1-5]] co/bwHofTxdC8[0] https[0] ://t[0] [[Sentence=-1,1=word max, 1-5]] co/kNoWRS5uz4[0] [[Sentence=-1,1=word max, 1-5]][[[1,-1 max of sentences]]]

Figure 4: Extract from the tweet database analyzed by SentiStrength. *[Personal elaboration].*

3.4.3 Issues with Sentiment Classifiers

SentiStrenght meets the requirements we set to find a valid and reliable sentiment classifier, however there are some factors, typical of the majority of sentiment analysis tools, that currently do not allow us to rely blindly on them, when it comes to interpreting human language. These factors are described by [Donkor 13].

- Semantic interpretation. Semantics, as the study of words meanings and relations between them, does not allow us to assume that if a sentence contains a positive sentiment word, that sentence is also positive. In fact, other words can alter the sentiment of a word, like negations (e.g. "I think this train is *not* reliable*"),* connectives (e.g. "This train is everything *but* reliable"), modals (e.g. "In theory, this train *should* be reliable"). SentiStrength's algorithm takes into account some of these words but it cannot consider every case.
- Context. Depending on the context, the same word can communicate opposite sentiments (e.g. "My bank does a *great* job when it comes to stealing money from me").
- Sentiment ambiguity. Phrases with sentiment words do not necessarily express any feeling (e.g. "What is the most *reliable* train company?"). Sentences without any sentiment words can express sentiments (e.g. "I'm over the moon!").
- Sarcasm. Sarcasm can invert the usual sentiment expressed by a word (e.g. "Of course I'm *happy* to go to work two hours early").
- Language. A word may change its meaning and sentiment based on the language used. It is common for dialect and slang (e.g. "*sick*" in urban slang is used as a synonym of awesome).

These factors represent restrictions and challenges of current sentiment analysis tools, much can still be done to develop and sharpen them, however there is likely an intrinsic limit, in fact as [Donkor 13] affirms:

> *"One thing I've learn studying Linguistics at university is that language is complex. In fact, it would be too naïve to oversimplify*

language thinking that its underlying sentiment can always be accurately examined by a machine or an algorithm."

3.4.4 Building an Indicator

Now every tweet in our database has two scores, one for positivity and one for negativity. We need to define a rule to determine if a tweet's general sentiment is positive negative or neutral. In order to do this, we computed what we called "Compensated Strength" *(CS)* which is defined as:

$$CS = Vp + Vn$$

Where Vp is the score for positivity of a given tweet and Vn is the score for negativity. With this value, we can determine the general sentiment of every tweet. In particular:

$$CS > 0 \Rightarrow POSITIVE$$

$$CS < 0 \Rightarrow NEGATIVE$$

$$CS = 0 \Rightarrow NEUTRAL$$

After this classification, our database presents 8,588 positive tweets, 8,855 negative tweets and 29,001 neutral tweets. With this information, we built a Twitter –based financial mood indicator about Apple Inc.'s stock called Twitter Mood Indicator *(TMI)*. Based on the number of positive (N_{pos}) and negative (N_{neg}) tweets on a given day, we define the Twitter Mood Indicator for a day t (TMI_t) as:

$$TMI_t = \frac{N_{pos}}{N_{pos} + N_{neg}}$$

A similar indicator, which inspired ours, can be found in [Counts 11]. It is called TIS – Twitter Investor Sentiment, and measures the investor sentiment score based on the number of tweets containing the words *"bullish"* and *"bearish"*.

We want to present an alternative version of the Twitter Mood Indicator, defined TMI*. It occurs in [Zhang 10] as polarity indicator and can provide additional information compared to its basic version. While *TMI* is always positive, if *TMI** goes in negative direction, it means that we have a higher number of negative tweets compared to the positive ones and vice versa. *TMI** is computed as follows.

$$TMI_t{}^* = \frac{(N_{pos} - N_{neg})}{(N_{pos} + N_{neg})}$$

As shown by the formulas above, the number of neutral tweets is not involved in building the *TMI* or *TMI** and considering that neutrality represents the majority of the sentiment of our data set, it can be seen as a limit of this approach. We will discuss this issue in section 3.4.5.

A second limit related to the methodology that we decided to use to build our indicator, is omitting the strength about positivity and negativity of tweets. In *Table 4* we present an example to clarify the problem.

	Tweet Text	Score from SentiStrength	Compensated Strength (*CS*)	Polarity
a	*I really love you but I dislike your cold sister.*	(4,-3)	1	Positive
b	*I love dogs so much. I want to have one.*	(3,-1)	2	Positive
c	*I left early because the film was boring. I am not a fan of musicals.*	(1, -2)	-1	Negative
d	*I am terrified of my neighbor. I think he may kill me.*	(1, -4)	-3	Negative

Table 4: *Examples of texts ranked by SentiStrength as positive and negative [Personal elaboration].*

Tweet *a* and *b* are both positive, however their compensated strength is different. Tweet *b* has a bigger value of *CS* (i.e. 2, compared to 1 of tweet *a*); this reflects a "more positive" sentiment than tweet *a*. Same for tweet *c* and *d*. In this case tweet *d*, shows a compensated strength equal to -3, compared to -1 of tweet *c*, for this reason, tweet *d*'s sentiment is "more negative".

To carry out the analysis we decided to consider two different indicators. We saw that TMI does not contain any information about strength and takes into consideration only polarity. In order to use both kinds of information, we decided to develop another indicator, following a different approach: including in the calculations the actual value of compensated strength for every tweet. The new indicator is called *CSI* (Compensated Strength Indicator), and it considers tweets' polarity and strength. Based on the value of positive tweets' compensated strength (CS^+) and negative tweets' compensated strength (CS^-) on a given day, we define the Compensated Strength Indicator for a day *t* (CSI_t) as:

$$CSI_t = \frac{{CS^+}_t}{{CS^+}_t + |{CS^-}_t|}$$

Where CS^+ is the value of compensated strength if CS > 0, and CS^- is the value of compensated strength if CS < 0.

As for the *TMI*, also the Compensated Strength Indicator has an alternative version that provides additional information (*CSI**). When *CSI** is positive the value of positive tweets' compensated strength is greater than the value of negative tweets' compensated strength and vice versa. *CSI** is defined as:

$$CSI_t{}^* = \frac{CS^+{}_t - |CS^-{}_t|}{CS^+{}_t + |CS^-{}_t|}$$

These new indicators (*CSI* and *CSI**) take into account the information about tweets' strength, contained in the values of compensated strength. The example in *Table 5* compares *TMI* and *CSI* and shows how the Compensated Strength Indicator reflects this information.

	Vp	***Vn***	***CS***	**Polarity**
e	3	-1	2	Positive
f	2	-3	-1	Negative

Table 5: *Examples of tweets ranked by SentiStrength on a given day [Personal elaboration].*

Consider tweet *e* and *f* as all the tweets extracted on a given day *t*. The indicators are equal to:

$$TMI_t = \frac{1}{1+1} = \frac{1}{2} \qquad TMI_t^* = \frac{1-1}{1+1} = 0$$ [16]

$$CSI_t = \frac{2}{2+|-1|} = \frac{2}{3} \qquad CSI_t^* = \frac{2-|-1|}{2+|-1|} = \frac{1}{3}$$

As we can see from the example shown in *Table 5,* on day *t* we extracted and analyzed two tweets: *e* is ranked positive and *f* is ranked negative. TMI and TMI* show that for day *t* we have an equal number of positive and negative tweets. However, in absolute terms, the positivity expressed by tweet *e* (*CS* = 2) is stronger than the negativity expressed by tweet *f* (CS = -1) and this information does not emerge from TMI or TMI*. Instead, looking at CSI and CSI*, it is possible to see that they reflect the greater strength of positivity compared to negativity, in fact, $CSI_t > TMI_t$ and $CSI_t^* > TMI_t^*$.

3.4.5 How to Treat Neutral Polarity

As pointed out by [Koppel 06], a common aspect of almost all the works that deals with models for sentiment analysis is the tendency to define the task as a two-category problem; positive versus negative. "In almost all actual polarity problems, including sentiment analysis, there are, however, at least three categories that must be distinguished: positive, negative, neutral" [Koppel 06]. It is true in fact that not every comment express purely a positive or negative sentiment. Some comments may report objective fact omitting any mood, while others may reflect mixed or opposing sentiments. Even though researchers obviously are aware of the existence of neutrality, the reason why they keep ignoring it relies on two assumptions discussed by [Koppel 06]:

[16] A value of *zero* for *TMI** means that the number of positive tweets is equal to the number of negative tweets. $-1 \leq TMI^* \leq 1$

- Neutral documents simply lie near the boundary of the binary model positive/negative.
- There is less to learn from neutral documents than from ones with defined sentiment.

In his paper [Koppel 06] wants to overcome these assumptions and shows their lack of foundation, proving that taking into consideration also neutral document can potentially improve the overall classification accuracy of sentiment analysis.

As we already mentioned, we decided not to take into consideration for our analysis the 29,001 observations ranked from SentiStrength as neutral. However since they represent the great part of the tweets we collected, it is reasonable to examine them a little bit more in depth. The example reported in *Table 6* is useful to distinguish between two fundamentally different types of neutrality.

	Tweet Text	Score from SentiStrength	Compensated Strength (*CS*)	Polarity
g	*I bought a new car but it is last year's model.*	(1,-1)	0	Neutral
h	*I like your apartment but I think your cats are ruining it.*	(2,-2)	0	Neutral

Table 6: *Examples of texts ranked by SentiStrength as neutral. [Personal elaboration].*

Both tweets *g* and *h*, after being analyzed by SentiStrength, showed a value of *CS* equal to z*ero*, so we considered them neutral. Nevertheless, tweet *g*'s score is (1,-1) which means (as we know from the methodologies of SentiStrength's algorithm) that the classifier considers it neither positive or negative, i.e. absence of both sentiments. Tweet *h* instead, got a score of (2,-2) which means that this text communicates two opposite sentiments with equal strength, and for this reason its compensated strength is 0. It is undeniable that the information contained in these

two kinds of tweets is very different: tweet *g* expresses an absence of sentiment while tweet *h* expresses conflicting opinions. When a text presents no sentiment at all we can consider it neutral, when we have the same number of positive and negative opinions in a text, we have ambivalent sentiment [Donkor 13].

With respect to this, we want to specify that among the 29,001 "neutral" tweets we identified, 27,751 are neutral (i.e. they express no sentiments) and 1,250 are ambivalent (i.e. they express conflicting sentiments with equal strength). This means that our indicators only exclude the information from the ambivalent tweets, since the neutral tweets do not communicate information about any sentiment.

3.4.6 Correlation Analysis

After computing TMI and CSI values for every day from December 12th 2016 to January 12th 2017, we investigated the relationship between these indicators and the financial data we collected about Apple Inc. (opening and closing price, high and low, trading volume and return). We also investigated how tweets volume behaves compared to the mentioned financial data.

Opening, closing, high and low are values highly correlated among them, and they generally present the same trend; we decided to take into consideration all of them to investigate whether there could be any particular case. However, we found out that their relationship with the sentiment indicators is typically the same, so in this discussion we will only refer to opening price. We also noticed that TMI and CSI behave in the same exact way of their alternative versions TMI* and CSI*. For this reason, we will only refer to TMI and CSI.

The main results of our correlation analysis are shown in *Table 7.* It is possible to examine the whole correlation matrix in *Table 1* of the Appendix.

	Opening Price	Return	Trading Volume	TMI	CSI	Tweet Volume
Opening Price	1					
Return	-0.118	1				
Trading Volume	-0.272	0.358	1			
TMI	-0.403 ∗	0.181	0.18	1		
CSI	-0.43 ∗	0.189	0.243	0.981 ∗	1	
Tweet Volume	-0.46 ∗	0.305	0.466 ∗	0.42 ∗	0.432 ∗	1

(∗ indicates p-value < 0.1)

Table 7: *Main results from the correlation matrix between Twitter sentiment, volume and Apple Inc. financial data. [Personal elaboration].*

We can observe that TMI is negatively correlated with opening price, this result is consistent with [Counts 11]. Opening price is also negatively correlated with tweet volume, and this may indicate that if Apple's stock price decreases, people will tweet more about the company. We observed a quite high correlation between TMI and tweet volume ($\gamma = 0.42$), and we know from TMI's formula (*see section 3.4.4*) that this indicator includes also information about tweets' quantity, not only sentiment. Thus, the fact that TMI and tweet volume display a very similar behave in respect to opening price, suggests that the information that explains the great part of these relationships is volume and not sentiment.

We identified the same negative correlation also between CSI and opening price. More in general we noticed that CSI presents the same behaviour of TSI when compared to the considered financial data, even displaying higher correlation coefficients. This may suggest that adding information about sentiment strength contribute to develop an indicator able to better explain the relationship between Twitter sentiment and Apple's financial data.

Tweet volume displays a positive correlation with trading volume, as showed also by [Mao 12] and [Oliveira 13]. This may suggest how the more Apple stocks are traded, the more people will discuss about them.

In *Figure 5,* it is possible to visually examine the aforementioned results through trend analysis and scatterplots.

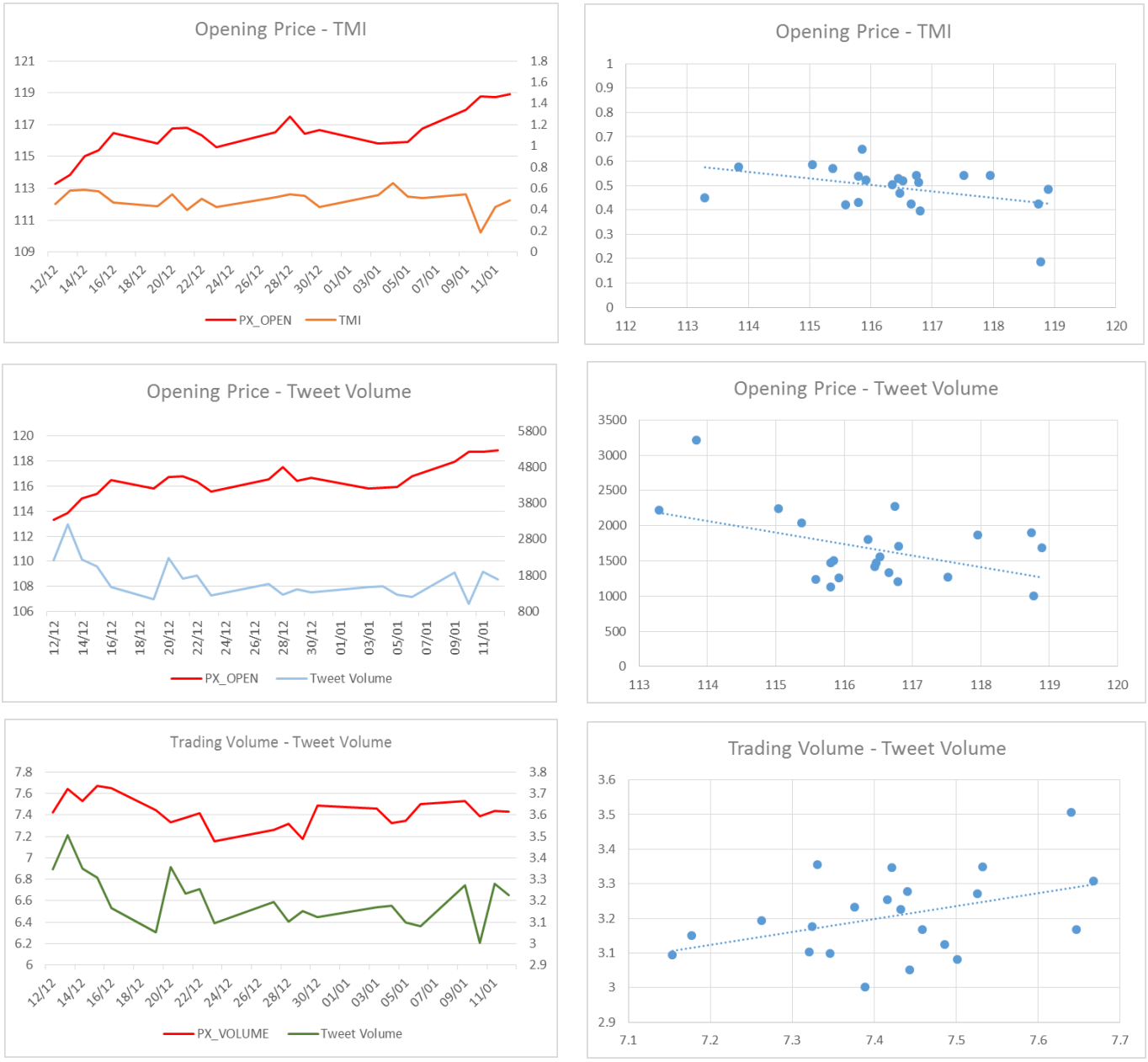


Figure 5: *Trend analysis and scatterplots of Apple Inc.'s opening price and trading volume vs. TMI and tweet volume [Personal elaboration].*

The correlation analysis we have just described considers financial data and tweets from the same day. In order to investigate whether one of this information influences the other with a time delay, we carried out another correlation analysis introducing a time lag of 1 day. First, we tested how Twitter information from today can influence tomorrow's stock data, then the impact that today's market values can have on tomorrow's tweets. The main results of our correlation analysis are shown in *Table 8.* It is possible to examine the whole correlation matrices in *Table 2* and *Table 3* of the Appendix.

Analyzing the case in which today's market values influence tomorrow's tweets, we noticed a higher correlation of TMI and tweet volume with the opening price. Therefore, this may indicate that if Apple Inc. price decreases, the next day people will tweet more. However, the correlation between tweet volume and trading volume loses its statistical significance in this case, while in the other scenario they show a high positive correlation *(γ = 0.44*). This may indicate that the right interpretation is that when people tweet more about Apple Inc., the next day the trading volume of the stock will increase.

As it is possible to notice in *Table 7 and Table 8*, the correlation analysis we conducted revealed many statistically insignificant values (*with a p-value < 0.1*), making impossible to interpret all results. For example, the correlations between our mood indicators and Apple's stock return is not statistically significant, and the same happens for the relationship between return and tweet volume. The reason why this happened is probably linked to the shortage of the sample, that only considers a 30 days time frame.

Matrix A	**Opening Price**	**Return**	**Trading Volume**	**TMI**	**CSI**	**Tweet Volume**
Opening Price	1					
Return	-0.314	1				
Trading Volume	-0.275	0.364	1			
TMI	-0.347	-0.068	0.115	1		
CSI	-0.379 ∗	-0.064	0.161	0.981 ∗	1	
Tweet Volume	-0.398 ∗	-0.072	0.436 ∗	0.42 ∗	0.433 ∗	1

(∗ indicates p-value < 0.1)

Matrix B	**Opening Price**	**Return**	**Trading Volume**	**TMI**	**CSI**	**Tweet Volume**
Opening Price	1					
Return	-0.029	1				
Trading Volume	-0.235	0.364	1			
TMI	-0.532 ∗	-0.058	-0.027	1		
CSI	-0.558 ∗	-0.066	0.018	0.981 ∗	1	
Tweet Volume	-0.522 ∗	0.009	0.095	0.46 ∗	0.464 ∗	1

(∗ indicates p-value < 0.1)

Table 8: *Correlation matrices between Twitter sentiment, volume and Apple Inc. financial data, with a 1-day time lag. [Personal elaboration].*

Matrix A considers today's tweets and tomorrow's market values.

Matrix B considers today's market values and tomorrow's tweets.

3.5 Analyzing the Sample: Search Volume-Based Analysis

The following section will present the methodologies behind the search volume-based analysis. In particular, we will discuss the result of the correlation analysis between search queries volume of financial terms and financial indicators about S&P 500 and FTSE 100.

Figure 6 displays the time series of S&P 500 closing price and volatility (VIX) during the period considered. *Figure 7* shows the time series of FTSE 100 closing price and volatility (VFTSE) during the same period.

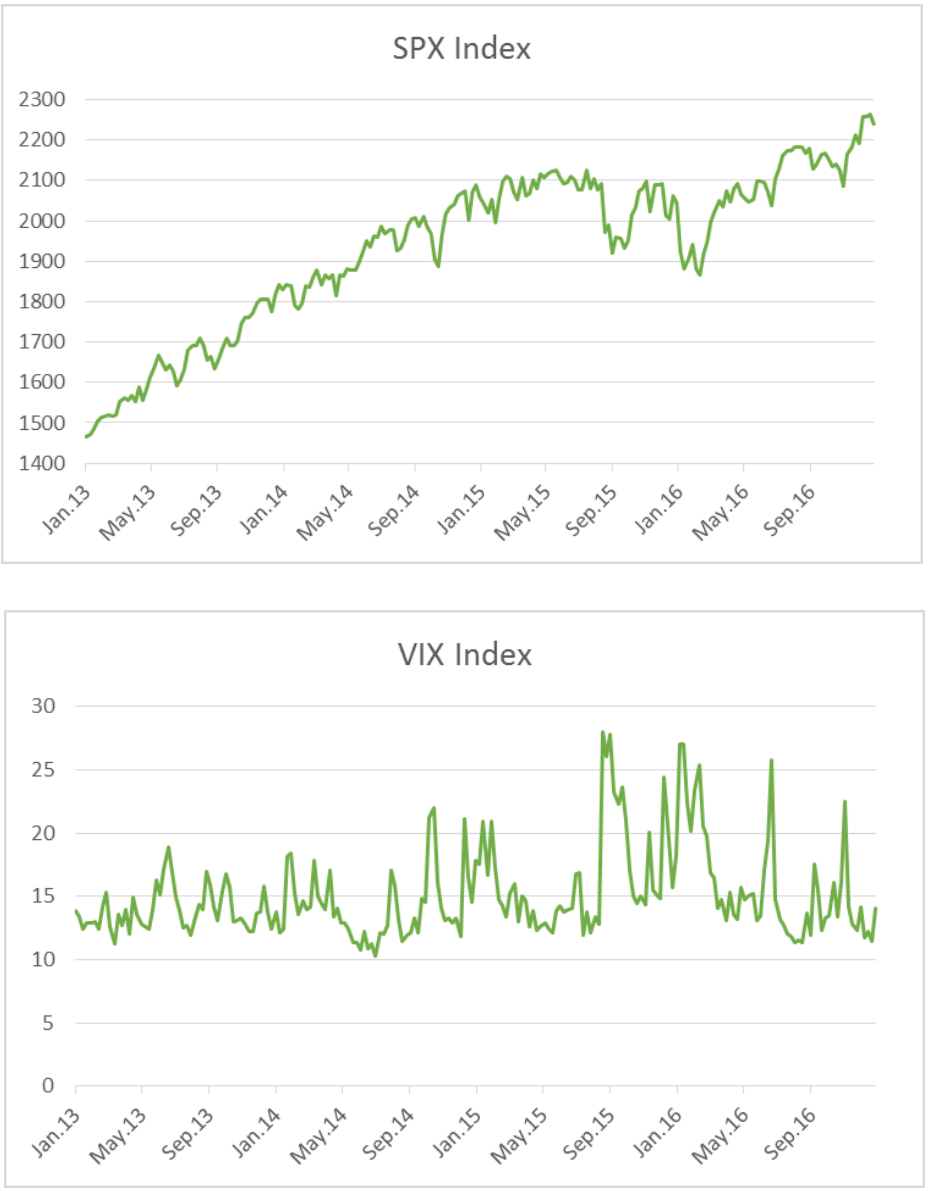
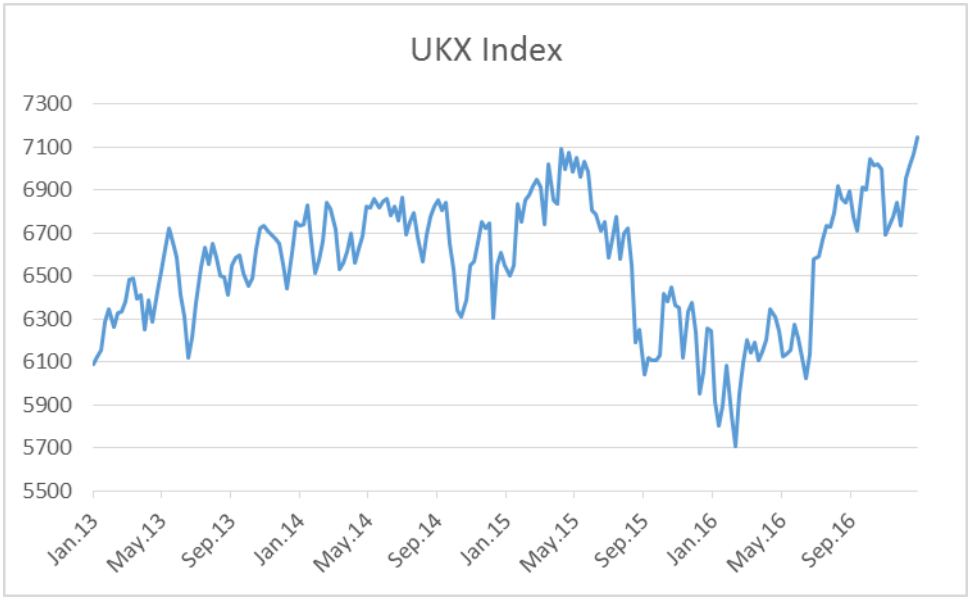


Figure 6: *Time series of S&P 500 closing price and volatility (VIX) [Personal elaboration].*



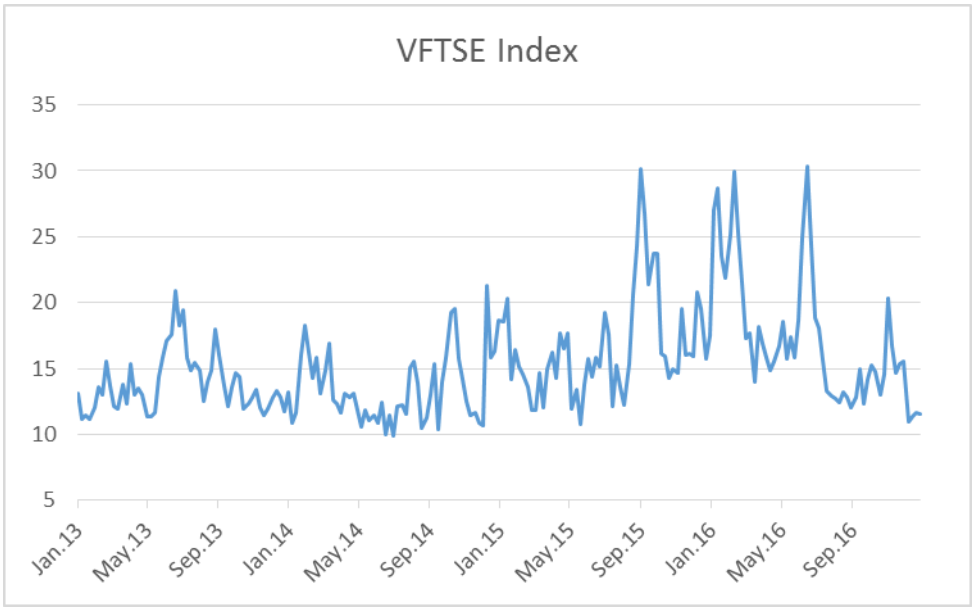


Figure 7: *Time series of FTSE 100 closing price and volatility (VFTSE) [Personal elaboration].*

3.5.1 Correlation Analysis

First, we computed a correlation analysis[17] between our 20 time series of search volume queries and the time series of the collected financial data of S&P 500 and FTSE 100. The outcome of the analysis is showed in *Table 9*.

The results showed a generally strong correlation, especially for those terms more close to the name of the index such as "*S&P 500"* and "*FTSE 100",* and those containing the word "*news"* like *"financial news", "stock market news",* etc. More in particular, the search word time series shows a positive correlation with S&P 500 trading volume and VIX. The same evidence has been found by [Counts 11], which took into consideration the DJIA – Dow Jones Industrial Average, and it may indicate that more people searching financial terms online, corresponds to an increase in market's volatility (higher values of VIX) and a raise in trading volume. A relatively strong correlation exists also between our 20 search terms and S&P 500 closing price, the majority of the words presents a positive coefficient (e.g. "*S&P 500", "stock price",* etc.) but some are negative, like "*stock exchange"* with a correlation coefficient equal to -0.32. For this reason, the interpretation of this relationship is not clear, furthermore this result is against what found by [Counts 11] that detected a negative correlation with the closing price. S&P 500 returns show a negative correlation with all the analyzed search terms.

[17] All listed correlation are statistically significant with p-value < 0.1.

	S&P 500				FTSE 100			
	Closing Price	**Return**	**Trading Volume**	**Volatility (VIX)**	**Closing Price**	**Return**	**Trading Volume**	**Volatility (VFTSE)**
Bearish	0.484 ∗	-0.165 ∗	0.259 ∗	0.283 ∗	-0.18 ∗	-0.13 ∗	0.33 ∗	0.399 ∗
Bullish	0.043	-0.071	0-093	0.079	-0.146 ∗	-0.02	0.151 ∗	0.056
Finance News	0.145 ∗	-0.309 ∗	0.267 ∗	0.525 ∗	-0.250 ∗	-0.259 ∗	0.358 ∗	0.513 ∗
Finance	-0.218 ∗	-0.251 ∗	0.015	0.304 ∗	-0.147 ∗	-0.256 ∗	0.095	0.238 ∗
Financial Market	0.012	-0.134 ∗	0.243 ∗	0.317 ∗	-0.184 ∗	-0.206 ∗	0.287 ∗	0.249 ∗
Financial News	-0.235 ∗	-0.304 ∗	0.056	0.361 ∗	-0.184 ∗	-0.268 ∗	0.15 ∗	0.302 ∗
FTSE 100	0.29 ∗	-0.129 ∗	0.305 ∗	0.347 ∗	-0.203 ∗	0.013	0.513 ∗	0.46 ∗
Market Value	0.179 ∗	-0.053	0.285 ∗	0.255 ∗	-0.164 ∗	-0.096	0.365 ∗	0.271 ∗
S&P 500	0.434 ∗	-0.31 ∗	0.393 ∗	0.649 ∗	-0.321 ∗	-0.232 ∗	0.358 ∗	0.593 ∗
Stock Decline	0.057	-0.303 ∗	0.302 ∗	0.445 ∗	-0.325 ∗	-0.328 ∗	0.193 ∗	0.329 ∗
Stock Exchange	-0.319 ∗	-0.261 ∗	-0.047	0.263 ∗	-0.097	-0.265 ∗	-0.024	0.15 ∗
Stock Fall	0.189 ∗	-0.331 ∗	0.233 ∗	0.474 ∗	-0.128 ∗	-0.292 ∗	0.208 ∗	0.332 ∗
Stock Market Crash	0.087	-0.361 ∗	0.195 ∗	0.48 ∗	-0.206 ∗	-0.344 ∗	0.113 ∗	0.336 ∗
Stock Market News	0.149 ∗	-0.390 ∗	0.332 ∗	0.64 ∗	-0.33 ∗	-0.333 ∗	0.24 ∗	0.537 ∗
Stock Market Today	0.184 ∗	-0.364 ∗	0.293 ∗	0.578 ∗	-0.262 ∗	-0.289 ∗	0.172 ∗	0.441 ∗
Stock Market	0.141 ∗	-0.342 ∗	0.263 ∗	0.563 ∗	-0.224 ∗	-0.321 ∗	0.181 ∗	0.426 ∗
Stock Price	0.709 ∗	-0.151 ∗	0.399 ∗	0.395 ∗	-0.117 ∗	-0.104	0.421 ∗	0.431 ∗
Stock to Buy	0.324 ∗	-0.14 ∗	0.169 ∗	0.315 ∗	0.023	-0.128 ∗	0.162 ∗	0.202 ∗
Stock Value	0.174 ∗	-0.057	0.232 ∗	0.192 ∗	-0.099	-0.116 ∗	0.306 ∗	0.158 ∗
Stock	0.491 ∗	-0.231 ∗	0.344 ∗	0.465 ∗	-0.140 ∗	-0.202 ∗	0.347 ∗	0.414 ∗

(∗ indicates p-value < 0.1)

Table 9: *Correlation matrix between search word volume and S&P, FTSE 100 financial data. [Personal elaboration].*

Analyzing the relationship between the search word time series and financial values we collected about FTSE 100 it is possible to notice how the results are consistent with the case of S&P 500, except for the closing price of the index. In this case, in fact, the words from our list are all negative correlated with FTSE 100 closing price, showing a result consistent with [Counts 11] and [Preis 13]. Therefore, considering the relationships between FTSE 100's mentioned values and word search volume, it is possible to notice a negative correlation with price and return and a positive one with trading volume and volatility. This may indicate that more people searching on financial terms, corresponds to an increase in market's volatility (higher values of VFTSE) and a raise in trading volume, while FTSE 100 prices will decrease, leading to a decline of return.

At this point, to narrow our data sample in order to conduct further tests, we computed a correlation analysis among the 20 financial terms we identified (see *Table 3*). From the outcome of this analysis we found a general positive correlation between the terms, so we decided to identify those couples of words with a correlation coefficient higher than 0.70 ($\gamma > 0.70$) and for every couple we retained only one word. After this process, our list was reduced to 11 terms, indicated in *Table 10.*

Bearish, Bullish, Finance, Financial News, FTSE 100, Market Value, S&P 500, Stock Decline, Stock Exchange, Stock Market Crash, Stock Price.

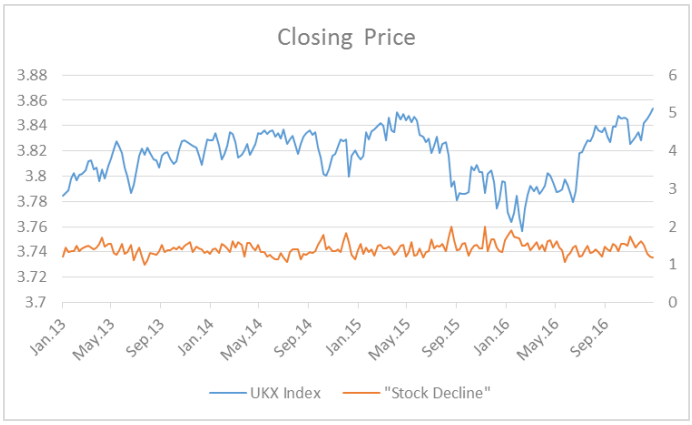
Table 10: *Updated list of search terms.*

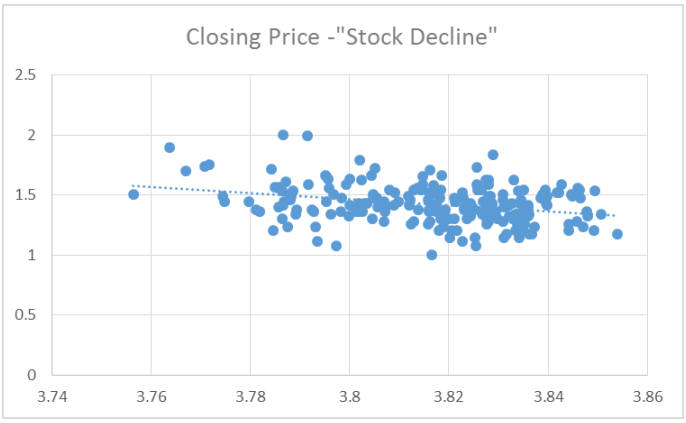
Among these terms, we chose the ones that showed the highest correlation with the mentioned financial data and we overlaid their time series with the financial indicators to visually examine the presence of any particular trend.

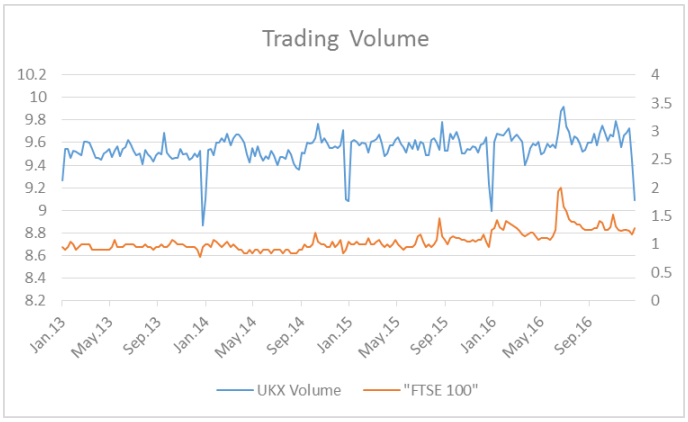
Figure 8 shows the actual time series of FTSE 100 financial data compared to search term time series (indicated in orange). A simple visual inspection reveals a similar trend between FTSE 100 trading volume and the search word time series, and the same it is true for volatility. In fact, peaks in search term time series often

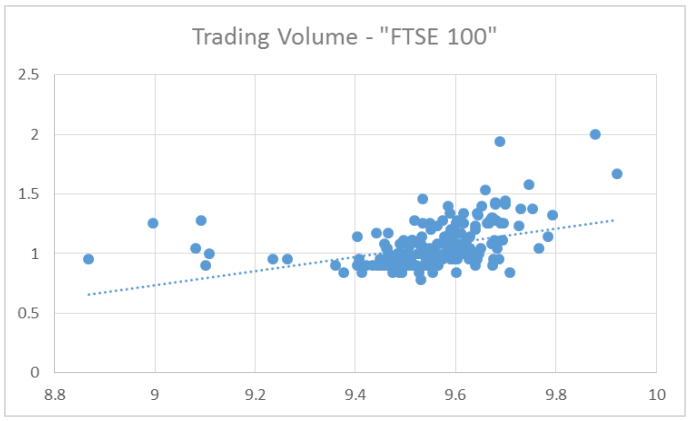
co-occur with those of VFTSE and trading volume. These positive correlations are clearly displayed by the respective scatterplots displayed in *Figure 8*. In the same way, FTSE 100 closing price's trend shows a negative correlation with search word time series, as indicated also by the scatterplot.

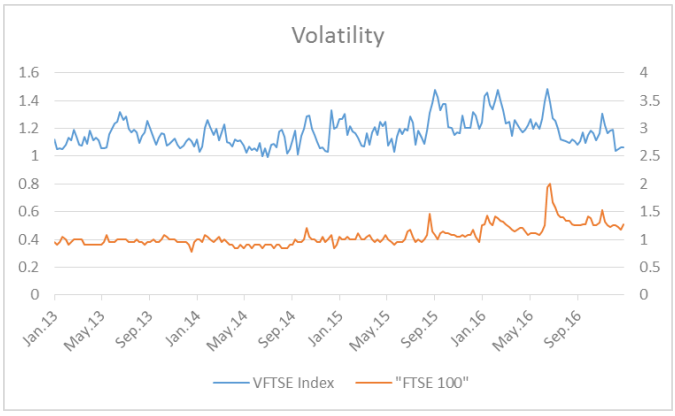
Figure 9 displays the time series of S&P 500 financial data in comparison to search terms time series (in orange). S&P 500 trading volume and volatility display a positive correlation with the search word time series considered, showing a similar behavior to their correspondent values of FTSE 100. Instead, as we already mentioned, S&P 500 closing price does not behave in a clear way, displaying positive correlation with some words (e.g. *stock price*) and negative correlation with others (e.g. *stock exchange*).











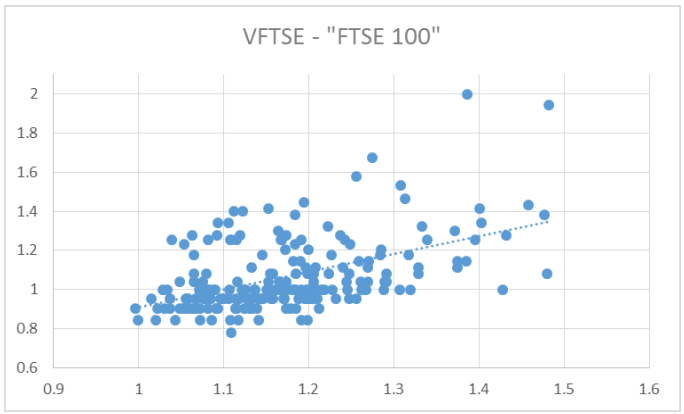


Figure 8: *Trend analysis and scatterplots of search word time series vs. financial indicators such as FTSE 100 closing price, trading volume and volatility (VFTSE) [Personal elaboration].*

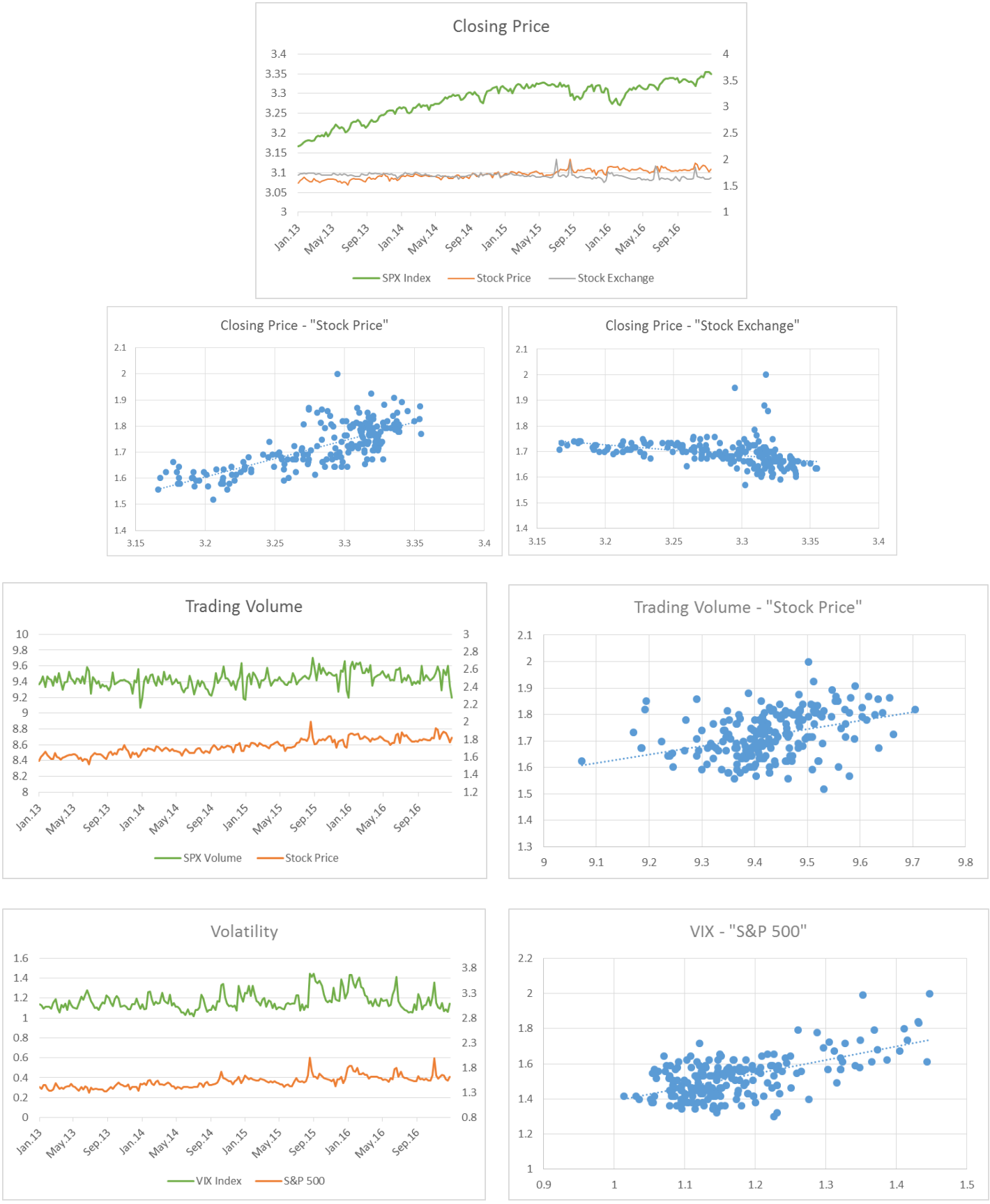


Figure 9: *Trend analysis and scatterplots of search word time series vs. financial indicators such as S&P 500 closing price, trading volume and volatility (VIX) [Personal elaboration].*

To further investigate the relationship between S&P 500 closing price and search volume, we plotted the time series of S&P 500 closing price and four words. In particular, we want to analyze how they react to each other in specific time periods. *Figure 10* shows, in the top panel, the time series of S&P 500 closing price from January 6th 2014 to January 1st 2017. The four time series in the lower panels display the trend of words such as *"bearish", "S&P 500", "stock exchange"* and *"stock price"* during the same time period.

The words we considered are the ones that showed a higher correlation with S&P 500 price: "*bearish"*, *"S&P 500" "stock price"* displayed positive correlation coefficients while *"stock exchange"* resulted negatively correlated. We identified four time periods when S&P 500 price fell sharply (October 2014, August 2015, January 2016, October 2016) and we marked them with rectangles.

When the price fell in October 2014, we notice an increase in the search volume of the word *"S&P 500"*, while there are no particular movements in the other words' trends. During the second bar (August 2015) we see all the words trending upward. The price fall of January 2016 coincides with an increase of all the search terms and also the last bar (October 2016) displays a similar situation. In conclusion, even though there is a considerable noise in the data, we realized that a decrease of S&P 500 closing price is often linked to an increase in search word volume. This may suggest a negative correlation between these sets of data, even though it is not represented by the correlation coefficients. A negative correlation between S&P 500 closing price and the search terms would have been consistent with our results about FTSE 100 as well as with [Counts 11] and [Preis 13].

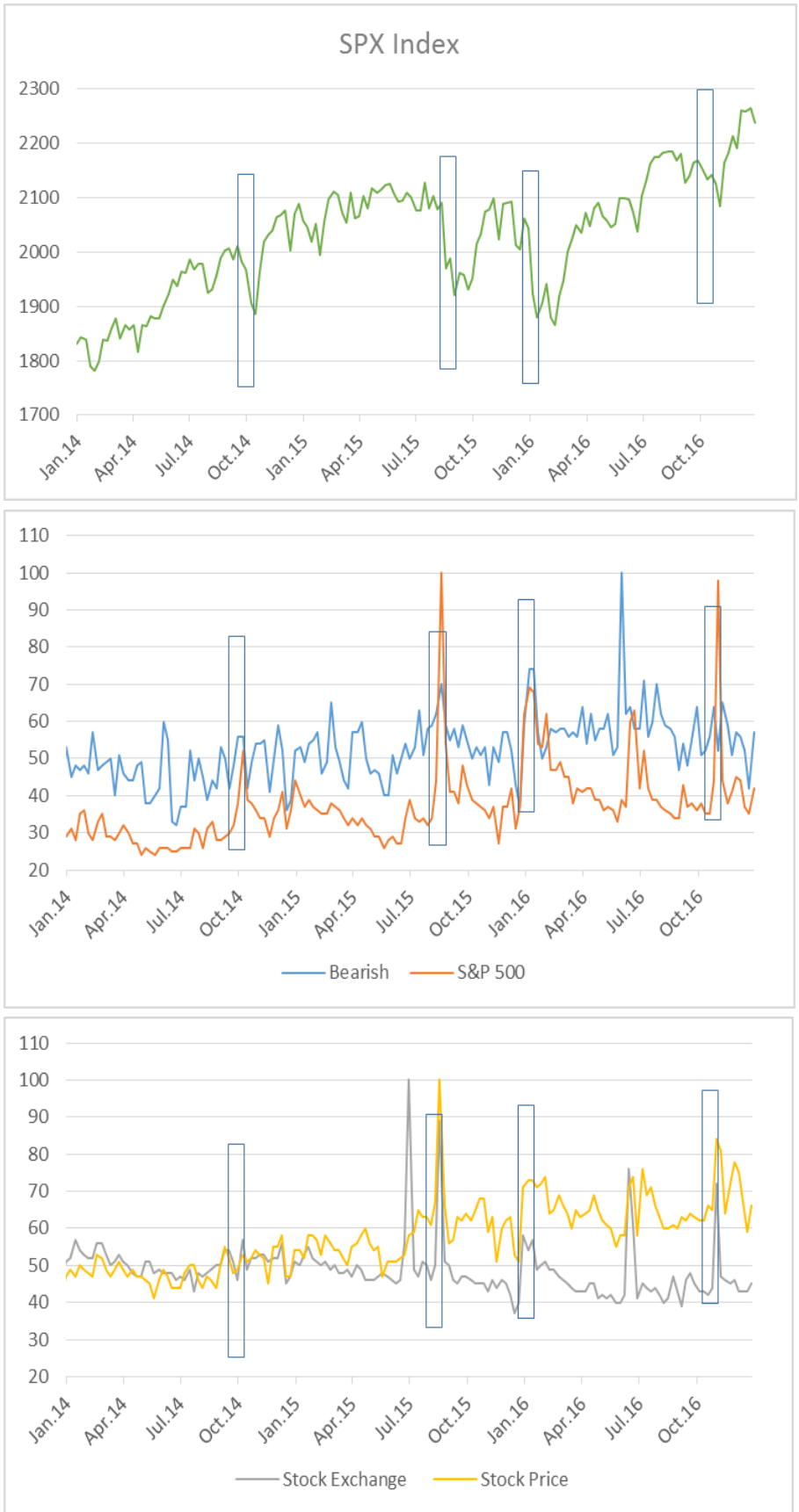


Figure 10: *Time series of S&P 500 closing price and four search words. [Personal elaboration].*

4 CONCLUSIONS

The main goal of this work is to present a preliminary assessment of the information contained in social media and whether it is possible to use them to track some financial markets variables. Moving from recent contributions in the literature, we carried out two analysis, one that considers an individual company stock while the other considers the whole stock market.

The first one deals with the analysis of Twitter messages to determine if the sentiment that emerges from them and their volume can influence Apple Inc. stock's trend. We collected financial data such as opening and closing price, high, low, return and trading volume, during a 30-days time period. We built a database of tweets about Apple Inc. and we submitted it to a sentiment classifier, *SentiStrength,* which ranked every message with a double score assessing their positivity and negativity. With this information about sentiments, we built two indicators that we called *Twitter Mood Indicator* (TMI) and *Compensated Strength Indicator* (CSI). Then we tested the correlation of tweets volume and these indicators with the stock values about Apple Inc. The results showed a negative correlation of mood indicators and tweet volume with prices and a positive one between tweet volume and trading volume. We noticed a similar behavior of our sentiment indicators and tweet volume, in relation to Apple Inc. stock's values. For this reason, we supposed that the information that explains the great part of these relationships is volume and not sentiment. It was not possible to comment the behavior of return and the relationship between the mood indicators and trading volume since these tests resulted not statistically significant.

The second analysis uses Google Trends as source of information and wants to investigate if the volume of some specific financial search words can influence financial indexes such as S&P 500 and FTSE 100. We collected the time series of 20 financial search terms during a time period of 4 years (2013-2016) then we analyzed the relationship between these data and closing price, return, trading volume and volatility of S&P 500 and FTSE 100. After conducting analytical and

graphical analysis, the results showed a negative correlation of search word volume with closing price and return and a positive one with volatility and trading volume.

The results presented here are promising, however, this research only has a preliminary nature, and given its purpose and the feature of the samples we used, it is important to take our conclusions with caution. Both the analysis we conducted showed interesting results but they deserve further research.

For organizational reasons, our Twitter-based analysis has been carried on during a particular time of the year, the period before, during and after Christmas 2016, so this could have been a distorting factor for our sample. In fact, our test produced many statistically not significant values and this could be the result of the mentioned distortion along with the short length of the considered period. To overcome these issues and enhance the research, it may be appropriate to consider a larger period of time, in order to collect a bigger data set and avoid any distortions due to too specific time frames. To further increase the sample of tweets, it is possible to add more stock related search words, other than just the cashtag of the company, and then filter out spurious tweets. Furthermore, instead of focusing on just one company it may be interesting to compare stocks from enterprises operating in different sectors, countries, etc.

We think that a big improvement of this analysis could be achieved using more sophisticated sentiment classifiers, for example more adjusted to stock market terminology. In fact, we think that among the large number of tweets ranked as neutral by SentiStrength, it is possible to extract additional sentiment information. This, along with an upgrade of the sentiment indicators' building process, may allow an enhancement of these indicators, make them able to better express Twitter's mood. The search volume-based analysis also presents many opportunities for improvement. For example, extending the analysis to other financial indexes and refining the list of search terms. Last, it may be interesting to combine both data about twitter sentiment and search volume and use them to build a trading strategy to assess whether and how it is possible to anticipate financial market's trends.

The study of the predictive power of online social data is still in its infancy. With our work, we tried to highlight strength and weakness of this new activity, discussing the limits and downsides we met during our analysis. We realized that extended research is needed in order to enhance our knowledge of why and how these data track and predict financial markets.

5 REFERENCES

[Adam 01] – Adams K. C, *The Web as Database: New Extraction Technologies and Content Management,* Online-Weston Then Wilton 25 (2), pp. 27-32, 2001.

[Alanyali 13] – Alanyali M., Moat H., S., Preis T., *Quantifying the Relationship Between Financial News and the Stock Market*, Sci. Rep. 3, 3578, 2013.

[Appelt 99] – Appelt D., *Introduction to Information Extraction.* AI Communications 12 (3), pp. 161-172, 1999.

[Baker 07] – Baker M., Wurgler J., *Investor Sentiment in the Stock Market,* The Journal of Economic Perspectives 21 (2), pp. 129-151, 2007.

[Baker 11] – Baker S., Fradkin A., *What Drives Job Search? Evidence from Google Search Data*, Discussion Papers 10-020, 2011.

[Basu 02] – Basu S., Banerjee A., Mooney R. J., *Semi-supervised Clustering by Seedings*, In Proceedings of the 19th International Conference on Machine Learning (ICML-2002), 2002.

[Batool 13] – Batool R., Khattak A. M., Maqbool J., Lee S., *Precise Tweet Classification and Sentiment Analysis,* 12th International Conference on Computer and Information Science (ICIS), pp. 461-466, 2013.

[Bordino 12] – Bordino I., Battiston S., Caldarelli G., Cristelli M., *Web Search Queries can Predict Stock Market Volumes,* PloS One 7 (7), 2012.

[Bollen 11] – Bollen J., Mao H., Zeng X. *Twitter Mood Predicts the Stock Market*, Journal of Computational Science 2 (1), pp. 1-8, 2011.

[Brill 92] – Brill E., *A Simple Rule-based Part of Speech Tagger*. In Proceedings of the Workshop on Speech and Natural Language, pp. 112-116, 1992.

[Cardie 06] – Cardie C., Farina C., Bruce T., Wagner E., *Using Natural Language Processing to Improve eRulemaking: Project Highlight,* In Proceedings of the 2006 International Conference on Digital Government Research, 2006.

[Challet 13] – Challet D., Bel Hadj Ayed A., *Predicting Financial Markets with Google Trends and not so Random Keywords*, arXiv 1307.4643, 2013.

[Choi 12] – Choi H., Varian H., *Predicting the Present with Google Trends*, Economic Record 88 (1), pp. 2-9, 2012.

[Cohn 94] – Cohn D., Atlas L., Ladner R., *Improving Generalization with Active Learning.* Machine Learning, 15 (2), pp. 201-221, 1994.

[Counts 11] – Counts S., Mao H., Bollen J., *Predicting Financial Markets: Comparing Survey, News, Twitter and Search Engine Data*, arXiv 1112.1051, 2011.

[Cristianini 12] – Cristianini N., Lampos V., *Nowcasting Events from the Social Web with Statistical Learning*, ACM Transactions on Intelligent System and Technology 3 (4), 2012.

[Da 15] – Da Z., Engelberand J., Gao P., *The Sum of All Fears: Investor Sentiment and Asset Prices*, Review of Financial Studies 28 (1), pp. 1-32, 2015.

[D'Andrea 15] – D'Andrea A., Ferri F., Grifoni P., Guzzo T., *Approaches, Tools and Applications for Sentiment Analysis Implementation,* International Journal of Computer Applications 125 (3), pp. 26-33, 2015.

[De Vries 12] – De Vries L., Gensler S., Leeflang P. S. H., *Popularity of Brand Posts on Brand Fan Pages: An Investigation of the Effects of Social Media Marketing,* Journal of Interactive Marketing 26 (2), pp. 83-91 2012.

[Dodds 11] – Doods P. S., Harris K. D., Kloumann I. M., Bliss C. A., Danforth C. M., *Temporal Patterns of Happiness and Information in a Global Social Network: Hedonometrics and Twitter,* Plos One 6 (12), 2011.

Donkor 13] – Donkor B., On Social Sentiment and Sentiment Analysis, 2013. http://brnrd.me/social-sentiment-sentiment-analysis/

[Downey 05] – Downey D., Etzioni O., Soderland S., *A Probabilistic Model of Redundancy in Information Extraction,* In Proceedings of the 19th International Joint Conference on Artificial Intelligence, pp. 1034-1041, 2005.

[Ederington 02] – Ederington L., Guan W., *Is Implied Volatility an Informationally Efficient and Effective Predictor of Future Volatility?*, Journal of Risk 4 (3), pp. 29-46, 2002.

[Eikvil 99] – Eikvil L., *Information Extraction from the World Wide Web: A Survey,* Technical Report 945, Norwegian Computing Center, 1999.

[Fan 11] – Fan T. K., Chang C. H., *Blogger-centric Contextual Advertising,* Expert Systems with Applications 38 (3), pp. 1777-1778, 2011.

[Feldman 13] – Feldman R., *Techniques and Applications for Sentiment Analysis,* Communications of the ACM 56 (4), pp. 82-89, 2013.

[Ginsberg 09] – Ginsberg J., Mohebbi M.H., Patel R.S., Brammer L., Smolinski M.S., Brilliant L. *Detecting Influenza Epidemics Using Search Engine Query Data*, Nature 457 (7232), pp. 1012-1014, 2009.

[Gonzalez 16] – Gonzalez C. B., Garcia-Nieto j., Navas-Delgado I., Anldana-Montes J. F., *A Fine Grain Sentiment Analysis with Semantics in Tweets,* International Journal of Interactive Multimedia and Artificial Intelligence 3 (6), pp. 22-28, 2016.

[Greenwood 16] – Greenwood S., Perrin A., Duggan M., *Social Media Update 2016,* Pew Research Center, 2016.

[Grishman 97] - Grishman R., *Information Extraction: Techniques and Challenges.* In "Information Extraction a Multidisciplinary Approach to an Emerging Information Technology", Springer Berlin Heidelberg, 1997.

[Hentschel 14] – Hentschel M., Alonso O., *Follow the Money: A Study of Cashtags on Twitter,* First Monday 19 (8), 2014.

[Huttunen 02] - Huttunen S., Yangarber R., Grishman *R., Complexity of Event Structure in Information Extraction*, In Proceedings of the 19th International Conference on Computational Linguistics, pp. 1-7, 2002.

[Kang 12] – Kang H., Yoo S. J., Han D., *Senti-lexicon and improved Naïve Bayes Algorithm for Sentiment Analysis of Restaurant Reviews,* Expert System with Applications 39 (5), pp. 6000-6010, 2012.

[Koppel 06] – Koppel M., Schler J., *The Importance of Neutral Examples for Learning Sentiment,* Computational Intelligence 22 (2), pp. 100-109, 2006.

[Lampos 12] – Lampos V. *On Voting Intentions from Twitter Content: a Case Study on UK 2010 General Election*, arXiv 1204.0423, 2012.

[Lampos 13] – Lampos V., Lansdall-Welfare T., Araya R.,Cristianini N., *Analysing Mood Patterns in the United Kingdom Through Twitter Content*, arXiv 1304.5507, 2013.

[Liu 06] – Liu B., *Web Data Mining,* Springer, 2006.

[Manning 99] – Manning C., Schutze H., *Foundations of Statistical Natural Language Processing,* MIT Press 999, 1999.

[Mao 11] – Mao H., Counts S., Bollen J., *Computational Economic and Finance Gauges: Polls, Search & Twitter,* In Meeting of the National Bureau of Economic Research-Behavioral Finance Meeting 11 (5), 2011.

[Mao 12] – Mao Y., Wei W., Wang B., Liu B., *Correlating S&P500 Stocks with Twitter Data*, In Proceedings of the 1st ACM International Workshop on Hot Topics on Interdisciplinary Social Networks Research, 2012.

[Maynard 11] – Maynard D., Funk A., *Automatic Detection of Political Opinions in Tweets,* Proceedings of the 8th International Conference on the Semantic Web, pp. 88-99, 2011.

[Mejova 09] – Mejova Y., *Sentiment Analysis: An Overview*, Comprehensive Exam Paper, 2009.

[Mitkov 05] – Mitkov R., *The Oxford Handbook of Computational Linguistics*, Oxford University Press, 2005.

[Mittal 12] – Mittal A., Goel A., *Stock Prediction Using Twitter Sentiment Analysis,* Stanford University, 2012.

[Moat 13] – Moat H.S., Curme C., Avakian A, Kenett D.Y., Stanley H.E., Preis T. *Quantifying Wikipedia Usage Patterns Before Stock Market Moves,* Scientific Reports 3, 2013.

[Newberry 16] – Newberry C., *Top Twitter Demographics That Matter to Social Media Marketers,* 2016. https://blog.hootsuite.com/twitter-demographics/

[Nofsinger 05] – Nofsinger J. R., *Social Mood and Financial Economics*, The Journal of Behavioral Finance 6 (3), pp.144-160, 2005.

[Norman 11] – Norman G. J., Norris C. J., Gollan J., Ito T. A., Hawkley L. C., *Current Emotion Research in Psychophysiology: The Neurobiology of Evaluative Bivalence,* Emotion Review 3 (3), pp. 349-359, 2011.

[Oliveira 13] – Oliveira N., Cortez P., Areal N., *Some Experiments on Modeling Stock Market Behaviour Using Investor Sentiment Analysis and Posting Volume from Twitter*, In Proceedings of the 3rd International Conference on Web Intelligence, Mining and Semantics, 2013.

[Paliouras 99] – Paliouras G., Karkaletsis V., Papatheodorou C., Spyropoulos C., *Exploiting Learning Techniques for the Acquisition of User Stereotypes and Communities,* UM99 User Modeling, pp. 169-178, Springer, 1999.

[Pang 08] - Pang B., Lee L. *Opinion Mining and Sentiment Analysis.* Foundation and Trends in Information Retrieval 2 (1-2), pp. 1-135, 2008.

[Pennebaker 07] – Pennebaker J.W., Chung C.K., Ireland M., Gonzales A., Booth R. J., *The Development and Psychometric Properties of LIWC2007*, The University of Texas at Austin, 2007.

[Pepe 11] – Pepe A., Bollen J., Mao H., *Modelling Public Mood and Emotion: Twitter Sentiment and Socio-Economic Phenomena,* International Conference on Web and Social Media 11, pp. 450-453, 2011.

[Piskorski 13] - Piskorski J., Yangarber R., *Multi-source, Multilingual Information Extraction and Summarization,* Springer Science & Business Media, 2013.

[Polgreen 08] – Polgreen P. M., Chen Y., Pennock D.M., Nelson F.D., Weinstein R.A. *Using Internet Searches for Influenza Surveillance,* Clinical Infectious Diseases 47 (11), pp. 1443-1448, 2008.

[Preis 10] – Preis T., Reith D., Stanley H. E., *Complex Dynamics of Our Economic Life of Different Scales: Insights from Search Engine Query Data,* Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences 368 (1933), pp. 5707-5719, 2010.

[Preis 13] – Preis T., Moat H.S., Stanley H.E., *Quantifying Trading Behavior in Financial Markets Using Google Trends*, Scientific Reports 3, pp. 1684, 2013.

[Quirk 10] – Quirk R., Greenbaum S., Leech G., Svartvik J., *A Comprehensive Grammar of the English Language,* Pearson Education India, 2010.

[Ranco 15] – Ranco G., Aleksovski D., Caldarelli G., Grcar M., Mozetic I., *The Effects of Twitter Sentiment on Stock Price Returns,* PLoS One 10 (9), 2015.

[Rice 13] – Rice D. R., Zorn C., *Corpus-Based Dictionary for Sentiment Analysis of Specialized Vocabularies,* In Proceedings of New Directions in Analyzing Text ad Data Workshop (NDATAD), pp. 98-115, 2013.

[Rijsbergen 79] – Rijsbergen Van C.J., *Information Retrieval,* Butterworths, London, 1979.

[Ronk 2014] – Ronk J., *Structured, Semi-structured and Unstructured Data*, 2014. https://jeremyronk.wordpress.com/2014/09/01/structured-semi-structured-and-unstructured-data/

[Ruiz 12] – Ruiz E., Hristidis V., Castillo C., Gionis A., Jaimes A, *Correlating Financial Time Series with Micro Blogging Activity*, In Proceedings of the 5th ACM International Conference on Web Search and Data Mining, pp. 513-522, 2012.

[Saavedra 11] – Saavedra S., Duch J., Uzzi B., *Tracking Traders' Understanding of the Market Using e-Communication Data*, PloS One 6 (10), 2011.

[Sakaki 16] – Sakaki T., Okazaki M., Matsuo Y., *Earthquake Shakes Twitter Users: Real-time Event Detection by Social Sensors,* In Proceedings of the 19th International Conference on World Wide Web, pp. 851-860, 2016.

[Schumaker 12] – Schumaker R. P., Zhang Y., Huang C. N., Chen H., *Evaluating Sentiment in Financial News Articles,* Decision Support System 53 (3), pp. 458-464, 2012.

[Sheth 16] – Sheth A., #Brexit: "There is a Big Trouble for #remain" — Some Lessons from Real-time #socialmedia Analysis, 2016. https://www.linkedin.com/pulse/brexit-big-trouble-remain-some-lessons-from-real-time-amit-sheth

[Smeaton 98] – Smeaton A.F., *Retrieving Images of Scanned Text Document,* In Proceedings of the Optical Engineering Society of Ireland & Irish Machine Vision and Image Processing Joint Conference, pp. 271-286, 1998.

[Carrière-Swallow 13] – Carrière-Swallow Y., Labbé F., *Nowcasting with Google Trends in an Emerging Market,* Journal of Forecasting 32 (4), pp. 289-298, (2013).

[Thelwall 10] – Thelwall M., Buckley K., Paltoglou G., Cai D., Kappas A., *Sentiment Strength Detection in Short Informal Text*, Journal of the American Society for Information Science and Technology 61 (12), pp. 2544-2558, 2010.

[Thelwall 12] – Thelwall M., Buckley K., Paltoglou G., *Sentiment Strength Detection for the Social Web,* Journal of the American Society for Information Science and Technology 63 (1), pp. 163-173, 2012.

[Thelwall 16] – Thelwall M., *The Hearth and Soul of the Web? Sentiment Strength Detection in the Social Web with SentiStrength*, Cybermotions, pp. 119-134, Springer International Publishing, 2016.

[Tianhao 02] - Tianhao W., *Theory and Applications in information extraction from unstructured text"*, Thesis and Dissertations, Paper 741, 2002.

[Turney 03] – Turney P. D., Littman M. L., *Measuring Praise and Criticism: Inference of Semantic Orientation from Association,* ACM Transactions on Information Systems (TOIS) 21 (4), pp. 315-346, 2003.

[Vosten 11] – Vosten S., Schmidt T., *Forecasting Private Consumption: Survey-Based Indicators vs. Google Trends,* Journal of Forecasting 30 (6), pp. 565-578, 2011.

[Wang 12] – Wang H., Can D., Kazemzadeh A., Bar F., Narayanan S., *A System for Real-time Twitter Sentiment Analysis of 2012 U.S. Presidential Election Cycle,* In Proceedings of the ACL 2012 System Demonstrations, pp. 115-120, 2012.

[Wang 13] – Wang D., Zhu S., Li T., *SumView: A Web-based Engine for Summarizing Product Reviews and Customer Opinions,* Expert Systems with Applications 40 (1), pp. 27-33, 2013.

[Wiebe 04] – Wiebe J. M., Wilson T., Bruce R., Bell M., Martin M., *Learning Subjective Language,* Computational Linguistic 30 (3), pp 277-308, 2004.

[Zhang 10] – Zhang W., Skiena S., *Trading Strategies to Exploit Blog and News Sentiment,* In Proceedings of the 4th International AAAI Conference on Weblogs and Social Media, pp. 375-378, 2010.

[Zhang 11] – Zhang X., Fuehres H., Gloor P.A., *Predicting Stock Market Indicators Through Twitter "I Hope it is not as Bad as I Fear"*, Procedia-Social and Behavioral Sciences 26, pp. 55-62, 2011.

[Zheludev 15] – Zheludev I., N., *When Can Social Media Lead Financial Markets,* PhD Dissertation, University College London, 2015.

Web Sites

AlchemyAPI, http://www.alchemyapi.com/

Google Trends, https://www.google.com/trends/

LIWC – Linguistic Inquiry and Word Count, http://www.liwc.net/

Sentiment140, http://www.sentiment140.com/

SentiStrength, http://sentistrength.wlv.ac.uk/

The Stock Sonar, http://www.thestocksonar.com

Tweetfeel, http://www.tweetfeel.com

Twitter Archiver, https://chrome.google.com/webstore/detail/twitterarchiver/pkanpfekacaojdncfgbjadedbggbbphi

6 APPENDIX

	Opening Price	High	Low	Closing Price	Return	Trading Volume	TMI	TMI*	CSI	CSI*	Tweet Volume
Opening Price	1										
High	0.937 ∗	1									
Low	0.965 ∗	0.95 ∗	1								
Closing Price	0.925 ∗	0.968 ∗	0.964 ∗	1							
Return	-0.118	0.141	0.063	0.208	1						
Trading Volume	-0.272	-0.194	-0.295	-0.237	0.358	1					
TMI	-0.403 ∗	-0.381 ∗	-0.356	-0,362 ∗	0.181	0.18	1				
TMI*	-0.403 ∗	-0.381 ∗	-0.356	-0.362 ∗	0.181	0.18	1 ∗	1			
CSI	-0.43 ∗	-0.419 ∗	-0.388 ∗	-0.401 ∗	0.189	0.243	0.981 ∗	0.981 ∗	1		
CSI*	-0.43 ∗	-0.419 ∗	-0.388 ∗	-0.401 ∗	0.189	0.243	0.981 ∗	0.981 ∗	1 ∗	1	
Tweet Volume	-0.46	-0.315	-0.384 ∗	-0.344	0.305	0.466 ∗	0.42 ∗	0.42 ∗	0.432 ∗	0.432 ∗	1

(∗ indicates p-value < 0.1)

Table 1: *Correlation matrix between Twitter sentiment, volume and Apple Inc. financial data. [Personal elaboration].*

	Opening Price	High	Low	Closing Price	Return	Trading Volume	TMI	TMI*	CSI	CSI*	Tweet Volume
Opening Price	1										
High	0.916 ∗	1									
Low	0.941 ∗	0.939 ∗	1								
Closing Price	0.897 ∗	0.961 ∗	0.952 ∗	1							
Return	-0.347	0.023	-0.123	0.069	1						
Trading Volume	-0.275	-0.172	-0.323	-0.246	0.364	1					
TMI	-0.347	-0.413 ∗	-0.349	-0.451 ∗	-0.068	0.115	1				
TMI*	-0.347	-0.413 ∗	-0.349	-0.451 ∗	-0.068	0.115	1 ∗	1			
CSI	-0.379 ∗	-0.445 ∗	-0.385 ∗	-0.489 ∗	-0.064	0.161	0.981 ∗	0.981 ∗	1		
CSI*	-0.379 ∗	-0.445 ∗	-0.385 ∗	-0.489 ∗	-0.064	0.161	0.981 ∗	0.981 ∗	1 ∗	1	
Tweet Volume	-0.398 ∗	-0.418 ∗	-0.382 ∗	-0.444 ∗	-0.072	0.436 ∗	0.42 ∗	0.42 ∗	0.433 ∗	0.433 ∗	1

(∗ indicates p-value < 0.1)

Table 2: *Correlation matrix between Twitter sentiment, volume and Apple Inc. financial data, with a 1-day time lag. This matrix* considers today's tweets and tomorrow's market values *[Personal elaboration].*

	Opening Price	High	Low	Closing Price	Return	Trading Volume	TMI	TMI*	CSI	CSI*	Tweet Volume
Opening Price	1										
High	0.92 ∗	1									
Low	0.95 ∗	0.934 ∗	1								
Closing Price	0.911 ∗	0.958 ∗	0.961 ∗	1							
Return	-0.029	0.243	0.155	0.324	1						
Trading Volume	-0.235	-0.146	-0.261	-0.201	0.364	1					
TMI	-0.532 ∗	-0.525 ∗	-0.591 ∗	-0.519 ∗	-0.058	-0.027	1				
TMI*	-0.532 ∗	-0.525 ∗	-0.591 ∗	-0.519 ∗	-0.058	-0.027	1 ∗	1			
CSI	-0.558 ∗	-0.607 ∗	-0.607 ∗	-0.552 ∗	-0.066	0.018	0.981 ∗	0.981 ∗	1		
CSI*	-0.558 ∗	-0.607 ∗	-0.607 ∗	-0.552	-0.066	0.018	0.981 ∗	0.981 ∗	1 ∗	1	
Tweet Volume	-0.522 ∗	-0.504 ∗	-0.504 ∗	-0.43 ∗	0.009	0.095	0.46 ∗	0.46 ∗	0.464 ∗	0.464 ∗	1

(∗ indicates p-value < 0.1)

Table 3: *Correlation matrix between Twitter sentiment, volume and Apple Inc. financial data, with a 1-day time lag. This matrix* considers today's market values and tomorrow's tweets *[Personal elaboration].*