Applying Natural Language Processing to Financial Sentiment Analysis

By Dr. Olga Petrova

Professor of Information Technology, Mälardalen University, Sweden

1. Introduction to Financial Sentiment Analysis

1. Introduction

Sentiment analysis is the process of recognizing, categorizing, and measuring sentiment expressed in texts. Within the field of finance, sentiment analysis is quickly becoming increasingly important. The finance industry has long been aware of how mood can alter the dynamics of proceedings in financial markets and the way in which investors make choices regarding trading and investment. However, the area of sentiment analysis for financial markets has expanded far beyond just investor sentiment and is currently an intrinsic part of daily business operations across a wide variety of fields and niche focuses. In the data-driven finance sector of today, accumulating numerous resources allows for the correct estimation of the sentiment of any subject.

In the past, company performance reports packaged in large annual documents were the main source of available information for decision-making in finance. However, while the nature of these documents is not changing, the methods of communication in a company are. Many announcements are made to the market through the use of press releases, emails, presentations, newspaper articles, news site articles, webcasts, and much more. This large expanse of resources allows entry into the world of unclear financial communication. Many businesses also hold a presence on social media in many forms. Consequently, the ability for financial managers to measure the sentiment around these discussions is of the utmost importance for gaining a realistic insight into what stakeholders and the public think of the company's products, services, or even the chief executive. Furthermore, the ability of investors to assess whether publicly traded companies and cash flows carry an overall positive or negative attitude offers numerous improvements in terms of risk assessment and prediction to company filings.

1.1. Definition and Importance of Sentiment Analysis in Finance

Sentiment analysis, also called opinion mining, refers to the interpretation of people's sentiments and emotions expressed in textual form. In finance, sentiment analysis involves the detection of financial sentiments and opinions to analyze investors' moods, which impact markets. One of the primary reasons to investigate the sentiment expressed in the texts is that it has the potential to influence trading strategies, investment positions, and the overall perception of the market. There are already different metrics and indicators to approximate the feelings of texts, for example, by three main indexes: Consumer Confidence Index, Business Climate Indicator, and the Investor’s Confidence Index.

Sentiment is seen by many researchers as a leading indicator that moves the stock markets after some period. Even the degree of interactivity between news and investors can be related to the stock market. One crucial and leading application is the possibility of predicting in advance some movements in the stock market, indexes, and currencies according only to the sentiment of people whose opinions are accessible to the trading agents. However, capturing the feelings is not always a trivial task, and for finance, especially, it becomes less proven yet. The most challenging part of mining and measuring sentiment is how to describe sentiments accurately among investors and, accordingly, say whether the information is a positive, negative, or neutral clue to the price of a commodity. Some of the software had to expand dictionaries and blacklists in order to properly label the words and exclamations provided by scheduled people. This classification is negotiating a considerable space in focusing on improving trading strategies and the possible predictions in financial markets.

2. Fundamentals of Natural Language Processing (NLP)

Natural Language Processing (NLP) is an interdisciplinary field of computer science, artificial intelligence, and linguistics that enables computers to understand, interpret, provide context to, and manage human language. NLP is a field of computer science that focuses on text data analysis, which is an important asset for businesses of all kinds across the globe. Some common tasks in NLP include tokenization, named entity recognition, relation extraction, part-of-speech tagging, parsing, and sentiment detection. Named entity recognition suggests the detection of specific entities present in text with respect to a predefined set of categories. Sentiment detection infers the prevailing public perception towards an organization, idea, or individual as indicated in text content. Linguistic structures and semantics are crucial elements for interpreting and processing unstructured data using NLP, owing to the fact that language use has contextual significance.

While the integration of data capturing techniques and the application of new algorithms have significantly sped up the advent of textual information, the evolution of related subfields of computer science, with an emphasis on deep learning in recent years, has transformed theoretical findings of the past decades into useful practices. Machine learning and deep learning are gaining popularity as techniques applied in various NLP functionalities such as sentiment analysis, named entity recognition, document classification, part of speech, and natural language understanding, among others. Despite the improvement in machine learning and deep learning techniques, understanding human language presents great challenges due to ambiguity and language dependency. Positive advancements are being made in industries such as healthcare, online retail, finance, e-commerce, and government due to the successful transformation of theoretical findings into NLP technologies, providing exciting new standards. Research findings also project that industries rich in data content, finance, and marketing will exhibit faster positive growth compared to the rest in the near future.

2.1. Key Concepts and Techniques in NLP

The material presented in this section is essential to understand how NLP can be taken advantage of for the purpose of financial sentiment analysis. Named entity recognition, sentiment classification, part of speech tagging, and syntactic parsing are common NLP tasks and techniques. Sentiment classification attempts to determine the sentiment or subjective evaluation expressed within text. Preprocessing steps are essential for the effectiveness of sentiment classification, such as text normalization, stop-word removal, and case conversion. Algorithms can be used for sentiment classification such as support vector machines, decision trees, naive Bayes, neural networks, and long short-term memory.

A sentiment classifier is developed by building a supervised learning method. To train, trade information has been labeled with the analyst's report from the datasets. This labeling is used to determine whether the return on the trade is positive, negative, or insignificant over price limit for the predictors. Additionally, it is also important to mention that good accuracy can only be achieved if most of the training data used is relevant. By selecting the correct features for training data, the accuracy of the model can also increase. Therefore, lexical analysis, time aspect, named entity, and confidence features are selected to be encoded in numerical values so that the machine learning algorithm can be adapted. Numerous studies demonstrate that NLP text mining techniques can successfully be applied in the field of finance to gain useful information for decision-making. However, more work relevant to the specific linguistic patterns used in finance remains. Further, applying NLP in relation to financial time series analysis to enhance trading is a relatively unused research topic to date but has great potential.

3. Integration of NLP in Financial Sentiment Analysis

The combination of NLP and financial decision-making represents a growing field of research at the intersection of finance, computer science, and psychology. Computational approaches have the potential to mine meaningful sentiments from the flood of financial narrative data more efficiently than humans. Moving beyond simple keyword matching, NLP techniques can extract important entities and expressions underlying company activities and investor responses. Computational analyses might also provide evidence of investor sentiment in financial data by integrating new psychological and lexical knowledge. In this way, linguistic tools can help inform our understanding of previously mysterious and complex financial data series on their own, as well as via a complementary process of creative integration of various data sources.

A common goal of financial sentiment analysis is to automate the extraction of sentiment from diverse texts discussing financial firms, markets, or the broader economy. There are a variety of ways to automate this process, with different techniques geared towards different sub-goals. Common status differences between financial sources can be exploited to assign sentiments that help balance various perspectives in the financial media. This text is concerned with financial events rather than markets in general, and as such, focuses on the literature at the intersection of NLP and financial analysis. Sentiment calculation techniques can be built using either supervised or unsupervised learning. Supervised approaches use labeled examples to find a mapping between document features and sentiment scores. Spin theory predicts that the addition of opinion is smaller than the average absolute value of the underlying fact. Sentiment expresses opinion or evaluation about the data fact. The recent literature uses NLP methods to mine sentiment hidden in traditional financial documents such as news, blogs, analyst reports, and company filings. Datasets from these financial events can be used in isolation or in various combinations. Challenges in parsing these heterogeneous sources often arise from differences in data quality, sentiment granularity, and extraction approaches. In finance, developing methods based on the analysis of company reports, news, and market chatter around firms can not only foster a better understanding of information flow but can also enhance market, economic, and political prediction models. For example, anticipating a company's earnings announcement might be used to make trading decisions on the firm or its peers, predict the stock market, or serve as a forecast of the underlying economy.

3.1. Challenges and Opportunities

Currently, two different types of sentiment analysis approaches dominate financial markets and empirical research, one being the classical approach and the other, the emerging news part of NLP. The literature on sentiment analysis using NLP in finance is not as mature as the classic approach, although there is a growing trend.

The efforts in applying NLP for financial sentiment analysis face several challenges. For instance, financial language is replete with ambiguities, resulting in increased discrepancies. That is, it displays the large variability of natural language. Different words might have different meanings in finance. Moreover, as the representation of sentiment is context-sensitive, it is commonly challenging to differentiate between the positive and negative meanings of a document based only on its textual context. In addition, some works claim that using NLP brings bias from the training data. In contrast to the challenges, the opportunities associated with employing NLP for sentiment analysis are promising regarding the supplemental technology in the finance area, such as better risk management, based on a fundamental and more precise understanding of the trading behavior arising from multi-modal fatigue sentiments. Additionally, the exploitation of textual data for sentiment analysis would lead to better probabilistic sentiment predictions, supporting more trustworthy advances in risk prediction, portfolio decisions, and range forecasting.

Financial sentiment analysis is receiving a lot of interest and attention for several reasons. Investors and financial analysts tend to make trading, investment, and risk management decisions based on the news and financial reports. The ability to automatically harness sentiment is important because relevant financial information volumes are too large to be analyzed manually. In existing literature, there are two dominating approaches to financial sentiment analysis. The first one is the classical approach, which uses language and machine learning techniques. The second one is the emerging news part of NLP. The news part of NLP is the key emphasis of the current study. In the following discussion, we analyze both the challenges and opportunities of utilizing NLP for financial sentiment analysis. Bringing the dual analysis gives a wider view. This opportunity can increase innovation in applying sentiment analysis.

4. AI Models for Market Forecasting

Stock market prediction is a core element of trading strategies. The adoption of intelligent agents for the analysis of stock markets and to predict their movements has long been of growing interest. By giving a machine the ability to behave or react to uncertain agent behavior in environments that require intelligence, intelligent agents are expected to automate trading, risk, and portfolio management to surpass humans in these tasks. Many models rely on historical data and sentiment analysis to forecast financial markets. One of the reasons for choosing NLP techniques for sentiment analysis is that these models learn and predict effectively in the preprocessed text data. Nowadays, with big data from various sources available, AI has gained much attention as a useful tool in building models that provide signals to help improve trading performance. However, be aware of the use of large datasets for training. The meaning extracted from raw text is encoded into features, which are used to train a model. Robust models require real-world features; in many cases, big datasets are needed for proper supervised models.

Incorporating AI into existing trading algorithms for executing market orders is an active research area in stock trading. AI models in financial forecasting can range from regression models to time series analysis. Sentiment as a market indicator is a non-price metric and is focused on market expectations. Market sentiment and media news have high correlations, leading to widespread use of related models. The use of AI is effective in doing this as it reacts similarly to a person interacting in a scenario. This is also the reason so much research has flourished in this direction. The effectiveness of using sentiment stems from studies that show the effect of news sentiment on inflation rates and on the predictive capabilities of analyst stock ratings. Embedding techniques can be effective in this context. Building accurate models is an art as well as a science. This is due to the fact that the success of an event is based on various aspects of the environment. Market sentiment is combined with economic indicators to show its impact. A higher number of features helps to improve predictive accuracy. The use of technical indicators with market sentiment increases model accuracy. AI models have been successful in stock market prediction. Better predictive models are achieved by combining quantitative analysis with qualitative analysis. Market reaction to news, historical movement of stock prices, and transaction volume have been effective in providing buy or sell signals. Market reaction to news is lower than the financial indicators. Integration of news with financial stock indicators is used to predict stock movement.

4.1. Machine Learning Approaches for Forecasting

There is a rich body of current research that shows that machine learning methods, often referred to as data mining techniques, might enhance the market's forecasting capabilities. In general, a simple learning algorithm and a standard statistical model can perform competitively. Among various approaches, neural networks, decision trees, and ensemble methods started to really shine when it comes to modeling in forecasting contexts, outperforming classical models and rising as one of the new directions in efficient time series forecasting. Such algorithms are able to find patterns in data, which enhance overall forecasting accuracy. Of course, there were cases where machine learning models have achieved around 80% accuracy in stock market prediction. Selecting an appropriate machine learning model and fitting it to the task of interest implies a number of considerations regarding training, model choice, and model validation and evaluation. The latter task is represented mainly by choosing a predictive model and providing some measures of the quality of the forecasts. The results generated by these models should be scrutinized to ensure that they really made the data generate the 'right' results.

A common issue in forecasting is model overfitting. One option to reduce its likelihood is model validation, where the model is tested out of a sample, over which the algorithm did not train. Another important task in machine learning is feature selection or feature engineering, where engineers select the most interesting data inputs to be used by the learning algorithm. This task is of particular interest in the sentiment analysis field, since recent investigations revealed that the incorporation of textual and sentiment information delivers improved predictive accuracy. Some researchers have used sentiment analysis as an independent variable. Pertinently, incorporating a sentiment-derived indicator into an asset-pricing model can lead to improved forecasting ability. Furthermore, trading models incorporating sentiment features can consistently outperform models based uniquely on the past price return. These results suggest new directions for the best application of text-mining techniques in financial economics.

Model validation and feature selection are key aspects of a machine learning approach aimed at financial forecasting. Rigorous testing is required in contemporary finance, because the world is now so competitive that new forecasting methods, however slight the improvement on standard practice, are invaluable. For the purpose of this paper, case studies on the machine learning model application to the same domain for two research areas. We found that using a simple random forest model along with risk-sensitive predicted non-sentimental returns is the best solution with the lowest mean squared error. Because our models are mid-term predictors, they cannot be directly applied in financial markets, whereby the utility is unknown.

5. AI Models for Investment Decision Support in Banking

Banks rely on insights into the future development of market prices to support the investment decisions of their clients. Consequently, investment firms are eager to develop strategies drawn from relevant data. One pathway for doing so is the implementation of supervised machine learning models that predict future asset performance. Distinguishing between different AI approaches, the first comprises normal contemporary algorithms, such as gradient boosted decision trees, support vector machines, and random forests. The second, more detailed approach includes state-of-the-art deep learning models that have gained traction in recent years.

In parallel to predicting asset performance more accurately for the purpose of portfolio management, investment banks around the globe turned to sentiment analysis to get a sense of the market's mood. Investment banks have been leveraging AI to improve various processes. In the field of finance, such AI algorithms autonomously assist with investment decisions, where the AI algorithm either makes a suggestion to the bankers, which bankers can then either agree or disagree with, or supports the banker with additional insights that are not included in the suggestion. As such, AI aims to improve the performance of the banker rather than replace his or her decision-making process altogether. AI assists with the personalization of risk management and investment strategies by providing small retail investors, as well as wealth management clients, with advice that is automated and tailored to one's unique risk profile and behavioral preferences. Essentially, machine learning can thus support a form of client relationship management that traditionally could only have been done manually.

AI challenges remain concerning the regulatory compliance with the General Data Protection Regulation, as well as with the Financial Supervisory Authority regulations regarding personal advice. Moreover, it is essential to ensure the explainability of the investment suggestions produced by the model, as a lack of model transparency is one of the main reasons for non-adoption by banks. It is still a work in progress to ensure that the computational process used for the personal suggestions can be explained to state authorities and the bank's clients.

5.1. Use Cases and Case Studies

Investment Decision Support in Banking

Four use cases describe investment decision support where advanced AI models were used to classify and rank potential investment opportunities. These were real-life client projects, and the banks are kept anonymous. 1. Sentiment Analysis for Stock Market Prediction. A bank wanted a model to support investment teams in identifying potentially relevant stocks to invest in. Using news sentiment data, we developed a model to evaluate the synthesized news sentiment and labeled press releases to train and validate it. A strategy to integrate the results into the tactical asset allocation was then developed. 2. Airport Rankings Using News Sentiment. A model was developed to rank airports in the US and internationally. Results were integrated into a global sourcing investment committee. News data comes from various sources. 3. Sentiment Analysis for M&A Integration Risk Management. M&A deals integrating two companies offer a large potential to create value as well as the risk of value impairment. A large proportion of the risk can be mitigated using a rigorous approach to mapping the areas that will impact the success of integration, prioritizing them and managing them post-completion. An innovative analysis was developed to quickly and accurately create that mapping using the factors that really count: key employees who will be working together on day one. It was considered whether financials would provide an early warning signal of successful integration. It was determined that while the people factors matter, financials capture the net result of all potential liabilities and are, in fact, a very good indicator of how well a solution could benefit a new M&A deal. Moreover, the analysis suggests that during integration, two companies need to manage this risk by producing financials that are better than the combined average of the two companies as a whole, and this is true even if they are not actually sharing workers or assets to the extent suggested by the financials.

6. Future Direction

In the future, techniques and methods from deep learning and natural language processing are likely to further advance the methods used for sentiment extraction and benchmark results, together with innovative types of data. As technology becomes ever more sophisticated, especially due to progress in natural language processing and machine learning, we will see a development towards more accurate (pseudo-) sentiment labels and tools for sentiment analysis that are based on various types of advanced AI models. This trend aligns with the growing importance of real-time sentiment analysis and machine reading for the automatic generation of strategies that decide on buying and selling decisions in the liquid, high-frequency part of financial markets. The connectivity and data continuum are now more important than model development. Therefore, it will likely become at least as important to have agility in integrating new data as it is to focus on the statistical and quantitative models developing what to do with it.

The analysis predicts rising trends in financial sentiment analysis, in terms of changes in tools, models, and methodologies. However, applications of NLP in financial sentiment analysis going forward largely depend on regulatory demand and market dynamics. As financial regulators around the world are increasingly interested in alternative data, which include real-time news and social media, corporate suits, and earnings calls, there is also an increasing tendency to influence algorithmic sales and operations through the desire to tap into individually attributable industry metrics. The academic literature investigating the use of NLP in financial services typically also draws findings related to ethical challenges and incentive problems when employing machines in predominantly human judgment and decision-making environments. While this discourse is especially important when developments are proposed and tested in areas of high societal impact, this study will not delve into ethical and regulatory considerations.

7. Conclusion

Natural Language Processing (NLP) has played a very crucial role in the field of stock market prediction and financial sentiment analysis. It has forced the finance industry to think beyond the traditional methods of decision making and the study of behavioral finance. NLP provides models that with the help of advances in Artificial Intelligence (AI) can assist in the limitless possibilities of market prediction, sentiment analysis and investment recommendation. There are many challenges posed to the application of AI and machine learning to financial sentiment analysis. For once, NLP and machine learning are data driven models that need to be reinforced with constant data to keep them updated with changing dynamics of the market. Many AI models and machine learning models have been proposed that need to be employed on a much larger scale for generalization of research. AI applications have posed many questions in finance, business and many other fields about the potential of their successful application and expansion of their scope of work over time.

Natural Language Processing (NLP) techniques can have a significant impact on finance. More specifically, NLP algorithms that have been combined with sentiment analysis have been shown to add value to the finance domain. NLP techniques can extract knowledge or valuable information in a readable format from various unstructured finance data sources including market news, expert opinions and social media. One of the main keys to make a useful decision is to be insightful and be objective about the situation. With the help of tools that contain NLP methodologies and model, investors or analysts are more likely to make useful predictions whether an increase or decrease of asset price in the future. Investors can also set price threshold and set the margin of the distance, either to average salary, debt, taxes sold. Finally, if the order of macroeconomic analysis is necessary it is presented as below: financial sentiment, short and long-term analysis and prediction that combines short and long-term results. This concludes the contributions of Sentiment Analytics in finance.

Reference:

1. Tamanampudi, Venkata Mohit. "NLP-Powered ChatOps: Automating DevOps Collaboration Using Natural Language Processing for Real-Time Incident Resolution." Journal of Artificial Intelligence Research and Applications 1.1 (2021): 530-567.
2. Sangaraju, Varun Varma, and Kathleen Hargiss. "Zero trust security and multifactor authentication in fog computing environment." Available at SSRN 4472055.
3. S. Kumari, "Kanban and Agile for AI-Powered Product Management in Cloud-Native Platforms: Improving Workflow Efficiency Through Machine Learning-

Driven Decision Support Systems", Distrib Learn Broad Appl Sci Res, vol. 5, pp. 867–885, Aug. 2019

4. Pal, Dheeraj Kumar Dukhiram, et al. "Implementing TOGAF for Large-Scale Healthcare Systems Integration." Internet of Things and Edge Computing Journal 2.1 (2022): 55-102.
5. Zhu, Yue, and Johnathan Crowell. "Systematic Review of Advancing Machine Learning Through Cross-Domain Analysis of Unlabeled Data." Journal of Science & Technology 4.1 (2023): 136-155.
6. J. Singh, "The Future of Autonomous Driving: Vision-Based Systems vs. LiDAR and the Benefits of Combining Both for Fully Autonomous Vehicles ", J. of Artificial Int. Research and App., vol. 1, no. 2, pp. 333–376, Jul. 2021
7. Gadhiraju, Asha. "Improving Hemodialysis Quality at DaVita: Leveraging Predictive Analytics and Real-Time Monitoring to Reduce Complications and Personalize Patient Care." Journal of AI in Healthcare and Medicine 1.1 (2021): 77-116.
8. Gadhiraju, Asha. "Empowering Dialysis Care: AI-Driven Decision Support Systems for Personalized Treatment Plans and Improved Patient Outcomes." Journal of Machine Learning for Healthcare Decision Support 2.1 (2022): 309-350.
9. Tamanampudi, Venkata Mohit. "Automating CI/CD Pipelines with Machine Learning Algorithms: Optimizing Build and Deployment Processes in DevOps Ecosystems." Distributed Learning and Broad Applications in Scientific Research 5 (2019): 810-849.
10. J. Singh, "Understanding Retrieval-Augmented Generation (RAG) Models in AI: A Deep Dive into the Fusion of Neural Networks and External Databases for Enhanced AI Performance", J. of Art. Int. Research, vol. 2, no. 2, pp. 258–275, Jul. 2022
11. S. Kumari, "Cloud Transformation and Cybersecurity: Using AI for Securing Data Migration and Optimizing Cloud Operations in Agile Environments", J. Sci. Tech., vol. 1, no. 1, pp. 791–808, Oct. 2020.
12. Sangaraju, Varun Varma, and Senthilkumar Rajagopal. "Applications of Computational Models in OCD." In Nutrition and Obsessive-Compulsive Disorder, pp. 26-35. CRC Press.

13. Tamanampudi, Venkata Mohit. "A Data-Driven Approach to Incident Management: Enhancing DevOps Operations with Machine Learning-Based Root Cause Analysis." Distributed Learning and Broad Applications in Scientific Research 6 (2020): 419-466.