A Framework for Facilitating Reproducible News Sentiment Impact Analysis

Weisi Chen
School of Software Engineering, Xiamen University of Technology, China
chenweisi@xmut.edu.cn

Islam Al-Qudah
Department of Computer Science, University of Sharjah, UAE
ialqudah@sharjah.ac.ae

Fethi Rabhi
School of Computer Science and Engineering, UNSW Sydney, Australia
f.rabhi@unsw.edu.au

ABSTRACT

The proliferation of outlets for news media in recent decades has contributed to faster issuance of news data. News analysis has been one of the key activities conducted by researchers in a broad variety of research disciplines. In general, the analysis process used in these studies includes interpreting the content of the news items, and then discovering their impact in a specific area. In this paper, we delve into the field of news analysis applied to the financial domain and explore news sentiment impact analysis in the context of financial markets. Existing studies lack systematic methods to assimilate financial context and evaluate the impact of a given news dataset on relevant entities financial market performance. We introduce an improved version of the framework called News Sentiment Impact Analysis (NSIA) that encompasses models, supporting software architecture and processes for defining various financial contexts and conducting news sentiment impact analysis. The framework is then evaluated using a prototype implementation and a case study that investigates the impact of extremely negative news on the stock price of the related entities. The results demonstrate the functionality, usability and reproducibility of the framework, and its capability to bridge the gap between generating news sentiment and evaluating its impact in selected financial contexts.

CCS CONCEPTS

• **Applied computing** → Service-oriented architectures; • **Software and its engineering** → Designing software; • **Information systems** → Enterprise information systems.

KEYWORDS

News Analysis, Sentiment Analysis, Software Architecture, Information Management

ACM Reference Format:
Weisi Chen, Islam Al-Qudah, and Fethi Rabhi. 2022. A Framework for Facilitating Reproducible News Sentiment Impact Analysis. In *2022 The 5th International Conference on Software Engineering and Information Management (ICSIM) (ICSIM 2022), January 21–23, 2022, Yokohama, Japan.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3520084.3520104

ICSIM 2022, January 21–23, 2022, Yokohama, Japan

ACM ISBN 978-1-4503-9551-9/22/01…$15.00
https://doi.org/10.1145/3520084.3520104

1 INTRODUCTION

The proliferation of outlets for news media in recent decades has contributed to faster publication of news data. News analysis has become one of the key activities of researchers in a broad variety of research disciplines. In general, the analysis process used in these studies includes interpreting the content of the news items, and then discovering their impact in a specific area. Various news analysis techniques have been applied, including text mining [1, 2] to identify patterns within text and find new insights, as well as opinion mining, also known as sentiment analysis, to figure out the sentiment expressed in the text, commonly via the calculations of sentiment scores [3]. In this paper, we delve into the field of news analysis applied to the financial domain, and explore news sentiment impact analysis in the context of financial markets, as this research area lacks efficient methods to evaluate the impact of any news dataset on the related financial market. Additionally, the impact analysis processes employed are not well defined and documented in many impact analysis studies, leading to difficulty in managing and reproducing such analysis studies.

The structure of the paper is as follows. Section 2 gives some background on sentiment analysis and discusses various approaches to evaluating the impact of news on financial markets. Section 3 describes the News Sentiment Impact Analysis Framework, designed for the purpose of facilitating reproducible news sentiment impact analysis with any news dataset, and Section 4 presents a prototype of this framework with a case study to validate it. Section 5 concludes the paper with some directions for future work.

2 RELATED WORK

Research on news sentiment analysis and the impact of news sentiment on the financial domain has been two-folds. Some studies concentrate on financial markets modelling and evaluation, whereas others focus on designing and validating innovative techniques for sentiment analysis, without sufficient attention to impact analysis models.

In the first type of research, many studies have analysed existing news sources (e.g. EDGAR [4, 5]) and used a variety of sentiment analysis techniques to determine sentiment scores in the target news corpus. The majority of the sentiment result datasets were generated using a lexicon-based approach [5, 6], whereas others [7, 8] were sourced directly from third-party providers such as TRNA (Thomson Reuters News Analytics) [9]. This type of research was limited in that a narrow range of computing methods was used to measure sentiment and evaluate impact. In most of these studies, determining the news sentiment's impact on financial markets was carried out using traditional statistical analysis like regression testing, sometimes involving trading strategy simulation. Often, daily stock returns were used in the analysis, together with other measures like price volatility, trading volumes, market capitalisation and earnings per share. In one particular study [5] as an example, more than twenty regression analysis tests were conducted. In addition to sentiment index, a number of other measures were applied to investigate the impact on stock returns, trading volumes, and company earnings. In fact, it is highly complicated to analyse and interpret the reaction and behaviour of financial markets. This complexity has been highlighted in some of the literature [2, 10].

On the other hand, the second type of related work mostly involved the use of novel sentiment analysis techniques. Often, a prototype implementation of the corresponding technique that consolidates the required steps and automates the sentiment analysis process for particular news datasets was presented. Text mining, machine learning or natural language processing approaches were employed to process the news content and to generate sentiment benchmarks [11-16]. Without the financial domain expertise seen in the first type of studies, the investigation of sentiment on the behaviour of financial markets, e.g. regression analysis and trading strategies, are not treated with sufficient depth, if done at all. For instance, [15, 16] focused on comparing various sentiment analysis algorithms without further impact analysis, and in [14] only limited statistical tests were employed for the prediction of stock returns for related companies.

To summarise, prior studies have some elements in common in terms of defining an impact analysis process but vary in many other aspects. For the employed techniques, the first type of studies primarily adopted lexicon-based approaches in the context of finance, whereas the second type of studies have concentrated on applying machine learning and natural language processing algorithms. In terms of sentiment impact analysis on financial markets, the first type conducted comprehensive regression testing whereas in the second type, no or limited impact analysis activities have been performed. Among all existing studies, there is a lack of a systematic method to characterise the financial context and evaluate the impact of any given news dataset on the relevant entities. For example, if the context is the US equities markets, a particular news impact needs to be evaluated in relation to this context (the same news items may have a different impact in the Chinese equities markets). This constitutes a barrier for researchers attempting to reproduce the results of a particular analysis process or to reuse it with varying financial contexts or news datasets. The research gap addressed in this paper is how to define systematic and reproducible news sentiment impact analysis processes that can be reused by diverse types of users with various levels of domain expertise, which also support performing the analysis automatically with various news corpus in different financial contexts.

3 PROPOSED FRAMEWORK – NSIA

In order to bridge the research gap identified in the previous section, a framework called News Sentiment Impact Analysis (NSIA) is proposed. Earlier versions of this framework are described in [17–19]. The NSIA framework comprises three components; NSIA Comparison Parameters Model (CPM), NSIA Architecture and NSIA Processes.

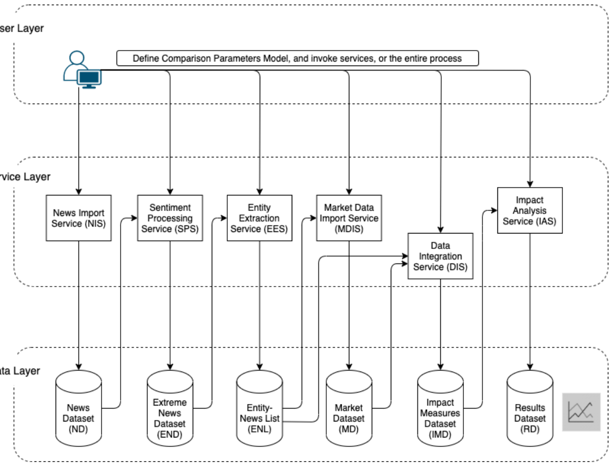

Figure 1: NSIA Architecture Overview.

3.1 NSIA Comparison Parameters Model (CPM)

The NSIA CPM defines the contexts of a defined analysis. It consists of three sets of parameters: context parameters of the domain (C), news sentiment parameters (NS) and impact measure parameters (IM).

Firstly, the C parameters define the financial contexts of a particular study. They are defined by the user or determined programmatically based on the content of the news in question. It is a vector with seven elements, C = (N, E, Ev, B, Bv, P, EP), described in Table 1

The NS parameters describe the sentiment analysis methods and algorithms used in a particular study. It is a vector with three elements, i.e., NS = (LX, FF, ESE), as shown in Table 2

Lastly, given the set of C parameters, the IM parameters specify the measures used to evaluate the news sentiment impact, as shown in Table 3

3.2 NSIA Architecture

The NSIA Architecture provides guidelines for the software implementation of the framework for developers. The design complies with the ADAGE Framework guidelines [20], aiming to facilitate reproducible news sentiment impact analysis with any given news dataset, as defined by the CPM parameters. As a Service-Oriented Architecture (SOA), it has three layers; the User Layer, the Service Layer and the Data Layer, as shown in Figure 1. Specifically, the User Layer features the user interface that facilitates user access and interaction with the Service Layer, e.g. end user definitions of the CPM parameters, and the invocation of services in the Service Layer via the user interface. The Service Layer encompasses a collection of services, which encapsulate the business logic of the impact analysis process. The Data Layer stores all the datasets generated by the services in the Service Layer. The detailed description of these services is as follows:

- News Import Service (NIS): This service enables the importation of the news data of interest from a broad range of

Table 1: Definitions of Context parameters (C)

Parameter Name	Definition	Example
News N	The news in question	Australian Financial Review news, BBC news, WSJ news
Entity E	The entity that is impacted by the news	a listed company, an industrial sector, the economy of a state
Entity Variable Ev	The variable of the selected entity E which is affected by the news in value	share prices (open, close, high, low), earnings per share, market capitalisation
Benchmark B	The point of reference against which the Ev values are compared to measure impact	a group of companies, an industrial sector, the economy of a state
Benchmark Variable Bv	The variable of the selected benchmark B	share prices (open, close, high, low) a market index, GDP per capita
Study Period P	The period during which the impact analysis is conducted	a set number of days, months, years before and/or after the news date
Estimation Period EP	The period against which the study period P is compared to measure impact	a set number of days, months, years before and/or after the news date

Table 2: Definitions of News Sentiment parameters (NS)

Parameter Name	Definition	Example
Lexicon (LX)	The dictionary used to calculate sentiment scores	Loughran and McDonald, Jockers, Nrcnet, Huliu, Sentiwordnet
Filtration Functions (FF)	The functions to filter news items based on the fields of the given news dataset	(sentiment score > 0), (news type = "Corporation")
Extreme Sentiment Extraction algorithms (ESE)	The algorithms that selects "extreme" news items out of the given news dataset	Identifying the top 10% negative news records based on the sentiment score

Table 3: Definitions of Impact Measure parameters (IM)

Parameter Name	Definition	Example
Impact Measure IM	The measures used to evaluate the news sentiment impact on the selected entity (E), compared with the selected benchmark (B), as defined in C parameters	mean cumulative abnormal returns (MCAR), trading volume, market depth

data sources (web sourced or locally stored) in various formats (CSV files, relational databases, etc.) based on the user-defined CPM parameters. The imported news data is then fed into the Data Layer as the News Dataset (ND).

- Sentiment Processing Service (SPS): This service uses the News Dataset (ND) in the Data Layer as input, calculates sentiment scores where necessary (except when the sentiment scores are already provided within the news dataset), identifies extreme news according to the user-defined CPM parameters, and commits it to the Extreme News Dataset (END) component of the Data Layer.
- Entity Extraction Service (EES): This service identifies related entities (e.g. companies) based on the user selections of entities or the content of selected news in the Extreme News Dataset (END), and generates the Entity-News List (ENL) component of the Data Layer, which contains the matched pairs of selected news and related entities. In the meantime, this service updates the user-defined C parameters in the CPM accordingly (i.e. the E, Ev, P and EP parameters) where necessary.
- Market Data Import Service (MDIS): According to the defined CPM parameters and the ENL from the Data Layer, this service imports corresponding market data from a broad range of data sources (web sourced or locally stored), in various formats (CSV files, relational databases, etc.), and saves it to the Data Layer, i.e. as the Market Dataset (MD).
- Data Integration Service (DIS): This service merges the Entity-News List (ENL) and the Market Dataset (MD) into the Impact Measures Dataset (IMD) in the Data Layer.
- Impact Analysis Service (IAS): This service executes the impact analysis based on the models defined in CPM, loading Impact Measures Dataset (IMD) and generating results, which are then committed to Results Dataset (RD) in the Data Layer. Data visualisation of the results is also produced as a chart together with the RD.

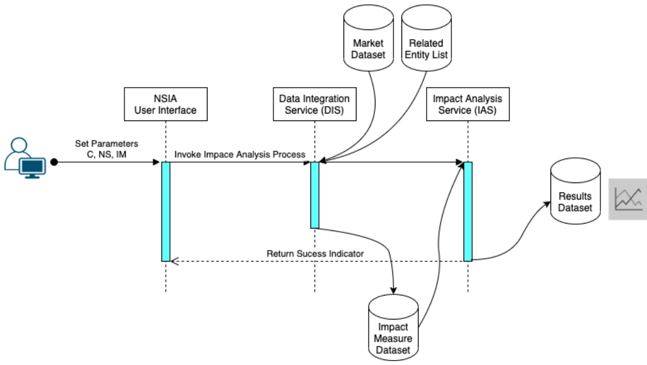

Figure 2: NSIA impact analysis sequence diagram.

The six components of the Data Layer retain all raw data and processed data throughout the impact analysis processes.

3.3 NSIA Processes

The NSIA Processes are the guidelines for users to conduct evaluation studies using any implementation of the framework. They are composed of the following set of steps:

- Importing targeted news data: This process loads the user-defined CPM parameters and invokes the News Import Service (NIS), whose role is to import the selected news data as defined in the C parameters.
- Identifying extreme news data: This process invokes the Sentiment Processing Service (SPS) either to obtain sentiment analysis results of the news from a sentiment data source or to calculate sentiment scores based on the content of the news, and then selects extreme items based on user-defined News Sentiment parameters (NS).
- Extracting related entities: This process reads the selected extreme news from the previous step and then invokes the Entity Extraction Service (EES) to generate a list of news and related entities, updating the C parameters where necessary.
- Importing relevant market data: This process loads the user-defined or updated Context parameters (C), and the Entity-News List (ENL), and runs the Market Data Import Service (MDIS) to import the relevant market data according to the C parameters.
- Performing impact analysis: Based on CPM parameters, this process runs the Data Integration Service (DIS) to merge all related datasets (ENL and MD) to prepare for the impact analysis, and then invokes the Impact Analysis Service (IAS) to conduct the impact analysis and generate final results and visualisation, as demonstrated in Figure 2

Note that one of the advantages of being a Service Oriented Architecture (SOA) design is that each of these processes can be either individually invoked, or pipelined and automated using workflow technologies (e.g. Apache Taverna [21]), which enhances the reproducibility of the impact analysis process.

4 CASE STUDY

A real-life impact analysis problem in the financial domain was used as a case study so as to evaluate the functionality, usability and reproducibility of the NSIA framework. A group of financial researchers guided the work on the case study. In this case study, we had access to news extracted from a mainstream finance-focused newspaper in Australia, the Australian Financial Review (AFR) [22], between July 2019 and July 2020 in XML format. The corpus had a total size of 131.2 MB and contained 25,891 news items in total. The financial researchers aimed to find out how extremely negative items from the news source affected the stock prices of the related Australian companies.

4.1 Prototype Implementation of NSIA

A prototype of the NSIA framework including all the components described earlier in Section 3.2 was implemented, so we can evaluate its application to the case study. The implementation has been carried out as follows:

- News Import Service (NIS): This service is implemented using the R programming language. It imports selected news data via published APIs of various data sources into News Dataset (ND). The data source used in this case study is Australian Financial Review.
- Market Data Import Service (MDIS): This service is a pre-existing element of the ADAGE framework [20]. Thanks to the portability inherent in the Service-Oriented Architecture, this service has been reused in this NSIA implementation. This service imports market data via the Yahoo Finance API into the Market Dataset (MD), based on the Context Parameters (C) defined in the Section 4.2.
- Sentiment Processing Service (SPS), Entity Extraction Service (EES), Data Integration Service (DIS) and Impact Analysis Service (IAS): These services are all built using the R programming language.

All these services were made accessible as web services via RESTful APIs, which facilitates automation of the entire process using workflow technologies. The Data Layer generated datasets stored using RDS format, which can then be exported as CSV format for further external use when requested by the user as part of the Impact Analysis Service (IAS).

To evaluate the functionality of the NSIA framework, we designed the experiment to allow the financial researcher to go through the entire impact analysis process by setting CPM parameters, invoking the services, and inspecting results. The evaluation of usability and reproducibility of NSIA involves another financial researcher reviewing the CPM defined by the former user, invoking the process, and comparing results.

4.2 CPM Settings

This section describes how the CPM parameters were set. First of all, Table 4 shows the Context parameters (C). In this case study, some of the financial context parameters are determined programmatically based on the content of the news in question rather than user-defined. The study period and the estimation period are predefined by the user.

Table 4: Context parameters (C) defined in the case study.

Parameter Name	Setting
Entity E Entity Variable Ev Benchmark B Benchmark Variable Bv	Determined by the Entity Extraction Service (EES), updated according to the Entity News List (ENL)
Study Period P	[-30 days, +30 days]: within 30 days before and 30 days after the news date
Estimation Period EP	[-130 days, -31 days]: between 31 to 130 days before the news date

Table 5: News Sentiment parameters (NS) defined in the case study.

Parameter Name	Setting
Lexicon (LX)	Loughran & McDonald
Filtration Functions (FF)	News Source = "AFR", and News Issuance Date = [2019/07/01, 2020/07/31], and News Type = "Corporation"
Extreme Sentiment Extraction (ESE) Algorithms	*average sentiment score < pos * (Pr + Ps * standard deviation), where: Pr = "mean sentiment score in the lexicon" Ps = 1, pos = -1*

Table 6: Details of the Extreme Sentiment Extraction (ESE) Algorithm in the case study.

ESE (Pos, Pr, Ps, ND)		Description
Pos	-1	To extract extremely positive news or not; extremely positive news: Pos = 1; extremely negative news: Pos = -1.
Pr	mean sentiment score in the lexicon	Threshold parameter to fetch news items above the threshold.
Ps	1	Threshold parameter multiplied by the standard deviation to define how far we want the algorithm to deviate from the mean.

Table 7: Impact Measure parameters (IM) defined in the case study.

Parameter Name	Setting
Impact Measure IM	Mean Cumulative Abnormal Return (MCAR)

The News Sentiment parameters (NS) are defined in Table 5. Specifically, ESE requires some user-defined parameters (see Table 6), to determine extreme news records (extremely negative news in this case) in the News Dataset. The impact measure (IM) used in this case study is mean cumulative abnormal returns (MCAR), as shown in Table 7

4.3 Experiments and Results

The News Import Service (NIS) in the prototype of the NSIA framework extracted all news records within the defined news period from the Australian Financial Review file archive. 55 extremely negative news records were identified by the Sentiment Processing Service (SPS) based on the user-defined NS parameters.

The Entity Extraction Service (EES) generated the Entity-News List (ENL), detailing the 55 pairs of related entities, i.e., Australian companies, and corresponding news dates (business days only), shown in Table 8

In the meantime, undefined Context parameters (C) are updated with actual values (see Table 9). In this case, it has 55 sets of Context parameters, each corresponding to one Entity-News pair as listed in Table 8
The Market Data Import Service (MDIS) then downloads the market data corresponding to the updated C parameters. Finally, the Data Integration Service (DIS) combines all datasets required for the impact analysis, ENL and MD, and the Impact Analysis Service (IAS) analysed the impact of selected news records. Figure 3 displays a visual summary of the experiment results.

As seen from chart, the Mean Cumulative Abnormal Return (MCAR) dropped sharply around Day 10, i.e. 10 days after the news date (Day 0). This indicates a delayed significant negative impact of the extremely negative news on the relevant Australian companies.

Table 8: Related Entity List (REL).

Code	News Date	Code	News Date	Code	News Date	Code	News Date	Code	News Date
ANZ.AX	2019/7/15	SUN.AX	2019/10/8	WBC.AX	2019/11/21	WBC.AX	2020/2/24	TWE.AX	2020/5/1
ANZ.AX	2019/7/18	QBE.AX	2019/10/8	BHP.AX	2019/11/27	ANZ.AX	2020/2/24	WES.AX	2020/5/8
WBC.AX	2019/7/18	BSL.AX	2019/10/9	WBC.AX	2019/12/19	NAB.AX	2020/2/24	SUN.AX	2020/5/12
WOW.AX	2019/8/6	ANZ.AX	2019/10/17	PTM.AX	2019/12/23	TER.AX	2020/3/2	ALQ.AX	2020/5/14
AMP.AX	2019/8/28	CBA.AX	2019/10/18	TWE.AX	2020/2/5	DCG.AX	2020/3/13	BLD.AX	2020/5/28
FLT.AX	2019/9/2	WBC.AX	2019/10/24	CBA.AX	2020/2/7	COH.AX	2020/3/18	WBC.AX	2020/6/5
IFM.AX	2019/9/9	CWN.AX	2019/10/25	WBC.AX	2020/2/7	CBA.AX	2020/3/18	WBC.AX	2020/6/9
CIM.AX	2019/9/16	CGC.AX	2019/10/29	ANZ.AX	2020/2/7	AMP.AX	2020/3/19	IAG.AX	2020/6/29
ANZ.AX	2019/9/23	MQG.AX	2019/11/12	NAB.AX	2020/2/7	RIO.AX	2020/4/2	FNP.AX	2020/7/2
PSQ.AX	2019/9/24	RIO.AX	2019/11/13	CWN.AX	2020/2/10	TWE.AX	2020/4/6	DCG.AX	2020/7/8
IAG.AX	2019/10/8	GNC.AX	2019/11/15	CBA.AX	2020/2/24	BHP.AX	2020/4/21	WBC.AX	2020/7/31

Table 9: Updated Context parameters C according to Entity-News List.

Context *C*	Entity *E*(Ev = Closing Price)	Benchmark *B*(Bv = Closing Price)	Benchmark Description	Study Period *P*(business days)	Estimation Period *EP*(business days)
C1	{ANZ} ∩ {AXJO}*	AXJO	AXJO = Australia's S&P/ASX 200	[2019/05/31, 2019/08/26]	[2019/01/07, 2019/05/30]
C2	{ANZ} ∩ {AXJO}	AXJO	AXJO = Australia's S&P/ASX 200	[2019/06/05, 2019/08/29]	[2019/01/10, 2019/06/04]
C3	{WBC} ∩ {AXJO}	AXJO	AXJO = Australia's S&P/ASX 200	[2019/06/05, 2019/08/29]	[2019/01/10, 2019/06/04]
C4	...	...	...	...	...

*{AXJO} represents all the constituents of market index AXJO, and Ev is set to the "closing price" field for every constituent in entity E.

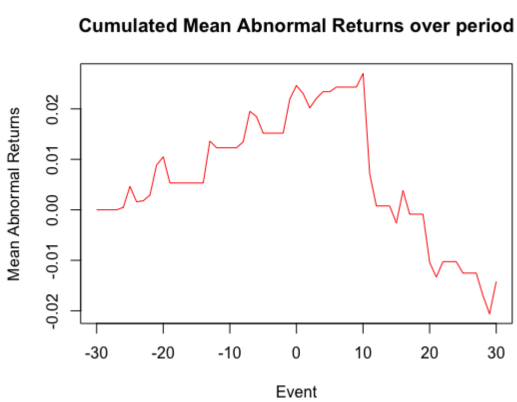

Figure 3: Data visualisation of the result, showing the MCARs of all 55 company-news pairs over the study period.

4.4 Discussion

This experiment was initiated by a financial researcher with financial expertise and a good knowledge of impact analysis studies. The raw news dataset contains 25,891 news items in total, among which 55 were found to be extremely negative based on selected News Sentiment parameters (NS). Each of the NSIA processes was executed via the user interface manually, and approximately 20 minutes was required for an impact analysis process at this scale to be completed. The prototype preserves all raw data, processed data and all the CPM parameters used in this study, which can then be reused by any other service or other users. To evaluate the reproducibility of the prototype, another financial researcher ran the same study with the selected CPM parameters successfully and got the same results in less than 5 minutes. The entire impact analysis process, from importing news data, processing news sentiment, extracting relevant entities, downloading market data, to data integration and executing impact analysis, can all be automated as mentioned earlier, taking advantage of workflow technologies like Apache Taverna. The overall results show a delayed response from the market, 10 days in this case, which is unusual. Financial expertise will come into play when interpreting the abnormality.

In summary, the case study was able to demonstrate the functionality, usability and reproducibility aspects of the NSIA framework.

5 CONCLUSION AND FUTURE WORK

This paper introduced in detail the News Sentiment Impact Analysis (NSIA) Framework to facilitate and manage reproducible sentiment impact analysis studies with any given news dataset. To evaluate the functionality, usability and reproducibility of the NSIA framework, a prototype was implemented. The performance of the prototype was then evaluated using a case study with the involvement of financial researchers. The result of the experiment was interesting from the perspective of the domain experts, and it could be easily reproduced or repeated with different parameters by another financial researcher.

The experiments and results demonstrated the advantages of the NSIA framework, i.e., functionality, usability and reproducibility. There is also a straightforward path to full automation. Therefore, our next goal is to ensure the robustness of the implemented system and to introduce workflow technologies so it will be easier and more time efficient for users to conduct a vast number of studies that correspond with various sets of CPM parameters with little human intervention.

An unresolved application domain issue is that it is unclear from the result which news record had the greatest impact, especially when there were multiple news records affecting the same company. This may lead to seemingly abnormal results, as different weightings of news items were not considered. Other future work will extend the NSIA framework to better represent all types of financial contexts, and to allow other impact measures to be used, such as intraday market measures like; market depth, trading volume and mean intraday returns across different financial contexts. The framework can be further evaluated with news datasets from various sources, and different sentiment scoring algorithms to allow comparison between different contexts and parameters.

ACKNOWLEDGMENTS

We wish to thank Nine Publishing for supplying us with the AFR News dataset used in this paper. This paper is funded by Xiamen Scientific Research Funding for Overseas Chinese Scholars, and the Research Fund by Xiamen University of Technology (YKJ21012R).

REFERENCES

[1] Bettina Berend. 2011. Text Mining for News and Blogs Analysis. In: Sammut C., Webb G.I. (eds) Encyclopedia of Machine Learning. Springer, Boston, MA. https://doi.org/10.1007/978-0-387-30164-8_827

[2] Arman Khadjeh Nassirtoussi, Saeed Aghabozorgi, Teh Ying Wah and David Chek Ling Ngo. 2014. Text mining for market prediction: A systematic review. Expert Systems with Applications, 41(16), 7653-7670.

[3] Bing Liu. 2012. Sentiment analysis and opinion mining. Synthesis Lectures on Human Language Technologies, 5(1), 1-167.

[4] Paul C. Tetlock. 2007. Giving content to investor sentiment: The role of media in the stock market. The Journal of Finance, 62(3), 1139-1168.

[5] Paul C. Tetlock, Maytal Saar-Tsechansky and Sofus Macskassy. 2008. More than words: Quantifying language to measure firms' fundamentals. The Journal of Finance, 63(3), 1437-1467.

[6] Tim Loughran and Bill McDonald. 2011. When Is a Liability Not a Liability? Textual Analysis, Dictionaries, and 10-Ks. The Journal of Finance, 66(1), 35-65.

[7] Michal Dzielinski. 2011. News Sensitivity and the Cross - Section of Stock Returns. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.1889030

[8] David E. Allen, Michael McAleer and Abhay K. Singh. 2019. Daily market news sentiment and stock prices, Applied Economics, 51:30, 3212-3235, https://doi.org/10.1080/00036846.2018.1564115

[9] Thomson Reuters. Thomson Reuters News Analytics (TRNA). Accessed Sep 2021, from https://www.thomsonreuters.com/content/dam/openweb/documents/pdf/tr-com-financial/report/thomson-reuters-news-analytics-2015.pdf

[10] Malcolm Baker and Jeffery Wurgler. 2006. Investor Sentiment and the Cross-Section of Stock Returns. The Journal of Finance, 61: 1645-1680. https://doi.org/10.1111/j.1540-6261.2006.00885.x

[11] Artur Lugmayr. 2013. Predicting the future of investor sentiment with social media in stock exchange investments: A basic framework for the DAX performance index. Handbook of social media management (pp. 565-589) Springer.

[12] Michael Siering. 2013. Investigating the Impact of Media Sentiment and Investor Attention on Financial Markets. In Enterprise Applications and Services in the Finance Industry. FinanceCom 2012. Lecture Notes in Business Information Processing, vol 135. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-36219-4_1

[13] Johan Bollen and Huina Mao. 2011. Twitter Mood as a Stock Market Predictor. Computer, 44(10), 91-94. https://doi.ieeecomputersociety.org/10.1109/MC.2011.323

[14] Robert P. Schumaker, Yulei Zhang, Chun-Neng Huang and Hsinchun Chen. 2012. Evaluating Sentiment in Financial News Articles. Decision Support Systems, 53(3), 458-464. https://doi.org/10.1016/j.dss.2012.03.001

[15] Soonh Taj, Baby Bakhtawer Shaikh and Areej Fatemah Meghji. 2019. Sentiment Analysis of News Articles: A Lexicon based Approach. In 2019 2nd International Conference on Computing, Mathematics and Engineering Technologies (iCoMET), 1-5.

[16] Antony Samuels and John Mcgonical. 2020. News Sentiment Analysis. *ArXiv, abs/2007.02238*.

[17] Islam Al Qudah, Fethi Rabhi and Maurice Peat. 2014. A Proposed Framework for Evaluating the Effectiveness of Financial News Sentiment Scoring Datasets. In FinanceCom2014, 29-47. https://doi.org/10.1007/978-3-319-28151-3_3

[18] Islam Al Qudah and Fethi Rabhi. 2017. News Sentiment Impact Analysis (NSIA) Framework. In: Feuerriegel S., Neumann D. (eds) Enterprise Applications, Markets and Services in the Finance Industry. FinanceCom 2016. Lecture Notes in Business Information Processing, vol 276. Springer, Cham. https://doi.org/10.1007/978-3-319-52764-2_1

[19] Islam Al Qudah and Fethi Rabhi. 2019. Systematic Approach to Quantify Impact of News Sentiment on Financial Markets. In 2019 International Conference on Computational Intelligence and Knowledge Economy (ICCIKE), pp. 60-65, https://doi.org/10.1109/ICCIKE47802.2019.9004384

[20] Fethi Rabhi, Lawrence Yao and Adnene Guabtni. 2012. ADAGE: A Framework for Supporting User-Driven Ad-hoc Data Analysis Processes, *Computing*, 2012. Volume 94, Issue 6, Page 489-519, https://doi.org/10.1007/s00607-012-0193-0

[21] Tom Oinn, Matthew Addis, Justin Ferris, Darren Marvin, Martin Senger, Mark Greenwood, Tim Carver, Kevin Glover, Matthew R. Pocock, Anil Wipat and Peter Li. 2004. Taverna: a Tool for the Composition and Enactment of Bioinformatics Workflows. Bioinformatics, 20(17), 3045–3054, https://doi.org/10.1093/bioinformatics/bth361

[22] Australian Financial Review. 2021. https://www.afr.com/