INF-UFG at FiQA 2018 Task 1: Predicting Sentiments and Aspects on Financial Tweets and News Headlines

Dayan de França Costa
Federal University of Goiás
Goiânia, GO, Brazil
dayanfcosta@gmail.com

Nadia Felix Felipe da Silva
Federal University of Goiás
Goiânia, GO, Brazil
nadia@inf.ufg.br

ABSTRACT

This paper describes our system which participate in Task 1 of FiQA 2018. The task's focuses was to predict sentiment and aspects of financial microblog posts and headlines. The sentiment analysis for a specific company had to be predicted using a scale between -1 and 1, while the aspect prediction had to be predicted using a set of aspects which was given in train data. We had used Support Vector Regression (SVR) to predict the sentiments in both cases (microblog posts and headlines).

KEYWORDS

Predictive sentiment analysis; Stock market; Sentiment analysis in the financial domain

ACM Reference Format:

Dayan de França Costa and Nadia Felix Felipe da Silva. 2018. INF-UFG at FiQA 2018 Task 1: Predicting Sentiments and Aspects on Financial Tweets and News Headlines. In *The 2018 Web Conference Companion, April 23–27, 2018, Lyon, France*. ACM, New York, NY, USA, 5 pages.

1 INTRODUCTION

The use and communication through microblogs as Twitter¹ and StockTwits² have been increasing in the last years, making the market to pay more attention in what is said about them in those microblogs [1]. This increase is given through the facility to use these platforms, the message formats and its accessibility[20]. People are using these communication vehicles to express sentiments about life, business, work, sports and other ones, therefore the brands has become more worried about what people say about them in these platforms to measure how people fell about the brand, attract more customers, etc. In financial domain, these texts full of opinions are able to change the financial sector, raising or falling stock values.

With the increasing use of these platforms, applications of technologies have been growing up too in last decade and it has been caused by the applications coming from academia to commercial domain and because sentiment the analysis tasks are full of challenges. In last years, sentiment analysis become an interest of financial

¹<https://twitter.com>

²<https://stocktwits.com>

This paper is published under the Creative Commons Attribution 4.0 International (CC BY 4.0) license. Authors reserve their rights to disseminate the work on their personal and corporate Web sites with the appropriate attribution.

WWW '18 Companion, April 23–27, 2018, Lyon, France

© 2018 IW3C2 (International World Wide Web Conference Committee), published under Creative Commons CC BY 4.0 License.

ACM ISBN 978-1-4503-5640-4/18/04.

<https://doi.org/10.1145/3184558.3191828>

research area [2, 4, 31, 32]. Since researchers have been showing that opinions expressed in microblogs and social media discussions can make a great impact on market[7].

This paper describes our systems submitted to Open Challenge – Financial Opinion Mining in FiQA (2018) - Task 1 which is related to aspect-based financial sentiment analysis³, which had 4 teams participating. This task aims to identify positive (bearish) and negative (bullish) sentiments and the aspects associated to companies stocks based on a dataset of microblogging posts and news statements and headlines. The challenges were that sentiment should be scaled between -1 (very negative/bullish) and 1 (very positive/bearish) instead of using the conventional positive, negative and neutral labels, as a classifying challenge, in addition the aspects should be classified using a big set of aspects given in the dataset, using two levels of classification.

In this work, we extract a series of elaborately designed features like word embeddings, n-grams, word replacements and multiple different regression and classification algorithms such as SVM in linear kernels, Bayesian models, ensemble and tree models. We also have tried an ensemble containing some of these models quoted.

The rest of this paper is structured as follows. Section 2 describes related work. Section 3 reports datasets, experiments and results discussion. Finally, Section 4 concludes our work.

2 RELATED WORK

Many research initiatives have focused on whether the sentiment analysis of social media can be used to predict the future of stock market indicators. In this section, we give an overview of related studies, which are focused on: (i) sentiment analysis of social media as a predictor of the future stock market indicators, and (ii) aspect-based sentiment analysis.

2.1 Financial tweet sentiment analysis

The most well-known publication in this area is by Bollen et al. [2]. They investigated whether measurements of collective moods derived from Twitter feeds are correlated to the value of the Dow Jones Industrial Average (DJIA) over time. They analyzed the tweets by two mood tracking tools, namely OpinionFinder that measures positive vs. negative mood and Google-Profile of Mood States (GPOMS) that measures mood in terms of 6 dimensions (Calm, Alert, Sure, Vital, Kind, and Happy). They cross-validated the resulting mood-time series by comparing their ability to detect the public's response to the presidential election and Thanksgiving day in 2008. The Granger causality analysis and a Fuzzy Neural Network were used to investigate the hypothesis that public mood states, as measured

³<https://sites.google.com/view/fiqa>

by the OpinionFinder and GPOMS mood time series, were predictive of changes in DJIA closing values. Their results indicated that the accuracy of DJIA predictions can be improved by the inclusion of specific public mood dimensions.

Smailovic et al. [31] also analyzed whether the sentiment expressed in tweets, which discuss selected companies and their products, can indicate their stock price changes. To address this problem, an active learning approach was developed and applied to sentiment analysis of tweet streams in the stock market domain. The paper first presents a static Twitter data analysis problem, explored in order to determine the best Twitter-specific text preprocessing setting for training the Support Vector Machine (SVM) sentiment classifier. In the static setting, the Granger causality test shows that sentiments in stock-related tweets can be used as indicators of stock price movements a few days in advance, where improved results were achieved by adapting the SVM classifier to categorize Twitter posts into three sentiment categories of positive, negative and neutral. These findings were adopted in the development of a new stream-based active learning approach to sentiment analysis, applicable in incremental learning from continuously changing financial tweet streams. To this end, a series of experiments was conducted to determine the best querying strategy for active learning of the SVM classifier adapted to sentiment analysis of financial tweet streams. The experiments in analyzing stock market sentiments of a particular company showed that changes in positive sentiment probability can be used as indicators of the changes in stock closing prices.

Pagalu et al. [19] observed how well the changes in stock prices of a company, the rises and falls, are correlated with the public opinions being expressed in tweets about that company. Their paper employed two textual representations, Word2vec and N-gram, for analyzing the public sentiments in tweets. The authors applied sentiment analysis and supervised machine learning principles to the tweets and analyzed the correlation between stock market movements of a company and sentiments in tweets. They showed that positive news and tweets in social media about a company would encourage people to invest in the stocks of that company and as a result the stock price of that company would increase. In addition, this study demonstrated a strong correlation between the rise and falls in stock prices with the public sentiments in tweets.

Given the link between sentiment and market dynamics, the analysis of public sentiment becomes a powerful method to predict the market reaction. However, the accuracy of machine learning-based sentiment analysis approaches rarely exceeds seventy percent [32].

It is still important to mention the Task 5 of Semeval⁴[4] which discusses the “Fine-Grained Sentiment Analysis on Financial Microblogs (subtask a) and News (subtask b)” is specifically under the “Detecting sentiment, humor, and truth” theme. This task contains two tracks, where the first one concerns Microblog messages and the second one covers News Statements and Headlines. The main goal behind both tracks was to predict the sentiment score for each of the mentioned companies/stocks. The sentiment scores for each text instance adopted floating point values in the range of -1 (very negative/bearish) to 1 (very positive/bullish), with 0 designating

neutral sentiment. This task attracted a total of 32 participants, with 25 participating in Track 1 and 29 in Track 2. Tables 1 and 2 summarize the top five teams that participated in task.

Table 1: Top five teams that participated in Task 5 of Semeval – predicting sentiment from financial news headlines

	Authors	Feature Set	Approach
1	Mansar et al. [15]	word embeddings (GloVe), lexicon	A deep learning architecture
2	Kar et al. [11]	word embeddings (word2vec), lexicon	Support Vector Regression (SVR) and a deep learning architecture
3	Rotim et al. [26]	String kernels, word embeddings (Glove and word2vec), named entity recognition (NER)	Support Vector Regression (SVR)
4	Moore and Rayson [18]	n-grams and word-replacements	Support Vector Regression (SVR) and a Bidirectional Long Short-Term Memory (BLSTM)
5	Jiang et al. [10]	linguistic, sentiment lexicon, domain-specific and word embedding features (word2vec)	regressor ensemble

Table 2: Top five teams that participated in Task 5 of Semeval – predicting sentiment from microblogs messages

	Authors	Feature Set	Approach
1	Jiang et al. [10]	linguistic, sentiment lexicon, domain-specific and word embedding features (word2vec)	regressor ensemble
2	Ghosal et al. [6]	Character ngrams, Word ngrams, POS-tag, Lexicons, Pointwise Mutual Information (PMI), and Microblog Specific Features	Deep Learning Ensemble
3	Deborah et al. [27]	bag-of-words- based features	Gaussian Process
4	Cabanski et al. [3]	word embedding and lexicon features	recurrent neural network
5	Kumar et al. [12]	Word Embedding (GloVe), Tf-Idf Score, Sentiment Lexicon	an ensemble of Support Vector Classifier and Logistic Regression.

2.2 Aspect-based sentiment analysis

Aspect-based sentiment analysis systems treat a set of texts (e.g., product reviews or messages from social media) discussing a particular entity (e.g., a new restaurant) to detect the main (e.g., the most frequently discussed) aspects (features) of this entity (e.g., ‘food’, ‘service’) to estimate the average sentiment of the texts per aspect (e.g., how positive or negative the opinions are on average for each aspect) [13].

One of the earliest studies on Aspect-based sentiment analysis is [9], which followed a frequency-based approach. The basic idea is that frequently mentioned nouns are more likely to be aspects.

⁴<http://alt.qcri.org/semeval2017/task5/>

To compensate for the errors resulting from ignoring infrequent names, the authors suggested exploiting opinion words to find aspects. For this part, they proposed to consider the nearest opinion word. This idea was used in papers [25, 33].

The main approaches use classifiers with expensive handcrafted features based on n-grams, parts-of-speech, and sentiment lexicon. For example, Popescu and Etzioni [24] suggested the use of part-of-relationship to remove frequent noun phrases that are not aspects. Benefiting from statistics about the use of nouns in the English language, Scaffidi et al. [30] improved the general approach of relying frequent nouns to extract aspects. Long et al. [14] added the use of information distance, and dependent words (adjectives).

The studies about aspect-based sentiment analysis has gained more propulsion in some editions of SemEval⁵⁶⁷ [21–23]. These Semeval competitions provided datasets of English reviews annotated at the sentence level with aspect terms (e.g., “mouse”, “pizza”) and their polarity for specific domains, as well as aspect categories (e.g., “food”) and their polarity.

It is interesting to present that a one of the best scores [22] was achieved based on a Liblinear model [8] and the second best score [5] was achieved with an SVM model trained on the restaurants training data. The model used features based on unigrams, sentiment lexica and PMI scores learnt from TripAdvisor data. The team of EliXa [29] (one of the challenge winners) used a multiclass SVM and features based on word clusters, lemmas, n-grams, POS tagging, and sentiment lexica.

3 EXPERIMENTAL EVALUATION

3.1 Datasets

The dataset provided consists of 675 microblog messages, and 438 news statements and headlines for training stage (see Table 3). There are 99 microblog messages, and 93 news statements and headlines for testing stage. The data was collected from web sites as Stocktwits⁸, Reddit⁹, Wikinews¹⁰, and other finance domain web pages. We used just the dataset available for participating systems.

Table 3: Statistis about training dataset

	Instances	Positive	Negative	Neutral
Microblog Posts	675	440	234	1
Headlines	438	282	144	12

The Table 3 refers to statistics about the sentiments of training dataset. The Instances columns refers to the amount of sentences available in the dataset. Positive, Negative and Neutral correspond to instances with sentiment score positive, negative or zero, respectively.

About the aspects, the datasets contains 95 aspect classes for the news statements and headlines and 83 aspect classes for the microblog posts. Into these classes, the most often classes in news statements and headlines dataset are ‘corporate’, ‘m&a’ and ‘stock’

⁵<http://alt.qcri.org/semeval2014/task4/>

⁶<http://alt.qcri.org/semeval2015/task12/>

⁷<http://alt.qcri.org/semeval2016/task5/>

⁸<https://stocktwits.com/>

⁹<https://www.reddit.com/>

¹⁰<https://www.wikinews.org/>

with 327, 106 and 101 classifications respectively. The most often classes into microblog posts dataset are ‘stock’, ‘price action’ and ‘bullish’ with 546, 379 and 203 classifications respectively. The datasets has a lot of classes that are classified just once in whole dataset, the news statements and headlines dataset there are 28 classes and in the microblog posts dataset there are 29 classes that were used just once.

The training dataset contains the followed attributes:

- **Target:** In microblog posts it refers to stocks companies without the cashtag (that is present in sentence). In headlines it refers to the main company in sentence
- **Sentence:** The news statements and headlines or the microblog posts which sentiment is expressed
- **Snippets:** The main piece of the sentence
- **Aspects:** The aspects associated to the sentences
- **Sentiment Score:** A scaled sentiment between -1 and 1, where 1 is very positive, -1 is very negative and 0 represents a neutral sentiment. It’s related to the sentences and targets.

3.2 Feature Engineering

While building the system, we tested a set of features and parameters trying get the best data fitting. The features used are described below:

(1) N-Grams

It is a sequence of N words in a text/sentence. We used unigrams, bigrams, and trigrams in the grid search to see which one would perform better.

(2) Tokenization

Was used a library called Unitok¹¹ [16] to help tokenize sentences. This library recognizes URLs, e-mail addresses, DNS domains and IP addresses, and an interesting characteristics of that is the recognition of abbreviations made in English.

(3) Word Replacements

It was tested replacing or not the company’s stock symbols presented in the sentences as well as positive and negative words. To replace negative and positive words, it was used the N most similar words based on a cosine similarity, in our case N was 10. With all the stock symbols we tested to replace them with a common word (e.g., replacing \$AAPL by ‘company’ string) and removing them from the sentences.

(4) Word Embeddings

As many tasks involving natural language processing, we decided to use word embeddings that performs better than a natural bag-of-words, which hardly capture the words’s semantic. As it was widely used by the winners of Task 5 of Semeval 2017 we decided to use this technique in our experiments through the use of word2vec¹² model. We have chosen to use word2vec because of its size (3 millions vocabulary entries) and because it is trained on Google News using the Mikolov’s [17] method.

(5) TF-IDF (Term Frequency - Inverse Document Frequency)

Using TF-IDF weighting [28] was used to scale down the impact of frequencies words in both cases, regression and classification. Was tested using both l1 (a simple sum of

¹¹<http://corpus.tools/wiki/Unitok>

¹²<https://code.google.com/archive/p/word2vec/>

the vector's components) and l2 (based on Euclidean Distance) normalizations and using or not inverse-document-frequency to reweighting.

3.3 Experimental Setup

In training stage, we split the dataset using randomly 80%-20% proportion, where the 80% part was used to train the algorithms and 20% was used to test and validate our algorithms according to competition measures. The results was evaluated using Mean Squared Error (MSE) and R Squared (R2) to measure the sentiment analysis score and precision, recall and F1-score for aspect classification task. We used, in both cases Support Vector Machine algorithms from sklearn¹³, so we applied the Support Vector Regressor in sentiment analysis' task and Support Vector Classifier in aspect's classification task.

In our experiments, we used a grid search cross validation with 10-folds for both cases. On regression case (sentiment) we tested the values 10, 1 and 0.1 for the penalty parameter C of LinearSVR algorithm, and we validated with the R Squared and Mean Squared Error, both metrics from sklearn metrics' package. We ran some sklearn's regressors to this case, including an ensemble of them and our best results were with LinearSVR¹⁴ only.

On classification case (aspects) we first binarized the aspects based on a set of all possible aspects in training set. To extract this set, we simple split the aspects given in "/" and cleared the values to remove punctuation and white spaces. Then we ran the OneVsRest¹⁵ meta Classifier with LinearSVC algorithm to classify the multi labeled aspects and the scorer function we used the mean of precision, recall and F1-score algorithms provided by sklearn, where greater were better. As it was done in regression, we tested the same parameters for the penalty parameter C of LinearSVC.

3.4 Results and Discussion

In this section we present the results for sentiment analysis and aspect classification tasks of FiQA (2018). The metrics used were Mean Squared Error(MSE) and R Squared for the sentiment analysis task, and Precision, Recall, and F1-score for the aspects classification. The formulas is shown below:

$$MSE = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2.$$

$$RS_{tot} = \sum_{i=1}^n (y_i - \bar{y})^2$$

$$Precision = \frac{tp}{tp + fp}$$

$$Recall = \frac{tp}{tp + fn}$$

$$F_1 = 2 \cdot \frac{1}{\frac{1}{recall} + \frac{1}{precision}} = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

In our tests, we obtained the following results – see Table 4.

¹³<http://scikit-learn.org/stable/index.html>

¹⁴<http://scikit-learn.org/stable/modules/generated/sklearn.svm.LinearSVR.html>

¹⁵<http://scikit-learn.org/stable/modules/generated/sklearn.multiclass.OneVsRestClassifier.html>

Table 4: Results of Aspects Classification in “unofficial test set” – The training dataset was split using randomly 80%-20% proportion, where the 80% part was used to train the algorithms and 20% was used to test and validate our algorithms according to competition measures.

	Precision	Recall	F1-Score
Microblog Posts	0.6673	0.5592	0.5775
Headlines and Statements	0.4992	0.4	0.4240

In the test dataset the all predicted values were 'stock', 'price action', 'coverage', 'bullish', 'corporate', 'analyst ratings' and 'bearish' for the posts, with the most often classified 'stock', 'price action' and 'coverage'. For the headline and news statements the predicted values were 'corporate', 'stock', 'm&a', 'appointment', 'bullish', 'price action' with the most often classified 'corporate', 'stock', and 'm&a'.

According to official results¹⁶, we scored the following results about sentiment analysis challenge – see Table 5.

Table 5: Official Results - Sentiment Analysis

	MSE	R Squared	Cosine
Microblog Posts	0.206794	0.1665593177	0.415379
Headlines and Statements	0.0958436	0.1642305099	0.533388

We scored the best MSE of the competition with the dataset of headlines and news statements, according to the official results¹⁷. As we provided the aspects in a wrong format, the results about them was not evaluated, becoming zero.

4 CONCLUSIONS

In this paper we presented our solution for FiQA 2018 Task 1 using Linear Kernel (SVR and SVC) models from sklearn. We have tested some regressors, classifiers and ensembles of them, but our best scores were just using only linear models for each part of the task. The enhancing of the results can be made by using different approaches of learning like simple neural networks or LSTMs (Long Short Term Memory). In the future we can improve our solution with some of these approaches.

REFERENCES

- [1] Sitaram Asur and Bernardo A. Huberman. 2010. Predicting the Future with Social Media. In *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology - Volume 01 (WI-IAT '10)*. IEEE Computer Society, Washington, DC, USA, 492–499. <https://doi.org/10.1109/WI-IAT.2010.63>
- [2] Johan Bollen, Huina Mao, and Xiao-Jun Zeng. 2010. Twitter mood predicts the stock market. *CoRR abs/1010.3003* (2010). arXiv:1010.3003 <http://arxiv.org/abs/1010.3003>
- [3] Tobias Cabanski, Julia Romberg, and Stefan Conrad. 2017. HHU at SemEval-2017 Task 5: Fine-Grained Sentiment Analysis on Financial Data using Machine Learning Methods. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. Association for Computational Linguistics, Vancouver, Canada, 832–836. <http://www.aclweb.org/anthology/S17-2141>

¹⁶Until the deadline for the submission of this paper, the organizers did not divulge the gold classes and values for the classification task of aspects and for the sentiment analysis.

¹⁷The official results were not publicly released until the deadline for the submission of this paper

- [4] Keith Cortis, André Freitas, Tobias Daudert, Manuela Huerlimann, Manel Zarrouk, Siegfried Handschuh, and Brian Davis. 2017. SemEval-2017 Task 5: Fine-Grained Sentiment Analysis on Financial Microblogs and News. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. Association for Computational Linguistics, 519–535. <https://doi.org/10.18653/v1/S17-2089>
- [5] Orphee De Clercq, Marjan Van de Kauter, Els Lefever, and Veronique Hoste. 2015. LT3: Applying Hybrid Terminology Extraction to Aspect-Based Sentiment Analysis. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. Association for Computational Linguistics, Denver, Colorado, 719–724. <http://www.aclweb.org/anthology/S15-2122>
- [6] Deepanway Ghosal, Shobhit Bhatnagar, Md Shad Akhtar, Asif Ekbal, and Pushpak Bhattacharyya. 2017. IITP at SemEval-2017 Task 5: An Ensemble of Deep Learning and Feature Based Models for Financial Sentiment Analysis. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. Association for Computational Linguistics, 899–903. <https://doi.org/10.18653/v1/S17-2154>
- [7] Rohitha Goonatilake, Ajantha Herath, Suvineetha Herath, Susantha Herath, and Jayantha Herath. 2007. Intrusion detection using the chi-square goodness-of-fit test for information assurance, network, forensics and software security. *Journal of Computing Sciences in Colleges* 23, 1 (2007), 255–263.
- [8] Hussam Hamdan, Patrice Bellot, and Frederic Bechet. 2015. Isislif: Feature extraction and label weighting for sentiment analysis in twitter. In *In Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval)*.
- [9] Minqing Hu and Bing Liu. 2004. Mining and Summarizing Customer Reviews. In *Proceedings of the Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '04)*. ACM, New York, NY, USA, 168–177. <https://doi.org/10.1145/1014052.1014073>
- [10] Mengxiao Jiang, Man Lan, and Yuanbin Wu. 2017. ECNU at SemEval-2017 Task 5: An Ensemble of Regression Algorithms with Effective Features for Fine-Grained Sentiment Analysis in Financial Domain. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. Association for Computational Linguistics, 888–893. <https://doi.org/10.18653/v1/S17-2152>
- [11] Sudipta Kar, Suraj Maharjan, and Thamar Solorio. 2017. RiTUAL-UH at SemEval-2017 Task 5: Sentiment Analysis on Financial Data Using Neural Networks. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. Association for Computational Linguistics, 877–882. <https://doi.org/10.18653/v1/S17-2150>
- [12] Abhishek Kumar, Abhishek Sethi, Md Shad Akhtar, Asif Ekbal, Chris Biemann, and Pushpak Bhattacharyya. 2017. IITPB at SemEval-2017 Task 5: Sentiment Prediction in Financial Text. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. Association for Computational Linguistics, 894–898. <https://doi.org/10.18653/v1/S17-2153>
- [13] Bing Liu. 2012. *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- [14] Chong Long, Jie Zhang, and Xiaoyan Zhut. 2010. A Review Selection Approach for Accurate Feature Rating Estimation. In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters (COLING '10)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 766–774. <http://dl.acm.org/citation.cfm?id=1944566.1944654>
- [15] Youness Mansar, Lorenzo Gatti, Sira Ferradans, Marco Guerini, and Jacopo Staiano. 2017. Fortia-FBK at SemEval-2017 Task 5: Bullish or Bearish? Inferring Sentiment towards Brands from Financial News Headlines. *CoRR abs/1704.00939* (2017). <http://arxiv.org/abs/1704.00939>
- [16] Jan Michelfeit, Jan Pomikálek, and Vit Suchomel. 2014. Text Tokenisation Using unitok. In *RASLAN 2014*, Aleš Horák and Pavel Rychlý (Eds.). Tribun EU, Brno, Czech Republic, 71–75.
- [17] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.
- [18] Andrew Moore and Paul Rayson. 2017. Lancaster A at SemEval-2017 Task 5: Evaluation metrics matter: predicting sentiment from financial news headlines. *CoRR abs/1705.00571* (2017). <http://arxiv.org/abs/1705.00571>
- [19] Venkata Sasank Pagolu, Kamal Nayan Reddy Challa, Ganapati Panda, and Babita Majhi. 2016. Sentiment Analysis of Twitter Data for Predicting Stock Market Movements. *CoRR abs/1610.09225* (2016). <http://arxiv.org/abs/1610.09225>
- [20] Alexander Pak and Patrick Paroubek. 2010. Twitter as a corpus for sentiment analysis and opinion mining. In *LREC*, Vol. 10.
- [21] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Ion Androutsopoulos, Suresh Manandhar, Mohammed AL-Smadi, Mahmoud Al-Ayyoub, Yanyan Zhao, Bing Qin, Orphée De Clercq, Veronique Hoste, Marianna Apidianaki, Xavier Tannier, Natalia Loukachevitch, Evgeniy Kotelnikov, Nürta Bel, Salud Maria Jiménez-Zafra, and Gülşen Eryigit. 2016. SemEval-2016 task 5 : aspect based sentiment analysis. In *Proceedings of the 10th International Workshop on Semantic Evaluation (SemEval-2016)*. Association for Computational Linguistics, 19–30. <http://www.aclweb.org/anthology/S16-1002>
- [22] Maria Pontiki, Dimitris Galanis, Haris Papageorgiou, Suresh Manandhar, and Ion Androutsopoulos. 2015. *SemEval-2015 Task 12: Aspect Based Sentiment Analysis*. The Association for Computational Linguistics, 486–495.
- [23] Maria Pontiki, Dimitris Galanis, John Pavlopoulos, Harris Papageorgiou, Ion Androutsopoulos, and Suresh Manandhar. 2014. SemEval-2014 Task 4: Aspect Based Sentiment Analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*. Association for Computational Linguistics and Dublin City University, Dublin, Ireland, 27–35. <http://www.aclweb.org/anthology/S14-2004>
- [24] Ana-Maria Popescu and Oren Etzioni. 2005. Extracting Product Features and Opinions from Reviews. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT '05)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 339–346. <https://doi.org/10.3115/1220575.1220618>
- [25] Guang Qiu, Bing Liu, Jiajun Bu, and Chun Chen. 2009. Expanding Domain Sentiment Lexicon Through Double Propagation. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI'09)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 1199–1204. <http://dl.acm.org/citation.cfm?id=1661445.1661637>
- [26] Leon Rotim, Martin Tutek, and Jan Snajder. 2017. TakeLab at SemEval-2017 Task 5: Linear aggregation of word embeddings for fine-grained sentiment analysis of financial news. In *Proceedings of the 11th International Workshop on Semantic Evaluation, SemEval@ACL 2017, Vancouver, Canada, August 3-4, 2017*. 866–871. <https://doi.org/10.18653/v1/S17-2148>
- [27] Angel Deborah S, S Milton Rajendram, and T T Mirnalinee. 2017. SSN_MLRG1 at SemEval-2017 Task 4: Sentiment Analysis in Twitter Using Multi-Kernel Gaussian Process Classifier. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)*. Association for Computational Linguistics, Vancouver, Canada, 709–712. <http://www.aclweb.org/anthology/S17-2118>
- [28] Gerard Salton and Christopher Buckley. 1988. Term-weighting Approaches in Automatic Text Retrieval. *Inf. Process. Manage.* 24, 5 (Aug. 1988), 513–523. [https://doi.org/10.1016/0306-4573\(88\)90021-0](https://doi.org/10.1016/0306-4573(88)90021-0)
- [29] Iñaki San Vicente, Xabier Saralegi, and Rodrigo Agerri. 2015. EliXa: A Modular and Flexible ABSA Platform. In *Proceedings of the 9th International Workshop on Semantic Evaluation (SemEval 2015)*. Association for Computational Linguistics, Denver, Colorado, 748–752. <http://www.aclweb.org/anthology/S15-2127>
- [30] Christopher Scaffidi, Kevin Bierhoff, Eric Chang, Mikhael Felker, Herman Ng, and Chun Jin. 2007. Red Opal: Product-feature Scoring from Reviews. In *Proceedings of the 8th ACM Conference on Electronic Commerce (EC '07)*. ACM, New York, NY, USA, 182–191. <https://doi.org/10.1145/1250910.1250938>
- [31] Jasmina Smailovic, Miha Grear, Nada Lavrac, and Martin Znidarsic. 2014. Stream-based active learning for sentiment analysis in the financial domain. *Information Sciences* 285 (2014), 181 – 203. <https://doi.org/10.1016/j.ins.2014.04.034> Processing and Mining Complex Data Streams.
- [32] Pyry Takala, Pekka Malo, Ankur Sinha, and Oskar Ahlgren. 2014. Gold-standard for Topic-specific Sentiment Analysis of Economic Texts. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)* (26-31), Nicoletta Calzolari (Conference Chair), Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis (Eds.). European Language Resources Association (ELRA), Reykjavik, Iceland.
- [33] Li Zhuang, Feng Jing, and Xiao-Yan Zhu. 2006. Movie Review Mining and Summarization. In *Proceedings of the 15th ACM International Conference on Information and Knowledge Management (CIKM '06)*. ACM, New York, NY, USA, 43–50. <https://doi.org/10.1145/1183614.1183625>