SENTIMENT ANALYSIS FOR FINANCIAL MARKET PREDICTION

T. H. H Methmal

168246D

Degree of Master of Science in Computer Science

Department of Computer Science and Engineering

University of Moratuwa
Sri Lanka

April 2020

SENTIMENT ANALYSIS FOR FINANCIAL MARKET PREDICTION

Thommaya Hewa Hasanga Methmal

168246D

Dissertation submitted in partial fulfillment of the requirements for the degree Master of Science in Computer Science specializing in Data Science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

April 2020

DECLARATION

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text.

Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my dissertation, in whole or in part in print, electronic or other medium. I retain the right to use this content in whole or part in future works (such as articles or books).

Signature: ……..... Date: …….....

Name: T. H .H Methmal

I certify that the declaration above by the candidate is true to the best of my knowledge and that this report is acceptable for evaluation for the CS6997 - MSc Research Project.

The above candidate has carried out research for the Masters Dissertation under my supervision.

Signature: ……..... Date: …….....

Name: Dr. Surangika Ranathunga

ABSTRACT

Today's world is highly dependent on financial markets. Financial markets are very dynamic, making it difficult to predict prices. If we can make good predictions, we can make financial gains without any risks. There are two main categories of data that we can use to predict the market price; historic market data and textual data. Most importantly, textual analysis on sources like news, social media and reports is more popular today among researchers. This process can be introduced as sentiment analysis. The whole idea behind sentiment analysis is checking the opinion behind the text; whether it is a positive, negative, or neutral polarity.

This research focuses on sentiment-analysis-based financial market prediction using deep leaning. Market prediction using sentiment analysis is a very challenging task. There are complex linguistic issues to solve and using a microblog dataset like Twitter for the prediction task makes it even more difficult. However, the current prediction approach rarely exceeds the seventy percent accuracy mark. This research is based on the SEMEVAL 2017 fifth task and will use the same dataset shared by the SEMEVAL team.

This thesis presents an improved version for above mention reported baseline. We experiment with different techniques in both machine leaning and deep learning domains. Lexicon based dictionaries are heavily used here in each model since this is a small dataset with train set (1693) and test set (793). We had to enlarge the dataset as much as possible to achieve good accuracy. We created mainly four models which are based on machine learning and deep leaning techniques. Support vector regression algorithm is used for the machine leaning model. Also we used convolutional neural network (CNN) , long short term memory (LSTM) and gated recurrent unit (GRU) as deep leaning architectures which are performed better than any of the baseline models on this dataset. Our deep leaning models are achieved maximum similarity scores than any of the single system.

Finally, we experiment three main ensemble techniques which are Averaging Linear Regression and multilayer perceptron. We achieved best results from averaging ensemble model.

Keywords: Financial Sentiment Analysis; Opinion Mining; Reviews; Text Analysis; Deep learning; Sentiment Analysis Challenges; Semeval.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my supervisor, Dr Surangika Ranathunga for the continuous support from the first day of my M.Sc. study, for her patience, motivation, enthusiasm, and immense knowledge. Besides my supervisor, I would like to thank the rest of my teachers throughout the journey for guiding me towards the right path in life. Last but not least, I would like to thank my mother, without whom I would not be here, and for her support in guiding me spiritually throughout my life.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

GDP	Gross Domestic Product
FOREX	Foreign Exchange Market
MSH	Morgan Stanley High-Tech Index
LDA	Latent Dirichlet Allocation
IG	Information Gain
CHI	Chi-square Statistics
DF	Document Frequency
TF	Term Frequency
IDF	Inverse Document Frequency
SVM	Support Vector Machine
CATS	Categorization And Trading System
SVC	Support Vector Classification
SVR	Support Vector Regression
RBF	Radial Basis Function Kernel
OLS	Ordinary Least Square
GI	The General Inquirer
LMFSD	Loughran-McDonald financial sentiment dictionary
CBOW	continuous bag-of-words
SG	skip-gram
GLOVE	Global Vectors
ABR	AdaBoost Regressor
BR	Bagging Regressor
RF	Random Forest
CNN	Convolutional Neural Network