UNIVERSITETI I EVROPËS JUGLINDORE
УНИВЕРЗИТЕТ НА ЈУГОИСТОЧНА ЕВРОПА
SOUTH EAST EUROPEAN UNIVERSITY

Applied Text-Mining algorithms for stock price prediction based on financial news articles

Adrian Besimi, Zamir Dika, Mubarek Selimi

Contemporary Sciences and Technologies

South East European University

Tetovë, R. of N. Macedonia

www.seeu.edu.mk

a.besimi@seeu.edu.mk

Cooperation at Academic Informatics Education across Balkan Countries and Beyond: The Impact of Informatics to Society. Jelsa, Croatia, 2019

Outline

Introduction

Our work

Applied Text-Mining algorithms for stock price prediction

Simulation

Conclusion

Can we predict stock price movements?

Short answer: NO

Long answer: NO, but...

INTRODUCTION

Stock market data and relevant news associated with fin-tech industry are increasing rapidly. Lots of investors are involved in stock market and they have a common interest in knowing more about the future of market in order to be able to have successful investments.

Information published in news articles influence, in a varying degree, the decision of the stock traders, especially if the given information is unexpected.

INTRODUCTION (2)

Sentiment analysis classifies textual data into positive, negative and neutral sentiments so this can be used to categorize a given textual article.

In our study we worked towards analysing data, concretely news articles and historical stock prices to make future prediction about stock direction.

Our work: Applied steps

Identifying the news sources and targeted companies

1. Data collection and data cleaning of news articles

1. Sentiment Analysis of news articles

1. Data collection of stock prices

1. Calculating Rate of Change (ROC)

1. Categorizing the data

1. Applying Naive Bayesian classifier

1. Training

Our work: Dataset totalling 20226 news articles

Variable	Categories	Frequencies	%
Source	BGR	1073	5.884
	Breitbart	435	2.385
	CNN	687	3.767
	Fox Business	813	4.458
	The Street	3810	20.893
	The Verge	2847	15.612
	The Washington post	6051	33.182
	market-watch	2520	13.819
Company	Apple	7591	41.626
	Facebook	7513	41.199
	Tesla	3132	17.175

Our work: Sentiment analysis

Vader Sentiment Analysis was used. VADER (Valence Aware Dictionary for sEntiment Reasoning) is a pre-built sentiment analysis model included in the NLTK package of Python.

VADER however is focused on social media and short texts, unlike Financial News which are almost the opposite. We updated the VADER lexicon with words plus sentiments from other sources/lexicons such as the Loughran-McDonald Financial Sentiment Word Lists, to be appropriate for our collected financial news

Our work: Dataset with sentiment calc

Author	Date	Content	link	Source	Company	neg	pos	neu	compound	5-day ROC	Future ROC	Sentiment WithoutNeutralClass	ROC_Sentiment	CLASS
Tauhid C	4/5/2018			The Washington post	Apple	0.031	0.08	0.889	0.9954	3.79626	0.77546	POSITIVE	POSITIVE	UP
Steve He	4/5/2018			The Washington post	Apple	0.064	0.036	0.9	-0.9818	3.79626	0.77546	NEGATIVE	POSITIVE	NEUTRAL
Juli Brisk	4/5/2018			The Washington post	Facebook	0.029	0.041	0.93	0.708	4.12337	2.84298	POSITIVE	POSITIVE	UP
Geoffrey	4/5/2018			The Washington post	Facebook	0.03	0.076	0.894	0.9966	4.12337	2.84298	POSITIVE	POSITIVE	UP
Amy Dick	4/5/2018			The Washington post	Facebook	0.028	0.021	0.952	-0.4404	4.12337	2.84298	NEGATIVE	POSITIVE	NEUTRAL
Eugene I	4/5/2018			The Washington post	Facebook	0.048	0.063	0.89	0.7114	4.12337	2.84298	POSITIVE	POSITIVE	UP
Joshua F	4/5/2018			The Washington post	Facebook	0.057	0.054	0.889	-0.7778	4.12337	2.84298	NEGATIVE	POSITIVE	NEUTRAL
David Ig	4/5/2018			The Washington post	Facebook	0.062	0.05	0.888	-0.8442	4.12337	2.84298	NEGATIVE	POSITIVE	NEUTRAL
Lenny Be	4/5/2018			The Washington post	Facebook	0.055	0.03	0.915	-0.9832	4.12337	2.84298	NEGATIVE	POSITIVE	NEUTRAL
	4/5/2018			market-watch	Apple	0.122	0.013	0.865	-0.8834	3.79626	0.77546	NEGATIVE	POSITIVE	NEUTRAL
	4/5/2018			market-watch	Apple	0.051	0.14	0.809	0.998	3.79626	0.77546	POSITIVE	POSITIVE	UP
Brett Are	4/5/2018			market-watch	Facebook	0.044	0.033	0.924	-0.7651	4.12337	2.84298	NEGATIVE	POSITIVE	NEUTRAL
	4/5/2018			market-watch	Facebook	0.015	0	0.985	-0.2382	4.12337	2.84298	NEGATIVE	POSITIVE	NEUTRAL
	4/5/2018			market-watch	Tesla	0.056	0.025	0.919	-0.6249	18.5973	-3.80741	NEGATIVE	NEGATIVE	DOWN
	4/5/2018			market-watch	Tesla	0	0.081	0.919	0.296	18.5973	-3.80741	POSITIVE	NEGATIVE	NEUTRAL
Paul R. La	4/5/2018			CNN	Apple	0.042	0.113	0.845	0.9944	3.79626	0.77546	POSITIVE	POSITIVE	UP
Richard C	4/5/2018			CNN	Facebook	0.049	0.089	0.862	0.9816	4.12337	2.84298	POSITIVE	POSITIVE	UP
Ivana Ko	4/5/2018			CNN	Facebook	0.033	0.073	0.894	0.9231	4.12337	2.84298	POSIVE	POSITIVE	UP
Zach Eps	4/5/2018			BGR	Apple	0.021	0.069	0.911	0.985	3.79626	0.77546	POSITIVE	POSITIVE	UP
Chris Smi	4/5/2018			BGR	Facebook	0.085	0.028	0.888	-0.9849	4.12337	2.84298	NEGATIVE	POSITIVE	NEUTRAL
Motley Fc	4/5/2018			Fox Business	Apple	0.009	0.046	0.945	0.9606	3.79626	0.77546	POSITIVE	POSITIVE	UP
Megan H	4/5/2018			Fox Business	Facebook	0.054	0.021	0.925	-0.791	4.12337	2.84298	NEGATIVE	POSITIVE	NEUTRAL
Motley Fc	4/5/2018			Fox Business	Tesla	0.034	0.103	0.863	0.9844	18.5973	-3.80741	POSITIVE	NEGATIVE	NEUTRAL
Kinsey Gr	4/5/2018			The Street	Apple	0.03	0.076	0.894	0.9568	3.79626	0.77546	POSITIVE	POSITIVE	UP

Applied Text-Mining algorithms for stock price prediction

TRAINING THE MODEL with categorizing the data from previous steps

Training Set of 18236 records/articles) and Test Set 1990 records/articles.

The following variables are used to train and test the first model: *Source, Company, Sentimentof_text* and the *5-day ROC*

The algorithm applied classifies 15.71% of the articles in the training set as "DOWN" , 50.71% is classified as "NEUTRAL" and 33.59% of the data as "UP" (meaning the stock will go up).
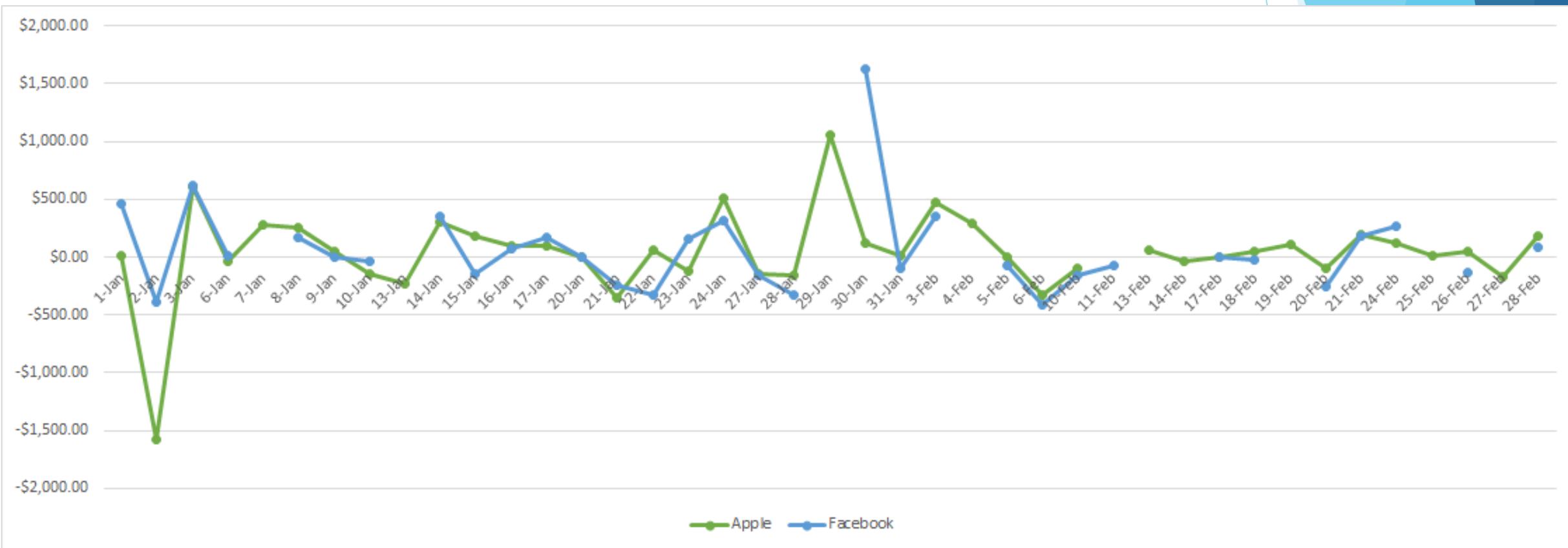
REMARK: Once the 5 days Rate of Change is removed as variable the whole efficiency of predicting goes down!

Simulation to see if algorithm makes sense?

SIMULATION: Profit/Loss simulation on Test set data based on classification model
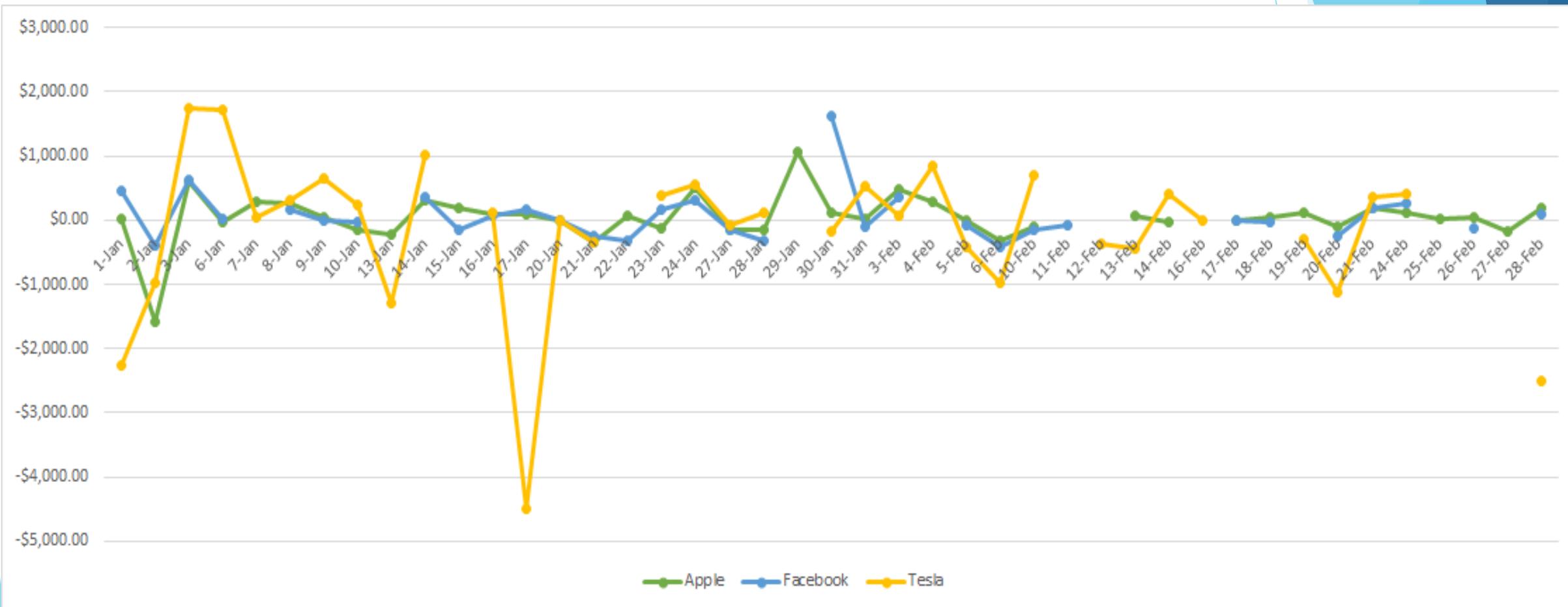
SIMULATION: Profit/Loss simulation on Test set data based on classification model (-Tesla)

CONCLUSION

► Adding more variables in top of the sentiment analysis of financial news articles provide more information for future movements of stock markets.

► Unfortunately, there is **no 100% prediction** for the future of stock prices, and the main reason is that there are too many variables included that can change and that are unpredictable.

► The simulation conducted does not show 100%-win case for the classification of stock prediction and as such it does not apply to all companies. The difference where there are better results relies on the targeted companies, such as **Apple** and **Facebook**, which are more stable ones rather than Tesla, which as a case had different fluctuations that in long term did not bring good results in our simulation.

► **The simulation resulted in $3,716.00 profit in a period of 2 months on daily basis investments of $20.000,00**

UNIVERSITETI I EVROPËS JUGLINDORE
УНИВЕРЗИТЕТ НА ЈУГОИСТОЧНА ЕВРОПА
SOUTH EAST EUROPEAN UNIVERSITY

Questions?

Assoc. Prof.
Dr. Adrian Besimi

Contemporary Sciences
and Technologies

South East European
University

Thank you

Tetovë, N. Macedonia

a.besimi@seeu.edu.mk