Report on **Bachelor / Master Thesis**

Institute of Economic Studies, Faculty of Social Sciences, Charles University

Student:	**B.Sc. Dejan Lazeski**
Advisor:	**prof. Ing. Evžen Kočenda M.A., Ph.D., DSc.**
Title of the thesis:	**Stock Market Prediction: A Multiclass Classification on Emotions and Sentiment Analysis for Tweets and News Headlines**

OVERALL ASSESSMENT *(provided in English, Czech, or Slovak):*

Contribution

The bachelor thesis focuses on extracting sentiment and emotions from news headlines and tweets addressing the top 5 Big Tech companies (Google, Apple, Facebook, Microsoft, and Amazon). Dejan constructs models to categorize the headlines tweets in five emotions (Anger, Fear, Joy, Surprise, and Sadness) or sentiment polarity (positive, negative, or neutral). He uses the supervised learning methods to train and evaluate their performance in predicting the outcome of observations. The methods for accuracy classification and classification analysis are well described.

In the empirical part, Dejan did not confirm any significant correlation between automatically classified emotions or sentiment and future stock price returns. However, this is not surprising as the author uses daily data, and the frequency of sentiment updates are also daily. In this setting, there is an elevated probability of a high noise level. Thus, even in case of strong predictive abilities of the sentiment measure, the prediction accuracy would be rather low.

Another empirical part is rather confusing and needs clarification during the defense. Dejan constructs trading strategy using classifiers' predictions. These predictions are compared to the buy-hold signals representing the historical trend lines of a company stock price movement. The result is mostly inconclusive or contradictory. Especially the Apple stock price predictions (Table 5.5) shows some fundamental problems that need to be addressed.

There is a clear contribution of the thesis in using five emotions classification for financial data. The topic is hot with great potential.

Methods

Dejan uses standard methods for sentiment and emotions classifications. The machine learning techniques are correctly applied.

Literature

Dejan uses relevant literature in the. Considering the work with literature, it is clear that Dejan understands the literature, and he quotes appropriately. Moreover, the literature overview section is well written.

Manuscript form

The manuscript form has a clear and logical structure. However, in the empirical parts, there are sections where it is difficult to follow the ideas and methods used for the analysis. On the other hand, sections covering sentiment classification is very well written. No evidence of significant text similarity with other sources has been found in the Urkund analysis.

Report on **Bachelor / Master Thesis**

Institute of Economic Studies, Faculty of Social Sciences, Charles University

Student:	**B.Sc. Dejan Lazeski**
Advisor:	**prof. Ing. Evžen Kočenda M.A., Ph.D., DSc.**
Title of the thesis:	**Stock Market Prediction: A Multiclass Classification on Emotions and Sentiment Analysis for Tweets and News Headlines**

Summary and suggested questions for the discussion during the defense
To conclude, the thesis is a nice piece of work. It has a contribution. However, there are parts that need a more careful approach. The thesis meets requirements for a master thesis at IES, Faculty of Social Sciences, Charles University. I, therefore, recommend it for the defense and suggest a grade C, with the possibility of improvement when the raised issues will be addressed. For the defense I suggest the two following questions:

- What can be the reason for the opposite direction of classifiers predictions and the buy-hold signal (Figure 5.3)?

- Why the buy buy-hold strategy uses t-3 lagged value?

SUMMARY OF POINTS AWARDED *(for details, see below):*

CATEGORY		**POINTS**
Contribution	*(max. 30 points)*	25
Methods	*(max. 30 points)*	25
Literature	*(max. 20 points)*	20
Manuscript Form	*(max. 20 points)*	10
TOTAL POINTS	*(max. 100 points)*	**80**
GRADE	**(A – B – C – D – E – F)**	**C**

NAME OF THE REFEREE:

DATE OF EVALUATION: 8.9.2020

Digitálně podepsáno (8.9.2020)
Lukáš Vácha

_____

Referee Signature