



Universität Paderborn  
Fakultät für Wirtschaftswissenschaften  
Department Wirtschaftsinformatik

Exposé zur Studienarbeit

## **Predictive Analytics**

### **Daily Financial News for 6000+ Stocks**

von

Benjamin Kasten

Matrikelnummer: 7154784

Warburger Str. 100, 33098 Paderborn

[bkasten@mail.uni-paderborn.de](mailto:bkasten@mail.uni-paderborn.de)

und

Dominik Höhr

Matrikelnummer: 7157684

Salentinstraße 16 A, 33102 Paderborn

[dhoehr@mail.uni-paderborn.de](mailto:dhoehr@mail.uni-paderborn.de)

vorgelegt bei

Prof. Dr. Oliver Müller

Prof. Dr. Erika Mustermann

15. August 2020

# 1 Expose

## 1.1 Forschungsfrage und Thema

Zu Grunde gelegter Datensatz: Daily Financial News for 6000+ Stocks (see Developer, 2020)

Der Datensatz liefert rund 1.410.000 Überschriften von Internetartikeln, über die Entwicklung von zu geordneten Aktien, welche an der Amerikanischen Börse notiert sind. Zu beachten ist, dass die Internetartikel hauptsächlich von auf benzinga.com veröffentlicht wurden, dabei bestehen die Artikel häufig nur aus den Überschriften oder einer Sammlung von aktuellen Zahlen der Aktie(n).

Angedacht ist eine Untersuchung der Zusammenhänge der Überschriften.

Dabei wird vor allem auf die Wirkung und Auswirkung der Überschriften eingegangen. Ziel der Studienarbeit ist es, ein geeignetes objektives Bewertungskriterium zu finden. Dabei geht es um eine ordinäre Ordnung.

Denkbar sind dabei das Zugrundelegen von Aktienmarktpreisen oder eines Sentiment-Score Wörterbuches.

Die konkrete Forschungsfrage und insbesondere der konkrete Nutzen für eventuelle Akteure hängt von den ersten Ergebnissen ab.

## 1.2 Hintergrund

*Wie sind Sie auf das Thema gekommen? Welche Hintergrundgeschichte führte zu der Forschungsfrage?*

Der Markt wird immer schnelllebiger, insbesondere der Aktienmarkt erlebt einen immer schnelleren Wandel.

Die Einflussnahme von Image und Schlagzeilen nimmt dabei ebenfalls ständig zu und die Auswirkungen sind nicht immer vollständig zu erfassen. Dabei muss sowohl von der redaktionellen als auch der unternehmerischen Seite reaktiv schnell gehandelt werden. Nach Möglichkeit der Forschungsergebnisse erhoffen wir uns zukünftig ein proaktives

Handeln auf Aktienentwicklungen und Überschriften von Schlagzeilen.

## 1.3 Motivation

*Warum ist dieses Thema bzw. diese Fragestellung untersuchenswert?*

Die Entwicklung, Gruppierung und Formulierung der Überschriften sowie deren Mengenverhältnis zu den Aktien können interessante Erkenntnisse bringen.

Eine Bewertung der Überschriften, jeglicher Form, sogt für eine schnellere Auswertung und daraus folgende schnellere Reaktion oder gar für ein proaktives Handeln.

Das schnelle agieren auf solche Internetartikel und insbesondere deren Überschriften ist insoweit wichtig, da diese kurzen Nachrichten durch die schnelllebigkeit des Internets rasch eine unüberschaubare Auswirkung nach sich ziehen.

Die Macht der redaktionellen Überschriften ist nicht zu vernachlässigen.

Ist ein Zusammenhang und eine ordinäre Ordnung zwischen den Überschriften erkennbar, so ist der jeweilige Nutzen daraus noch zu konkretisieren.

Abhängig des evaluierten Bewertungsschema, können weitere Forschungsfragen angeschlossen werden.

So ist zum Beispiel eine Vorhersage der Aktienentwicklungen denkbar.

## 1.4 Akteure und Profit

*Wer sind die möglichen Akteure? Wer könnte von unserer Forschung profitieren?*

Die entsprechenden börsennotierten Unternehmen könnten ein berechtigtes Interesse an deren Kategorisierung der Überschrift mit anderen Wettbewerbern haben. Dabei ist das Image des Unternehmens zu nennen, welches durch gemeinsame kategorisierung mit anderen Unternehmen auf dem Markt stark beeinflusst werden kann. Wie oben schon einleitend genannt ist es auch denkbar die Marktpreise heranzuziehen, hier ist es im Interesse des börsennotierten Unternehmens einer positive Markt Entwicklung gegenüber zu stehen.

## 1.5 Bisher

*Was wurde in diesem oder einem ähnlichen Gebiet bereits getan?*

Einige Anstrengungen, insbesondere mit dem zu Grunde gelegten Datensatz, ergänzten

den Datensatz mit den Finanzdaten der genannten Aktien oder haben ein ersten Topic Modelling vorzogen.

Wir möchten abschließend darauf hinweisen, dass eine konkrete Bewertung der Überschriften anhand der Aktienentwicklung nur mit entsprechenden Finanzdaten möglich ist. Die entsprechenden APIs sind dabei in der Größenordnung des Datensatzes nur kostenpflichtig zu erreichen.

(see Stegeman, 2020) (see Zhang, 2020)

## Literaturverzeichnis

Developer, B. (2020). Daily Financial News for 6000+ Stocks. <https://www.kaggle.com/miguelaelnle/massive-stock-news-analysis-db-for-nlpbacktests>.

Stegeman, D. (2020). YahooFinance. <https://www.kaggle.com/danielstegemann/yahoofinance>.

Zhang, L. (2020). ECon 7930 Topic Modelling. <https://www.kaggle.com/luckyzx/econ-7930-topic-model>.