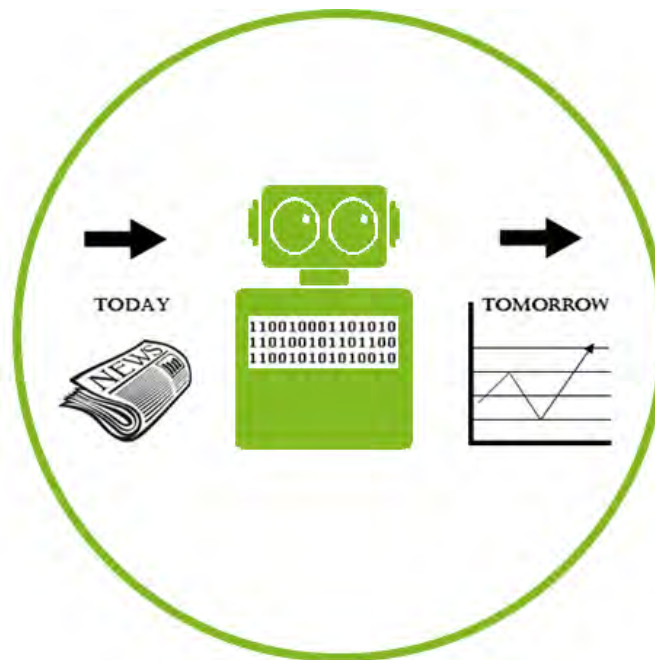


Predicting stock price movements based on newspaper articles using a novel deep learning approach

Internship Report



Emil Rijcken

Rob van der Mei, 1st supervisor

Antske Fokkens, 2nd supervisor

Maarten Ijlstra, Deloitte supervisor

VU University, Amsterdam



The project

The project

Problem

While stock price movements are known to be mostly influenced by news updates, most financial companies only include stock price information in their models. By not including news updates, a huge opportunity is missed.

Approach

In this thesis a state-of-the-art model is developed that 'reads' news articles from a large news source and outputs whether the stock of a particular company will go up or down, based on it.

Findings

It is indeed possible to make stock price predictions based on news updates. The model with the highest performance managed to predict 56.25% of all news articles correctly. Training this model is time intensive. However, it can make predictions of stock movements in fraction of seconds, once the model is trained.

Also, the technique used for this model can be extended to other domains as it is not specific to stock price prediction. The model is trained to classify different articles and thus to assign a label to each article that is inputted. The task of text classification can also be very useful when analyzing contracts, annual reports or other documents.

Host organization

Deloitte

Deloitte is the overarching brand name, operating in 150 countries, under which over 244 000 employees work globally in the fields of Accountancy, Consulting, Financial Advisory, Risk Advisory and Tax advisory (see figure 1). Within the Netherlands it is one of the largest professional services provider in these fields as well. Spread over 14 offices, Deloitte employs over 5 500 professionals in this country.

The vision of Deloitte is: ‘We aspire to be the Standard of Excellence, the first choice of the most sought-after clients and talent’.

Furthermore, its shared values describe the core principles within the company:

1. Integrity
2. Outstanding value to markets & clients
3. Commitment to each other
4. Strength from cultural diversity

Risk Advisory

Every company is facing risks at a daily basis. Through Risk Advisory, clients are taught how to utilize these risks and use it as a driver for further growth. The goal of Deloitte Risk Advisory is to enable companies to utilize risk management to become more agile. Within Risk Advisory there are different service areas [7]:

- Strategic & Reputation Risk
- Regulatory Risk
- Financial Risk
- Operational Risk
- Cyber Risk
- Managed Risk

Each service area might consist of several service lines.

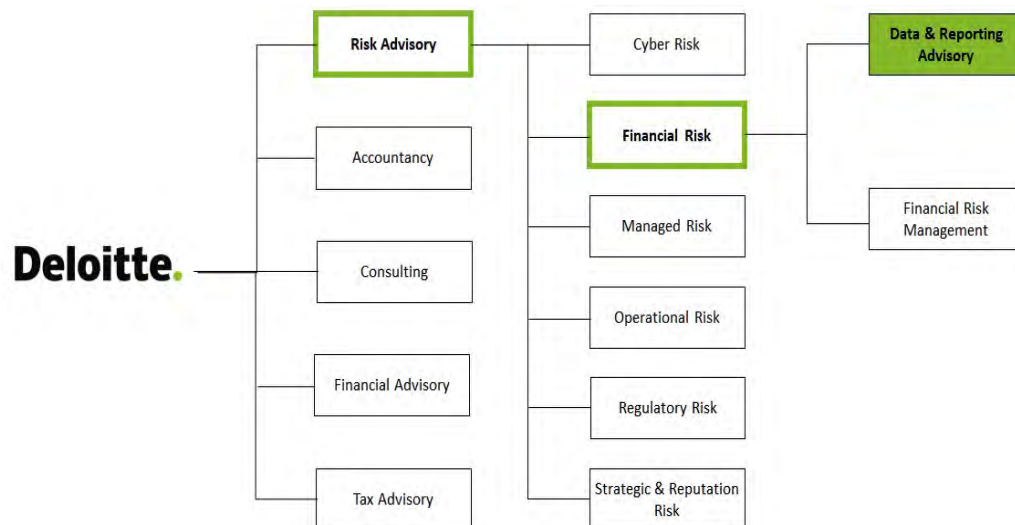


Fig. 1. Organization chart Deloitte

Data & Reporting Advisory

The Data & Reporting (DRA) service line is part of the Financial Risk service area. Due to changing demands of regulators, financial institutions have an urging need for support in the reporting process. DRA bundles its expertise in the fields of analytics, finance and accountancy to establish high quality reports, using data-driven solutions

Relationship between the project and relevant business

At Risk Advisory, the goal of the company is to help clients deal better with risks in the market. This thesis has the potential to contribute to this goal.

Furthermore, one of Deloitte's main focusses in the next years is to grow its capabilities in the machine learning domain. The use case of obtaining valuable information out of plain textual data can be used to demonstrate Deloitte's capabilities in this domain to current and potential clients and is in line with Deloitte's strategic goals.

Project conduction

Writing a thesis at Deloitte, means having a lot of freedom in determining and execution of the approach. Initially, a large dataset containing many financial articles from a large Dutch newspaper was made available. Based on this dataset, I formulated a relevant research question and approach based on my interests (text mining) and my personal learning goals (getting to know more about specific types of neural networks). After I an approval for this approach was given by both Deloitte as well as my VU supervisors, the newspaper withdrew and did not approve on using the same dataset anymore. Therefore, a new approach needed to be formulated.

Instead, a similar dataset was retrieved from the LexisNexis database by manually downloading 26 876 relevant articles. The master BA curriculum did not cover the usage of different types of neural networks (convolutional and recurrent in this case), therefore I did an Udemy course to get up to speed on this specific topic.

Under the supervision of Maarten Ijlstra, my supervisor from Deloitte, I created a project planning with appropriate deadlines. Besides the initial problems with the dataset, the approach and execution of this project went very smoothly as I met all my deadlines and obtained satisfactory results.

End products

Next to writing my thesis at Deloitte, I helped my colleagues on improving their knowledge in different/related fields. The following products are a result of my thesis period:

- *Written report*
 - o Contains an extensive description of the full approach, sources and results.
- *Thesis presentation*
 - o During the presentation I explained the (technical) approach in an intuitive, non-technical way so that all of my colleagues got an intuition as to how my algorithm was developed and why it works
 - o After the presentation I was invited by several partners to explain the approach one on one. They kept the slides for future usage.
- *Trainings*
 - o At my service line, not a lot of colleagues knew how to program. Together with a colleague I developed and gave a training in the R programming language. The slides of this training are also kept by the service line and the training will be given to more colleagues in the future.
 - o After going to a blockchain hackathon with a group of Deloitters, I developed and gave a workshop on Blockchain to my colleagues. In the workshop, I explained how it works intuitively and elaborated on relevant use cases for our service line.