

# Report 1

Tomi Theodor Kuusik, Kris Katmuk, Kristen Surva

29.11.2021

## Business understanding

### Background

News can be catalysts for stock market fluctuations or the other way around. The hypothesis is that some keywords in news headlines can predict market movement trends e.g “Oil prices rise to record levels” would correlate with the rise energy market stocks (in this case the keywords “oil” and “rise” would be the words). However correctly identifying such keywords is very difficult because the keywords do not always lead to market movement and correctly determining such keywords need more situational awareness which can be hard to achieve with machine learning. Hence the aim of the task is to determine if such model can be trained and if this task is possible.

### Business goals

1. Determine if a model that can predict stock market movement based on news headlines can be trained
2. Train such a model

### Business success criteria

1. Insight if a model that can predict market movement based on headlines can be trained
2. If possible, implemented

## Assessing your situation

### Inventory of resources

- Stock market data (809MB)
- Worldnews reddit headlines (82MB)

### **Requirements, assumptions, constraints**

- Regularize the data by date (i.e headlines for weeks and price movement based on simple moving average for 2 weeks data)
- For stock data etf's data or bigger stocks grouped by category should be used to reduce the data points.
- Smalles time range to be analysed should not get bigger than 1 week.
- Time constraint = deadline of the course project

### **Risks and contingencies**

- Stock market movement and keywords do not correlate, therefore predictions cannot be made.
- Data is too narrow and market movements lag behind/ahead headlines therefore connections do not seem apparent.

### **Terminology**

- TBD

### **Consts and benefits**

- TBD

## **Defining your data mining goals**

### **Goals**

- Deliver a model that given a set of keywords outputs most probable securities/sectors which this headline will impact.
- Get a list of best indicators for every security or sector.
- Get a list of the keywords with the highest correlation in market fluctuation

### **Sucess criteria**

- The model has to be atleast 60% accurate.