Capstone 3: News Classification

Problem statement formation

What opportunities exist for investment firms and hedge funds that trade stocks based on news sentiment?

Context

News classification is a hot topic in finance. Stocks move partially because of public interest, or their sentiment towards the news. The ability to identify which news to move a stock is usually one of the first steps to capture opportunities.

Criteria for success

In order to succeed, a high probability of classifying news must be achieved.

Scope of solution space

In order to understand what will drive the stock price, we need to understand simply classifying positive, negative or neutral sentiment would not provide a sustainable model for predicting stock movement. Thus, classifying news into different types is the first step of a trading strategy.

Constraints

The textual data we obtain may be unstructured and contain characters in other languages or encoding. And we need to keep in mind that most of the structures are imperfect. And ongoing supervision and correction is needed.

Stakeholders

Anyone who pays attention to news before trading and investing in a stock.

Data sources

Kaggle, Finviz

DSM steps

Step 1 Business understanding

We are trying to solve the problem of classifying news into different categories. This will give us the idea of which news will have a higher probability of moving a stock.

Step 2 Data wrangling

Textual data collected is highly unstructured, it most likely will have missing values, different languages, different encoding. We will clean the financial data for training and classify them into appropriate types.

Step 3 Preprocessing, training and modeling

We will split the data into training and test sets. And we build a neural network model for prediction.

Step 4 Documentation

A slide deck and report will be written to communicate the key takeaway of this project.