News Headlines as Predictors of Stock Market Movement

# Study Objective

While it is arguable what constitutes market information that affects the stock market, it is an interesting issue to investigate whether news headlines are a predictor of stock market movement. This paper investigates whether the Flex stock prices, traded on the NASDAQ stock market, move up or down based on the news headlines. How good are the news headlines as predictors of stock market movement?

# Audience for Study Results

This researcher will present this data science research report to the Chief Information Officer of Flex, a global manufacturing company. It is this researcher’s hope that this report will serve as a proof of concept (POC) of how data science can be used by Flex for predictive analytics. Hopefully, the CFO will make further investments to champion and sponsor additional data science projects to study organizational situations which benefit the company.

# Data Acquisition Approach

The following, two publicly available, datasets will be used for this study.

1. News headline data: From Newsmax archive at <http://www.newsmax.com/Archives/Newsfront/16/2008/1/>

These headlines are NOT ranked by readers.

Each date has more than one headline.

Date Range: 01-02-2008 to 12-01-2016

1. Stock data: NASDAQ data from <http://www.nasdaq.com/symbol/flex/historical>

Date Range: 01-02-2008 to 12-01-2016

# Methodology

Based on initial research, machine learning algorithms will be used.

1. Prepare Headline data table:
   1. Prepare data from <http://www.newsmax.com/Archives>
   2. Download and or copy/paste headlines from the website, into an excel spreadsheet.
   3. Prepare news headline data. Date in one column. Headline text in the second column. If one trading date has more than one headline, then combine multiple headlines into one blob of text for each trading day in a single cell.
2. Prepare stock data table
   1. Download data from <http://www.nasdaq.com/symbol/> into excel
   2. Date, Close, Volume, Open, High, Low
   3. Add column “Label”
   4. Assign binary classification value in the Label column for each row.
      1. Label value is “1” if the “Open” value is greater than or equal to “Close” value.
      2. Label value is “0” if the “Open” value is less than “Close” value.
3. The final data table will have three columns: Date, Label, HeadlineTxt
4. At this point,
   1. I am proposing to use XGBoost Algorithm for prediction. Don’t yet know how and why.
   2. Based on what I have read, XGBoost works only on numeric data.
   3. I am not sure what I need to do to convert the headlines into numeric values. Some numeric encoding will be required here. Not sure what needs to be done.
5. I want to understand how to use then XGBoost algorithm for predictive analytics. I want to promote the use of predictive analytics within Flex. I want to learn how to articulate the output of XGBoost in business terms.

# Deliverables

I will write a paper and also prepare a power point slide deck.