**Guidelines for the project proposal**

# Project team number (from Canvas) and the names of team members

# Anurag Sharma

# Priya Ashok - 1160974383

# Shruti Gupta - 116092517

# Vinamra Jain - 116348362

# Yanfang Huo

# Name of the project

**Prediction of Stock Movements using News**

# Abstract

**The project’s goal is to predict the stock prices through analyzing news data. The Stock Market Data is picked from *Intrino* while the news data is provided by *Thomson Reuters*. We shall be be evaluating how useful news data is to decide whether to invest in certain stocks.**

**We plan to apply regression models followed by classification to evaluate the performance of our model.**

# Main question

1. Is analyzing news data helpful in predicting stock prices ?

# Main variables

**Dependenet Variable -**

**Market Data**

1. **returnsOpenNextMktres10 – market return**

**Main Predictor Variables –**

1. **Open - the open price for the day**
2. **Close - the close price for the day**
3. **sourceTimestamp - UTC timestamp of this news item when it was created**
4. **irstMentionSentence - the first sentence, starting with the headline, in which the scored asset is mentioned.**

**1: headline**

**2: first sentence of the story body**

**3: second sentence of the body, etc**

# Main methods

1. Sentence Tokenization, lemmatization, stop work removal
2. Linear Regression
3. Classification

# Data source and link

* Market Data – Intrino
* News Data – Thomson Reuters
* https://www.kaggle.com/c/two-sigma-financial-news/data

# Data dictionary

Create a table for **all** the variables in your dataset (put the main DV first and *make it italic*)

|  |  |  |  |
| --- | --- | --- | --- |
| **#** | **Variable** | **Description** | **Type** |
| * 1111 | returnsOpenNextMktres10 | Market Return | Categorical - Three letter string |
| * 2 | time | UTC timestamp showing when the data was available on the feed | Continuous |
| * . | sourceTimestamp | UTC timestamp of this news item when it was created | Continuous |
|  | firstCreated | UTC timestamp for the first version of the item | Continuous |
|  | sourceId | an Id for each news item | Continuous |
|  | headline | the item's headline | Categorical |
|  | urgency | differentiates story types (1: alert, 3: article) | Categorical |
|  | time | the current time | Continuous |
|  | assetCode | a unique id of an asset | Continuous |
|  | assetName | the name that corresponds to a group of assetCodes | Categorical |
|  | volume | trading volume in shares for the day | Continuous |
|  | close | the close price for the day | Continuous |
|  | open | the open price for the day | Continuous |