

Group 2: Final Report

Di Ye (diye2), Jingyi Xie (jingyi3), Zhenyu Gu (zgu15), Zihe Wang (zwang199)

November 29, 2018

Contents

Introduction	1
Data Description	1
Data Cleaning	2
Missing Values	2
Data Preparation	2
Exploratory Data	2

Introduction

The stock data is from four companies, CME, CBOT, NYMEX and Comex.

CME Group Inc. (Chicago Mercantile Exchange & Chicago Board of Trade) is an American financial market company operating an options and futures exchange. It owns and operates large derivatives and futures exchanges in Chicago, New York City, and exchange facilities in London, using online trading platforms. It also owns the Dow Jones stock and financial indexes, and CME Clearing Services, which provides settlement and clearing of exchange trades. The exchange-traded derivative contracts include futures and options based on interest rates, equity indexes, foreign exchange, energy, agricultural commodities, rare and precious metals, weather, and real estate.

CBo Territoria SA engages in the development, promotion and management of residential and business real estate properties based on Reunion Island. The company operates through the following segment: Property development and Land. The Property development segment engages in the development and management of its own real estate property, consisting primarily of business real estate: activity centers, offices and businesses. The Land segment offer land management and regional development. CBo Territoria was founded on January 16, 2004 and is headquartered in Sainte-Marie, Reunion Island.

The New York Mercantile Exchange (NYMEX) is a commodity futures exchange owned and operated by CME Group of Chicago. NYMEX is located at One North End Avenue in Brookfield Place in the Battery Park City section of Manhattan, New York City. Additional offices are located in Boston, Washington, Atlanta, San Francisco, Dubai, London, and Tokyo.

COMEX is the primary futures and options market for trading metals such as gold, silver, copper and aluminum. Formerly known as the Commodity Exchange Inc., COMEX merged with the New York Mercantile Exchange (NYMEX) in 1994 and became the division responsible for metals trading.

Data Description

The datasets contain AlgoSeek Level 2 data from CME, CBOT, NYMEX and Comex. The Level 2 data is from the CME FIXFast feed which provides 10 levels of the book for the bid and the ask. The variables in the dataset are listed as below: - UTC: Timestamps in HHMMSSMMM in UTC time [MMM is for milliseconds] - Ticker: Instrument name (up to 8 characters) - Side: BUY/SELL side of the book - 10 levels of: Price x Contracts (Number of Orders). For example "1.38 x 12 (5)" means there are 12 contracts at price 1.38 from a total of 5 orders. Each row has 10 levels but there will only be data for the number of Depth levels (see depth field above).

Data Cleaning

Missing Values

We grouped the datasets by **s**, **m**, **h**, and **Side** to match the datasets with the label. We planned to use the mean price, open price (the best price, i.e. Level 1 price), volume (i.e. contracts) for each second as our features. We used `na.interpolation` and `na.replace` from R package `imputeTS` to fix the price of each level and set the volume to 0 if the price and volume are not available at a certain time.

Data Preparation

We calculated the mean and standard deviation of all the prices and volumes for all levels to scale the data for data preparation. We used 60% of the data as our train data, 20% validation data, and 20% test data to train and test our models.

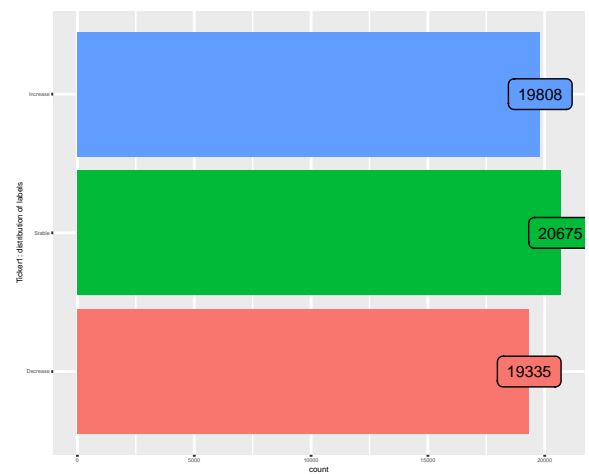
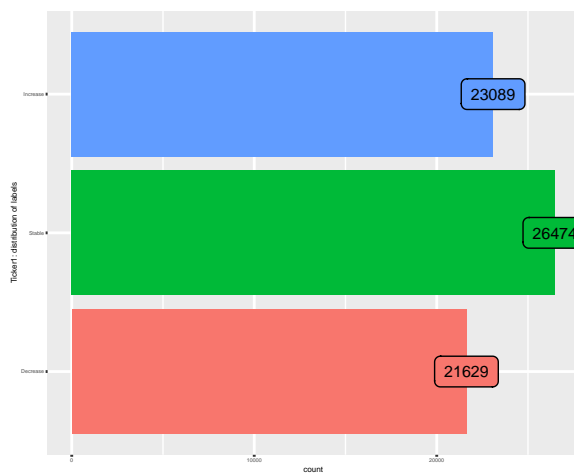
Exploratory Data

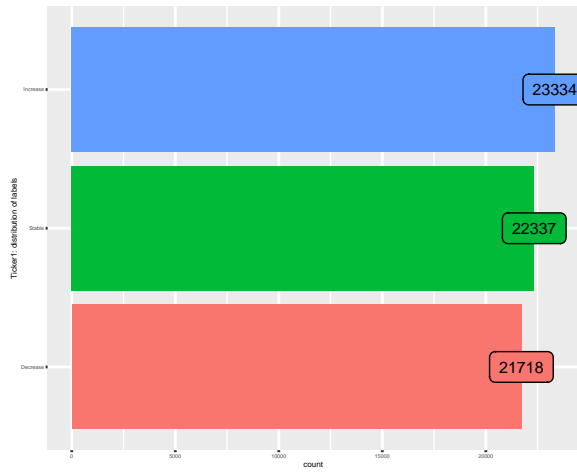
Before we started models, we have checked the labels and features to better understand the data.

```
library(ggplot2)
```

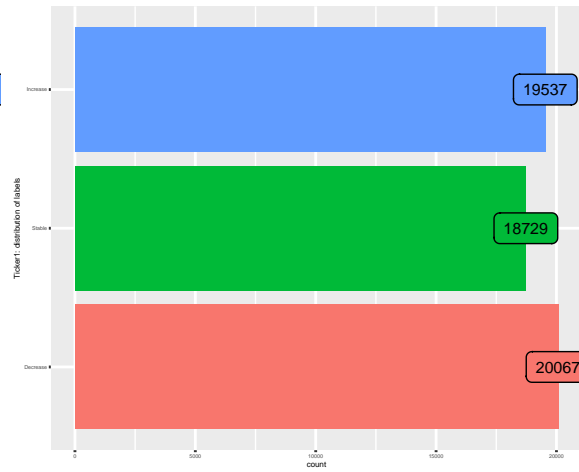
```
## Warning: package 'ggplot2' was built under R version 3.3.2
```

```
library(data.table)
ticker1_label <- fread("project data/Ticker1_label.csv")
ticker2_label <- fread("project data/Ticker2_label.csv")
ticker3_label <- fread("project data/Ticker3_label.csv")
ticker4_label <- fread("project data/Ticker4_label.csv")
```





direction a Decrease a Stable a Increase



direction a Decrease a Stable a Increase