



T.C.

MARMARA UNIVERSITY
FACULTY of ENGINEERING

CSE4062 Introduction to Data Science and Analytics

Spring 2021

Group #1

Title of the Project

“Machine Learning Approach to US Stock Investments”

Group Members

150118825 Ahmet Hakan Ekşi ahe9953@gmail.com

150215036 Canberk Köroğlu canberkkoroglu@hotmail.com

150215045 Erim Varış ermvrs@live.com

Lecturer

Doç. Dr. Murat Can Ganiz

Project Description

Successful predictions of stock movements have always been utterly important in the market. Billions of dollars were spent in infrastructures and R&D departments in order to predict the future outcomes of share-stocks. Can a computer master the art of finding values in stocks and beat the buy-hold strategies [1]?

It's no secret that machines are taking up a bigger and bigger share of investing, but the extent of their influence is approaching shocking proportions. It is as high as 80%, according to one major investing firm [2].

This means so much of stock trading is now in the hands of automated buyers and sellers that the market is increasingly sensitive to headlines and more prone to sharp price swings, many notable investors believe [2][3].

Aim of the project is to predict for US stocks future performances by utilizing 200+ indicators and data throughout 2014-2018. Such is that prediction will decide if one should invest or not in a specific stock.

Data Statistics

Our dataset consists of 4 csv files [1]:

- 2014_Financial_Data.csv → 3808 rows, 225 columns
- 2015_Financial_Data.csv → 4120 rows, 225 columns
- 2016_Financial_Data.csv → 4797 rows, 225 columns
- 2017_Financial_Data.csv → 4960 rows, 225 columns
- 2018_Financial_Data.csv → 4392 rows, 225 columns

First column indicates specific stock name.

The attribute named "Sector" indicates sector of corresponding stock, nominal type.

The attribute named "PRICE VAR [%]" indicates the percent price variation of the corresponding stock for the next year, numeric type.

The attribute named "class" is for classification and it's generated from "PRICE VAR [%]" column, binary type. If "PRICE VAR [%]" positive for corresponding stock then class attribute is 1, else is 0.

Other attributes are financial indicators, numeric type.

References

- [1] 200+ Financial Indicators of US stocks (2014-2018)
Available: "<https://www.kaggle.com/cnic92/200-financial-indicators-of-us-stocks-20142018>" (Date of Access: 16/04/2021)
- [2] 80% of the stock market is now on autopilot
Available: "<https://www.cnbc.com/2019/06/28/80percent-of-the-stock-market-is-now-on-autopilot.html>" (Date of Access: 16/04/2021)
- [3] The stock market is now run by computers, algorithms and passive managers
Available: "<https://www.economist.com/briefing/2019/10/05/the-stockmarket-is-now-run-by-computers-algorithms-and-passive-managers>" (Date of Access: 16/04/2021)

P.S. our group consists of only 2 departments: CSE, ENVE but you stated that each group must consist of students from at least three different departments. Therefore, we are open for students from another department that is not CSE or ENVE.