Homework 11 – Project Report

MONEYTALKS

Adriano Augusto – Grace Achenyo Okolo

https://github.com/nemo-91/moneytalks

EX.1

NOTE: our project is not meant to benefit directly a specific business, we refer to investors in general. These latter can be either individuals or associations (i.e. funds management companies).

1. **Identifying Business Goals**

Background

Often, companies need to rise funds to finance their future projects, or cover their debts. Although there are many alternatives to rise funds, one of the most popular alternative is to list the company on a financial market. Whenever a company is listed on a financial market (i.e. the New York Stock Exchange, NYSE), it is fragmented into shares (identified by a universal code of the type: company-code[.|:]market-code, e.g. RACE.NYSE, called stock, equity or security). Each share is either retained by the company itself (owner, chiefs, administrators etc.) or sold to investors at a front value decided **only initially** by the company (i.e. 1$). Daily (Monday-Friday, 9.30-16.00), investors can buy or sell their shares on the financial market the company has been listed. Investors are free to set their **bid** and **ask** prices, however, only when **ask price** match **bid price,** a trade takes place, and the stock price is updated to the price of last trade occurred. *The first trade of the day* determines the so-called: **opening price** of the stock. Whilst, *the last trade of the day* determines the so-called: **closing price** of the stock. During a trading day, the stock price is also estimated in terms of percentage change with respect to the previous day closing price. E.g. if the last trade occurred on Monday for RACE.NYSE was at 100.00$, and the first trade occurred on Tuesday for RACE.NYSE is 101.00$, RACE.NYSE had a change of 1%. Investors buy a stock when they think its price will grow in the short or long term, i.e. forecast of company growth. Vice-versa, they sell a stock when they either had a profit (i.e. they bought a stock at a lower price than the actual price) or they think its price will sink.
Although several researches have showed to be impossible the forecast of a stock price, rules of thumbs, investors' behaviour understanding, and financial markets knowledge provide a huge help for driving investors decisions to maximize their profits and minimize their losses.

Business Goals

Investors goal is only one: **maximizing the profit, reducing the losses.** Quantifying such profit and losses is usually difficult, and it varies according to the investor financial background. However, on average, investors (not traders) expect to have a 10% profit with a 2% loss risk.

Business Success Criteria

In a real case study, the success of this project could be quantified according to the annual return of an investment, that is expected to be between -2% and 10%.

2. Assessing your Situation

Inventory of Resources

- 2 students, with 25h workload.
- 2 HP laptops with Intel Core i5-6200U CPU @ 2.3GHz and 16GB RAM, running Windows 10 Pro.
- Microsoft Office Excel and Word, R and RStudio, Notepad++.

Requirements, Assumption, and Constrains

The project will be completed by the 8th of January. No security neither legal obligations, requirements for acceptable finished work are: good understanding and use of data mining techniques studied during the course, discovery of interesting facts out of the datasets analysed and clear presentation of what has been achieved.

Risks and Contingencies

The project may not lead to interesting results. The time estimated may not be enough. Christmas is ahead, and holiday mood may negatively affect the performance of the team.

Terminology

Share – a fragment of a company that is sold or bought by the company itself or by investors.
Stock, Security, Equity – refers to a company shares.
Stock Price – the last price the stock was bought on the stock market.
Stock Opening Price (of a day) – the price the stock is bought during the first trade of the day.
Stock Closing Price (of a day) – the price the stock is bought during the last trade of the day.
Volume – the number of shares traded during the day.
Stock % Change (on a timeframe) – the change in price of a stock over a specific timeframe.

We assumed that data-mining terminology is not necessary in our context.

Costs and Benefits

Costs – at least 50h of time, struggles and nightmares.
Benefits – up to 20 points on the final grade of the course.
We have a conflict of interests in determining whether the costs exceed the benefits or not.

3. Defining your data-mining goals

Data-mining Goals

1 – Identify patterns in the price movements of the stocks composing the NASDAQ-100 index.
2 – Draw general advices for the investors according to the patterns identified.
3 – Analyse the relation between *opening price* and *closing price* in terms of % change.
4 – Train and evaluate a classifier able to provide BUY or SELL advices for investors.

Data-mining success criteria

1 – Achieving at least 75% of accuracy with the designed classifier.

EX.2

1. Gathering data

Outline data requirements

- Gather the historical data of price movements of stocks.
- Data-format is CSV.
- Timeframe for each historical data is 5 years: 30-Nov-2012 to 30-Nov-2017.

Verify Data Availability

All the data is available and public. One of the easiest way to access it is through the website https://investing.com/. This latter is a popular website used by individual investors, it sources the data straight from the financial markets information systems.

Selection Criteria

Studying all the available stocks in the world would be just impossible, since it would be more than 100 thousand. Therefore, we selected for this project the 107 stocks composing the well-known index NASDAQ-100, on 9th Dec 2017. For further information about this index, please refer to: https://en.wikipedia.org/wiki/NASDAQ-100.

2. Describing Data

Each historical dataset of each stock contains up to 1262 entries. Some stocks have less entries due to the fact the corresponding companies were listed on the market after 30-Nov-2012. Each dataset contains 7 features, here described:

date [30-Nov-2012 to 30-Nov-2017] – the date of the trading day the other features refer to;
price [0 to +∞] – the closing price of the stock;
open [0 to +∞] – the opening price of the stock;
high [0 to +∞] – the highest price of the stock recorded during the trading day;
low [0 to +∞] – the lowest price of the stock recorded during the trading day;
vol. [0 to +∞] – the volume of the stock;
change % [-∞ to +∞] – the stock % change at the end of the trading day.

3. Data Exploration

Data exploration is not doable at this stage, since we do not have the resources (i.e. time) to explore separately 107 datasets, which are assumed to be uncorrelated. Data exploration will be performed after data preparation. Please refer to the timeline of the project in EX.3. For the data ranges, refer to the previous subsection.

4. Verifying Data Quality

The data is reliable and complete. However, we will need to heavily manipulate it to obtain extra useful information. For example, the data does not report the change in % of a stock price at the opening time, but only at the closing time.

EX.3

The repository is on GitHub: https://github.com/nemo-91/moneytalks

The project slide is available both on GitHub and at the following link: https://docs.google.com/presentation/d/1veA_WQcfRRx7hQnE8qklmsLYceWzQGrPHrieSEWqcaI/edit#slide=id.g2a54911d76_0_0

Project Plan by Steps, work allocation in brackets, timeframes are based on previous homework individual performances.

1. Gathering the data *(Adriano), 1h 40m* – downloading of all the datasets from the website. It took roughly one minute per dataset, since for each download it was necessary to setup the timeframe and then check the data downloaded was correct.
2. Data refactoring *(Grace), 3h 20m* – for each entry of each dataset the feature *date* was divided into 3 unique features: *day, month, year.* At the same time, we added 4 new features that will be used later: the *stock name*, the *stock price % change at the opening,* the *stock price % change in a week,* the *stock price % change in a month (20 days).*
3. Data merging *(Adriano), 20m* – the 107 datasets are merged into a unique dataset, to ease the upcoming analysis.
4. Data exploration *(Adriano, Grace), 8h* – getting acquainted with the new and final dataset, understanding (via experiments) how the data can be represented in a useful manner to identify patterns.
5. Data representation *(Adriano, Grace), 6h* – generating plots according to the insights of the previous step. Highlighting of the stocks' prices movement patterns. Report these latter as results of the first goal.
6. Data analysis vol.1 *(Adriano, Grace) 6h* – analysing the plots and report about interesting findings, which can help investors to take better decisions. This task completes the second goal. Its outcome is supposed to be part of the final poster presentation.
7. Data analysis vol.2 *(Adriano, Grace) 10h* – in this second part of the data analysis, we focus on the stock price % change between opening and closing times. We will analyse the data, to find a possible relation between the two numbers, that can be supported by logical statements.
8. Preparing the dataset for the training *(Adriano, Grace) 2h* – the outcome of the previous step, should help us to identify a rule that can let us label each entry with a suggestion of *Buy* or *Sell.* This suggestion should be driven by the possible relation between opening and closing price % change. This step completes the third goal.
9. Training the classifiers *(Adriano, Grace) 4h* – finally we would give as training set the 80% of the dataset to different type of classifiers (Tree, Random Forest, K-NN).
10. Testing *(Adriano, Grace) 2h* – the classifiers are then tested on the remaining 20% of the dataset, and their accuracy is evaluated and compared. According to the evaluation results, we will choose one of the classifiers as the best, or if none of them performs enough well, we will cry and go back to step 7.

This document is available at: https://github.com/nemo-91/moneytalks
Named: *first_steps.pdf*