TITLE

Ethan Graham, Scott Gozdzialski, Huynh Vu

Graduate Students Master of Data Science

Southern Methodist University

Dallas, Texas

**1. Abstract** We endeavor to pull in several years of financial securities (stocks) information and save them into a data management system for easy querying and analysis. We can use the stocks package for Kaggle, found at <https://www.kaggle.com/benjibb/lstm-stock-prediction-20170507/data>. We can export the data into a data warehousing and querying system, such as MySQL.

Once the data is loaded into a database, we can perform several queries on the data and report the results.

**2. Introduction** As we move forward with the growth of data science, we have to look into ways data science and machine learning will continue to expand into the future. One of these markets is in stock trading, in the current market between 50% and 84% of all trading action is completed by high-frequency computers, and it is predicted as time goes on computer trading will only continue to expand. But, as computers are taking on more and more of this role in our society we still have to continue with human interactions with the market. A big way this will be done is in refinement of the stock trading algorithms. We will be focusing on the first step in the algorithm refinement process, understanding. We will be looking through stock market data, building a database to house the data, and gleaming insights into what this data shows.

**3. Data and Database construction** The data we will be using is found at kaggle.com, it contains different data point of the stock market from 2010 till 2016. The data is broke up into four different files. They cover the raw price of the different stocks not adjusted for splits, the prices of the stocks with adjustment for splits, a securities file that covers “general description of each company with division for sectors” (Kaggle.com), and a fundamentals files with “metrics extracted from annual SEC 10K fillings.” (Kaggle.com) This data will be placed in a SQL database to be used for analysis and to build the insights of this data.

**4. In-sights**

**4.1 In-sight 1**

**4.2 In-sight 2**

**4.3 In-sight 3**

**4.4 In-sight 4**

**4.5 In-sight 5**

**5. Visualization**

**6. Conclusiton**

7. References

<http://www.zerohedge.com/contributed/2012-17-26/84-all-stock-trades-are-high-frequency-computers-%E2%80%A6-only-16-are-done-human-tra>

https://www.kaggle.com/benjibb/lstm-stock-prediction-20170507/data

2. General Instructions

The maximum length of a manu­script is 4 pages, printed single-sided. Print all text, including section titles and figures, in two-column format where each column is 8.5 cm by 24.5 cm (3.35 in by 9.65 in) and there is a 0.6 cm (0.24 in) space between the two columns. Excep­tions to the two-column format include the title at the top of the first page and any full-width figures or tables. Start all pages directly under the top margin. Text should be centered on each page. On A4 paper, this roughly means leaving 1.7 cm (0.67 in) margins on left and right sides of each page as well as a 2.5 cm (1 in) margin on the top and bottom of each page. Type single-spaced. Indent when starting a new paragraph. Use standard fonts such as Times New Roman or Computer Modern Roman, 10 points for text, 11 points (bold) sub­section headings, 12 points (bold) for section headings, 14 points (bold) for title, 11 points for authors’ names, and 10 points for their affiliations.

**2.1 The First Page**

Center the title across both columns. Use the two-column format only when you begin the abstract.

**Title:** Place the title at the top of the first page, followed by the authors’ names and their affiliations. Long title should be typed on two lines without a blank line intervening. Leave approximately 1 cm (0.39 in) between the title and the body of the first page.

**Abstract:** Type the abstract at the beginning of the first column. The abstract should be no longer than 200 words.

**Text:** Begin typing the main body of the text immediately after the abstract, observing the two-column format as shown in this example.

**2.2 Sections**

**Headings:** Type and label section and subsection headings in the style shown on these pages. Use numbered sections, in order to facilitate cross references.

**References:** Citations within the text appear in brackets as [ref. number]. Gather the full set of references together under the heading **References**; place the section before any **Appendices**, unless they contain references. Arrange the references in the order that they are cited in the text. Provide as com­plete a citation as possible, using a consistent format.

**Appendixes:** Appendixes, if any, directly follow the text and the references (but see above). Letter them in sequence and provide an informative title: **Appendix A Title of Appendix.**

**2.3 Footnotes**

Put footnotes at the bottom of the page. They may be numbered or referred to by asterisks or other symbols.[[1]](#footnote-1) Footnotes should be separated from the text by a line.[[2]](#footnote-2)

**2.4 Graphics**

**Illustrations:** Place figures, tables, and photographs in the paper near where they are first discussed, rather than at the end, if possible. Wide illustrations may run across both columns.

**Captions:** Provide a caption for every illus­tration; number each one sequentially in the form: "Figure 1. Caption of the Figure." "Table 1. Caption of the Table." Type the captions for figures below the figures. Type the captions for tables above the tables.

**3. Length of Camera-ready Manuscript**

For the length of camera-ready manuscripts, **a paper is limited up to 4 pages**. All illustrations, references, and appendices must be accommodated within these page limits. Any extra page beyond the first four pages will be deleted. **Please DO NOT put a page number in each pag**

**4. Submission Process**

1. Format your paper using this template.

2. Turn the hardcopy by Dec 4th before the lecture starts

### References

[1] T.A. Jones, “Writing a good paper,” *IEEE Trans. on General Writing*, Vol. 1, no. 2, pp.1-10, May 2002.

[2] K. Hwang, *Computer Arithmetic*, John Wiley, 1997.

You have to refer to the following rubric to get the good score as the isntructor will evaluate your paper based on it.

**Team Presentation Rubric: 100%**

1.      Presentation Slide Format: 30%

a.      Font Size (at least 24 font size): 9%

b.      Data Size (How much MB or GB): 7%

c.      Data Source URL: 6%

d.      H/W experimental Specifications (Database server, CPU speed, Memory Size): 8%

2.      Originality: 20%

* 1. How unique your idea different from the existing data analysis (10%)
  2. and your deliverable, that is,  what insights you find out (10%)

3.      Relevance with the topic in the class: 35%

a.      Data Analysis using RDB or NoSQL DB: 10%

b.      Implementation using SQL or other scrpts: 15%

c. Github link and the content that has your presentation file (ppt) and term paper: 10%

4.      Communicate with the instructor about the topic to get approval about the topic: 15%

**NOTE:**

1. **Peer Evaluation (Optional)**: You have to email to the instructor the peer evaluation about your team members for the term project and the presentation. It should be composed of:

If you don’t email me peer evaluation, I assume, all of you contribute the work fairly well.

**For example,** your team score is 95% and your peer evaluation by your team members are 100%, your score is 95 (= 95 x 100%)

|  |  |  |  |
| --- | --- | --- | --- |
| **Group NAME: ?**  **Member** **Name** | **Weight (100%)**  **Ex: 90% (of 100%)** | **1-2 lines of explanation what the member has done and the member’s contribution is** | **NOTE** |
| Yourself |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

1. **Plagiarism**: If you make a copy of others, it should violate the academic integrity so that you should get 0 in the term project or F in the course in the worst case.

**Term Paper Rubric: 100%**

It should be almost same as the team presentation. But, mostly, I will take a look at if you revise the content per my **comment** at the presentation. Thus, any penalty at the presentation can be recovered. You also need to email the instructor the **peer evaluation** for the term paper. If you don’t email me peer evaluation, I assume, all of you contribute the work fairly well.

**For example,** your team score is 95% and your peer evaluation by your team members are 100%, your score is 95 (= 95 x 100%)

|  |  |  |  |
| --- | --- | --- | --- |
| **Group NAME: ?**  **Member** **Name** | **Weight (100%)**  **Ex: 90% (of 100%)** | **1-2 lines of explanation what the member has done and the member’s contribution is** | **NOTE** |
| Yourself |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

1. **Plagiarism**: If you make a copy of others, it should violate the academic integrity so that you should get 0 in the term project or F in the course in the worst case.

1. This is how a footnote should appear [↑](#footnote-ref-1)
2. Note the line separating the footnotes from the text [↑](#footnote-ref-2)