How to Use Stock Data for Data Science Education: A Simulated Trading Platform in Classroom

Yunkai Liu, Ph.D.
Department of Computer and Information Science
Gannon University
Erie, PA, United States
liu006@gannon.edu

Xiangjing Wei, Ph.D.
Department of Management
Gannon University
Erie, PA, United States
wei002@gannon.edu

***Abstract*— The research hereby presents an innovate practice to enhance data science education by integrating an in-house-developed stock trading platform. Students of Data Science area usually face the difficulty of understanding the complex of real-time data and sophisticated statistical indicators and models. Another difficulty is students are easy to lost interests and patience during the process of learning programming and analysis process, such as R-language. Stock market, as a huge data source and an important data disciplinary, is comparatively easy to attract student attentions. Thus, we developed an intuitive trading platform for education purpose. The platform contains three components: an exploring window, a control windows and a report window. The exploring window shows current and historical stock price trends and related indicators. Several representative stocks from different sectors can be picked and specific time frames can be assigned. The control window allows students to develop their own trading strategies. A trading strategy can be created by either intuitive way, or through single or combinations of indicators, or be built generically though plug-in R-programing module. The report window demonstrates the expected return of a stock in a specific time frame through a specific strategy. A more comprehensive report with detailed transaction information is also provided for back-testing purpose. Students are involved into the development of the software and get experience for R-programming. Preliminary version of the product has been tested and surveyed in a data science classroom. About ten junior Data Science students have practiced and provided feedback. The positive survey results show the feasibility of the approach. In the future, artificial intelligent component will be integrated into the platform.**

Keywords— Data Science; Trading Platform; R-Programming

I. INTRODUCTION

We are facing the age of data, enriched with big data, unstructured data, mixed media data, you name it. The academia and industry communities are requesting for a new class of experts who can extract actionable knowledge from rich and varied datasets. Huge amounts of data with complex structures in the form of text, video, and streaming data are routinely collected in social networks (e.g., Google, Facebook, Yahoo, Twitter), biological and health sciences (e.g., drug discovery, patient care), sciences and engineering (e.g., astronomy, networks, smart buildings), business and industry (e.g., automotive, robotics, banking, insurance, ad networks) as well as by government and society at large. Data engineers will help quantify and address the pressing concerns of modern society, including those in healthcare, sustainability, security, equity, and economics. However, undergraduate programs in the area of Data Science or Data Engineering are still few. And most existing programs are focusing on theories and algorithms, which are generally hosted by Mathematics or Statistics department. The software development and database management challenges are usually ignored by those schools.

Faculty at Gannon University, especially those in the Department of Computer and Information Science (CIS), have offered selective courses and have been active in scholarship in Data Science and Engineering. A series of data-project-based collaborations between CIS faculty and faculty from other departments, such as Criminal Justice, Physical Therapy and Environmental Engineering, has been occurring on campus. Due to the growing demands for graduates in the data area, a group of CIS faculty worked together to establish this pioneer educational program. The program focuses on data science (data analytics data modeling, machine learning, and etc.) and data engineering (cloud computing, database management, business intelligent systems, and etc.). The curriculum was developed by three factors: 1) the benchmark of other universities that have similar, but different programs; 2) the understanding of the development trend of big data in both academia and industry worlds; and 3) the reality of the faculty's expertise and scholarship focus within the CIS department.

During the enhancement of Data Science and Engineering program, one of the challenges in our actual teaching experiences is how to involve students into real-world data and inspire their interests. Therefore, a simulated trading platform is designed and developed for education goal.

II. ALGORITHIMIC TRADING AND RELATED TREND

As banks continue to update themselves and enter the modern era, there is a big need for profitable and reliable algorithms that can earn better returns than humans can do alone. The New York Times [1], Wall Street Journal [2] and many other media have noticed the power of applying data science and engineering into stock market. According to C. Metz, "Nytimes, (2017), artificial Intelligence - powered stock trading is an increasingly hot topic".

The idea of using computers to trade stocks is hardly new. Algorithmic trading has been around for well over a decade and rapidly gaining in popularity. Figure 1 demonstrates the historical trend of algorithmic trading vs percentages of market volume. Today, upwards of 90% of trading is conducted by computer programs. [4]

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on July 21,2020 at 19:02:54 UTC from IEEE Xplore. Restrictions apply.

Fig. 1. Trend of Algorithmic Trading.

Recently, more companies offer AI-driven investing for the average person. For a single investor with – tens of thousands of dollars to invest, they don't need advanced hedge fund technology – they just need some software that can allocate their funds into a simple and smart investment portfolio, and for them to consult as their needs, or the market changes [3].

III. Software and Data Design

A. Software Architecture

The simulation software (as Fig 2) contains three components: 1) Yahoo Finance server, which provides free and real-time stock data. 2) a local server, which is a server hosted in department. Its major role is to collect historical data and provide necessary computing for students. 3) user interface, which is the self-developed trading platform to demonstrate stocks information and to enable student to implement trading strategy. The local server and user interface are implemented based on the web framework in R, named Shiny.

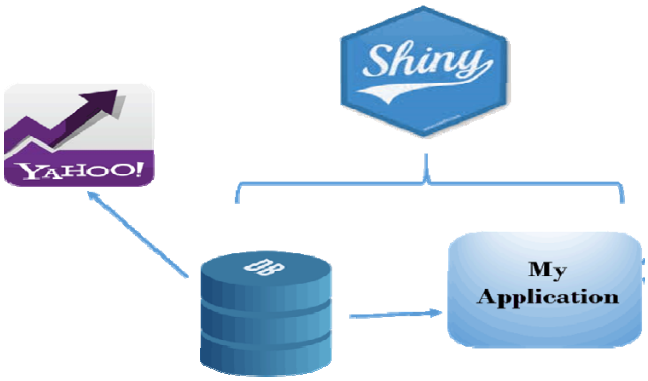

Fig. 2. Software Architecture

B. R Packages

In this project, R studio is used as development environment. The version of R used in this project is 3.3.2, which was released in October 2016. Some key R packages are listed as follows.

- Shiny: Open source R package that provides an elegant and powerful web framework for building web applications using R.
- Quantmod: The package designed to assist the quantitative trader in the development, testing, and deployment of statistically based trading models.
- Dygraphs: An R interface to the dygraphs JavaScript charting library. It provides rich facilities.
- Magrittr: The package providing a mechanism for chaining commands with a new forward-pipe operator, %>%. This operator will forward a value, or the result of an expression, into the next function call/expression. There is flexible support for the type of right-hand side expressions.

C. Picked Stocks

Due to the variation of stock types, only five different stocks are picked to represent the different sectors and industries. Reputed stocks are selected to inspire student interests.

- AAPL: Apple, Inc. engages in the design, manufacture, and marketing of mobile communication, media devices, personal computers, and portable digital music players. It represents the "Consumer Goods" sector in "Electronic Equipment" industry.
- CELG: Celgene Corp. is an integrated global biopharmaceutical company, which engages in the discovery, development and commercialization of therapies for the treatment of cancer and inflammatory diseases. It represents the "Healthcare" sector in "Biotechnology" industry.
- JPM: JPMorgan Chase & Co. operates as a financial services company worldwide. It represents the "Financial" sector in "Money Center Banks" industry.
- XOM: Exxon Mobil Corporation is engaged in energy business. The Company is engaged in the exploration, production, transportation and sale of crude oil and natural gas, and the manufacture, transportation and sale of petroleum products. It represents the "Basic Materials" sector in "Major Integrated Oil & Gas" industry.
- ABX: Barrick Gold Corporation (Barrick) is a gold mining company. It represents the "Basic Materials" sector in "Gold" industry.

IV. User Interface Design

Due to the educational goal of the project, the design of the trading platform is to fulfill two objectives: 1) to integrate enriched content for knowledge delivery, and 2) to be friendly to students, who may not be comfortable with busy data environment. Thus, four panels are implemented into the user interface.

A. Exploring panel

The exploring panel (as Figure 3) is designed to demonstrate real-time stock information. The content includes real-time or historical stock prices and multiple indicators, such as BBand, Volumn, RSI and CCI.

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on July 21,2020 at 19:02:54 UTC from IEEE Xplore. Restrictions apply.

Fig. 3. Exploring Panel

Further functionalities are provided as follows.

- Different Time Interval Selection: the time interval could be for 1 day, 1 week, 1 month, 3 months or 1 year. The default setting is for 3 months.
- Zoom In or Zoom Out: the user can zoom in or zoom out the plot easily by clicking and draging the area.
- Buy Points and Sell Points: trading transactions are also demonstrated on the plot. Buying transaction is shown as a green dot; and selling transaction is shown as a red dot.

B. *Control Panel*

Control panel (as Figure 4) is designed to enable students to pick fundamental parameter of the trading platform.

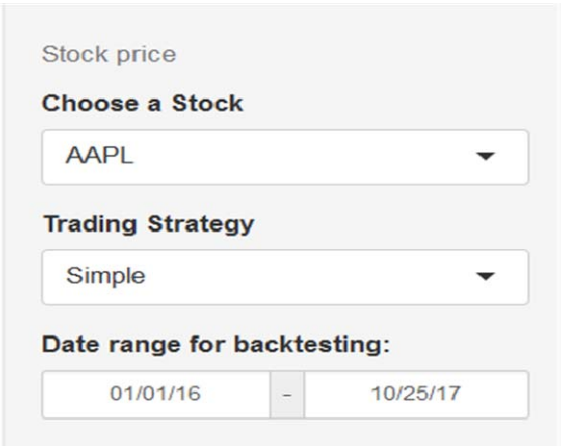

Fig. 4. Control Panel

The first drop-down menu allows student to pick one of the five stocks. The second drop-down menu is for student to select one trading strategy. There are three trading strategies. The third sub component is optional, which is the time range that the user wants to do back-testing. The default range is 1/1/2016 to current. However, user can extend the starting time to 1/1/2001.

C. *Reporting Panel*

Reporting panel (as Figure 5) displays the summary of back-testing results. It presents the time range of back-testing, and gives the number and amount of buy and sell transactions. The unrealized earning and total earning amount are also demonstrated.

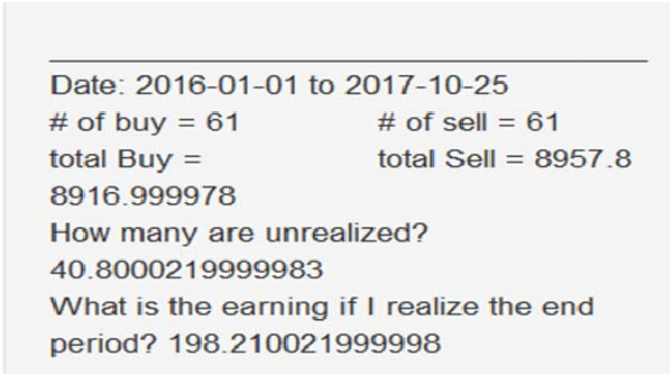

Fig. 5. Reporting Panel

Furthermore, a more detailed report is provided for further analysis during back-testing. Our system automatically generates a .txt file, which contains detail information of every transaction. The transaction time and buy/sell prices could be used to either verify the accuracy of strategy or to evolve into more comprehensive strategy later.

D. *Strategy Panel*

Strategy panel (as Figure 6) allows student select one of three trading strategies in a drop-down menu.

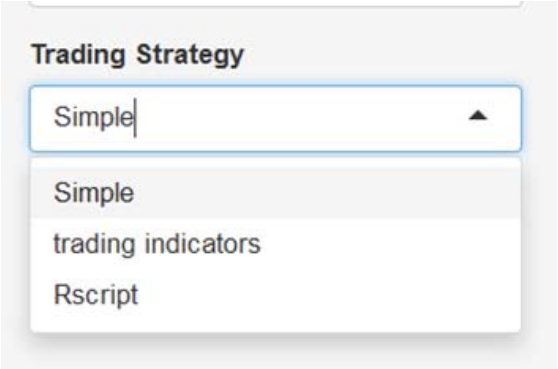

Fig. 6. Strategy Panel

The Simple Strategy provides a most intuitive way to trade in stock market. If today is an up day (close price higher than open price), buying order is set in the beginning of next trading day. If today is a down day (close price lower than open price), selling order is set in the beginning of next trading day. The strategy is for student without any experience of trading.

The Indicator Strategy is to simulate trading based on benchmarks of technical indicators. Students can build his/her own strategy by giving a value for each indicator. The selection condition is based on arithmetic operations, such as <, =, or >. Multiple condition statements can be connected by logical operations, such as AND or OR. In addition, a previous time slot (T-1) can be considered in the Indicator Strategy.

In order to provide the more flexibility of trading strategy, RScript Strategy is designed to allow self-developed trading methods. Students can build their own strategy as a R function and upload the code into the system. Even though a little limitation is required in format, the approach is for students with advanced experiences of trading and R-programming.

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on July 21,2020 at 19:02:54 UTC from IEEE Xplore. Restrictions apply.

V. User Survey

The preliminary software was tested in a "Statistical Computing" course. The course contains totally 10 students, 7 graduate-level and 3 undergraduate-level. All of students have entry-level experience of R-Programming. But most of them have limited experience with stock trading.

Two surveys were given to students and results were analyzed. The first survey is to introduce the software. Students learned how to run the system and observe related performance. The second survey focuses on the functionalities of the platform and related features. The trading strategies were emphasized. Eventually, students were able to develop their own trading strategies using R-programming. An in-class competition were performed based on earnings and the winner was award bonus credit.

Based on survey results, 100% of the students successfully ran the system and all of them experienced smooth installation and performance of the system. All the students agreed that the platform is a user-friendly interface (see Figure 7).

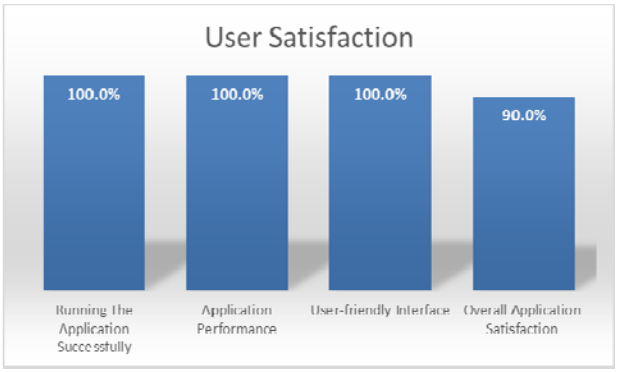

Fig. 7. Survey Results of User Satification

Regarding the functionality survey, 100% of students practice the price panel while 80% understand it. Exploring panel also has as the same statistics as price panel. 90% of them could try different stocks and see the related prices and graphs. 100% of them performed zooming in and out function inside the plot successfully. 100% of the students could change the time range and gets the data on that period (see Figure 8).

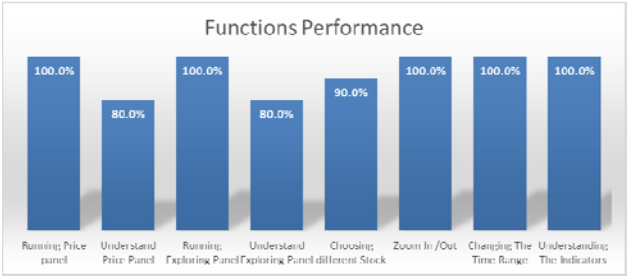

Fig. 8. Survey Results of Function Performance

During the learning of trading strategy, 100% students successfully practiced the simple trading strategy while 90% understood it. How to build a trading strategy using the indicators was tested by 90% of the students and the process was understood by 70% of them. 90% of them tested the R-Programming for building a trading strategy and 90% of them understood it and built their own strategies successfully. Finally, when asked about the user satisfaction, 90% of the students liked the overall system and satisfied about it (see Figure 9).

Fig. 9. Survey Results of Trading Strategy

VI. Conclusion

During the education experiment, a simulated trading platform was developed and deployed using RShiny. Related back testing system was also provided. The system connects YAHOO Finance server for real-time and historical stock information. Five classic stocks are selected for the diversity of the market. Three well-defined indicators are used for technical trading concept. Furthermore, three different trading strategies are designed and delivered to students. It targets students with different backgrounds in trading or programming. After adopting the project into classroom, students gave positive feedbacks on the software. It approves the feasibility of the approach.

In future, more functionalities and indicators will be integrated into the system. Further data analysis or data science components will be emphasized and added. Artificial intelligent system is considered as another big practice. Due to the lack of similar free systems, we can consider this system has big potential as a the base of an education software for data science and financial engineering.

Acknowledgment

Special thanks for Ms. Eman Alghamdi's devoted contribution to this project. She was involved into programming and student survey. Part of the results has been Ms Alghamdi's Master Thesis. Also, thanks MS. Omniyah Y. Alansari for her helping during student survey.

References

[1] C. METZ, "NYTimes," 22 Oct 2017. [Online]. Available: https://www.nytimes.com/2017/10/22/technology/artificial-intelligence-experts-salaries.html. [Accessed 4 Nov 2017].

[2] I. Wladawsky-Berger, "The Wall Street Journal," 15 Sep 2017. [Online]. Available: https://blogs.wsj.com/cio/2017/09/15/artificial-intelligence-is-ready-for-business-are-businesses-ready-for-ai/. [Accessed 4 Nov 2017].

[3] A.Semeney, Jan 2017. [Online]. Available: https://www.devteam.space/blog/artificial-intelligence-in-stock-trading-future-trends/. [Accessed 1 Nov 2017].

[4] "Machine Learning For Stock Trading Strategies," 14 Apr 2016. [Online]. Available: https://www.nanalyze.com/2016/04/machine-learning-for-stock-trading-strategies/. [Accessed 7 Nov 2017].

[5] S. Greg Walters, 22 Mar 2017. [Online]. Available: https://www.livescience.com/58364-ai-investors-rack-up-massive-returns-in-stock-market-study.html. [Accessed 12 Oct 2017].

Authorized licensed use limited to: UNIVERSITY OF BIRMINGHAM. Downloaded on July 21,2020 at 19:02:54 UTC from IEEE Xplore. Restrictions apply.