Data Analytics and Business Intelligence Framework for Stock Market Trading

Batool AlArmouty
Computer Science Department
King Hussein School of Computing Sciences
Princess Sumaya University for Technology
Amman, Jordan
armoutib@gmail.com,

Salam Fraihat
Computer Science Department
King Hussein School of Computing Sciences
Princess Sumaya University for Technology
Amman, Jordan
s.fraihat@psut.edu.jo

Abstract*— Business intelligence is an umbrella term that combines architectures, tools, databases, analytical tools, applications, and methodologies. The efficiency of making decisions can increase significantly using business intelligence solutions, by taking advantage of the existing historical or real-time data of the business. Trading in stock markets is imminent with taking risks of losing money, which requires extensive experience in the market, to make efficient decisions. In this paper, we propose a framework that makes use of stock prices historical data, to help investors in making more efficient trading decisions.

Keywords*— Business Intelligence, Business Analytics, Data Analytics, Decision Support Systems, Stock Market.

I. INTRODUCTION

Business Intelligence (BI) is defined as a collective term that combines different technologies, applications, and tools used for the gathering of data from sources, storing, analyzing and visualizing it, with the purpose of helping users to make better decisions [1]. In the last few years, data has been increasing rapidly, and with the ease of acquiring and storing this data, organizations have started to leverage it for enhancing the decision making the process.

BI's objective is transforming data to information through analysis to meet the business objective of the user [2], by enabling the user to interactively manipulate the data, and apply different analysis in the way she/he needs for extracting information, and get valuable insights from the data.

The stock market is a public market with strict regulation for the trading of companies' stocks, where each stock is a share of company permitted to be traded called listed company ownership. Investors make money by buying stocks at a lower price while selling them at a higher price, stocks prices are determined by the success of the company, supply and demand, and external factors like government regulations.

Investors take risks in determining the best time for selling or buying stocks, that's why they need an efficient help to reduce the risks of this decision. In this paper, a business intelligence framework is proposed using the historical stock market data, with the objective of analyzing stock market attributes for a collection of companies, and enhance the efficiency of choosing the appropriate time for buying or selling specific company stocks, in order to reduce the risk of losing money.

The proposed framework covers the conversion of the data to useful information satisfying the business objective. The framework contains the processes of acquiring the data from its source, transformations applied on the data, storing the data into an appropriate data warehouse, the analytics applied on the data after the transformation to extract information, and finally the presentation of the extracted information to the end-user, most likely the decision-maker.

The rest of the paper is structured as follows, section II presents the related work, section III explains the requirements of the framework, the proposed architectures in section IV, the design in section V, and business intelligence presentation in section VI, finally the conclusion in section VII.

II. RELATED WORK

A. Stock Market Analysis

Umadevi et al [3] applied analytical techniques on stock market data and tried to design to a prediction model. The authors obtained Google, Apple and Microsoft stock prices over six months, with four attributes (low, high, open and close). The analysis applied on the stock market data involves stock scores and candlelight plot to visualize all the parameters.

Alraddadi [4] made analysis using the stock prices data of John Wiley & Sons company over one year, the data contains six attributes (open, high, low, closing, and adjusted close). The author applied descriptive statistics to explore the nature of the data, and analysis measures including measures of central tendency, and measures of variability. Moreover, they made use of plots to fully understand the nature of the data, like histogram and time series plots.

Sen et al [5] made analysis on the Indian stock market, by decomposition the time series data into three components; the trend, the seasonal component, and the random component. The decomposition was done to help understand whether the buys are short-term or long-term, and discover the pattern of the stocks trading. Based on this analysis, the months in which the seasonal component plays a major role were discovered, and have an idea about the trends of the stocks. Moreover, the decomposition results were used to forecast the values for 12 months.

Bhoopathi et al [6], proposed a framework to discover the trends in stock trading by finding casual relationships in stock dataset, in the form of direct, indirect, and exception association rules in the stock dataset, the framework also considers the events and government decision that may influence the stock trading.

B. Business Intelligence

Martin et al [7] proposed a business intelligence framework consist of Quantitative bankruptcy prediction components, where financial features found using Genetic

978-1-7281-2882-5/19/$31.00 ©2019 IEEE

Algorithm are applied to predict the business performance quantitatively, Qualitative Bankruptcy prediction components, in which the features are found using expert analysis and predict the business performance qualitatively using Ant Miner Algorithm and a Customized reporting where the right information is delivered to the right user in the requested presentation using Fuzzy Multi-Criteria Decision Support System (FMCDS).

Jadi et al [8] suggested a framework for collecting data, as the first step of implementing an e-government business intelligence system. For the purpose of taking benefits of the immense data collected in enhancing decision making and effectively enhance public services, the authors took the morocco e-government system as a case study for collecting the data. They suggested three sources of data: First, government-to-government, where they suggest enabling the interaction between government departments' databases. Second, government-to-business, this source of data depends on the organization's way of storing data, instead of storing each government organization separately, store them in one database. Third, government-to-citizen, the authors proposed an approach for collecting information related to citizen interaction with government organizations.

Khedr et al [9] proposed a framework of a business intelligence system for healthcare analytics. The framework contains six tiers: First, Data source tier, in this tier, twelve different sources are proposed to be used in the business intelligence system. Second, Extract Transform and Load (ETL), the data extracted from the data sources are integrated and transformed in the staging area, to ensure the data quality. Third, Data storage tier, this tier composed of two components, data warehouse and three types of data mart (Operational data mart, Medical claims data mart, and financial data mart). Fourth, Analytics tier, in this tier, the authors suggested applying diagnostic, descriptive and predictive analytics on the data. Fifth, Optimization tier, the results obtained from the analytics tier is modified in this tier. Finally, Presentation tier, in this tier the result of the system is presented visually to the user, which makes it easier to make decisions.

Olexova [10] presented a case study by applying BI in the retail chain, the study was conducted in a sports-fashion multi-brand chain of retail stores. In this study, the BI life cycle was analyzed, besides evaluating the factors impacting the BI adoption. The main findings of this study are considering the most important benefit of BI adoption is improving decision-making in both speed and quality, and according to the managers, the customization of the BI system is the more important factor for a successful BI adoption.

III. REQUIREMENTS

The system requirements describe the capabilities and functions that will be implemented to satisfy the user's business objectives [11]. Requirements are supposed to define what the system is supposed to do and to be, which distinguishes each system from its competitors, they range from high-level requirements defining the system performance to the very basic functionalities that will be needed in the system. Defining high-quality requirements is considered a critical phase in the Business Intelligence system developments, because of their importance in developing effective systems, and detecting errors in early stages of the development process [12]. In the proposed framework, the requirements are divided as follows:

A. Business Requirements

- Show each company's stock price and quantity trend over time.
- Compare companies' trends.
- Predict the better investment in terms of expected profit.
- Dashboard to perform visual analytics and prediction.

B. Data Requirements

1- **Data Source:** The data can be collected from Google Finance.

2- **Data Acquisition:** Historical data of four companies' stock prices will be collected from the source and saved in a place for preparations.

Company names: Apple, Nike, Disney, Microsoft.

3- Data Transformation: the collected data should be prepared for the analysis before storing it, the preparation consists of:

- **Data Quality:** Data must be assured to have good quality, which is achieved when the data embodies the "Five Cs": clean, consistent, conformed, current, and comprehensive.
  The data contains missing dates, these dates are weekends and holidays, in which the market closes, therefore we will not fill them.
- **Feature Selection:** Data must be reduced to contain only the features that will be used for the analysis and prediction.

4- **Data Storage:** Data must be stored to be ready for being viewed or used for reports and analytics anywhere and at any time.

C. Functional Requirements

- The system should allow the user to view the trend of each company for a period chosen by the user (year, season, or month).
- The system should allow the user to show which is bigger, the opening or the close for each day.
- The system should allow the user to show the trend for more than one company in the same Fig, companies chosen by the user.
- The system should allow the user to view basic descriptive statistics (maximum, minimum, first, second, and third quantiles) for each attribute of the company.
- The system should allow the user to show the distribution for each company for specific years or months.
- The system should allow the user to compute basic statistics (Mean, Standard Deviation, and Range) for each company in a specific year.
- The system should allow the user to compute basic statistics for each company from the start of the current year.

- The system should provide a recommendation to a user based on the prediction of the best action to take (strongly sell, sell, neutral, buy, strongly buy).

D. Non-functional Requirements

- **Interactive Visualization:** User should be able to interact directly with the BI application to display the wanted companies, plot, and metric she/he wants.
- **Performance:** the dashboard must be fast when changing companies or functionalities, quick responses for the user requests, and use a minimum memory.
- **User-friendly:** User should not have difficulties when interacting with the dashboard.
- **Portability:** User should be able to access the system using (Windows, Linux, Mac, Android, IOS) operating systems, installed on any hardware.
- **Reliability:** System should be tested to determine the probability of failure, and ensure that the system can handle these failures without disturbing the user.
- **Availability:** User should be able to access the system anytime and anywhere.
- **Scalability:** The system should have the ability to improve, by adding new functionality, or companies without disturbing existing activities.

E. Technical Requirements

- **ETL** (Extract, Transform, Load) Data should be extracted from the data sources, transformed in the proper form, then will be stored in a data warehouse.
- **Data Acquisition:** Data should be extracted from different sources e.g. the internet, DB server, Excel File.
- **Staging:** Data will be temporary stored between the data source and the data warehouse, to be transformed and analyzed before storing in the Data warehouse.
- **Data Warehouse:** Transformed data will be stored after the staging in a structured form, where the information will be retrieved for applying the requested analytics.
- **OLAP** (Online Analytics Processing) used for analyzing the data using the Multidimensional model.
- **Dashboard:** is the business intelligence application, where results of the analytics applied on the data, will be displayed to the user and allow the interaction with the dashboard.

IV. Architecture Framework

The proposed business intelligence system is represented through a set of architectures identifying system components and the relations between them [13]. The system requirements defined earlier are interpreted into structures to meet the business objectives. In our business intelligence framework, we provide four types of architectures:

A. Information Architecture

Defines the processes needed to transform the data obtained from the source, to readable information in the BI system (Fig 1).

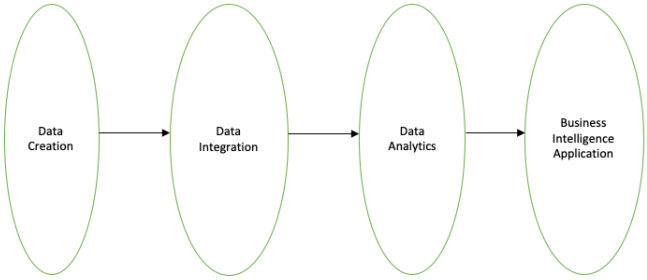

Fig. 1. Information Architecture.

1- **Data Creation:** This process is done by the data source, where the data is created.
2- **Data Integration:** This process is done after collecting the data from sources, where the data is integrated, cleaned, and ensure the quality of the data.
3- **Data Analytics:** Analytics is applied to the data after storing it in the data warehouse.
4- **Business Intelligence Application:** The results of the analytics will be represented to the decision-makers.

B. Data Architecture

Define the processes of data integration, transformation, storage and workflow needed to meet the requirements of the information architecture (see Fig 2).

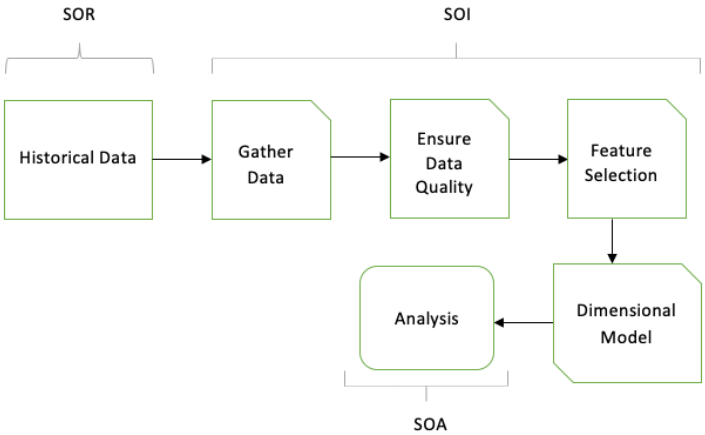

Fig. 2. Data Architecture.

1- **System of Record (SOR):** The authorized data source for a particular data that can be found in multiple sources. In this framework the data type that will be used is historical data, the historical data contains the stock price at the opening of the market, the price at the closing of the market, the highest and the lowest prices of the stock at the day, and the volume of the traded stocks.
2- **System of Integration (SOI):** The process of integrating the data. After the data gathering, techniques to ensure the quality of the data, like detect

missing data, and duplication will be applied, then the relevant features will be selected before storing the data in the data warehouse.

3- **System of Analytics (SOA):** The discovery of meaningful patterns in the data, the requested analysis will be applied in this process.

C. *Technical Architecture*

Fig.3 illustrate the techniques that will be used in implementing the business intelligence system.

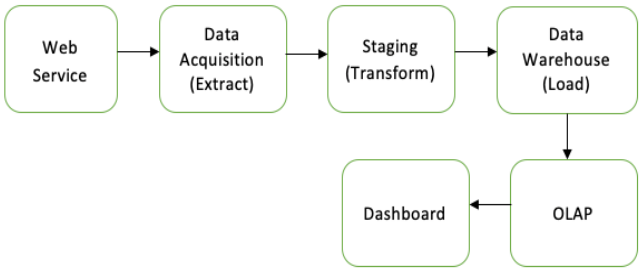

Fig. 3. Technical Architecture.

The required data is available on the internet, therefore it will be extracted, transformed, and then loaded in the data warehouse before OLAP can be applied on the data, finally, the data will be displayed to the user using the dashboard.

D. *Product Architecture*

In Fig.4, we define the products that could be used in implementing the techniques required for the business intelligence system.

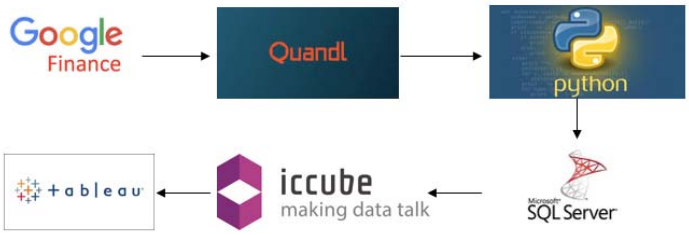

Fig. 4. Product Architecture.

The historical stock market data is provided by google finance[1] for free and are extracted using python Quandt library[2], using this library the number of years' historical data can be determined, and then the transforming stage can be done using python. Then the data warehouse can be build using sql server[3], OLAP can be done using IcCube[4], and finally, the dashboard can be implemented using Tableau[5].

[1] www.google.com/finance
[2] www.quandl.com/tools/python
[3] www.microsoft.com/en-us/sql-server/sql-server-2017
[4] www.iccube.com/
[5] www.tableau.com/

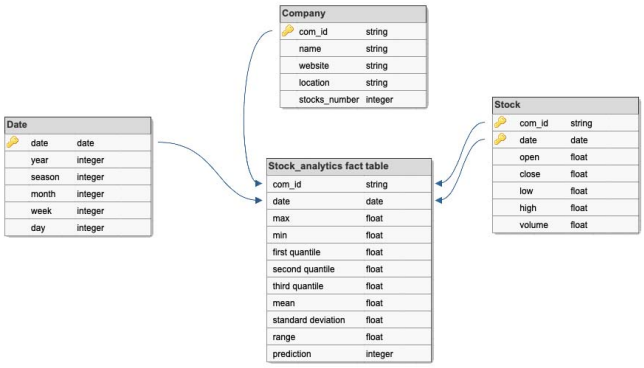

Fig 5. Data Schema.

V. Design

The proposed schema is provided in Fig 5. In this framework we used the dimensional model, with three dimensions for company, stock and date.

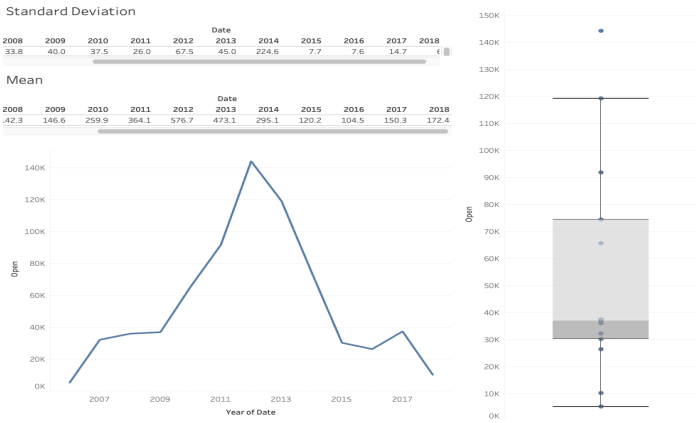

Fig 6. Dashboard prototype.

VI. Business Intelligence Application

The information extracted from the data must be presented to the end-user in an understandable way, using an application of BI to display the requested metrics, the application is implemented to help monitoring the progress of the business, and embower making efficient decisions to improve the business.

In the proposed framework, we suggest using a dashboard, which is a data visualization tool that displays several metrics on the same page, allowing the user to compare the results of different metrics. For the obtained data, the dashboard will contain companies' different attributes plots; a line plot to show the trend of the company stocks, and boxplot to show the quantiles and make it easier to spot outliers easily, in addition to the presentation of the standard deviation, and mean for each year. The dashboard will also have the ability to compare more than one company, and year together in the same plot. A prototype of the proposed dashboard is presented in Fig 6.

VII. Conclusion

Business Intelligence plays an important role in the success and survival of the business, nowadays it became easier to apply the business intelligence, because of the easiness of collecting data using the internet. In this work, we proposed a framework that makes use of the historical

stock market of companies, to help the investors in making future trading decisions. The framework proposed the techniques and tools to collect data, transform, store, analyze and present them to the end-user, in our case the investor.

REFERENCES

[1] Azeroual O., Theel H., "The Effects of Using Business Intelligence Systems on an Excellence Management and Decision-Making Process by Start-Up Companies: A Case Study", International Journal of Management Science and Business Administration, 2018, pp.30-40.

[2] Chang V., Larson D., "A Review and Future Direction of Agile, Business Intelligence, Analytics and Data Science", International Journal of Information Management, 2016, pp.700-710.

[3] Gaonka A., Kulkarni R. et al, "Analysis of Stock Market using Streaming Data Framework", International Conference on Advances in Computing, Communications and Informatics, 2018, pp.1388-1390.

[4] Alraddadi R., "Statistical Analysis of Stock Prices in John Wiley & Sons", Journal of Emerging Trends in Computing and Information Sciences, 2015, pp. 38-47.

[5] Sen J. and Chaudhuri T., "A framework for Predictive Analysis of Stock Market Indices – A Study of the Indian Auto Sector", arXiv, 2015, pp. 1-19.

[6] Bhoopathi H. and Rama B., "A Novel Framework for Stock Trading Analysis Using Casual Relationship Mining", 2017, International Conference on Advances in Electrical, Electronics, Information, Communication, and Bio-Informatics (AEEICB), pp. 1-6.

[7] Martin A., Lakshmi T., and Venkatesan V., "A Business Intelligence Framework for Business Performance using Data Mining Techniques", International Conference on Emerging Trends in Science, Engineering and Technologies, 2012, pp. 373-380.

[8] Jadi Y., Lin J., "An Implementation Framework of Business Intelligence in e-government systems for developing countries: Case study: Morocco e-government system", International Conference on Information Society, 2017, pp.138-142.

[9] khedr A., Kholeif S. et al, "An Integrated Business Intelligence Framework for Healthcare Analytics", International Journal of Advanced Research in Computer Science and Software Engineering, 2017, pp. 263-270.

[10] Olexova C., "Business Intelligence Adoption: A Case Study in the Retail Chain", WSEAS Transactions on Business and Economics, 2014, pp. 95-106.

[11] Bahill, A.T. and Dean, F.F. 2009. Discovering system requirements. Handbook of Systems Engineering and Management. A.P. Sage and W.B. Rouse, eds. John Wiley & Sons. 205–266.

[12] Chakraborty A., Baowaly M. et al, "The Role of Requirement Engineering in Software Development Life Cycle", Journal of Emerging Trends in Computing and Information Sciences, 2012, pp.723-729.

[13] Bass L., Clements P. et al, "Software Architecture in Practice", Second Edition, Chapter 1, Addison-Weley,2003.