International Journal of Advanced Smart Convergence Vol.5 No4 32-39 (2016)
https://doi.org/10.7236/IJASC.2016.5.4.32

IJASC 16-4-6

An Effective Data Model for Forecasting and Analyzing Securities Data

Seung Ho Lee[1*], Seung Jung Shin[2]

[1]*Graduate School of HanSei University*
[2]*Dept. of IT convergence, Hansei University*
[*]*shoo2222@naver.com, expersin@ gmail.com*

Abstract

Machine learning is a field of artificial intelligence (AI), and a technology that collects, forecasts, and analyzes securities data is developed upon machine learning. The difference between using machine learning and not using machine learning is that machine learning—seems similar to big data—studies and collects data by itself which big data cannot do. Machine learning can be utilized, for example, to recognize a certain pattern of an object and find a criminal or a vehicle used in a crime. To achieve similar intelligent tasks, data must be more effectively collected than before. In this paper, we propose a method of effectively collecting data.

Keywords: *EDATE date, ETRADING volume, ECLOSEPRICE closing price*

1. Introduction

1.1 The Reason We Need Machine Learning for Securities Data Prediction

The greatest advantage of using securities data collected for securities data prediction and analysis is to predict an investing item's outlook which allows to minimize a loss and maximize a profit. After learning a certain pattern based on data analysis of securities items and sectors, an investor can utilize the information to predict and analyze securities when a related event occurs. A forecast and an analysis of securities data can bring very crucial data to predict a percentage of fall and rise when an investor tries to buy or sell certain securities. For this research, we used past data of last three years to forecast and analyze securities data.

Russian born American economist Simon Kuznets is a good example of machine learning for data forecast and analysis. He collected all data from past three hundred years in order to develop a model for measuring national income statistics. More accurate forecasts and analysis can be made if past data covers a wider range of older periods, but the data for past data analysis is beyond count, and the volume of it is too large. Big data analysis (machine learning) is used in such cases.

Manuscript Received: Sep. 30, 2016 / Revised: Oct. 12, 2016 / Accepted: Oct. 25, 2016
Corresponding Author: shoo2222@naver.com
Tel: + 82-031-450-5274
Graduate School of HanSei University

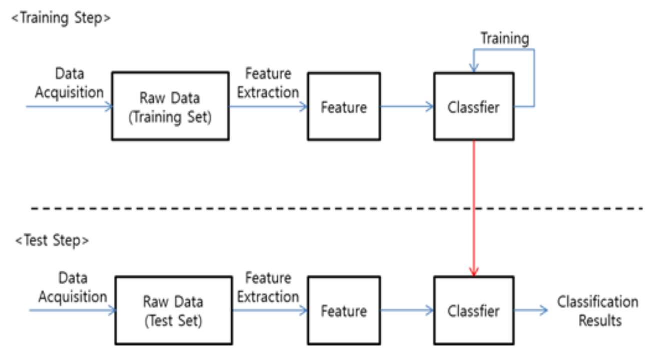

Figure 1. Conceptual Map of Machine Learning

Figure 1 represents a simple conceptual map of machine learning. Machine learning is divided into two steps: 1) training step, where a classifier is trained using prepared data, and 2) test step, where a result of input data is reviewed using a trained classifier.

How can we proceed data analysis using machine learning like this? First, we need to derive a singularity of data based on past data.

After calculating a real rate of stock price fluctuation, we conduct an analysis of times of changes and changes in stock price influenced by a change of system or other factors.

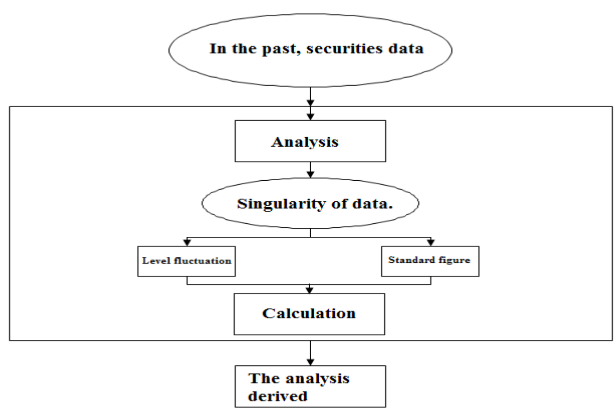

Figure 2. Conceptual Map of Securities Data Analysis

Real rate of stock price fluctuation is a rate of stock price fluctuation derived from only considering a market price trend regarding market supply and demand over a certain period.

A calculation of real rate of stock price fluctuation is

(fluctuation value – standard value) / standard value * 100,

and with this calculation, we can get a rate of fluctuation.

1.2 Drawback of Securities Data Collection

A drawback of securities data collection comes from a question of how to extract data required for an analysis. To acquire securities data, many people use APIs or web crawling, but most of tools for web crawling are overseas API such as Google and Yahoo. Using this method makes data collection and analysis difficult due to time differences. This causes a large volume of duplicate data and a pile of threads on a server which eventually stops a system from a server overload.

Even though a domestic API exists, it can still be a problem if we use only one API. If that API becomes suspended or terminated due to some reasons such as an inspection, then a forecast and analysis service for securities data collection will be terminated as well.

1.3. Solution of Securities Data Collection

Some methods for securities data collection are:

1) API of Kiwoom Securities

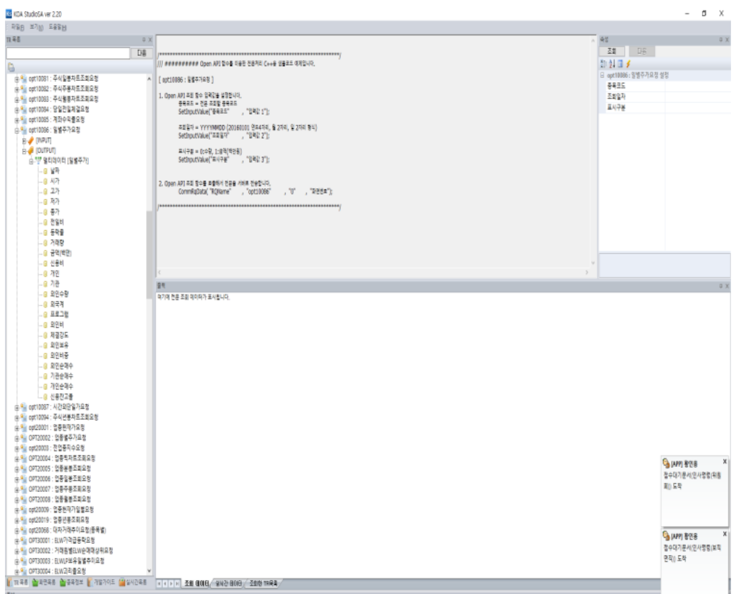

Figure 3. API Screen Shot of Kiwoom Hero S

2) R Code web crawling of Google

```
+
+   #### downloading data from the Google Finance ####
+
+   output <- as.data.frame(readHTMLTable(url,stringsAsFactors=FALSE)[4])
+   output[,1] <- as.Date(output[,1],format="%b %d, %Y")
+   names(output) <- c("Date",paste(ticker,c("Open","High","Low","Close","Volume"),sep="."))
+   for (k in 2:6){
+       output[,k] <- as.numeric(gsub(",","",output[,k]))
+   }
+
+   #### Checking whether all data are retrieved or not ####
+
+   end <- as.Date(output[length(output[,1]),1])
+   ll <- 200-length(output[,1])
+   if ((end-ll)<start_date) end <- start_date
+   if (end > start_date) {
+       output <- rbind(output,g.hist(ticker,start_date,as.character(end-1)))
+   }
+   return(output)
+ }
> g.hist("KOSDAQ:035720","2014-01-02","2015-07-11")
         Date KOSDAQ:035720.Open KOSDAQ:035720.High KOSDAQ:035720.Low KOSDAQ:035720.Close
1   2015-07-10             131500             135200            128500              130500
2   2015-07-09             136200             137400            127500              131300
3   2015-07-08             141200             141900            137000              138100
4   2015-07-07             142000             143400            133000              138400
5   2015-07-06             137500             145200            137100              138000
6   2015-07-03             138700             139700            134600              139200
7   2015-07-02             139000             141500            135900              138300
8   2015-07-01             128300             139600            128100              136600
9   2015-06-30             128000             128700            122500              126200
10  2015-06-29             121400             128500            121400              126000
11  2015-06-26             120500             126200            119000              123500
12  2015-06-25             118000             120400            117100              119800
13  2015-06-24             107300             115500            107300              114300
14  2015-06-23             108500             108500            106600              107000
15  2015-06-22             108200             109400            106100              106600
16  2015-06-19             108200             109300            106500              107800
17  2015-06-18             107000             107100            104400              105000
```

Figure 4. R Code Web Crawling Screen Shot of Google

3) data collection web crawling using python of Daum Securities

```
<title>Daum</title>
<meta name="verify-v1“
content="iXgHW7UooMcyeiV/Zb0Tk/yK2yB+IuA5/5GgIzGBEns=" >
<meta property="og:image“
content="http://i1.daumcdn.net/img-media/mobile/meta/finance.png"/>
<link rel="stylesheet“
href="http://s1.daumcdn.net/stock/css/common.css?ver=201602202142
49" type="text/css">
<link rel="stylesheet“
href="http://s1.daumcdn.net/stock/css/newHeader.css?ver=2016022021
4249" type="text/css">
<link rel="stylesheet“
href="http://s1.daumcdn.net/stock/css/column.css?ver=2016022021424
9" type="text/css">
<link rel="stylesheet“
href="http://s1.daumcdn.net/stock/css/trade.css?ver=20160220214249"
type="text/css">
<script type="text/javascript“
src="http://s1.daumcdn.net/stock/js/common.js?ve
```

Figure 5. Web Crawling Screen Shot of Daum Securities

As we see, there are many ways to collect securities data; but in case of web crawling, it could be mistaken for D-dos even after completing a code and successfully launching. It is an era where data is generated constantly and limitlessly, and extracting all data overloads a server which causes a misidentification a right result for D-dos. If one uses an API, the one can utilize everything the company provides and the API supports.

Especially, API has almost no possibility of being mistaken for D-dos like web crawling and provides many functions that reduces code design and a burden on a server; it is possible to more effectively design codes.

With these reasons, we considered that using an API provided by Kiwoom Securities would be the most appropriate way to extract securities data.

2. Securities Data Extraction Using API

2.1. Data Extraction using API of Kiwoom Securities

Open API of Kiwoom Securities helps an investment strategy programmed by a user get connected to a module provided by Kiwoom and allows a user to check a market price and an account balance and make a purchase and so on. In this paper, we try to make it possible to view a market price and extract corresponding data using Open API.

Kiwoom Securities offers TR codes as following.

Figure 6. List of TR Codes of Kiwoom Securities

One can view the data after finding a list that one wants to extract and writing code.

```
Const stGRID lstOPT10081 [] =
{
{"Current", "20", -1, 0, DT_ZERO_NUMBER, FALSE,
DT_CENTER, "", ""}
{"Volume", "10", -1, 1, DT_ZERO_NUMBER, TRUE, DT_RIGHT,
"", ""}
{"Transaction Value", "25", -1, 2, DT_ZERO_NUMBER, TRUE,
DT_CENTER, "", ""}
{"Date", "11", -1, 3, DT_DATE, TRUE, DT_RIGHT, "", ""}
{"Cigar", "13", -1, 4, DT_ZERO_NUMBER, FALSE, DT_RIGHT,
"", ""}
{"High", "14", -1, 5, DT_ZERO_NUMBER, FALSE, DT_RIGHT, "",
""}
{"Low", "14", -1, 6, DT_ZERO_NUMBER, FALSE, DT_RIGHT, "",
""}
{"Modify stock quotes", "14", -1, 7, DT ZERO NUMBER, FALSE,
```

Figure 7. C++ Code Design of the TR Code List

2.2. Difference Between Web Crawling and API

We should know why API is used for more effective crawling first. Web crawling begins from a URL called seed and recognizes a link of website and renews a URL. Since web crawling recursively asks for a renewed URL, it is easy to be mistaken for D-dos. And to extract securities data for data of wanted items, web crawling has to go find a value of every developer mode of the Explorer to get. So it takes a lot of time writing codes, and a code gets longer because of that.

When one uses an API, however, one can enjoy all items offered by an API. API has codes of user's wanted items, and a user just simply uses them.

```
var list = {
 "a" = "events"
 "b" = "trading volume"
     ....
}
```

Figure 8. C++ Code Design of Web Crawling

Web crawling has to write a list of every wanted item as shown on the above.

```
{"trading volume", − 1 − 1, 1, dt true, number, _ _ zero dt right, "" and ""}
```

Figure 9. C++ Code Design of API of Kiwoom

When extracting data using API, one can put all data of wanted items in one line of code and request the data. In this way, using API instead of web crawling allows one to concisely design a code. Also, API can minimize a dangerous situation where a server gets overloaded. During the extraction process of web crawling and API, the speeds of I/O of DB were not very distinguishable. However, in case of web crawling, it only provides past data within a month, not over a year. Noticeable differences of API and web crawling are the simplification of code and the extraction of past data.

Therefore, in this paper, we will use a more effective extraction method API to extract securities data.

3. Expected Effect

In a stock market, one has to make a decision on when to sell or buy and collects a lot of securities data for forecasting and analyzing of the market to decide an investment strategy. However, individual investors depend on uncertain information and gain comparably lower profits than other investors. A tool for securities data extraction prosed in this paper allows an individual to manage stock price data and create a stock price forecast model based on technical data using API. In other words, among many crawlings used in various fields, a user can utilize API of Kiwoom Securities to build one's own stock price forecast model. A stock price forecast model suggests a stock expected to rise at a certain time and helps a user to decide items to sell or buy. The model eventually enables a user to judge a certain situation objectively to avoid risks and safely earn profits based on securities data. A proposed tool for securities data extraction let the user practice all steps of wanted data and develop securities data regarding one's investment preference. Therefore, a securities data model proposed in this paper will help stock investors make decisions on many parts that must be done in a stock market.

4. Test and Result

The study suggested web crawling and a tool to do web crawling for securities data extraction. This method develops a mechanism of securities data extraction using a tool, effectively manages individual stock data, and helps an individual investor to make an investment. Big data has been in use very widely, and securities data will be converted into big data and used as widely as big data for data extraction. Investors can create data of wanted items and verify the wanted items using past data. This mechanism enables past data extraction to support a decision making process during an actual trade based on data of proper prices of buy, sell, trading volume, and closing price.

LG		
EDATE	ETRADING	ECLOSE PRICE
20161004	-64300	288780
20160930	64700	251040
20160929	-64300	193396
20160928	-65000	180716
20160927	65400	173744

Figure 10. Some Extracted Securities Data

Table on the above represents the values of past data extraction of date, trading volume, and closing price of LG.

As a result, using web crawling with API is more effective for securities data extraction for machine learning than only using web crawling, and it helps investors to determine their investment strategies based on extracted data. Following that, investors are expected to aim for more profits and less loss using a securities data extraction model.

5. Conclusion

We studied a data extraction model for helping investors more effectively set their investment strategies via machine learning.

To collect securities data for machine learning, we need to use a crawling method. And comparing web crawling and web crawling using API, we found that using API is more effective which does not overload a server and save time thanks to code simplification. Using API, past data of last three years API was extracted and organized in a table for machine learning; this mechanism is expected to be very helpful for investors to design their investment strategies. A data collection model is necessary for investors to forecast and analyze which item will go up or down by what percent; therefore, based on the research result in this paper, we confidently expect that more accurate forecast and analysis can be conducted using API.

References

[1] Optimal Traffic Signal, Traffic Signal Cycle, Neural Network, Data Mining, Algorithm, Decision Tree Algorithm 2007-03-05

[2] Cross-lingual Tet Retrieval, Multilingual Information Retrieval, Thesaurus, Information Retrieval, Knowledge Base 2008-01-08

[3] Cyber University, Fuzzy rules, Rfid tag, IRT Theory 2010-01-19
[4] Machine Learning Techniques for Stock Prediction New York University, 2007
[5] Forecasting the volatility of KOSPI 200 using data mining 2008
[6] KOSPI directivity forecasting by time series model 2009
[7] Stock-Index Invest Model Using News Big Data Opinion Mining 2012
[8] Analysis of trading performance on intelligent trading system for directional trading 2011