PYTHON AND WEB SCRAPER FOR STOCK MARKET ANALYSIS

KODAVATI MOHAN KARUNA KUMARI[1], S D SHIVAPRIYA[2], BUCHI AVINASH[3],
DR. K. SRUJAN RAJU[4], DR. B. KRISHNA[5]

Student, Department of Computer Science and Engineering, CMR Technical Campus, Medchal, Hyderabad, Telangana, India[1,2,3]

Professor, Department of Computer Science and Engineering, CMR Technical Campus, Medchal, Hyderabad, Telangana, India[4]

Professor, Department of Computer Science and Engineering, CMR Technical Campus, Medchal, Hyderabad, Telangana, India[5]

ABSTRACT

Web scraping is a term used to describe the use of a program or algorithm to extract and process large amounts of data from the web. Whether you are a data scientist, engineer, or anybody who analyzes large amounts of datasets, the ability to scrape data from the web is a useful skill to have. Let's say you find data from the web, and there is no direct way to download it, web scraping using Python is a skill you can use to extract the data into a useful form that can be imported. Data extraction from the web using Python's Beautiful Soup module. Data manipulation and cleaning using Python's Pandas library. Data visualization using Python's Matplotlib library. This project is a application where users can give their desired company names for stocks information to know the best company to invest their money in stocks.

1.INTRODUCTION

Stock Market Analysis : Use high quality financial web data to make profitable decisions. This project is a application where users can give their desired company names for stocks information to know the best company to invest their money in stocks. We use web scraper that is able to access and extract data from websites using a web application as an interface for user interaction. The web scraper will extract stocks of various companies from their respective websites. Then compares them based on different parameters and suggests the user the best stock to invest their money in. This project will help the users who are investing in stocks take more efficient decisions with accurate comparison results provided. The time taken in decision making is also reduced by a great extent. The project will not only provide comparison results but it also graphically shows the variations and deviations in the stock prices being compared. It is user friendly and faster than the traditional way of researching for investing in stocks.

2.LITERATURE REVIEW
EXISTING SYSTEM

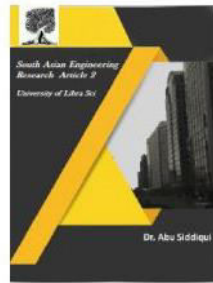If a user needs to collect information from the web today, without web scraping, it will

be a very long and repetitive process. The current system prior to Web scraping is the user has to open the browser, enter the keywords required, visit a website and then search through the website for a specific set of information.

The process of going into each and every website and searching for the stock manually is a time taking process. Even though there are multiple websites that show a comparison of various company stocks, these websites are not standardized or approved by anybody therefore it is not always advisable to trust these comparisons blindly. Additional to this, these websites only compare a set of companies and not the ones the user specifically wants.

PROPOSED SYSTEM

The proposed system for this project is a application where users can give their desired company names for stocks information to know the best company to invest their money in stocks. We use web scraper that is able to access and extract data from websites using a web application as an interface for user interaction. The web scraper will extract stocks of various companies from their respective websites. Then compares them based on different parameters and suggests the user the best stock to invest their money in.

3.SYSTEM REQUIREMENTS

HARDWARE REQUIREMENTS :

Processor: P4 or higher RAM: 2GB and above Hard disk: 4GB and above

SOFTWARE REQUIREMENTS:

LANGUAGE :Python Programming Language

PLATFORM: Anaconda (Data Science) Jupiter Notebooks (live code, equations, visualizations and narrative text)

LIBRARIES:

Pandas : Pandas has a neat concept known as a Data Frame. A Data Frame can hold data and be easily manipulated. We can combine Pandas with Beautiful soup to quickly get data from a webpage.

NumPy : NumPy is another library we will use here. It lets us handle array and matrices especially those multidimensional. It also provides several high-level mathematical functions to help us to operate.

Matplotlib : Matplotlib is a plotting library for the Python programming language and its numerical mathematics extension NumPy. It provides an object-oriented API for embedding plots into applications using general-purpose GUI toolkits .

Seaborn: Seaborn is a Python data visualization library based on matplotlib. It provides a high-level interface for drawing attractive and informative statistical graphics.

4.ARCHITECTURE
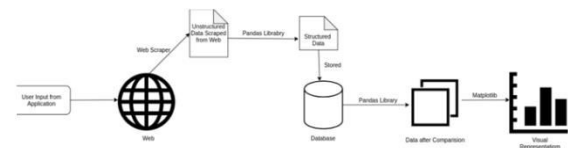
5.WORKING

The web scraper will extract stocks of various companies from their respective websites. Then compares them based on different parameters and suggests the user the best stock to invest their money in. Data extraction from the web using Python's
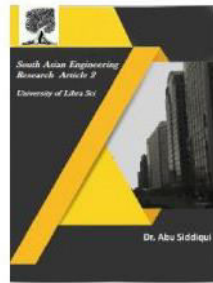
Beautiful Soup module. Data manipulation and cleaning using Python's Pandas library. Data visualization using Python's Matplotlib library

FLOW OF EXECUTION :

1.The user will enter the companies that he wants stock comparison for from the User Interface.

2.The application will then navigate to the desired company website and extract all the information from the website using beautiful soup.

3.The retrieved data will be given to the applications HTML Parser to extract only the useful information i.e., the stock prices.

4.The extracted stock data from each website is now put into the data frame using the Pandas Library of Python.

5.The data in the data frame may have multiple anomalies like null values, missing values etc. Such data will be cleaned using the NumPy library of Python.

6. The data is now compared and shown in graphical format using Matplotlib Library.

7.It also gives the best performing stocks to the user. The user can then make an accurate decision to invest in the appropriate stocks.

Storing data in data frames:

```
number of companies 7
{'CompanyName': 'Hero Motocorp Ltd.', 'StockValue': 2048.1}
{'CompanyName': 'Reliance Industries Ltd.', 'StockValue': 1328.65}
{'CompanyName': 'State Bank of India', 'StockValue': 302.9}
{'CompanyName': 'HDFC Bank Ltd.', 'StockValue': 1177.6}
{'CompanyName': 'Tata Consultancy Services Ltd.', 'StockValue': 2000.95}
{'CompanyName': 'ICICI Bank Ltd.', 'StockValue': 496.05}
{'CompanyName': 'Yahoo', 'StockValue': 38297.29}
                      CompanyName  StockValue
0              Hero Motocorp Ltd.     2048.10
1        Reliance Industries Ltd.     1328.65
2             State Bank of India      302.90
3                  HDFC Bank Ltd.     1177.60
4  Tata Consultancy Services Ltd.     2000.95
5                 ICICI Bank Ltd.      496.05
6                           Yahoo    38297.29
```

Comparing data and showing the stocks
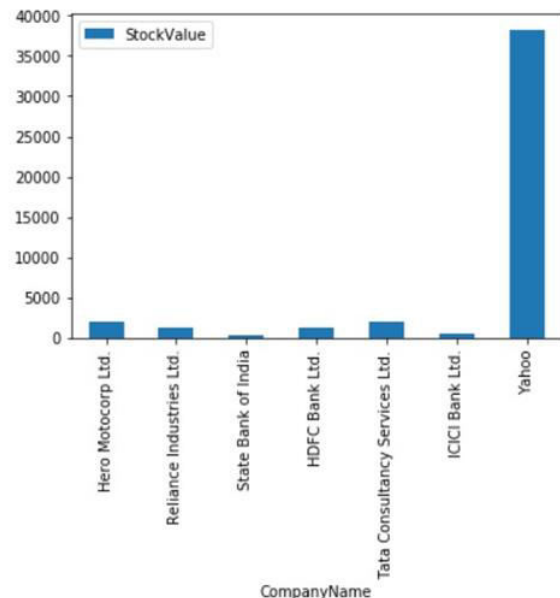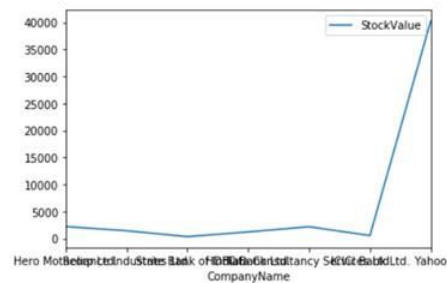
6.CONCLUSION

This project will help the users who are investing in stocks take more efficient decisions with accurate comparison results provided. The time taken in decision making is also reduced by a great extent. The project will not only provide comparison results but it also graphically shows the variations and

deviations in the stock prices being compared. It is user friendly and faster than the traditional way of researching for investing in stocks.

7.ACKNOWLEDGEMENT

8.REFERENCES

Bailey, W., Mao, C. X., & Rui, Z. (2003). Exchange rate regimes and stock return volatility: some evidence from Asia's silver era. Journal of Economics and Business, 55, 557-584.

Kim, J. R. (2003). The stock return: inflation puzzle and the asymmetric causality in stock returns, inflation and real activity. Economic Letters, 80, 155-160.

Mohanty, M., & Bota, G. (2010). Oil shocks and stock returns: the case of the central and eastern European (CEE) oil and gas sectors. Emerging Markets Review, 11, 358-372.