Journal of New Results in Science
https://dergipark.org.tr/en/pub/jnrs
Research Article
Open Access

E-ISSN:1304-7981 https://doi.org/10.54187/jnrs.1185912

A web scraping-based approach for fundamental analysis platform in financial assets

Yunus Santur[1], Mustafa Ulaş[2], Murat Karabatak[3]

Keywords:
Django,
Financial forecasting,
Fundamental analysis,
Selenium,
Web scraping

Abstract — There are two main benefits of using fundamental analysis for investors and portfolio managers. First, investing in a company with good ratios has lower risks. The second reason is that it is possible to evaluate share prices with internal valuation methods based on ratios. These price valuations can be more meaningful when combined with technical analysis data. Many data terminals provide processes such as fundamental analysis data and price valuation on a paid and licensed basis. However, the balance sheet data of publicly traded markets are publicly available and can be obtained and interpreted by web scraping methods. This study presents an approach in which basic analysis and price evaluation are made with balance sheets and ratios using open-source tools and web scraping.

Subject Classification (2020): 91G15, 91G80.

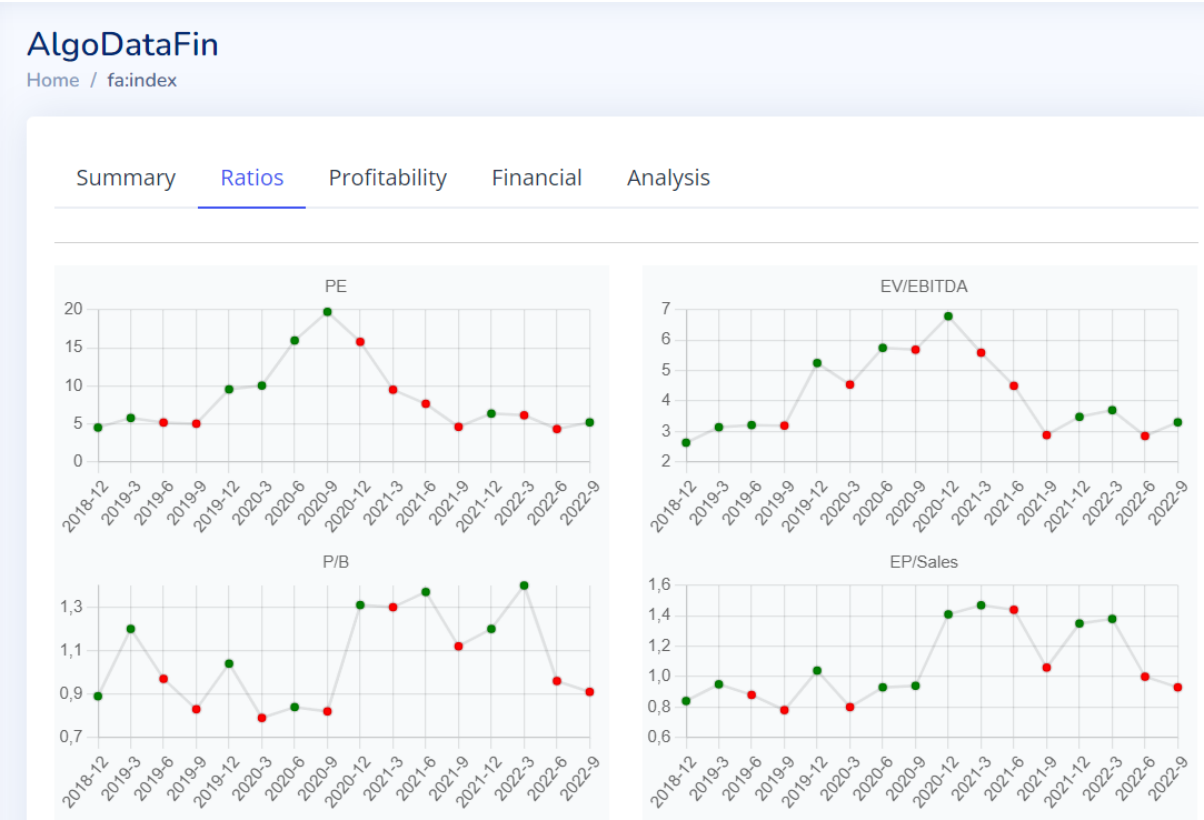
1. Introduction

Two main approaches are used in financial markets: technical analysis and fundamental analysis. Technical analysis mainly prefers by short-term investors called traders [1] and uses statistical methods-based approaches on time series consisting of opening, closing, lowest, highest price, and volume information of financial assets. These approaches are based on moving averages and indicators obtained with similar approaches to these averages [2]. On the other hand, fundamental analysis is an approach based on the ratios announced quarterly. These ratios are the company's quarterly profit or loss, indebtedness, stock status, cash flow, Price/Earnings ratio (P/E), and market Price-to-Book value (P/B) ratio [3]. Unlike technical analysis data, fundamental analysis data makes sense for long-term investment. For example, while a company's profitability increases, the decrease in the stock holding period and the decrease in the receivable collection period can indicate a significant acceleration in the company's growth figures. Using these ratios, the company's intrinsic value, that is, the required price can be determined. Different approaches are used for this valuation process. [4]. It is common to choose the statistically most probable of these approaches or to average them. In this way, if there are significant differences between the instant value of the company in the market in which it is traded and its real value, investment opportunities arise in the markets that are traded in two directions.

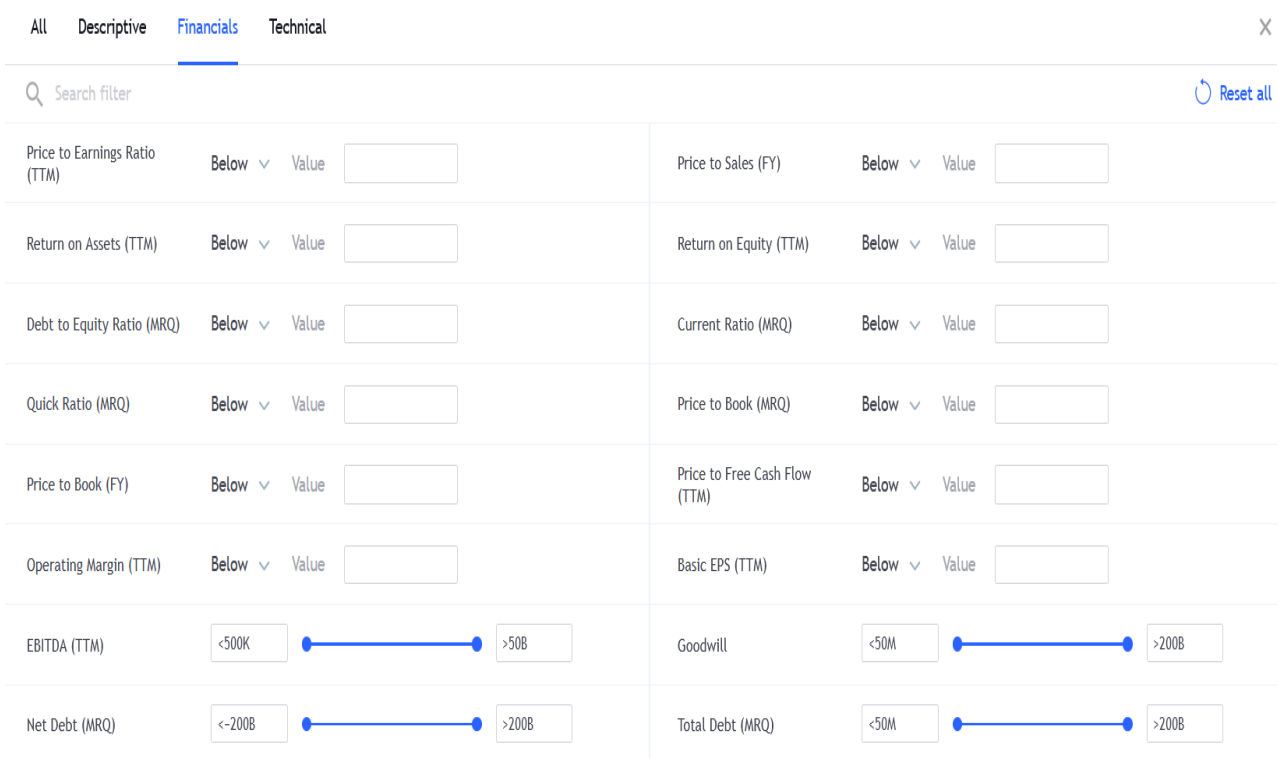
[1]ysantur@firat.edu.tr (Corresponding Author); [2]mustafaulas@gmail.com; [3]mkarabatak@firat.edu.tr
[1,2,3]Department of Software Engineering, Faculty of Technology, Fırat University, Elazığ, Türkiye
Article History: Received: 07 Oct 2022 — Accepted: 26 Dec 2022 — Published: 31 Dec 2022

The instant price of the company, which is traded on the stock market, is meaninglessly above these internal values, which provides an opportunity for short selling or short trading in the futures markets. As an opposite example, if the company's current price is below the stock market, there is an opportunity for a long transaction in the futures market [5]. Subfigures obtained from data terminals for technical and fundamental analysis are given in Figure.1. While technical analysis is used to determine the trend direction and momentum with price and price averages, fundamental analysis filters out hundreds of financial assets traded and identifies companies to invest in.

(a)



(b)



(c)

Figure 1. Fundamental analysis screener (a) Tradingview technical analysis [6] (b) Fundamental analysis example (c) Tradingview platform fundamental analysis filtering [6]

The main contribution of this study to the literature is the analysis of financial assets based on basic ratios and the establishment of a backend structure for integrating artificial intelligence in the future. The study is organized as Background, Material and Method, Results and Discussion and Conclusion sections.

2. Background

Fundamental analysis data of a company is obtained from the quarterly balance sheets announced by the company. Since it is generally prepared as an excel table, the ratios can be created automatically with a programming language. However, it can sometimes occur in extraordinary situations on data such as profitability, debt, stock transfer, and balance sheet. For example, the sale of an immovable property the company already owns can be seen as a profit in the balance sheet period. Still, the company may even have sold the immovable property for debt payment. Even if the balance sheet figures appear positive, such exceptional cases, called footnotes, should also be interpreted.

Similarly, in a situation where the amount of borrowing of the company increases, the company may have grown through acquisitions. Therefore, just like in technical analysis for financial assets, it cannot be expected that the forecasts to be made with the results to be obtained from the figures and formulas for the future in the fundamental analysis will show 100% consistency. For this purpose, text processing and interpretation approaches can be used. In this case, the most general approach that can be used is to make sense of the related footnotes by reading them automatically and classifying the emotions [7].

The ratios used in the basic analysis are given in groups in Figure.2 below. There are dozens of ratio data. All ratios can be obtained from excel tables, easily analysed and visualised with data science approaches, and converted into .csv format. As a general opinion, it is accepted that the P/E and P/B

values are small, and the Earnings Before Interest, Tax, Depreciation and Amortisation (EBITDA) ratio grows positively. However, no single ratio is meaningful on its own. For example, investing in a company whose P/E value has decreased only according to this ratio may involve high risk. The main point here is to extract the association rules between the ratios, interpret them by experts, and make a price valuation [8].

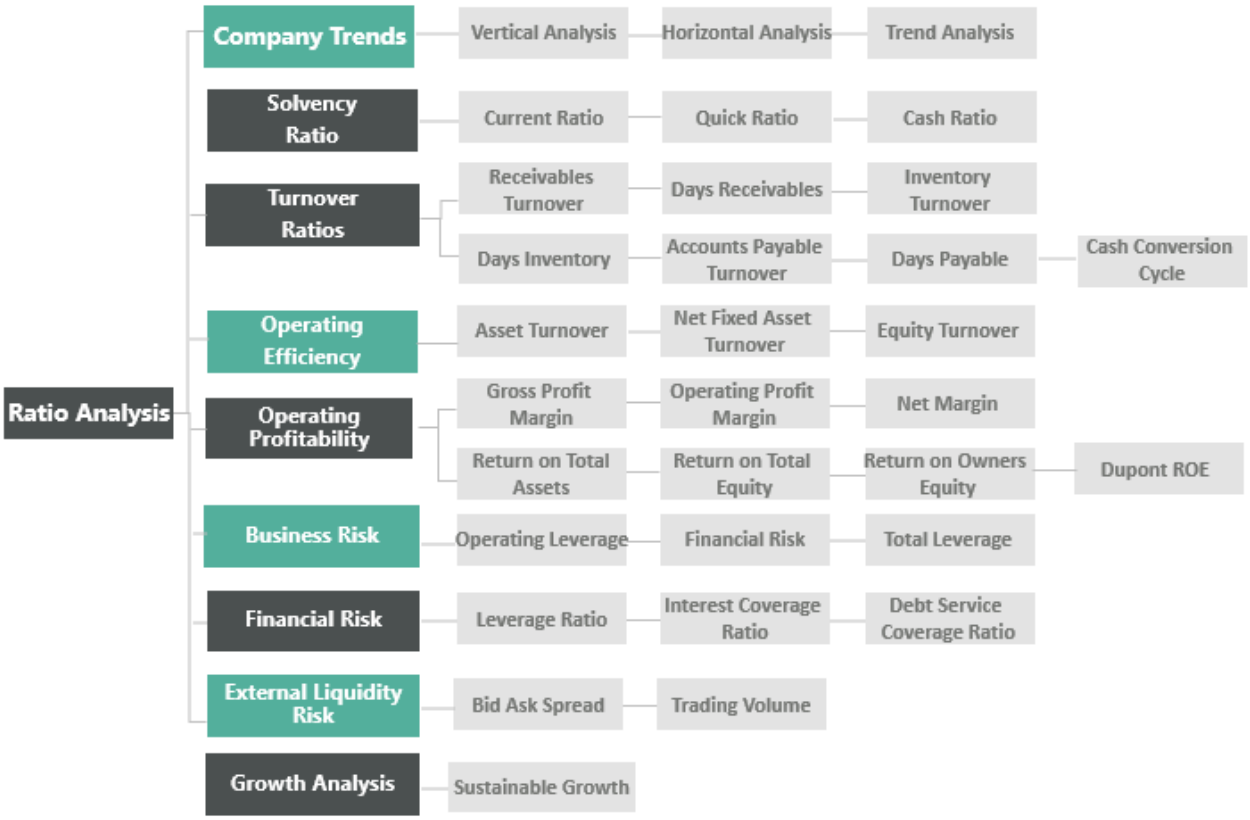


Figure 2. Grouped fundamental analysis ratios [9]

3. Material and Method

In the development of this approach, web-based software has been developed using the Python programming language and Selenium library to visualise and price valuation by pulling automated data from the Public Disclosure Platform (KAP), Yahoo Finance and similar sources that contain financial data and balance sheets [10-12]. The general diagram of the developed approach is given in Figure 3.

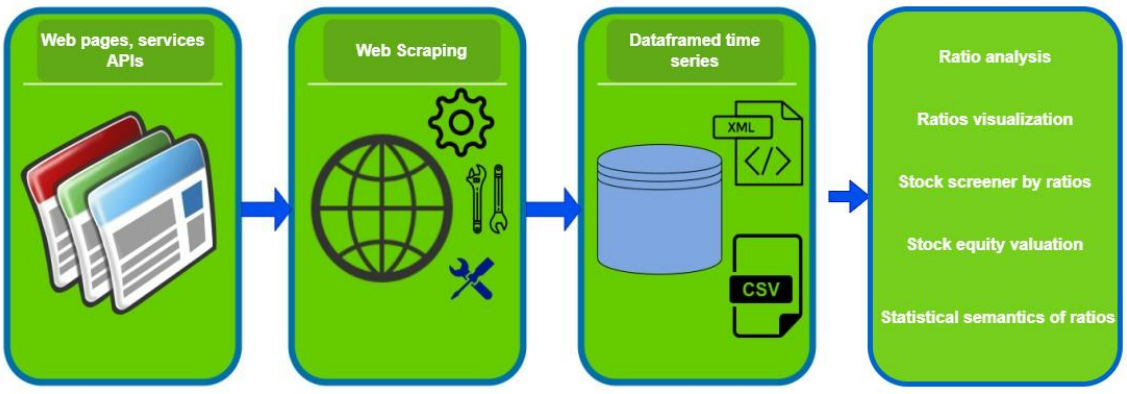
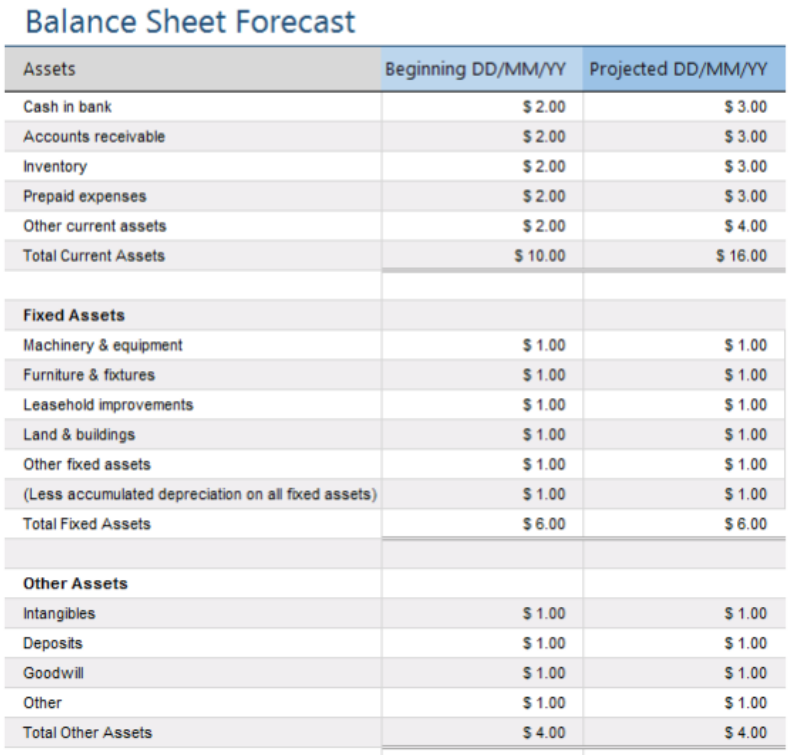


Figure 3. The architecture of the proposed model

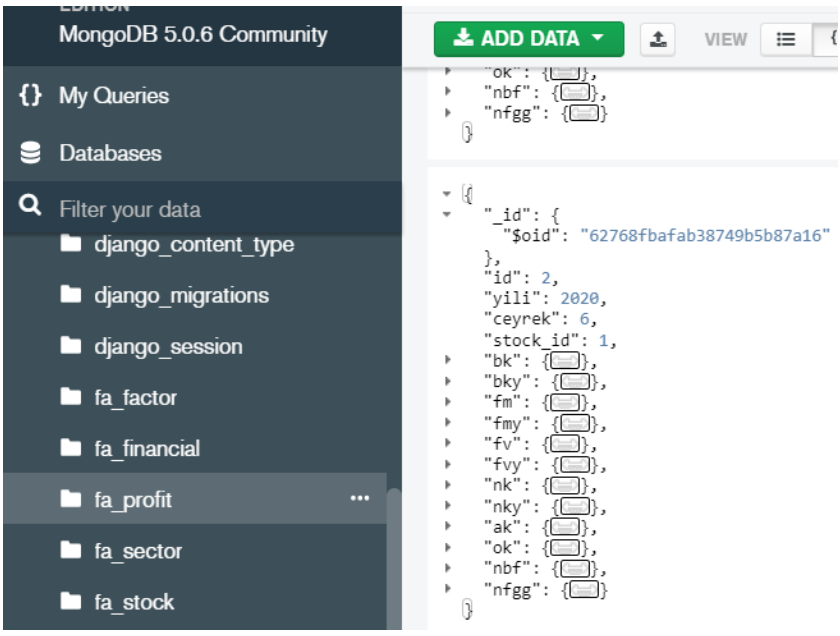
Collecting data from websites and sorting through the collected data is called web scraping. Selenium is a web browser testing and automation software. It performs the operations according to the desired conditions. It is generally used for automated tests, but thanks to these features, it can be used for downloading and saving data from web resources within certain rules [13]. In Figure 4, raw data and the balance sheet, a typical excel document, are transformed into a structured data structure. For this purpose, MongoDB was used in the No-SQL structure. MongoDB keeps data in a .json/.bson structure [14, 15].



Balance Sheet Forecast

Assets	Beginning DD/MM/YY	Projected DD/MM/YY
Cash in bank	$ 2.00	$ 3.00
Accounts receivable	$ 2.00	$ 3.00
Inventory	$ 2.00	$ 3.00
Prepaid expenses	$ 2.00	$ 3.00
Other current assets	$ 2.00	$ 4.00
Total Current Assets	$ 10.00	$ 16.00
Fixed Assets		
Machinery & equipment	$ 1.00	$ 1.00
Furniture & fixtures	$ 1.00	$ 1.00
Leasehold improvements	$ 1.00	$ 1.00
Land & buildings	$ 1.00	$ 1.00
Other fixed assets	$ 1.00	$ 1.00
(Less accumulated depreciation on all fixed assets)	$ 1.00	$ 1.00
Total Fixed Assets	$ 6.00	$ 6.00
Other Assets		
Intangibles	$ 1.00	$ 1.00
Deposits	$ 1.00	$ 1.00
Goodwill	$ 1.00	$ 1.00
Other	$ 1.00	$ 1.00
Total Other Assets	$ 4.00	$ 4.00

(a)



(b)

Figure 4. Stock raw data and structured database image (a) Balance sheet raw data (b) No-Sgl MongoDB

In this study, the ratios were obtained with Selenium, and their structural storage was performed with MongoDB. Matplotlib was used in the Python environment for the visualisation of the ratios. Python and MongoDB map-reduce paradigm is used for stock filtering [16]. For the filtering process, the 1-Many structure has been designed to establish relations such as stocks, the sector it belongs to and indices. In Figure 5, the Django web framework and the model file of the application are given. Below is an example of obtaining a semantic data summary based on statistical data using the Map-Reduce paradigm and Aggregation framework. Pandas library data frame structure is used for filtering. This structure, which is kept as a data frame in the server environment, is sent to the client in .json format. Data sending is done only once. Filtering according to the selected criteria is performed on the client using Javascript. In this context, dynamic filtering and sorting can be done by combining date, string, integer, and float fields with and/or conditions with numerous criteria such as small, large, between, containing or not.

```
class Stock(models.Model):
    son = models.FloatField(null=True, blank=True)
    name = models.CharField(max_length=6, null=False)
    title = models.CharField(max_length=50, null=False)
    web = models.CharField(max_length=50, null=True, blank=True)
    f_k = models.FloatField(null=True,blank=True)
    pd_dd = models.FloatField(null=True,blank=True)
    fd_favok = models.FloatField(null=True,blank=True)
    hao = models.FloatField(null=True,blank=True)
    ypay = models.FloatField(null=True,blank=True)
    pd = models.FloatField(null=True,blank=True)

    sector=models.ForeignKey(Sector, on_delete=models.CASCADE)
    slug = models.SlugField(null=True, unique=True, db_index=True,default=0, allow_unicode=True)

    def save(self, *args, **kwargs):
        self.slug=slugify(self.name)
        super().save(*args, **kwargs)

    def __str__(self) -> str:
        return self.name
```

Figure 5. Example of using Django model and Aggregation framework Map-Reduce

The last application carried out within the scope of the study is the share valuation scenarios, which are very important for investors. In this context, three basic approaches are used in the example given. It can be assumed that the ratio presented here is P/E or P/B. In the first approach given in Equality 3.1, the share valuation is done by taking the proportional value according to the ratio. The approach used here is the simple ratio-proportion approach. In Equality 3.2, the sector average of the relevant ratio value is found instead of taking the constant of 1 as in Equality 3.1. For example, there is a relative value of the stock according to the average of all stocks in the "Food" sector. This approach can be extended recursively to the European and world averages. There is multiple valuation management according to the selected ratio feature. Equality valuation can be obtained by taking the average or weighted average of all valuations, as in Equality 3.3, with the simplest approach [4]. The approaches presented are based on statistical calculations. The motivation of this study is expanded with artificial intelligence-based prediction models in the future. For this purpose, recent literature studies will be used in price valuation of financial assets based on basic ratios [17-19].

$$Value\ of\ Equity\ (VoE) = Stock_{price}\ \frac{1}{Ratio_{stock}} \tag{3.1}$$

$$VoE = Stock_{price} \frac{\mu(Ratio_{sector})}{Ratio_{stock}} \tag{3.2}$$

$$VoE = \mu(VoE_1, VoE_2, \ldots, VoE_n) \tag{3.3}$$

4. Results

In this study, a software library has been introduced to enable investors to perform the basic analysis transactions they need for financial assets using the Python programming language and the Django web library. With the implemented application, a backend was created with MongoDB and Django. Then, stock information and balance sheets were automated with Selenium. The application has the features of obtaining and visualising the basic analysis ratios, making price valuations according to the basic ratios and making sense of them according to statistical values. Figure 6 displays basic information such as the last price, most frequently used ratios, free float, foreign share and price valuation based on more than one ratio value.

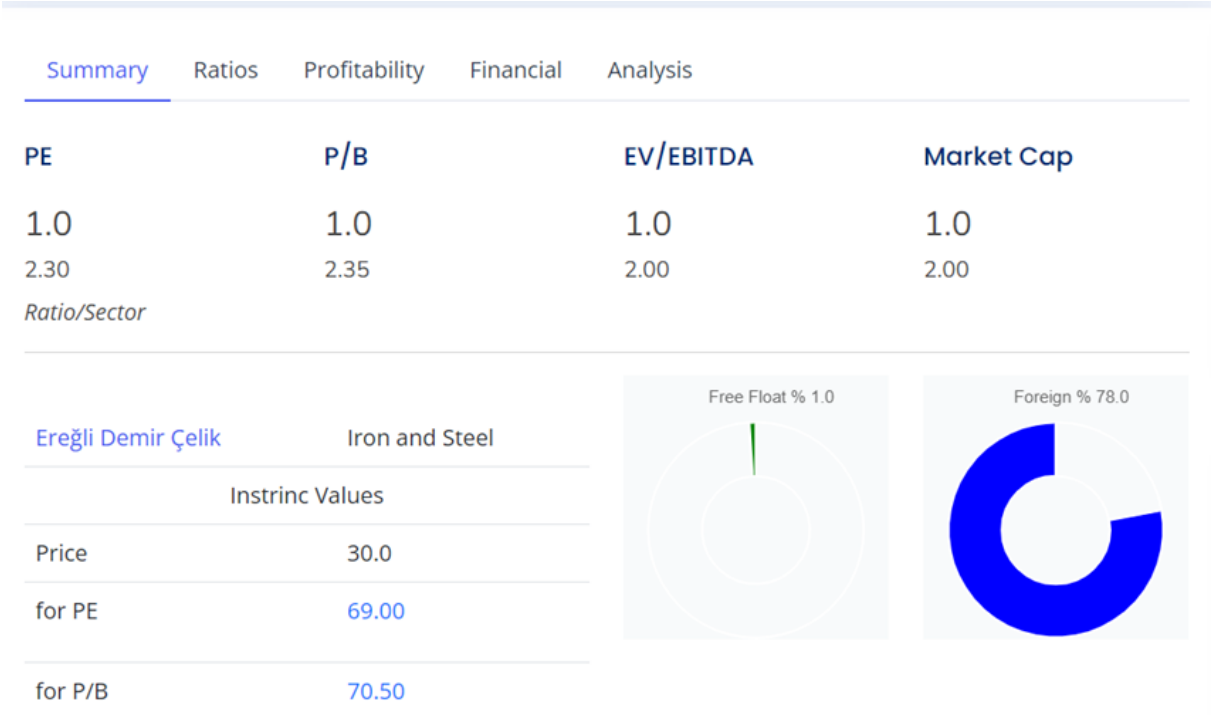


Figure 6. Display of stock basic ratios and price valuation according to more than one ratio

In Figure 6, the basic ratios and the groups they belong to are given in tabs. The ratios obtained from the quarterly balance sheets can be visually displayed in each tab. In Figure 7, ratios such as market value, earnings per share, dividend per share, stock holding period, receivable collection period, debt and collection period are given in this group of multipliers.

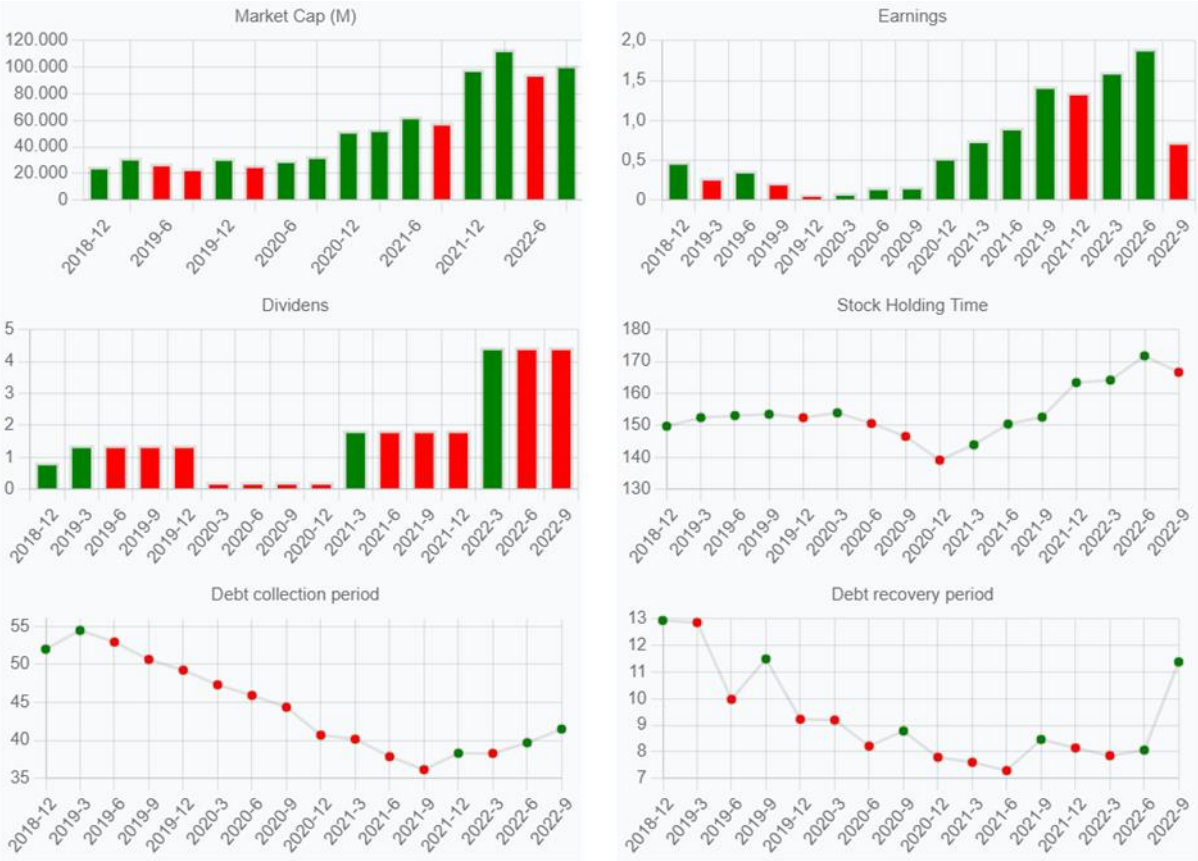


Figure 7. Visualisation of stock base ratios

Figure 8 shows the sectoral analysis screen. The selected sector or index shares in this tab can be sorted practically according to the selected ratio value.

AlgoDataFin

Home / fa:sektor

Sectors

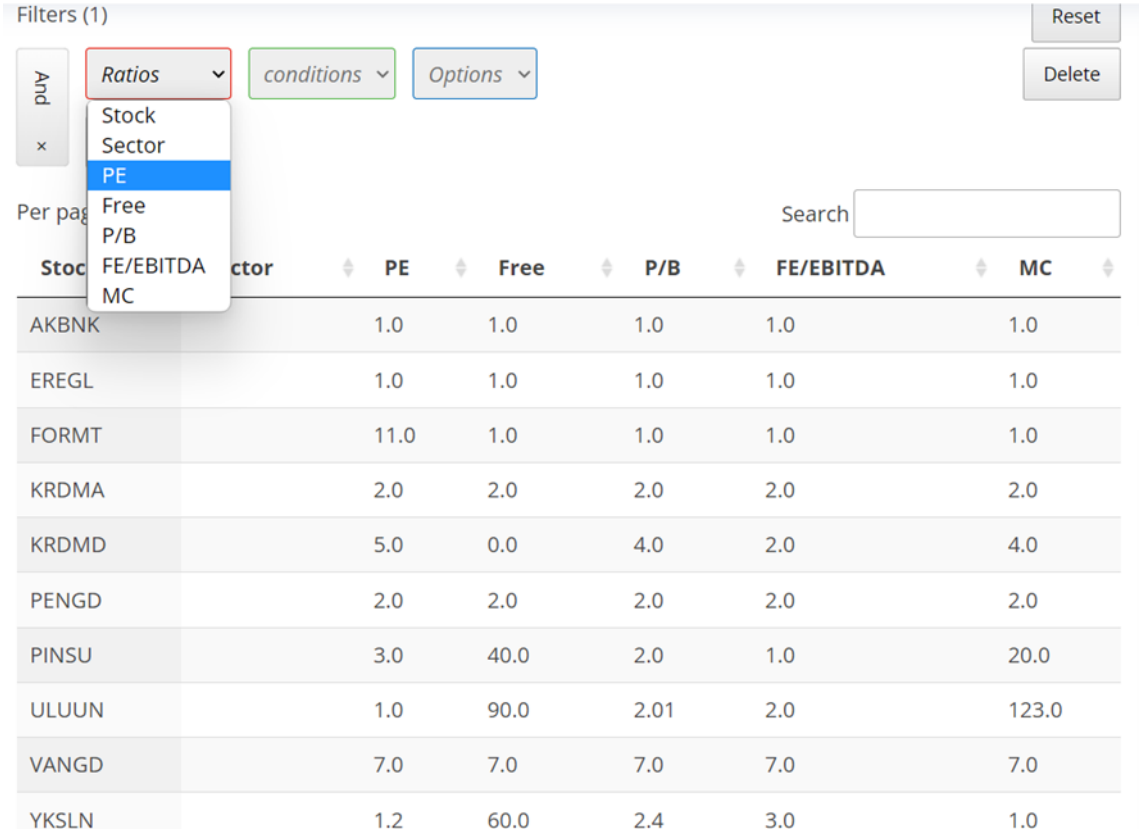
Select sector

Demir çelik | 4 hisse bulundu

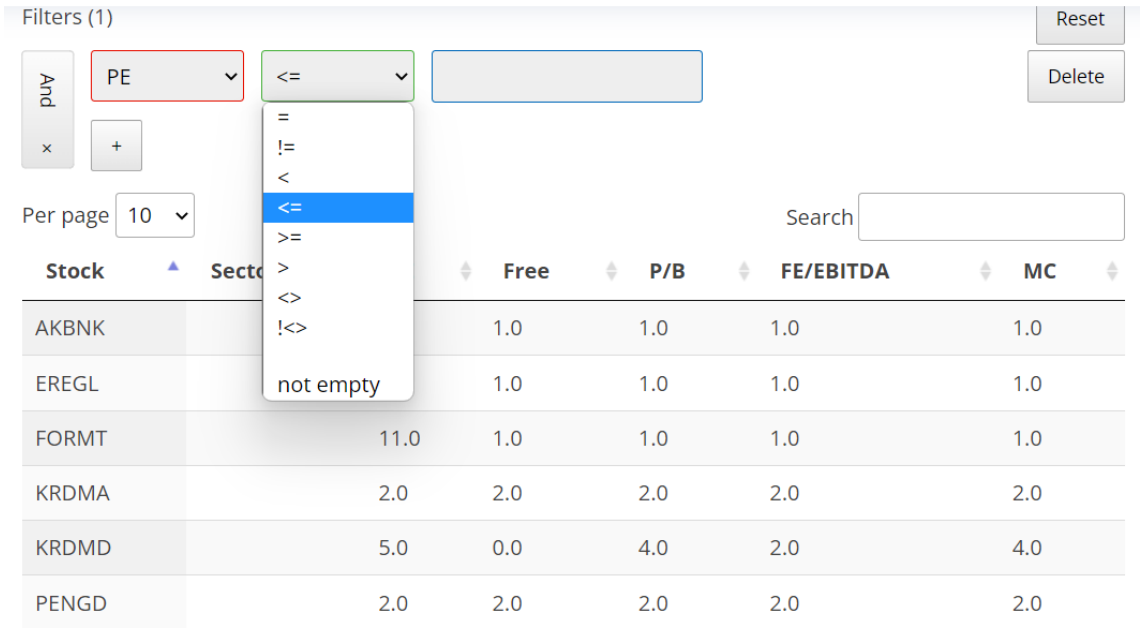
Stock	Last	P/E	P/B	FE/EBITDA	Free Float	Foreign	Market Cap
#EREGL	30.0	1.0	1.0	1.0	1.0	78.0	1.0
#YKSLN	24.0	1.2	2.4	3.0	60.0	20.0	1.0
#KRDMA	23.0	2.0	2.0	2.0	2.0	2.0	2.0
#KRDMD	16.2	5.0	4.0	2.0	0.0	40.0	4.0

Figure 8. Analysis of basic share ratios by sectors

In Figure 9, on the other hand, the shares can be filtered according to the criteria of more than one ratio by combining the and/or conditions according to the sector, index and date. Among these criteria, there are many options, such as being small, large, inclusion, not containing, being in between and not being empty.



(a)



(b)

Figure 9. Advanced stock filtering (a) and (b) based on fundamental ratios

In the last tab of the application, statistical data (such as the smallest, largest and above average) and simple semantic data are given in Figure 10.

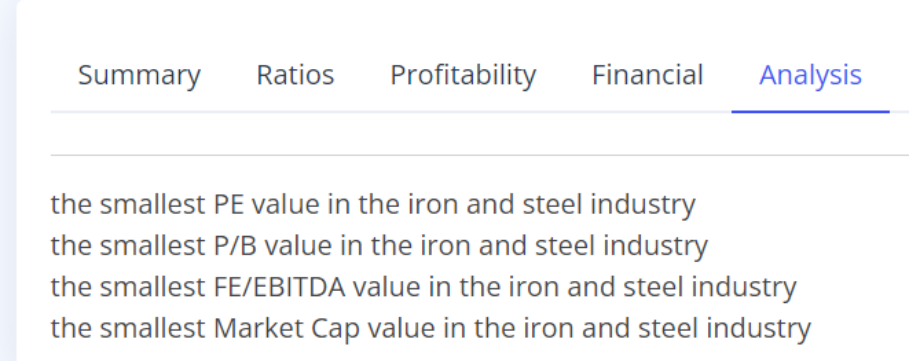


Figure 10. Obtaining simple statistical semantics

5. Discussion and Conclusion

The main contribution of this study to the literature is the analysis of financial assets in basic ratios and the establishment of a backend structure for integrating artificial intelligence in the future. The biggest disadvantage of working is that it currently works on localhost. It is not put into service on the cloud or any platform. experimental developments continue to be developed in the local environment. This study proposed an approach to obtain and visualise publicly traded companies' ratios, statistically interpret companies' relative and within their sectors, find intrinsic value, and make real-time and semantic queries among dozens of ratios.

The proposed approach is coded in Python and developed as a Django platform. This study will be designed to make semantic queries and artificial intelligence-supported stock suggestions based on fundamental analysis in the future. At this stage of the study, the MongoDB database and scientific python environment are designed as a backend structure. Artificial intelligence models developed within the scope of the project will be integrated into this structure in the future and will work as a smart recommendation platform at near time.

Author Contributions

All the authors contributed equally to this work. They all read and approved the last version of the paper.

Conflicts of Interest

The authors declare no conflict of interest.

Acknowledgement

This work was supported by the Scientific and Technological Research Council of Turkey (TUBITAK, Grant No. 121E733).

References

[1] A. Altan, S. Karasu, *The effect of kernel values in support vector machine to forecasting performance of financial time series*, The Journal of Cognitive Systems, 4(1), (2019) 17-21.

[2] O. B. Sezer, M. U. Güdelek, A. M. Özbayoğlu, *Financial time series forecasting with deep learning: A systematic literature review: 2005–2019*, Applied Soft Computing, 90, (2020) 106181.

[3] C. Rudin, C. Chen, Z. Chen, H. Huang, L. Semenova, C. Zhong, *Interpretable machine learning: Fundamental principles and 10 grand challenges*, Statistics Surveys, 16, (2022) 1-85.

[4] I. K. Nti, A. F. Adekoya, B. A. Weyori, *A systematic review of fundamental and technical analysis of stock market predictions*, Artificial Intelligence Review, 53, (2020) 3007-3057.

[5] Y. Huang, Y. Gao, Y. Gan, M. Ye, *A new financial data forecasting model using genetic algorithm and long short-term memory network*, Neurocomputing, 425, (2021) 207-218.

[6] Tradingview, Tradingview platform, retrieved from http://www.tradingview.com. Accessed on August 13, 2022.

[7] Y. Santur, *Candlestick chart based trading system using ensemble learning for financial assets*, Sigma Journal of Engineering and Natural Sciences, 40(2), (2022) 370-379.

[8] J. Roeder, M. Palmer, J. Muntermann*, Data-driven decision-making in credit risk management: The information value of analyst reports,* Decision Support Systems, 158, (2022) Article ID: 113770.

[9] Fundamental Analysis Ratios, retrieved from https://www.wallstreetmojo.com/ratio-analysis. Accessed on August 13,2022.

[10] N. R. Mustika, N. Novrina, *Automated Black Box Testing using Selenium Python*, International Journal of Computer Science and Software Engineering, 7(9), (2018) 201-204.

[11] KAP, Public Disclosure Platform, retrieved from https://www.kap.org.tr/en. Accessed on August 13, 2022.

[12] Yahoo Finance. Finance, retrieved from yahoo.com. Accessed on August 13, 2022.

[13] Ü. Aycel, Y. Santur, *A new moving average approach to predict the direction of stock movements in algorithmic trading*, Journal of New Results in Science, 11(1), (2022) 13-25.

[14] M. Etemad, A. Soares, P. Mudroch, S. A. Bailey, S. Matwin, *Developing an advanced information system to support ballast water management*, Management of Biological Invasions, 13(1), (2022) 68-80.

[15] Y. Santur, *Deep learning-based regression approach for algorithmic stock trading: A case study of the Bist30*, Gümüşhane University Journal of Science and Technology, 10(4), (2020) 1195-1211.

[16] E. Özhan, E. Uzun, The Analysis of Big Financial Data Through Artificial Intelligence Methods, Impact of Artificial Intelligence on Governance, Economics and Finance, Springer Nature Singapore, 1, (2021) 51-79.

[17] U. Erkan, A *precise and stable machine learning algorithm: eigenvalue classification (EigenClass)*, Neural Computing and Applications, 33(10), (2021) 5381-5392.

[18] S. Memiş, S. Enginoğlu, U. Erkan, *Numerical data classification via distance-based similarity measures of fuzzy parameterized fuzzy soft matrices*, IEEE Access, 9, (2021) 88583-88601.

[19] Ü. Aycel, Y. Santur, *A new algorithmic trading approach based on ensemble learning and candlestick pattern recognition in financial assets*, Turkish Journal of Science and Technology, 17(2), (2022) 167-184.