SETSCI Conference Proceedings
17, 25-29, 2024

*Cognitive Models and Artificial Intelligence Conference*

*May 25-26, 2024, İstanbul-Türkiye*

https://doi.org/10.36287/setsci.17.1.0025
2687-5527/ © 2024 The Authors. Published by SETSCI

# Stock Portfolio Selection with Supervised Classification

Hasan Taş[1*], Esra Karasakal [1]

[1]*Industrial Engineering, Middle East Technical University, Ankara, Türkiye*
[*]*tash@metu.edu.tr*

***Abstract –*** This study employs a supervised learning approach for the Stock Portfolio Selection (SPS) problem. The proposed approach solves three problems defined in the literature review: the narrow period selection problem, the proper objective selection problem, and the comparable peer selection problem. Three classification methods are utilized to categorize the stocks into select (purchase) and ignore (do nothing) classes. The genetic algorithm is used for training the classification parameters. Trained individuals together form a portfolio of stocks based on a voting mechanism. Forty-quarters of the data of 29 of 30 DJI index stocks are chosen for experiments. Statistical analysis of experimental results shows that focusing on the defined problems helps in choosing a portfolio of stocks that beats the market.

*Keywords – Stock portfolio selection, portfolio optimization, genetic algorithm, finance, fundamental analysis*

## I. INTRODUCTION

Investment is the act of purchasing an asset with the expectation of gaining returns in the future. For this purpose, various publicly traded investment instruments, such as bonds/bills, commodities, stocks, and derivatives, are available in the capital markets. Stocks are one of the most common investment instruments that are publicly traded shares of the companies. Stock investing involves systematic and systemic risks that may cause capital loss. To avoid these risks, a portfolio is formed with various unrelated stocks. Forming the stock portfolio involves Stock Portfolio Selection (SPS) and portfolio optimization. The former means selecting the best (or a relatively good) subset of many stocks regarding the return and risk criteria. The latter focuses on the optimal allocation of the budget into selected stocks. Classification and clustering methods are among the most suitable tools in SPS studies since they categorize data instances (like stocks) into predefined groups.

SPS problem involves the valuation of the stocks. There are two valuation methods: absolute and relative (or comparative). Absolute valuation is based on predicting the intrinsic value of a stock by investigating its assets, liabilities, and potential future cash flows. Relative valuation involves meaningful financial ratios (relative valuation metrics or financial ratios) to compare the companies to each other and determine the value of the corresponding company based on the value of the other(s) (Plenborg & Pimentel, 2016).

There is no one best relative valuation metric that best predicts the future price of a stock. Therefore, a combination of a set of relative valuation metrics is employed in this process. Various papers employ relative valuation metrics as features (in machine learning) or criteria (in multicriteria classification) to compare the stocks. Some studies regress a potential future price of a stock (e.g., Geertsema and Lu, 2023). Others classify/categorize stocks as attractive-unattractive (to purchase) or undervalued-overvalued classes (e.g., Khedmati and Azin, 2020). This study focuses on classification approaches to the SPS problem.

This study identifies three problems in SPS that are explained in Section 2.3: the narrow training period selection, proper objective selection and comparable peer selection problems (practical concerns in finance, (Plenborg & Pimentel, 2016)).

Two settings are studied as the proposed method. The first (Base) setting focuses on the narrow period and proper objective selection problems. The second (Extended) setting focuses on the Base setting's problems and the practical concerns in finance (comparable peer selection) problems. The methods are compared based on thirty random replications, and the Base setting outperforms the market in each replication with significantly higher returns than the benchmark (index return). The Extended setting outperforms the Base setting. Therefore, we recommend considering the usage of the Extended setting.

The remainder of this report is organized as follows. Section 2 is a review of the studies with supervised and unsupervised categorization of stocks in Machine Learning (ML), Preference Disaggregation Analysis (a form of supervised learning; abbr. PDA), and MultiCriteria Decision Making (MCDM) areas. Section 3 continues with listing the gaps in the literature and descriptive statistics about the data. Section 4 presents the proposed approach. The experimental setting and results are presented in Section 5.

## II. LITERATURE REVIEW

Three types of approaches are employed in evaluating stocks and forming portfolios in ML: regression, classification, and clustering. In MCDM literature, two methods are employed in evaluating stocks: multicriteria ranking (ordinal regression) and MultiCriteria Sorting (abbr. MCS, also called ordinal classification). Ordinal classification differs from nominal classification in the definition of classes. In ordinal classification, classes respect a preference order

### A. ML Approaches

There are several studies focusing on the categorization of stocks and markets (Huang et al., 2005; Khedmati and Azin, 2020; Goudarzi et al., 2017). The categorization is based on the financial performance of the stocks and technical indicators of the stock price. Previous studies focus on stocks or the stock market (an index composed of a group of stocks).

Huang et al. (2005) argue that the trades based on small-error price forecasts may not be as profitable as the ones based on market direction prediction. Huang et al.'s (2005) study predicted the stock market's direction. The market (NIKKEI 225 Index) is categorized with Support Vector Machines as bullish (increasing price) or bearish (decreasing price).

Khedmati and Azin (2020) employ a clustering approach to select an attractive subset of stocks to form a portfolio. Goudarzi et al. (2017) cluster stocks into groups and rank the groups from best to worst using the Technique for Order of Preference by Similarity to the Ideal Solution (TOPSIS) method (a popular distance-based ranking or ordinal regression method). Also, some studies are regressing the value of a stock with relative valuation metrics to aid purchase and sell decisions (e.g., Geertsema and Lu, 2023).

### B. MCDM Approaches

PDA is the learning of a decision-maker's preferences by training a preference function with preference examples or a reference set (preference examples and reference sets are substitute phrases for training data in PDA). Supervised learning is a well-suited and commonly employed approach for this purpose. MCS is the ordinal classification of data instances to predefined preference-ordered classes. This approach is also among the most suitable approaches for SPS. Various studies train monotonic utility functions to infer a decision-maker's preference in SPS (Zopounidis et al., 1999; Zopounidis and Doumpos, 2002).

The features utilized in these studies are relative valuation metrics and profitability indicators. However, these studies consider classification accuracy as the success criteria and ignore portfolio profitability, which is not parallel to the aim of investing. Also, comparable peer selection problem and practical considerations in using relative valuation metrics are omitted. Furthermore, training data consists of a single period that could impair learning and validation (enhance overfitting).

In both ML and MCDM approaches, comparable peer selection problem in using relative valuation metrics are omitted, which are critical in finance literature (Plenborg & Pimentel, 2016).

### C. Gaps In the Literature

This paper categorizes the gap in the literature into three problems: proper objective selection, narrow period selection and comparable peer selection problems.

1. The first problem: Studies focusing on training the model with a single period are categorized to narrow the selection problem.
2. The second problem: Studies that do not consider the financial success of the models categorized to proper objective selection problem.
3. The third problem: Studies that omit practical concerns on the best usage of the features are categorized as comparable peer selection problem.

Table 1: The three problems defined in the literature review

| Papers | The narrow period selection problem | The proper objective selection problem |
|---|---|---|
| Zopounidis et al., (1999) | + | + |
| Zopounidis and Doumpos (2002) | + | + |
| Khedmati and Azin (2020) | | |
| Huang et al. (2005) | | |

These problems are tackled by employing the following:

1. Collecting multi-period data in the data collection phase.
2. The model evaluation phase involves adding the financial success metrics to the models, such as return or portfolio profit.
3. Implement data preprocessing to cope with practical concerns in finance (comparable peer problem) in the data preprocessing phase.

## III. The Data Set and The Features

### A. The data set and the data collection phase

The data set is the 10-year (6/2013-6/2022) quarterly stock price and relative valuation metrics (appendix B) data in Dow Jones Industrial Average and 10-year return, specific risk and systemic risk data from yahoo finance. The data set is obtained by web scraping from stockanalysis.com and Yahoo Finance.The features used to evaluate the stocks for classifications are relative valuation metrics and profitability indicators obtained from companies' balance sheets, income statements, and cash flow statements (Appendix B). More detail can be found in (Zopounidis et al., 1999; Zopounidis and Doumpos 2002; Plenborg & Pimentel, 2016; Geertsema and Lu, 2023).

The data set involves 29 stocks (out of 30) from the Dow Jones Industrial average index (DJI 30). One of the stocks is excluded from the set due to the absence of data in narrow periods (full list of included stocks is given in Appendix A). The features utilized for stock selection are given in Appendix B. Their meanings and economic interpretations are explained. The data is scaled to [0,1] range to avoid scaling affect using ML methods.

### B. Descriptive statistics about the features and the comparable peer selection

Descriptive statistics is employed to analyze the features and their scales. Boxplot is chosen for the analysis since it directly reflects the estimated mean and the spread of the data and gives meaningful insights about categorical/qualitative evaluations.

A stock is undervalued (better or more attractive to purchase) compared to another stock if its Price-book ratio (P/B) is lower than the other. That is, all of the stocks of a sector that have P/B above the mean are overvalued compared to the ones under the mean.

Figure 1 is the boxplot of the P/B ratio of stocks from different sectors. Consider the technology sector in Figure 1. In Figure 1, many overvalued (red region) "technology sector" stocks (above mean) have less P/B ratio than the undervalued (green region) "industrials sector" stocks. When these stocks are compared without concerning their industry membership, almost all of the technology sector stocks would seem more undervalued than the industrial sector stocks. Thus, such an evaluation would be misleading. Therefore, one of the three school of thoughts of relative valuation suggests comparing the stocks in their related sectors. This suggestion is immediately verified with the descriptive statistics in Figure 1. This is called the comparable peer problem, meaning one cannot bundle all stocks without their sector information and use them analytically.

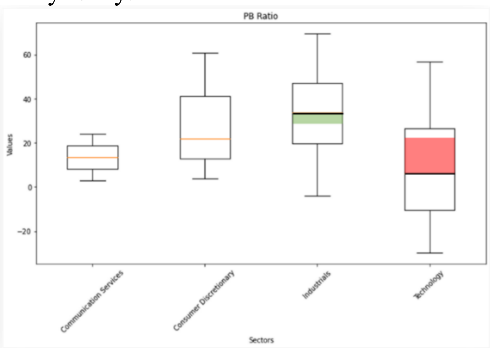


Fig. 1: Boxplot of the P/B ratio of stocks from different sectors

In all of the ML and MCDM studies included in the literature review of this report, comparable peer problems are omitted. Dividing all stock features by their mean (or median, weighted average) would be a viable approach to overcoming the comparable peer problem. Plenborg & Pimentel (2016) research possible courses of action in comparing the stocks and using median instead of average results in more accurate results. Because the median is less sensitive to outliers in the data, in this study, stocks will be re-scaled with the sector medians to be in line with comparable peer selection. Figure 2 presents the resulting boxplots. Now, the means are aligned, and overvalued companies of one sector are not undervalued compared to stocks of other sectors. This action directly focuses on the third problem. However, in the experiments, the third problem has not been considered yet and will be considered in the final report.

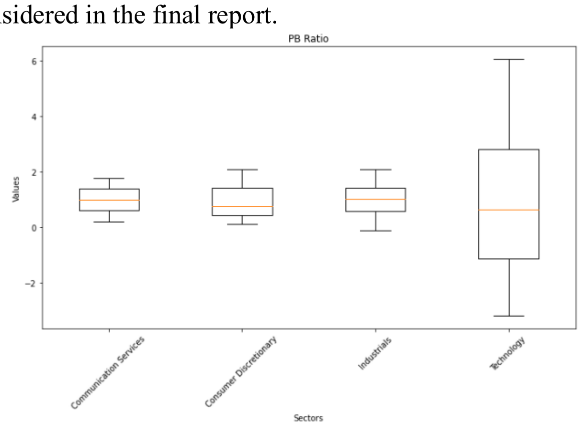


Fig. 2: The sectors after scaling with division by median

## IV. Proposed Approach

The proposed approach is a supervised learning approach involving labelling stocks that outperform the DJI to purchase class and others to ignore class to enable supervised classification. It employs the multiple-period scheme and minimizes the sum of the classification error of all periods at once. In this kind of problem, there is no class cardinality problem where one of the two (purchase and do not purchase/ignore) classes is empty. The empty class situation has an economic interpretation. If the classifier assign all of the data instances to the purchase class then it likes to mimic the market (or the index). If the purchase class is empty then the budget is invested in the risk-free asset for that period.

The linear, polynomial, and Nearest-Centroid classifiers (NC) are chosen for experimentation with different classification methods. To resolve the comparable peer selection problem, stocks of each industry is evaluated according to its industry standards. Based on the comparison made in Section 3.2, in the Extended setting, features are converted to indicator variables. In conversion, as industry standard, industry average, weighted average and median is used. Since the median is less sensitive to outliers, the median is preferred in this study. The features are relabeled as follows:

$$New\ feature\ value = \begin{cases} 1, \text{if the feature value is below its industry median} \\ -1,\ \text{otherwise} \end{cases} \quad (1)$$

The genetic algorithm is employed for the optimization of linear, polynomial, and NC classifiers' parameters (weights ($w$), and thresholds ($w_0$)). The motivation behind choosing a population-based method is that we can train many individuals (agents) in communication with each other (crossover) and then combine their portfolios to count on the diversified opinions of individuals. When we normalize the frequency of stocks repeating in different agents' portfolios, we can distribute the budget according to a voting mechanism. The portfolio allocation function is as equation (2).

$$Proportion\ of\ stock\ i = \frac{number\ of\ agents\ including\ stock\ i\ in\ the\ portfolio}{Total\ number\ of\ stocks\ held\ by\ all\ agents} \quad (2)$$

To resolve the second problem, the training model evaluation performance measure (fitness of individuals) is chosen as the classification accuracy plus a small constant times the portfolio profit as below.

$$Augmented\ Fitness = Classification\ Accuracy + 0.001 * Portfolio\ Profit \quad (3)$$

This formulation enables us to evaluate the model success as a multi-objective approach that chooses a non-dominated solution that not only minimizes the classification error but also chooses a more profitable portfolio between alternative non-dominated solutions. That is, a set of classification parameters may result in the same classification accuracy, but one may result in a more profitable portfolio. This approach avoids a weakly dominated solution.

## V. EXPERIMENTS

The proposed methods are named the Base and Extended settings for ease of comparison. In the Base setting, the method focuses on the narrow period selection and proper objective selection problems, and the Extended setting focuses on the Base setting plus the comparable peer selection problem. The Base setting uses the augmented fitness function in Equation (3) and randomly selects multiple past training periods to train the model. The Extended setting treats features as indicator variables as in Equation 1.

Beating the market (i.e., having above-index returns in the test data set) is a difficult task. Therefore, the first performance measure is the success rate. The test success rate is not the classification accuracy but the percentage of times that the test profit is higher than the market profit. The second performance measure is the test average upside. The test upside is the excess (lacking) return above (below) the market when the method achieves above market return (loses).

### A. *Experimental Setting*

The training and test partition is as follows. Out of 40 periods, the first 25 to 35 periods are partitioned into the training set randomly with discrete uniform distribution. The remaining periods are partitioned into test data. Momentum learning is used between generations The momentum learning formulation is presented in equation (4).

$$\theta_n = \beta * \theta_n + (1-\beta) * \theta_{n-1} \qquad (4)$$

The classifiers used in this setting are linear, polynomial (with 4th-degree polynomial weight at most), Nearest-Centroid (NC) with rectilinear, Euclidean and Tchebycheff distances. The features with a higher than 0.9 correlation with each other are removed to avoid multicollinearity problem. In the following subsection, summarized results are given, compare and commented.

### B. *Experimental Results Discussion and Comparison*

The success rate is the percentage of times that the method achieves above-index returns. The Base setting performs below 50% success rate with LDF and PDF methods that is those methods are profitable less than half of the replications. NC method achieves a 100% success rate with the rectilinear and Euclidean distances. NC with Tchebycheff distance is profitable for 60% of the replications. The Extended setting significantly improves the success rate. It improves the success rate from 40% to 50%, 40% to 56% and 60% to 100% for LDF, PDF and NC with Tchebycheff respectively. The performance of the NC with the rectilinear distance decreases insignificantly. The results are given in Table 2.

Table 2: The success rate and average upside of the methods under the Base and Extended settings

| | Success Rate | | Average Upside | |
|---|---|---|---|---|
| **Methods** | **Base** | **Extended** | **Base** | **Extended** |
| LDF | 40.0% | 50.0% | -2.3% | 0.0% |
| PDF | 40.0% | 56.7% | -0.6% | **-0.7%** |
| NC with Rectilinear | 100.0% | **96.7%** | 28.9% | 34.4% |
| NC with Euclidean | 100.0% | 100.0% | 41.9% | 50.0% |
| NC with Tchebycheff | 60.0% | 100.0% | 9.5% | 32.0% |

Similar to the improvements in the success rate, the average upside improves with the Extended setting. Although the success rate decreases by 3.3% for NC with the rectilinear distance the average upside improves by 5.5%. The loss in the average upside is 0.01% for the PDF method. All of the other methods improve the average upside when the Extended setting is applied.

With 5 methods and 30 replications for each method, 150 results are compared with Paired T-Test. Figure 3 presents the comparison of the 150 results for the Base and the Extended settings with boxplot and interval plots (95% confidence interval).

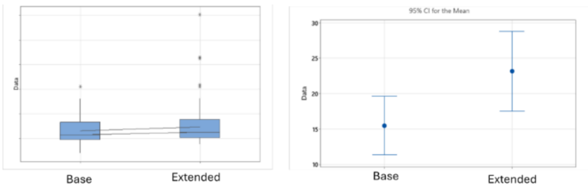


Fig. 3: Boxplot and interval plot of the Base and Extended settings

Figure 4 presents the T-test results of the comparison. According to the overall comparison of the 150 experimental results (with a 95% confidence interval), the Extended setting outperforms the Base setting. This indicates that the consideration of the comparable peer problem improves the success of the classification methods under consideration.

**Descriptive Statistics**

| Sample | N | Mean | StDev | SE Mean |
|---|---|---|---|---|
| Extended | 150 | 23.16 | 34.86 | 2.85 |
| Base | 150 | 15.47 | 25.62 | 2.09 |

**Estimation for Paired Difference**

| Mean | StDev | SE Mean | 95% Lower Bound for μ_difference |
|---|---|---|---|
| 7.69 | 26.71 | 2.18 | 4.08 |

*μ_difference: population mean of (Extended - Base)*

**Test**

Null hypothesis H₀: μ_difference = 0
Alternative hypothesis H₁: μ_difference > 0

| T-Value | P-Value |
|---|---|
| 3.53 | 0.000 |

Fig. 4: Paired T-test results of the comparison of the Base and Extended settings

## VI. CONCLUSION

GA is a population-based algorithm that enables ensemble learning to perform portfolio allocation without a portfolio allocation method. Narrow period selection and proper objective selection problems are handled well with above-market returns for NC method, PDF and LDF underperform the market. Involving comparable peer selection problem improved the model performances. The method is specifically designed for mid and long-term investments. It can be extended for short-term investing or trading with different metrics. Class cardinality problem is accounted to index funds and risk-free investments. A multi-objective optimization approach is utilized for fitness function. The recommended setting is to consider the comparable peer problem and utilize the multiple-period selection and multi-objective optimization approach (the Extended setting). Of the methods utilized, we recommend using NC with Euclidean distance since it is the best for the success rate and average upside performance measures. Other classification approaches must be considered under the Extended setting for future work. Also, NC with other distance functions can be analyzed.

## Appendix A

The list of the DJI 30 tickers (companies) involved in this study are: 'UNH', 'MSFT', 'GS', 'HD', 'MCD', 'CAT', 'V', 'AMGN', 'CRM', 'BA', 'HON', 'AAPL', 'AXP', 'TRV', 'JNJ', 'CVX', 'WMT', 'PG', 'JPM', 'IBM', 'MRK', 'NKE', 'MMM', 'DIS', 'KO', 'CSCO', 'VZ', 'INTC', 'WBA'.

## Appendix B

Market Capitalization (MC), Price to Earnings ratio (P/E), Price to Book value ratio (P/B), Quick Ratio (QR), Debt to Equity ratio (D/E), Current Ratio (CR), Net Profit Margin (NPM%), Return on Assets (ROA), Return on Equity (ROE), Return (R), Risk (V: variance), Beta (systematic risk, abbr. B)

## References

Doumpos, Michael, and Constantin Zopounidis. "Multi–criteria classification methods in financial and banking decisions." International Transactions in Operational Research 9, no. 5 (2002): 567-581.

"Free Online Stock Information for Investors." Stock Analysis. Accessed November 13, 2023. https://stockanalysis.com/.

Geertsema, P., & Lu, H. (2023). Relative Valuation with Machine Learning. Journal of Accounting Research, 61(1), 329-376.

Goudarzi, S., Jafari, M. J., & Afsar, A. (2017). A hybrid model for portfolio optimisation based on stock clustering and different investment strategies. International Journal of Economics and Financial Issues, 7(3), 602-608.

Huang, W., Nakamori, Y., & Wang, S. Y. (2005). Forecasting stock market movement direction with support vector machine. Computers & operations research, 32(10), 2513-2522.

Khedmati, M., & Azin, P. (2020). An online portfolio selection algorithm using clustering approaches and considering transaction costs. Expert Systems with Applications, 159, 113546.

Plenborg, Thomas, and Rene Coppe Pimentel. "Best practices in applying multiples for valuation purposes." The Journal of Private Equity 19, no. 3 (2016): 55-64.

"Yahoo Finance - Stock Market Live, Quotes, Business & Finance News." Yahoo! Finance. Accessed November 13, 2023. https://finance.yahoo.com/.

Zopounidis, C., Doumpos, M., & Zanakis, S. (1999). Stock evaluation using a preference disaggregation methodology. Decision Sciences, 30(2), 313-336.