ISSN 1816-6075 (Print), 1818-0523 (Online)
Journal of System and Management Sciences
Vol. 13 (2023) No. 5, pp. 260-275
DOI:10.33168/JSMS.2023.0517

Portfolio Construction of Good Defensive Malaysian Shariah-Compliant Stocks Using Data Mining Techniques

Nur Sara Zainudin[1, *], Keng-Hoong Ng[2], Choo-Yee Ting[1], Kok-Chin Khor[3], Gee-Kok Tong[1], Suraya Nurain Kalid[1]

[1]Faculty of Computing & Informatics, Multimedia University, Cyberjaya, 63100, Malaysia
[2]School of Computing, Asia Pacific University of Technology and Innovation, Kuala Lumpur, 57000, Malaysia
[3]Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Kajang, 43000, Malaysia

1171100244@student.mmu.edu.my; ng.kenghoong@apu.edu.my; cyting@mmu.edu.my; kckhor@utar.edu.my; gktong@mmu.edu.my; suraya.nurain@mmu.edu.my

Abstract. Investors, particularly Muslims who constitute a significant portion of Malaysia’s population, seek Shariah-compliant stocks that provide stable returns and exhibit defensive behaviour during economic downturns. However, the analysis of these stocks can be challenging due to the large number of Shariah-compliant stocks listed on Bursa Malaysia. This could lead to investors refraining from investing in these stocks due to a lack of financial knowledge and time. To address this issue, this study proposes a practical approach that employs data mining techniques to aid stock portfolio construction. The approach uses the Beta coefficient to identify less volatile Shariah-compliant stocks on Bursa Malaysia from 2018-2021. The study then utilises *k*-Means clustering to group stocks with similar financial characteristics and selects well-performing clusters based on their financial performance. The researchers then form equal-weighted portfolios using the stocks frequently selected as members of the well-performing clusters and evaluate their performance by comparing their returns with various sectors’ price indexes. The results show that most stock portfolios outperformed the indexes. The study highlights the importance of data mining techniques in identifying good Shariah-compliant stocks and forming portfolios. Furthermore, the study provides a practical solution to investors looking to invest in Malaysian Shariah-compliant stocks.

Keywords: Stock profiling, Defensive Shariah-compliant stock, Beta coefficient, *k*-means clustering.

1. Introduction

Recently, the Malaysian government has encouraged citizens to invest in stock. It is known that 63.5% of Malaysian citizens are Muslim (Department of Statistics Malaysia, 2022), and they make up the majority of potential investors in Malaysia. Due to a large number of Muslim investors in Malaysia, there is a demand for investing in Shariah-compliant stocks. Shariah-compliant stocks are known to be involved in socially responsible business, which improves social justice, the environment, and clean technology (Chee, 2019). The stock must also abide by the Shariah law and Islamic principles, such as prohibiting interests or products/services with gambling, alcohol, and tobacco. Shariah-compliant stocks are usually less risky and stable because of their low reliance on debt, net leverage, and intangible assets (Cheong, 2021). Previous studies have shown that Shariah-compliant stocks tend to be less volatile than conventional stocks, even when facing a global financial crisis (Abduh, 2020).

Stock investments aim to obtain income through dividend payments or capital appreciation. However, an individual must possess good financial knowledge to conduct extensive research to identify stocks that could yield lucrative returns. As of 2022, more than 900 common stocks, including 783 Shariah-compliant stocks, are listed on Bursa Malaysia (Securities Commission Malaysia, 2020). As a result of the immense stock data, many amateur investors often find difficulties coming up with good investment decisions. Though they have extra cash in hand and wish to begin investing, they are refrained from doing so because lacking relevant financial knowledge and time to analyse them. In Malaysia, experienced investors commonly apply either (1) fundamental analysis that determines the real stock value or (2) technical analysis that observes stock patterns to find investment opportunities (Naz´ario et al., 2017). It requires in-depth knowledge of both analyses to gain insight into additional stock details, such as their growth, defensive, aggressive, and cyclical characteristics.

Financial Technology (FinTech) today is employing data mining techniques to analyse the stock market (Gai et al., 2018). Although FinTech has entered the Islamic financial market in Malaysia, it is still in its infancy stage (Abdullah, 2017). Currently, very little research is conducted to identify defensive Malaysian Shariah-compliant stocks using data mining techniques. Further, analysing such stocks requires Herculean efforts since the number of Shariah-compliant stocks listed on Bursa Malaysia is large.

Researchers formed stock portfolios from stock markets using various clustering methods. However, research in the context of defensive Shariah-compliant stocks is lacking, especially ones from the Malaysian stock market using defensive stock-related financial ratios. The paper presents a practical approach that uses data mining techniques to form stock portfolios consisting of good defensive Shariah-compliant stocks from Bursa Malaysia that can provide stable returns and exhibit defensive behaviours during economic downturns. The approach requires minimal user input and financial literacy from investors and can easily identify good defensive Shariah-compliant stocks. The research outcome shall promote stock investment among Malaysians, particularly Muslims and possibly provide personal financial stability by increasing their passive income. The research also aligns with the Malaysian government's initiative to promote and encourage more Malaysians (particularly Muslims) to invest in the stock market.

The paper is divided into five sections. Section 2 discusses the related research. Section 3 details the study's methodology, and Section 4 evaluates the resulting output. Finally, the research is concluded in Section 5.

2. Related Works

We first discuss existing labelling methods for defensive stocks since our study aims to identify good defensive stocks. We then review clustering algorithms which are used to group similar objects. Our study applies a clustering algorithm to stock data to form clusters. Descriptive analysis is crucial in describing the financial ratios of each cluster formed. The selected stocks from the proposed method are used to form stock portfolios then compare their performances with the sectors' indexes.

2.1. Labelling Defensive Stocks

Defensive stocks are generally less volatile in price movement. Different measures are used to judge the volatility of stocks, such as the Beta coefficient, standard deviation, and R-squared (Chaudhary et al., 2020; Cortes and Weidenmier, 2019; Rout and Panda, 2020). A common method is calculating the Beta value of stocks and then labelling them according to the computed value, whereby stocks with Beta value < 1 are defensive and stocks with Beta value > 1 are cyclical (Bardh and Haglund, 2021; Retha and Budiarti, 2021; Yelamanchili, 2019). Other than using the Beta coefficient, some researchers classify stocks based on Morgan Stanley Capital International's (MSCI) Cyclical and Defensive Sectors Indexes, which were developed to gauge defensive/cyclical stock performance (Asinas, 2018; Azam et al., 2022). MSCI Cyclical and Defensive Sectors Indexes classify sectors instead of individual companies (Morgan Stanley Capital International, 2022). For example, the Consumer Staples sector is defensive, whereas the Consumer Discretionary sector is cyclical. The sectors listed by both indexes follow the Global Industry Classification Standard (GICS), which slightly differs from the sector classification in Bursa Malaysia. The sector classification in Bursa Malaysia is not uniform with the sectors listed in GICS, for example the Consumer Products & Services sector in Bursa Malaysia were split into Consumer Staples sector and Consumer Discretionary sector in GICS. Due to this issue, directly using the Beta value of stocks to label Malaysian defensive stocks is suitable for this study to ensure labelling consistency.

2.2. Clustering Methods for Grouping Similar Stocks

Clustering methods such as k-Means clustering, Fuzzy clustering, and *k*-Medoids clustering were used on stock data to group similar stocks or sectors. Among the clustering methods, *k*-Means clustering is a popular method applied in research. For example, a group of researchers proposed a portfolio optimization strategy using *k*-Means clustering (Kedia et al., 2018). For stock creation, they used Indian stock data and relevant investment parameters, such as financial ratios and liquidity ratios. The generated portfolio's rate of return after a year was 16.21 per cent greater than the Sensex and 16.89 per cent greater than the BSE100.

Another research (Korzeniewski, 2018) applied the *k*-Means clustering and Partitioning Around Medoids (PAM) for portfolio construction. The data set consists of time series data concerning 85 companies from 1st February 2011 to 30th April 2016. Although the PAM algorithm obtained lower returns than *k*-Means clustering, the results for both methods were satisfactory as the return rate was better than the market. In the study by Buszko et al. (2021), *k*-Means and Ward techniques were used to investigate the stability and resistance of different industries towards the Coronavirus Disease 2019 (COVID-19) pandemic. The data set consists of stock data from 16 sectors and from different periods, of which six variables were related to price stability, volume, oversold conditions, and volatility. Both techniques obtained the identical mean silhouette coefficient. The cluster consisting of chemical, energy, mining, and oil and gas sectors outshined the other clusters in overall stability for all periods.

Aside from *k*-Means clustering, other algorithms were utilised in clustering financial data, such as the study by Zainol Abidin et al. (2020) which used Fuzzy clustering to perform stock selection derived from the investors' preferences and confidence levels. The data of 30 Shariah-compliant stocks, including their return rates, standard deviations, and Treynor index values, were used in the study. The evaluation was conducted by calculating the correlation between the actual stock performance ranking and the clustering results. Their proposed Fuzzy clustering surpassed k-Means clustering and Hierarchical clustering as it produced greater correlation values.

Other examples include another research which applied the subtractive clustering-based adaptive neuro-fuzzy approach to predict stock prices (Chandar, 2019). The data set consisted of Apple stock from 3rd January 2005 to 30th January 2015, which was used by the algorithm to predict the opening price of Apple's stock data for the following day. The proposed approach performed better than adaptive neuro-fuzzy inference system (ANFIS) training and Subtractive clustering in training and testing. Similarly, Nakagawa et al. (2019) used *k*-Medoids clustering on dissimilarity matrix with Indexing Dynamic Time Warping (IDTW) method for stock price prediction. Financial time series data from January 1989 to March 2017 was used for the prediction. As a result, *k*-Medoids clustering with IDTW obtained a higher prediction accuracy than the Auto-Regressive model, *k**-NN, and *k*-Medoids clustering with DTW.

We observed that *k*-Means is the most extensively used clustering method and has been used in many research works (Zhu and Liu, 2021). *k*-Means clustering has many advantages, such as easy implementation and the ability to easily scale to large data sets and adapt to new examples (Google Developers, 2022). The ability of k-Means clustering proves that it is a reliable method for clustering stock data and is suitable for our study. To our best knowledge, it has not been used to identify defensive Shariah-compliant stocks listed on Bursa Malaysia.

2.3. Elbow Method in Clustering

k-Means clustering requires the number of *k* clusters to be defined beforehand (Hussein et al., 2021). One can solve the problem of defining the *k* value by comparing the performance of the clustering algorithm using a range of *k* values. There are several optimisation methods to find the optimal number of clusters such as the Elbow method, Silhouette Coefficient, and Calinski-Harabasz index.

The Elbow method is a common cluster optimisation method suitable for finding a small range of k number of clusters (Cui, 2020). The *k* selection uses a graph showing the difference of every cluster's sum of square error (SSE), where the point forming the elbow angle in the graph represents the optimal number of *k* clusters (Umargono et al., 2020). It is a simple but effective way of selecting the number of clusters through visualisation and was implemented in this research.

2.4. Descriptive Analysis and Stock Portfolio Evaluation Methods

One may study the descriptive analysis (minimum, maximum, average, quartiles, etc.) to understand the characteristics of stocks (Carter et al., 2022; Malaysia et al., 2022; Sholichah et al., 2021). The study by Carter et al. (2022) used descriptive analysis to analyse the stock market performance of different industries during the COVID-19 pandemic, namely, airline, hotel, and tourism. On the other hand, Sholichah et al. (2021) also used it to describe the sample stocks for each financial variable. Research by Malaysia et al. (2022) utilised descriptive analysis and box plots to interpret and visualise the data effectively. They were used to compare the opening prices of Malaysian stocks. Essentially, box plots are elegant for visualising the 5-number summary of data as a straightforward method to compare data (Potter et al., 2006). Such visualisation is appropriate for comparing the financial ratios of defensive stocks against other stocks in our study.

As for evaluation, there are various portfolio performance measures: Sharpe ratio, Treynor Index, and Jensen Index factor in risk, whereby a higher value generally means a better-performing portfolio (Hertina et al., 2021). These measures are built under different assumptions: the Sharpe ratio assumes the distribution of the returns is normal, Treynor Index is under the assumption that the stocks are greatly diversified, and the Jensen Index assumes that a constant rate of returns is estimated using the Jensen Alpha rate of return (Hertina et al., 2021). Besides this, some researchers evaluated portfolio performance by comparing the returns of the portfolio against the returns of benchmarks, such as the BSE100 index and OMX Stockholm Benchmark GI index (Bardh and Haglund, 2021; Kedia et al., 2018). When calculating the returns of a stock portfolio, the investor must consider the weightage of each stock in a portfolio, and equal-weighted portfolios are commonly used (Jiang et al., 2019). Such a portfolio is straightforward and easily understood from a layman's point of view.

After reviewing the previous research, we decided to filter stocks using the Beta coefficient to obtain defensive stocks instead of using MSCI's Cyclical and Defensive Sectors Indices. The Beta coefficient ensures consistency in labelling defensive stocks since the stock sector classification in Bursa Malaysia does not conform to the GICS used in MSCI's index. We were also driven to use *k*-Means clustering, a popular method with many benefits like easy implementation and its ability to scale to large data sets. Likewise, box plots were selected to describe clusters and study the financial ratios as it deduces the 5-number summary of data visually in a straightforward manner. The stock portfolios shall be evaluated by comparing the returns of equal-weighted portfolios and their corresponding price index returns. As per the above justification, the proposed method applied in the study shall be reflected in the next section.

3. Methodology

The study follows a series of steps, which can further be divided into three main components: (i) data preparation and preprocessing, (ii) *k*-Means clustering, and (iii) stock portfolio construction and evaluation.

3.1. Data Preparation and Preprocessing

We collected data from 2871 Shariah-compliant stocks listed on Bursa Malaysia using DataStream. The data contains financial data of the stocks from 2018 to 2021, with 703, 713, 726, and 729 stocks, respectively. The stocks are from 12 different sectors: (i) Consumer Products & Services, (ii) Industrial Products & Services, (iii) Property, (iv) Construction, (v) Health Care, (vi) Technology, (vii) Telecommunications & Media, (viii) Transportation & Logistics, (ix) Plantation, (x) Utility, (xi) Energy, and (xii) Financial Services.

Table 1: The Four Financial Ratios Used in This Study.

Financial Ratio	**DataStream Datatype**	**Formula**
Dividend Yield	DY	Dividend per Share / Price per Share
Price to Book Value	PBV	Stock Market Price / Book Value per Share
Price-earnings Ratio	PE	Price per Share / Earnings per Share
Beta Value	WC09802	Covariance (Re, Rm) / Variance (Rm), Re = return on the stock & Rm = return on the market

The financial ratios in the data set are shown in Table 1, with their corresponding DataStream data type (Institute of Financial Management, 2017; Financial, 2007). The financial ratios extracted from DataStream are deemed important when considering defensive stocks. Dividend yield (DY) is the amount a company pays its dividends to shareholders annually and can be in the form of cash, stocks, or other assets (Arslan et al., 2014). Price-earnings (PE) ratio represents the relationship between stock price and earnings per share (EPS). A low PE ratio indicates that the stock is low in price and promises a high return in future (Doblas et al., 2020). Price to book value (PBV) compares the market price to the book value of shares, whereby a high PBV indicates good company growth (Kusmayadi et al., 2018). However, another research states that low PBV stocks are undervalued, so the general rule is to invest in such stocks since the prices will increase once the value rises (Doblas et al., 2020). Beta value describes the volatility of the stock price compared to the stock market, whereby a stock with a Beta value of less than 1 shows its defensive quality (Asinas, 2018).

In this study, the negative values of the Price-earnings Ratio in the data were set to zero. Certain companies' data were unavailable and not found from other sources, such as the Bursa Malaysia portal. Hence, we removed those companies' data. Algorithm 1 details the data preparation and preprocessing steps.

Algorithm 1: Data Preparation and Preprocessing

```
for all Shariah-compliant stocks listed on Bursa Malaysia do
    Collect financial ratios of stock from DataStream
    if stock has incomplete data then
        Delete stock data from the data set
    end
    if stock is beta < 1 then
        Keep stock data for clustering to get companies with low volatility in price
    else
        Filter out stock data
    end
    Separate stock with low volatility in price based on sector and year
    Normalise data with min-max normalisation
end
Return preprocessed data
```

Before proceeding to the clustering phase, we filtered the whole data set to contain only companies with a Beta value of less than 1. The Financial Services sector contains 24 companies. However, only two companies remain after filtering using the Beta value < 1. The Energy sector, on the other hand, contains four companies in total, and two remain after filtering. Thus, the initial 12 sectors in the data set were reduced to 10 since the Energy and Financial Services sectors did not contain enough data to form clusters. Finally, the company data was then separated into data frames based on sector and year.

The range of values was different for the financial ratios. So, we applied data min-max normalisation to ensure each variable carried the same value range [0,1] and would not outweigh those with smaller ranges when clustering was applied.

3.2. *k*-Means Clustering

k-Means clustering is one of the most widely used unsupervised clustering algorithms. Its purpose is to group stocks with similar characteristics as a precursor to profiling stocks based on their financial ratios. The implementation of *k*-Means clustering, and the Elbow method are detailed in Algorithms 2 and 3. Initially, *k* initial centroids were set with *k* ranged [1,10], and nearby points were assigned to the nearest centroid to form clusters. The centroids of each cluster were then updated according to the mean of all points in said cluster. The clustering process repeated until the clusters faced no more changes (Wu, 2012). The number of *k* clusters must be defined as the input when using k-Means clustering (Ernawati et al., 2022). We used the Elbow method to find the optimal number of k clusters.

Algorithm 2: *k*-Means Clustering

```
// k is the number of clusters
Set k number of clusters
Choose k centroids at random
repeat
    for each point do
        Reassign the point to the cluster with the closest centroid
    end
    for each k cluster do
        Recalculate new centroid (mean)
    end
until convergence criteria are fulfilled
```

Algorithm 3: Generate Optimal *k* Clusters

```
// k is the number of clusters
repeat
  for different k values do
    Apply k-Means clustering on data
  end
  Use Elbow Method to find optimal k
  Generate k clusters with k-Means clustering
until loop through each sector by year
Return generated k clusters
```

3.3. Stock Portfolio Construction and Evaluation

The equal-weighted stock portfolio was constructed then evaluation was conducted as shown in Algorithm 4. We determined the well-performing cluster based on the average financial ratios of each cluster. High DY and low Beta values were the main criteria when screening the clusters. Other considerations were low PBV and low PE. For risk-averse investors, stocks with high DY and low Beta are preferable because they provide steady income and have less price fluctuation. The aforementioned reasoning leads us to prioritise these two financial ratios when selecting well-performing defensive clusters. PE ratio and PBV are relative valuation metrics that compare a stock's value against its peers in the industry (Agudze & Ibhagui, 2020). This information is a good indication of a stock's 'health'. Low PE ratio and PBV stocks were chosen because they are considered undervalued. Risk-averse investors favoured these ratios since they are associated with less risk (Svanberg & Max, 2019).

Algorithm 4: Stock Portfolio Construction and Evaluation

```
for each sector by year do
  Determine the best cluster(s) based on their financial ratios
end
for each sector do
  Combine companies' data from the best clusters for all years by sector
  Display box plots for combined data
  Analyse the defensive companies using results
  Create frequency table based on number of times a stock was in the selected
  clusters
  Construct equal-weighted portfolio with frequently selected stocks
  Compare portfolio returns with benchmark price indexes
end
```

Following the selection of the well-performing clusters, data of the selected clusters for all years were combined based on sector. The box plots of the combined data for each sector were displayed in a figure. The box plots show the minimum value, first quartile, median, third quartile, maximum, and average value for every financial ratio involved. The main goal is to profile the stock clusters based on their similar financial characteristics.

We then created an equal-weighted portfolio based on the frequency a stock was being selected from 2018 to 2021. The assumption is that stocks with high frequency are showing defensive behaviour more consistently. The evaluation was done by calculating the portfolio's returns and comparing it to the price index's return of its corresponding sector from Bursa Malaysia. The equal-weighted portfolio return was calculated as per Equation 1.

$$Return_p = \sum W_i \times \left(\frac{P_{t+1} - P_t}{P_t}\right)_i \times 100\% \tag{1}$$

The portfolio return was the sum of the product between the weight and the stock returns multiplied by 100%. The weight (W_i) was the fraction of the number of stocks in the portfolio, whereas the return was calculated by subtracting the stock price from different dates ($P_{(t+1)}$, P_t) divided by the price during the earlier date (P_t).

4. Results and Discussion

We filtered Shariah-compliant stocks listed on Bursa Malaysia from 2018-2021 using a Beta value of < 1. The stock numbers remaining were 217, 221, 224, and 225 stocks from 2018-2021. Hence, the stock data from ten sectors were used to perform the steps in the proposed methodology of this study.

4.1. Clustering Results

Clusters were generated after applying k-Means clustering to the stocks' data. We determined the best *k* clusters for all ten sectors using the Elbow method. For instance, Table 2 shows that the optimal *k* cluster for the Health Care sector were 5, 4, 3, and 4 from 2018-2021 (a total of 16 clusters). We cannot show all the clustering results in this paper because there were 40 clustering results. Thus, only several clustering results are shown (Tables 3-6).

Table 2: The optimal *k* number of clusters for the sectors and the corresponding years.

	2018	**2019**	**2020**	**2021**
Consumer Products & Services	5	4	5	5
Industrial Products & Services	5	6	4	5
Property	4	5	4	4
Construction	4	5	5	4
Health Care	5	4	3	4
Technology	4	4	4	5
Telecommunications & Media	4	4	4	4
Transportation & Logistics	4	5	4	5
Plantation	4	4	4	4
Utility	5	4	4	6

For instance, k-Means generated four clusters for the Technology sector using 2018 and 2020 data (Tables 3-4). The resulting clusters differ in their cluster members due to the difference in the financial characteristics of stocks each year. However, there were some constants, i.e., 0097.KL, 0021.KL, and 7022.KL (bolded stock codes in the tables) were grouped into the same cluster for both years by the clustering algorithm.

Table 3: The Clustering Result for the Technology Sector Year 2018.

Cluster	**Stock Code**
0	0166.KL, 3867.KL, 5005.KL, 0083.KL, 0040.KL
1	0068.KL, 0191.KL, 0107.KL, 9393.KL, 0113.KL, 0056.KL, 7160.KL, 4359.KL, 0086.KL
2	0128.KL, 0041.KL, 9334.KL, 0132.KL
3	**0097.KL, 0021.KL, 7022.KL**

Table 4: The Clustering Result for the Technology Sector Year 2020.

Cluster	**Stock Code**
0	0041.KL, 0083.KL, 0132.KL, 4359.KL
1	3867.KL, 5005.KL, 0191.KL, 0113.KL, 0056.KL, 0040.KL
2	0128.KL, 0166.KL, **0097.KL, 0021.KL, 7022.KL**, 9334.KL
3	0068.KL, 0107.KL, 9393.KL, 7160.KL, 0086.KL

Table 5: The Average Financial Ratios of the Resulted Clusters from the Property Sector Year 2019. Cluster 0 is a well-performing cluster based on its financial performance.

Cluster	DY	PE	Beta	PBV
0	**8.165**	**4.4**	**0.165**	**0.685**
1	0.536	3.72	0.877	0.525
2	4.253333	13.722222	0.662222	0.582222
3	0.459231	9.230769	0.406154	0.651538
4	0	76.7	0.15	2.55
Average	2.043218	8.791954	1.107471	0.607816

Table 6: The Average Financial Ratios of the Resulted Clusters from the Consumer Products & Services Sector year 2018. Cluster 2 and Cluster 4 are well-performing clusters based on their financial performance.

Cluster	DY	PE	Beta	PBV
0	0.907692	14.007692	0.263077	1.573846
1	1.204211	11.189474	0.735789	0.753158
2	**4.7675**	**9.525**	**0.3875**	**1.3**
3	1.62	58.333333	0.383333	31.356667
4	**4.99**	**12.071429**	**0.821429**	**1.237143**
Average	2.389037	12.165185	0.977037	1.999037

After forming the clusters, the average values of every financial ratio for each cluster were calculated, as shown in Tables 5-6. The well-performing cluster was determined based on its financial performance. As shown in Table 5, Cluster 0 (bolded) was selected and determined to be the well-performing cluster of the Property sector for 2019. The cluster showed the highest average DY (8.165) compared to the other clusters while retaining a low average Beta value (0.165). Furthermore, the cluster also possessed low average PE (4.4) and medium average PBV (0.685).

The same criteria for selecting the best cluster were also applied to the other sectors. However, there was an exception for the Consumer Products & Services sector in 2018, as it was the only case with two selected well-performing clusters (Table 6). Their average DY (Cluster 2 = 4.7675 & Cluster 4 = 4.99) were similar and significantly higher than the other clusters. Their average PBVs were also higher than 1, meaning that the market value is above the accounting value. Furthermore, their average Beta values were lower than 1, thus showing defensive quality. In addition, their average PE (Cluster 2 = 9.525 & Cluster 4 = 12.071429) were also not high. Overall, both clusters were quite similar and were selected due to their good financial performance.

4.2. Analysing Stocks from the Selected Clusters with Box Plots

The well-performing clusters of every sector for all four years (2018-2021) were combined to form a defensive stock pool. We then used box plots to describe the overall financial characteristics of the stocks in the clusters and compared them with the overall stocks. Two box plots were used to represent each financial ratio of (1) overall stocks regardless of Beta values (denoted as Overall) and (2) stocks from the selected well-performing clusters (denoted as Selected) (see Fig. 1). The red line of the box plots represent the median value, whereas the diamond shape denotes the average value. The extreme outliers obscuring the box plot details were hidden in the following box plots.

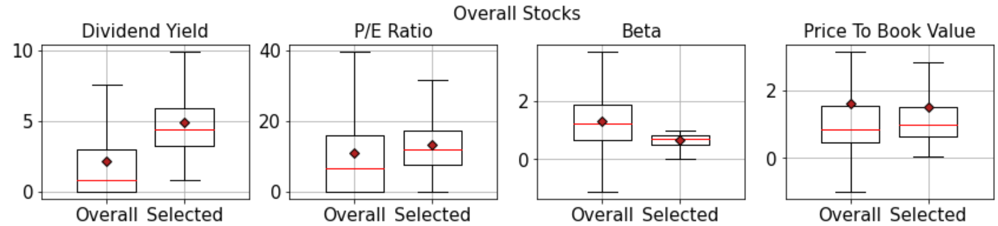

Fig. 1: The Selected Defensive Stocks vis-à-vis Overall Stocks regardless of Beta values.

All the generated box plots can be seen in Fig. 2 and Table 9 of Appendix A. Most box plots of the sectors exhibited similar results. To generalise the findings, the average DY, PE, and PBV of the selected stocks were higher than the overall stocks. As for Beta value, the selected stocks always had a lower average value than the overall stocks. It shows that the stocks chosen from the clustering process were defensive.

4.3. Selecting Stocks from the Selected Clusters to Form Portfolios

Table 7 shows the number of times a stock was selected as a member of a well-performing cluster throughout 2018-2021. The list of stocks includes the Consumer Products & Services sector (CONSU), Industrial Products & Services sector (INDUS), Property sector (PROP), Technology sector (TECH), Health Care sector (HEALTH), Transportation & Logistics sector (TRANS), Construction sector (CONST), and Utility sector (UTIL). Only the top 10 were shown in the table. We excluded the Telecommunications & Media sector and Plantation sector since none of their stocks exceeded a sum of one. Table 7 shows that many stocks were selected for several years. For instance, 7617.KL and 6017.KL were selected as a member of well-performing clusters for all four years from 2018 to 2021. The higher the frequency, the more likely a stock is defensive.

The results from Table 7 were used to produce equal-weighted stock portfolios, whereby the weight of investment for each stock was equal, based on the number of times a stock is selected. We set the threshold to three and above, resulting in 22 stocks from eight sectors. We then compared these portfolios' returns against the corresponding sectors' price index returns of Bursa Malaysia. The data used was from 3rd January 2022 to 30th June 2022 (first half of 2022).

The price of stocks listed on Bursa Malaysia showed a decreasing trend since the start of 2022, but the returns of the equal-weighted portfolios were better than the benchmark sectors' price index returns (Table 8). Though some of the portfolios' returns were negative, the losses were still lower than the corresponding benchmark price indexes, i.e., the Industrial Products & Services and the Technology sectors.

Table 7: The Frequency (Freq) a Stock was Selected as a Member of a Well-performing Cluster.

CONSU Sector	**Freq**	**INDUS Sector**	**Freq**	**PROP Sector**	**Freq**	**TECH Sector**	**Freq**
9997.KL	3	7229.KL	3	7617.KL	4	5005.KL	3
5131.KL	3	7248.KL	3	6017.KL	4	0040.KL	3
9369.KL	3	7073.KL	3	5200.KL	3	0166.KL	2
7089.KL	3	5143.KL	2	5049.KL	3	3867.KL	2
7202.KL	3	7232.KL	2	5236.KL	3	0083.KL	2
6939.KL	3	1368.KL	2	5020.KL	2	0191.KL	2
5160.KL	2	7034.KL	2	5084.KL	2	0113.KL	2
7082.KL	2	0054.KL	2	8494.KL	2	0056.KL	1
7134.KL	2	7100.KL	1	5789.KL	2	7022.KL	1
7237.KL	2	8125.KL	1	5827.KL	1	5286.KL	1
HEAL Sector	**Freq**	**TRANS Sector**	**Freq**	**CONST Sector**	**Freq**	**UTIL Sector**	**Freq**
7148.KL	3	6645.KL	4	9598.KL	4	8524.KL	3
0002.KL	3	5246.KL	2	8311.KL	2	5264.KL	2
7178.KL	3	3816.KL	2	0206.KL	1	5347.KL	1
03023.KL	2	5032.KL	1	5006.KL	1	5209.KL	1
5878.KL	1	5140.KL	1				

Table 8: The Returns of the Equal-Weighted Portfolios are Better than The Returns of the Price Indexes, except for the Construction sector.

Sector	**Equal-Weighted Return**	**Price Index's Return (Bursa Malaysia)**
Consumer Products & Services	6.66	-3.73
Industrial Products & Services	-7.52	-9.78
Property	8.27	-9.03
Construction	-13.94	0.96
Health Care	1.16	-26.05

Technology	-24.04	-35.84
Transportation & Logistics	32.81	-6.18
Utility	9.05	-4.23

However, there was one exception. The stocks selected from the Construction clusters made more losses than the price index. 9598.KL was the sole stock chosen to form the portfolio, and the stock suffered losses in the first half of 2022. The company was possibly not doing well, causing project delays due to the COVID-19 pandemic (Zakaria and Singh, 2021). Furthermore, as this only concerns a single stock, the results showed the importance of portfolio diversity as it shall reduce unsystematic risk and the portfolio's volatility (Zaimovic et al., 2021). The overall results were satisfactory since most portfolios' returns were higher than the sectors' price index returns.

5. Conclusions

The study proposed a practical portfolio construction that utilises k-Means to cluster defensive Shariah-compliant stocks (filtered using Beta Coefficient) listed on Bursa Malaysia. After clustering stocks with similar financial characteristics, we determined the best clusters per sector and analysed their financial ratios using box plots. The generalised results of the research were that the selected defensive stocks often exhibit higher average dividend yield, PE ratio, and PBV compared to the overall stocks in the same sector.

We created equal-weighted stock portfolios based on the times a stock was selected from 2018 to 2021. The portfolios outperformed each corresponding sector's price index in returns except for the Construction sector. The results prove the superiority of our proposed approach in identifying good defensive Shariah-compliant stocks.

Investors may have different views when considering whether a financial ratio of a particular sector is low, medium, or high in value. This study is limited by the manual selection of the best defensive stock clusters. In this study, we determine the best defensive stock cluster by comparing the average financial ratios among the generated clusters. In the future, we may consider automating the identification of good defensive Shariah-compliant stocks using a classification method.

Acknowledgments

We are grateful for the project funding provided by the Ministry of Higher Education Malaysia under the Fundamental Research Grant Scheme grant number FRGS/1/2020/SS01/MMU/03/1.

References

Agudze, K. & Ibhagui, O. (2020). Do fundamentals drive relative valuation? Evidence from global stock market indices. *Journal of Financial Management, Markets and Institutions*, 8, 2050002.

Abduh, M. (2020). Volatility of Malaysian conventional and Islamic indices: does financial crisis matter?. *Journal of Islamic Accounting and Business Research*, 11, 1-11.

Abdullah, O. (2017). *Impact of fintech to islamic finance*. The Malaysian Reserve. https://themalaysianreserve.com/2017/04/03/impact-of-fintech-to-islamic-finance/.

Arslan, M., Zaman, R. & Phil, M. (2014). Impact of dividend yield and price earnings ratio on stock returns: A study non-financial listed firms of Pakistan. *Research Journal of Finance and Accounting*, 5, 68-74.

Asinas, M. A. S. (2018). Stock market betas for cyclical and defensive sectors: A practitioner's perspective. *Philippine Management Review*, 25, 99-114.

Azam, M. Q., Hashmi, N. I., Hawaldar, I. T., Alam, M. S. & Baig, M. A. (2022). The covid-19 pandemic and overconfidence bias: The case of cyclical and defensive sectors. *Risks*, 10, 56.

Bardh, P. & Haglund, J. (2021). *An investment approach built on systematic risk: A performance analysis based on the characteristics of defensive and cyclical sectors on the swedish stock market.* [Bachelor's thesis, Jönköping University]. Diva Portal. https://www.diva-portal.org/smash/record.jsf?dswid=-8437&pid=diva2%3A1562547.

Buszko, M., Orzeszko, W. & Stawarz, M. (2021). Covid-19 pandemic and stability of stock market—a sectoral approach. *PLoS ONE*, 16, e0250938.

Carter, D., Mazumder, S., Simkins, B. & Sisneros, E. (2022). The stock price reaction of the covid-19 pandemic on the airline, hotel, and tourism industries. *Finance Research Letters*, 44, 102047.

Chandar, S. K. (2019). Stock market prediction using subtractive clustering for a neuro fuzzy hybrid approach. *Cluster Computing*, 22, 13159-13166.

Chaudhary, R., Bakhshi, P. & Gupta, H. (2020). Volatility in international stock markets: An empirical study during covid-19. *Journal of Risk and Financial Management*, 13, 208.

Chee, J. E. (2019). *Shariah vs conventional: Which funds are better?* FocusM. https://www.maybank-am.com/documents/15019/0/syariahconventional.pdf/.

Cheong, C. W. (2021). Risk, resilience, and shariah-compliance. *Research in International Business and Finance*, 55, 101313.

Cortes, G. S. & Weidenmier, M. D. (2019). Stock volatility and the great depression. *The Review of Financial Studies*, 32, 3544-3570.

Cui, M. (2020). Introduction to the k-means clustering algorithm based on the elbow method. *Accounting, Auditing and Finance*, 1, 5-8.

Department of Statistics Malaysia. (2022). *Key findings population and housing census of Malaysia 2020*. https://www.mycensus.gov.my/index.php/census-product/publication/census-2020/key-findings-of-census-2020.

Doblas, M., Enriquez, J. & Lagara, M. (2020). Price to earnings and price to book ratios as determinants of stock return: The case of financial institution listed in bahrain bourse. *Journal of Applied Economic Science*, 15, 532–539.

Ernawati, S. S. K. B., Fauziah Kasmin, H., & Purwanugraha, A. (2022). Geo-marketing Promotional Target Selection using Modified RFM with Spatial and Temporal Analysis: A Case Study. *Journal of System and Management Sciences*, 12, 156-180.

Financial, T. (2007). Worldscope database: datatype definitions guide. *Thomson Financial*, 1-492.

Gai, K., Qiu, M. & Sun, X. (2018). A survey on fintech. *Journal of Network and Computer Applications*, 103, 262-273.

Google Developers. (2022). *k-means advantages and disadvantages*. https://developers.google.com/machine-learning/clustering/algorithm/advantages-disadvantages.

Hertina, D., Destriani, N., Naufal, M. D., Dauliah, R. D., Sinaga, D. S., Nursapriti, A. W. & Saudi, M. H. (2021). Sharpe, treynor and jensen methods in doing stock portfolio performance analysis. *Review of International Geographical Education Online*, 11, 829-834.

Hussein, A., Ahmad, F. K. & Kamaruddin, S. S. (2021). Cluster Analysis on covid-19 outbreak sentiments from twitter data using K-means algorithm. *Journal of System and Management Sciences*, 11, 167-189.

Institute of Financial Management. (2017). Introduction to datastream. https://www.ifm.unibe.ch/e39710/e39716/e274093/files495520/DATASTREAM.pdf.

Jiang, C., Du, J. & An, Y. (2019). Combining the minimum-variance and equally-weighted portfolios: Can portfolio performance be improved?. *Economic Modelling*, 80, 260-274.

Kedia, V., Khalid, Z., Goswami, S., Sharma, N. & Suryawanshi, K. "Portfolio generation for indian stock markets using unsupervised machine learning," *2018 Fourth International Conference on Computing Communication Control and Automation (ICCUBEA)*, IEEE, 2018.

Korzeniewski, J. (2018). Efficient stock portfolio construction by means of clustering. *Acta Universitatis Lodziensis Folia Oeconomica 1*.

Kusmayadi, D., Rahman, R. & Abdullah, Y. (2018). Analysis of the effect of net profit margin, price to book value, and debt to equity ratio on stock return. *International Journal of Recent Scientific Research*, 9, 28091-28095.

Malaysia, B. P. S., SANTHANASAMY, D. K. & ABDUL-RAHMAN, A. (2022). Robustification of shewhart control chart by median based estimators: A study on malaysia stock data. *Journal of Quality Measurement and Analysis JQMA*, 18, 13-26.

Morgan Stanley Capital International. (2022). Msci cyclical / defensive indexes. https://www.msci.com/documents/1296102/30991361/MSCI+Cyclical+and+Defensive+Indexes.pdf.

Nakagawa, K., Imamura, M. & Yoshida, K. (2019). Stock price prediction using k-medoids clustering with indexing dynamic time warping. *Electronics and Communications in Japan*, 102, 3-8.

Naz´ario, R. T. F., e Silva, J. L., Sobreiro, V. A. & Kimura, H. (2017). A literature review of technical analysis on stock markets. *The Quarterly Review of Economics and Finance*, 66, 115-126.

Potter, K., Hagen, H., Kerren, A. & Dannenmann, P. (2006). Methods for presenting statistical information: The box plot, In *VLUDS*, 97-106.

Retha, H. M. A. & Budiarti, R. (2021). Pengaruh beta terhadap return saham defensif dan agresif guna membantu investor dalam keputusan investasi. *GREENOMIKA*, 3, 56-62.

Rout B. & Panda J. (2020). Construction of optimal portfolio on selected stocks of bse using sharpe's single index model. *Srusti Management Review*, 12, 27-41.

Securities Commission Malaysia. (2020). List of shariah-compliant securities by the shariah advisory council securities commission malaysia. https://www.sc.com.my/api/documentms/download.ashx?id=dd80630a-6dd5-45ae-bedb-75f0586fe3ec.

Sholichah, F., Asfiah, N., Ambarwati, T., Widagdo, B., Ulfa, M. & Jihadi, M. (2021). The effects of profitability and solvability on stock prices: Empirical evidence from Indonesia. *Journal of Asian Finance, Economics and Business*, 8, 885-894.

Svanberg, J. & Max, D. (2019). *The Moat of Finance: Does Complexity Reward the Private Investor?* [Bachelor's thesis, KTH Royal Institute of Technology]. Diva Portal. https://www.diva-portal.org/smash/record.jsf?dswid=3641.

Umargono, E., Suseno, J. E. & Gunawan, S. V. "K-means clustering optimization using the elbow method and early centroid determination based on mean and median formula," *The 2nd International Seminar on Science and Technology (ISSTEC 2019)*, Atlantis Press, 2020.

Wu, J., 2012. Advances in K-means clustering: a data mining thinking. Springer Science & Business Media.

Yelamanchili, R. K. (2019). Impact of consumer sentiment on defensive and aggressive stock returns: Indian evidence. *International Journal of Economics and Financial Issues*, 9, 109.

Zaimovic, A., Omanovic, A. & Arnaut-Berilo, A. (2021). How many stocks are sufficient for equity portfolio diversification? a review of the literature. *Journal of Risk and Financial Management*, 14, 551.

Zainol Abidin, S. N., Jaaman, S. H., Ismail, M. & Abu Bakar, A. S. (2020). Clustering stock performance considering investor preferences using a fuzzy inference system. *Symmetry*, 12, 1148.

Zakaria, S. A. S. & Singh, A. K. M. (2021). Impacts of covid-19 outbreak on civil engineering activities in the malaysian construction industry: A review. *Jurnal Kejuruteraan*, 33, 477-485.

Zhu, Z., Liu, N. (2021). Early warning of financial risk based on k-means clustering algorithm. *Complexity*, 202.

Appendix A. Financial Ratios of Selected Clusters

The box plot results of each sector are shown in Fig. 2. In addition, the average (Avg), Interquartile Range (IQR), minimum value (Min), and maximum value (Max) of each financial ratio by sector are recorded in Table 9.

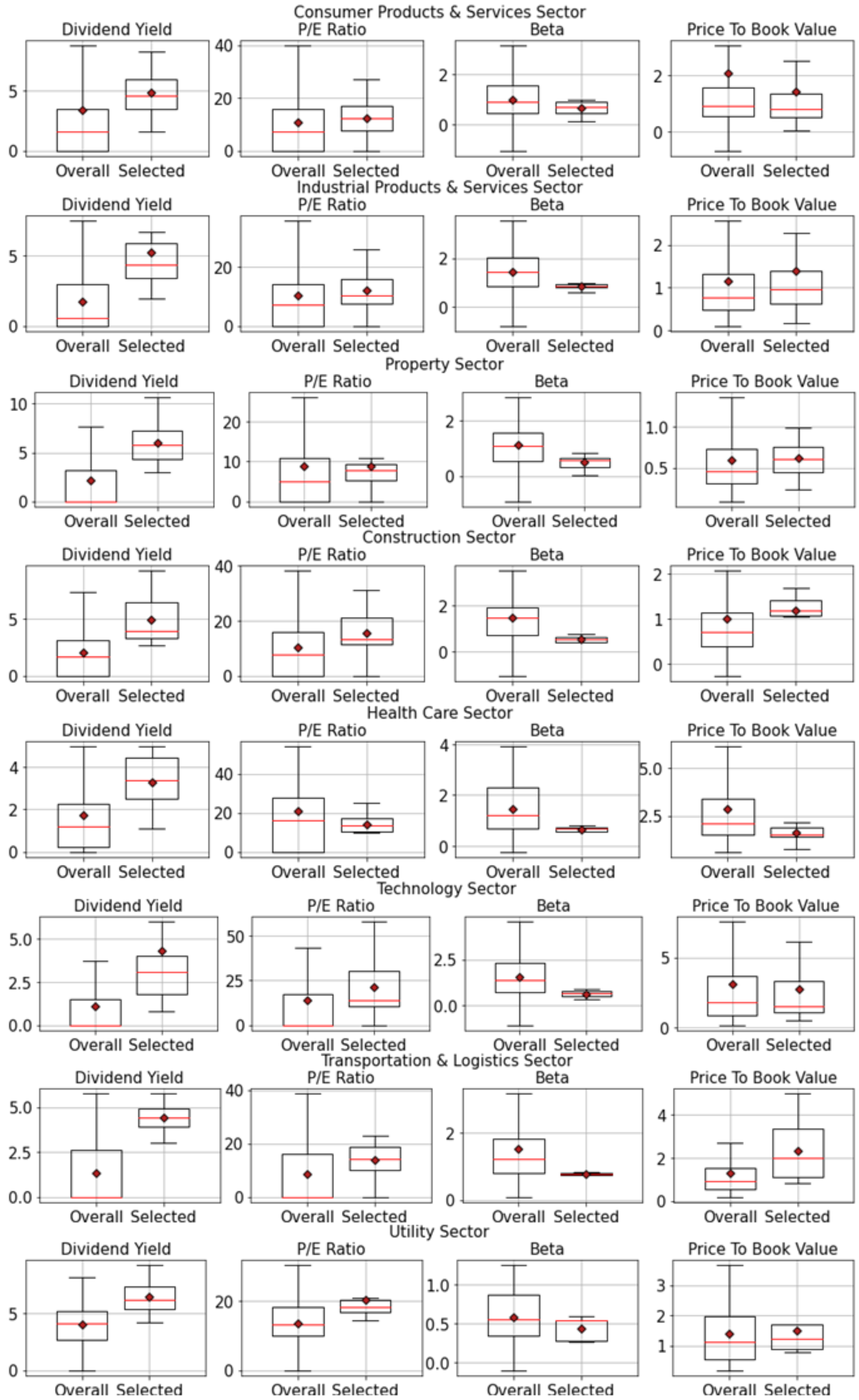

Fig. 2: The Box Plots of The Financial Ratios by Sector.

Table 9: The Summary of The Financial Ratios by Sector.

	DY				PE Ratio				Beta				PBV			
	Avg	IQR	Min	Max	Avg	IQR	Min	Max	Avg	IQR	Min	Max	Avg	IQR	Min	Max
Consumer (All)	3.4	3.49	0	250	10.67	15.9	0	85.7	0.99	1.09	-3.43	4.29	2.06	0.1	-11.69	57.03
Consumer (Select)	4.8	2.45	1.63	13.33	12.18	9.55	0	31.8	0.66	0.42	0.14	0.97	1.42	0.84	0.05	14.02
Industrial (All)	1.78	3.01	0	16.67	10.19	14.3	0	196.6	1.45	1.2	-4.01	6.79	1.14	0.84	0.1	19.78
Industrial (Select)	5.26	2.46	1.96	16.67	12.17	8.45	0	29.4	0.84	0.15	0.59	0.99	1.4	0.76	0.16	7.37
Property (All)	2.12	3.17	0	33.33	8.74	10.83	0	108.3	1.11	1.02	-1.6	5.33	0.59	0.42	0.09	2.74
Property (Select)	5.97	2.89	2.99	10.64	8.85	4.23	0	37.1	0.48	0.33	0.02	0.83	0.62	0.32	0.23	1.36
Construct (All)	2	3.12	0	20	10.18	15.8	0	53.2	1.47	1.2	-1.05	6.06	0.1	0.74	-0.26	6.96
Construct (Select)	4.94	3.06	2.63	9.17	15.49	9.6	0	31.4	0.53	0.21	0.41	0.75	1.19	0.36	0.49	1.68
Health (All)	1.72	2.04	0	10.47	20.9	27.8	0	230.1	1.44	1.59	-0.24	3.91	2.86	1.92	0.6	9.81
Health (Select)	3.29	1.94	1.09	4.95	13.7	6.33	0	30.7	0.63	0.19	0.05	0.8	1.61	0.53	0.73	2.85
Technology (All)	1.09	1.51	0	28.58	14.03	17.45	0	681.1	1.52	1.62	-3.55	5.14	3.06	2.76	0.17	63.4
Technology (Select)	4.29	2.23	0.8	28.58	21.17	19.85	0	66.8	0.63	0.28	0.36	0.86	2.72	2.23	0.52	8.05
Transport (All)	1.32	2.62	0	7	8.63	16.15	0	51.7	1.53	1.02	-1.65	7.67	1.28	0.99	-2.2	7.56
Transport (Select)	4.4	1.03	2.99	5.75	13.65	8.85	0	22.9	0.76	0.08	0.59	0.83	2.33	2.26	0.81	4.98
Utility (All)	4.04	2.54	0	9.2	13.54	8.3	0	36.3	0.58	0.53	-0.1	1.25	1.38	1.42	0.19	3.65
Utility (Select)	6.44	1.99	4.21	9.2	20.4	3.35	14.5	36.3	0.44	0.26	0.27	0.6	1.48	0.79	0.78	3.19