RESEARCH ARTICLE

Performance profiling of the unit trust funds in Malaysia with data mining techniques [version 1; peer review: 2 approved]

Aida Farah Khairuddin ¹, Keng-Hoong Ng ¹, Kok-Chin Khor ²

¹Faculty of Computing and Informatics, Multimedia University, Cyberjaya, 63000, Malaysia

²Lee Kong Chian Faculty of Engineering and Science, Universiti Tunku Abdul Rahman, Sungai Long, 43200, Malaysia

V1 First published: 13 Dec 2021, 10:1272
<https://doi.org/10.12688/f1000research.73467.1>

Latest published: 13 Dec 2021, 10:1272
<https://doi.org/10.12688/f1000research.73467.1>

Abstract

Background: Millennials are exposed to many investment opportunities, and they have shown their interest in gaining more income via investments. One popular investment avenue is unit trusts. However, analysing unit trusts' financial data and gaining valuable insights may not be as simple because not everyone has the required financial knowledge and adequate time to perform in-depth analytics on the numerous financial data. Furthermore, it is not easy to compile the performance of each unit trust available in Malaysia. The primary objective of this research is to identify unit trust funds that provide higher returns than their average peers via performance profiling. **Methods:** This research proposed a performance profiling on Malaysia unit trust funds using the two data mining techniques, i.e., Expectation Maximisation (EM) and Apriori, to assist amateur retail investors to choose the right unit trust based on their risk tolerance. EM clustered the unit trust funds in Malaysia into several groups based on their annual financial performances. This was then followed by finding the rules associated with each cluster by applying Apriori. The resulted rules shall serve the purpose of profiling the clustered unit trust funds. Retail investors can then select their preferred unit trust funds based on the performance profile of the clusters. **Results:** The yearly average total return of the financial year 2018 and 2019 was used to evaluate unit trust funds' performance in the clusters. The evaluation results indicated that the profiling could provide valuable and insightful information to retail investors with varying risk appetites. **Conclusions:** This research has demonstrated that the financial performance profiling of unit trust funds could be acquired via data mining approaches. This valuable information is crucial to unit trust investors for selecting suitable funds in investment.

Open Peer Review

Approval Status

	1	2
version 1
13 Dec 2021	view	view

1. **Lay Ki Soon** , Monash University Malaysia, Subang Jaya, Malaysia
2. **Wen Cheong Chin**, Xiamen University Malaysia, Sepang, Malaysia

Any reports and responses or comments on the article can be found at the end of the article.

Keywords

unit trust funds, expectation maximisation, apriori, performance profiling

This article is included in the **Research Synergy** Foundation gateway.

Corresponding author: Keng-Hoong Ng (khng@mmu.edu.my)

Author roles: **Khairuddin AF:** Conceptualization, Data Curation, Formal Analysis, Investigation, Methodology, Resources, Software, Visualization, Writing – Original Draft Preparation; **Ng KH:** Conceptualization, Project Administration, Resources, Software, Supervision, Validation, Writing – Original Draft Preparation; **Khor KC:** Conceptualization, Software, Supervision, Validation, Visualization, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: The author(s) declared that no grants were involved in supporting this work.

Copyright: © 2021 Khairuddin AF *et al.* This is an open access article distributed under the terms of the **Creative Commons Attribution License**, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Khairuddin AF, Ng KH and Khor KC. **Performance profiling of the unit trust funds in Malaysia with data mining techniques [version 1; peer review: 2 approved]** F1000Research 2021, 10:1272 <https://doi.org/10.12688/f1000research.73467.1>

First published: 13 Dec 2021, 10:1272 <https://doi.org/10.12688/f1000research.73467.1>

Introduction

In Malaysia, different investment securities or schemes are publicly available to investors. The common securities include stocks, bonds, deposits, properties, unit trusts and commodities. Youngsters nowadays are more aware of the importance of earning passive income from investment. Bursa Malaysia recorded a 36% rise in the new Central Depository System (CDS) account holders aged below 25 in 2016, totalling more than 25.2 thousand account holders at the end of this year.¹ Recently, the fast-growing trend has driven more financial products to be brought to the market to attract investors.

Unit trusts are investment options with lower risk² than equities because they are well-diversified financial instruments that fund managers handle. Each unit trust fund has its investment objective and strategy. Hence an amateur needs to pick the suitable funds to be included in their portfolio. However, it is not easy to choose suitable unit trust funds that meet the investor's requirements, especially from a large pool of unit trust funds available in Malaysia. As of December 2019, the Security Commission of Malaysia (www.sc.com.my/analytics/fund-management-products) reported that Malaysia's total launched unit trust funds are 685.

Amateur investors always find it challenging to identify appropriate unit trust funds that meet their investment strategies and risks. The task requires much time and effort to complete, looking at the abundance of data to be searched and analysed. Most amateur investors are usually busy with their full-time jobs or running their own business. As a result, they cannot screen and perform in-depth analysis on many unit trust funds.

Another typically encountered problem is the low financial literacy among amateur investors. Thus, they cannot carry out a thorough analysis to extract meaningful information or knowledge from the annual reports they read. Subsequently, they find it challenging to analyse unit trust funds' performance, and usually, their findings are inconclusive and indecisive. Amateur investors also face obstacles when they are looking for unit trust funds to fulfil their risk appetite. Hence, it is difficult for them to make investment decisions.

Considering the mentioned problems, this article aims to propose a data mining model that could build the performance profiles of unit trust funds using clustering and association rules mining algorithms. The model could mitigate the problems because: (i) the performance profiles could be generated rapidly, (ii) the financial information in the profile is intuitive and easy to understand by amateur investors, and (iii) the profiles also provide risk information.

Literature review

The research by F. Cai *et al.*³ evaluated 904 investment funds from financial datasets (time series and transactions) retrieved from the Morningstar using *k*-means and density-based clustering. The finding indicated that *k*-means performed better than the latter, giving the best number of clusters. In the research by T. Sakakibara *et al.*,⁴ clustering of mutual funds was done based on investment similarity instead of using historical performance similarity. The proposed approach was tested on 551 Japanese mutual funds. The result claimed that the approach could acquire the optimal number of clusters, even better than the classification provided by Morningstar Inc.

K. H. Ng *et al.*⁵ adopted association rules mining to find frequent financial patterns in outlier stocks listed on Bursa Malaysia. The outliers were identified via a score-based approach, and they were manually grouped into superior and poor outliers. The research produced nine rules associated with outstanding stocks and four rules for the poor performing stocks. Investors can refer to these rules for decision making. Another study applied clustering to construct an efficient stock portfolio in the Warsaw Stock Exchange.⁶ The study utilised *k*-means and Partitioning Around Medoids (PAM) methods for this purpose and showed satisfactory returns.

Methods

This research applies data mining techniques to profile the unit trust funds available in Malaysia. The entire process consisted of five key steps. The first step was to collect the financial data of the unit trust funds from annual reports. The next step involved a normalisation technique to ensure that the financial attributes share the same scale. The third step was grouping unit trust funds using a clustering algorithm. This was followed by profiling the clusters to find the common rules/characteristics associated with each cluster. In the final step, evaluation was conducted to study the efficiency of the proposed profiling method. The research overview is shown in Pseudocode 1.

Pseudocode 1. Profiling of unit trust funds.**Input:** Annual financial reports of unit trust funds, *AFRs***Output:** Financial profiling of unit trust funds

```

1. //collect unit trust funds financial data
2. for each unit trust fund's report,  $ut \in AFRs$  do
3.   for  $Fy \in \{fromYr, toYr\}$  do//financial year  $Fy$ 
4.      $FinD \leftarrow$  extract  $Fy$  data from  $ut$ 
5.   //append financial data,  $FinD$  to dataset  $Db_{Fy}$ 
6.      $Db_{Fy}.append (FinD)$ 
7.   end for
8. end for
9. //normalize the two financial datasets
10. perform z-score normalization on  $Db_{2017}$  and  $Db_{2018}$ 
11. //apply EM clustering to group unit trust funds
12.  $\{C_1, C_2, C_3, \dots, C_m\} \leftarrow EMCluster (Db_{2017})$ 
13. //EM generated  $m$  clusters,  $C_1$  until  $C_m$ 
14.  $\{E_1, E_2, E_3, \dots, E_n\} \leftarrow EMCluster (Db_{2018})$ 
15. //EM generated  $n$  clusters,  $E_1$  until  $E_n$ 
16. //discretize the attributes in the datasets
17. calculate quartile 1 - 3 values for all attributes in  $Db_{2017}$  and  $Db_{2018}$ 
18. for each attribute,  $att \in \{Db_{2017}, Db_{2018}\}$  do
19.   if  $att.value \geq q3$ //quartile 3 value
20.      $att.value \leftarrow$  high
21.   else if  $att.value \geq q1$ 
22.      $att.value \leftarrow$  moderate
23.   else
24.      $att.value \leftarrow$  low
25.   end if
26. end for
27. //find associated rules in each cluster
28. for each cluster generated in line 12 and 14
29.   apply apriori algorithm to find frequent itemset for the cluster
30.   profile the cluster based on the frequent itemset
31. end for

```

Unit trust funds datasets

The financial data of the unit trust funds were gathered from the annual report of each unit trust fund. From the annual report, six common financial variables were identified as crucial in the study: Net Asset Value (NAV) per unit, total growth rate, capital growth rate, income distribution rate, management expense ratio (MER), and portfolio turnover ratio (PTR). The detail of each financial attribute is shown in [Table 1](#). In this study, 326 local unit trust funds' data from 26 financial institutions were collected (Underlying data).⁷ Examples of financial institutions were CIMB Bank, Maybank, Prudential Insurance, Kenanga Investment Bank, etc.

Data normalisation

The created dataset underwent the z-score normalisation before the clustering process. This was to prevent features with a larger scale from outweighing smaller-scaled features in the data mining process. This normalisation technique produced equal weight features within the range of $[-1, 1]$. The formula for the z-score normalisation⁸ is shown in the following:

$$Z = (p - \mu) / \sigma$$

Variable p denotes the original financial data value. The mean and the standard deviation of the financial data are represented by μ and σ , respectively.

Expectation Maximisation (EM) clustering

In this step, the normalised dataset was partitioned into clusters based on their feature similarity. The use of EM clustering in this study could be justified with the following: (i) No predefined number of clusters, and (ii) It could handle missing values.⁹ EM clustering assigns each unit trust fund a probability distribution that represents the probability it belongs

Table 1. The six selected financial attributes that were used in this study.

Financial attribute	Description/Formula
NAV per unit (RM)	The net asset value of a fund divided by the number of units in circulation at the valuation point. $NAV\ per\ unit\ (RM) = total\ NAV\ of\ fund / total\ units$
Total Growth (%)	The actual rate of return of an investment at the valuation point. $Total\ Growth = Capital\ Growth\ (\%) + Income\ Return\ (\%)$
Capital Growth (%)	Increase in the value of an asset over time. $Capital\ Growth = (NAV\ per\ unit\ end / NAV\ per\ unit\ begin - 1) \times 100$
Income Distribution (%)	Income declared and distributed back to its investors in a year $Income\ Distribution = (Gross\ Income\ Distribution\ declared / NAV\ per\ unit\ begin - 1) \times 100$
Management Expense Ratio (MER) (%)	A measure to see how expensive a unit trust fund is to investors. ⁷ $MER = (operating\ costs + management\ fee) / Total\ Assets\ under\ management$
Portfolio Turnover (PTR) (times)	A measure of how frequently assets within a fund are bought and sold by the managers. ⁷ $X = total\ purchases\ or\ total\ sales\ (choose\ higher)$ $PTR = X / average\ monthly\ assets$

to each cluster. By maximising the log-likelihood of the data, EM finds the optimised parameters of a probability distribution.

EM has two major steps, i.e., E-step (expectation) and M-step (maximisation). The detailed explanation of the algorithm is illustrated in Pseudocode 2. EM begins with initialising random values to parameters mean, variance, and a fraction of the data in each cluster c (line 2). Subsequently, E-step (line 3-7) computes the expected likelihood for the unit trust dataset. This is followed by maximising the likelihood of the data by re-estimating the parameter values in the M-step (lines 8-10). Both steps are repeated until the likelihood converges and reaches a local maximum. This means that the iteration will halt once the likelihood cannot be improved further.

Pseudocode 2. EM clustering on the unit trust dataset.

Input: Unit trust dataset for the financial year, Db_{FY}

Output: Unit trust clusters

```

1. //  $Fr_c$  represents the fraction of the data represented by cluster  $c$ 
2. assign random values to  $mean^0(c)$ ,  $variance^0(c)$  and  $Fr_c^0$ 
3. for each unit trust,  $utr$  in  $Db_{FY}$  do //expectation step
4.   for each cluster  $c$  do
5.     compute the probability of  $utr$  in the cluster  $c$ 
6.   end for
7. end for
8. for each cluster  $c$  do //maximisation step
9.   re-estimate  $mean^{i+1}(c)$ ,  $variance^{i+1}(c)$  and  $Fr^{i+1}_c$  to maximise the likelihood of
     the unit trusts
10. end for
11. repeat steps 3 - 10 until the parameters converge
12. return the final clusters,  $c_1, c_2, \dots, c_k$  //  $k$  denotes the number of clusters

```

Class Association Rules (CAR) mining

The original format of the financial features in the dataset is numeric. Thus, these data were required to undergo discretisation before being processed by the association rules algorithm. The binning method was applied to discretise the financial features into three categories, i.e., low, medium and high. They were divided using the quartile range. The data was discretised to “low” if it was less than Q1 of the data (quartile 1). “Moderate” referred to data that was more than or equal to Q1 and less than or equal to Q3. Any data greater than Q3 was discretised as “high”.

Association Rules Mining (ARM) is useful to discover frequent rules or patterns among groups of objects in a dataset. In this study, CAR¹⁰ mining was adopted to find a subset of rules associated with each cluster. The primary intention was to

uncover a set of financial behaviours that are associated with each cluster uniquely. The financial behaviours can be used to build the cluster profile. CAR mining could be divided into two major steps, i.e. (i) discover all frequent k -itemsets in a dataset that comply with the user-defined minimum support, (ii) find the frequent $k + 1$ itemsets with the help of k -itemsets by applying a self-join rule. The detailed steps of the CAR are described in Pseudocode 3.

Pseudocode 3. Mining rules associated with the cluster.

Input: Discretised financial data of unit trusts in a cluster

Output: Financial profiles of the unit trust cluster

```

1. //mSup denotes minimum support
2.  $L_1 \leftarrow$  search all frequent 1-itemset with support count  $\geq c\_size * mSup$ 
   //c_size is the cluster size
3. Initialize  $j \leftarrow 1$ 
4. repeat
5.    $j \leftarrow j+1$ 
6.    $CandIS_j \leftarrow$  find candidate itemsets from  $L_{j-1}$ 
7.   for each unit trust  $utc$  in the cluster do
8.      $Cand_{utc} \leftarrow$  subset ( $CandIS_j, utc$ )
9.     for each candidate itemset in  $Cand_{utc}$  do
10.      add support count of the candidate itemset
11.    end for
12.  end for
13.  $L_j \leftarrow$  extract the frequent  $j$ -itemsets with support count  $\geq c\_size * mSup$ 
14. until  $L_j$  is null
15. return all  $L$ 

```

Results and discussion

This study used a unit trust dataset that contained six normalised financial variables of the year 2017. It started with the clustering process using the Expectation Maximisation (EM) method. As a result, eight clusters were produced. Table 2 shows the detail of each cluster, including the number of assigned unit trust funds and the means of the six financial features. Cluster 2 was the largest cluster with 63 unit trust funds, and the smallest was Cluster 5 with only 14 members.

Table 2. EM clustering yielded 8 clusters. The smallest cluster has 14 funds (cluster 5), and the largest cluster contains 63 funds (cluster 2).

NAV	TR	CG	ID	MER	PTR
Cluster 1 (42 unit trust funds)					
-0.177	-0.033	-0.318	-0.356	-0.178	-0.339
Cluster 2 (63 unit trust funds)					
-0.239	-0.102	-0.208	-0.967	-0.139	-0.513
Cluster 3 (52 unit trust funds)					
-0.572	-0.027	-0.364	-0.264	-0.039	-0.454
Cluster 4 (41 unit trust funds)					
-0.444	-0.196	-0.429	-0.622	-0.399	-0.727
Cluster 5 (14 unit trust funds)					
-0.426	-0.124	-0.487	0.055	0.002	-0.375
Cluster 6 (38 unit trust funds)					
-0.776	-0.096	-0.239	-0.870	-0.128	-0.850
Cluster 7 (49 unit trust funds)					
0.207	-0.207	-0.449	-0.600	-0.588	-0.754
Cluster 8 (27 unit trust funds)					
-0.133	0.514	0.409	-0.810	-0.003	-0.233

Even though the mean of each financial variable is produced and displayed in the table, it is not an easy task for an amateur investor to analyse and interpret the generated information correctly. This is because the financial variables have been normalised before the clustering. Hence, a further step was taken to process this information by employing the Association Rules Mining technique so that the generated profile information on each cluster would be more intuitive and easier to understand. To perform the ARM, the continuous financial data was discretised into three bins, i.e., high, moderate, low, using the binning method.

ARM was conducted on each cluster to discover a set of rules associated with each cluster. Two parameters, i.e., minimum support and minimum confidence, were predefined. The first parameter was set to 50%,¹¹ and a higher threshold of 90% was applied to the minimum confidence. ARM only yielded meaningful outcomes on three clusters: Cluster 8, Cluster 4, and Cluster 5. The remaining clusters did not demonstrate any strongly associated rules after the process.

Cluster 8 produced two frequent 3-itemsets (Table 3). The result strongly indicates that unit trust funds in this cluster belong to the type of high total return, high capital growth, and high portfolio turnover. Upon examining the portfolios of some unit trust funds in the cluster, it was discovered that these funds prioritise stock investment in their investment baskets. This finding was sufficient to justify the derived rules associated with the cluster because the return and capital growth of stock investment were relatively higher than the fixed income securities such as bonds, fixed deposits, etc. High portfolio turnover is also expected in these unit trust funds because the fund managers must always fine-tune their portfolios to maximise the returns. As such, the financial profile of this cluster can be summarised as “Aggressive”. This cluster is appropriate for investors who are risk-takers and aim for a high growth rate and return.

The rules associated with Cluster 4 are illustrated in Table 4, displaying the three frequent 3-itemsets derived from the cluster. The financial profile of Cluster 4 might not have been as outstanding as Cluster 8, but it is noteworthy to describe the unit trust funds in Cluster 4 as average and Defensive. Even though they have average performance among their peers in total return, capital growth, and income distribution, they give stable returns over the years.

A thorough analysis of some of the unit trust funds in Cluster 4 revealed conservative investment approaches. They adopted a more diverse investment strategy by pooling their funds into a basket of securities consisting of shares and bonds, deposits, and properties. Combining these investment securities can lower the downside risk due to market volatility. In short, Cluster 4 was considered “Defensive”, and therefore suitable for investors with a low-risk tolerance.

Cluster 5 exhibited inferior financial performance. Table 5 shows the six frequent 2-itemsets derived after the ARM process. The associated rules in this cluster included low total return, low capital growth, high portfolio turnover, and

Table 3. ARM on the discretised financial data of the unit trust funds in Cluster 8.

Itemsets	Support count
Total Return = High Capital Growth = High Income Distribution = Moderate	23
Total Return = High Capital Growth = High Portfolio Turnover = High	18

Table 4. Association rules mining applied to Cluster 4.

Itemsets	Support count
NAV per unit = Moderate Capital Growth = Moderate Income Distribution = Moderate	24
NAV per unit = Moderate Total Return = Moderate Income Distribution = Moderate	23
NAV per unit = Moderate Total Return = Moderate Capital Growth = Moderate	21

Table 5. Six frequent 2-itemsets were produced after analysing Cluster 5 with the ARM.

Itemsets	Support count
NAV per unit = Moderate Income Distribution = High	10
Total Return = Low Capital Growth = Low	8
NAV per unit = Moderate Total Return = Low	7
Total Return = Low Income Distribution = High	7
NAV per unit = Moderate Portfolio Turnover = High	7
Income Distribution = High Portfolio Turnover = High	7

high-income distribution. The phenomena of low total return but high-income distribution was due to poor capital growth. Stagnant or negative growth in the investment capital significantly outweighed the income distribution. Hence, this cluster was deemed as an inferior type. Investors are advised to avoid such unit trust funds at all costs.

The financial profiles created for the three clusters were further assessed and validated using the average total return (yearly) for the financial year 2018 and 2019 (Table 6). For 2018, the best performer was the Defensive cluster, which was the only one that still delivered a positive return (+1.17%). Both Aggressive and Inferior clusters were in the negative territory. Thorough investigation revealed that the unsatisfactory performance in 2018 was mainly attributed to the tumble of the Bursa market (Figure 1) affected by the three factors: general election, funds pulled out by foreign investors, and negative investor sentiment.

The 14th Malaysia general election has shaken investors' confidence to invest in Malaysia's equity markets. As a result, many investors withdrew their securities investments, and their action caused a sharp fall in the Bursa Index.¹² The stock market was still in higher volatility mode in the second half of 2018. Many stocks could not recover their prices in the first half of the year.¹³ Funds pulled out by foreign fund managers also attributed to the stock market fall. It has been reported that the net foreign fund outflow in 2018 for Malaysia totalled RM11.65 billion.¹⁴ The last factor that adversely impacted the stock market was investor sentiment. Issues such as unresolved trade wars, lower oil prices, and geopolitical tension had refrained investors from securities investment.

In 2019, the volatility of the Malaysian stock market had subsided. Hence, all three clusters showed improvement in the total return as compared to the preceding year. The average total return (+4.65%) of Cluster 4 outperformed the others. Cluster 8 managed to gain a positive return of +3.52%. The worst performer (−2.98%) was still Cluster 5. The evaluation of two financial years for the three cluster profiles has strongly demonstrated that Cluster 4, with a Defensive profile, was resilient in the harsh economic climate that included market downturn, high market volatility, etc. This cluster may not have performed as well as Cluster 8 with an Aggressive profile during the economic boom, but it did provide steady passive income with lesser risk in the long run. Lastly, it could be summarised that investors with lower risk tolerance could aim at the defensive unit trust funds in Cluster 4. On the other hand, high-risk investors could identify the Aggressive unit trust funds in Cluster 8.

Table 6. Three identified clusters were evaluated using average total returns (%) in 2018 and 2019.

Cluster	Average total return (%)		
	2017	2018	2019
Cluster 8 Aggressive	+21.56	−9.41	+3.52
Cluster 4 Defensive	+3.07	+1.17	+4.65
Cluster 5 Inferior	+3.94	−8.69	−2.98

Figure 1. Bursa Malaysia KLCI Index lost more than 100 points in the year 2018.

Conclusions

This study used data mining techniques to generate a financial profile for unit trust funds. The profile information is intuitive and easily understood by an investor. The investor could narrow down the number of unit trust funds for analysis. Investors could focus on just a single cluster for making an investment decision. For instance, the investor could identify aggressive unit trust funds from Cluster 8 if he/she is a high risk-taker.

On the contrary, investors could use Cluster 4 for steady and average passive incomes. The finding from this study will have significant implications for unit trust investors. To the best of our knowledge, no published work has applied data mining approaches to profile unit trust funds in Malaysia. We hope to continue and extend the study to include non-financial factors, e.g., fund manager profile and time-series data, in the future.

Data availability

Underlying data

Zenodo: Malaysia unit trust funds dataset, <https://doi.org/10.5281/zenodo.5291931>.⁷

Local unit trust checked version 1: The dataset contained 326 unit trust funds with their six financial attributes.

Data are available under the terms of the [Creative Commons Zero “No rights reserved” data waiver](#) (Attribution 4.0 International)

Author contributions

All authors contributed equally to the conceptualisation of this study. Data curation, methodology, analysis and investigation, were performed by Aida Farah Khairudin. Keng-Hoong Ng and Kok-Chin Khor supervised the study

and validated the evaluation results. Keng-Hoong Ng and Aida Farah Khairudin wrote the first draft of the manuscript. Kok-Chin Khor reviewed and edited the manuscript. All authors checked and approved the final manuscript.

References

1. Aruna P: **More young investors entering into Malaysia market.** *The Star Online*. May 2017.
[Reference Source](#)
2. Annamalah S, Raman M, Marthandan G, *et al.*: **An empirical study on the determinants of an investor's decision in unit trust investment.** *Economies*. 2019; **7**(3): 80.
[Publisher Full Text](#)
3. Cai F, Le-Khac NA, Kechadi T: **Clustering approaches for financial data analysis: a survey.** *arXiv preprint arXiv:1609.08520* 2016.
4. Sakakibara T, Matsui T, Mutoh A, *et al.*: **Clustering mutual funds based on investment similarity.** *Procedia Comput. Sci.* 2015; **60**: 881–890.
[Publisher Full Text](#)
5. Ng KH, Khor KC, Tong GK: **Class Association Rules for Profiling Outlier Stocks.** *Int. J. Advance Soft Compu. Appl.* 2017; **9**(3): 114–131.
6. Korzeniewski J: **Efficient Stock Portfolio Construction by Means of Clustering.** *Acta Universitatis Lodzensis. Folia Oeconomica*. 2018; **1**(333): 85–92.
[Publisher Full Text](#)
7. Ng KH, Khairuddin AF, Khor K-C: **Malaysia unit trust funds dataset [Data set].** *Zenodo*. 2021.
[Publisher Full Text](#)
8. Othman J, Asuta M, Jamilan N: **Comparing the determinants of fund flows in domestically managed Malaysian Islamic and conventional equity funds.** *J. Islam. Account. Bus. Res.* 2018; **9**(3): 401–414.
[Publisher Full Text](#)
9. Saranya C, Manikandan G: **A study on normalization techniques for privacy preserving data mining.** *Int. J. Eng. Technol.* 2013; **5**(3): 2701–2704.
10. Jung YG, Kang MS, Heo J: **Clustering performance comparison using K-means and expectation maximisation algorithms.** *Biotechnol. Biotechnol. Equip.* 2014; **28**(sup 1): S44–S48.
[PubMed Abstract](#) | [Publisher Full Text](#)
11. Nguyen D, Nguyen LT, Vo B, *et al.*: **Efficient mining of class association rules with the itemset constraint.** *Knowl.-Based Syst.* 2016; **103**: 73–88.
[Publisher Full Text](#)
12. Abdulsalam SO, Adewole KS, Akintola AG, *et al.*: **Data mining in market basket transaction: An association rule mining approach.** *International Journal of Applied Information Systems (IJ AIS)*. 2014; **7**(10): 15–20.
13. Nadzri MMN: **The 14th General Election, the Fall of Barisan Nasional, and Political Development in Malaysia, 1957-2018.** *J. Curr. Southeast Asian Aff.* 2018; **37**(3): 139–171.
[Publisher Full Text](#)
14. Misman FN, Roslan S, Aladin MIM: **General Election and Stock Market Performance: A Malaysian Case.** *Int. J. Financ. Res.* 2020; **11**(3): 139–145.
[Publisher Full Text](#)
15. Murugiah S: **Foreign selling narrowed to RM127.6m last week, says MIDF Research.** *The Edge Markets*. Dec 2018.

Open Peer Review

Current Peer Review Status:

Version 1

Reviewer Report 27 June 2022

<https://doi.org/10.5256/f1000research.77120.r109294>

© 2022 Chin W. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Wen Cheong Chin

Department of Mathematics, Xiamen University Malaysia, Sepang, Malaysia

This article aims to use 2 data mining methodologies to profile the performance of 326 unit trust funds in Malaysia. 6 financial attributes have been selected and has being normalized as an Z-score before the clustering processes. The outcome of the analysis is able to assist investors in their investments based on their risk and return preferences by selecting their desired Cluster.

Suggestion:

1. Provide more literature review related to this similar study in profiling performance.
2. It is good to provide some risk indicator (eg. Sharpe ratio) for each cluster. Therefore investors can consider how risky each cluster.
3. It is good for this study to extend to predictive analysis, where investors are given some guides for their future investments.

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Financial time series and risk management.

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Reviewer Report 23 February 2022

<https://doi.org/10.5256/f1000research.77120.r109286>

© 2022 Soon L. This is an open access peer review report distributed under the terms of the [Creative Commons Attribution License](#), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Lay Ki Soon

School of Information Technology, Monash University Malaysia, Subang Jaya, Malaysia

The article reports the experiments on profiling selected unit trust funds in Malaysia using EM and CAR algorithms. While the results were as expected, the overall work is decent and the experimental validation is quite thorough. Some minor suggestions to improve the article:

1. More details should be provided on the dataset collection. For instance, the source for the raw data (annual report), the year of the annual report collected, and how were the 326 funds selected.
2. In terms of the features, why were only six variables used? Any empirical results to support this?
3. For the discretization, which binning method was used?
4. In results analysis, details were given on the impact of general elections to the stock market in Malaysia. It would be good to include relation between the stock market and unit trust funds.
5. The results were good in the sense that similar unit trust funds are grouped in the same clusters. These were expected. Any outliers (something that is unexpected but discovered by the method used) found in the experimental results? That would be interesting to discuss.

Is the work clearly and accurately presented and does it cite the current literature?

Partly

Is the study design appropriate and is the work technically sound?

Yes

Are sufficient details of methods and analysis provided to allow replication by others?

Yes

If applicable, is the statistical analysis and its interpretation appropriate?

Yes

Are all the source data underlying the results available to ensure full reproducibility?

Yes

Are the conclusions drawn adequately supported by the results?

Yes

Competing Interests: No competing interests were disclosed.

Reviewer Expertise: Text and data mining

I confirm that I have read this submission and believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

The benefits of publishing with F1000Research:

- Your article is published within days, with no editorial bias
- You can publish traditional articles, null/negative results, case reports, data notes and more
- The peer review process is transparent and collaborative
- Your article is indexed in PubMed after passing peer review
- Dedicated customer support at every stage

For pre-submission enquiries, contact research@f1000.com

F1000Research