Journal of Asia Entrepreneurship and Sustainability

Volume XIX,
Issue 2, July 2023

Refereed Special Conference Edition

Entrepreneurship, and Management for a Sustainable Business Growth in the Age of Web 4.0 and Beyond - Challenges and Opportunities

Special Edition Editors:
Bala Subramanian R (Managing)
Lakshmi, J
Rajagiri Business School
Kochi, India

Print: ISSN 1177-4541
On-Line: ISSN 1176-8592
www.asiaentrepreneurshipjournal.com

RossiSmith Academic Publishing

Special Edition Editors:

Dr. Bala Subramanian R is an Assistant Professor with Rajagiri Business School, Rajagiri College of Social Sciences, Kochi, India. His area of specialization is Organizational Behaviour and Human Resource Management. Bala Subramanian R has an overall 15 years of experience in academics and two years in the industry. His research area includes positive psychology, personality, group, and well-being.
Bala.mbahr@gmail.com

Dr Lakshmi Jayaprakash is an Associate Professor at Rajagiri College of Social Sciences, Kochi India. Her area of specialisation is Accounting and finance. Dr Lakshmi Jayaprakash have over 17 years of experience in teaching. Her area of research is behavioural finance.
Email id: lakshmi@rajagiri.edu

Table of Content

RossiSmith Academic Publications, Oxford/London UK, www.publicationsales.com

Journal of Asia Entrepreneurship and Sustainability
www.asiaentrepreneurshipjournal.com

Towards a sustainable India: Clustered Evaluation of state performance on SDG indicators

Nagendra BV, Joseph Durai Selvam J, Anil Kumar B, Renuka Prasad A, Atul Kattakayam, Prabhkirat Kaur, Mehak Deep Kaur, Tania Chakraborty

School of Business and Management, CHRIST (Deemed To Be University), Bengaluru, India

Abstract

Purpose: The purpose of this study is to explore the application of cluster analysis on the statewide SDG performance of India. The key objective is to cluster the regions of India based on the SDG scores to enable the government to carry out development projects with minimized logistical load and resource allocation.

Method: This paper employs a K-means clustering algorithm based on the optimal number of clusters identified by the elbow method in Hubert statistics and Dindex values. The model is evaluated using the silhouette score.

Findings: The findings reveal that India can be primarily divided into three clusters based on SDG performance. The first cluster consists mostly of the Central-Eastern regions and performs considerably poorly across five to six SDGs. This cluster requires immediate and major attention from the central government. The second cluster consists of the Southern states and the neck region of the country. This cluster is performing well in general, with a slight concern regarding one or two SDGs. The final cluster includes regions defined by unconventional geography. Its SDG performance is quite average, mainly restricted by the uniqueness of its geography.

Research limitations/implications: This research suggests there is potential for further research in this domain. The research is limited by the irrelevancy of a few SDG goals, not accounting for subgoals within.

1. Introduction

Sustainable development goals (SDGs) have become increasingly important in India. They are a set of goals by the United Nations to achieve a more sustainable and equitable global society. In India, these goals are increasingly being implemented at the state level. India is a rapidly growing economy with an expanding population and increasing resource demand. Sustainable development goals can help the states of India to address this challenge by promoting renewable energy, reducing air pollution, and providing better access to healthcare and education.

There are, in total, 17 SDG goals known as Global goals. SDG 1: No Poverty ensures to end poverty in all its forms everywhere, SDG 2: Zero Hunger, ensures to end hunger, achieve food security and improved nutrition and promote sustainable agriculture, SDG 3: Good Health and Well-being, ensure healthy lives and promote well-being for all at all ages, SDG 4: Quality Education, ensure inclusive and equitable quality education and promote lifelong learning opportunities for all, SDG 5: Gender Equality, ensures to achieve gender equality and empower all women and girls, SDG 6: Clean Water and Sanitation, ensure availability and sustainable management of water and sanitation for all, SDG 7: Affordable and Clean Energy, ensure access to affordable, reliable, sustainable and modern energy for all, SDG 8: Decent Work and Economic Growth, ensures to promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all, SDG 9: Industry, Innovation and Infrastructure, ensures to build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation, SDG 10: Reduced Inequalities,ensures to reduce income inequality within and among countries, SDG 11: Sustainable Cities and Communities, ensures to make cities and human settlements inclusive, safe, resilient and sustainable, SDG 12: Responsible Consumption and Production, ensures sustainable consumption and production patterns, SDG 13: Climate Action, ensures to take urgent action to combat climate change and its impacts by regulating emissions and promoting developments in renewable energy, SDG 14: Life Below Water, ensures to conserve and sustainably use the oceans, seas, and marine resources for sustainable development, SDG 15: Life on Land, ensures to protect, restore and promote the sustainable use of terrestrial ecosystems, forests, and biodiversity, SDG 16: Peace, Justice, and Strong Institutions, ensures to promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable, and inclusive institutions at all levels, SDG 17: Partnerships for the Goals, ensures to strengthen the means of implementation and revitalize the global partnership for sustainable development.

Sustainable development goals provide a framework for states to prioritize economic growth and sustainable development. By setting targets and timelines for achieving these goals, states can ensure that their economies are growing sustainably and equitably. The SDGs are also important for reducing poverty in India. They provide a roadmap for states to focus on poverty alleviation, gender equality, and improving access to basic services. By addressing these challenges, states can reduce poverty and improve the lives of their citizens.

Sustainable development goals are an important tool for states in India to implement sustainable development. They provide a framework for states to address many pressing issues and promote equitable economic growth. By setting targets and timelines for achieving these goals, states can ensure that their economies grow sustainably and equitably. Moreover, if it was figured out that similar states lack any particular Sustainable development goal, then it is possible to make a common plan for similar states that will save time and provide a clear plan to work on a sustainable development goal in all these common states.

Clustering is a data-driven approach that can help ensure that sustainable development initiatives are effective in each state and can help identify areas of the states that may need additional support to reach their sustainable development goals. Moreover, it can be used to identify and group states that share common characteristics, such as clustering a few of the states into one cluster that has not worked on common sustainable development goals and, as a result, provide a better understanding of the future projects to be done. Additionally, clustering can be used to identify states which are outliers or do not fit into the overall pattern of the data, thus allowing for a more informed decision-making process. Ultimately, clustering can be a useful tool for solving complex problems and providing more efficient solutions.

2. Literature Review

The global pandemic of COVID-19 has the potential to undermine the world's commitment to the 2030 Agenda for Sustainable Development. This study focuses on the impact of the pandemic on the interdependencies between Sustainable Development Goals (SDGs), particularly SDG3, SDG4, SDG8, SDG12, and SDG13. Based on moderated focus group discussions, the study reveals a unique pattern of interconnectedness between SDGs that can be related to the consequences of COVID-19 and highlights the potential impact on SDG5, SDG9, SDG10, SDG17, and SDG11 (Shulla et al., 2021).

The 17 Sustainable Development Goals (SDGs) aim to promote sustainable practices and solutions to address the main issues faced by society. Circular Economy (CE) and Industry 4.0 (I4.0) are two emerging topics that have gained interest for their potential to contribute to achieving the SDG. This study performed a systematic literature review to identify how the combination of CE practices and I4.0 technologies (CE-I4.0 nexus) could contribute to achieving the SDG. The study found that the CE-I4.0 nexus directly benefits several SDG targets and is a crucial step in reaching the SDG. Further research could investigate the impacts and secondary effects of the CE-I4.0 nexus on the SDG (Dantas et al., 2021).

The United Nations 2030 Agenda introduced the Sustainable Development Goals (SDGs) in 2015 as a global solution to sustainability challenges. Despite being a new field, digitalization offers the potential for a greener economy and society. However, little research exists on the contribution of digital paradigms to sustainability and the SDGs. The article aims to address these gaps by examining the main SDGs research gaps and exploring the potential of Big Data and Artificial Intelligence to overcome these gaps (Del Río Castro et al., 2021).

Investment in the Sustainable Development Goals (SDGs) is falling short of the target to close the $2.5 trillion annual financing gap for developing countries. The COVID-19 crisis has only worsened the situation and may reverse the progress made in the last six years. This paper evaluates the global trends in investing and financing the SDGs, including the various financing instruments launched to address the COVID-19 crisis. It analyzes the challenges of mobilizing funds, channeling investment into SDG sectors, and maximizing positive impact (Zhan & Santos-Paulino, 2021).

Artificial intelligence is starting to deliver real value in the fourth industrial revolution due to the availability of data, computational power, and algorithms. The study investigated AI's impact on attaining Sustainable Development Goals, specifically poverty reduction and industry, innovation, and infrastructure development. The results showed that AI has a significant influence on poverty reduction, infrastructure reliability, and economic growth in emerging economies, particularly in the agriculture, education, and finance sectors. Governments in emerging economies should invest more in AI to achieve the SDGs related to innovation, infrastructure, and poverty reduction (Mhlanga, 2021).

Artificial Intelligence (AI) has emerged and is increasingly impacting various sectors, which requires an evaluation of its effect on Sustainable Development

Goals. The research found that AI can enable the achievement of 134 targets but may also hinder 59 targets, with current research missing important aspects. The rapid development of AI needs regulatory insight and oversight to support sustainable development; otherwise, there may be gaps in transparency, safety, and ethical standards (Vinuesa et al., 2020).

The Sustainable Development Goals (SDGs) and the Paris Agreement on Climate Change require significant changes in every country that would need coordinated efforts from governments, science, civil society, and business. However, there is a lack of a shared understanding of how the 17 SDGs can be implemented. This paper presents six SDG Transformations as building blocks for achieving the SDGs, which are education, gender and inequality; health, well-being, and demography; energy decarbonization and sustainable industry; sustainable food, land, water, and oceans; sustainable cities and communities; and digital revolution for sustainable development. These transformations outline the necessary investments and regulatory challenges and require actions from specific parts of the government in collaboration with businesses and civil society (Sachs et al., 2019).

The insights into interpretable artificial intelligence (AI) models are discussed in this context. The models play a crucial role in the development of ethical AI systems and data-driven solutions that align with Sustainable Development Goals (SDGs). The potential for extracting truly interpretable models from deep-learning methods, such as symbolic models obtained through inductive biases, is highlighted to ensure the sustainable development of AI (Vinuesa & Sirmacek, 2021).

Data is widely available, and much of it is open to the public. Analyzing this data can reveal hidden and previously unknown information that can be put to purposeful use. Census data is a valuable source of information that provides data on the people living in a country, including the socio-economic status of the country. This study analyzed the 2011 Indian census to determine the socioeconomic status of different states in India by examining literacy rates, categories of workers, and gender-based working populations. The study also used data mining and machine learning techniques to identify the population's economic status, including those living below and above poverty. The results showed that k-mediods clustering method had better performance compared to K-means (Balasankar* & Varma, 2020).

In India, urban districts play a crucial role in sustainability, particularly in the context of a developing nation. India is divided into twenty-eight states and eight union territories, each of which has multiple districts governed by state laws but with varying climatic and demographic features. This study uses census data and satellite imaging data to gather information from 96 districts in India and groups them using k-means clustering. The results show that some districts in North India are grouped together based on their geography, while others are clustered together despite being geographically far apart (Budamagunta, Gaur, Verma, Anand, & Deb, 2022).

The novel coronavirus originating from Wuhan, China, led to a pandemic known as COVID-19 that spread rapidly worldwide, including in India through people with travel histories to affected countries and their contacts. The pandemic resulted in serious respiratory illness and deaths for millions of people across all states and union territories in India. In a current study, two unsupervised clustering algorithms, k-means clustering, and hierarchical agglomerative clustering, are applied to a COVID-19 dataset to group Indian states/union territories based on the impact of the pandemic and vaccination programs from March 2020 to early June 2021 (Chakraborty, 2022).

Covid-19 was declared a global pandemic by the World Health Organization (WHO) due to its rapid spread among humans. In India, one of the Asian countries that experienced a surge in Covid-19 cases, the transmission of the virus reached over 400,000 cases in one day, setting a record high. However, the problem of the spread of Covid-19 in India, which has the second-largest population in the world, continues to worsen, with a total of 21 million cases. The government of India grouped the Covid-19 cases into regions to map the spread of the virus and create strategies to prevent further spread based on the results obtained from K-Means Clustering method research in 38 regions (Gustian, Zaenal Abidin, Handayani, Hasbi, & Muslih, 2021).

In this study, National Sample Survey (NSS) data was used to measure the extent of interstate divergences in India by analyzing per capita consumption expenditure on various items, specifically food, in rural and urban regions. The aim was to raise awareness of the gravity of these divergences in a knowledge economy. To accomplish this, it was conducted Wilks' general classificatory analysis and hierarchical and k-means clustering of the states. The results showed a high degree of overlap between rural regions of certain leaders and laggard states, highlighting the need for region-specific and state-specific strategies to close the gaps between leading and laggard states (Sethi & Pandhi, 2014).

Identifying and forming groups in infectious disease data sets is a major challenge. To overcome this, data mining techniques like cluster analysis have become popular for analyzing large volumes of infectious disease data. In this study, cluster analysis was used to classify real groups of COVID-19 data in different states and union territories (UTs) in India based on their high similarity. The results allowed us to understand the clusters of affected states and UTs in India. The aim of clustering was to optimize monitoring techniques and inform government policies, medical facilities, and treatments to reduce the number of infected and deceased persons (Kumar, 2020).

The National Institution for Transforming India Ayog (NITI AYOG) is an innovation of the Modi Government, replacing the 60-year-old planning commission known as the "Super Cabinet". Unlike the planning commission, NITI

AYOG is considered a "Policy Think Tank" and provides for the state's active involvement in developing developmental policies. The Prime Minister heads the Commission, which comprises specialists from various fields. Instead of the top-down approach of the Planning Commission, NITI AYOG aims to adopt a bottom-up approach to achieve the goal of establishing "Cooperative Federalism" in the country (Shukla, 2015).

The Sustainable Development Goals (SDGs) have gained attention as a continuation of the partially successful Millennium Development Goals (MDGs) in reducing poverty and promoting human development. India has made progress in meeting its SDG targets, building on the success of the MDGs in gaining public support for development aid. The present study aims to examine the performance of the MDGs and compare it to the SDGs in India, and analyze the progress of the SDGs in the country and among Indian states using the 2019 SDG India Index report. The study found that achieving the SDGs in India requires focusing on inclusive economic growth, access to social services, infrastructure investment, and women's empowerment (K P & P, 2020).

In the planning commission era, the role of states was limited to annual interaction with the planning commission and limited participation in the National Development Council. However, with the establishment of NITI Aayog, which includes all chief ministers and administrators, states are expected to play a greater role in planning and policy implementation. The government has replaced the

Planning Commission with NITI Aayog to separate the process of governance from its strategy. The study aims to understand the challenges and opportunities of NITI Aayog and compare its functions to those of the Planning Commission (A. M, 2019).

The NITI Aayog, the new entity established to replace the Planning Commission in India, has been assigned to develop a 15-year vision, a 7-year strategy, and a 3-year implementation framework. This marks the return of planning, crucial in creating a shared commitment to a common vision and an integrated strategy. However, it is not enough to simply articulate the vision and strategy in a document and communicate it to all stakeholders. The process through which the vision and the final strategic plan are evolved and implemented is equally important (Sen, 2021).

The k-means algorithm is a widely used clustering method in machine learning and pattern recognition. Several extensions of k-means have been proposed in the literature. However, k-means and its extensions require prior knowledge of the number of clusters, making it a semi-supervised method. This study proposed a novel unsupervised k-means (U-k-means) clustering algorithm that eliminates the need for prior knowledge of the number of clusters and reduces the dependence on initializations and parameter selection. The computational complexity of the U-k-means algorithm is also analyzed, and results show that the algorithm outperforms existing methods (Sinaga & Yang, 2020).

Clustering is the grouping of data based on their similar properties, where data objects are categorized into similar groups or clusters. One of the most widely used clustering algorithms is K-means, which assigns data objects to the nearest cluster by measuring the distance between the data objects and the centroids of the clusters. In this research, the performance of the K-means algorithm with three different distance metrics was evaluated in terms of execution time with various datasets and the number of clusters. The results showed that the Manhattan distance measure metrics produced the best results in most cases and that distance metrics can impact the execution time and the number of clusters generated by the K-means algorithm (M. Ghazal et al., 2021).

The K-means clustering algorithm is widely used for its simplicity and fast convergence, but it requires the K-value to be selected in advance. To address this issue, four K-value selection algorithms were analyzed: Elbow Method, Gap Statistic, Silhouette Coefficient, and Canopy. Also, provided pseudo code for each algorithm and conducted experiments using the standard data set Iris. The results were evaluated, and the strengths and weaknesses of each of the four algorithms were discussed, as well as the appropriate clustering range for the data set (Yuan & Yang, 2019).

Clustering is a data mining technique that groups data based on similarities and variations. In the context of SMEs in Indonesia, customer mapping is necessary to identify loyal customers and improve the production of goods, specifically batik sales. To achieve this, researchers used the K-Means method combined with the Elbow technique to determine the best number of clusters for the data. The K-Means algorithm can be sensitive to the selection of the starting position, and Elbow helps to find the optimal number of clusters. Based on the results, K = 3 was the best number of clusters for studying batik visitors, with a sharp decrease in SSE values (Syakur, Khotimah, Rochman, & Satoto, 2018).

Clustering is a crucial unsupervised learning technique widely used in natural language processing. K-means is a classical algorithm in this field. However, the efficiency, correctness, and stability of K-means can be affected by the processing mode of abnormal data and the similarity calculation method used. To address these issues, this paper presents a new similarity calculation method based on weighted and Euclidean distance, which outperforms traditional K-means algorithms in terms of efficiency, accuracy, and stability (Zhao & Zhou, 2021).

The research shows that the commonly used k-means clustering algorithm, Lloyd's heuristic, can result in biased outcomes for subgroups of data, particularly in human-centric applications like resource allocation. To address this issue, an algorithm named Fair-Lloyd Was introduced, a modification of Lloyd's heuristic that provides equitable costs for different groups. The Fair-Lloyd algorithm is simple, efficient, and stable and exhibits unbiased performance compared to standard Lloyd's. The results from benchmark datasets show that Fair-Lloyd ensures equal costs for all groups with a negligible increase in running time, making it a viable and fair option for k-means clustering (Ghadiri, Samadi, & Vempala, 2021).

In this study, a clustering framework based on the transform learning paradigm was presented. The K-means clustering is performed using the representation from transform learning, and the K-means clustering loss is integrated into the transform learning framework. The joint problem is then solved with the alternating direction method of multipliers. The results on document clustering show that the approach outperforms the state-of-the-art (Goel & Majumdar, 2021).

This study presents a new initialization algorithm to improve the random initialization issue in the rough k-means clustering algorithm refined by Peters. A new method to select appropriate zeta values in the Peters algorithm is also proposed, along with a new performance measure S/O index. The performance of the proposed algorithm was compared to random initialization and the Peters algorithm, as well as popular initialization algorithms such as k-means++, Peters Π, Bradley, and Ioannis. The results showed that the proposed initialization algorithm performed better than existing algorithms with the Peters refined rough k-means clustering algorithm on various datasets and zeta values (Murugesan & Murugesan, 2020).

The article investigates the role of population migration in ensuring sustainable territorial development. It analyzes the socio-economic and migration indicators of the Russian Federation and the Republic of Kazakhstan, uncovering the features and trends of their migration processes and socio-economic development. Using the pair correlation method in SPSS 20, the research found a strong correlation between migration and socio-economic indicators. These results can guide socio-economic policy in both countries (Manukovskaya, Zhukova, & Mashukov, 2020).

In this article, the Sustainable Development Goals (SDG) framework is analyzed as a political project that faces tension between its universal and multilateral aspirations and narrow populist visions dominating global politics. Based on Laclau and Mouffe's theory of populism and their concept of 'radical democracy', the SDGs are viewed as a struggle for hegemony and competition with other political styles over what is considered 'development'. The struggle for hegemony plays out in attempts to form political constituencies behind development slogans, where religious actors become significant, given their role in organizing communities and expressing values, visions, and aspirations. By examining faith actors' engagement in India and Ethiopia and how these states have engaged with religion in defining development, the article argues that a 'radical democracy' of sustainable development requires intentional integration of religious actors in implementing the SDGs (Haustein & Tomalin, 2021).

As the deadline for achieving the United Nations Sustainable Development Goals (SDGs) approaches, the need for harmonizing economic, environmental, and social progress in economies, particularly those affected by the COVID-19 pandemic, becomes crucial. India, with 17.7% of the world's population, has a significant responsibility in achieving the SDGs and is in a strong position to create positive spillover effects. This article aims to assess the progress made toward the SDGs in India at both the global level and state levels. The global-level analysis compares India's progress with other developed and developing economies, while the state-level analysis identifies which states are doing well and which are lagging behind. The analysis is based on the Global Sustainable Development Report and the Sustainable Development Report by NITI Ayog, Government of India. The results of this analysis can aid in course correction and optimize time and resources toward achieving SDGs (Chatterjee, 2021).

Since the 1940s, the United Nations (UN) and its technical agencies and funds have led the international development agenda. The Millennium Development Goals (MDGs) generated new partnerships, increased public awareness, and demonstrated the importance of setting ambitious goals. India has made substantial progress toward the MDGs, but work is still to be done. The momentum generated by the MDGs must be sustained, and India-specific goals, targets, and indicators should be developed by the concerned ministry and states and union territories (UTs). Improving accountability at all levels is crucial. The next 15 years will see an unprecedented mobilization of resources and efforts to make the world a better place for all, particularly marginalized and disadvantaged groups. This paper explores the adoption of the MDGs to the Sustainable Development Goals (SDGs) in light of the unfinished targets of the MDGs (Jariwala, 2017).

The aim of sustainable development is to not only maintain economic growth but also improve quality of life by addressing environmental issues. The increasing concern over pollution has led international organizations such as the UN to hold various meetings, conferences, and accords. A review of international conventions showed that achieving sustainable goals is not a simple task. The study also analyzed the 17 SDGs and the policy initiatives taken by the government of India. Results showed that among the 28 states in India, Kerala, Himachal Pradesh, Andhra Pradesh, Tamil Nadu, and Telangana ranked in the top three of the SDG Index (Bharathi & Mohanasundaram, 2021).

3. Research Methodology

This research adopts a quantitative research approach, as it uses numerical data analysis to draw conclusions.

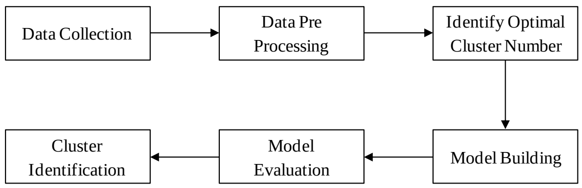


Figure 1. Proposed Methodology of Research

Source: Authors own

Data Collection: This process involved the identification of a reliable and trustworthy source. The data was collected from the NITI Aayog website of the Indian Government https://sdgindiaindex.niti.gov.in/#/. The dataset contains the latest 15 SDG scores of all the states and union territories of the country.

Data Pre-processing: This step involves cleaning, transforming, and preparing the data for analysis. Unnecessary columns and rows were removed, and the data was converted into a format that was suitable for analysis.

Identify Optimal Cluster Number: Various techniques, such as the elbow method, second differences of Hubert statistic and Dindex values, etc., were employed. These methods are appropriate as they help to determine the optimal number of clusters that best represents the data.

Model-Building: K-Means clustering was used to build the model. It is a very flexible model and one of the simplest and easiest to implement. Considering the classification of states based on their SDG scores has not been implemented before, a simple technique is enough to explore the viability and uncover the potential of similar techniques in the domain.

Model Evaluation: The silhouette score is used to evaluate the quality of clusters formed. It measures how similar an object is to its own cluster compared to others. It ranges from -1 to 1, where a score close to 1 indicates that the object is well-matched to its own cluster and poorly matched to other clusters, while a score close to -1 indicates that the object is poorly matched to its own cluster and well-matched to other clusters. The average silhouette score over all the samples in the dataset provides a general measure of the quality of the clustering solution.

Cluster Identification: This step involves identifying the different clusters formed by the model and analyzing the characteristics of each cluster. This step is important as it provides insights into the conditions of states in each cluster, creating a clearer picture of how these clusters should be approached in terms of addressing their SDG needs.

4. Problem Statement

Attending to the needs of improvement of specific SDG goals of each state/union territory of India is very resource intensive and time-consuming for the central government.

The goal of the research is to understand the regional disparities in SDG implementation across the country and to identify clusters of states with similar SDG scores. It is designed to provide value to the government by generating insights into the SDG performance of states. By performing clustering on this data, the government can better allocate resources and implement policies that effectively address the challenges faced by different regions in achieving the SDGs. This information can target resources and interventions in clusters of states with similar needs, thus reducing the logistical burden on the central government to plan and carry out the solutions, which is a novel and valuable contribution. The ultimate aim is to improve the composite scores, leading to a more sustainable and equitable future for all citizens of India.

5. Objectives

- To provide a comprehensive analysis of the state-level performance on Sustainable Development Goals (SDGs) in India.

RossiSmith Academic Publications, Oxford/London UK, www.publicationsales.com

- To identify strengths and challenges faced by different clusters of states in achieving SDGs and develop strategies for improvement.
- To understand the regional disparities in SDG implementation and suggest measures for reducing such gaps.

By achieving these objectives, the research paper can help the government in several ways. Firstly, it can provide valuable insights into the current status of SDG implementation in the country, which can be used to develop more effective policies and strategies. Secondly, it can identify the areas where specific states need to improve and provide recommendations for doing so. This can help the government to allocate resources more efficiently and effectively. Finally, by highlighting regional disparities, the research can suggest ways to reduce such gaps, which is essential for promoting sustainable development across the country.

6. Analysis

The dataset consists of scores for the 15 Sustainable Development Goals set by the United Nations. The scores for twenty-eight states and eight union territories in the Indian subcontinent are taken for analysis with no missing values. The scores are on a scale of 100 for each SDG. The clustering technique divides the data into

Journal of Asia Entrepreneurship and Sustainability
www.asiaentrepreneurshipjournal.com

groups based on the similarities in the data points within the same group, which helps in analyzing and understanding the patterns and relationships within the data.

6.1. Exploratory Data Analysis

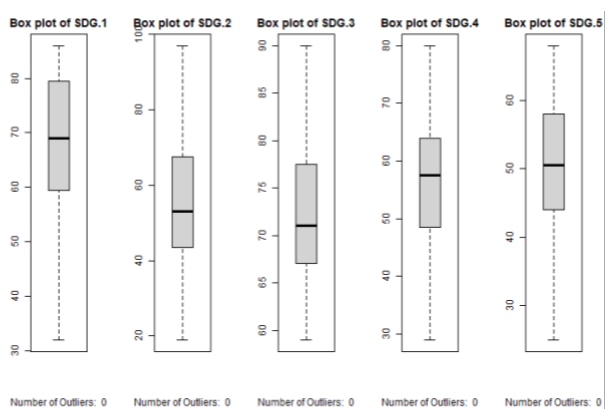
```
     SDG.1          SDG.2          SDG.3          SDG.4          SDG.5          SDG.6
 Min.   :32.00  Min.   :19.00  Min.   :59.00  Min.   :29.00  Min.   :25.00  Min.   : 54.00
 1st Qu.:59.75  1st Qu.:43.75  1st Qu.:67.00  1st Qu.:48.75  1st Qu.:44.50  1st Qu.: 82.75
 Median :69.00  Median :53.00  Median :71.00  Median :57.50  Median :50.50  Median : 87.00
 Mean   :66.50  Mean   :55.47  Mean   :71.81  Mean   :57.11  Mean   :50.11  Mean   : 84.56
 3rd Qu.:79.25  3rd Qu.:66.75  3rd Qu.:77.25  3rd Qu.:64.00  3rd Qu.:58.00  3rd Qu.: 90.25
 Max.   :86.00  Max.   :97.00  Max.   :90.00  Max.   :80.00  Max.   :68.00  Max.   :100.00
     SDG.7            SDG.8          SDG.9           SDG.10           SDG.11          SDG.12
 Min.   : 50.00  Min.   :36.00  Min.   :23.00  Min.   : 41.00  Min.   :39.00  Min.   :47.00
 1st Qu.: 84.50  1st Qu.:53.75  1st Qu.:36.75  1st Qu.: 64.75  1st Qu.:64.00  1st Qu.:65.75
 Median :100.00  Median :59.50  Median :46.50  Median : 67.50  Median :76.00  Median :76.50
 Mean   : 92.33  Mean   :59.83  Mean   :48.14  Mean   : 67.53  Mean   :72.22  Mean   :74.33
 3rd Qu.:100.00  3rd Qu.:65.25  3rd Qu.:60.25  3rd Qu.: 72.50  3rd Qu.:81.00  3rd Qu.:82.00
 Max.   :100.00  Max.   :78.00  Max.   :72.00  Max.   :100.00  Max.   :98.00  Max.   :99.00
     SDG.13          SDG.15          SDG.16
 Min.   :16.00  Min.   :27.00  Min.   :46.00
 1st Qu.:43.75  1st Qu.:57.50  1st Qu.:69.75
 Median :58.00  Median :64.00  Median :73.00
 Mean   :53.28  Mean   :64.53  Mean   :72.58
 3rd Qu.:63.50  3rd Qu.:72.25  3rd Qu.:77.50
 Max.   :77.00  Max.   :93.00  Max.   :86.00
```

Figure 2. Summary Statistics.   Source: Authors own

The exploratory data analysis on the dataset gives insights into the overall score of each sustainable development goal for India and where it needs improvement. The lowest score for all SDGs is clearly 16 to 59, with the highest score ranging from 68 to 100. This illustrates that achievement for each SDG varies greatly between states and union territories.

For all SDGs, the first quartile (25th percentile) spans from 43.75 to 82.75, suggesting that at least 25% of states and union territories have a score below this range. For each SDG, the median score (50th percentile) runs from 47 to 87, and the third quartile (75th percentile) extends from 64 to 90.25, meaning that half of the states and union territories have a score above the median and at least 25% have a score above the third quartile. The average (mean) score for all SDGs varies from 50.11 to 84.56. This means that the average performance for each SDG is in the center of the range of scores for all states and UTs.

SDG 7, which stands for ‘Affordable and Clean Energy’, has the highest mean score of 92.33 among other SDG scores, and SDG 9, which is ‘Industry, Innovation and Infrastructure’ has the least score of 48.13 and is the only SDG that has a score below fifty. This indicates that India, in general, has to focus more on its Industry, Innovation, and Infrastructure development and should form policies and take measures to boost the same.



6.2. Outlier Analysis

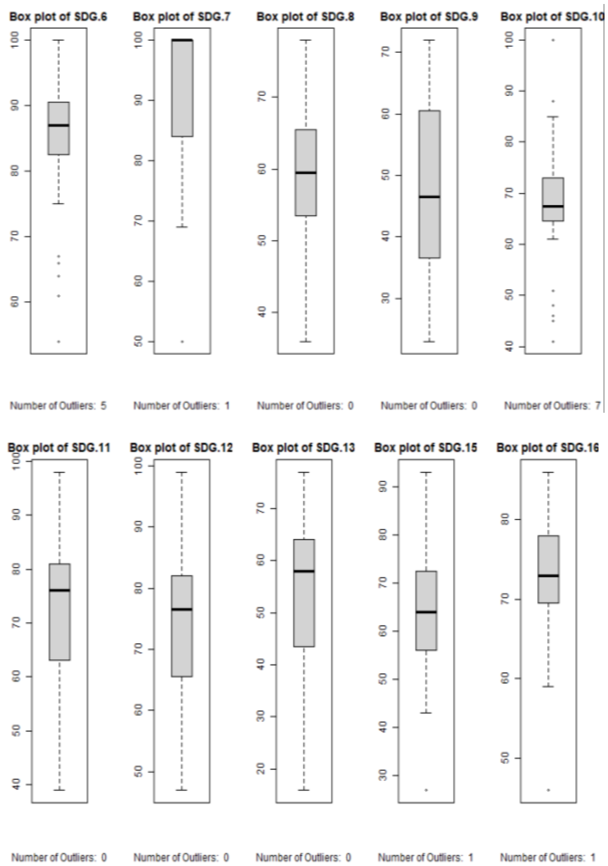


Figure 3. Boxplots Showing Outliers Among Each SDG.   Source: Authors own

From the above box plots, various insights can be drawn for different Indian states. The following list discusses specific SDGs and their underlying outlier states in detail:

a. Considering SDG 6, which guarantees accessibility and sustainability in the provision of water and sanitation for everyone, the outlier states are Andhra Pradesh, Punjab, Rajasthan, Delhi, and Assam. These states have a considerably lower score for this goal than other Indian states. This includes water scarcity, contaminated water sources, insufficient infrastructure, poor sanitation practices, overuse of groundwater, and unplanned urbanization. To achieve the goal of sustainable water and sanitation management for all, it is important to address these challenges.

b. Considering SDG 7, the goal is to provide access to sustainable and modern energy for all. To achieve the goal of providing access to sustainable and modern energy for all, Indian states like Meghalaya face several challenges that contribute to low scores for SDG 7. These challenges include limited access to electricity, inadequate energy infrastructure, dependence on traditional energy sources, insufficient investment in renewable energy, and inefficient energy use. Financial constraints also pose a challenge, limiting the state's ability to invest in more sustainable forms of energy. Addressing these issues will be crucial in ensuring the achievement of SDG 7 in Indian states like Meghalaya.

c. SDG 10 aims to reduce income inequality within and among countries. Indian states can be divided into two segments with regard to their performance on SDG 10 - some states perform exceptionally well while others perform poorly compared to the rest. In Chandigarh, the administration has taken active steps to drive SDG initiatives. A dedicated team headed by the Finance Secretary has been formed, and implementing departments of SDGs are working under the Chairpersonship of the Advisor to the Administrator. The administration has also allocated sector-wise resources to identify any resource gaps. To address the issue of income inequality, the Chandigarh Administration has designed strategies for identifying vulnerable populations. Based on these strategies, various schemes are being implemented to support the vulnerable population. The government of Meghalaya has formed an SDG cell at the state level to coordinate the preparation of its Vision document and SDG implementation. The state has completed the mapping of government schemes against the SDGs. However, in Indian states like Bihar, Madhya Pradesh, Nagaland, Rajasthan, and Uttar Pradesh, there are several factors that contribute to lower scores for SDG 10. These include high poverty levels, unequal distribution of wealth, lack of access to education and resources, weak social protection programs, and inefficient labor policies. Addressing these challenges is essential in order to reduce income inequality and achieve the goals of SDG 10.

d. The Indian state of Ladakh faces several challenges that result in lower scores for SDG 15, which aims to protect and promote sustainable use of terrestrial ecosystems. Unsustainable land practices, the impact of climate change, limited resources, ineffective policies, and lack of public engagement are some of the factors. Addressing these issues is crucial in achieving the goal of sustainable management of forests, combating desertification, halting land degradation and biodiversity loss, and promoting sustainable use of terrestrial ecosystems.

e. The Union Territory of Andaman and Nicobar stands out when it comes to SDG 16, which promotes peace, justice, and effective governance. The territory has developed information, education, and communication (IEC) materials for distribution to various departments. Awareness campaigns, advertisements, street plays, and community meetings have been conducted to educate stakeholders. However, a challenge still remains in accessing knowledgeable individuals.

6.3. Optimal Number of Clusters

Proceeding with further analysis, the elbow method helped to identify the optimal number of clusters that can be formed using the dataset under consideration. Based on the Hubert Statistic and Second Difference Dindex values, it is determined that the optimal number of clusters should be three. The below images show the different test results.

RossiSmith Academic Publications, Oxford/London UK, www.publicationsales.com

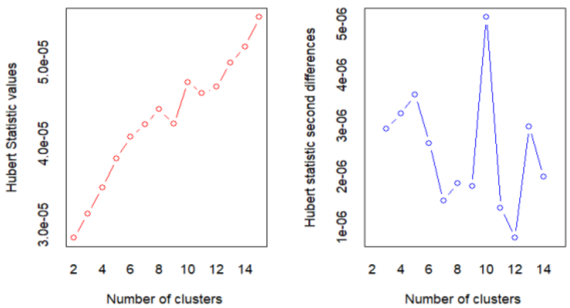


Figure 4. Hubert Statistics

Source: Author’s own

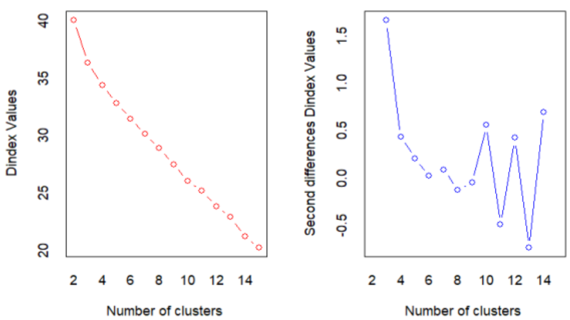


Figure 5. Dindex Values. Source: Authors own

7. Results and Discussion

In the analysis of SDG performances of twenty-eight states and eight union territories of India, three major clusters were identified. These clusters have a silhouette score of 0.3439728, suggesting that there is a reasonable amount of ambiguity in the positioning of certain states/union territories in the clusters. Hence, it is important to note that the clusters identified in this study may not be the most optimal classification. A related study using a different clustering technique might result in better results.

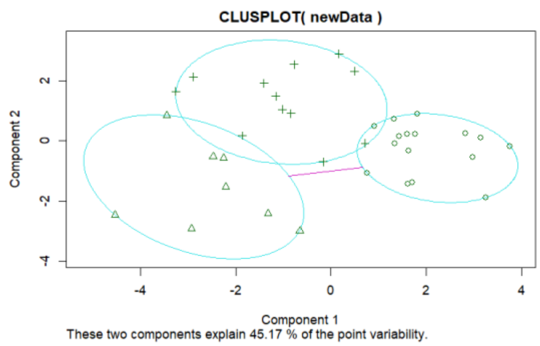


Figure 6. Bivariate Cluster Plot of the three Clusters.   Source: Authors own

Journal of Asia Entrepreneurship and Sustainability
www.asiaentrepreneurshipjournal.com

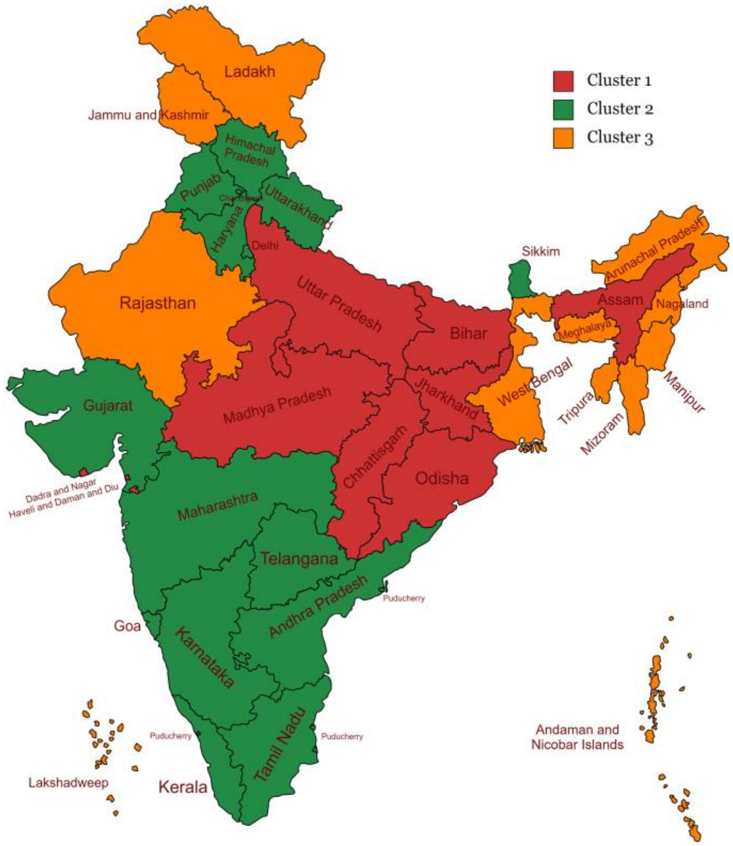


Figure 7. Visualization of the three Clusters. Source: Authors own

Clusters	SDG.1	SDG.2	SDG.3	SDG.4	SDG.5	SDG.6	SDG.7	SDG.8	SDG.9	SDG.10	SDG.11	SDG.12	SDG.13	SDG.15	SDG.16
1	44	39	64	45	50.5	87	79	53.5	38	65	74	65	38.5	68	70.5
2	75	60	77	66.5	56	89	100	66.5	62.5	70	79	76.5	60.5	65.5	73
3	70	64	70	51.5	46	84.5	97	57	35	68	57	83	62.5	61.5	74

Figure 8. SDG Median Scores of each Cluster. Source: Authors own

7.1. Cluster 1

The following table contains the list of states/union territories in the first cluster.

Cluster 1	
Assam	Madhya Pradesh
Bihar	Odisha
Chhattisgarh	Uttar Pradesh
Jharkhand	Dadra and Nagar Haveli and Daman and Diu

Table 1. Cluster 1 Regions

Source: Authors Own

Cluster 1 consists of most regions in Central-Eastern India and is doing considerably worse than the other two clusters. These regions perform quite poorly with SDGs 1, 2, 4, 9, and 13 while barely scoring above fifty in SDGs 5 and 8. These regions require major support and focus from the central government to address poverty, hunger, education, industry infrastructure, and climate-related issues. The government should give first priority to the regions under this cluster.

7.2. Cluster 2

The following table contains the list of states/union territories in the second cluster.

Cluster 2	
Andhra Pradesh	Punjab
Goa	Sikkim
Gujarat	Tamil Nadu
Haryana	Telangana
Himachal Pradesh	Uttarakhand
Karnataka	Chandigarh
Kerala	Delhi
Maharashtra	Puducherry

Table 2. Cluster 2 Regions. Source: Authors own

Cluster 2 consists of Southern India and the neck region of the country. This cluster of regions seems to be performing well overall, with slight concerns related to SDGs 2 and 5, which are associated with hunger and gender equality. The government does not have to divert much focus and resources into these regions. These regions are also optimal for different businesses to set up, meaning the government can promote these regions to outside businesses.

7.3 Cluster 3

The following table contains the list of states/union territories in the third cluster.

Cluster 3	
Arunachal Pradesh	Tripura
Manipur	West Bengal
Meghalaya	Andaman and Nicobar Islands
Mizoram	Jammu and Kashmir
Nagaland	Ladakh
Rajasthan	Lakshadweep

Table 3. Cluster 3 Regions. Source: Authors own

This cluster consists of regions that are relatively hard to traverse due to their challenging terrains. For example, Rajasthan is mostly desert; Jammu & Kashmir and Ladakh are frozen deserts; the Northeast states are predominantly mountainous; and finally, Andaman & Nicobar and Lakshadweep are islands. Although the terrain is not favorable to this cluster, the regions are doing relatively better than the first cluster.

These regions are particularly performing poorly in SDG 9, related to industry innovation and infrastructure. This makes perfect sense considering that the geography of these regions makes it extremely difficult to make progress in this specific scenario. The government has limited capacity to tackle this problem, considering it a geographical issue. But, some strategies could be explored that could make use of this unconventional geography to the maximum extent. Moreover, these regions seem to be struggling with SDG 5, which is gender equality. This could be attributed to how most of these regions are still rooted in their traditions, and a slow transition through policies and legislation could improve this factor.

8. Limitations

This study is meant to explore the potential application of clustering techniques on the SDG performance of regions and hence is limited in many ways. Following are some of the limitations identified:

- Irrelevant SDG Goals: SDGs 14 and 17 are missing from the dataset. SDG 14 deals with "Life Below Water" which has certain parameters that are only applicable to states containing coastlines, making the generalization of the SDG with other states invalid. SDG 17, being "Partnership for Goals", is only applicable to countries and not relevant to states. Because of these conditions, a complete state-wise analysis of SDGs is drastically limited.
- Cluster Quality: The quality of clusters generated in this research is found to be sub-optimal. Although the clusters are serviceable, better results may be possible using different techniques.

- Geographic Data: The dataset used for this study does not include the geographic data of the regions. Because of this, the clusters are formed without considering the spatial distribution of the regions, which could be an important factor in determining the logistical stress of plans catering to these clusters.
- SDG Subgoals: This study has not taken into account the many subgoals of each SDGs. This study is looking at the overall SDGs individually and hence is not taking into consideration the granular details that are contained within the subgoals.

9. Conclusion

This paper presents a novel attempt to explore the application of clustering techniques in identifying major clusters of regions in India based on their SDG performance. This information is useful for the central government to carry out development programs, mainly assisting in driving down the logistical pressure of planning and executing said programs. Some businesses can also use this information to identify the optimal regions to invest in and expand their businesses into.

The study identified three main clusters of regions using the K-means algorithm. This study is meant to explore the potential of this domain and so is limited by many constraints, which are discussed earlier in the limitation section. This implies there is huge scope for further research that addresses these limitations. For example, instead of using K-means clustering, which is the simplest format of clustering techniques, a more complex algorithm might be able to create better-quality clusters that could address the logistical issues more efficiently. Similarly, a follow-up study could take into account the sub-goals within each SDG and do a more extended analysis of the topic, which could give more detailed and accurate clusters.

10. References

Shulla, K., Voigt, B. F., Cibian, S., Scandone, G., Martinez, E., Nelkovski, F., & Salehi, P. (2021, March 17). Effects of COVID-19 on the Sustainable Development Goals (SDGs). Discover Sustainability, 2(1). https://doi.org/10.1007/s43621-021-00026-x

Dantas, T., de-Souza, E., Destro, I., Hammes, G., Rodriguez, C., & Soares, S. (2021, April). How the combination of Circular Economy and Industry 4.0 can contribute towards achieving the Sustainable Development Goals. Sustainable Production and Consumption, 26, 213–227. https://doi.org/10.1016/j.spc.2020.10.005

Del Río Castro, G., González Fernández, M. C., & Uruburu Colsa, N. (2021, January). Unleashing the convergence amid digitalization and sustainability towards pursuing the Sustainable Development Goals (SDGs): A holistic review. Journal of Cleaner Production, 280, 122204. https://doi.org/10.1016/j.jclepro.2020.122204

Zhan, J.X., Santos-Paulino, A.U. Investing in the Sustainable Development Goals: Mobilization, channeling, and impact. J Int Bus Policy 4, 166–183 (2021). https://doi.org/10.1057/s42214-020-00093-3

Mhlanga, D. (2021, May 21). Artificial Intelligence in the Industry 4.0, and Its Impact on Poverty, Innovation, Infrastructure Development, and the Sustainable Development Goals: Lessons from Emerging Economies? Sustainability, 13(11), 5788. https://doi.org/10.3390/su13115788

Vinuesa, R., Azizpour, H., Leite, I., Balaam, M., Dignum, V., Domisch, S., . . . Fuso Nerini, F. (2020). The role of artificial intelligence in achieving the Sustainable Development Goals. Nature Communications, 11(1). https://doi.org/10.1038/s41467-019-14108-y

Sachs, J. D., Schmidt-Traub, G., Mazzucato, M., Messner, D., Nakicenovic, N., & Rockström, J. (2019). Six Transformations to achieve the Sustainable Development Goals. Nature Sustainability, 2(9), 805–814. https://doi.org/10.1038/s41893-019-0352-9

Vinuesa, R., & Sirmacek, B. (2021). Interpretable deep-learning models to help achieve the Sustainable Development Goals. Nature Machine Intelligence, 3(11), 926–926. https://doi.org/10.1038/s42256-021-00414-y

Balasankar*, V., & Varma, D. P. S. (2020). Socio-Economical Status of India using Machine Learning Algorithms. International Journal of Recent Technology and Engineering (IJRTE), 8(5), 3804–3811. https://doi.org/10.35940/ijrte.e6610.018520

Budamagunta, M., Gaur, A., Verma, A., Anand, A., & Deb, C. (2022). Clustering Indian districts based on multidimensional demographic and climate data. Proceedings of the 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation. https://doi.org/10.1145/3563357.3566146

Chakraborty, S. (2022). Monitoring COVID-19 Cases and Vaccination in Indian States and Union Territories Using Unsupervised Machine Learning Algorithm. Annals of Data Science. https://doi.org/10.1007/s40745-022-00404-w

Gustian, D., Zaenal Abidin, M., Handayani, S., Hasbi, A., & Muslih, M. (2021). Mapping Spread Virus COVID-19 in India Using K-Means Clustering. 2021 IEEE 7th International Conference on Computing, Engineering and Design (ICCED). https://doi.org/10.1109/icced53389.2021.9664839

Sethi, A. S., & Pandhi, R. (2014). Interstate Divergences in Nutritional Expenditure in India: A Cluster Analysis Approach. Poverty & Public Policy, 6(1), 80–97. https://doi.org/10.1002/pop4.63

Kumar, S. (2020). Monitoring Novel Corona Virus (COVID-19) Infections in India by Cluster Analysis. Annals of Data Science, 7(3), 417–425. https://doi.org/10.1007/s40745-020-00289-7

Shukla, O. (2015). Niti Ayog-A step towards transformation. Asian Man (the) - an International Journal, 9(2), 267. https://doi.org/10.5958/0975-6884.2015.00041.9

K P, V. C., & P, S. (2020). Transition From Millennium Development Goals To Sustainable Development Goals: A Close Scrutiny From India. International Journal Of Research -Granthaalayah, 8(5), 78–88. https://doi.org/10.29121/granthaalayah.v8.i5.2020.72

A. M, M. S. (2019). The Challenges and Opportunities of NITI Aayoga. International Journal for Research in Applied Science and Engineering Technology, 7(6), 472–475. https://doi.org/10.22214/ijraset.2019.6084

Sen, P. (2021). Plan, but Do Not Over-plan: Lessons for NITI Aayog. Planning in the 20th Century and Beyond, 264–282. https://doi.org/10.1017/9781108859448.013

Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-Means Clustering Algorithm. IEEE Access, 8, 80716–80727. https://doi.org/10.1109/access.2020.2988796

M. Ghazal, T., Zahid Hussain, M., A. Said, R., Nadeem, A., Kamrul Hasan, M., Ahmad, M., . . . Tahir Naseem, M. (2021). Performances of K-Means Clustering Algorithm with Different Distance Metrics. Intelligent Automation & Soft Computing, 29(3), 735–742. https://doi.org/10.32604/iasc.2021.019067

Yuan, C., & Yang, H. (2019). Research on K-Value Selection Method of K-Means Clustering Algorithm. J, 2(2), 226–235. https://doi.org/10.3390/j2020016

Syakur, M. A., Khotimah, B. K., Rochman, E. M. S., & Satoto, B. D. (2018). Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. IOP Conference Series: Materials Science and Engineering, 336, 012017. https://doi.org/10.1088/1757-899x/336/1/012017

Zhao, Y., & Zhou, X. (2021). K-means Clustering Algorithm and Its Improvement Research. Journal of Physics: Conference Series, 1873(1), 012074. https://doi.org/10.1088/1742-6596/1873/1/012074

Ghadiri, M., Samadi, S., & Vempala, S. (2021). Socially Fair k-Means Clustering. Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency. https://doi.org/10.1145/3442188.3445906

Goel, A., & Majumdar, A. (2021). Transformed K-means Clustering. 2021 29th European Signal Processing Conference (EUSIPCO). https://doi.org/10.23919/eusipco54536.2021.9616177

Murugesan, V. P., & Murugesan, P. (2020). A new initialization and performance measure for the rough k-means clustering. Soft Computing, 24(15), 11605–11619. https://doi.org/10.1007/s00500-019-04625-9

Manukovskaya, L., Zhukova, M., & Mashukov, V. (2020). The impact of the socio-economic policies of states on population migration as a factor in the

sustainable development of territories. E3S Web of Conferences, 208, 08014. https://doi.org/10.1051/e3sconf/202020808014
Haustein, J., & Tomalin, E. (2021). Religion, Populism, and the Politics of the Sustainable Development Goals. Social Policy and Society, 20(2), 296–309. https://doi.org/10.1017/s147474642000072x
Chatterjee, S. (2021). Implementing Sustainable Development Goals in India: Progress So Far. South Asian Journal of Social Studies and Economics, 1–9. https://doi.org/10.9734/sajsse/2021/v12i330304
Jariwala, V. (2017). Marching Towards Millennium Development Goals (MDGs) to Sustainable Development Goals (SDGs) in Indian Context. SSRN Electronic Journal. https://doi.org/10.2139/ssrn.3069427
Bharathi, R., & Mohanasundaram, V. (2021). International Conventions On Sustainable Development And india's Strategies. Indian Journal Of Applied Research, 46–48. https://doi.org/10.36106/3217429

Journal of Asia Entrepreneurship and Sustainability
www.asiaentrepreneurshipjournal.com

MOOCs: A Tool for Education 4.0 and Its Appeal to B-School Students

Latha K
Post-Doctoral Fellow of KSHEC, Govt. of Kerala, Mahatma Gandhi University, Kottayam, Kerala, India (corresponding) lathabrijesh2@gmail.com

Siby Yohannan
FPM Scholar, Rajagiri Centre for Business Studies, Autonomous Institute affiliated with Mahatma Gandhi University, Kochi, Kerala. fpm22m0010@rajagiri.edu

Abstract

In order to prepare the young talents for Industry 4.0, the vast opportunities made possible by technological advancements must be effectively explored through education 4.0, a revolution in the education sector. Massive Open Online Courses (MOOCs) are a cutting-edge educational approach that has received a lot of attention recently. Despite an increasing number of enrolled users, many stopped utilising MOOC platforms after their initial experience; as a result, the reduced attractiveness of MOOCs has seriously hampered the long-term growth of MOOC platforms. The purpose of this study was to identify the components that B-school students need for MOOCs to be more appealing as well as the effects of various variables on students' intrinsic motivation to enrol in a MOOC, based on Kirkpatrick’s Model. A survey among B-school students in Kerala who had used MOOC platforms yielded a total of 550 validly completed questions. Across different income levels, there is no statistically significant difference in the intrinsic desire of students to enrol in online learning. The learning analytics results show that Self Directed Learning Environments (SDLE) and User-Friendly Design (UFD) have a substantial relationship with the Attractiveness of MOOCs (AOM), whereas Participation, Interaction, and Instructor Guidance (PIIG) and Internet speed (IS) have a moderate relationship. A PLS bootstrapping algorithm with 550 samples reveals statistically significant associations between each factor and the dependent variable (AOM) with t values significant at t > 1.96 at P 0.05. The R square number of the dependent variable reveals that all of the independent variables account for more than 52% of the total variance. The study's practical application entails assisting online learning platform developers as it fills in the gaps in theory regarding the factors influencing MOOCs' allure.

1. Introduction

India, which has the world's most extensive higher education system, with approximately 40,000 colleges and 993 universities serving 37.4 million students, bears a great deal of responsibility in the educational sector (Salwan et al., 2021). India used traditional teaching methods to train young minds in higher education institutes, but that time has passed, as online learning for students has become a popular method of providing education at the postgraduate and undergraduate levels. We have a small impact because the mode of learning is online, i.e., an extension of distance learning, which essentially requires a medium of new modes of presentation and interaction with the young generation.Despite being the world's second largest market for e-learning, India's gross enrolment ratio in higher education is 26%, just above the global average of 24% and far lower than China's gross enrolment ratio of 51%. (Lee et al., 2016). The Indian higher education sector is facing enormous challenges as a result of recent developments such as COVID-19, fee increases at higher education institutions, and so on. While the issue of higher education fee increases was raging in our country, the COVID-19 pandemic fueled the fire by disrupting the entire education system (Association of Indian Universities, 2020). According to a UNESCO (2020) online study, the closure of universities, colleges, and schools has harmed more than 70% of the world's student population. The lockdown, followed by an extension, and subsequent pandemic waves exacerbated the difficulty in accessing necessary study materials and other basic needs in the education industry. The emergence of online education for students provided solace in the midst of the pandemic. However, the question is how to provide online education to all students in the country, as there are still students who do not have full internet connectivity, just as there are homes that do not have electricity. As a result, the government and universities face a significant challenge

With Industry 4.0 completely disrupting manufacturing, services, and entrepreneurship and making them almost entirely reliant on technology, it is time for business schools to catch up and prepare their cohorts for the digital future. To fully capitalize on the opportunities created by advanced technology, a similar revolution in education, education 4.0, is required—not only to meet the needs of industry, but also to ensure the best possible student experience, use of staff time, and investment in infrastructure and facilities.

Even though pedagogy, curriculum, and assessment still require a lot of fine tuning to understand and appreciate how the industry of the future may work, the COVID years actually helped the education community from the ground up to use digital tools. Technology has unquestionably played an important role in the development and expansion of online education, and it continues to do so. MOOCs (Massive Open Online Courses) are gaining popularity in today's educational system. Because of the COVID pandemic, student acceptance of MOOCs has increased, and it is critical to understand the various factors that contribute to the attractiveness and quality of such courses. The current study aims to determine which factors are required to make MOOCs more appealing to B-school students. The study is being conducted among B-school students in order to identify the factors required to improve the attractiveness of MOOCs and the impact of selected variables on students. The study is more relevant in the current scenario because the educational system changed existing education methods during the COVID-19 period, and students are required to find online education methods for studying. In this situation, it is necessary to identify which factors need to be improved in order to have a better study experience.

2. Review of Literature

According to a BCG (2014) study, advanced teaching methods combined with employer concerns about graduates' skill sets will undoubtedly propel online teaching and learning to new heights. However, just as a live music concert attracts a larger audience than an online platform, online education has inherent drawbacks because education is not the sole goal of students attending a traditional business school or university. It is always preferable to see our favourite hero on stage rather than on an online platform, and the same is true for online learning (Kaplan 2012). Online education does not cover socialising or gaining life experience.

It is known that positive motivation leads to positive engagement and negative motivation leads to negative engagement in MOOC learners.Chen et al., (2015) looked into whether positive motivation leads to positive engagement and negative motivation leads to negative engagement. The findings show a strong relationship between these engagements and MOOC learners. Laxmisha et al. (2016) investigated learner interests and comprehension in such modes. The data is based on results available from the world's most prestigious universities since the inception of MOOCs, as well as general opinions derived from the perspectives of MOOC learners. Individual student behaviour has a large influence on learning. This may not be the case in classroom-based learning or time-tested traditional learning practises, where multiple students learn together and teachers influence underperforming students to perform better

Jena et al. (2020) emphasized the importance of online learning during times of crisis, such as job absences or pandemics. The Indian government's emerging approaches to online learning are also discussed. This study also considers the benefits and drawbacks of using an online learning platform. John Francis et al. (2021) attempted to identify the factors influencing student acceptance of an online learning platform during the COVID-19 pandemic. The findings of this study include a list of various factors that a student should consider when selecting a course. The research of Abdulsalam K. (2021) has two main goals. To begin, it was necessary to determine whether university students' use of online learning platforms during the COVID-19 period posed any learning challenges. Second, in response to proposed solutions, develop a conceptual model to mitigate the impact of such difficulties.

Online learning is needed to address student learning activities during the COVID pandemic. According to Hasan et al. (2020), online learning is one of the solutions required by educators and students in the COVID pandemic that occurred in February 2020. The purpose of this study was to determine the effectiveness of online learning using the Zoom cloud meeting application in resolving the problem of student learning activities at Nurul Jadid University during the COVID-19 pandemic.

During COVID-19, Mohd Hafiz Hanafiah (2020) conducted research on the use of Massive Open Online Courses (MOOCs) by Malaysian higher education institutions. MOOCs have proven to be a stable platform capable of providing independent and flexible learning experiences to higher education students, according to the findings of this paper. Because of this unprecedented occurrence, university administration has implemented a backup plan to move all academic classes online. Aayat Amin Aljarrah's (2020) research delves into the definition, characteristics, and patterns of MOOCs. This paper also discusses the specific advantages and disadvantages of participating in MOOCs, as well as the various challenges that MOOCs face. Furthermore, various improvement suggestions will be made to aid in the process of improving MOOCs. Finally, in the COVID-19 era, this paper discusses the value of MOOCs. This paper will assist instructors in understanding how to make MOOCs more efficient, as well as learners in becoming more organised and understanding MOOCs. It will also help education system experts find appropriate solutions to the problems that MOOCs are experiencing.

Stakeholders play an important role in developing intellectual human assets to increase enrollment and student participation rates of MOOCs. Anand et al. (2020) investigated the effects of massive open online courses (MOOCs) on students at higher educational institutions in India using the COVID-19 pandemic scenario. The government, higher education institutions, and MOOC providers must all play a significant role in the development of intellectual human resources for the growth of the country. MOOCs must improve sound design, quality and accessible delivery, multilingual facilitation, and efficient regulation to be successful in India. Despite their negative connotation, they help many students and working professionals gain knowledge and land a good job. Rekha et al., 2022 examined MOOC characteristics from an Indian perspective. They have identified that that Indian MOOCs may include subjects that Western universities do not typically offer, such as Indian classical music and dance. Furthermore, MOOCs are critical to the successful implementation of a choice-based credit system (CBCS). Trehan (2017) states: “I looked into the critical discussions about MOOCs in India and China and came to the conclusion that certain aspects needed to be improved.

Sound design, quality and accessible delivery, multilingual facilitation, and effective regulation of MOOC credits are just a few of the components that need to be improved. When looking at MOOCs from an Indian perspective, Haumin et al. (2019) found that the infrastructure facilities must be improved for MOOCs to be successful there. Despite their negative connotation, MOOCs help many students and working professionals gain additional knowledge and land a good job.

MOOCs are becoming increasingly popular in India. MOOCs contain a wide range of web-based learning materials and assets; in this particular circumstance, considering the development of MOOCs in India, it is now critical to assess the level of awareness of MOOCs among understudies overall amid this COVID-19 pandemic Hanifa Arrumaisha (2022) found that a client's satisfaction with a MOOC has a direct impact on their desire to continue learning.

Madhuri et al.,2018 in their studies attempted to educate those who are unfamiliar with MOOCs and to discuss their impact on management students. The paper will go over why many students prefer MOOCs to traditional classrooms, their impact on management students, what they think about them, whether they help with employment opportunities, and other MOOC-related topics.

3. Research Methodology

The Kirkpatrick model is used to perform the study among B-school students and the student population who participate in MOOCs in Kerala, India. According to the Kirkpatrick model, the first level of evaluation looks at how the students responded to the programme, the second level gauges "the extent to which students have learned," level three assesses "participant's behaviour," and the final level considers the course's overall effect.

3.1 The model's framework

The Goh et al., (2018) paradigm was modified and put to the test in assessing the appeal of MOOCs provided to B-Schools. Due to the evaluators' perception of their relative significance, two constructs—participation, interaction, and instructor guidance—were combined into a single category in the framework. Two additional constructs were added for a more thorough assessment.

3.2 Research Model

A methodology is suggested to evaluate how appealing MOOCs are (Fig. 1). The model's objective is to identify the factors that collectively affect AOM and the factors that can be enhanced to make MOOCs more alluring. Self-Directed Learning Environment (SDLE), User-Friendly Design of Course Content (UFD), Participation Interaction and Guidance from Teachers (PIIG), Internet Speed (IS), Course Level (CL), and Perceived Difficulty are the main features included in the model (PD).

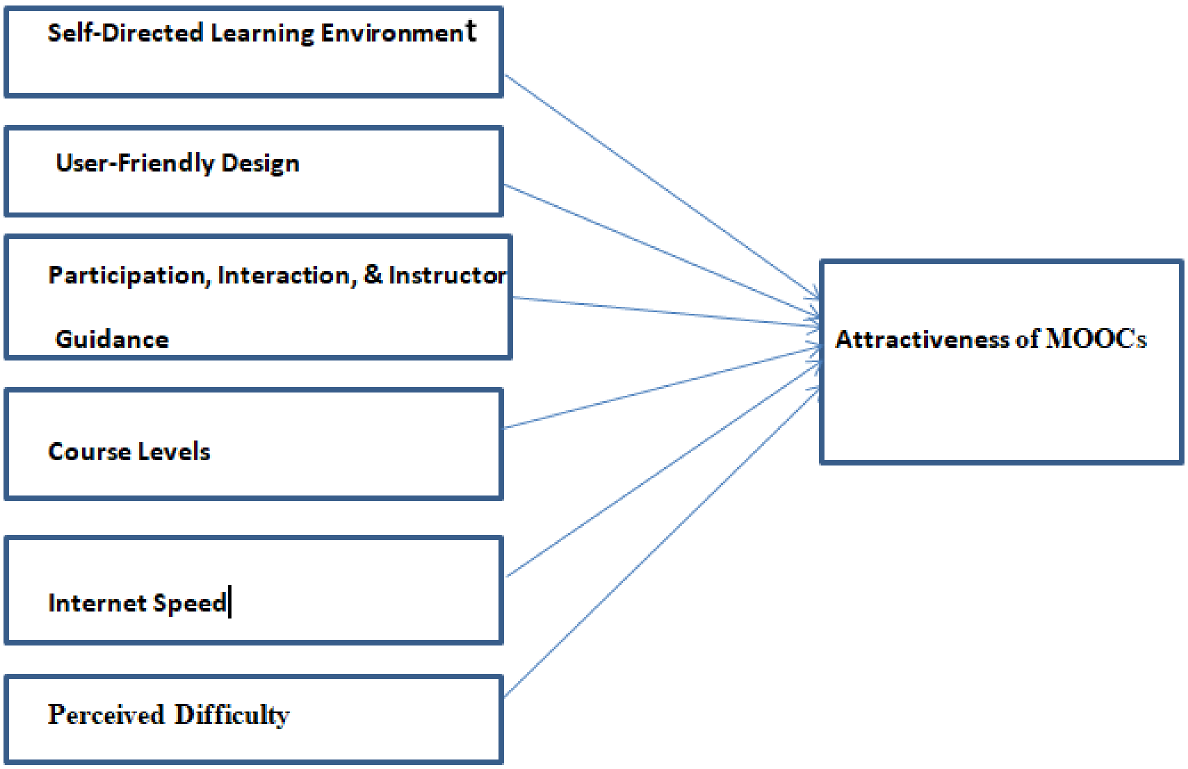


Fig.1 Research Model

3.3 Research Hypotheses

The effective completion of MOOCs due to a self-directed learning environment has been studied by Zhu et al. (2020). (SDLE) The roles and responsibilities of teachers and learners in MOOC learning environments are very different from those in conventional classrooms, according to Garrison's (1997) SDLE model; effective learners in MOOC learning environments must be self-directed. As the three main elements of self-directed learning, he considered incentive, self-monitoring, and self-management. showed that self-monitoring was directly impacted by motivation, and self-management was indirectly impacted by motivation. Additionally, self-monitoring had a favourable impact on self-management. Therefore, encouraging pupils and helping them develop self-monitoring skills are crucial. On how to help MOOC students support and facilitate their own self-monitoring, more study is required. So, using the Krickspatric model, the current research looked at how a self-directed learning environment and MOOC attraction are related, with a focus on learner behaviour in MOOCs. Consequently, the following first theory is put forth:

H1: A Self-Directed Learning Environment (SDLE) is positively associated with the Attractiveness of MOOCs (AOM).

Learners in MOOCs will be able to engage more freely and on their own terms as more potent and adaptable web technologies are released (Lin et al., 2013). According to De et al. (2021), one method for reducing learners' lack of social contact, boredom, fatigue, and lack of motivation is through the course's personalised design. The incorporation of gamification into an online course appears to boost both the completion rate and user engagement rates. Therefore, the current research examined the link between MOOC appeal and the user-friendly design of course material. Consequently, the following first theory is put forth:

H2: User Friendly Design of course content (UFD) is positively associated with the Attractiveness of MOOCs (AOM).

Studies on the MOOC retention rate have found a number of variables that can be examined from various angles. From the participants' point of view, time constraints, lack of self-motivation, lack of engagement, and monotonous course delivery are the main issues that can be resolved by designing MOOCs as individualised learning experiences (Rodrguez et al., 2022). The suggested "gamification approach" is founded on how students engage with the material. When creating the courses, the instructors consider elements like student involvement, readiness, and assessments. The instructors' main challenges when creating MOOCs include making sure that participants are actively participating, preparing the course materials, and developing assessment methods in addition to video developments. 2020 (Sari et al.). As a result, the current research examined the connection between MOOC attraction and Participation Interaction and Guidance from Instructors (PIIG). As a result, the following third theory is proposed:

H3: "Participation Interaction and Guidance from Instructors" (PIIG) is positively associated with the Attractiveness of MOOCs (AOM).

In nations with a comparatively advanced technological infrastructure, MOOCs are relatively appealing (Kaplan et al., 2016). The high speed of the internet is another element that influences learners' acceptance of MOOCs, according to Ayub et al. (2017)'s investigation into this topic using Kirkpatrick's model. The relationship between internet speed (IS) and the popularity of MOOCs was the focus of the current research. As a result, the following fourth theory is proposed:

H4: Internet Speed (IS) is positively associated with the Attractiveness of MOOCs (AOM).

One aspect of a MOOC's success rate is its long-term use or participants' desire to keep taking the course (Wu et al.,2017).Additionally, even if a participant does not complete the assessment procedure, the amount of course content that participant views is valuable to the MOOC designers. The current research investigated the connection between MOOC appeal and course level (CL). As a result, the following fifth theory is proposed:

H5: Course Level (CL) is positively associated with the Attractiveness of MOOCs (AOM).

The intention to continue the chosen MOOCs depends on a number of key variables, including the course schedule, participant workload, video quality, instructor charisma, subject-matter expertise, and the evaluation process. When it comes to the apparent difficulty and dropout rate of MOOCs, there are two opposing viewpoints. The first school of thought claims that the perceived difficulty of course material increases the drop rate, while the second school of thought disputes that this is the case (Eriksson et al., 2017; Du, B., 2022). The current research investigated how perceived difficulty (PD) and the allure of MOOCs related to one another. As a result, the following sixth theory is proposed:

H6: Perceived Difficulty (PD) is positively associated with the Attractiveness of MOOCs (AOM).

4. Data Analysis

A descriptive study was carried out to investigate the demographic makeup of the respondents. There is no statistically significant difference between income levels in students' intrinsic motivation to enrol in online learning. A partial least squares (PLS) regression was used to validate variable measurements and put theories to the test. The use of PLS is justified by the lax sample size and residual distribution criteria (Kim, Oh, Shin, and Chae, 2009; Chin, 1998). A survey of Kerala B-school students who had used MOOC platforms yielded 550 correctly answered questions. Table 1 shows the demographic breakdown of the interviewees.PSL was chosen as the best fit for the proposed model for the attraction of MOOCs due to its prediction-oriented, prediction accuracy, applicability to a small sample size, and resilience to multivariate analysisMikalef, P., and Pateli et al. (2017).PLS can also be used to examine a model with shaky theoretical underpinnings or weak relationships between factors (Benlian et al., 2009).

Table 1 Demographic Profile of the Respondents

Variables	Frequencies	%
Age		
Less than 22	nil	-
22-25	450	82
26-29	100	18
Gender		
Male	200	44
Female	350	64
Programme		
MBA	450	82
Ph.D	100	18

The reliability of the constructs is tested through Cronbach's alpha (>.70) and composite reliability (>.70). The proportion of variation in each factor explained

Journal of Asia Entrepreneurship and Sustainability
www.asiaentrepreneurshipjournal.com

by all the other constructs is tested through communality. The values are shown in Table.2 which explains the variations for factors under each construct..

Table 2: Reliability and Validity Tests

Constructs	Cronbach α	Composite Reliability	Communality
SDLE	0.912	0.921	0.712
UFD	0.856	0.867	0.657
PIIG	0.895	0.956	0.602
CL	0.751	0.812	0.599
IS	0.778	0.896	0.565
PD	0.712	0.796	0.651
AOM	0.891	0.902	0.712

The item's discriminant validity is checked by seeing how well the different constructs fit together. Details are furnished in table 3. It provided evidence of discriminant validity.

Table 3: Discriminant validity through Inter-correlations

Constructs	AVE	SDLE	UFD	PIIG	CC	IS	PD	AOM
SDLE	0.678	0.632						
UFD	0.712	0.587	0.665					
PIIG	0.673	0.567	0.672	0.678				

CL	0.667	0.654	0.543	0.456	0.564			
IS	0.713	0.745	0.654	0.634	0.563	0.435		
PD	0.601	0.645	0.589	0.547	0.589	0.589	0.578	
AOM	0.661	0.734	0.602	0.439	0.602	0.602	0.496	0.802

5. Results and Discussion

5.1 Testing of the Research Model

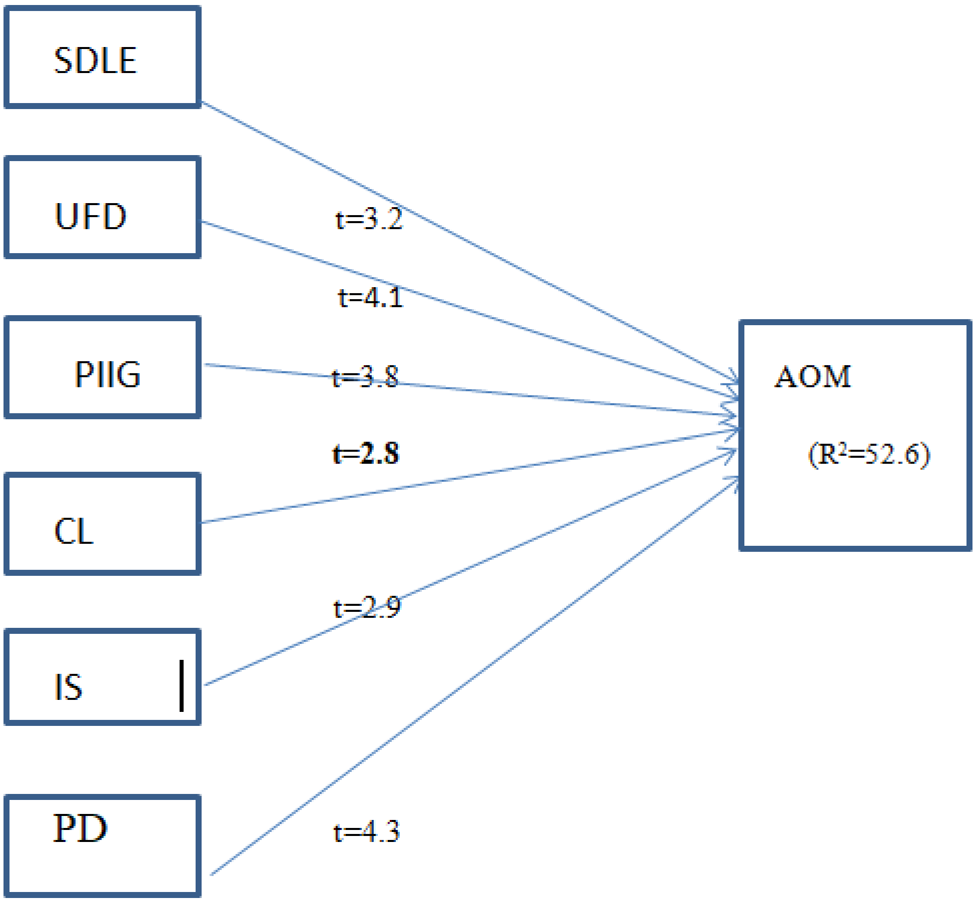
This study delves into the perspectives of B-school students in India who used MOOCs. The Krickpatrick's evaluation model, used in the study, which is based on four levels of evaluation viz, reactions,learning,behavior, and outcomes, is empirically tested to explore the factors of attractiveness of MOOCs. The current research interrogate the major factors like, Self-Directed Learning Environment (SDLE), User-Friendly Design of Course Content (UFD), Participation Interaction and Guidance from Teachers (PIIG), Internet Speed (IS), Course Level (CL), and Perceived Difficulty. Out of six, four hypotheses, which establishes the link with factors of attractiveness of MOOCs, postulated are supported in the study. The learning analytics results show that Self Directed Learning Environments (SDLE) and User-Friendly Design (UFD) have a substantial relationship with the Attractiveness of MOOCs (AOM), whereas Participation, Interaction, and Instructor Guidance (PIIG) and Internet speed (IS) have a moderate relationship. The dependent variable AOM is significantly impacted by each independent variable, including SDLE,UFD,PIIG, and SI. However prdectors like perceived diificulty (PD) and course level (CL) A total of 52.6% of the variance in the

Journal of Asia Entrepreneurship and Sustainability
www.asiaentrepreneurshipjournal.com

dependent variables can be accounted for by the independent variables. This suggests that the user-friendly course design, participation interaction, and guidance from instructors, as well as the fast internet and self-directed learning environment, have a favourable influence on a student's intrinsic motivation to engage in MOOCs. The findings related to all hypotheses are consistent with other empirical studies in the selected domain (Hone et al., 2016, Ayub et al.,2017. Alario-Hoyos et al., 2017, Goh et al., 2018).

The statistical significance of partial least squares results, such as path coefficients, Cronbach's alpha, and R2 values, is tested using bootstrapping, a non-parametric technique. The results of a PLS bootstrapping algorithm based on t-statistics of the path coefficient on 550 samples show that each factor has a statistically significant relationship with the dependent variable AOM. Based on the t-value threshold recommended by Hassanein (2007) and Gefen, 1.96 was chosen as the cut-off value (2005). Figure 2 depicts the hypothesis testing results, which are summarised in Table 4.

Fig.2 Results of the Model.



A PLS bootstrapping algorithm with 550 samples reveals statistically significant associations between each factor and the dependent variable (AOM) with t values significant at $t > 1.96$ at P 0.05. The R square number of the dependent variable reveals that all of the independent variables account for more than 52% of the total variance.

5.2 Testing of Hypotheses

The hypotheses H1,H2,H3, and H5 are supported, indicating that the most relevant factors are SDLE, UFD, PIIG, and IS. While hypotheses H4 and H6 are not supported, this suggests that course levels and perceived difficulty are the least important factors.

Table 4 Hypotheses Testing

Hypotheses	Pathes	Coefficient	SD	t-statistics	Results
H1	SDLE → AOM	0.278	0.032	4.298**	Supported
H2	UFD → AOM	0.348	0.036	5.71**	Supported
H3	PIIG → AOM	0.079	0.042	6.03**	Supported
H4	CL → AOM	0.005	0.039	0.103	Not Supported
H5	IS → AOM	0.352	0.041	2.19*	Supported
H6	PD → AOM	0.006	0.034	0.112	Not Supported
* p < 0.05, ** p < 0.01.					

5.3 Cross loading in PLS

Cross loading values in PLS has demonstrated a higher loading on parent construct for a particular item when compared to other constructs.

Table Cross Loading in PLS

RossiSmith Academic Publications, Oxford/London UK, www.publicationsales.com

Constructs	Items	SDLE	UFD	PIIG	CC	IS	PD	AOM
SDLE	SDLE1	0.7190	0.4970	0.5399	0.5660	0.3890	0.3230	0.3790
	SDLE2	0.7960	0.4740	0.2984	0.5430	0.2990	0.4145	0.3457
	SDLE3	0.7670	0.4510	0.3852	0.5200	0.3790	0.3000	0.3407
	SDLE4	0.8460	0.4280	0.2531	0.4970	0.3457	0.2000	0.3168
	SDLE5	0.7300	0.4050	0.1757	0.4740	0.3407	0.4000	0.2976
UFD	UFD1	0.5399	0.8150	0.3664	0.4380	0.4100	0.5500	0.4660
	UFD2	0.4984	0.7190	0.4569	0.4100	0.3820	0.5220	0.4380
	UFD3	0.3569	0.7960	0.5399	0.3820	0.3540	0.4940	0.4100

	UFD4	0.2821	0.7670	0.2984	0.3540	0.3260	0.4660	0.3830
	UFD5	0.1906	0.8460	0.3852	0.3260	0.2980	0.4380	0.3840
PIIG	PIIG1	0.4984	0.4569	0.7190	0.4660	0.4800	0.5900	0.3850
	PIIG2	0.3569	0.5399	0.7960	0.4380	0.3700	0.4800	0.3860
	PIIG3	0.2154	0.3569	0.7670	0.4100	0.3600	0.3700	0.3870
	PIIG4	0.0739	0.2154	0.8460	0.3820	0.2833	0.2600	0.3880
CL	CL1	0.5399	0.0739	0.3970	0.8780	0.4569	0.3200	0.5500
	CL2	0.5399	0.4940	0.3729	0.8690	0.5980	0.3200	0.5220
	CL3	0.5952	0.4660	0.3488	0.8600	0.5890	0.3200	0.4940

	CL4	0.6137	0.4380	0.3247	0.8510	0.5110	0.3200	0.4660
	CL5	0.6413	0.4100	0.3006	0.8420	0.6367	0.3200	0.4380
IS	IS1	0.4530	0.5399	0.5283	0.5890	0.8150	0.3200	0.6341
	IS2	0.5421	0.5870	0.5754	0.5110	0.7190	0.3200	0.6812
	IS3	0.4210	0.6341	0.6225	0.4330	0.7960	0.3200	0.5283
	IS4	0.4130	0.6812	0.6696	0.3550	0.7670	0.3200	0.5754
PD	PD1	0.3970	0.5500	0.4940	0.4380	0.8460	0.7190	0.4380
	PD2	0.3729	0.5220	0.4660	0.4100	0.4940	0.7960	0.4100
	PD3	0.3488	0.4940	0.4380	0.3820	0.4660	0.7670	0.3820

	PD4	0.3247	0.4660	0.4100	0.3540	0.4380	0.8460	0.3540
AOM	AOM1	0.3006	0.4380	0.5220	0.4100	0.4100	0.5110	0.7960
	AOM2	0.2765	0.4100	0.4940	0.3820	0.4660	0.5310	0.7670
	AOM3	0.2523	0.3820	0.4660	0.3540	0.4380	0.5510	0.8460
	AOM4	0.2282	0.3540	0.4380	0.3260	0.4100	0.5710	0.8190

6. Implications and suggestions

Technological advancements also allow for the testing of new educational theories or approaches in the MOOC context. Andersen et al. (2014), for example, used Salmon's (2000) five-stage model to facilitate students' online collaboration. This addresses institutional motivations for MOOC provision, as MOOCs have been viewed as a testing ground for educational innovations (e.g., Urrutia et al., 2015; Wong et al., 2015).

This study focused on Krickpatrick's evaluation model, which is based on four levels of evaluation: reactions, learning, behavior, and outcome. It was made possible by the adoption and validation of the key constructs identified by Goh et al. (2018) in the context of the effectiveness of MOOCs, as well as the addition of two new constructs: course level (CL) and perceived difficulty (PD). Factors such as the self-directed learning environment (SDLE), user-friendly course content design (UFD), participation interaction and guidance from instructors (PIIG), and internet speed (IS) are important, and MOOCs' allure was not enhanced by course level (CL) or perceived difficulty (PD). The study adds to the literature by highlighting specific implications and making suggestions for course design and improvement.

The possibility of top business schools offering MOOCs at a reduced cost is critical to the future of the education sector in general and business school education in particular. However, there are inherent disadvantages to online education, such as illegal file sharing, licencing violations, illegal downloading, and other issues. To increase their attractiveness, online course providers should use tactics such as user interaction, making a supporting staff member or a partner available to the key teachers, modifying the pace of the class for each course participant using learning diagnostics, and awarding points and badges to boost intrinsic motivation. Additionally, "experts as online teachers" play an important role in the success of an online course.

7. **Conclusion**

The increasing popularity and need for e-learning or web-based learning, as well as the increase in the number of courses and platforms offering e-learning and MOOCs, have necessitated an understanding of the factors that contribute to MOOCs' attractiveness. The current study sought to provide an answer to the question, "What can online distance learning providers do to make MOOCs more appealing to a wider audience and transform them into a platform that can genuinely educate the masses?" This article investigates the factors that contribute to selecting a specific MOOC as well as a few additional MOOC characteristics, namely perceived utility and course levels, using Krickpatrick's model. As the study result suggests, MOOC providers may consider the factors of user-friendly course content design (UFD) and self-directed learning environment (SDLE), which are highly correlated and affect MOOC attractiveness, when creating the structure of MOOCs in order to boost students' intrinsic motivation. Other factors that may have a greater impact on how appealing MOOCs are are participation, interaction, instructor guidance, and internet speed (IS).

This paper presents factors from the relevant literature that lead to effective MOOC teaching and their presence in MOOCs in practice. It contributes to our understanding of the differences between teaching traditional online courses and MOOCs, as well as the extent to which the methods and approaches proposed in the literature for teaching MOOCs are applicable.

The current findings do not impose a specific teaching strategy that must be used in order to be effective. The purpose of this paper is to draw attention to the MOOC context in various areas, in order to address practical issues arising from students' very large and diverse backgrounds, which impede effective teaching. MOOCs can now attract students from all over the world thanks to advances in information and communication technology and openness. However, without new approaches to teacher-student interactions, the concept of openness and the desired teaching and learning outcomes may not be realized (Anderson et al., 2002). There is no short-cut to effective MOOC teaching. It is an adventure of discovery, experimentation, and reflection on various teaching strategies, course design, and educational technology (Montgomery et al., 2015). It is worth noting that there is no evidence that the various features, or lack thereof, in course delivery result in differences in teaching and learning effectiveness (Glance et al., 2013). Given the exploratory nature of this study, more research is needed, particularly on the proper implementation of these features in teaching, their effectiveness, and methods for measuring such effectiveness. The researchers should look for other factors to improve the appeal of MOOCs.

References

Andersen, R. K., Garp, K., Nellemann, K., Nielsen, M. F., & Ørngreen, R. (2014, October). A framework for the analysis of collaborative and interactive elements in MOOCs. In The European Conference on e-Learning, Academic Conferences International Limited (p. 19).
Anderson, L. E., & Carta-Falsa, J. (2002). Factors that make faculty and student relationships effective. College Teaching, 50(4), 134-138.
Alario-Hoyos, C., Estévez-Ayres, I., Pérez-Sanagustín, M., Kloos, C. D., & Fernández-Panadero, C. (2017). Understanding learners' motivation and learning strategies in MOOCs. The International Review of Research in Open and Distributed Learning, 18(3).
Ayub, E., Wei, G. W., & Yue, W. S. (2017, January). Exploring factors affecting learners' acceptance of MOOCs based on Kirkpatrick's model. In Proceedings of the 8th International Conference on E-Education, E-Business, E-Management and E-Learning (pp. 34-39).
Benlian, A., Hess, T., & Buxmann, P. (2009). Drivers of SaaS-adoption–an empirical study of different application types. Business & Information Systems Engineering, 1, 357-369.
Chen, Y. H., & Chen, P. J. (2015). MOOC study group: Facilitation strategies, influential factors, and student perceived gains. Computers & Education, 86, 55-70.
De Notaris, D., Canazza, S., Mariconda, C., & Paulon, C. (2021). How to play a MOOC: Practices and simulation. Entertainment Computing, 37, 100395.
Du, B. (2022). Research on the factors influencing the learner satisfaction of MOOCs. Education and Information Technologies, 1-21.
Eriksson, T., Adawi, T., & Stöhr, C. (2017). "Time is the bottleneck": a qualitative study exploring why learners drop out of MOOCs. Journal of Computing in Higher Education, 29(1), 133-146.
Glance, D. G., Forsey, M., & Riley, M. (2013). The pedagogical foundations of massive open online courses. First monday

Goh, W. W., Wong, S. Y., & Ayub, E. (2018). The effectiveness of MOOC among learners based on Kirkpatrick's model. In Redesigning Learning for Greater Social Impact: Taylor's 9th Teaching and Learning Conference 2016 Proceedings (pp. 313-323). Springer Singapore.
Hone, K. S., & El Said, G. R. (2016). Exploring the factors affecting MOOC retention: A survey study. Computers & Education, 98, 157-168.
Jena, P. K. (2020). Impact of pandemic COVID-19 on education in India. International journal of current research (IJCR), 12.
Kaplan, A. M., & Haenlein, M. (2016). Higher education and the digital revolution: About MOOCs, SPOCs, social media, and the Cookie Monster. Business horizons, 59(4), 441-450.
Lee, J. W., & Lee, H. (2016). Human capital in the long run. Journal of development economics, 122, 147-169.
Long, L. K., DuBois, C. Z., & Faley, R. H. (2008). Online training: The value of capturing trainee reactions. Journal of Workplace Learning, 20(1), 21-37.
Lin, S., & Overbaugh, R. C. (2013). Autonomy of participation and ICT literacy in a self-directed learning environment (SDLE). Quality & Quantity, 47, 97-109.
Montgomery, A. P., Hayward, D. V., Dunn, W., Carbonaro, M., & Amrhein, C. G. (2015). Blending for student engagement: Lessons learned for MOOCs and beyond. Australasian Journal of Educational Technology, 31(6).
Mikalef, P., & Pateli, A. (2017). Information technology-enabled dynamic capabilities and their indirect effect on competitive performance: Findings from PLS-SEM and fsQCA. Journal of Business Research, 70, 1-16.
Pham, M., Singh, K., & Jahnke, I. (2021). Socio-technical-pedagogical usability of online courses for older adult learners. Interactive Learning Environments, 1-17.
Rekha, I. S., Shetty, J., & Basri, S. (2022). Students' continuance intention to use MOOCs: empirical evidence from India. Education and Information Technologies, 1-22.
Rodríguez, I., Puig, A., & Rodríguez, À. (2022). Towards Adaptive Gamification: A Method Using Dynamic Player Profile and a Case Study. Applied Sciences, 12(1), 486. MDPI AG. Retrieved from http://dx.doi.org/10.3390/app12010486
Salmon, G. (2012). E-moderating: The key to online teaching and learning. Routledge.

Salwan, P. (2021). Economic & Social Impacts of Online Learning in Higher Education: Building Resilience. Indian Journal of Industrial Relations, 56(3).
Sari, A. R., Bonk, C. J., & Zhu, M. (2020). MOOC instructor designs and challenges: what can be learned from existing MOOCs in Indonesia and Malaysia?. Asia Pacific Education Review, 21, 143-166.
Tham, C. M., & Werner, J. M. (2005). Designing and evaluating e-learning in higher education: A review and recommendations. Journal of leadership & organizational studies, 11(2), 15-25.
Urrutia, M. L., White, S., & White, S. (2016). MOOCs in Higher Education magazines: a content analysis of internal stakeholder perspectives. In Computer Supported Education: 7th International Conference, CSEDU 2015, Lisbon, Portugal, May 23-25, 2015, Revised Selected Papers 7 (pp. 395-405). Springer International Publishing.
Wong, B. T., Li, K. C., & Lam, H. (2015). Motivations and deterrents to MOOC offerings. Studies and Practices for Advancement in Open and Distance Education, 16.
Wu, B., & Chen, X. (2017). Continuance intention to use MOOCs: Integrating the technology acceptance model (TAM) and task technology fit (TTF) model. Computers in human behavior, 67, 221-232.
Zhu, M., Bonk, C. J., & Doo, M. Y. (2020). Self-directed learning in MOOCs: Exploring the relationships among motivation, self-monitoring, and self-management. Educational Technology Research and Development, 68, 2073-2093.

Journal of Asia Entrepreneurship and Sustainability
www.asiaentrepreneurshipjournal.com

Artificial Intelligence (AI) in Promoting Sustainable Entrepreneurship: A Bibliometric Analysis

Siby M Yohannan
Rajagiri Centre for Business Studies, Autonomous Institute affiliated with Mahatma Gandhi University, Kochi, Kerala
fpm22m0010@rajagiri.edu

Latha K (corresponding)
Mahatma Gandhi University, Kottayam, Kerala, India
lathabrijesh2@gmail.com

Abstract

This study aimed to assess the intellectual structure of the knowledge base on artificial intelligence (AI) in promoting sustainable entrepreneurship by conducting a bibliometric analysis. The study analysed 1023 publications from various sources and authors, between 2013 and 2023. The analysis was carried out using statistical software, namely the VOS viewer, to construct and visualise a structure map of

source coupling networks of researchers and co-authorship, which aided in identifying the main themes and concepts present in the publications. The results of the analysis revealed that the direction of the study related to artificial intelligence, sustainability, and entrepreneurship is gaining momentum, and this article also gives insight into the countries that focus on the research in this area and the major journals that promote research in this domain. This study also helps researchers understand the collaboration of authors in this area as well as the authors whose articles are most cited in this area of research. The bibliometric analysis of 1023 research documents from the Scopus database showed that the topics of artificial intelligence, entrepreneurship, and sustainability gained attention in 2013 and have seen a significant increase in publications and citations since then, with a sharp increase in 2022. The analysis also found a growing concentration of studies in the Asia-Pacific region, particularly in developing countries such as India and China. The most commonly used keywords included smart city, sustainability, digital transformation, and COVID-19. Collaboration among researchers in the same geographic region was found to lead to a synergistic effect and better academic output.

1.Introduction

Entrepreneurship and innovation are gaining significant popularity in the ever-changing business environment of today. As the business landscape continues to evolve rapidly with technological advancements, evolving customer preferences, and intense global competition, it has become crucial for businesses to adapt and innovate to stay relevant. Entrepreneurship offers a range of benefits, such as creating jobs for both entrepreneurs and others, driving innovation through the introduction of new products, services, and business models, and stimulating economic growth by generating revenue and contributing to the overall economy. Entrepreneurs are also known for their flexibility, which allows them to quickly respond to changing market demands and adapt to evolving economic conditions. Finally, successful entrepreneurship can lead to wealth creation for entrepreneurs, investors, and employees, which can have a positive impact on the wider economy. The central concept of sustainable entrepreneurship revolves around the actions and strategies employed by entrepreneurs in their pursuit of economic growth and business development, without causing harm to the social and ecological systems within which they operate. In other words, sustainable entrepreneurship is characterised by business practises that balance economic prosperity with social responsibility and environmental sustainability. The ultimate goal of sustainable entrepreneurship is to create long-term value for all stakeholders, including investors, employees, customers, and the wider community (Schaefer et al., 2015). Research on sustainable entrepreneurship is necessary to investigate how entrepreneurial activities can be used as a means to sustain nature and ecosystems, while also delivering economic and non-economic benefits for investors, entrepreneurs, and communities.

With the availability of big data, advanced algorithms, and enhanced computing and storage power, the popularity of AI has surged. AI systems are now an integral part of digital systems and are increasingly influencing human decision-making in significant ways. Using AI in business can enable the industry to depend on faster, more cost-effective, and more accurate methods of marketing. By employing AI in marketing strategies, an entrepreneur can obtain better responses from their audience and gain a competitive advantage over other brands. Additionally, AI has the potential to revolutionise businesses with innovative ideas and provide solutions to complex tasks, leading to substantial growth in the industry. Incorporating AI into products presents a chance to enhance companies' environmental sustainability, which can lead to higher purchasing intentions among consumers. Furthermore, it can appeal to new consumer segments that may not be attracted to conventional approaches to environmental sustainability (Frank, 2021). The use of AI alone cannot solve all of the complex environmental sustainability problems that we face. In the past, new technologies have promised to solve issues quickly, but have ultimately proven to be unsustainable over time. As a result, it is essential that we acknowledge the limitations of this innovation, look for ways to bypass some of the constraints, and come up with innovative approaches to make the most of AI's potential. It is very important to assess the intellectual structure of the knowledge base on artificial intelligence (AI) in promoting sustainability and entrepreneurship, and this study primarily focuses on how research is carried out in this area and how researchers across the world collaborate to conduct research on this topic.

2. Literature review

Dwivedi et al. (2019) present the combined knowledge and expertise of numerous prominent experts to examine the various opportunities, challenges, and potential research directions associated with the rapid rise of artificial intelligence (AI) in various domains such as business and management, government, the public sector, science, and technology. By recognising the impact of societal and industrial factors on the development of AI, it provides valuable and timely insights into the future implications of this technology on both industry and society as a whole. Manavalan et al. (2018) conducted a thorough analysis of the different factors that impact sustainable supply chains and documented their findings. Using these insights, a framework was developed to evaluate the preparedness of supply chain organisations for the Fourth Industrial Revolution, taking into account various perspectives. The framework model is based on five key aspects of supply chain management: business, technology, sustainable development, collaboration, and management strategy. By providing criteria that can be assessed by companies, this study offers a means of determining their readiness to undertake the transformative changes associated with Industry 4.0. Ceschin et al. (2018) examine the development of "design for sustainability" (DFS) over time, using a quasi-chronological approach to categorise past design approaches into four levels of innovation: product, product-service system, spatio-social system, and socio-technical system. By doing so, the paper proposes an evolutionary framework that maps out the various DFS approaches that have been reviewed. This framework illustrates how the DFS field has progressed from a narrow technical and product-oriented focus to a broader understanding of sustainability as a sociotechnical challenge requiring large-scale system-level changes. Additionally, the framework highlights how the different DFS approaches contribute to specific sustainability aspects and demonstrates linkages, overlaps, and complementarities between these approaches. Kiel et al. (2017) provide a comprehensive and structured overview of the economic, ecological, and social benefits and challenges associated with the Industrial Internet of Things (IIoT). The research aims to contribute to the limited existing literature on the IIoT by analysing its implications based on the Triple Bottom Line (TBL) framework. They suggest that for sustainable industrial value creation, the IIoT must be considered in the context of three additional dimensions beyond the established TBL framework: technical integration, data and information, and public context. Hughes et al. (2019) build on the idea that blockchain technology has significant potential in numerous areas, by providing a detailed narrative on its various applications and future direction across key themes, while also acknowledging the barriers to its adoption. The study highlights that blockchain technology can play a vital role in contributing to several United Nations Sustainable Development Goals and has the ability to bring about significant transformation in established industries and practises. This study also highlights the vast potential for blockchain technology to bring about positive change, with an increasing number of use cases emerging where its unique attributes can be leveraged to great effect. Diez-Olivan et al. (2019) present a thorough overview of the latest advancements in data fusion and machine learning for industrial prognosis, with a focus on identifying research trends, areas of opportunity, and unexplored challenges. The study presents a structured categorization of feature extraction techniques and machine learning methods based on their intended purpose, which includes descriptive analysis to determine the root cause of failure, predictive analysis to forecast when an asset will fail, and prescriptive analysis to minimise the impact of failure on the industry. Upward et al. (2019) introduce a framework of strongly sustainable business model propositions and principles, which were derived from a comprehensive trans-disciplinary literature review. A comparative analysis was conducted between this framework and the profit-oriented ontology for business models proposed by Osterwalder et al. (2005) The researchers also present an ontology that facilitates the description of successful, strongly sustainable business models, addressing existing weaknesses and incorporating functionally necessary relationships. Zimmer et al. (2015) critically analysed and evaluated the scientific literature on sustainable supplier management (SSM), with a particular emphasis on formal models that facilitate decision-making in sustainable supplier selection, monitoring, and development. To accomplish this goal, the paper proposes a framework for SSM and conducts a comprehensive content analysis, which includes a criteria analysis. The primary findings of the analysis indicate a rapid increase in academic interest in this topic, the prevalence of formal models such as the analytical hierarchy process, the analytical network process, and fuzzy-based approaches, a focus on the final evaluation and selection process, and a limited investigation of social and quantitative metrics. Wong et al. (2019) explored the impact of various factors, including relative advantage, complexity, upper management support, cost, market dynamics, competitive pressure, and regulatory support, on the adoption of blockchain technology for operations and supply chain management by small and medium enterprises (SMEs) in Malaysia. They analysed how these factors influence the adoption of blockchain technology and its potential benefits for SMEs in enhancing their operations and supply chain management. They conducted empirical research on 194 small and medium-sized enterprises (SMEs) and used a nonlinear, non-compensatory PLS-ANN approach to rank and analyse the data. The study found that competitive pressure, complexity, cost, and relative advantage have a significant impact on the behavioural intentions of SMEs towards blockchain adoption for operations and supply chain management. On the other hand, market dynamics, regulatory support, and upper management support did not have a significant effect on the behavioural intentions of SMEs towards blockchain adoption. Elia, et al. (2020) present a proposed definition of a digital entrepreneurship ecosystem, emphasising both the digital output and digital environment perspectives. The authors adopt a collective intelligence approach to develop a descriptive framework that identifies the distinguishing characteristics of a digital entrepreneurship ecosystem. Specifically, the framework defines and discusses four dimensions related to digital actors (who), digital activities (what), digital motivations (why), and digital organisation (how). To illustrate the framework's application, the authors analyse nine real-world cases of companies and initiatives, which are viewed as digital entrepreneurship ecosystems along the four key dimensions presented. In the present digital age, the comprehension of big data analytics (BDA) and artificial intelligence (AI) has become a tacit resource, as their successful utilisation relies heavily on the skill set of the workforce, particularly programming and data analytics abilities. Bag et al. (2020) combine institutional theory and resource-based view theory to explain how automotive companies utilise tangible resources and workforce skills to drive technological advancement and promote sustainable manufacturing practices, including the development of circular economy capabilities. The results indicate that both coercive and mimetic pressures are positively correlated with tangible resources, while coercive and normative pressures are positively correlated with workforce skills. In turn, tangible resources and workforce skills are positively correlated with the adoption of BDA-AI. BDA-AI adoption is also positively correlated with sustainable manufacturing practices (SMP) and circular economy (CE) capabilities, with SMP having a positive influence on CE capabilities. Additionally, their study reveals that organisational flexibility and industry dynamism positively moderate the enabling effects of BDA-AI and CE capabilities. This information provides guidance for managers seeking to enhance BDA-AI adoption and improve SMP and CE capabilities. Dubey, et al. (2019) utilised the dynamic capabilities view of firms and contingency theory to create and evaluate a model that elucidated the impact of entrepreneurial orientation (EO) on the adoption of big data analytics powered by artificial intelligence (BDA-AI) and operational performance (OP). The outcomes of data analysis reveal that EO is a crucial organisational characteristic that enables the effective utilisation and exploration of BDA-AI capabilities to achieve superior OP. Moreover, the results provide empirical evidence that EO is strongly linked to higher-level capabilities, such as BDA-AI, and OP, especially in the presence of diverse environmental dynamics (ED).

Consequently, this paper seeks to provide a bibliometric approach to assess the intellectual structure of the knowledge base on artificial intelligence (AI) in promoting sustainability and entrepreneurship, identifying the state of the art in this field, and more precisely, identifying trends and other relevant indicators by surveying the articles published on Scopus with subsequent treatment using VOSviewer software. Based on a content analysis of recent publications, we identify gaps and opportunities in the intellectual structure of the knowledge base on artificial intelligence (AI) in promoting sustainability and entrepreneurship, and the following research questions are formulated.

(1) To analyse the patterns of research literature creation over time.

(2) To analyse the trend of publications over time in the specific area of study.

(3) To find out which authors have written the most and most important books in the field of study.

(4) To determine the most frequently used keywords in the area of study

(5) To determine the most impactful and significant publications in the field of study.

(6) To identify the patterns of cooperation among authors and countries

(7) Figure out how the role of AI in helping entrepreneurs reach their sustainability goals changes over time and what the intellectual structure of research in this area is.

3. Methodology

This study looks at research on the intersection of artificial intelligence, sustainability, and entrepreneurship using a method called bibliometric review. The bibliometric approach allows for an objective categorization of publications based on specific criteria. Additionally, the study employs VOS viewer software to visualise data through category maps. The data was collected in March 2023 from Scopus, one of the most significant bibliographic databases, specifically from the Scopus Core Collection database and its sub databases. The search query "Artificial Intelligence, Sustainability, and Entrepreneurship" resulted in a total of

1023 publications. The search string used in the VOS viewer is as follows: ALL (artificial AND intelligence AND sustainable AND entrepreneurship) A ND (LIMIT TO (PUBYEAR 2013 to 2023) .The study employed bibliometric analysis using bibliometric indicators as a mechanism to collect and interpret data. The study utilized simultaneous occurrence of publications by year, keyword trends, co-citation, bibliographic coupling, and co-author analysis between countries and institutions. The results identified the state of development and main trends in terms of influence, main journals, articles, topics, authors, institutions, and countries. Analysis and graphical representation can aid academics and professionals in better understanding the research conducted in the field of sustainability, specifically related to artificial intelligence and entrepreneurship, and map out the main trends in the area. The study also employed citation analysis by document, journal, and author, author co-occurrence of keywords, co-authorship, and bibliographic coupling, which were applied to institutions and countries The study employed bibliometric analysis using bibliometric indicators as a mechanism to collect and interpret data. The use of Boolean operators resulted in 1023 bibliographic materials, which were analyzed using similarity visualization software (VOSviewer) to graphically present potential results. The study utilized simultaneous occurrence of publications by year, keyword trends, co-citation, bibliographic coupling, and co-author analysis between countries and institutions. The results identified the state of development and main trends in terms of influence, main journals, articles, topics, authors, institutions, and countries. Analysis and graphical representation can aid academics and professionals in better understanding the research conducted in the field of sustainability, specifically related to artificial intelligence and entrepreneurship, and map out the main trends in the area. The study also employed citation analysis by document, journal, and author, author co-occurrence of keywords, coauthorship, and bibliographic coupling, which were applied to institutions and countries.

3.1 Publications by Year

According to the data, sustainability, artificial intelligence, and entrepreneurship were emerging topics in 2013, with a significant increase in both the total number of publications and citations. Regular publications on this topic began in 2017, with 18 articles published. Since then, the number of publications has grown

Journal of Asia Entrepreneurship and Sustainability
www.asiaentrepreneurshipjournal.com

substantially, as shown in Table 1. Between 2013 and 2014, only four articles were published per year, while seven were published in 2015, 10 in 2016, 18 in 2017, 37 in 2018, 67 in 2019, 131 in 2020, 240 in 2021, and 406 in 2022. Figure 1 provides an overview of the annual trends in publications on this topic, based on a sample of 1023 articles. In 2022, there was an exponential increase in the number of publications, marking the "take-off phase" and resulting in 406 published articles. These numbers are a clear indication of the growing interest in sustainability, artificial intelligence, and entrepreneurship.

4. Data Analysis

Year-wise publication and citations are mentioned in Table 1.

Table 1: Year-wise publication and citations

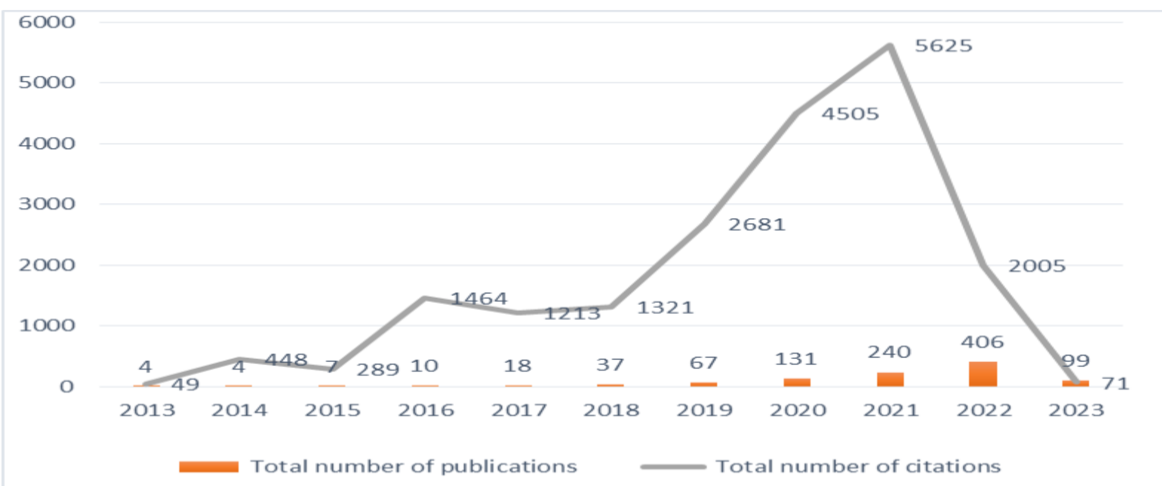
Year	**Total number of publications**	**Total number of citations**	**% of Publications**	**% of citation**

2013	4	49	0.39	0.25
2014	4	448	0.39	2.28
2015	7	289	0.68	1.47
2016	10	1464	0.98	7.44
2017	18	1213	1.76	6.17
2018	37	1321	3.62	6.72
2019	67	2681	6.55	13.63
2020	131	4505	12.81	22.90
2021	240	5625	23.46	28.60
2022	406	2005	39.69	10.19
2023	99	71	9.68	0.36

4.1 Publications by Year

According to the data, sustainability, artificial intelligence, and entrepreneurship were emerging topics in 2013, with a significant increase in both the total number of publications and citations. Regular publications on this topic began in 2017, with 18 articles published. Since then, the number of publications has grown substantially, as shown in Table 1. Between 2013 and 2014, only four articles were published per year, while seven were published in 2015, 10 in 2016, 18 in 2017, 37 in 2018, 67 in 2019, 131 in 2020, 240 in 2021, and 406 in 2022. Figure 1 provides an overview of the annual trends in publications on this topic, based on a sample of 1023 articles. In 2022, there was an exponential increase in the number of publications, marking the "take-off phase" and resulting in 406 published articles. These numbers are a clear indication of the growing interest in sustainability, artificial intelligence, and entrepreneurship.

Figure 1: Year-wise publications and citations



4.2 Publication by country

It is observed that Table 2 depicts the geographical distribution of authors. Among the 1023 articles, China tops the list with 144 articles, followed by Italy with 70 articles to its credit and the United Kingdom (UK) with 60 articles. India contributed 58 articles to bibliometric research between 2013 and 2023, ranking fourth, and South France published 18 articles, ranking fifteenth. Figure 2 shows the publication status of various countries on a map.

Figure 2: Publication Status of Countries

The top 15 countries in terms of publication are represented in table 2

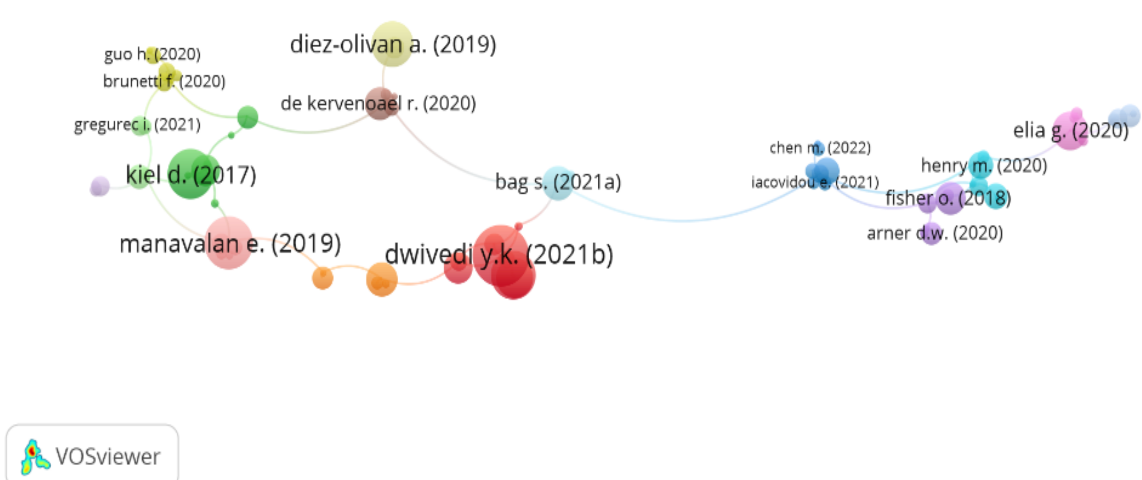
Table 2: The top 15 countries in terms of publication

Country name	Number of publications
China	144
Italy	70
United Kingdom	60
India	58
Poland	43
Spain	42
Germany	36
USA	34
Portugal	31
Romania	23
Malaysia	22
Australia	21
Russia	21
UAE	19
France	18

4.3 Citation analysis

Figure 3 provides an analysis of the citation structure in the relevant research area. It reveals the most cited articles in this field, with "Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agendas for research, practice, and policy" being the most referenced publication, with a total of 561 citations. The citation is produced when two documents refer to the same document. This approach is implemented for documents, journals, and authors and serves to demonstrate the relevance of a document for a thematic area. The Indian author Dwivedi Y.K. appears in the red cluster as one of the authors mentioned most often, in addition to other names, such as manavalan e., kiel d are also mentioned many number of times. By taking minimum number of citations as 6 of the 1023 articles 504 articles meet the threshold. The citation analysis is represented in Figure 3.

Figure 3: Citation Analysis



4.4 Publications by Journal

Upon analysing the journals in which the 1023 articles were published (as shown in Table 3), it was observed that out of the 378 journals, 70.89% published only one article on the subject, indicating that these journals are likely not from the areas of artificial intelligence, sustainability, and entrepreneurship. Furthermore, 13.22% (50 journals) published only two articles, 7.67% (29 journals) published three articles, 2.11% (8 journals) published four articles, and 6.084% (23 journals) published five or more articles, thereby being considered as journals on the role of artificial intelligence in sustainability and entrepreneurship (as listed in Table 3). Additionally, as illustrated in Table 4, this research topic is highly interdisciplinary and can be published in journals from various fields, each with a unique perspective.

Table 3: Production volume by journal

Production Volume by Journal	**Journals**	**% of 378**
1 Published Article	268	70.89947
2 Published Article	50	13.22751
3 Published Article	29	7.671958
4 Published Article	8	2.116402
5 or more Published Article	23	6.084656

Table 4: Top journals in terms of publications

Journal Name	**Number**	**% of 1023**
Sustainability (Switzerland)	272	26.588
Technological Forecasting and Social Change	57	5.572
Journal of Cleaner Production	38	3.715
Energies	27	2.639

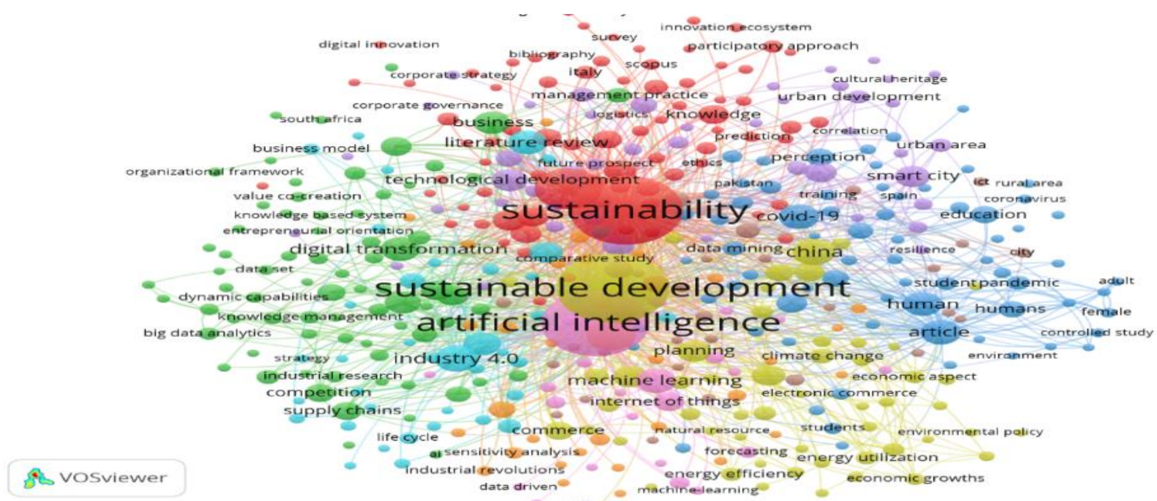
RossiSmith Academic Publications, Oxford/London UK, www.publicationsales.com

International Journal of Environmental Research and Public Health	20	1.955
Technology in Society	15	1.466
Journal of Business Research	14	1.369
Business Strategy and the Environment	13	1.271
International Journal of Information Management	8	0.782
Environmental Science and Pollution Research	7	0.684
Frontiers in Psychology	7	0.684
International Journal of Production Economics	7	0.684
International Journal of Technology Management	6	0.587

4.5 Keyword Analysis

The most frequently used keywords were identified and analysed to classify the 1023 articles that are part of the sample. From this analysis, the topics arising more often in the analysed area stand out. The map represented in Figure 4 groups the keywords into six clusters. The main keyword per cluster is smart city (purple cluster), sustainability (red cluster), sustainable development (yellow cluster), digital transformation (green cluster), artificial intelligence (pink cluster), digital transformation (green cluster), industry 4.0 (light blue cluster), and COVID-19 (blue cluster). From the 1023 articles, 6660 keywords were identified. The minimum number of occurrences of a keyword is taken as 5. 450 articles meet the threshold. For each of the 450 key words, the total strength of the co-occurrence links with other key words will be calculated. The keyword with the greatest total link strength (449) is selected. The largest set of connected items includes 449 items. Of these, 2509 words appeared only once.

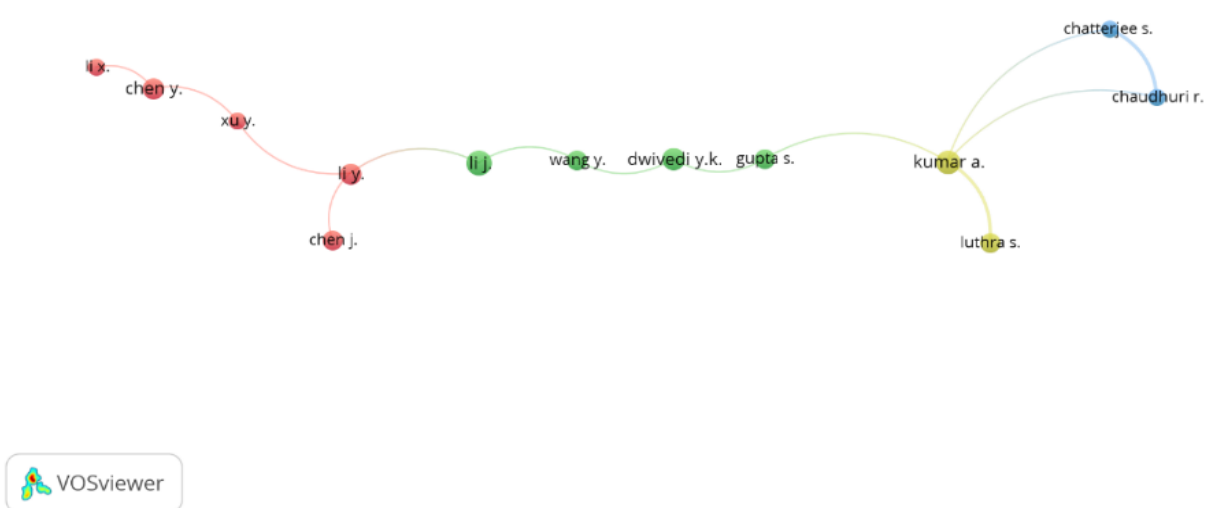
Figure 4: Keyword Analysis



4.6 Co-authorship Analysis

The minimum number of documents an author can have is 5. The minimum number of citations for an author is taken as 5. Of the 3301 authors, 19 authors met the threshold. For each of the 19 authors, the total strength of the co-authorship links with other authors is calculated. The authors with the greatest link strength are given a score of 19. Some of the items in the network are not connected to each other. The largest set of connected items consists of 13 items, as shown in the network visualisation in Figure 5.

Figure 5: Co-authorship analysis



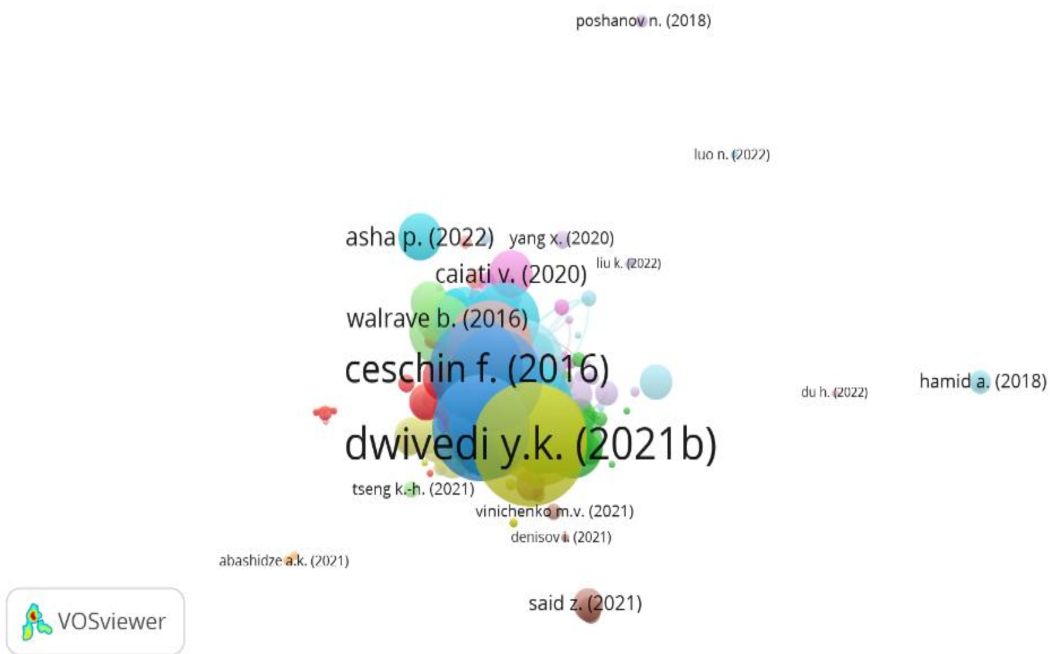
Journal of Asia Entrepreneurship and Sustainability
www.asiaentrepreneurshipjournal.com

4.7 Bibliographic coupling

The idea of bibliographic coupling was introduced by Kessler (1963), which refers to two articles being linked if they both cite a common third work in their reference lists. The degree of bibliographic coupling is strengthened when the number of references that two articles share in their reference lists increases. This method allows for the grouping of thematically similar documents into clusters. To illustrate this concept, VOS Viewer was used to cluster 1023 studies into distinct clusters. The resulting bibliographic coupling of the documents is shown in Figure 6; it is observed that some of the 1023 documents are not connected to each other, and the largest set of connected items consists of 976 items. The most cited article in the yellow cluster is by Dwivedi Y.K. et al. (2021), cited 561 times, in the dark blue cluster, the most cited article is by Ceschin F. (2016); in the light blue cluster, the most cited article is by Asha P. (2022); and in the pink cluster, the most cited article is by Caiati V. et al. (2020)

Journal of Asia Entrepreneurship and Sustainability
www.asiaentrepreneurshipjournal.com

Figure 6: Bibliographic coupling



5. Discussion

The authors have summarised our current study and outlined the results of the eight research questions that were introduced earlier in this section. The following list provides an account of the findings pertaining to each question: The first research question pertains to the patterns observed in the production of research literature. The results indicate that research on artificial intelligence, sustainability, and entrepreneurship has gained momentum and is likely to continue growing rapidly, particularly given the disruptive impact of recent technological developments. Significant contributions have been made in these fields over the last few years, with researchers increasingly publishing their work in reputable peer-reviewed journals since 2013. A country-wise analysis of scientific production reveals a different story, showing a growing concentration of studies in the Asia-Pacific region (APAC), which includes major developing countries like India and China.

For the second research question, the data shows that artificial intelligence and sustainable entrepreneurship were emerging topics in 2013 and saw a significant increase in publications and citations. Regular publications began in 2017, and since then, the number of publications has grown substantially, with a sharp increase in 2022. The growing number of publications in these fields shows that more and more people are interested in them.

To determine the authors who have produced the most work and have had the greatest impact in the field of study, the article with the highest number of citations is obtained, and it is observed that the research by Dwivedi Y.K. et al. (2021), which has been cited 561 times, falls into this category. The other most-cited articles are by Ceschin F. et al. (2016), Asha P. et al. (2022), and Caiati V. et al. (2020).

A study of 1023 articles was done to find out which keywords were used most often in the field of study. The most common keywords were found and put into six groups.The clusters represent the most important topics in the area that was studied. The most important keywords for each cluster are smart city, sustainability, sustainable development, digital transformation), artificial intelligence, industry 4.0, and COVID-19.

The most impactful and significant publications in the field of study is identified based on the citations. The article with the highest number of citations is identified as an article by Dwivedi Y.K. et al. (2021), The study brings together the collective insight from a number of leading expert contributors to highlight the significant opportunities, realistic assessment of impact, challenges, and potential research agenda posed by the rapid emergence of AI within a number of domains: business and management, government, the public sector, and science and technology.

An attempt is made to identify the patterns of cooperation among authors and countries. It highlights the absence of collaboration among countries and researchers in the fields of artificial intelligence, sustainability, and entrepreneurship. The author's analysis of the cooperation network reveals that researchers from the same geographic region tend to be more active in collaborative work, resulting in a synergistic effect and better academic output.

To determine the periodic progression of the role of AI in helping entrepreneurs achieve sustainability goals, the network map, is utilised. The resulting map revealed the emergence of several clusters, aiding in the identification of the intellectual structure of the research area.

6. Conclusion

This article presents a bibliometric analysis of the research related to the role of artificial intelligence in sustainability and entrepreneurship to determine the areas within which researchers are studying the role of artificial intelligence in sustainability and entrepreneurship, the trend in the number of publications from year to year, the most relevant journals for a literature review, and the most prolific and most cited authors who are concentrating their research in this area. This analysis offers a guide to those who are entering the field of the role of artificial intelligence in sustainability and entrepreneurship, providing information on which journals to consult and which authors are most eminent. The bibliometric analysis of 1023 research documents gathered from the Scopus database field shows that the topics of artificial intelligence, entrepreneurship, and sustainability started gaining attention in the year 2013 and have seen a significant increase in publications and citations since then. There has been a sharp increase in the number of publications in these fields in 2022. The analysis of scientific production by country shows a growing concentration of studies in the Asia-Pacific region (APAC), which includes major developing countries such as India and China. The most prominent keywords used are smart city, sustainability, sustainable development, digital transformation, artificial intelligence, industry 4.0, and COVID-19. This analysis has also found that researchers from the same geographic region tend to collaborate more, leading to a synergistic effect and better academic output.

7. Limitations and future research

Our study has certain limitations. Although we examined contributions for the whole study period, it may be helpful to perform the same type of analysis for each of the stages identified. This could provide a different perspective on the evolution of indicators and offer a different way to understand possible trends. It would be interesting to gain further insights into a broader range of industries and to perform more analyses from an international perspective. Finally, we think that our contribution provides a resourceful foundation on which to develop a systematic research review on this topic. This study does not include the major works indexed in another significant database (i.e., the Web of Science), which is the major limitation of the study. In the future, artificial intelligence based tools will be used for formulating strategic decisions, which is an interesting pathway for conducting research. Why China is focusing on this area is another important study to carry out, and why countries like Russia and France are contributing so little to this area of research is another topic that needs to be discussed.

References

Frank, P. (2021), "A proposal of personal competencies for sustainable consumption", *International Journal of Sustainability in Higher Education*, Vol. 22 No. 6, pp. 1225-1245. doi:https://doi.org/10.1108/IJSHE-01-2020-0027

Chowdhury, P., Paul, S. K., Kaisar, S., & Moktadir, M. A. (2021). COVID-19 pandemic related supply chain studies: A systematic review. Transportation research. Part E, *Logistics and transportation review*, 148, 102271. https://doi.org/10.1016/j.tre.2021.102271

Elia, G., Margherita, A., & Passiante, G. (2020). Digital entrepreneurship ecosystem: How digital technologies and collective intelligence are reshaping the entrepreneurial process. *Technological Forecasting and Social Change*, 150, 119791. https://doi.org/10.1016/j.techfore.2019.119791

Dwivedi, Y., Bag, S., Pretorius, J. H. C., & Gupta, S. (2021). Role of institutional pressures and resources in the adoption of big data analytics powered artificial intelligence, sustainable manufacturing practices and circular economy capabilities. *Technological Forecasting and Social Change*, 163, p. 120420. doi:https://doi.org/10.1016/j.techfore.2020.120420

Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., Williams, M. D. (2021). Artificial Intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 57, 101994. doi:https://doi.org/10.1016/j.ijinfomgt.2019.08.002

Dubey, R., Gunasekaran, A., Childe, S.J., Bryde, D.J., Giannakis, M., Foropon, C.R., Roubaud, D., & Hazen, B.T. (2020). Big data analytics and artificial intelligence pathway to operational performance under the effects of entrepreneurial orientation and environmental dynamism: A study of manufacturing organisations. International *Journal of Production Economics*, 226, 107599. doi:https://doi.org/10.1016/j.ijpe.2019.107599

Hughes, L., Dwivedi, Y., Misra, S. K., Rana, N., Raghavan, V., & Akella, V. (2019). Blockchain research, practice and policy: Applications, benefits, limitations, emerging

research themes and research agenda. *International Journal of Information Management*, 49, pp. 114-129. https://doi.org/10.1016/j.ijinfomgt.2019.02.005
Litvinenko, V. (2019). Digital Economy as a Factor in the Technological Development of the Mineral Sector. Natural Resources Research, 29, 1521-1541. https://doi.org/10.1007/s11053-019-09568-4
Wong, L.-W., Leong, L.-Y., Hew, J.-J., Tan, G. W.-H., & Ooi, K.-B. (2020). Time to seize the digital evolution: Adoption of blockchain in operations and supply chain management among Malaysian SMEs. *International Journal of Information Management*, 52(Complete). doi:https://doi.org/10.1016/j.ijinfomgt.2019.08.005
Diez-Olivan, A., Ser, J.D., Galar, D.P., & Sierra, B. (2019). Data fusion and machine learning for industrial prognosis: Trends and perspectives towards Industry 4.0. *Inf. Fusion, 50*, 92-111. doi:https://doi.org/10.1016/j.inffus.2018.10.005
Tavana, M., Yazdani, M., & Di Caprio, D. (2017). An application of an integrated ANP–QFD framework for sustainable supplier selection. *International Journal of Logistics Research and Applications, 20*, 254 - 275. doi: https://doi.org/10.1080/13675567.2016.1219702
Danvila-del-Valle, I., Estévez-Mendoza, C., & Lara, F. J. (2019). Human resources training: A bibliometric analysis. *Journal of Business Research*, 101, 627–636. doi:https://doi.org/10.1016/j.jbusres.2019.02.026
Ethirajan, M., & Kandasamy, J. (2020). Assessment of sustainability in supply chain management for industry 4.0 requirements. *AIP Conference Proceedings*. doi:https://doi.org/10.1063/5.0024872
Müller, D., Kiel, D., Müller, J. M., Arnold, C., & Voigt, K.-I. (2017). Sustainable industrial value creation: benefits and challenges of industry 4.0. *International Journal of Innovation Management* , 21(08), 1–34. doi:https://doi.org/10.1080/00207543.2019.1652777
Ceschin, F., & Gaziulusoy, I. (2016). Evolution of design for sustainability: From product design to design for system innovations and transitions. *Design Studies, 47*, 118–163. doi:https://doi.org/10.1016/j.destud.2016.09.002
Upward, A., & Jones, P. (2015). An Ontology for Strongly Sustainable Business Models. Organization & Environment, 29(1), 97-123 . doi:https://doi.org/10.1177/1086026615592933

Zimmer, K., Froehling, M., & Schultmann, F. (2015) Sustainable supplier management – a review of models supporting sustainable supplier selection, monitoring and development. International Journal of Production Research. 54. 1-31. doi:https://doi.org/10.1080/00207543.2015.1079340.

Kessler, M. M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, *14*(1), 10–25. doi:https://doi.org/10.1002/asi.5090140103

Journal of Asia Entrepreneurship and Sustainability
www.asiaentrepreneurshipjournal.com

Self-Efficacy: Is It Really Important for Adapting with Banking 5.0 Working environment? – A Case of the Indian Banking Sector

Isani Gazalabanu Abdul Gafar
International School of Management Excellence, (ISME), Research Centre, Bengaluru, Affiliated to the University of Mysore
gazalaisani@gmail.com

Irshad Nazeer
International School of Management Excellence (ISME), Research Centre, Bengaluru
nazeer.irshad@gmail.com

Abstract

The traditional banking sector had been a dream job sector for the job seekers in the pre-COVID era though the Digital Banking was becoming popular and even traditional banks were also adopting it gradually. This paper analysed the skill needs for the new Banking 5.0 environment enriched with Fin-Tech.

The study was conducted in two levels: Among retired employees and current employees. The need of the study is to analyse the adaption of employees with the dynamic environment where the traditional positions are already replaced with new positions that needs advance technical skills. The features of Smartphones along with Fin-tech attributes made the financial services quick, easy and less chances for mediation and fraudulence.

The advanced technology increased the productivity per employee very high and cyber security system monitor the applications based transactions. This research analyse how the employees cop up with this higher work stress and the effect of emotional intelligence in managing work stress.

The results shows that 23% of the employees emphasized the support of spouses in managing stress while the life commitment and survival thoughts motivated the higher performance. This study analyses the balancing skill of employees in maintaining a balance in work-life psychological wellbeing.

Introduction

Self –efficacy is an important attribute needed for every employee and self-learning or self self mastering is the first level of self-efficacy (Brown, Malou, & Schutte, 2005). The four factors that influence the self-efficacy are, emotional arousal, verbal persuasion, vicarious experience, and own experience (Lippke, 2020). Self-efficacy is an individual level attribute that enables an individual to adapt with new technologies. There is a drastic change in the functional, level in the post CoVID period, especially in adopting industry 4.0 standards. It is visible all levels including administration and human resource management. E-HRM is a digital twin of the process of human resource management in real time to digital form from the assessment of talent need to the performance appraisal and appreciation.

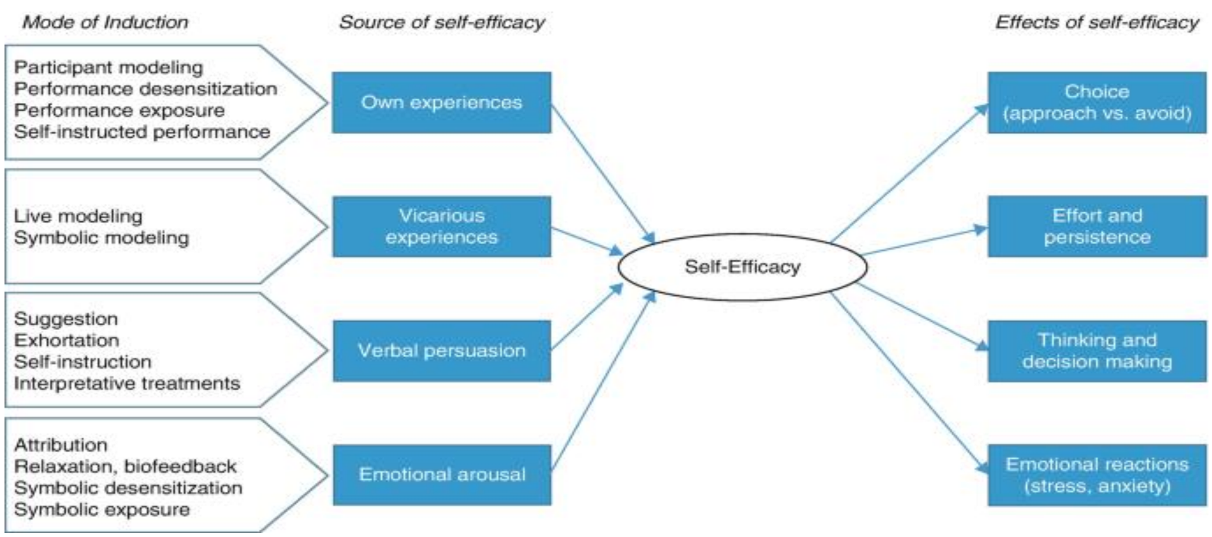
Banking sector is a technology driven manpower intense industry in which the modern banking is a shared platform in which the customers can use banking platforms for business and fund transfer without visiting any branches. The role bank employees deviated from fundamental banking services to the role of watchdog to ensure the security of funds of customers as the custodian and to promote value added services to enhance customer convenience and satisfaction along with increasing the bank profitability. This study analyses the role of self-efficacy in need of technology adaption of employees for an effective performance and use of EHRM in managing resources.

Journal of Asia Entrepreneurship and Sustainability
www.asiaentrepreneurshipjournal.com

Theoretical background

Self-efficacy theory

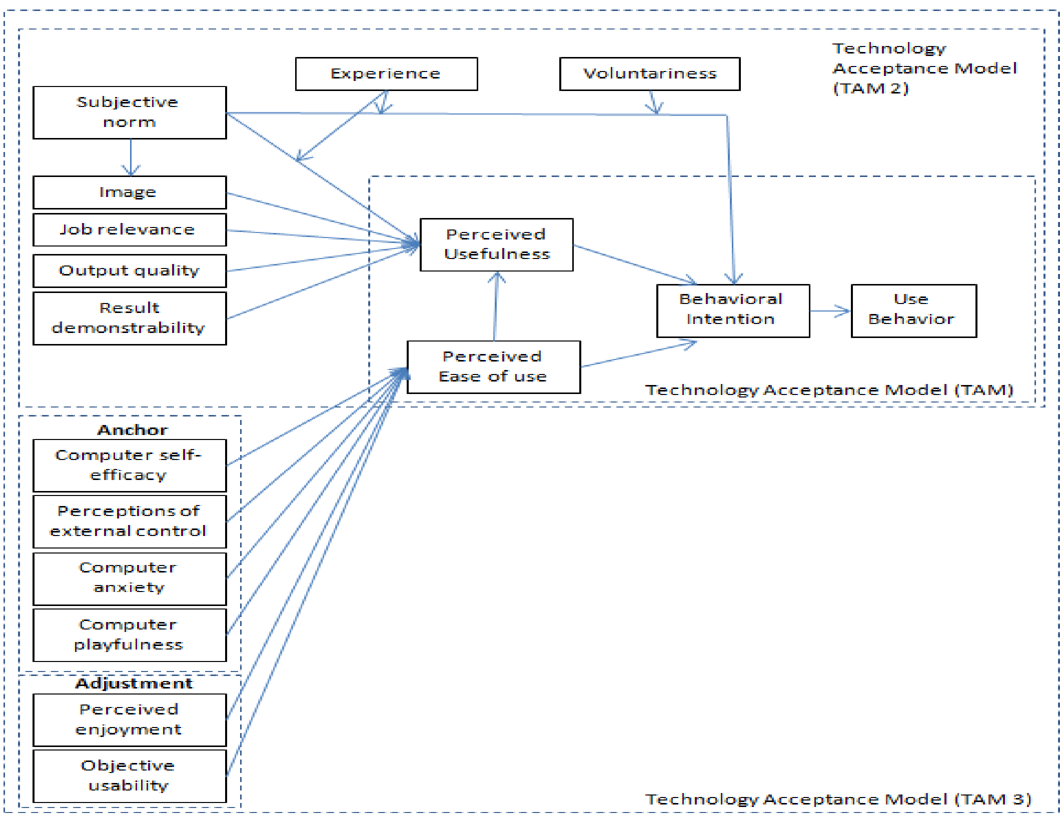
Fig 1: Self-efficacy (Lippke, 2020)



There are four decision making conditions, choice (do or not), efforts, persistence, perseverance (determination and involvement), though process and decision making (decision making) and emotional reactions. These attributes of individuals helps to adapt with new technologies. Choice is based on Maslow's need theory and it is a motivation to do the best choices. Persistence shows that determination and perseverance is the high level of persistence. Decision making is the outcome of efforts, perseverance, and need. The thinking process consists of knowledge, skill and other personal attributes to understand a critical situation and react to it to either get benefit or reduce loss. Emotional intelligence of a person helps to understand self to fit into a situation based on awareness, regulation and interaction with the environment. The four sources of self-efficacy are, own experience, vicarious experience, verbal persuasion, and emotional arousal. Self-Efficacy plays a significant role in adopting E- HRM (Ayyash, Herzallah, & Alkhateeb, 2021).

Technology Acceptance Model

Fig 3.1 : Technology Acceptance Model (TAM3) (Venkatesh & Bala, 2008)



(Davis & DeWitt, 2021) explained the need of technology as the central platform in managing resources keeping all stake holders in one platform. Industry 4.0 is the central point in the technology adaption in banks. Technology Acceptance Model (TAM3) integrates anchors and adjustments with the Technology Acceptance Model (TAM2) to include the human role in new platform. Internet of Service, and cloud computing enabled the banks to make the banking service more effective and productive.

The three inputs include subjective norm, technology inputs (Anchors) and adjustments (enjoyment and usability). The first component of anchor is self-efficacy that helps the employees to adapt technology. Selection, crite4rion for the adoption

In the TAM2 , the subjective norms are linked to behavioural intension in which the experience, attitude and importance of employees in the technology driven environment. The anchoring factors explain the self-efficacy of employees to adapt with computer technology and employee attitude on using technology.

Digital awareness is important in adoption of EHRM. It depends on employee potential to adapt with the technology oriented EHRM (Giri, Chatterjee, Bag, & Paul, 2019).

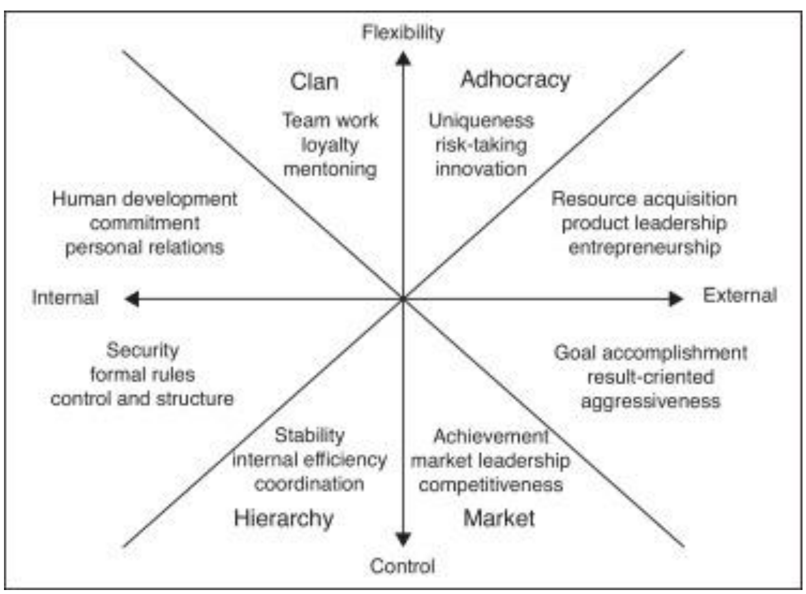
EHRM Model (Ruël & Bondarouk, 2004)

EHRM is a digital twin that makes an 'application' in mobile so that the human resource can be managed effectively. Most of the banks have multiple branches and the use of technology in customer involved banking services reduced the customer interaction up to a very great extent . In every bank , two digital systems are predominant, banking application and EHRM. The banking application helps the customers to do transactions, generate reports, apply for new services etc. (Milon, Alam, & & Pias, 2022).

Being a technology driven and labour intense sector, E-HRM has a high relevance in coordinating and controlling work in banks. EHRM has three levels: operational, relational and transformational. In operational eHRM, applying leave, pay role, maintaining personal data and get updated reports on available leave, applying leave , get salary updates etc. In most of the bank the salary is credited to the employee salary account . Relational EHRM includes, recruitment, training, performance management etc. The transformable eHRM include the strategies to improve the organizational performance and employee performance (Poreddy, 2018).

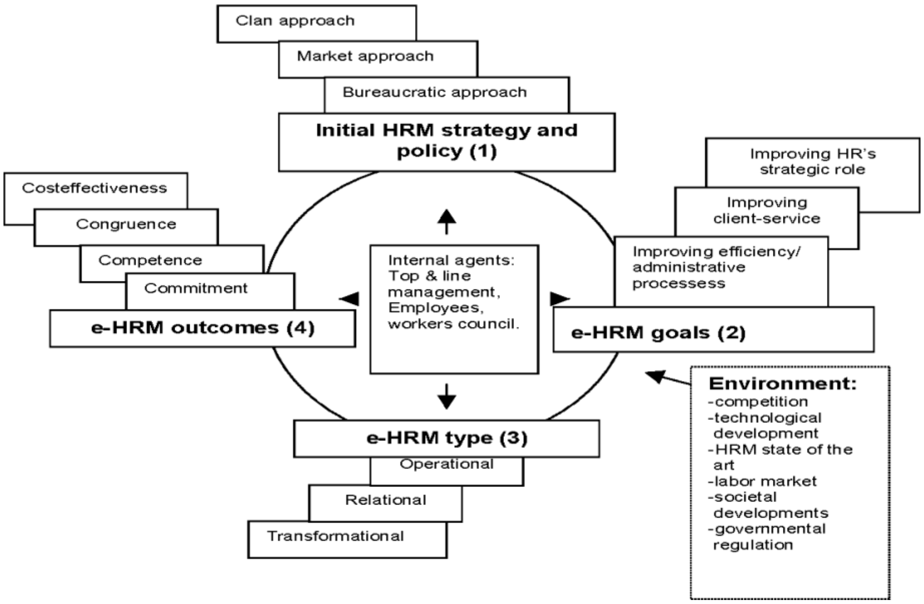
The EHRM Model explain the aims and expected outputs of EHRM accomplished through the internal agents called employees (Rathee & Bhuntel, 2022). If we consider the organization as an organism with definite boundaries, there exists two strategies: internal and external. Internal strategies are to maximise profit using optimum internal resources while external strategies are to use opportunities available in the environment or market. From the block diagram, it is explained the strategies (Clan approach, Market approach and Bureaucratic approach) to implement changes in the organization. The concept is derived from the organizational typology framework (Marín, Hernández, Valle, & Castillo, 2016). Horizontal axis represent internal and external strategies. In the vertical axis, have flexibility and control. The second quadrant of the diagram gives the role of individual employees in clan segment (left of Flexibility- Control axis) as teamwork, loyalty, mentoring, human resource training, commitment and personal relationships.

Fig:3.3 Organizational cultural typology framework. (N.Stock, L.McFadden, & R.Gowen, 2008)



Employee efficacy in using technology helps to understand customer needs and solve them. It increases customer loyalty and reduce employee stress. Every firm is looking for the efficient employes that the productivity will increase (AlHamad, et al., 2022). The organizational; culture is important in identfying the organizational need based on marke need, HR policies and strategies in HRM and organizational performance based on employee performance (Thathsara & Sutha, 2021). The EHRM will be useful to convey organizational goals, targets, internal policies, strategies to overcome market competition, safety and security of system, etc. to the employees while employees can share their innovative and unique ideas that benefit the organization (Shamout, Elayan, Rawashdeh, Kurdi, & Alshurideh, 2022)

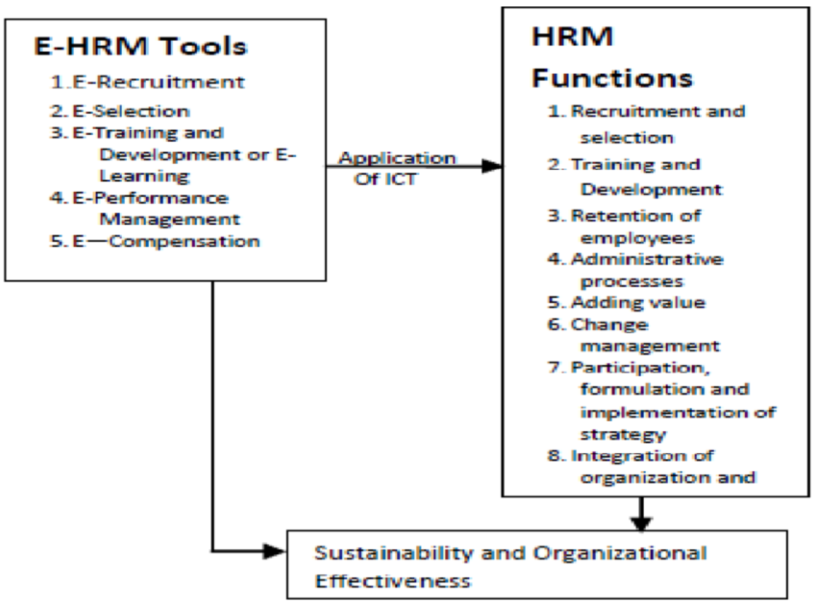
Fig 4: EHRM Model (Ruël & Bondarouk, 2004)



Functional differentiation is another criteria that integrate using software platform, the advantages of E HRM include cost effectiveness, congruence, competence, and commitment

EHRM in developing Organizational effectiveness and sustainability (Deshwal, 2015)

Fig 3.5 Impact of HRM on organizational effectiveness



The effect of EHRM in managing resources are gives in Fig 3.5. EHRM is a digital platform that integrates all Human Resource Practices to the Enterprise Resource Planning platform. This helps to integrate employee responsibilities, authorities, contributions, performance evaluation and remuneration in one platform that both employee and management are transparent in their roles. It also motivate to contribute more

Relevance of EHRM in Banking sector

Most of the banks are a networked system with a large number of branches, but the control and transactions are centralised. Though the software applications reduced the dependency of customers on branches, the need of large number of branches and employees reduced. Since the human resource is spread over a wide geographical area and the operational domain a wide subdomains, maintaining adequate employees is not easy. The EHRM helps to manage the adequate talents through training, recruitment, and promotions. EHRM is highly apt in banking sector. It useful in both branch level HRM and Corporate level.

Experience, emotional arousal, and verbal persuasion cause self-efficacy leading to a choice or action. EHRM is a technological solution to reduce delay in HRM practices and to involve employees to use it for getting HR information or do HR needs. It needs two skills: Technology adaption and information use. EHRM is a virtual platform in which the employees can update information through blogs, messages etc..

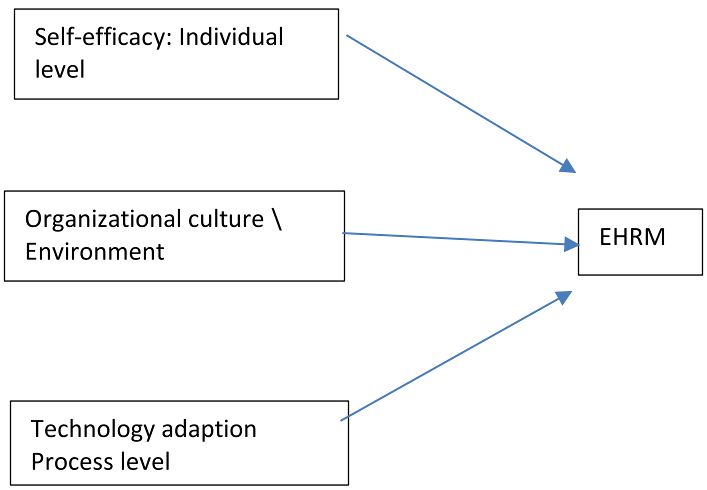
EHRM is useful in quick identification of talents, background verification of employees, online interviewing, on-boarding activities, generating appointment documents, employee benefit updates etc. The real-time integration of performance reports and evaluation process is another advantage. In the case of training, the banks can provide online trainings, certifications, tests for evaluating knowledge, motivating verses, etc. for employees.

RossiSmith Academic Publications, Oxford/London UK, www.publicationsales.com

In both aspects, employee competency and system effectiveness are important

Proposed model

The proposed model for this research can be explained as below



Hypothesis Development

Objective : 1 :-To analyse the current level of usage of E HRM in Banks

Technology-Organization-Environment Theory (TOE Theory) (DePietro, Wiarda, & Fleischer, 1990) explained the processes and procedures that assist firms in adopting and to assimilate technological innovation. It depends on, decisionmakers to manage the challenges, use apt technology matching to the characteristics and features of the firm, the areas where to speed-up the adoption of ICTs and enhance effective employee management. In the post COVID period, the significance of eHRM has increased and the firms, especially corporates who have multi-locational operation could manage operations effectively while small single unit firm who used work premises for work failed to integrate the operations in one platform though they managed with What's app or Email, yet full consolidation of information was not possible. This is the point , a single platform is essential.

Hence, the hypothesis,

$H_{o1:}$ There is no significant of EHRM on Employee usefulness

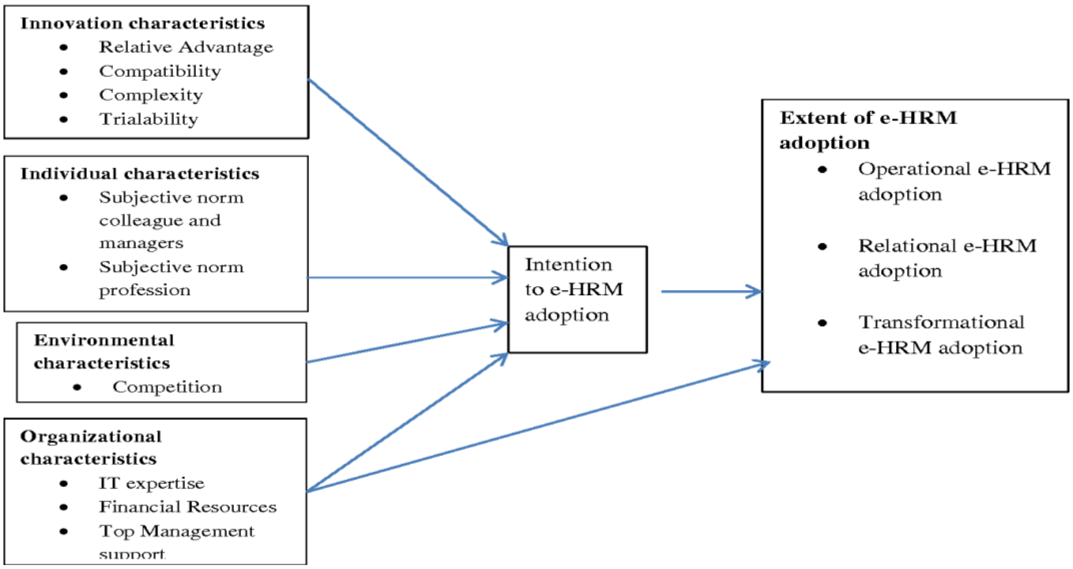
Objective 2: Effectiveness of EHRM in IT industries (Galhena, 2015)

The effectiveness of EHRM gives three outcomes: adoptions in operational, rational and transformational level. Innovation characteristics (relative advantage and compatibility), environmental characteristics (competition), organizational characteristics (top management support) explained the organizational e-HRM adoption intention. IT expertise is explained the extent of relational and transformational e-HRM adoption.

Hence, the Hypothesis

Journal of Asia Entrepreneurship and Sustainability
www.asiaentrepreneurshipjournal.com

$H_{o2:}$ There is no significant effect of EHRM on IT expertise and EHRM adoption



Research Methodology

Primary Data:

The study involves the collection of data directly by Respondents through structured questionnaire.

The sample size is calculated for the confidence level of ±5% for a maximum variance of 0.5

$N= Z^* (p^*q)/ (d^2) = 1.96^* (.5^*.5)/(.05^2) = 385$ (Israel, 2003)

To ensure the sampling adequacy for Multiple Linear Regression Model , the number of cases needed must be N> 5m+80 where m is the number of variables

(Green, 1991). In the case of factor model, the sample size may be 3 to 20 times of the number of independent variables (J, Mundfrom, Shaw, & Ke, 2005).

Considering all the above cases, the sample size is taken as 390. The response rate 56%   in which 8% responses were rejected due to incompleteness or redundant data.

The data was validated using test-retest method and found that there is consistency in response. The period of data collection is 2021-2022 from Bank employees. Combach alpha is above 0.721 for all questions.

Multiple linear regression model and Multi criteria decision model are used to analyse the level of using EHRM in Banks.

Population and Sample space

The population of this research is employees from different bank branches in Bangalore city. The distribution of respondents are given below

Level	Private	Scheduled	Public	Total
Executive	23	52	62	137
Low Management	53	45	51	149
Mid Management	23	19	35	77
Top Management	9	7	11	27
Total	108	123	159	390

Analysis and Interpretation

4.1 Regression Analysis

4.1.1 The regression analyse to determine the factors that influence the Employee usefulness of eHRM

$H_{04.1}$: There is no significant effect of EHRM on employee's usefulness

This analyse explains how EHRM helps to an employee useful performance system. It is taken for different positions. The use of EHRM at different levels varies with the information retrieved.

Motivating factors	Mean
Employee usefulness	2.20
Recruitment / Hiring Process	3.49
On-boarding Process	3.31
Learning and Development Programs	3.51
Performance Management System	2.45
Payroll Process	2.25
Leave & Attendance Management System	3.50

Employee Communication Process	2.56
Employee Database Management	3.20
Travel and Mobility Management	3.23
Employee Separation Process	2.23
Rewards and Recognition	2.69

Mean of the variables shows that, learning and development programs (3.51), leave and attendance Management system (3.50), Recruitment / Hiring Process (3.49), On-boarding Process (3.31) , Travel and mobility Management (3.23) and Employee Data Base Management (3.20) have higher mean.

Table 4.1.1 Correlation coefficients of Regression Analysis of Age on factors influencing purchase decision

Level	R	R^2	F value	Significance level	Hypothesis Accepted
Executive	0.586	0.339	2.871	.008	H_1
Low Management	0.533	0.284	8.867	.000	H_1
Mid Management	0.524	0.275	3.878	.000	H_1
Top Management	0.179	0.032	1.711	.103	H_0

The 'R' value is 0.586 for executive level and 33.9% variation (R^2) in response explained. The F value is statistically significant and it shows that the variation in response is consistent. Similarly, 28.4% and 27.5% variances were represented for the Low and mid-level management. Alternate hypothesis is accepted as the F is statistically significant.

The regression model for top management is not statistically significant and hence, null hypothesis accepted for top management

(Employee usefulness)$_{\text{Executive level}}$ = 2.12+ 0.30*Employee Communication Process -0.30

Travel and Mobility Management+ 0.36*Employee Separation Process +0.32*Rewards and Recognition + e

(Employee usefulness)$_{\text{low management}}$= 1.49 -0.13*On-boarding Process +0.17* Payroll Process+ 0.15* Leave & Attendance Management System+0.28*Employee Communication Process+0.36*Employee Database Management- 0.23* Employee Separation Process+0.15

Rewards and Recognition + e

(Employee usefulness)$_{\text{(mid management)}}$= 1.89 +0.23*Recruitment / Hiring Process+ 0.15* Performance Management System +0.23*Employee Communication Process - 0.12* Employee Separation Process+ 0.13* Rewards and Recognition + e

The employee usefulness of the eHRM is important in the present context that the technology adaption as the maintenance of details of talent capital in a branch along with the tasks assigned to them, performance, rewards and motivations from management, communications and notifications with bank and central admin system, leave attendance management, etc.

The regression constant of the equation is high for the executives (2.12) compared to other levels of employees. The coefficient of employee communication is high for executives while there is a negative coefficient for travel and mobility of employees. It is observed that, only for executives, the travel and mobility is important while it is not significant for other employee levels. Top management value more on employee data base than other employee levels. Employee separation process has a negative effect on employee usefulness.

4.1.2 The regression analyse to determine Technology adoption on implementation of EHRM

$H_{04.1}$: There is no significant effect of Technology adoption eHRM

This analyse explains how EHRM helps to an employee useful performance system. It is taken for different positions. The use of EHRM at different levels varies with the information retrieved.

Motivating factors	Mean

Readiness to adopt new technologies	3.21
Continuously changing work environment	3.39
Managing technology based customer complaints	3.36
Encourage customer adoption of new technologies	3.54
Managing safety of vital information	2.95
Improvising EHRM platform with the needs in technology driven organization culture	2.65
Clarity in roles and responsibilities	2.50
Increase in work density , but manageable at employee level	3.16

Customer and employee management	3.10
Monitoring the application based operation	3.13
Centralised resource planning	2.63
Maintenance of system	2.59

In Banking sector , the primary benefit of the technology adoption is to increase financial inclusion. But one of the big challenge is the slow innovation diffusion in rural areas. Also, the availability connectivity in rural areas are also critical. The rural population is adopting the new technologies with the increase in use of smart phones and increase in migration of rural youth to urban areas.

Table 4.4.1 Correlation coefficients of Regression Analysis of Age on factors influencing purchase decision

RossiSmith Academic Publications, Oxford/London UK, www.publicationsales.com

Level	R	R^2	F value	Significance level	Hypothesis Accepted
Executive	0.623	0.38813	3.698	0.008	H_1
Low Management	0.596	0.35522	9.652	0.000	H_1
Mid Management	0.589	0.34692	8.896	0.000	H_1
Top Management	0.139	0.019	1.569	0.192	H_0

Null hypothesis was accepted for top management.

RossiSmith Academic Publications, Oxford/London UK, www.publicationsales.com

(Readiness to adopt technology)$_{\text{Executive level}}$ = 2.32+ 0.28* Continuously changing work environment - 0.32* Managing technology based customer complaints + 0.33* Encourage customer adoption of new technologies +0.32* Maintenance of system + e

(Readiness to adopt technology)$_{\text{(low management)}}$ = 2.49 -0.15* Continuously changing work environment +0.17* Customer and employee management + 0.15* Clarity in roles and responsibilities +0.28* Monitoring the application based operation +0.35* Centralised resource planning - 0.23* Maintenance of system + e

(Readiness to adopt technology)$_{\text{(mid management)}}$= 2.39 +0.23* continuously changing work environment + 0.19* Managing safety of vital information +0.23* Improvising EHRM platform with the needs in technology driven organization culture - 0.12* Centralised resource planning + e

The model shows that the regression coefficient is high in adapting technology for executives due to the lower familiarity. But as the experience increases and continuous use of EHRM brings the coefficients low. In Mid management, strategic policy implementation is important.

Expert Opinion Analysis

Experts were selected in such a way that adequate number of respondents participate in the data collection. The representation was ensured in three domains : role , experience and occupational domain in the industry.

Selection of Experts

Table 1 Functional experience of experts

Journal of Asia Entrepreneurship and Sustainability
www.asiaentrepreneurshipjournal.com

	Position	Years of Experience	Number Experts in the team	%
Expert 1	Executive	6-14 years	52	26
Expert 2	Low level Management	7-12 years	56	28
Expert 3	Mid-level Management	5-18 years	58	29
Expert 4	Top Management	4-14 years	34	17

Total	200	100

Table 2:Expert selection : Domain and sector

	Executive	Low	Medium	High	Total	Percentage
Technical	11	13	12	9	45	22.5%
Managerial	14	11	17	8	50	25%
Administrative	14	15	14	9	52	26%
Customer Care	13	17	15	8	53	26.5%
Total	52	56	58	34	200	100%

Percentage	26%	28%	29%	17%	100%	

59% of the experts participated are from IT enabled services and 41% are from the consultancy services

Comparison Scale

Table 3: Comparison scale:

Comparison Scale of DEMATEL	
0	No influence
1	Low influence
2	Medium Influence
3	High influence

4	Very high influence

Comparison scale is described in a five point scale from 0 to 4 to express '0 for No influence' to 4 for 'Very high influence'.

Algorithm

1. Create Direct relationship matrix with work force attributes on rows and personal attributes on column
2. The frequency of each cell calculated as the weighted average of the response from the experts
3. Calculate row sum
4. Identify the largest value of the rows
5. Divide all the cell values by large number of the row sums to get neutralised direct relation matrix
6. Execute the Formula : T= X(I-X) where X is normalised direct relation matrix , I is the Identity matrix and T is the total relation matrix
7. Find the sum of Row elements and term it as D
8. Find the sum of column elements and term it as R
9. Find D+R and D-R

D+R gives the importance of criteria and D-R gives degree of relation of one criteria with other criteria.

Analysis & Interpretation

Weighted mean is used for comparing the variation of sub-factors across functional domains. The response is taken in five point scale.

Weighted mean= $\frac{\sum w*n}{\sum n}$ where w is the weight of the response and n is the number of response in variable.

Table 4: Expert response on sub-dimension

Dimensions	Sub Dimensions	Technical	Managerial	Administrative	Customer Care
Self-efficac	Perseverance of employees still helps maintain a resilient customer loyalty	4.1	3.3	2.9	3.2

y of employees	Employee shows a high emotional stability in solving customer issues	3.8	2.9	3.4	3.4
	Quick learning and computer anxiety improve adoption	3.9	3.3	3.5	2.9
Need of EHRM	Reduction in operating time	3.6	3.9	4.2	2.9
	Easy Coordination with all employees	3.9	3.1	2.3	3.7
	Centralised data management and decision making	2.8	2.3	2.9	3
Implementa	Number of leads for integrating all operations	2.6	2.1	2.7	2.3

tion of EHRM	Training of employees	3.3	2.6	3.9	2.1
	Aptness of EHRM design to the organization	2.5	2.8	2.3	2.9
	Report generation and employee notification	2.8	2.9	3.6	2.3
Standardizations in evaluations	Provision to evaluate unique performance	2.6	2.8	3.6	2.6
	Employee grievance/ complaint management	2.3	2.8	2.1	2.6
	Monthly salary, attendance updates	2.3	2.6	2.8	2.6

	Routine assignments, updates	3.2	2.8	2.7	3.1

Table 5: DEMATEL : Direct relationship matrix (Relative preference table)

	Executive	Low	Medium	High
Technical	0	2	1	4
Managerial	2	0	3	3
Administrative	1	3	0	3

Customer Care	4	3	4	0

(The cell values are calculated based on the weighted average formula

$A = \frac{\sum n_i * q_i}{\sum n_i}$ A is the value in each cell to represent the relative preference i_n is the frequency each numerical value of preference and q_i is the numerical value of the preference)

The Total relation matrix : Attributes of fund verses utility of fund

Personal / Work force	Executive	Low	Medium	High	D
Technical	0.526	0.5	0.	0.7	2.3

		97	52	41	84
Managerial	0.849	0.588	0.717	0.769	2.923
Administrative	0.975	0.89	0.613	0.897	3.375
Customer care	0.796	0.773	0.777	0.635	2.981
Index	3.146	2.8	2.6	3.0	

		48	27	42	

From the above table, executive and high management employees have more priority while administration and process high influence on eHRM. Technical team has the least effect.

Importance & Relationship table

Fund attributes	D	R	D+R	D-R
Technical	2.384	3.146	5.53	-0.762
Managerial	2.923	2.848	5.771	0.075
Administrative	3.375	2.627	6.002	0.748
Customer care	2.981	3.042	6.023	-0.061

The D+R value is the highest for Customer care (6.023) and then for Administrative (6.002) and then managerial (5.771) followed by risk and Technical (5.53).

. In the case of relative importance (D-R) , Administration (.817) followed by Managerial (.075) while Technical has a negative relative important value (-.762) followed return (-.061).

The relative importance shows that Administration is the top priority followed by managerial activity. Executive have more concern in eHRM followed by top management.

Conclusion

EHRM is one of the tools used internally for managing the HR resources that the tasks and responsibilities are assigned according to the talent demand at that time. Mean of the variables shows that learning and development programs (3.51), leave and attendance Management system (3.50), Recruitment / Hiring Process (3.49), On-boarding Process (3.31) , Travel and mobility Management (3.23) and Employee Data Base Management (3.20) have higher mean. The regression constant of the equation is high for the executives (2.12) compared to other levels of employees. The coefficient of employee communication is high for executives while there is a negative coefficient for travel and mobility of employees. It is observed that, only for executives, the travel and mobility is important while it is not significant for other employee levels. Top management value more on employee data base than other employee levels. Employee separation process has a negative effect on employee usefulness.

In the case effect of technology adoption, Readiness to adopt new technologies (3.21), Continuously changing work environment (3.39), Managing technology based customer complaint (3.36), Encourage customer adoption of new technologies (3.54), Increase in work density , but manageable at employee level (3.16), Customer and employee management (3.10) and Monitoring the application based operation (3.13) have high mean. This shows that the TAM factors have a high influence on EHRM.. The variables that contribute to the variation in dependent variable, readiness to adopt technology include, Continuously changing work environment , Customer and employee management , Clarity in roles and responsibilities , Monitoring the application based operation , Centralised resource planning , Maintenance of system, etc. The effective HRM strengthen the organization performance and customer satisfaction as well.

The importance is high for customer care followed by administration and managerial roles. In the case of relative importance (D-R) , Administration (.817) followed by Managerial (.075) while Technical has a negative relative important value (-.762) followed return (-.061).

The relative importance shows that Administration is the top priority followed by managerial activity. Executive have more concern in eHRM followed by top management.

The results shows the EHRM is effective in Banks that the employees can manage their HR related needs by themselves and coordinate effectively.

In brief, the self-efficacy, organization culture and technology adoption improve organization performance. This is supported by eHRM

RossiSmith Academic Publications, Oxford/London UK, www.publicationsales.com

Bibliography

Ajzen, I., & Fishbein, M. (2005). The influence of attitudes on behavior. In D. B. Albarracín, Johnson, & M. Zanna (Eds.), The Handbook of Attitudes (pp. 173-222).

AlHamad, A., Alshurideh, M., Alomari, K., Kurdi, B., Alzoubi, H., Hamouche, S., & Al, H. S. (2022). The effect of electronic human resources management on organizational health of telecommuni-cations companies in Jordan. (429-438, Ed.) International Journal of Data and Network Science, 6(2).

alriwali, S. s. (2020). The Impact of Emotional Intelligence on The Job Performance to Managers on Saudi Banks sector. Multi Knowledge Electronic comprehensive journal for education and science publications.

Ayyash, M. M., Herzallah, F. A., & Alkhateeb, M. A. (2021). Determinants of Employees' E-HRM Continuous Intention to Use: The Moderating Role of Computer Self-EfficacyInternational Journal of E-Business Research. International Journal of E-Business Research. doi:10.4018/IJEBR.309393

Boyle, G. J., Mathews, G., & Saklofske, D. H. (2008). Personality Theory and Assessment: Personality Theories and Models Vol 1. SAGE. Retrieved from The%20SAGE%20Handbook%20of%20Personality%20Theory%20and%20Assessment_%20Personality%20Theories%20and%20Models%20(%20PDFDrive%20).pdf

Brown, L., Malou, J., & Schutte, N. (2005). The effectiveness of self-efficacy intervention for helping adolescents cope with sport competition loss. Journal of Sports Behavior, 136–150. doi:https://samples.jbpub.com/9781449689742/chapter2.pdf

Butt, Z. I., Rashid, K., Mahanoor1, T. A., & Saeed, N. (2016). RELATIONSHIP BETWEEN PSYCHOLOGICAL WELL BEING AND EXERCISE. Science International.

Cherniss, C., & Goleman, D. (Eds.). (2001). Emotionally intelligent work place: How to Select for, Measure,and Improve Emotional , Intelligence in Individuals, Groups, and Organizations. JOSSEY-BASS. Retrieved from The%20Emotionally%20Intelligent%20Workplace_%20How%20to%20Select%2

0For,%20Measure,%20and%20Improve%20Emotional%20Intelligence%20in%20Individuals,%20Groups,%20and%20Organizations%20(%20PDFDrive%20).pdf
Cole, D., & Tibby, M. (2012). Defining and developing your approach to employability. Retrieved from https://www.heacademy.ac.uk/sites/default/files/resources/Employability_framework.pdf
Dacre Pool, L., & Sewell, P. (n.d.). . The key to employability: Developing a practical model of graduate employability. Education and Training,. 49(4). Retrieved from https://www.emerald.com/insight/content/doi/10.1108/00400910710754435/full/html
Davis, G. F., & DeWitt, T. (2021, September). Organization Theory and the Resource-Based View of the Firm: The Great Divide. Journal of Management, 47(7), 1684-1697. doi:10.1177/0149206320982650
DePietro, R., Wiarda, E., & Fleischer, M. (1990). The context for change: Organization, technology and environment. In L. Tornatzky, & M. Fleischer (Eds.), The Process of Technological Innovation . Lexington, MA: Lexington Books.
Deshwal, P. (2015). Role of E- HRM in Organizational Effectiveness and Sustainability. International Journal of Applied Research, 1(12), 605-609.
Dorsey, D., & Mueller-Hanson, R. (2019). PERFORMANCE MANAGEMENT THAT MAKES A DIFFERENCE: AN EVIDENSE BASED APPROACH. Retrieved from https://www.shrm.org/hr-today/trends-and-forecasting/special-reports-and-expert-views/documents/performance%20management.pdf
Fiori, M., & Vesely-Maillefer, A. K. (2018). Emotional Intelligence as!an!Ability: Theory, Challenges, and!New Directions. doi:https://doi.org/10.1007/978-3-319-90633-1_2
Galhena, B. L. (2015). Behavior of e-HRM Adoption: Empirical Evidence form Organizations in Developing Context. 2nd International Symposium on Partial Least Squares Path Modeling, Seville (Spain).
Gautam, A., & Khurana, C. (2019). DEMOGRAPHIC VARIABLES AS INDICATORS OF EMOTIONAL INTELLIGENCE: A STUDY OF SELECTED ENTERPRISES OF UTTARAKHAND. Journal of Management .

Giri, A., Chatterjee, S., Bag, M., & Paul, P. (2019). Determining 'Acceptability of E-HRM (Electronic - Human Resource Management)' in Indian Food Processing Industry using Augmented Technology Acceptance Model (TAM). International Journal of Recent Technology and Engineering (IJRTE), 8(1).
Goleman, D. (1998). Working with emotional intelligence. New York: BantamBooks.
Green, S. B. (1991). How Many Subjects Does It Take To Do A Regression Analysis. Multivariate behavioral research., 26(3), 499-510. doi:10.1207/s15327906mbr2603_7
Imandin, L., Bisschoff, C. A., & Botha, C. (2014). A model to measure employee engageme. Problems and Perspectives in Management.
Israel, G. D. (2003). Dertermining Sample saize. IFAS , University of Florida.
J, D., Mundfrom, Shaw, D. G., & Ke, T. L. (2005). Minimum Sample Size Recommendations for Conducting Factor Analyses. International Journal of Testing, 5(2).
JAIN, S., JAIN, V., & DAS, S. (2018). Relationship analysis between emotional ntelligence and service quality with special evidences from Indian banking sector. REVISTA ESPACIOUS.
Karoliny, M., & Poór, J. (2017). Emberi erőforrás kézikönyv – Rendszerek és alkalmazások. Budapest: Wolters Kluwer.
Kim, J., & Jung, H.-S. (2022). The Effect of Employee Competency and Organizational Culture on Employees' Perceived Stress for Better Workplace. International Journal of Environmental Research and Public Health, 4468. doi:https://doi.org/10.3390/ijerph19084428
Koopmans, L., Bernaards, C. M., Hildebrandt, V. H., & Henrica C.W. de Vet, A. J. (2013). Measuring Individual Work Performance Identifying and Selecting Indicators. In A Journal of Prevention, Assessment & Rehabilitation (Vol. 45). doi:DOI 10.3233/WOR-131659
Lippke, S. (2020). Self Efficacy Theory. In V. Zeigler-Hill, & T. Shackelford (Eds.), Encyclopedia of Personality and Individual Differences. Springer. doi:https://doi.org/10.1007/978-3-319-24612-3_1167
Marín, G., Hernández, A. J., Valle, I. d., & Castillo, M. Á. (2016). Organizational culture and family business: A configurational approach A configurational

approach. European Journal of Family Business. doi:https://doi.org/10.1016/j.ejfb.2017.05.002
Mayer, J. D., Caruso, R., D., & Salovey, P. (2004). Emotional intelligence meets traditional standards for an intelligence. Intelligence, 267-298.
Milon, M., Alam, M. A., & & Pias, M. H. (2022). Exploring the Key Practices of E-HRM in Place of Traditional HRM: A Study on Private Industry of Bangladesh. Human Resource and Sustainability Studies, 10(3), 403-417.
Muda, I., Rafik, A., & Harahap, M. R. (2014). Factors Influencing Employees' Performance: A Study on the Islamic Banks in Indonesia. International Journal of Business and Social Science.
N.Stock, G., L.McFadden, K., & R.Gowen, C. (2008). Organizational culture, critical success factors, and the reduction of hospital errors. International Journal of Production Economics. doi:https://doi.org/10.1016/j.ijpe.2006.07.005
Pant, S., Chatterjee, A., & Jaroliya, D. (2012). e-HRM Systems Implementation:A Conceptual Framework. Retrieved from https://www.iimidr.ac.in/wp-content/uploads/e-HRM-Systems-Implementation.pdf
Poreddy, D. (2018). Role of Electronic Human Resource Management in Banking Sector : A Study on SBI. Splint International Journal of Professionals, 5(2).
Rathee, R., & Bhuntel, M. R. (2022). A study on employee perception about the use of e HRM in IT. SCMS Journal of Indian Management, 18(1).
Ruël, H., & Bondarouk, T. (2004). E-HRM: innovation or irritation. An empirical explorative study on web-based HR in five large companies. Proceedings of the 12th European Conference on Information Systems (ECIS). Retrieved from . http://www.ecis2004.fi
Serrat, O. (2017). Understanding and Developing Emotional Intelligence. Asian Development Bank. doi:10.1007/978-981-10-0983-9-37
Shamout, M., Elayan, M., Rawashdeh, A., Kurdi, B., & Alshurideh, M. (2022). E-HRM practices and sustainable competitive advantage from HR practitioner's perspective: A mediated moderation analysis. International Journal of Data and Network Science, 6(1), 165-178.
Shekhar, A. (2011). Theories of Personality and Personality Testing. Retrieved from https://www.pdfdrive.com/theories-of-personality-and-personality-testing-e19434673.html

Thathsara, A., & Sutha, J. (2021). Investigating the influence of e-hrm practices on organizational performance: The mediating role of organizational agility (with special reference to financial institution). International Journal of Engineering and Management Research (IJEMR),, 11(1), 1-8.
Venkatesh, V., & Bala, H. (2008). 'Technology Acceptance Model 3 and a Research Agenda on Interventions. Decision Sciences, 39(2), 273-315.
Wright, P. M., & Snell, S. A. (1996). Human resource management, manufacturing strategy, and performance. Academy f Management Journal, 836-866.
Yorke, M., & Knight, P. (2006). Embedding employability into the curriculum; Learning and Employability. Series 1 No. 3. Higher Education Academy. Retrieved from http://www.heacademy.ac.uk/assets/York/documents/ourwork/employability/id460_ embedding_employability_into_the_curriculum_338.pdf

Journal of Asia Entrepreneurship and Sustainability
www.asiaentrepreneurshipjournal.com

Enhancing Portfolio Stability Through Financial Profile Analysis of Stocks

Nagendra B V, K Niharika Reddy, J.Joseph Durai Selvam, Rakshantha A, Sakshi Sharma, Arjun Muralisharan, Arsha Narayan, Rakesh Tigadi, Abhay Krishnan
CHRIST (Deemed to be) University, Bangalore
Karnataka, India

nagendra.bv@christuniversity.in,niharika.reddy@mba.christunivesity.in, joseph.selvam@christuniversity.in, rakshantha.a@mba.christuniversity.in, sakshi.sharma@mba.christuniversity.in, arjun.muralidharan@mba.christuniversity.in, arsha.narayan@mba.christuniversity.in, rakesh.tigadi@mba.christuniversity.in, abhay.krishnan@mba.christuniversity.in

Abstract

A key component of stock investing is having a portfolio of diversified stocks, which can help to mitigate the risk of investing in a single stock or sector. A common approach to stock diversification is constructing a portfolio based on the sector or industry. However, this strategy approach may not be ideal for all investors, as stocks within a sector or industry can have vastly different financial

profiles due to the varied business models and positioning of the company. This research paper explores the use of machine learning algorithms for diversifying a portfolio of stocks based on the financial profile of companies. In this context, clustering algorithms in machine learning can provide a solution to diversifying a portfolio based on the financial profile of companies. The study will begin by reviewing the importance of diversification in stock investing and the limitations of sector-based portfolio development. This will be followed by a step-by-step guide to implementing the clustering algorithm approach, including data collection, data cleaning and preprocessing, clustering, portfolio construction, and portfolio evaluation. The data collection stage involves collecting financial data on stocks, such as Earnings per share, Dividend per share, Net profit margin, Return on equity, Return on assets, PE ratio, Market cap, and return performance of the stock over a period. The data cleaning and preprocessing stage involinvolvening and preparing the financial data for use by the clustering algorithm. The clustering stage involves applying a clustering algorithm, such as K-Means or Hierarchical Clustering, to group stocks into clusters based on their financial profiles.

Research implications and limitations: The results of this research paper will demonstrate the feasibility of using clustering algorithms to diversify a portfolio of stocks based on their financial profiles by performing clustering algorithms on the set of sample stocks data collected from different platforms such as Grow platform, Screener and Moneycontrol. The paper will also provide insights into the strengths

and limitations of this approach to portfolio construction, and highlight the potential benefits for investors. The results of the study demonstrate the potential benefits of using clustering algorithms for diversifying portfolios, providing a useful tool for investors looking to improve their risk-return profile.

Keywords: Diversifying portfolio, Clustering algorithms, Financial profile, Portfolio construction, Risk-return profile, Investment strategy

Introduction

Investing in stocks is a common strategy for building wealth over the long-term. With the stock market offering the potential for high returns, many investors are attracted to the idea of investing in stocks to grow their wealth. However, investing in a single stock or a narrow range of stocks can be risky, as stock prices can be volatile and are subject to a range of economic, political, and market factors and the reason for loss is due to several factors, including

Market Volatility	The stock market is inherently volatile and stock prices can fluctuate rapidly due to a range of economic, political, and market factors. If a stock's price drops significantly, this can result in substantial losses for investors.

Company-Specific Risks	Investing in a single stock or a narrow range of stocks exposes the investor to company-specific risks. For example, if the company experiences financial difficulties, such as declining sales or increasing debt, this can result in a drop in its stock price and a loss for investors.
Economic Cycles	The stock market is influenced by economic cycles and changes in the economy can impact the performance of individual stocks. For example, a recession can result in a drop in stock prices and a loss for investors.
Political Risk	Political events and changes can also impact the stock market and the performance of individual stocks. For example, changes in government policies or regulations can impact the financial performance of a company, which in turn can impact its stock price.
Market Risks	The stock market is subject to a range of risks, such as market trends and changes in market sentiment, which can impact the performance of individual stocks. For example,

	a bear market, where the overall market trend is downward, can result in a drop in stock prices and a loss for investors.

Table 1: Factors responsible for loss in investing in a stock

The above listed factors can result in substantial losses for investors who have invested in a single stock or a narrow range of stocks. To reduce the risk associated with investing in stocks, it is important to diversify a portfolio of stocks by including stocks from a range of sectors and industries. The idea behind diversification is to ensure that the portfolio is not overly exposed to any one particular sector or industry. If one sector or industry performs poorly, the overall performance of the portfolio should be cushioned by the positive performance of other sectors or industries.

The traditional approach to diversifying a portfolio of stocks is based on sector or industry classifications. This approach involves dividing the stock market into broad categories, such as technology, healthcare, financials, etc., and then investing in a range of stocks from each sector. This approach can be effective in reducing the risk associated with investing in a single sector, as it spreads the portfolio across a range of sectors and industries. However, while this approach can be effective in reducing risk, it may not always be suitable for all investors. For instance, stocks within a sector can have vastly different financial profiles, making it difficult to achieve an optimal level of diversification. For example, a portfolio of stocks from the technology sector may include companies that specialize in hardware, software, and services. These companies can have vastly different financial profiles, making it difficult to achieve an optimal level of diversification using sector classifications alone.

Some of the factors that contribute to different financial profiles of companies within a same sector are

Business Model	Companies within a sector can have different business models, which can result in different financial profiles. For example, companies within the technology sector can specialize in hardware, software, or services, and each of these areas can have different financial profiles.
Market Position	Companies within a sector can have different market positions, which can result in different financial profiles. For example, companies that have a dominant market position may have higher profits, while companies that are struggling to compete may have lower profits.
Financial Metrics	Companies within a sector can have different financial metrics, such as revenue, profits, earnings per share, and

RossiSmith Academic Publications, Oxford/London UK, www.publicationsales.com

	debt-to-equity ratios, which can result in different financial profiles.
Growth Prospects	Companies within a sector can have different growth prospects, which can result in different financial profiles. For example, companies that are growing rapidly may have a higher stock price, while companies that are struggling may have a lower stock price.
Regulatory Environment	Companies within a sector can be subject to different regulatory environments, which can result in different financial profiles. For example, companies in regulated industries, such as healthcare or financials, may face more restrictions and regulations, which can impact their financial profile.

Table 2: Factors that contribute to different financial profiles of companies

To overcome this limitation, this research paper explores the use of machine learning algorithms for diversifying a portfolio of stocks based on the financial profile of companies.

RossiSmith Academic Publications, Oxford/London UK, www.publicationsales.com

Literature Review

Incorporating financial profile information into the stock selection process improved portfolio diversification and risk-adjusted returns compared to a traditional stock selection method based on market capitalization (Kim et al., 2020). (Liu et al. ,2019) through their study on China A -shares stock data found that using financial profile information in combination with technical indicators improved portfolio diversification and risk-adjusted returns compared to a portfolio based on technical indicators alone. (J. Martinez et al.,2010) found that in order to accurately reflect a company's financial health and future performance new stock selection method that incorporates financial profile information, such as a company's profitability, liquidity, and solvency ratios. This information is used to create a financial profile for each company and then compared to the financial profiles of other companies in the same industry. Another study by (Brown, T.,2007) found that portfolios constructed using financial profile information achieved higher returns and greater diversification compared to portfolios constructed using traditional methods, suggesting that incorporating financial profile information into the stock selection process can lead to improved investment outcomes.

Most recently, (Lhabitant et al.,2017) presented a more extensive and technical examination of portfolio diversification. His paper also discusses portfolio diversification concerns such as the household diversification conundrum, the diversification vs concentration quandary, and the interpretation of the (Pearson) correlation. Other papers contain brief but informative analyses of portfolio diversification strategies or measurements. (Deguest et al.,2013), for example, divide portfolio diversification metrics into two categories: weight-based measures and risk-based measures. This classification is used by (Carli et al.,2014) in their review of portfolio diversification measures. In his study of portfolio diversification measures, (Sharma,2018) distinguishes between factor-based and risk-based measures. As a result, he divides measures into three categories: weight-based measures, risk-based measures, and factor-based measures. Portfolio diversification is the process of spreading money across a variety of assets. Its advantage is risk reduction, which reduces both the likelihood and severity of portfolio loss through multilateral insurance in which each asset is protected by the other assets.

Statman points out the misunderstanding of (Pearson) correlations in terms of diversification benefits based on the studies of (Statman, 2005) and (Scheid, 2008).

Even those who believe Markowitz is incorrect continue to utilise mean-variance portfolio theory terminology.

Correlation is not a good measure, according to (Statman,2008) and (Scheid, 2013), for two reasons. First, we have a tendency to misinterpret correlations. High correlations result in limited diversity benefits, however correlations greater than 0.90 result in substantial diversification benefits. Second, the benefits of diversity are determined not only by the correlations between investment returns, but also by the standard deviations of the returns of each investment. Return gaps are superior measurements of the benefits of diversification than correlations because they account for the effects of both correlations and standard deviations while also offering intuitive measures of the benefits of diversification.

Volatility and unpredictability are simply two characteristics of modern economies, therefore we can conclude that risk has become an essential aspect of business life. Investing in a large number of companies decreases unsystematic (idiosyncratic) risk, i.e., risk peculiar to a certain company, and so reduces portfolio volatility. According to research, these two components of overall risk are not fully independent dimensions (Lubatkin et al., 1994), and idiosyncratic risk is the most significant contribution to total volatility (Marcelo et al., 2012).

However, even with a large number of assets in the portfolio, total risk cannot be totally avoided. There will always be a portion of the overall risk relating to systematic risk variables, which can only be diversified further through international diversification. Due to the law of large numbers, the exposure to any source of risk is modest because the sources of risk are independent and the investment is diversified across a large number of securities.

Numerous studies have shown that the traditional rule of thumb of 8-10 stocks established by (Evans et al., 1968) is indeed sufficient to achieve optimal diversification effects, but numerous works, particularly recently, have challenged this fact by demonstrating that 30-50 stocks are required for maximum diversification effect (Benjelloun 2010; Chong and Phillips 2013; Alexeev and Tapon 2014; Bradfield and Munro 2017; Oyenubi 2019).

(Renata Paola Dameri et al., 2020) attempted the use of neural networks, specifically self-organizing maps (SOMs), to assess and cluster the financial performance of firms. SOMs are commonly used to analyse financial statement data; however, in this paper, SOMs were used to overcome several limitations encountered in previous works on financial reporting indicators, such as the small number of companies in the sample, the limited number of ratios, the homogeneity of the economic sector, and the lack of explanation and further analysis of the SOM outputs.

In a paper by (SR Nanda et al,2010) a data mining approach for clustering stocks into groups is provided. Following the classification, stocks from these groupings could be chosen to form a portfolio. The results of the investigation reveal that K-means cluster analysis produces the best compact clusters for stock categorization data when compared to SOM and Fuzzy C-means.

(Markowitz,1952) created a model for generating an efficient portfolio. The return of a stock is the mean return in the Markowitz model, and the risk of a stock is the standard deviation of the stock returns. The portfolio return is the sum of stock returns. The collection of portfolios with the highest return for each degree of risk is known as the efficient frontier of portfolios (or equivalently, portfolios with the

Journal of Asia Entrepreneurship and Sustainability
www.asiaentrepreneurshipjournal.com

lowest risk for a given level of return). Risk reduction is measured by investors through investment diversification. Since then, much work has been done on portfolio management. (Topaloglou et al., 2008) developed a dynamic stochastic programming model for international portfolio management, which identifies capital allocations to foreign markets, asset selection within each market, and appropriate currency hedging levels. Oh, Kim, and Min used genetic algorithms for portfolio optimization in index fund management (2005). (Fernandez, 2005) proposes a stochastic control model that accounts for both ecological and economic uncertainty in the management of both types of natural resources. Fuzzy models for dynamic portfolio management (Stermark, 1996) have also been implemented.

Hence from the literature study, there are very few studies on clustering stock data, however there has been a lot of work on portfolio optimization. Initial cluster indexing of stock data can be beneficial to optimization models, increasing their efficiency.

Research Methodology

System design

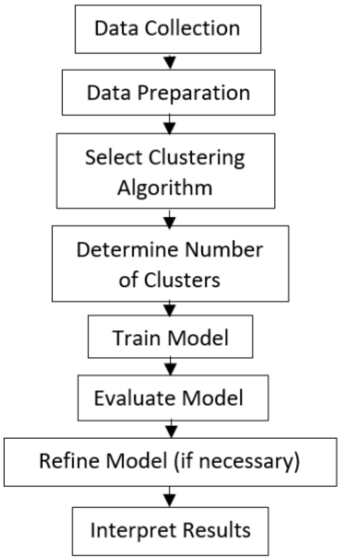


Figure 1: Flowchart of the methodology

The proposed system design is shown in Figure 1 and is discussed in detail as follows.

1. Data Collection: The first step in the process is to collect the data that will be used for clustering. Data was collected from various sources and combined together. The sources are:

   - Groww – It is an online trading platform. It was used to get data about different companies listed in the Stock exchange.
   - Screener - Screener is a tool for stock analysis. It provides the investor with a variety of facts on the listed firms on the Indian Stock Exchange. The Website does stock analysis using contemporary technologies.

- Money control – Financial ratios of different companies are extracted from moneycontrol.com.

2. Data Preparation: After the data has been collected, it must be prepared for use in the clustering analysis. This includes cleaning and transforming the data, handling missing values, and normalizing the data to ensure that all variables are on the same scale which is performed accordingly.

3. Select Clustering Algorithm: The next step is to choose a suitable clustering algorithm for the data and problem at hand. There are several types of clustering algorithms, including centroid-based methods (such as K-Means), hierarchical methods, density-based methods (such as DBSCAN). Here we use kmeans and hierarchical clustering for which the detailed pre-processing steps are given below:

- K-Means Clustering: K-Means is a centroid-based method of clustering. It works by dividing a dataset into a specified number of clusters (K), with each cluster represented by its centroid. The following are the steps involved:
  - Data Pre-processing: Prepare the data by cleaning and normalizing it, if necessary.
  - Specify the number of clusters (K): Choose the number of clusters that you want to divide the data into.
  - Initialize Centroids: Randomly initialize the centroids for each cluster.

  - Assign Points to Clusters: Assign each data point to the nearest centroid based on Euclidean distance or another similarity measure.
  - Update Centroids: Calculate the mean of the data points assigned to each centroid and update the centroids.
  - Repeat Steps 4 and 5 until the centroids no longer change or a stopping criterion is met.
  - Evaluate Clusters: Evaluate the quality of the clusters using metrics such as silhouette score or within-cluster sum of squares.
  - Re-run the algorithm if necessary: If the results are not satisfactory, re-run the algorithm with different initial centroids or a different number of clusters.
- Hierarchical Clustering: It is a tree-based method of clustering. It works by creating a hierarchical representation of the data, in which individual observations are merged into larger and larger clusters until a single cluster contains all the observations.
  - Data Pre-processing: Prepare the data by cleaning and normalizing it, if necessary.
  - Choose the linkage method: Choose the linkage method (single, complete, average, etc.) that we want to use for defining the proximity between clusters.
  - Calculate the proximity matrix: Calculate the proximity matrix that measures the distance between each pair of data points.

- Create the dendrogram: Use the linkage method and the proximity matrix to create the dendrogram, which represents the hierarchical structure of the data.
- Cut the dendrogram: Choose the number of clusters (K) by cutting the dendrogram at a certain height or by choosing a certain number of clusters from the dendrogram.
- Assign points to clusters: Assign each data point to the corresponding cluster based on the cut dendrogram.
- Evaluate Clusters: Evaluate the quality of the clusters using metrics such as silhouette score or cophenetic correlation coefficient.
- Re-run the algorithm if necessary: If the results are not satisfactory, re-run the algorithm with different linkage methods or a different number of clusters.

- Gaussian Mixture Model:
  - Data Pre-processing: Prepare the data by cleaning and normalizing it, if necessary.
  - Specify the number of clusters (K): Choose the number of clusters that you want to divide the data into.
  - Initialize Parameters: Initialize the parameters of the Gaussian distributions (mean, covariance matrix, and mixture weights) for each cluster.

- Expectation-Maximization (EM) Algorithm: Repeat the following two steps until convergence: a. Expectation (E) step: Compute the responsibility of each Gaussian distribution for each data point. b. Maximization (M) step: Re-estimate the parameters of the Gaussian distributions based on the responsibilities.
- Assign Points to Clusters: Assign each data point to the cluster with the highest responsibility.
- Evaluate Clusters: Evaluate the quality of the clusters using metrics such as the log-likelihood, Bayesian information criterion (BIC), or Akaike information criterion (AIC).

4. Determine Number of Clusters: Once the algorithm has been selected, the next step is to determine the number of clusters to use. This can be done using methods such as the elbow method, silhouette analysis, or the gap statistic. These methods help to determine the optimal number of clusters that can be used to best represent the structure of the data.

5. Train Model: This involves fitting the model to the data, and calculating the cluster assignments for each data point. The algorithm will iteratively adjust the centroids of the clusters until they best represent the structure of the data.

6. Evaluate Model: The next step is to evaluate the quality of the clustering solution. This can be done using metrics such as the silhouette score, the adjusted Rand index, or the Calinski-Harabasz index. These metrics provide a

quantitative measure of the quality of the solution and can be used to determine whether the model is performing well or if it needs to be refined.

7. Refine Model (if necessary): If necessary, the model can be refined by adjusting the parameters of the algorithm or the number of clusters. This process can be repeated until the quality of the solution is acceptable.

8. Interpret Results: Finally, the results of the clustering analysis can be interpreted and used to gain insights into the structure of the data. This may involve visualizing the clusters using techniques such as scatter plots, dendrograms, or heatmaps. The results can also be used to make predictions or to inform further analysis which is explained below.

Problem Statement

"The traditional approach to diversifying a portfolio of stocks based on sector or industry classifications may not always be suitable for all investors, as stocks within a sector can have vastly different financial profiles, making it difficult to achieve an optimal level of diversification. As a result, there is a need for a new approach to diversifying a portfolio of stocks that takes into account the financial profile of individual companies."

The approach of using clustering algorithms for diversifying portfolios of stocks based on the financial profile of individual companies is novel for several reasons:

- Integration of Machine Learning: This approach combines the power of machine learning algorithms with traditional portfolio optimization methods, leveraging the ability of clustering algorithms to identify patterns and relationships in financial data that may not be immediately apparent.
- Focus on Financial Profiles: By focusing on the financial profile of individual companies, rather than sector classifications, this approach takes into account the unique characteristics and risk factors associated with each stock, resulting in a more diversified portfolio.
- Improved Diversification: Clustering algorithms can group stocks into clusters based on their financial profiles, providing a more nuanced and effective approach to diversifying a portfolio of stocks than traditional sector-based approaches.
- Better Risk Management: By considering the financial profiles of individual companies, this approach can help to manage and reduce the risk associated with investing in stocks, resulting in improved risk-return trade-offs.

Diversifying portfolios of stocks based on the financial profile of individual companies is important for following reasons:

Improved Investment Decisions	By considering the financial profiles of individual companies, this approach provides investors with a more comprehensive and nuanced understanding of the risks and potential returns associated with investing in stocks.
Better Portfolio Performance	The ability of clustering algorithms to identify patterns and relationships in financial data can lead to improved portfolio diversification and risk management, resulting in better portfolio performance over time.
Increased Confidence	By using machine learning algorithms to diversify portfolios of stocks, investors can feel more confident in their investment decisions and better equipped to manage risk and uncertainty in the financial markets
Reduced Portfolio Risk	The use of clustering algorithms for diversifying portfolios of stocks can help to reduce the risk associated with investing in a single stock or a narrow

	range of stocks, providing a more effective and efficient approach to portfolio risk management.
Better Return on Investment	By combining the power of machine learning algorithms with traditional portfolio optimization methods, this approach can lead to improved risk-return trade-offs, providing investors with a better return on investment over time.

Table 3: Reasons to Diversify a Portfolio

Objectives

- Demonstrate the potential benefits of using clustering algorithms to diversify a portfolio based on the financial profile of companies.
- Provide insights into the strengths and limitations of adopting clustering algorithms for diversifying portfolios.

Journal of Asia Entrepreneurship and Sustainability
www.asiaentrepreneurshipjournal.com

- Investigate the effectiveness of this approach for portfolios of different sizes and levels of risk tolerance.
- Provide recommendations for investors on the use of machine learning algorithms for diversifying portfolios of stocks based on the financial profile of individual companies.

Analysis

Models used:

-K-means:

K-means is a popular algorithm for clustering data points into K clusters. Some of the key formulas used in K-means are:

1. Distance/Similarity Measure: To determine the similarity between a data point and the centroid of a cluster, Euclidean distance is often used in K-

d(X, Y) = sqrt((X1 - Y1)^2 + (X2 - Y2)^2 + ... + (Xn - Yn)^2)

means. The formula for Euclidean distance between two points X and Y is:

RossiSmith Academic Publications, Oxford/London UK, www.publicationsales.com

2. Centroid Computation: The centroid of a cluster is the mean of all the data

$$\boxed{\text{centroid}(C) = (\text{sum}(X1) / N, \text{sum}(X2) / N, ..., \text{sum}(Xn) / N)}$$

points in the cluster. The formula to compute the centroid of a cluster C containing N data points is:

3. Assignment Step: In each iteration of the K-means algorithm, data points are

$$\boxed{C = \text{argmin}(d(X, \text{centroid}(C)))}$$

assigned to their closest centroid. The formula to assign a data point X to a cluster C is:

4. Updating Centroids: After assigning all data points to clusters, the centroids are updated to the mean of all data points in the cluster. The formula to update

$$\boxed{\text{centroid}(C) = (\text{sum}(X1) / N, \text{sum}(X2) / N, ..., \text{sum}(Xn) / N)}$$

the centroid of a cluster C is:

These formulas are repeated until the centroids no longer change, or a maximum number of iterations is reached.

RossiSmith Academic Publications, Oxford/London UK, www.publicationsales.com

Journal of Asia Entrepreneurship and Sustainability
www.asiaentrepreneurshipjournal.com

-Agglomerative Hierarchical Clustering:

Agglomerative hierarchical clustering is a bottom-up clustering method that starts with individual data points as separate clusters and merges them into larger clusters in a sequential manner. The following are some of the key formulas used in agglomerative hierarchical clustering:

1. Similarity or Distance Matrix: The first step in agglomerative hierarchical clustering is to calculate a similarity or distance matrix between the data points. The matrix is a square matrix of size n x n, where n is the number of data points. The entry (i, j) in the matrix represents the similarity or distance between data points i and j.

2. Linkage Criteria: The linkage criteria determines the way in which two clusters are merged into a single cluster. The most common linkage criteria are:

The "ward.D" method in agglomerative hierarchical clustering refers to a specific linkage criteria that is used to merge clusters. The "ward.D" method is an abbreviation for Ward's method of linkage, which uses the sum of squared differences in cluster means as the criterion for merging clusters.

Ward's method tries to minimize the variance of the distances between the new cluster formed by merging two smaller clusters and the data points in the two

d(C_i, C_j) = sqrt((|C_i| + |C_j|) * d(C_i, C_j)^2 / |C_i| / |C_j| - (|C_i| * mean(C_i)^2 + |C_j| * mean(C_j)^2) / (|C_i| + |C_j|))

smaller clusters. The formula for Ward's method is:

where C_i and C_j are the two clusters being merged, d(C_i, C_j) is the Euclidean distance between the cluster means, mean(C_i) is the mean of the data points in cluster C_i, mean(C_j) is the mean of the data points in cluster C_j, |C_i| is the number of data points in cluster C_i, and |C_j| is the number of data points in cluster

These are the key formulas used in agglomerative hierarchical clustering. The choice of linkage criteria will depend on the characteristics of the data and the desired properties of the clustering solution.

- GMM (Gaussian Mixture Model):

GMM (Gaussian Mixture Model) is a probabilistic model-based clustering method that assumes that the data is generated by a mixture of several Gaussian distributions. The following are the key formulas related to GMM:

1. Gaussian Distribution: A Gaussian distribution, also known as a normal distribution, is defined by its mean μ and covariance Σ. The formula for a Gaussian distribution is:

$$f(x) = (1 / (\mathrm{sqrt}(2 * pi)\text{^}d * \det(\Sigma)\text{^}0.5)) * \exp(-0.5 * (x - \mu)\text{^}T * \Sigma\text{^}{-1} * (x - \mu))$$

where x is a d-dimensional data point, d is the number of features, Σ is the covariance matrix, $\det(\Sigma)$ is the determinant of the covariance matrix, and exp is the exponential function.

2. Likelihood: The likelihood of a data point x given the Gaussian distribution f is defined as the probability of observing the data point x given the parameters of the Gaussian distribution. The formula for the likelihood is:

$$p(x \mid f) = f(x)$$

3. Mixture Model: A mixture model is defined as a linear combination of several Gaussian distributions. The formula for a mixture model with K Gaussian distributions is:

$$f(x) = \mathrm{sum}(w\_k * f\_k(x)) \text{ for } k = 1 \text{ to } K$$

RossiSmith Academic Publications, Oxford/London UK, www.publicationsales.com

where w_k is the weight of the k-th Gaussian distribution and f_k(x) is the k-th Gaussian distribution.

4. Expectation Maximization (EM): The EM algorithm is used to estimate the parameters of the Gaussian mixture model (GMM). The EM algorithm iteratively updates the estimates of the mean μ, covariance Σ, and weight w of the Gaussian distributions until the log-likelihood of the data given the GMM converges.

5. The E-step computes the responsibilities, which are the probabilities of each

r_ik = p(z_k = 1 | x_i, Θ) = w_k * f_k(x_i) / sum(w_l * f_l(x_i)) for l = 1 to K

data point belonging to each of the Gaussian distributions. The formula for the responsibilities is:

where Θ is the set of parameters of the GMM, z_k is the latent variable indicating if the data point x_i belongs to the k-th Gaussian distribution, and f_k(x_i) is the probability density function of the k-th Gaussian distribution evaluated at x_i.

RossiSmith Academic Publications, Oxford/London UK, www.publicationsales.com

Journal of Asia Entrepreneurship and Sustainability
www.asiaentrepreneurshipjournal.com

6. The M-step updates the estimates of the parameters of the GMM based on the responsibilities. The formulas for updating the mean and covariance of the k-

$$\mu_k = (\text{sum}(r_{ik} * x_i)) / (\text{sum}(r_{ik}))\quad \Sigma_k = (\text{sum}(r_{ik} * (x_i - \mu_k)(x_i - \mu_k)^T)) / (\text{sum}(r_{ik}))$$

th Gaussian distribution are:

7. The formula for updating the weight of the k-th Gaussian distribution is:

$$w_k = \text{sum}(r_{ik}) / n$$

where n is the number of data points.

8. The EM algorithm alternates between the E-step and M-step until convergence. The log-likelihood of the data given the GMM is a measure of

$$L(\Theta) = \text{sum}(\log(\text{sum}(w_k * f_k(x_i)))) \text{ for } i = 1 \text{ to } n$$

the goodness of fit of the GMM to the data and can be used as a stopping criterion for the EM algorithm. The formula for the log-likelihood is:

where Θ is the set of parameters of the GMM.

These are the key formulas related to GMM based model based clustering.

Results and Discussion

There are several measures that may be used to measure how effectively the clustering has worked. The minimum intra-cluster distance and the maximum inter-cluster distance define ideal clustering.

To evaluate the performance of clustering, there are primarily two types of measurements. Ground truth labels are needed for extrinsic measures. Examples include the V-measure, the adjusted Rand index, the Fowlkes-Mallows scores, the mutual information-based scores, the homogeneity, and the completeness. Ground truth labels are not necessary for intrinsic measures. The Silhouette Coefficient, the Davies-Bouldin Index, the Calinski-Harabasz Index, and others are examples of clustering performance measurements.

Due to the unlabelled nature of the data, not all assessment measures can be used to evaluate the clustering models. Consequently, the two metrics employed are:

- Silhouette Score: The silhouette value evaluates an object's cohesiveness with its own cluster in relation to other clusters (separation). The silhouette varies from 1 to +1, with a high value indicating that the item is well matched to its own cluster and poorly matched to nearby clusters.

- Calinski Harabasz Score: The ratio of the sum of between-cluster dispersion and inter-cluster dispersion for all clusters is the Calinski-Harabasz index, sometimes referred to as the Variance Ratio Criterion. The greater the score, the better the performances, although there is no cut-off point for the values.

Model	Average Silhouette Width	Calinski Harabasz Score
K-means	0.13	22.84
Agglomerate Hierarchical	0.02	6.73
Gower Distance clustering	0.05	11.36
Model Based	-0.1	6.34

Table 5: Comparing the measures of different clustering methods

The table above illustrates the values of both the evaluation metrices for all the models that have been created. By observing the values of both matrices, we can say that K-means clustering model outperforms the other three clustering models. Hence, it has been chosen as the final model for interpretation.

Journal of Asia Entrepreneurship and Sustainability
www.asiaentrepreneurshipjournal.com

Determining the number of clusters (k):

<u>Silhouette Method</u>:

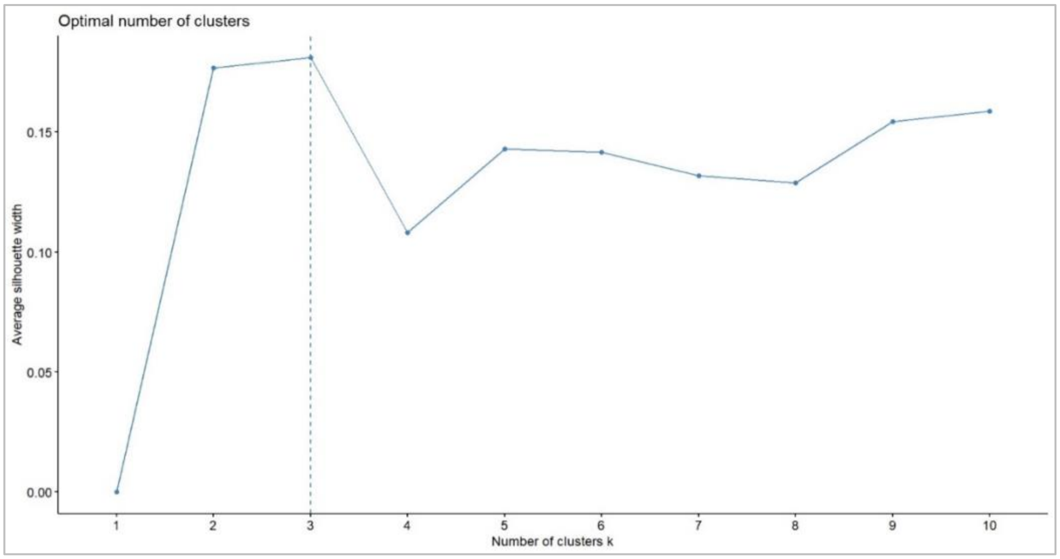


Figure 2: Plot to find the number of optimal clusters

The silhouette method is based on the idea that the optimal number of clusters is the value of k that results in the highest average silhouette width. The method involves fitting the k-means model for a range of values of k and calculating the silhouette width for each model. The optimal number of clusters is typically considered to be the value of k that corresponds to the highest average silhouette width.

We have used fviz_nbclust function to plot optimal number of clusters graph using silhouette method. According to the above screenshot, the optimal number of clusters is 3 as per silhouette. Therefore, the value k=3.

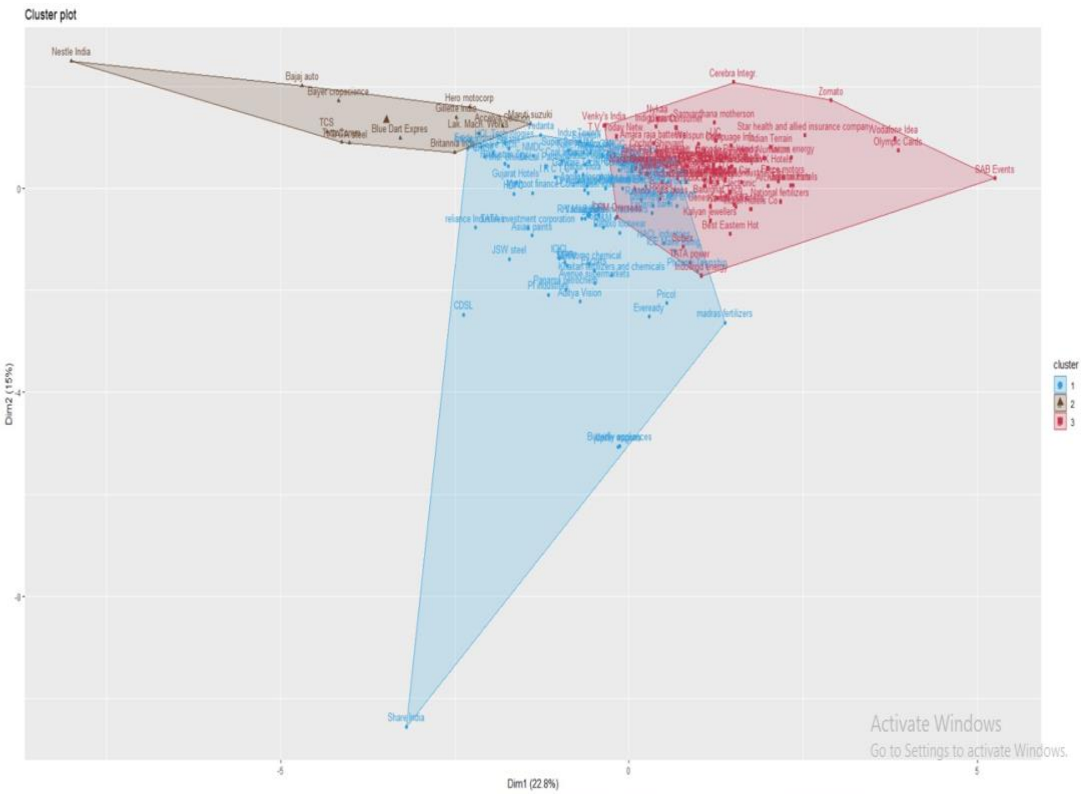


Figure 3: Cluster plot

Profiling

a) <u>Cluster 1(High Risk Low Return)</u>: This cluster represents stocks which are low dividend paying companies and have low ROA. A low return on assets (ROA) generally indicates that a company is not using its assets efficiently to generate profits, which is a sign of risk. This cluster is also suited for long term investments. Hence this could be labelled as risky companies for investors and be cautious while investing as EPS is also low.

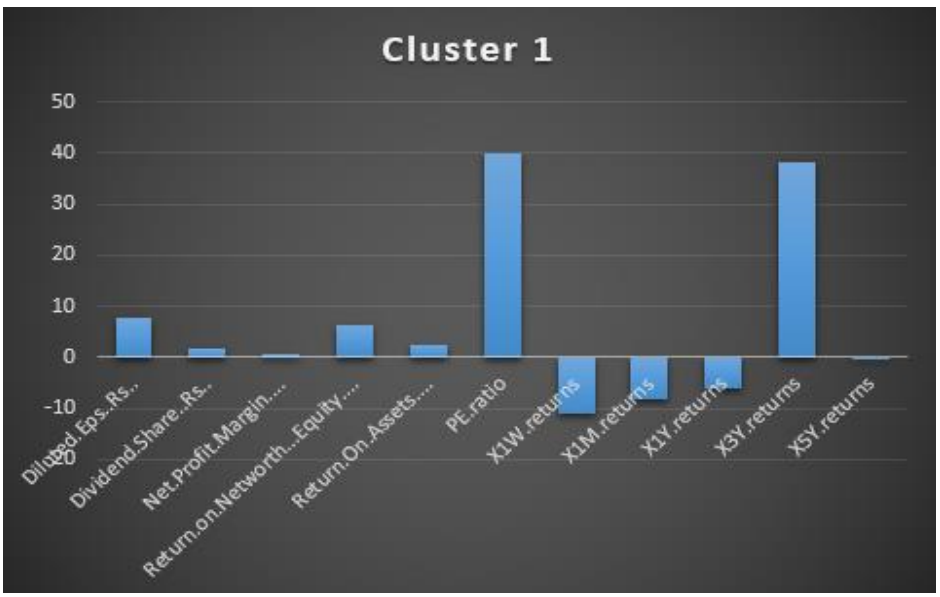


Figure 4: Companies in Cluster 1

b) <u>Cluster 2(Low Risk High Return)</u>: This cluster represents stocks which are fundamentally strong, with High EPS, ROA and ROE. This cluster also benefits the investor by paying dividends. Hence these could be considered as better investment options for the investors as they are less riskier than cluster one and three.

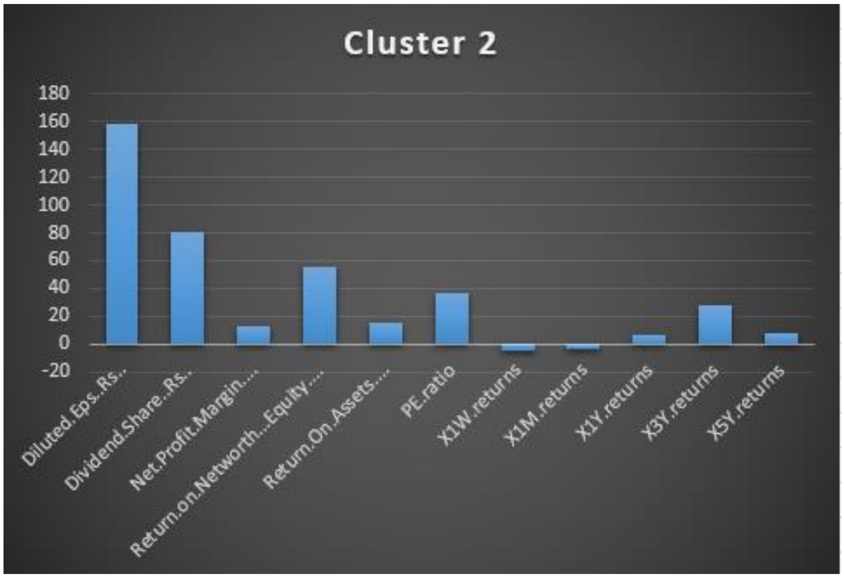


Figure 5: Companies in Cluster2

c) <u>Cluster 3(Moderate Risk and Return)</u>: This cluster represents stocks that have low EPS. A low EPS may be a sign that the company is struggling to control its costs, increase its revenue, or maintain its profitability. It may also indicate that the company is facing challenges in its industry or that it is not performing as well as expected. This cluster is also having low ROE. A low ROE may be a sign that the company is struggling to generate profits or that it is taking on more risk than its peers. It may also indicate that the company is not effectively utilizing the capital invested in it by its shareholders to generate returns. Hence they could be considered as potential stocks to get good return but more riskier than cluster 2 companies and less riskier than cluster one companies.

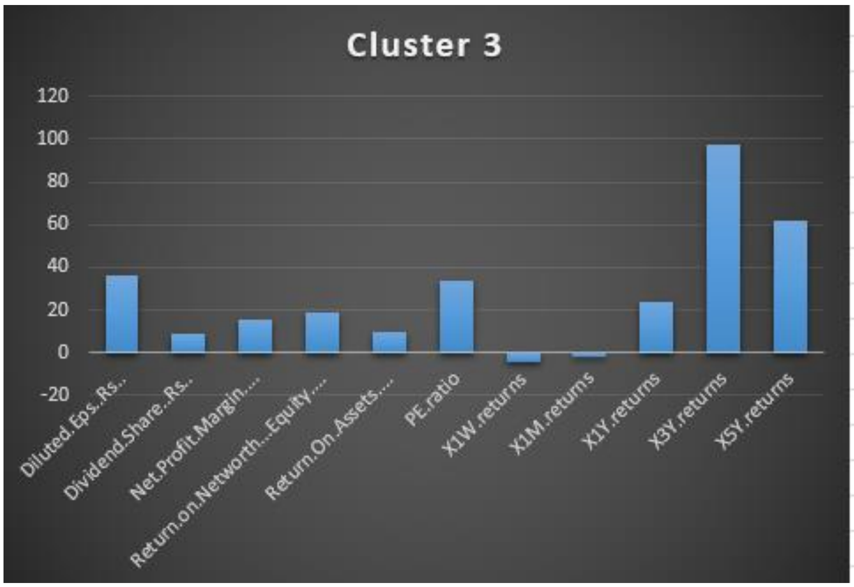


Figure 6: Companies in Cluster 3

<u>Example</u>: In cluster 2 we can see companies like TCS which are fundamentally strong and are paying good dividends to its investors.

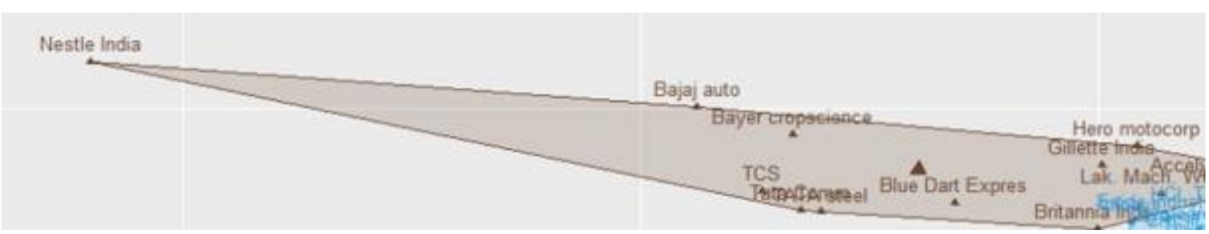


Figure 7: A example of cluster 2

Limitations

It is important to keep in mind that clustering is just one tool among many that investors can use to analyse and make decisions about the stock market. It is essential to conduct thorough research and consider a range of factors, including economic conditions, company performance, and risk tolerance, before making any investment decisions. The partitioning of clusters can be improved with much more parameters. In short, the model could be further improved with additional parameters.

Conclusion

By establishing a stable set of starting points, the clustering was performed using the k-means, hierarchical clustering, Gower distance clustering, and model-based clustering methods. Due to its superior Silhouette and Calinski scores compared to other models, K-means was selected as the superior model. The main concern in this case is the number of clusters, which was addressed by using the cluster silhouette metric along with visual interpretation of the parallel coordinate results. Three clusters were the end result of the partitioning. On truncated subsets of the original data, this partitioning is reliable and repeatable.

It is generally concluded that clustering can be a useful tool for identifying groups of stocks with similar characteristics and trends. By applying clustering techniques

RossiSmith Academic Publications, Oxford/London UK, www.publicationsales.com

to stock data, investors can potentially gain insights into the overall structure and dynamics of the market, and use this information to inform their investment decisions and portfolio diversification strategies. Overall, the conclusion is that clustering can be a valuable tool for investors looking to diversify their portfolio and manage risk, but it should be used in conjunction with other analysis and decision-making approaches.

RossiSmith Academic Publications, Oxford/London UK, www.publicationsales.com

References

Renata Paola Dameri, Roberto Garelli, Marina Resta; Neural Networks in Accounting: Clustering Firm Performance Using Financial Reporting Data. Journal of Information Systems 1 June 2020; 34 (2): 149–166

Nanda, S. R., Mahanty, B., & Tiwari, M. K. (2010). Clustering Indian stock market data for portfolio management. Expert Systems with Applications, 37(12), 8793-8798.

Lim, K. P., Brooks, R. D., & Kim, J. H. (2008). Financial crisis and stock market efficiency: Empirical evidence from Asian countries. International review of financial analysis, 17(3), 571-591.

Bae, J., Yang, X., & Kim, M. I. (2021). ESG and stock price crash risk: role of financial constraints. Asia-Pacific Journal of Financial Studies, 50(5), 556-581.

Markowitz, H. (1952). Portfolio selection. Journal of Finance, 7, 77–91. Michaud, P. (1997). Clustering techniques. Future Generation Computer System, 13(2), 135–147. Mingoti, S. A., & Lima, J. O. (2006).

Comparing SOM neural network with fuzzy cmeans, K-means and traditional hierarchical clustering algorithms. European Journal of Operational Research, 174, 1742–1759.

Lhabitant, F.-S.: Portfolio Diversification. Elsevier, Saint Louis, available from: ProQuest Ebook Central. [1 March 2019] (2017)

Deguest, R., Martellini, L., Meucci, A.: Risk Parity and Beyond-From Asset Allocation to Risk Allocation Decisions. SSRN Working Paper (2013)

Carli, T., Deguest, R., Martellini, L.: Improved Risk Reporting with Factor-Based Diversification Measures. EDHEC-Risk Institute Publications (2014)

Sharma, P.: Improving portfolio diversification: identifying the right baskets for putting your eggs. Manag. Decis. Econ. 39, 698–711 (2018)

Statman, M., Scheid, J.: Correlation, return gaps and the benefits of diversification. J. Portf. Manag. 34, 132–139 (2008)

Statman, M.: Is Markowitz wrong? Investment lessons from the financial crisis. J. Portf. Manag. 40, 8–11 (2013)

Shin, H. W., & Sohn, S. Y. (2004). Segmentation of stock trading customers according to potential value. Expert Systems with Applications, 27, 27–33.

Journal of
Asia
Entrepreneurship
and Sustainability
www.asiaentrepreneurshipjournal.com

Shu, G., Zeng, B., Chen, Y. P., & Smith, O. H. (2003). Performance assessment of kernel density clustering for gene expression profile data. Comparative and Functional Genomics, 4(3), 287–299.
Tari, L., Baral, C., & Kim, S. (2009). Fuzzy c-means clustering with prior biological knowledge. Journal of Biomedical Informatics, 42, 74–81.
Topaloglou, N., Vladimirou, H., & Zenios, S. A. (2008). A dynamic stochastic programming model for international portfolio management. European Journal of Operational Research, 185, 1501–1524.
Xie, X. L., & Beni, G. A. (1991)Validity measure for fuzzy clustering. IEEE Transactions on Pattern Analysis and Machine Intelligence, 3(8), 841–846.
Fallahpour, S., Zadeh, M. H., & Lakvan, E. N. (2014). Use of clustering approach for portfolio management. International SAMANM Journal of Finance and Accounting, 2(1), 115-136.
Baser, P., & Saini, J. R. (2015). Agent based stock clustering for efficient portfolio management. International Journal of Computer Applications, 116(3), 36-41.
Yan, X., Yang, H., Yu, Z., & Zhang, S. Return-on-Equity Network Analysis for Portfolio Optimization: An Application to the Chinese Stock Market.
Barber, B. M., & Odean, T. (2000). Too many cooks spoil the profits: Investment club performance. Financial Analysts Journal, 56(1), 17-25.
Gasser, S. M., Rammerstorfer, M., & Weinmayer, K. (2017). Markowitz revisited: Social portfolio engineering. European Journal of Operational Research, 258(3), 1181-1190.
Irala, L. R., & Patil, P. (2007). Portfolio size and diversification. SCMS Journal of Indian Management, 4(1).

RossiSmith Academic Publications, Oxford/London UK, www.publicationsales.com