STM JOURNALS

Journal of
Artificial Intelligence
Research & Advances

ISSN: 2395-6720
Volume 11, Issue 1, 2024
DOI Journal: 10.37591/JoAIRA

http://computers.stmjournals.com/index.php?journal=JoAIRA&page=index

Review JoAIRA

Country Cluster Visualization based on Agricultural Imports: Unsupervised Learning Approach

R.S. Kamath[1,*], P.G. Naik[1], S.S. Jamsandekar[1]

Abstract

The clustering algorithm used in this analysis makes it easier for policymakers to understand performance metrics. K-means clustering has been demonstrated to be a tool for analyzing countries' agricultural imports and a visualization tool for interpreting the results in this study. Authors have used agricultural imports from 190 nations from Knoema, a web-based open data platform. Cereals, meat, and coffee imports from 190 nations in 2016 are included in the dataset. Based on their imports of cereals, meat, and coffee, the nations were divided into three clusters of sizes 24, 13, and 159. According to the findings of this study, 159 nations fall into the cluster category, with average imports of 6.64 million tonnes of cereals, 0.72 million tonnes of meat, and 1.1 million tonnes of coffee, respectively. This shows agrarian prosperity of these nations since the imports are less contrasted with the other two bunches. According to the findings, K-means has the potential to become the most widely used tool for country cluster analysis.

Keywords: K-means, cluster visualization, unsupervised learning, artificial intelligence, cluster designs

INTRODUCTION

Artificial Intelligence (AI) tools are speeding up data analysis process in all sectors and making them even more effective. AI and machine learning often leverage historical data to forecast and anticipate future outcomes (Han & Kamber, 2022) [10]. As import of agricultural products depend on several factors such as international and domestic demand & supply situation, prices, quality concerns and food security concern no targets for imports are fixed. A broad study on machine learning techniques has convinced the digital transformation in agriculture (Benos et al, 2021) [2]. The K-Means Clustering Method is used to discuss the application of data mining to rice imports by primary country of origin in this study (Supriyatna et al, 2020) [12]. The research findings can help ascertain the quantity of rice imported by the primary country of origin. Application of data mining methods illustrated with clustering customers of agriculture products (Chen et al, 2022) [4]. The findings demonstrated that clustering with three groups can lengthen a customer's lifespan and increase their purchase value.

In this context this paper visualizes 190 countries based on their imports of cereals, meat and coffee by applying K-means clustering approach. A comprehensive analysis has been offered by employing unsupervised learning to design prominent clusters among these countries. In this division, agricultural imports of 190 countries for the year 2016 are considered [1]. The authors aimed

***Author for Correspondence**
R.S. Kamath
E-mail: rskamath@siberindia.edu.in

[1]Associate Professor, Department of Computer Studies, Chhatrapati Shahu Institute of Business Education and Research (CSIBER), Kolhapur, Maharashtra, India
[2]Professor, Department of Computer Studies, Chhatrapati Shahu Institute of Business Education and Research (CSIBER), Kolhapur, Maharashtra, India
[3]Assistant Professor, Department of Computer Studies, Chhatrapati Shahu Institute of Business Education and Research (CSIBER), Kolhapur, Maharashtra, India

Received Date: November 17, 2023
Accepted Date: December 02, 2023
Published Date: January 03, 2024

Citation: R.S. Kamath, P.G. Naik, S.S. Jamsandekar. Country Cluster Visualization based on Agricultural Imports: Unsupervised Learning Approach. Journal of Artificial Intelligence Research & Advances. 2024; 11(1): 1–8p.

to extract latent insights from agricultural imports datasets and generate clusters of countries grounded on their individual values.

AGRICULTURAL IMPORTS DATA

In this comprehensive study, we used agricultural imports of 190 countries' from Knoema, a free to use web based open data platform. The dataset includes cereals, meat and coffee imports of 190 countries for the year 2016 (Knoema). These imports are by quantity in terms of Tonnes [3].

Figure 1 provides exploratory data analysis of the dataset designed for this study. Japan, China and The United States of America are the top countries by cereals, meat and coffee imports quantity in the world respectively. As reported in Knoema, the world's total cereals, meat and coffee imports quantity was estimated at 435 million tonnes, 44.7 million tonnes and 30.2 million tonnes in 2016 respectively.

```
    Cereals            Coffee            Meat
 Min.    :      26   Min.    :      1   Min.    :    115
 1st Qu.:  136871   1st Qu.:    475   1st Qu.:   7979
 Median :  591236   Median :   4596   Median :  41944
 Mean    : 2299421   Mean    : 161756   Mean    : 237693
 3rd Qu.: 2102560   3rd Qu.:  81498   3rd Qu.: 160609
 Max.    :23400454   Max.    :5746364   Max.    :3271819
```

Figure 1. Basic statistical summary of dataset

K-means Clustering for Agricultural Imports

We have carried out unsupervised classification for the visualization of countries based on their agricultural imports. The K-means procedure determines an assortment of k groups utilizing a heuristic search beginning with a determination of k randomly picked groups every one of which before all else addresses a cluster mean (Kamath et al, 2019) [11]. Further the cluster analysis depends on estimating likeness between data points by finding the distance between each pair. The likeness is estimated concerning the mean worth of the information things in a group.

The present experiment is simulated in R, an open source data mining platform. 190 countries selected for the study are clustered based on their quantity of cereals, meat and coffee imports. Each record Ni addressed as multidimensional vector and is characterized as:

$N_i = [Cereal_i, Meat_i, Coffee_i]$

Where i = 1 to 190

K-means technique requires the researcher to determine the number of groups to be generated. R code for this process is displayed in Figure 2. We plotted the "quantity of groups against" the "inside clusters sum of squares", which is the parameter should be limited during the unsupervised learning. Figure 3 illustrates "within clusters sum of squares" for various quantities of groups. Plot reveals that value of within-cluster sum of squares is less for 3 clusters as compared to 4.

The K-means clustering technique divides the data items into k partitions in such a way that the resulting within cluster similarity is more but the inter cluster similarity is less. The R function *kmeans()* is applied to get clusters of countries (Kamath & Kamat, 2019) [7]. The study is adjusted by altering the number of clusters while maintaining a constant number of iterations at 15. The results of this experiment are briefed in Table 1. Performance is assessed based on "within-cluster sum of square errors" and "BetweenSS/TotalSS". Preferably, a clustering method that exhibits both inside cohesion and outside separation would have a BSS/TSS ratio approaching 1 [8].


```
wss <- (nrow(import)-1)*sum(apply(import,2,var))

for (i in 2:10)
  wss[i] <- sum(kmeans(import, centers=i)$withinss)
  plot(1:10, wss, type="b", xlab="Number of Clusters",
                   ylab="Within groups sum of squares")
```

Figure 2. R code for determining number of clusters.

RESULTS AND DISCUSSIONS

Considering cluster performance explained in Table 1 and Figure 3, it has been decided to design three clusters of 190 countries. The Table shows that for 3 clusters BSS/TSS ration is 87% indicating a good fit (Badadare et al, 2022) [6]. However, the clusters size also taken into the consideration. K-means algorithm is applied with three clusters. The k-means clustering yielded three clusters with sizes 24, 13, and 153. Figure 4 displays the cluster means and clustering vectors obtained in the R environment. Within-cluster variation measures how much each event within a cluster differs from the others in the same cluster. (Jamsandekar et al., 2022) [6]. Figures 5 and 6 shows cluster plot for 1st and 2nd principle components and discriminant functions respectively. Plot reveals that cluster 3 is dense and some of the data points of cluster 1 are closer to cluster 3 [5, 9]. The matrix plot of countries clusters based on agricultural imports shown in Figure 7. The clusters and participating countries are tabulated in Table 2.

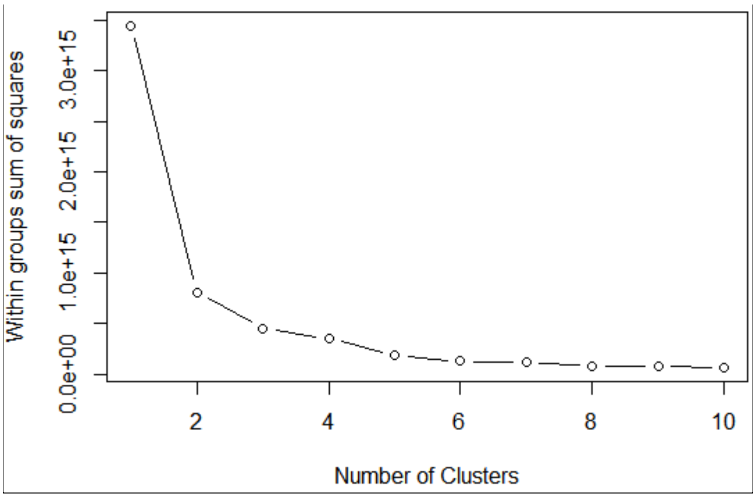

Figure 3. Number of clusters against within clusters sum of squares.

```
K-means clustering with 3 clusters of sizes 24, 13, 153

Cluster means:
     Cereals    Coffee      Meat
1  5666743.5 530449.79  494052.2
2 15541001.4 537434.62 1263993.7
3   646111.6  72001.05  110277.2

Clustering vector:
  1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18  19  20  21  22  23  24 
  3   3   3   3   2   1   3   2   3   3   3   3   3   3   3   3   3   3   3   3   3   3   3   3 
 25  26  27  28  29  30  31  32  33  34  35  36  37  38  39  40  41  42  43  44  45  46  47  48 
  3   3   3   3   3   3   3   3   3   3   3   3   3   3   3   3   3   3   3   3   3   1   3   1 
 49  50  51  52  53  54  55  56  57  58  59  60  61  62  63  64  65  66  67  68  69  70  71  72 
  1   3   3   3   3   2   3   3   3   1   3   1   3   3   3   3   3   3   3   3   3   2   3   2 
 73  74  75  76  77  78  79  80  81  82  83  84  85  86  87  88  89  90  91  92  93  94  95  96 
  3   3   3   3   3   3   3   3   3   3   3   3   3   3   3   3   3   3   3   3   3   3   3   3 
 97  98  99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 
  3   1   2   2   2   3   1   3   3   3   3   3   1   3   3   3   3   3   3   3   3   3   3   3 
121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 
  3   3   3   3   3   3   3   3   3   1   3   3   2   1   3   3   3   3   3   3   3   3   3   3 
145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 
  3   3   3   3   3   1   1   1   1   3   2   3   3   3   3   3   2   3   2   3   3   1   1   3 
169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 
  3   3   3   3   3   3   3   1   3   1   3   3   3   1   1   1   1   3   1   3   3   2 

Within cluster sum of squares by cluster:
[1] 153764252636126 195977230412473  96704266960079
 (between_SS / total_SS =  87.0 %)

Available components:

[1] "cluster"      "centers"      "totss"        "withinss"     "tot.withinss" "betweenss"   
[7] "size"         "iter"         "ifault"
```

Figure 4. Cluster means and cluster vector for countries clusters.

Table 1. Performance evaluation for accuracy of cluster designs.

No. of clusters	Cluster Size	Between_SS/Total_SS
2	17, 173	76.4%
3	24, 13, 153	87%
4	12, 28, 139, 11	89.8%
5	3, 13, 34, 18, 122	94.6 %

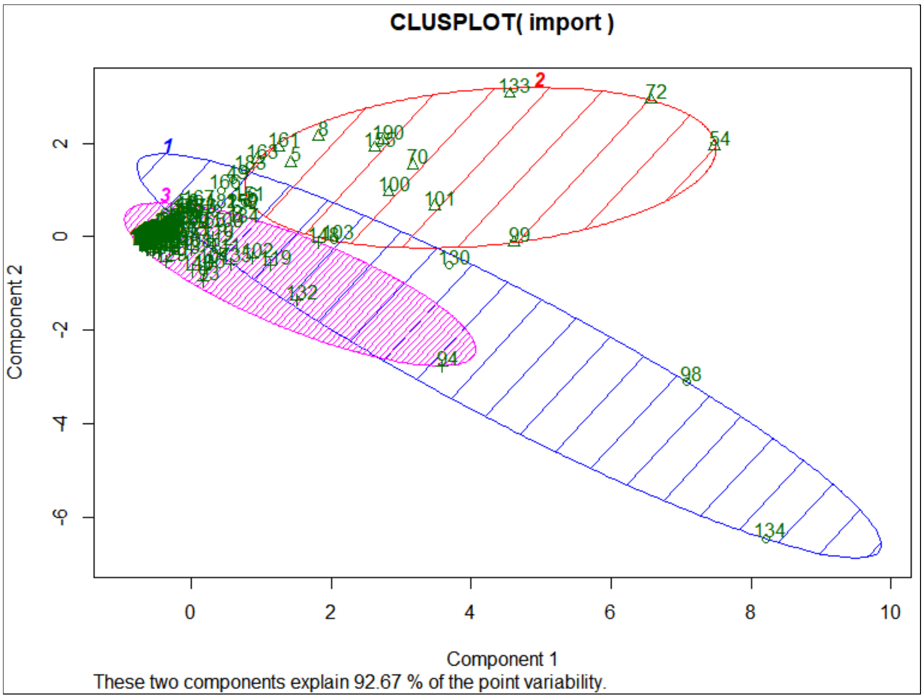

Figure 5. Cluster plot for 1st and 2nd principle components.

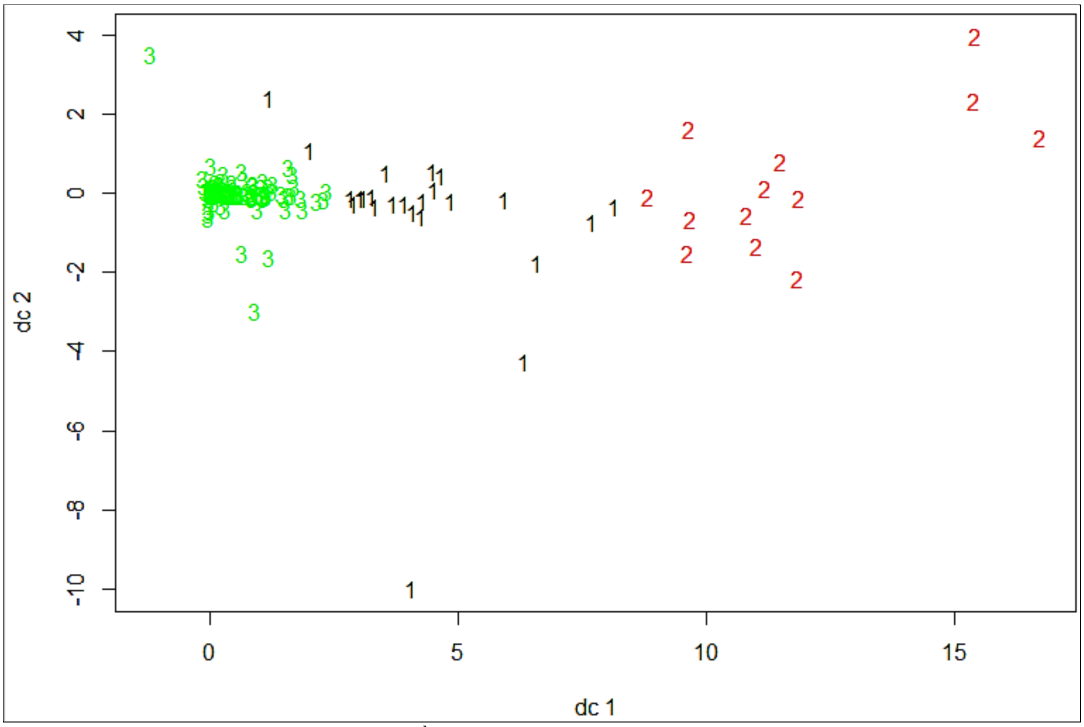

Figure 6. Cluster plot for 1st and 2nd discriminant functions.

Table 2. Clusters of countries based on agricultural imports.

Cluster 1	Cluster 2	Cluster 3		
South Africa	Algeria	Democratic Republic of the Congo	Armenia	Latvia
Tunisia	Egypt	Congo	Sri Lanka	Norway
Nigeria	Japan	Rwanda	Tajikistan	Russian Federation
Morocco	Republic of Korea	Ghana	India	Croatia
Bangladesh	China, mainland	Libya	Turkmenistan	Ukraine
Afghanistan	Italy	Guinea	Georgia	Republic of Moldova
Germany	Spain	Lesotho	Kyrgyzstan	The former Yugoslav Republic of Macedonia
Belgium	Netherlands	Angola	Uzbekistan	Albania
Portugal	Mexico	Central African Republic	Azerbaijan	Montenegro
United Kingdom	Viet Nam	Somalia	Democratic People's Republic of Korea	Bosnia and Herzegovina
United States of America	Indonesia	Burkina Faso	Maldives	Serbia
China, Taiwan Province of	Brazil	Gabon	Nepal	Iceland
Philippines	Saudi Arabia	Togo	Antigua and Barbuda	Finland
Malaysia		Sierra Leone	Barbados	Slovenia
Thailand		Liberia	Dominica	Canada
Colombia		Djibouti	Grenada	Oceania
Peru		Eritrea	Saint Kitts and Nevis	Micronesia
Turkey		Kenya	Saint Lucia	Kiribati
Yemen		Gambia	Saint Vincent and the Grenadines	Nauru

Iraq		Comoros	Cuba	Tonga
Iran (Islamic Republic of)		Cabo Verde	Guatemala	Tuvalu
United Arab Emirates		Côte d'Ivoire	Dominican Republic	Vanuatu
Israel		Cameroon	Jamaica	Samoa
Jordan		Burundi	Costa Rica	Australia
		Mali	Panama	New Zealand
		Chad	Honduras	Papua New Guinea
		Benin	Nicaragua	Solomon Islands
		Swaziland	Belize	Fiji
		Equatorial Guinea	El Salvador	China, Hong Kong SAR
		Zimbabwe	Trinidad and Tobago	China, Macao SAR
		Sudan	Bahamas	Singapore
		Malawi	Haiti	Lao People's Democratic Republic
		Mozambique	Switzerland	Myanmar
		Niger	France	Timor-Leste
		Guinea-Bissau	Luxembourg	Brunei Darussalam
		United Republic of Tanzania	Malta	Cambodia
		Senegal	Belarus	Venezuela (Bolivarian Republic of)
		Uganda	Poland	Argentina
		Zambia	Bulgaria	Chile
		Madagascar	Hungary	Ecuador
		Mauritius	Sweden	Paraguay
		Seychelles	Czechia	Uruguay
		Botswana	Denmark	Falkland Islands (Malvinas)
		Sao Tome and Principe	Romania	Bolivia (Plurinational State of)
		Mauritania	Austria	Guyana
		Namibia	Greece	Suriname
		Kazakhstan	Slovakia	Cyprus
		Mongolia	Estonia	Bahrain
		Pakistan Syrian Arab Republic	Ireland	Kuwait
			Lithuania Lebanon	Oman
				Qatar
				Occupied Palestinian Territory

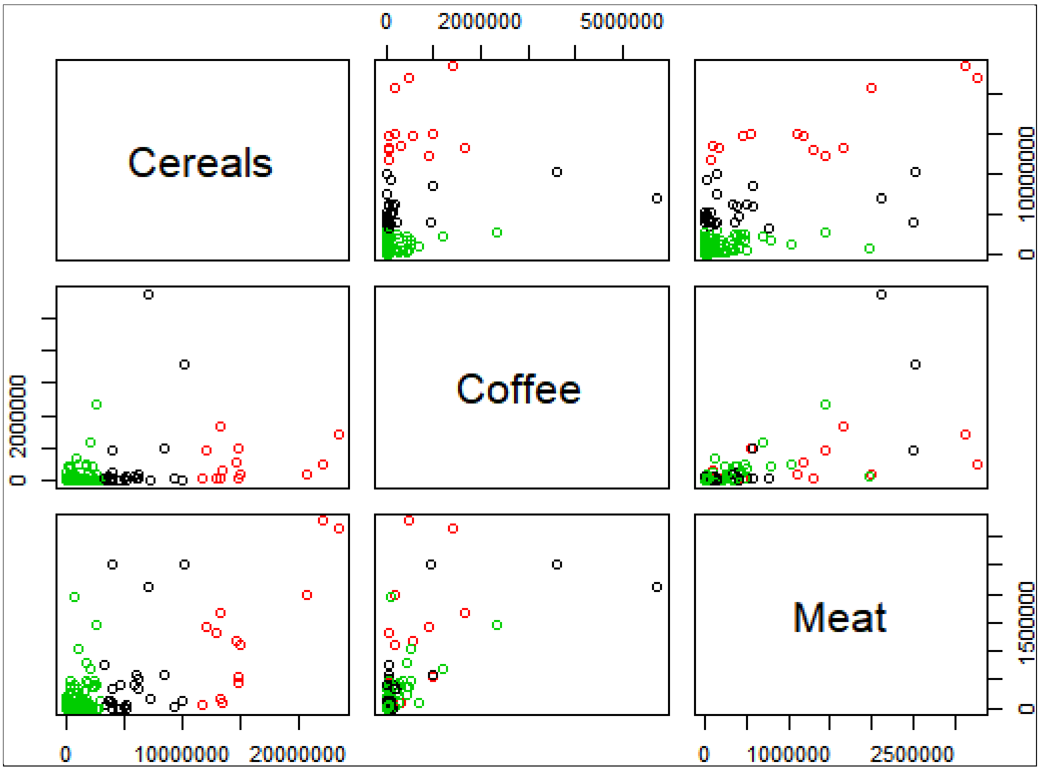

Figure 7. Matrix plot of country clusters.

Upon studying the cluster means, we can establish correlations between each group and the three classes of countries:

- Cluster 2 has highest value for imports of cereals, meat and coffee. There are 13 countries in this cluster.
- Cluster 1 is average for cereals and coffee imports and higher for meat imports. There are 24 countries in this cluster.
- The values for all the three agricultural imports are lowest for cluster 3 where 159 countries falls under this category. This shows agricultural wellbeing of these countries since the imports are less.

CONCLUSION

This research of employing unsupervised learning, offers insights into performance metrics crucial for policymakers. This study has demonstrated the K-means clustering as tool for analysing agricultural imports of countries together with visualization for interpreting the results. The countries grouped into three clusters of sizes 24, 13 and 159 based on imports of cereals, meat and coffee. This study concluded that 159 countries falls under the category of cluster with mean imports 6.64Lakh tonnes, 0.72Lakh tonnes and 1.1Lakh tonnes for cereals, meat and coffee respectively. This shows agricultural wellbeing of these countries since the imports are less compared to the rest of the two clusters. The findings indicate that K-means holds promise as a primary tool for conducting cluster analysis of countries grounded on their agricultural imports.

REFERENCES

1. Altuntas, S., Selim, S., and Altuntas, F. (2022). A hierarchical clustering based panel data approach: A case study of regional incentives. *International Journal of Information Management Data Insights*, *2*(2), 100098. https://doi.org/10.1016/j.jjimei.2022.100098
2. Benos, L., Tagarakis, A.C., Dolias, G., Berruto, R., Kateris, D. and Bochtis, D. (2021). Machine learning in agriculture: A comprehensive updated review. *Sensors*, 21(11), p. 3758. Available at: https://doi.org/10.3390/s21113758.

3. Cereals imports quantity by country, 2022. Knoema. https://knoema.com/atlas/topics/Agriculture/Trade-Import-Quantity/Cereals-imports-quantity (Accessed: March 14, 2023).
4. Chen, T., Alazzawi, F.J., Mavaluru, D., Mahmudiono, T., Enina, Y., Chupradit, S., Ahmed, A.A., Syed, M.H., Ismael, A.M., and Miethlich, B. (2022). Application of Data Mining Methods in Grouping Agricultural Product Customers. *Mathematical Problems in Engineering*, 2022. https://doi.org/10.1155/2022/3942374
5. Coffee imports quantity by country, 2022. *Knoema*. https://knoema.com/atlas/topics/Agriculture/Trade-Import-Quantity/Coffee-imports-quantity (Accessed: March 14, 2023).
6. Badadare, V.L., Jamsandekar, S.S., Kamath, R.S. (2022). Insights of Exploratory Data Analysis (EDA) of Covid-19 Vaccine Administration in India, *South Asian Journal of Management Research (SAJMR)*, 12(1), pp. 36-44.
7. Kamath, R.S. and Kamat, R.K. (2019). Visualization of University Clusters based on NIRF and NAAC Scores: K-means Algorithm Approach. *University News*, 57(03).
8. Jamsandekar, S. S., Kamath, R.S., Patil, M.B. (2022) Time series and cluster analysis of Covid-19 growth curve in India, *South Asian Journal of Management Research (SAJMR)*, 12(1), pp. 45-53.
9. Meat imports quantity by country, 2022. *Knoema*. https://knoema.com/atlas/topics/Agriculture/Trade-Import-Quantity/Meat-imports-quantity (Accessed: March 14, 2023).
10. Han, J., & Kamber, M. (2022). *Data Mining: Concepts and Techniques*. 4th ed. Elsevier.
11. Kamath, R.S., Jamsandekar, S.S., Kharade, K.G., Kamat, R.K. (2019). *Data Analytics in R: A Case Study Based Approach.* Himalaya Publishing House Pvt. Ltd.
12. Supriyatna, A., Carolina, I., Widiati, W., Nuraeni, C. (2020). Rice Productivity Analysis by Province Using K-Means Cluster Algorithm. *IOP Conf. Series: Materials Science and Engineering.* 771 (2020) 012025.

