Table 4: Number of Countries for each Cluster in Socio-Economic and Health Datasets

Techniques/clusters	Without risk	Low risk	Medium risk	High risk
k-means++	31	61	2	73
Farthest First	7	92	28	40
Hybrid	**32**	**46**	**27**	**62**

5. Conclusion

In this paper, a hybrid technique based on K-mean algorithm and farthest first algorithm is proposed to cluster the countries with HELP International socio-economic and health datasets. For this purpose, a set of data is obtained from Kaggle.com website. K-means++ and farthest first are utilized to categorize the countries based on the collected data. The experiments show that the hybrid performs better than K-means++ and the farthest first for clustering countries separately. The proposed hybrid approach significantly improves the clustering by minimizing the risks related to socio-economic and health data for each country.

6. References

[1] Help international organiztion, https://www.help-international.com/, Accessed 2020,

[2] MacQueen, J. B. (1967). Some Methods for classification and Analysis of Multivariate Observations. Proceedings of 5th Berkeley Symposium on Mathematical Statistics and Probability. 1. University of California Press. pp. 281–297

[3] Gonzalez, T. F. (1985), "Clustering to minimize the maximum intercluster distance", Theoretical Computer Science, 38 (2–3): 293–306

[4] Rao Muzamal Liaqat, Bilal Mehboobb, Nazar Abbas Saqibc, and Muazzam A Khand, A "Framework for Clustering Cardiac Patient's Records Using Unsupervised Learning Techniques", In Proceeding of the 6th International Conference on Current and Future Trends of Information and Communication Technologies in Healthcare (ICTH 2016).

[5] Adam Andrzejuk, "Classification of Agricultural Emissions Among OECD Countries with Unsupervised Techniques", Problems of World Agriculture, volume 18 (XXXIII), number 4, 2018: 80–91.

[6] Rodrigo M. Carrillo-Larco and Manuel Castillo-Cara, "Using country-level variables to classify countries according to the number of confirmed COVID-19 cases: An unsupervised machine learning approach", Wellcome Open Research, 2020.

[7] Ahmed Anwar and Ussama Yaqub, "Bot detection in twitter landscape using unsupervised learning", In Proceeding of the 21st Annual International Conference on Digital Government Research, June 2020.

[8] R. S. Kamath and R. K. Kamat, "K-MEANS CLUSTERING FOR ANALYZING PRODUCTIVITY IN LIGHT OF R & D SPILLOVER", International Journal of Information Technology, Modeling and Computing (IJITMC) Vol. 4, No.2, May 2016.

[9] Aiyshwariya Paulvannan Kanmani, Renee Obringer, Benjamin Rachunok and Roshanak Nateghi, "Assessing Global Environmental Sustainability Via an Unsupervised Clustering Framework", Sustainability 2020, vol. 12(2).

[10] P.Viveka, F. Kurus Malai Selvi, "A Modified K- Means Algorithm for Big Data Clustering", ADALYA JOURNAL, Volume 9, Issue 2,pp.507-512, 2020.

[11] Weipeng Wang, Shanshan TU and Xinyi Huang," IKM-NCS: A Novel Clustering Scheme Based on Improved K-Means Algorithm", Engineering World Journal, volume 1, pp.103-108, 2019.

[12] Wisan Tangwongcharoen, "Comparison of Methods for Analyzing Shoulder Blades", International Journal of Simulation Systems, Science & Technology ,(IJSSST), Volume 21, Number 3, pp. 3.1-3.8, 2020.

[13] Sungjin Im, Mahshid Montazer Qaem, Benjamin Moseley, Xiaorui Sun, Rudy Zhou," Fast Noise Removal for k-Means Clustering", Proceedings of the 23rdInternational Conference on Artificial Intelligence and Statistics (AISTATS) 2020, Palermo, Italy.

[14] Mirpouya Mirmozaffari, Azam Boskabadi, Gohar Azeem, Reza Massah, Elahe Boskabadi, Hamidreza Ahady Dolatsara, and Ata Liravian," A comparison of Machine learning Farthest First and Expectation Maximization Clustering Algorithms Based on the DEA Cross-efficiency Optimization Approach for Banking System", European Journal of Engineering Research and Science(EJERS), Vol 5, Issue 6, pp. 651-658,2020.

[15] Vivek, S. (2018). "Clustering algorithms for customer segmentation", https://towardsdatascience.com/clusteringalgorithms- for-customer-segmentation-af637c6830ac. (Access date: 19.09.2018).

[16] K. Mumtaz1 and Dr. K. Duraiswamy "A Novel Density based improved k-means Clustering Algorithm – Dbk-means" International Journal of Computer Science and Engineering ISSN: 0975-3397 213, Vol. 02, No. 02, 2010.

[17] Sharmila and Mukesh Kumar, "An Optimized Farthest First Clustering Algorithm", In Proceeding of 2013 Nirma University International Conference on Engineering (NUiCONE), 2013.

[18] Heba Ayeldeen, Mahmood A Mahmood, Aboul Ella Hassanien," Effective Classification and Categorization for Categorical Sets: Distance Similarity Measures", Information Systems Design and Intelligent Applications, Advances in Intelligent Systems and Computing, pp. 359-368, springer,2015. DOI 10.1007/978-81-322-2250-7_36.

[19] David Arthur and Sergei Vassilvitskii. 2007. "K-means++: the advantages of careful seeding". In Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms (SODA '07). Society for Industrial and Applied Mathematics, USA, pp. 1027–1035, 2007.

[20] [kaggle.com dataset] https://www.kaggle.com/rohan0301/unsupervised-learning-on-country-data, Accessed, july-2020.