An Approach for Finding Anthropogenic Prediction Using Novel Cluster Analysis Technique over Exploratory Approach Based on Accuracy

S. Sumanth and N. Deepa
Saveetha University, Chennai, India

Keywords: Artificial Intelligence, Civilization, Novel Cluster Analysis Technique, Human Intervention, Machine Learning, Resilience.

Abstract: The objective of this research is to make a specialty of anthropogenic prediction by way of using a better cluster evaluation approach on a dataset by means of evaluating with the E exploratory Approach. Materials and using an exploratory approach method with a sample size of 27 and a novel cluster analysis technique with a sample size of 27, respectively, and a G power of 80%, an accurate anthropogenic prediction was made. Results: The accuracy of the Innovative Cluster Analysis Method is 77.26%, which is somewhat more than the accuracy of the Exploratory Approach, which is 73.07%. With p=0.294 (p>0.05), it demonstrates there is no statistically significant difference comparing New Cluster Analysis Method as well as Exploratory technique. Conclusion: When compared to the subsets of variables in the marginal distribution and the exploratory approach, which has an accuracy of 73.07%, novel cluster analysis technique has a greater accuracy of 77.26%.

1 INTRODUCTION

A fundamental aspect of artificial intelligence is machine learning, which equips systems with the capability to perform tasks and execute functions without explicit programming. This facet of AI focuses on the utilization of computers, as highlighted by Hottung and Tierney (2022), to access data and execute tasks autonomously. Much like the human mind accrues expertise and knowledge, machine learning, as elucidated by Rouhanizadeh, Kermanshachi, and Nipa (2020), relies on input such as educational records and knowledge graphs to comprehend entities, domains, and the relationships between them. As entities define civilization, the application of deep learning can contribute to enhancing urbanization.

The machine learning process initiates with observations and resilience from challenges, encompassing examples, direct experiences, or training. It seeks patterns within data to subsequently make inferences based on the provided examples. The primary objective of machine learning, as emphasized by Kopec, Shetty, and Pileggi (2014) and G. Ramkumar et al. (2022), is to empower computers to autonomously learn without human intervention, enabling them to adapt and perform tasks independently. This autonomy is especially beneficial for computers in executing programs without requiring human involvement, as noted by Qochuk, Sharma, and Chatterjee (2020) and Padma, S et al. (2022).

Numerous research papers in the field of machine learning are frequently published in IEEE and Science Direct. The IEEE Xplore digital library, for instance, houses 5877 journals, while Science Direct boasts 3455 articles. Additional citations are available in 6203 papers from Springer and 7990 articles from Google Scholar.

The issue of environmental pollutants is closely tied to urbanization and commercial trends across the entire spectrum of human intervention. Air pollutants, highlighted by Sinnott and Guan (2018), are identified as primary concerns in metropolitan regions globally, exacerbated by increased urbanization and societal development. Notably, Tehran, the capital of Iran, grapples with air pollution challenges, affecting both the well-being of its residents and the city's resilience (Sinnott and Guan, 2018). Delavar et al. (2019) emphasize that a significant portion of Tehran's air pollutants is attributed to PM10 and PM2.5 pollution.

Sumanth, S. and Deepa, N.
An Approach for Finding Anthropogenic Prediction Using Novel Cluster Analysis Technique over Exploratory Approach Based on Accuracy.
DOI: 10.5220/0012567000003739

In *Proceedings of the 1st International Conference on Artificial Intelligence for Internet of Things: Accelerating Innovation in Industry and Consumer Electronics (AI4IoT 2023)*, pages 96-102
ISBN: 978-989-758-661-3

Metallic performance, as studied by Mele and Magazzino (2020), holds a prominent position as one of the most widely used resources in human intervention. This prominence traces back to the industrial revolution, coinciding with the advent of machine learning techniques and the establishment of large iron and steel production units (Nicodemi, 1994). The material's resilience activities are facilitated by its high resistance to stress at elevated temperatures, atmospheric and corrosive agents, as well as its ductility and ability to undergo plastic deformation. This adaptability aligns with the technological advancements in civilization.

In the realm of cluster analysis strategies, existing conventional algorithms exhibit low precision and sensitivity. Cluster analysis, a statistical method organizing similar items into respective categories, serves as a fundamental approach in human intervention (Zheng, 2022). This method aims to group devices based on chosen traits and attributes of civilization. For exploratory analysis, it is imperative that data sets do not contain redundancies, missing values, or null values (Cipresso et al., 2018). The primary objectives of exploratory analysis include identifying faulty data points to facilitate their removal, thus ensuring data cleanliness. Moreover, it aids in understanding the relationships between variables, offering a comprehensive perspective on the data.

The Novel Cluster Analysis Technique, in comparison to the exploratory approach, specifically targets identifying air pollution levels in the atmosphere, contributing to a more refined understanding of environmental conditions.

2 MATERIALS AND METHODS

The study takes place in the data analytics lab of the SIMATS School of Engineering, which is well-equipped with advanced facilities for both output production and research purposes. Two assessments of assembly numbers were conducted with a sample size of 27 (Williams et al., 2019). The calculations are secured using 80% G power, a 95% confidence level, an alpha value of 0.05, and a beta value of 0.2, implemented through clinical software.

The dataset provides information on the condition of contaminated states and their corresponding air quality levels. Analyzing particulate matter levels in the dataset simplifies the determination of air quality (Borbet, Gladson, and Cromar, 2018).

Cluster evaluation is a method used in gadget studying that attempts to locate clusters of observations within a dataset. The intention of cluster analysis is to find clusters such that the observations within each cluster are quite similar to every different, whilst observations in exclusive clusters are quite specific from each other.

Table 1: Shows procedure for proposed algorithm.

Input: Assign the clusters present in the data set.
Output: Accuracy of the data set's existing clusters.
Step 1: First need to define some clusters and assign variable k. Step 2: Choose at random a few clusters for computation from those defining clusters. Step 3: After computing the clusters, assign data points to each cluster. Step 4: Now define squared difference between information factors of the clusters Step 5: Now assign statistical variables to each cluster present in the data. Step 6: Find mean of all clusters present in the data set after computing the data set.

Table 2: Exploratory Approach is coded as procedure.

Input: Give insights to the data set
Output: Accuracy of the optimal factors present in the data set.
Step 1: Maximum insights should be obtained from the data set. Step 2: Uncover the unlayered structure present in the data set. Step 3: Important variables present in the data set need to be extracted. Step 4: Find out if any outliner should be identified inside the data set. Step 5: Underlying assumptions should be tested in this stage. Step 6: Finally have to determine the optimal factors present in the data set.

Table 3: Displays the accuracy raw data table for both the exploratory approach and the novel cluster analysis technique.

Sl.no	Novel Cluster Analysis Technique Accuracy (%)	Exploratory Approach Accuracy (%)
1	50	98
2	52	97
3	55	94
4	57	92
5	59	89
6	61	86
7	64	85
8	66	83
9	69	80
10	71	79
11	73	77
12	75	76
13	78	74
14	81	72
15	82	70
16	84	69
17	85	68
18	86	67
19	87	65
20	89	63
21	90	60
22	92	59
23	93	57
24	95	56
25	96	53
26	97	51
27	99	50

Table 4: Displays the N (27), Mean (73.07%), Std.Deviation (13.862), Std.Error Mean values (2.668) and N (27), Mean (74.26%), Std.Deviation (15.096), Std.Error Mean values (2.905) for Innovative Cluster Analysis Method.

	Algorithm	N	Mean	Std.Deviation	Std.Error Mean
Accuracy	Novel Cluster Analysis Technique	27	77.26	15.096	2.905
	Exploratory Approach	27	73.07	13.862	2.668

Google Colab, akin to the Jupyter notebook, offers a convenient platform for program execution. Its browser-based interface allows users to write and execute arbitrary Python code, making it particularly useful for data analysis. This feature facilitates code sharing, enabling others to easily run the provided code snippets (Borbet, Gladson, and Cromar, 2018).

Novel Cluster Analysis Technique

Group 1 in sample preparation utilizes an innovative cluster analysis technique, treating a cluster of statistical objects as a singular group. Before assigning labels to organizations in cluster analysis, it is imperative to initially divide the data collection into groups based on information similarity for robustness. The primary advantage of category-based clustering lies in its adaptability, enabling the differentiation of valuable services that distinguish various businesses.

Clustering analysis finds extensive applications in various fields, including market research, resilience, pattern recognition, data analysis, and image processing. Additionally, it assists entrepreneurs in identifying human intervention and understanding the distinct groups within their customer base due to increased urbanization. Entrepreneurs can effectively categorize customer groups based on their purchasing habits. In the realm of biology, clustering analysis proves useful for classifying genes with similar functionalities, providing insights into population structures. It also aids in deriving taxonomies for plants and animals through artificial intelligence. In databases containing Earth statements, clustering facilitates the discovery of areas with similar land uses. Employing machine learning, it contributes to the recognition of property groups in a city based on housing type, cost, and geographic location.

Table 1 outlines the Novel Cluster Analysis Technique procedure.

2.1 Exploratory Analysis

Exploratory analysis is sample preparation group 2 and analyzes data with the help of visual techniques, through this it can discover patterns or check assumptions with the help of graphical representation and statistical representation. Exploratory analysis' principal goal is to examine the data before adopting any civilizational assumptions. It can aid in locating glaring faults, in addition to better understanding data trends, locating outliers, and locating intriguing relationships between the variables. Exploratory analysis consists of data of the improved urbanization. It is divided into three parts: data

summarization, data visualization, and data normalization is a part of artificial intelligence. The data summarization is when summarizing the essence of the particular dataset using certain key statistical measures such as mean and standard deviation of the human intervention. Data visualization as the name suggests is going to be visually exploring the data by plotting various types of graphs such as histograms and bar plots. The data normalization is to adjust the scales of the data. Table 2 represents the procedure for exploratory analysis.

2.2 Statistical Analysis

The following statistics were calculated using IBM SPSS version 26.0: standard deviation mean, standard error mean, mean difference, significance, as well as F value. The amount of particle matter and the air quality are independent variables. Deforestation is a dependent variable.et al. 2021).

3 RESULTS

The procedure for the New Cluster Analysis Method is displayed in Table 1. The number of clusters that are present in the data collection must first be specified. After discovering the data set, it is important to concentrate on computing the clusters and determining the average of all the clusters as a sign of some pointers to the data set. This ought to aid in determining correctness. present in the data set that contains this data. In order to locate the common points across all the clusters, the dataset is first partitioned into two sets of models. These models are then assigned to various functions in order to calculate accuracy.

The procedure for the exploratory approach is shown in Figure 2. Get the data first, then look for any interesting findings. Discover the data set's unlayered structure after data collection. The data for the unlayered structure should be shown in two sets. This should be used to validate the described structures and determine accuracy.

Table 3 displays the accuracy raw data table for both the exploratory approach and the novel cluster analysis technique.

Table 4 displays the N (27), Mean (73.07%), Std.Deviation (13.862), Std.Error Mean values (2.668) and N (27), Mean (74.26%), Std.Deviation (15.096), Std.Error Mean values (2.905) for Innovative Cluster Analysis Method.

Table 5 represents the statistical independent sample T-Test values. Mean difference for the New Cluster Analysis Method and Exploratory Approach is 4.185, Standard Error difference is 3.944, and 95% Confidence Interval, respectively. With $p=0.294$ ($p>0.05$), it demonstrates that there is no statistically significant difference between the New Cluster Analysis and Exploratory method.

By looking at Figure 1, you can compare the mean accuracy of two algorithms—the exploratory approach and the novel cluster analysis technique. Using a bar graph, the accuracy of both lost and gained value is determined. When compared to the exploratory approach, which has an accuracy of 73.07%, the Innovative Cluster Analysis Technique has an accuracy of 77.26%.

4 DISCUSSION

The relevance of the air quality stages is determined based on the results obtained using the enhanced Cluster evaluation approach. The significance rate is calculated entirely based on the outcomes attained using unbiased factors. The accuracy of 77.26% is exceeded by the significance value of 0.520 ($p>0.05$) so there is no significance difference between them. The accuracy of the New Cluster Analysis Method is better than 73.07% for the exploratory approach. The basic concept of anthropogenic prediction is to find the quality of the air. This is the method to analyze the data (Rouhanizadeh, Kermanshachi, and Nipa 2020) with the help of visual techniques. unfavorable herbal disasters, have an effect on nations round the world. To recover from a catastrophe, the managers need to make diverse decisions using artificial intelligence; however, what makes the submit-disaster selection precise is lack of time to make the optimal civilization decision. A complete expertise of obstacles to recuperation can result in development of rules that help prevent delays in healing technique and therefore (Rouhanizadeh and Kermanshachi 2021; M R et al. 2019) effects in resiliency. Although numerous studies had been achieved to become aware of the healing obstacles, they did no longer offer a complete overview of limitations and their classification based on their techniques and also by the improved urbanization. This paper aimed to identify as well as categorize boundaries to effective submit-disaster recuperation with a focal point on hurricanes. The (M R et al. 2019) Researchers, scientists, and educators are becoming more interested in plant-based fibers that have been extracted because of their application in polymer composites, as well as their long-term viability in machine learning and environmental friendliness.

Table 5: Shows the statistical independent sample T-Test values. Mean difference for the New Cluster Analysis Method and Exploratory Approach is 4.185, Standard Error difference is 3.944, and 95% Confidence Interval, respectively. With p=0.294 (p>0.05), it demonstrates that there is no statistically significant difference between the New Cluster Analysis as well as Exploratory method.

		F	sig	t	dt	Sig (2-tailed)	Mean difference	Std.Error Difference	Lower	Upper
Accuracy	Equal variables assumed	.41	.52	1.06	52	.29	4.18	3.94	-3.73	12.10
	Equal variables not assumed			1.06	51.62	.29	4.18	3.94	-3.73	12.10

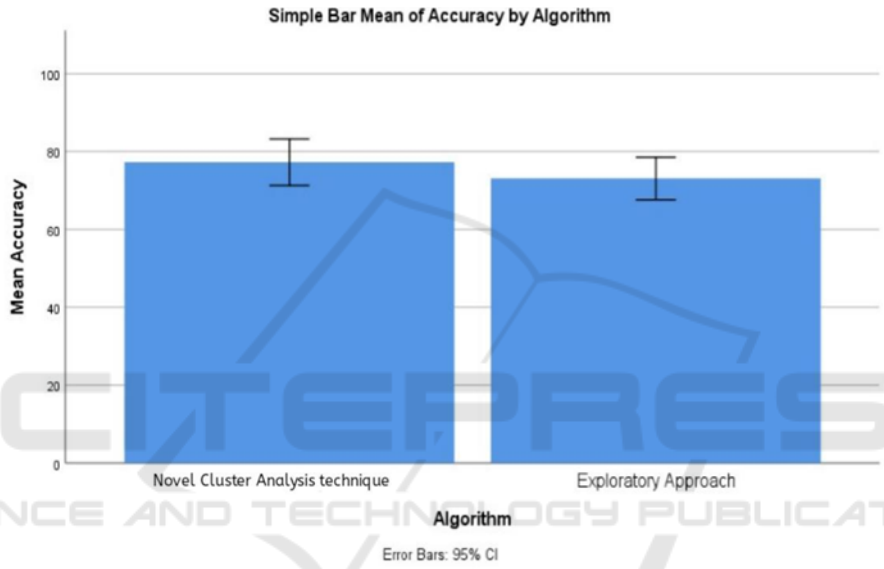

Figure 1: Comparison among New Cluster Analysis Method (77.26%) as well as Explorative Approach (73.07%) depending on accuracy. New Cluster Analysis Method has a higher average accuracy than Exploratory Approach. New Cluster Analysis Method vs Exploratory Approach on X axis, Mean Accuracy on Y axis. +/-2 SE Error Bar.

The learning and extraction of the fibers affect the characteristics of herbal fibers. Mechanical retting, dew retting, or water retting techniques are used to remove natural plant fibers. Mainly (Vardhan, Kumar, and Panda 2019) For life to exist on earth, water is an incredibly important supply. The overpopulation, human activities, rapid industrialisation, unskilled exploitation of natural water resources, and unplanned urbanization have all had a negative impact on the water quality. According to Kaushik et al. (2022), "heavy metals" are a class of metals and metalloids with atomic densities greater than 4000 kg/m3. At their extremely low concentrations, heavy metals are poisonous by nature and can seriously harm both humans and animals. This is relevant to artificial intelligence. Runoff from agriculture and industrial discharges are two ways that these heavy metals get into the aquatic system, according to machine learning. The use of antibiotics to treat bacterial infections has long been a tenet of contemporary medicine (Kraemer, Ramachandran, and Perron 2019). Antibiotic abuse and widespread overuse, however, have resulted in unexpected consequences that call for significant adjustments in policy to be mitigated. Two major categories of antibiotic misuse and overuse corollaries are discussed in this review. The factors affecting the anthropogenic conditions are a main issue of anthropogenic climate change is international warming, which refers to a sluggish warming of the earth caused by an unnatural (human-precipitated) growth of the greenhouse impact by the help of the artificial intelligence, as concentrations of greenhouse gasses growth generally from the burning of fossil fuels (coal, oil, and herbal fuel). The limitations of anthropogenic prediction are

deforestation, rising of earth atmosphere based on machine learning. The future work of anthropogenic prediction is used to find out the polluted air present in the atmosphere.

5 CONCLUSION

Anthropogenic prediction makes determining the amount of air pollution in the atmosphere easier. The accuracy rate of the Novel Cluster Analysis Technique (77.26%) is significantly higher than the Exploratory Approach (73.07%) while using the Novel Cluster Analysis Technique and Exploratory Approach data sets, as well as when comparing the two techniques.

REFERENCES

Borbet, Timothy C., Laura A. Gladson, and Kevin R. Cromar. (2018). "Assessing Air Quality Index Awareness and Use in Mexico City." *BMC Public Health* 18 (1): 538.

Cipresso, Pietro, Irene Alice Chicchi Giglioli, Mariano Alcañiz Raya, and Giuseppe Riva. (2018). "The Past, Present, and Future of Virtual and Augmented Reality Research: A Network and Cluster Analysis of the Literature." *Frontiers in Psychology* 9 (November): 2086.

Delavar, Mahmoud, Amin Gholami, Gholam Shiran, Yousef Rashidi, Gholam Nakhaeizadeh, Kurt Fedra, and Smaeil Hatefi Afshar. (2019). "A Novel Method for Improving Air Pollution Prediction Based on Machine Learning Approaches: A Case Study Applied to the Capital City of Tehran." *ISPRS International Journal of Geo-Information*. https://doi.org/10.3390/ijgi8020099.

G. Ramkumar, G. Anitha, P. Nirmala, S. Ramesh and M. Tamilselvi, (2022) "An Effective Copyright Management Principle using Intelligent Wavelet Transformation based Water marking Scheme," International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI), Chennai, India, 2022, pp. 1-7, doi: 10.1109/ACCAI53970.2022.9752516.

Hottung, André, and Kevin Tierney. (2022). "Neural Large Neighborhood Search for Routing Problems." *Artificial Intelligence*. https://doi.org/10.1016/j.artint.2022.103786.

Kaushik, Pallavi, Renu Khandelwal, Neha Rawat, and Mukesh Kumar Sharma. (2022). "Environmental Hazards of Heavy Metal Pollution and Toxicity: A Review." *FLORA AND FAUNA*. https://doi.org/10.33451/florafauna.v28i2pp315-327.

Kopec, Danny, Shweta Shetty, and Christopher Pileggi. 2014. *Artificial Intelligence Problems and Their Solutions*. Mercury Learning and Information.

Kraemer, Susanne A., Arthi Ramachandran, and Gabriel G. Perron. (2019). "Antibiotic Pollution in the Environment: From Microbial Ecology to Public Policy." *Microorganisms* 7 (6). https://doi.org/10.3390/microorganisms7060180.

Lavrentieva, Olena O., Ihor O. Arkhypov, Olexander I. Kuchma, and Aleksandr D. Uchitel. (2020). "Use of Simulators Together with Virtual and Augmented Reality in the System of Welders' Vocational Training: Past, Present, and Future." https://doi.org/10.31812/123456789/3748.

Mele, Marco, and Cosimo Magazzino. (2020). "A Machine Learning Analysis of the Relationship among Iron and Steel Industries, Air Pollution, and Economic Growth in China." *Journal of Cleaner Production*. https://doi.org/10.1016/j.jclepro.2020.123293.

M R, Sanjay, Suchart Siengchin, Jyotishkumar Parameswaranpillai, Mohammad Jawaid, Catalin Iulian Pruncu, and Anish Khan. (2019). "A Comprehensive Review of Techniques for Natural Fibers as Reinforcement in Composites: Preparation, Processing and Characterization." *Carbohydrate Polymers* 207 (March): 108–21.

Qochuk, Benjamin, Priyanshi Sharma, and Aditya Chatterjee. 2020. *Problems with AI (Artificial Intelligence)*.

Padma, S., Vidhya Lakshmi, S., Prakash, R., Srividhya, S., Sivakumar, A. A., Divyah, N., ... & Saavedra Flores, E. I. (2022). Simulation of land use/land cover dynamics using Google Earth data and QGIS: a case study on outer ring road, Southern India. Sustainability, 14(24), 16373

Rouhanizadeh, Behzad, and Sharareh Kermanshachi. (2021). "Barriers to an Effective Post-Recovery Process: A Comparative Analysis of the Public's and Experts' Perspectives." *International Journal of Disaster Risk Reduction*. https://doi.org/10.1016/j.ijdrr.2021.102181.

Rouhanizadeh, Behzad, Sharareh Kermanshachi, and Thahomina Jahan Nipa. (2020). "Exploratory Analysis of Barriers to Effective Post-Disaster Recovery." *International Journal of Disaster Risk Reduction*. https://doi.org/10.1016/j.ijdrr.2020.101735.

Sinnott, Richard O., and Ziyue Guan. (2018). "Prediction of Air Pollution through Machine Learning Approaches on the Cloud." *2018 IEEE/ACM 5th International Conference on Big Data Computing Applications and Technologies (BDCAT)*. https://doi.org/10.1109/bdcat.2018.00015.

Vardhan, Kilaru Harsha, Ponnusamy Senthil Kumar, and Rames C. Panda. (2019). "A Review on Heavy Metal Pollution, Toxicity and Remedial Measures: Current Trends and Future Perspectives." *Journal of Molecular Liquids*. https://doi.org/10.1016/j.molliq.2019.111197.

Williams, A. Park, A. Park Williams, John T. Abatzoglou, Alexander Gershunov, Janin Guzman-Morales, Daniel A. Bishop, Jennifer K. Balch, and Dennis P.

Lettenmaier. (2019). “Observed Impacts of Anthropogenic Climate Change on Wildfire in California.” *Earth’s Future*. https://doi.org/10.1029/2019ef001210.

Zheng, Zhida. 2022. “A Novel Air Quality Prediction Method Based on GAF and Dense Net.” (*2022) International Conference on Machine Learning and Intelligent Systems Engineering (MLISE)*. https://doi.org/10.1109/mlise57402.2022.00029.