JURNAL TIFDA
Technology Information and Data Analytic
Website Jurnal : http://jeptira.id/index.php/tifda
e-ISSN : 3064-0660

Implementation of Data Mining to Support Waste Reduction Program in Special Region of Jakarta Using Time Series Algorithm and K-Means Clustering

Muhammad Krisna Adiputro[1], Afri Yudha[2]*
[1.2] Information Technology Study Program, Darma Persada University
[1]Jl. South Malaka Park, East Jakarta, DKI Jakarta 13450, Indonesia
ibnugazali@gmail.com

Abstract — This study aims to analyze the trend of waste growth in Jakarta using the ARIMA method and to group areas based on waste volume using the K-Means Clustering algorithm. The waste accumulation problem at the Bantargebang TPST continues to worsen each year, with increasing volumes from various sub-districts. Data used in this study were obtained from the DKI Jakarta Environmental Agency, covering the period from January 2022 to April 2024, focusing on organic waste, plastic, and household hazardous waste (B3). The research applies the CRISP-DM methodology, consisting of business understanding, data understanding, data preparation, modeling, evaluation, and implementation. Data processing includes cleaning, normalization, and splitting into training and testing sets. The analysis results show that the ARIMA model achieves good forecasting accuracy, with MAPE, MAE, and RMSE values around 3652. The K-Means algorithm successfully classifies Jakarta areas into three main clusters dominated by organic, plastic, and mixed waste types. A web-based system was developed using Streamlit and MongoDB Atlas to facilitate data analysis and visualization for policymakers, especially the Environmental Agency. The study concludes that ARIMA is effective in forecasting waste growth, while K-Means supports more targeted waste management strategies. It is recommended to enhance the system by incorporating external variables such as policy changes and socio-economic factors, and to improve model accuracy using more advanced machine learning techniques. Additionally, the system should be continuously updated and expanded to support more optimal and sustainable waste management across Jakarta.

Keywords – ARIMA, K-Means Clustering, CRISP-DM, MAPE, MAE.

Copyright © 2025 JURNAL TIFDA
All rights reserved.

I. INTRODUCTION

The problem of waste accumulation at the Bantar Gebang Integrated Waste Management Site (TPST) is indeed an increasingly pressing issue. According to data from the DKI Jakarta Environmental Agency, in 2021, the daily volume of waste entering the Bantargebang TPST reached around 7,800 tons, and is expected to continue to increase along with population growth and urbanization in Jakarta [1]. One of the main factors contributing to the increase in waste volume is changes in people's consumption patterns. Around 60% of the total waste produced in Jakarta is organic waste, especially food waste [2]. In addition, plastic waste, which is around 15% of the total waste, is also a serious problem because it is difficult to decompose and can pollute the environment [3]. Data from the World Bank (2021) shows that Indonesia is one of the largest contributors of plastic waste in the world, with Jakarta as one of the most affected cities.

The DKI Jakarta government has taken steps to address this issue, including plans to build an intermediate waste processing facility (ITF) which is expected to reduce dependence on the Bantar Gebang TPST. In this context, the use of the Time Series method with the ARIMA approach to predict waste growth in Jakarta is very relevant. By analyzing historical data on the volume of waste entering the Bantargebang TPST, the government can plan the management capacity needed in the future [4]. The ARIMA model can provide accurate predictions of waste volume based on historical data, allowing managers to take proactive steps in waste management [5][6].

In addition, the application of K-Means Clustering to map the tonnage of each region is also very important. By using this algorithm, the government can identify areas with high waste contributions [7]. K-Means Clustering analysis can help in understanding people's consumption patterns and provide useful information for related agencies to design more effective waste reduction programs [8].

By integrating data-based approaches, community participation, and technological innovation, it is hoped that the problem of waste accumulation at the Bantargebang TPST can be resolved comprehensively. This will not only improve environmental quality, but also provide economic and social benefits for the people of Jakarta.

II. METHODOLOGY

A. Basic Research Design

In this study, the field to be studied is the forecasting of waste growth in the Jakarta area. Using the Time Series algorithm and Grouping the amount of waste with the K-Means Clustering algorithm. By taking data directly from the DKI Jakarta Environmental Service office located at Jl. Mandala V No.67 1, RT.1 / RW.2, Cililitan, Kramat jati District, East Jakarta City and also at the Bantar Gebang Integrated Waste Processing Unit (UPST) office, Bekasi.

B. Research Methodology Design

The methodology used in this study follows the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework. This methodology includes six stages [9], namely:

1. Business understanding, using the K-Means algorithm, we will group provinces based on the type and tonnage of waste produced, so that we can provide more effective program recommendations such as Reduce, Reuse, Recycle (3R). In addition, the Time Series ARIMA approach will be used to predict the amount of waste in the future based on historical data from 2022 to April 2024.
2. Data understanding, the data used in this study is waste data generated by various sub-districts in DKI Jakarta from 2022 to April 2024. This data includes monthly tonnage of biodegradable and non-biodegradable waste, as well as various types of waste such as paper, plastic, metal, B3 waste, masks, and others. The ARIMA approach will be used to identify seasonal patterns and trends in waste tonnage data.
3. Data preparation, including data cleaning, normalization and standardization, feature engineering, and data separation.
4. Modeling, the main modeling method that will be used in this research is K-Means Clustering and ARIMA.
5. Evaluation, in the K-Means Clustering method, evaluation will be carried out using the Silhouette Score metric to measure the quality of clustering. The ARIMA prediction model will be carried out by comparing the prediction results with actual data using error metrics such as Root Mean Squared Error (RMSE).
6. Implementation, the results of this study will be applied to provide recommendations to local governments and sanitation services in improving the Reduce, Reuse, Recycle (3R) program to be more effective and targeted. In addition, the prediction of the amount of waste using the ARIMA and K-Means Clustering models will help in better planning and management of cleanliness, especially in the allocation of waste processing facilities Application Interface Design (Waterfall).

The waterfall method, or often called the waterfall model, is a development process into several sequential stages, like water flowing from top to bottom [9]. Each stage must be completely completed before proceeding to the next stage. In developing this system, the author uses a Software Development Life Cycle (SDLC) Waterfall methodology.

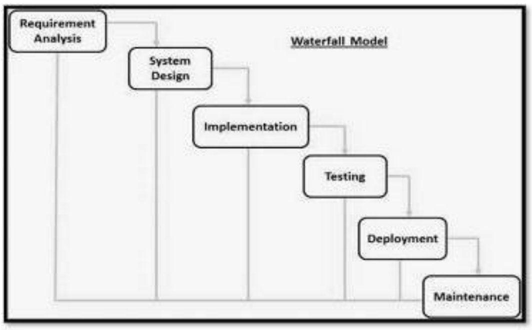

Figure 1. Waterfall Methodology

UML design, developers can create a blueprint or system design before implementation, thereby reducing the risk of errors and improving the quality of the resulting software. UML also serves as an effective communication tool between developers, analysts, and other stakeholders, because UML diagrams can be easily understood by all parties involved in the development project [10].

NoSQL database structure design has a different approach from relational databases (such as MySQL, PostgreSQL). If relational databases are highly structured with rigid tables, rows, and columns, NoSQL offers higher flexibility in storing unstructured or semi-structured data [11]. DesignFlow Charts to visualize the sequence of steps in a process, assist in understanding the workflow, as well as identifying potential improvements or problems in the system [12].

https://doi.org/10.70491/tifda.v2i1.74

III. RESULTS DISCUSSION

A. System Specification

The hardware and software specifications used by researchers in the process of creating this web-based inventory system application were carried out using the following specifications:

Devices	: Asus TUF Gaming Laptop
Operating system	: Windows 11
Processor	: AMD Rayzen 5
Memory	: 16 GB
Storage	: 512GB SSD
Database	: MongoD

This implementation section covers the operation of each page that has been created in the customer loyalty analysis system.

B. Black box test result

The testing focuses on system input and output to ensure that the application behaves as expected, as shown in the following test results:

a) Login page, users must login first to be able to enter the main page before making a forecast.

b) Data page, users will be asked to import the dataset first before making a forecast.

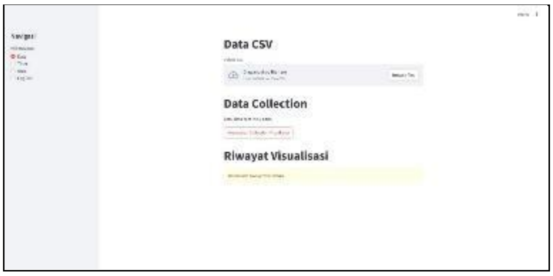

Figure 2. DATA Page

After importing the dataset, a summary of the dataset will be displayed.

c) Time page, the user will perform Time Series or ARIMA forecasting. The user must ensure that the dataset has been imported. If the dataset has not been imported, the user is asked to return to the data page as shown in the image below.

After making a prediction, the forecasting results will appear as shown in the image below. What will be displayed are graphs and also the evaluation results of MAE, MAPE, RMSE from ARIMA.

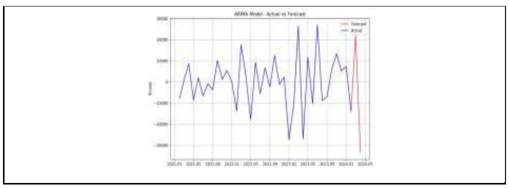

Figure 3. ARIMA Graph Display

d) Area Page (K-Means Clustering), the user will perform K-Means Clustering forecasting to group the waste. The user must ensure that the dataset has been imported. If the dataset has not been imported, it will be asked to return to the data page.

After importing the dataset, the user can perform clustering forecasting show on Figure 4.

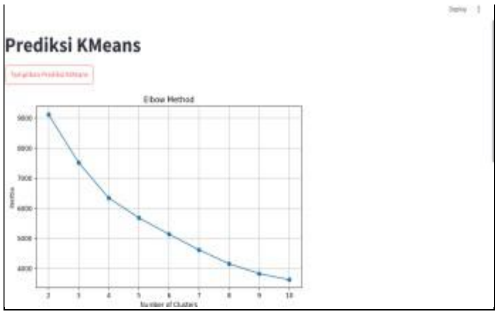

Figure 4. Elbow View Page Area

From the image, the Elbow Method shows clusters in 3 to 5 clusters.

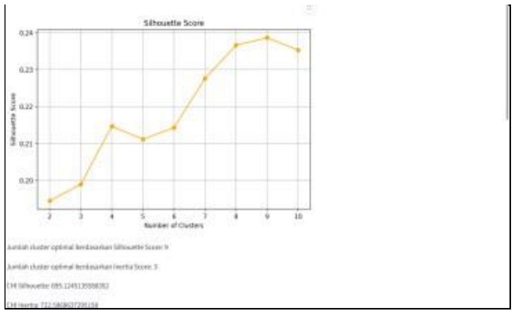

Figure 5. Area Page View

Figure 6 describe the Silhouette Score shows the results of 9 clusters at a value of 0.24.

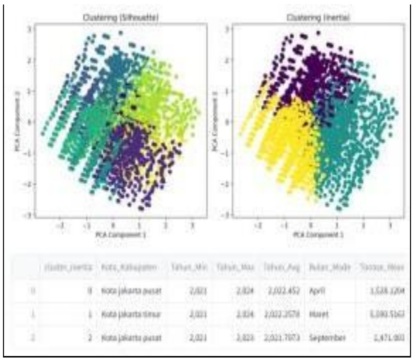

Figure 6. Area Page Plot View

From the image there are two comparison plots of cluster results from the Elbow Method and also clusters from the Silhouette Score.

A. Analysis results

a) ARIMA Analysis Results

The results of the ARIMA analysis show that the MAPE results at 10% to 14% RMSE are at 3652 units. The graphic results show stationary data according to the ARIMA criteria.

b) K-Means Clustering Analysis Results

Get 3 clusters from the Elbow Method results. Here are the cluster results.

Table 1. K-Means Clustering Results Table

Cluster	Kota/Kabupaten	Tahun Awal	Tahun Akhir	Bulan	Tonase Rata-Rata
0	Jakarta Pusat	2021	2024	April	1,528.1204
1	Jakarta Timur	2021	2024	Maret	5,590.5163
2	Jakarta Pusat	2021	2024	September	1,471.003

c) From the table, the results of K-Means Clustering can be interpreted as follows:

1. Cluster 0 is Central Jakarta with a cluster of moderate waste contribution from 2021 to 2024 and the highest average is 1,528 tons in April each year.
2. Cluster 1 is the cluster with the largest waste contributing area in Jakarta, located in East Jakarta, with the highest average tonnage of 5,590 which occurs every March every year.
3. Cluster 2 is an area that contributes little waste, namely Central Jakarta, with the largest average tonnage of waste being 1,471 tons in September each year.

IV. CONCLUSION

The ARIMA prediction results can be concluded that stationary data with an RMSE value of 3652 with an ARIMA prediction of (5,1,0) which is selected is able to capture a fairly significant pattern in the Tonnage data. This data produces predictions of rising and falling trends every month. From the results of this prediction, the Environmental Service can create a program for every month that experiences an increase in waste growth.

The results of three clusters from the grouping of waste volume. Cluster one shows the smallest amount of waste in Central Jakarta in the period of April every year, while Cluster two shows the largest amount of waste in East Jakarta in the period of March every year, and cluster three shows the medium amount of waste in the Central Jakarta area in the period of September every year. From the results of this Cluster, the Environmental Agency can make program planning according to the volume of waste distribution in priority areas or areas that contribute the most waste.

REFERENCES

[1] Artawan, KN, Dharma, A., & Sudestra, IMA (2023). Meta-Analysis Of Genetic Algorithm Implementation For Optimization Of Artificial Neural Network Methods. TIERS Information Technology Journal, 4(2), 129–133. https://doi.org/10.38043/tiers.v4i2.5515

[2] Assef, FM, Steiner, MTA, & Lima, EP de. (2022). A review of clustering techniques for waste management. In Heliyon (Vol. 8, Issue 1). Elsevie Ltd. https://doi.org/10.1016/j.heliyon.2022.e08784

[3] Bagwan, WA (2024). Electronic waste (E- waste) generation and management scenario of India, and ARIMA forecasting of E-waste processing capacity of Maharashtra state till 2030. Waste Management Bulletin, 1(4), 41–51. https://doi.org/10.1016/j.wmb.2023.08.002

[4] Kamal, A., Meidiana, C., & Yudono, A. (2024). Potential for Waste Reduction Through Waste Management at TPS, Kembangan District, West Jakarta City (Vol. 13, Issue 1).

[5] Magriaty, R., Murtilaksono, K., Anwar, S., Perencanaan, B., Penelitian, P., Pengembangan, D., Tapin, K., & Selatan, K. (nd). Journal of Regional and Rural Development Planning (Jurnal Perencanaan Pembangunan Daerah dan Perdesaan) February. 2023(1), 79–90. https://doi.org/10.29244/jp2wd.2022.7.1.79-90

[6] Onwe, J.C., Ridzuan, AR, Uche, E., Ray, S., Ridwan, M., & Razi, U. (2024). Greening Japan: Harnessing energy efficiency and waste reduction for environmental progress. Sustainable Futures, 8.1 0030. https://doi.org/10.1016/j.sftr.2024.100302

[7] Suryanti, T., Setyono, P., Ramelan, AH, Magister, (, & Lingkungan, I. (nd). Surakarta City Waste Management With Data Mining. National Data Science Seminar, 2024.

[8] Bagwan, WA (2024). Electronic waste (E- waste) generation and management scenario of India, and ARIMA forecasting of E-waste processing capacity of Maharashtra state till 2030. Waste Management Bulletin, 1(4), 41– 51. https://doi.org/10.1016/j.wmb.2023.08.002

[9] Kamal, A., Meidiana, C., & Yudono, A. (2024). Potential for Waste Reduction Through Waste Management at TPS, Kembangan District, West Jakarta City (Vol. 13, Issue 1).

[10]Magriaty, R., Murtilaksono, K., Anwar, S., Planning, B., Research, P., Development, D., Tapin, K., & South, K. (nd). Journal of Regional and Rural Development Planning (Jurnal Perencanaan Pembangunan Daerah dan Perdesaan) February. 2023(1), 79–90.https://doi.org/10.29244/jp2wd.2022.7.1.79-90

[11]Onwe, J.C., Ridzuan, AR, Uche, E., Ray, S., Ridwan, M., & Razi, U. (2024). Greening Japan:

Harnessing energy efficiency and waste reduction for environmental progress. Sustainable Futures, 8, 10030

[12]Suryanti, T., Setyono, P., Ramelan, AH, Magister, (Lingkungan, I. (nd). Surakarta City Waste Management With Data Mining. National Data Science Seminar, 2024.