International Journal of Informatics and Information Systems
Vol. 7, No. 2, March 2024, pp. 63-75

Prediction of Waste Generation in Yogyakarta Special Region Province Using ARIMA Model

Resti Nur Azizah[1,*], Ranggi Praharaningtyas Aji[2]

[1,2] *Information System Department, Amikom Purwokerto University, Indonesia*

(Received: September 16, 2023; Revised: December 9, 2023; Accepted: January 30, 2024; Available online: March 31, 2024)

Abstract

Waste management poses a significant challenge in densely populated urban areas of Indonesia, particularly in the Special Region of Yogyakarta, which grapples with both a large populace and robust tourism industry. The region witnesses a staggering 11.53% annual uptick in waste accumulation, exacerbating the strain on existing waste management infrastructure. Addressing this pressing issue necessitates a comprehensive analysis to forecast waste generation trends, enabling the formulation of effective management strategies. This study employs the CRISP-DM framework, facilitating a structured approach to analysis and informed decision-making. Utilizing a time series dataset spanning from 2016 to 2022, detailing waste generation across various districts and cities, the research employs the ARIMA model for predictive analysis. This model, renowned for its suitability in time series forecasting, emerges as the preferred choice. Projections derived from the ARIMA model reveal a notable surge in waste generation across the Special Region of Yogyakarta from 2023 to 2025. Specifically, it is anticipated that the aggregate waste output will escalate from 638 thousand tons in 2022 to 642 thousand tons in 2023, with an anticipated annual increase ranging between 5 to 7 thousand tons thereafter. These forecasts underscore the urgency of implementing proactive measures to mitigate the burgeoning waste management challenges facing the province.

Keywords: ARIMA, CRISP-DM, Time Series, Prediction, Waste

1. Introduction

Waste is one of the environmental issues that has received a lot of attention. According to the Waste Management Law No. 18 of 2008, it is stated that waste is the residue of daily human activities and or derived from natural processes. Waste is any form of material that is considered worthless or unwanted and is usually discarded by humans. These materials can come from various sources, such as home, industrial, commercial, or institutional operations[1]. The rate of population growth and changes in the consumption habits of the population, such as income changes, economic growth, urbanization, and industrial development, are other causes of the increasing volume of waste in a region[2]. The increasing prosperity of society and the advancement of human civilization are accompanied by an increasingly consumptive lifestyle. Consumption is becoming more and more prevalent. The needs of human life are becoming more diverse, and people are becoming more consumptive. This consumptive culture exploits natural resources while degrading environmental quality. The growth of solid waste (garbage), liquid waste, and gaseous waste damages environmental quality and also depletes resources[3].

Waste is a big problem in big cities. The Special Region of Yogyakarta with a population of 4,073,907 is one of the cities in Indonesia that has a waste problem. The Special Region of Yogyakarta (DIY) as an area with great tourism potential is a lot of tourism potential ranging from beach tourism, site tourism, temple tourism and nature tourism. This triggers the waste problem experienced by the Special Region of Yogyakarta, according to data from the Yogyakarta

***Corresponding author: Resti Nur Azizah (restinurazizah152@gmail.com)**
DOI: https://doi.org/10.47738/ijiis.v7i2.200

ISSN 2579-7069

City Environmental Service (DLH), the volume of waste increases by an average of 11.53% per year[3]. Then based on data from the DIY Environmental Agency, the volume of waste reaches 2,117 tons per day.

The waste system in the Special Region of Yogyakarta is carried out in landfills (TPA), namely the Banyuroto Landfill located in Kulon Progo Regency, the Wukirsari Landfill located in Gunungkidul Regency and the Piyungan Landfill located in Bantul Regency, but becomes a regional TPST for three districts, namely Sleman Regency, Bantul Regency and Yogyakarta City. According to the DIY Environmental Agency, the volume of waste generated in each landfill is very high when compared to the waste management capabilities of the existing landfills in DIY.

The waste management process must be handled properly because poor municipal waste collection and disposal reduces the aesthetic value of the environment, causes flooding, and pollutes the ecosystem (air, water, and soil). It can also impact human health, damage land and marine biota, and impact economic development, such as tourism[4]. In this case, an analysis related to the potential waste generation in the Special Region of Yogyakarta is needed to project the amount of waste generation in the future to help the provincial and district governments to formulate appropriate management policies and strategies. Efficient waste management planning requires proper forecasting, especially the amount of waste that must be controlled[4].

Forecasting is the act of estimating something that will happen in the future over a relatively long period of time[5]. Time series forecasting is the most important quantitative modeling model in which data from the same variable is collected and evaluated to build a model that can predict the future[4]. One of the methods used for forecasting is the time series method using ARIMA. In general, the ARIMA (Autoregressive Integrated Moving Average) method will provide better results than other forecasting methods because this method does not ignore the rules of time series data[6]. This research uses the ARIMA method and the CRISP-DM framework. CRISP-DM is used to facilitate effective decision making on the prediction results of waste generation in the Special Region of Yogyakarta[7].

2. Literature Review

2.1. Waste Generation

Waste generation constitutes a crucial metric in waste management, representing the quantity or mass of waste originating from various sources over a specified timeframe. This parameter serves as a cornerstone for devising effective waste planning and management strategies, necessitating meticulous measurement and analysis. Conventionally, waste generation is quantified through diverse units, enabling a multifaceted understanding of waste dynamics. One widely adopted approach involves measuring waste generation in weight units, typically expressed as kilograms per person per day. This method enables the estimation of waste production on a per capita basis, offering insights into the consumption patterns and disposal habits of the local populace. By extrapolating these figures, policymakers and waste management authorities can gauge the magnitude of waste generated within a community, thereby informing the design and implementation of tailored waste management initiatives.

Alternatively, waste generation can be assessed in volume units, commonly denoted as cubic meters per day. This metric assumes significance in planning the infrastructure and capacity of waste disposal facilities such as landfills or incineration plants. By quantifying waste generation volumetrically, stakeholders can ascertain the spatial requirements and operational parameters necessary for accommodating the burgeoning volume of waste, ensuring the efficacy and sustainability of waste management infrastructure. In essence, the measurement and analysis of waste generation encompass a multifaceted approach, encompassing both weight and volume units. By leveraging these methodologies, policymakers and waste management practitioners can gain comprehensive insights into waste dynamics, facilitating informed decision-making and the formulation of sustainable waste management strategies tailored to the unique needs and challenges of a given locality.[8].

2.2. Time Series Prediction

Prediction or forecasting stands as a pivotal statistical tool, wielding significant influence in guiding decision-making processes. By extrapolating trends from historical data, forecasting endeavors to provide insights into future events. Among the array of methods utilized for forecasting, the time series method holds prominence [9]. Time series data, characterized by sequential observations recorded over a defined period, serves as the foundation for this method.

ISSN 2579-7069

Central to time series analysis is the notion that current observations (Zt) are influenced by one or more preceding observations (Zt-k).

The overarching objective of time series analysis encompasses a multifaceted approach. Firstly, it seeks to unravel and elucidate underlying mechanisms governing observed trends. By dissecting past patterns, analysts aim to discern the driving forces behind data fluctuations. Additionally, time series analysis endeavors to furnish forecasts of future values, leveraging insights gleaned from historical data to extrapolate potential trajectories. This predictive capability empowers decision-makers with invaluable foresight, aiding in strategic planning and risk mitigation endeavors. Moreover, time series analysis plays a pivotal role in optimizing control systems, facilitating the fine-tuning of operational parameters to enhance efficiency and efficacy [5].

2.3. ARIMA

The moving average (MA) method was first used by Slutsky (1937) and the auto regressive (AR) method was first used by Yule (1927). Later, Wold (1938) created the theoretical basis for the ARMA combination process. He created an ARMA method based on three-way identification and estimation procedures (for AR, MA, and mixed ARMA processes). The results were extended to include periodic seasonal series and a simple development that includes nonstationary processes (ARIMA)[9]. In forecasting, the ARIMA model completely ignores the independent variables. To produce accurate short-term forecasts, ARIMA uses past and present values of the dependent variable. If the observations of the time series are statistically related to each other (dependent), ARIMA is suitable. In general, the Box-Jenkins ARIMA method will give better results than other forecasting methods because this method does not ignore the rules of time series[6]. Modeling time series data can use traditional statistical models, such as moving averages, exponential smoothing, and ARIMA, to model time series because their values are condensed into a linear function of the previous data. ARIMA (p, d, q) is the name of the ARIMA model where p is the number of autoregressive terms, d is the number of differences, and q is the number of moving averages. The general form of the ARIMA equation is as follows :

$$Z_t = c + \varphi 1 * Z_{t-1} + \varphi 2 * Z_{t-2} + \dots + \varphi p * Z_{t-p} + \theta 1 * e_{t-1} + \theta 2 * e_{t-2} + \dots + \theta q * e_{t-q} + e_t$$

Description :

Zt	: Observation value in period t.
C	: Constant or intercept
φ1, φ2, ..., φp	: The p-th autoregressive coefficient
θ1, θ2, ..., θq	: Coefficient of qth moving average
et	: Error at period t

2.4. Local Factors Affecting Waste Generation in the Special Region of Yogyakarta Province

The study's findings, as supported by [10], illuminate the multifaceted factors influencing waste generation dynamics within the Special Region of Yogyakarta. While population growth emerges as a prominent driver, its impact intertwines with various socio-economic variables, delineating a complex web of interdependencies. Notably, economic development and industrial activities wield considerable influence, as burgeoning economies often coincide with heightened levels of consumption and production, consequently amplifying waste generation rates [11]. Moreover, the evolving landscape of consumer behavior plays a pivotal role in shaping waste composition and volume. Shifts in purchasing patterns, disposal habits, and recycling practices collectively contribute to the nuanced tapestry of waste generation patterns. Furthermore, the efficacy of waste management practices emerges as a critical determinant. Robust waste management frameworks, characterized by efficient collection systems, recycling initiatives, and public awareness campaigns, wield the potential to curtail waste generation by fostering responsible disposal habits and minimizing wasteful consumption practices [10]. Conversely, inadequacies in waste management infrastructure may exacerbate the waste burden, perpetuating a cycle of escalating waste generation[10].

 ISSN 2579-7069

3. Method

The method used in this research is CRIPS-DM or Cross Industry Standard Process for Data Mining, where the actions in these phases are highly dependent on each other[12]. According to CRISP-DM, these phases consist of business understanding, data understanding, data preparation, data modeling, evaluation, and deployment. The framework that will be used in analyzing the prediction of waste generation in the Special Region of Yogyakarta Province is as follows:

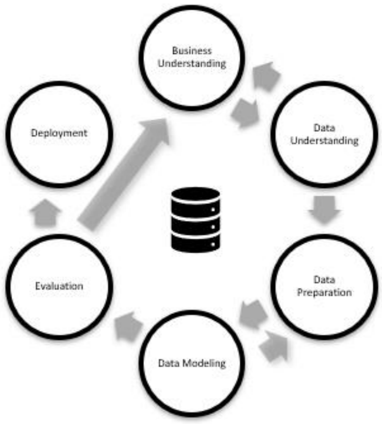


Figure 1. CRISP-DM Process

3.1. Business Understanding

In the initial stage of the research process, depicted in Figure 1, the focus is on comprehensively grasping the intricacies of the pertinent business processes, laying the groundwork for addressing prevalent challenges effectively. This step involves a methodical approach towards understanding the project's overarching objectives, delineating the specific business dilemmas at hand, and establishing clear success metrics [13]. In the context of this study, the primary aim is to harness predictive insights into future waste generation patterns, a pivotal undertaking essential for informed waste management planning within the Special Region of Yogyakarta. By elucidating the trajectory of waste production over forthcoming years, the research endeavors to furnish decision-makers with invaluable foresight, empowering them to devise strategic interventions geared towards optimizing waste management practices.

This phase entails a meticulous exploration of the multifaceted dimensions surrounding waste management dynamics, encompassing factors such as demographic trends, socio-economic indicators, tourism influxes, and industrial activities. Through a nuanced understanding of these contextual nuances, the research seeks to delineate a comprehensive framework that encapsulates the intricate interplay between various drivers influencing waste generation patterns [14]. Furthermore, this stage necessitates active engagement with stakeholders, including local authorities, waste management agencies, environmental experts, and community representatives. By fostering collaborative dialogues and soliciting diverse perspectives, the research endeavors to cultivate a holistic appreciation of the prevailing challenges and aspirations vis-à-vis waste management within the region. This inclusive approach not only enriches the research endeavor but also engenders a sense of ownership and collective responsibility towards fostering sustainable waste management practices.

3.2. Data Understanding

Once the business understanding stage is completed, the subsequent step, as illustrated in Figure 1, delves into a multifaceted process encompassing the identification and comprehensive understanding of the data at hand, data collection or acquisition, meticulous data analysis, and rigorous validation of the prevailing data quality [14]. This pivotal stage serves as the bedrock for the ensuing analytical endeavors, aiming to foster a profound comprehension of the dataset earmarked for analysis.

Central to this stage is the discernment of the intricacies embedded within the data, elucidating its inherent structure, patterns, and potential idiosyncrasies. Thorough exploration and scrutiny of the dataset are imperative to unearth underlying insights and nuances crucial for subsequent analytical procedures. Moreover, this phase entails the

ISSN 2579-7069

meticulous curation and acquisition of pertinent data sources, ensuring their alignment with the overarching analytical objectives [15]. The process of data collection entails the systematic retrieval of diverse datasets pertinent to the scope of the analysis. This may involve sourcing data from a myriad of repositories, ranging from governmental databases and archival records to proprietary datasets and third-party sources. The selection of data sources is guided by their relevance, reliability, and comprehensiveness in encapsulating the pertinent facets of the analytical inquiry.

Subsequent to data acquisition, meticulous scrutiny and evaluation of the data's quality ensue. This entails a comprehensive assessment of various quality attributes, encompassing accuracy, completeness, consistency, and timeliness [15]. Rigorous data validation techniques are employed to identify and rectify discrepancies, anomalies, or outliers that may compromise the integrity and reliability of the dataset. Furthermore, this stage necessitates the establishment of robust data governance frameworks and protocols to ensure adherence to data quality standards and regulatory compliance. This encompasses the implementation of data validation checks, data cleansing procedures, and data profiling techniques to fortify the integrity and veracity of the dataset.

3.3. Data Preparation

The subsequent phase, as delineated in Figure 1, involves meticulous data preparation, a pivotal step preceding any substantive data processing. This preparatory phase is indispensable in ensuring that the ensuing analyses effectively address the prevailing business challenges in a timely and pertinent manner. Data preparation encompasses multifarious tasks, prominently including data cleaning and transformation, each of which plays a crucial role in refining the dataset to render it amenable to analysis [16].

Data cleaning, the initial facet of this preparatory endeavor, entails the identification and rectification of any anomalies or inconsistencies within the dataset. This may encompass the detection and handling of empty, erroneous, or extraneous data entries, which could potentially distort the analytical outcomes if left unaddressed. Furthermore, data cleaning involves the judicious application of techniques such as outlier detection and removal, aimed at enhancing the overall quality and integrity of the dataset.

Subsequent to data cleaning, the process of data transformation ensues, which entails the systematic restructuring or manipulation of the dataset to render it conducive to analysis. One prevalent transformation technique involves the amalgamation of disparate variables or features to mitigate the impact of missing or inadequately represented data points. By consolidating related variables or synthesizing composite features, analysts can mitigate the deleterious effects of data lacunae, thereby fostering a more comprehensive and robust analytical framework.

Moreover, data transformation may encompass the normalization or standardization of variables to ensure comparability and facilitate meaningful interpretation of the ensuing analyses. This may involve rescaling numerical variables to a standardized range or encoding categorical variables into a more analytically tractable format. Additionally, data transformation techniques may encompass the creation of derived variables or features through mathematical operations or domain-specific knowledge, thereby enriching the analytical potential of the dataset.

3.4. Data Modeling

Furthermore, within the framework of CRISP-DM [17][18], the modeling stage plays a pivotal role in refining the methodology to be applied, aligning it with the intricacies of the existing data and the specific challenges at hand. In this study, meticulous consideration was given to selecting an appropriate model to forecast waste generation trends in the Special Region of Yogyakarta.

The ARIMA (Autoregressive Integrated Moving Average) model emerges as a robust choice owing to its adeptness in handling time series data, particularly in scenarios characterized by temporal dependencies and fluctuations. This model encompasses three key components: autoregression (AR), differencing (I), and moving average (MA). The autoregressive component captures the relationship between an observation and a number of lagged observations, reflecting the influence of past values on future outcomes. Meanwhile, the differencing operation helps in stabilizing the time series data by removing trends or seasonality, thereby rendering it stationary. Finally, the moving average component accounts for the correlation between an observation and a residual error from a moving average model applied to lagged observations.

In applying the ARIMA model to the dataset encompassing waste generation from 2016 to 2022 across various districts and cities, several steps were undertaken to ensure its efficacy. This included data preprocessing to address any anomalies or inconsistencies, such as missing values or outliers, which could potentially skew the forecasting outcomes. Additionally, model parameter selection involved rigorous testing and optimization to strike a balance between model complexity and predictive accuracy. Techniques such as grid search or iterative parameter tuning may have been employed to identify the optimal configuration for the ARIMA model.

Furthermore, model validation was conducted to assess the performance and reliability of the ARIMA model in capturing the underlying patterns and dynamics of waste generation. This likely involved partitioning the dataset into training and testing subsets, with the former utilized to train the model and the latter reserved for evaluating its predictive capabilities. Metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), or Root Mean Squared Error (RMSE) may have been employed to quantify the disparity between predicted and actual values, providing insights into the model's accuracy and precision.

3.5. Evaluation

Following the completion of data modeling, as depicted in Figure 1, the subsequent crucial step involves a meticulous evaluation of the employed model. This evaluation phase is pivotal in ensuring alignment with predefined project success criteria and overarching business objectives. It encompasses rigorous model testing and performance assessment to ascertain its efficacy in predicting waste generation patterns.

One of the fundamental evaluation metrics utilized within the CRISP-DM framework is the Root Mean Squared Error (RMSE) and Mean Absolute Percentage Error (MAPE) [19][20]. These metrics serve as quantitative indicators to gauge the accuracy and reliability of the predictive model. RMSE, derived from the square root of the mean square error, provides insights into the dispersion of prediction errors, thus offering a comprehensive understanding of the model's precision. By adjusting the error scale to match the target scale through the square root, RMSE facilitates a nuanced assessment of forecasting accuracy.

Similarly, MAPE, also known as the average mean absolute error, offers a holistic perspective on model performance by computing the average of absolute differences between predicted and actual values. This metric quantifies prediction accuracy by measuring the percentage deviation between observed and forecasted values for each period. Specifically, MAPE calculates the absolute error for each timeframe relative to the corresponding actual value, subsequently normalizing these errors to derive an average absolute percentage error..

3.6. Deployment

In the CRISP-DM (Cross-Industry Standard Process for Data Mining) framework, depicted in Figure 1, the deployment phase represents a pivotal stage wherein the waste generation prediction model, crafted from historical data of the DIY Province, transitions from development and validation to practical application. Preceding phases within the CRISP-DM process, including business understanding, data understanding, modeling, and evaluation, collectively lay the groundwork for this crucial deployment endeavor. During the deployment phase, the meticulously crafted waste generation prediction model is primed for real-world implementation. This involves integrating the model seamlessly into the operational fabric of waste management practices within the DIY Province. The transition from theoretical development to practical application demands meticulous planning and coordination to ensure the smooth execution of the predictive model within existing waste management frameworks.

As the model is rolled out into the operational environment, stakeholders are tasked with overseeing its integration and functionality. This entails ensuring that the model operates effectively within the operational constraints and dynamics of waste management practices in the province. Additionally, measures are taken to address any potential challenges or discrepancies that may arise during the deployment process, thereby optimizing the model's efficacy and utility. Crucially, the deployment phase culminates in the dissemination of analysis results and insights to pertinent stakeholders within the business environment. These insights serve as actionable intelligence, empowering decision-makers with the knowledge needed to enact informed strategies and initiatives aimed at enhancing waste management practices. Through collaborative engagement and knowledge-sharing, stakeholders leverage the predictive capabilities of the model to drive improvements and innovation within the realm of waste management in the DIY Province.

In essence, the deployment phase represents the tangible manifestation of data-driven insights, bridging the gap between theoretical modeling and real-world application. By seamlessly integrating the predictive model into operational workflows and providing stakeholders with actionable insights, this phase serves as a catalyst for transformative change and continual improvement in waste management practices within the DIY Province.

4. Results and Discussion

4.1. Business Understanding

Waste generation is the main focus of this research using time series analysis and ARIMA models based on historical data from each district in Yogyakarta Special Region Province. This analysis will be carried out on waste generation data in each district of Yogyakarta Special Region, namely Sleman, Bantul, Kulon Progo, Gunungkidul and Yogyakarta City. This prediction is intended to provide a basis for planning management policies that are efficient and in accordance with the dynamic changes in waste in Yogyakarta Province. This analysis is aimed at the authorities in Yogyakarta Special Region Province who are the main stakeholders as it relates to waste management policies. It also provides insights for the general public to increase their involvement in waste management improvement efforts.

4.2. Data Understanding

The research methodology relies on robust data acquired from two primary sources: the National Waste Information System (SIPN) and the Regional Development Planning Agency (Bappeda) of the Yogyakarta Special Region. This comprehensive dataset encompasses a range of pertinent variables, including the year of observation, the specific district or city under consideration, the daily volume of waste generated measured in tons, and the corresponding annual aggregate waste generation figures.

The National Waste Information System (SIPN) serves as a vital repository of nationwide waste data, offering insights into the broader trends and patterns of waste generation across Indonesia. Drawing from this centralized database ensures the research benefits from a nationally representative sample, enhancing the validity and generalizability of the findings. Moreover, leveraging data from the SIPN enables comparisons between waste management practices and outcomes across different regions within the country, facilitating a nuanced understanding of the factors influencing waste dynamics. Additionally, collaboration with the Regional Development Planning Agency (Bappeda) of the Yogyakarta Special Region provides localized insights tailored to the unique socio-economic and environmental context of the area. By accessing data directly from the local governing body responsible for regional development planning, the research gains access to granular, fine-grained information essential for crafting targeted waste management strategies. This partnership ensures that the research remains rooted in the realities and intricacies of waste management practices at the local level, thereby enhancing its relevance and applicability.

The dataset obtained from these sources undergoes rigorous validation and preprocessing to ensure data integrity and consistency. Quality assurance measures are implemented to address any discrepancies or anomalies within the dataset, thereby safeguarding the reliability of the subsequent analysis. By meticulously curating the dataset, the research endeavors to minimize bias and error, laying a solid foundation for robust and credible findings. Furthermore, the utilization of both daily and annual waste generation metrics enables a comprehensive analysis of temporal trends and seasonal variations in waste production. This multifaceted approach facilitates a nuanced understanding of the underlying drivers shaping waste dynamics, empowering policymakers and stakeholders to devise targeted interventions that address specific challenges and capitalize on emerging opportunities.

In summary, the integration of data from the National Waste Information System (SIPN) and the Regional Development Planning Agency (Bappeda) of the Yogyakarta Special Region underpins a rigorous and comprehensive research approach. By harnessing the insights gleaned from these rich datasets, the research endeavors to advance our understanding of waste management dynamics in urban contexts, ultimately informing evidence-based policy decisions and fostering sustainable development.

Tahun	Kabupaten/Kota	Timbulan Sampah Harian(ton)	Timbulan Sampah Tahunan(ton)
2016	Kab. Sleman	665.79	243013.41
2017	Kab. Sleman	685.00	250025.00
2018	Kab. Sleman	685.00	250025.00
2019	Kab. Sleman	699.12	255180.17
2020	Kab. Sleman	701.95	256210.07
2021	Kab. Sleman	735.57	268481.59
2022	Kab. Sleman	741.89	270789.85
2016	Kab. Bantul	488.46	150287.90
2017	Kab. Bantul	424.55	154960.86
2018	Kab. Bantul	429.15	156643.01
2019	Kab. Bantul	433.81	158343.16
2020	Kab. Bantul	526.09	160061.99

Figure 2. Waste Generation Datasets

Referring to Figure 2, the dataset utilized in this study encompasses a comprehensive compilation of waste generation data sourced from individual districts or cities within the Special Region of Yogyakarta (DIY) spanning the years 2016 through 2022. This dataset constitutes a vital resource in understanding the nuanced dynamics of waste accumulation patterns within the region over a significant timeframe. Each data point within the dataset encapsulates the quantitative representation of waste generation, meticulously disaggregated to reflect the distinct contributions from various administrative units comprising DIY. By capturing the temporal evolution of waste production across multiple spatial scales, ranging from densely urbanized municipalities to more rural districts, the dataset offers a granular insight into the heterogeneous nature of waste generation phenomena prevalent within DIY.

4.3. Data Preparation

At this stage, missing value checking is carried out, to find out whether there is empty data in each data row. After that, checking the data type is done to adjust to the existing variables and facilitate the modeling process.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 35 entries, 0 to 34
Data columns (total 4 columns):
 #   Column                        Non-Null Count  Dtype
---  ------                        --------------  -----
 0   Tahun                         35 non-null     int64
 1   Kabupaten/Kota                35 non-null     object
 2   Timbulan Sampah Harian(ton)   35 non-null     float64
 3   Timbulan Sampah Tahunan(ton)  35 non-null     float64
dtypes: float64(2), int64(1), object(1)
memory usage: 1.2+ KB
```

Figure 3. Data Preparation

The Figure 3 is the result of checking the dataset and no missing values are found. The next stage of data preparation is by splitting the data into training data and test data. The division of training data uses year variables because it is a time series analysis. The training data consists of data from 2016 to 2020, while the test data consists of 2020 to 2022. Training data is used to train Time Series models based on historical data, while test data is used to test the performance of the model in predicting the amount of annual waste generation in each district or city.

4.4 Data Modeling

The model used in this time series analysis is ARIMA with this technique applied using the statsmodels library and the Python programming language. The parameter used in this modeling is the amount of annual waste generation in tons. This research uses the ARIMA (Autoregressive Integrated Moving Average) model, which is a combination of the AR model, difference process, and MA model. The ARIMA (p, d, q) model will consider the influence of waste generation data in the previous period, overcome non-stationary problems through the difference process, and examine the relationship between forecasting errors in the current period and the previous period.

ISSN 2579-7069

District Best Order (p, d, q) and AIC values are the result of the process of selecting the best ARIMA model for each district in the time series analysis. The Best Order (p, d, q) refers to the combination of ARIMA parameters that gives the lowest AIC (Akaike Information Criterion) value, which indicates better model quality. An ARIMA model with such a combination of parameters will yield the lowest AIC value. The parameters p (order autoregressive term), d (order differencing), and q (order moving average term) in ARIMA modeling determine the characteristics and complexity of the model. The ideal p, d, and q values for each district or city are obtained based on the lowest AIC assessment. Then, the resulting ARIMA model can be evaluated using the best parameter combination. The results for each district/city in Yogyakarta Special Region Province based on the Best Order model are as follows:

	Kabupaten/Kota	Best Order (p, d, q)	AIC Value
0	Kab. Sleman	(1, 2, 1)	49.925752
1	Kab. Bantul	(1, 2, 0)	47.168195
2	Kab. Kulon Progo	(0, 2, 1)	37.546371
3	Kab. Gunungkidul	(0, 2, 0)	-20.413226
4	Kota Yogyakarta	(0, 2, 0)	1.869990

Figure 4. Best Order and AIC Value

The Figure 4. above is the result of the parameters p (order of autoregressive term), d (order of differencing), and q (order of moving average term) done by district / city, as well as the AIC value of each district / city from the ARIMA model that has been made.

4.5 Evaluation

In the evaluation stage, the quality and performance of the ARIMA model that has been built using the best order that has been determined previously are measured. The following are the results of the RMSE and MAPE Evaluation of Best Order (p, d, q) from each district/city.

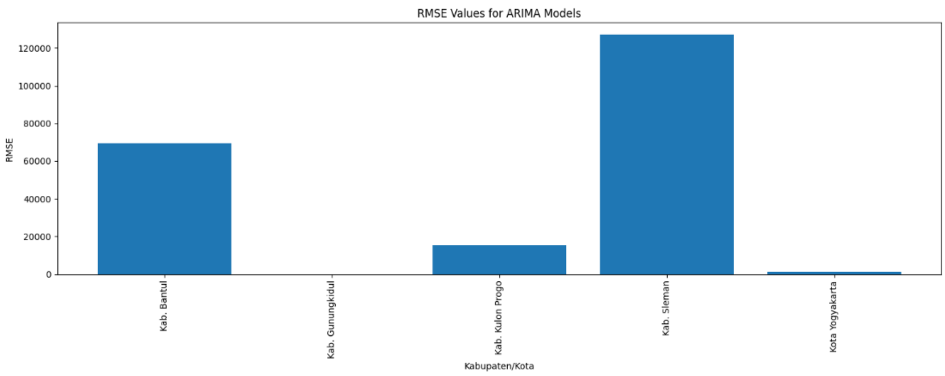


Figure 5. RMSE Value

Figure 5., shows that the RMSE value by district/city is significant enough to conclude that the model built has the adequacy to make predictions if the RMSE results are small. The waste generation prediction model has a high RMSE, meaning there is a large difference between the model prediction and the actual data. If the model has a low RMSE, it means that the difference between model predictions and actual data tends to be small. In the modeling results, Gunung Kidul Regency has the smallest RMSE value.

ISSN 2579-7069

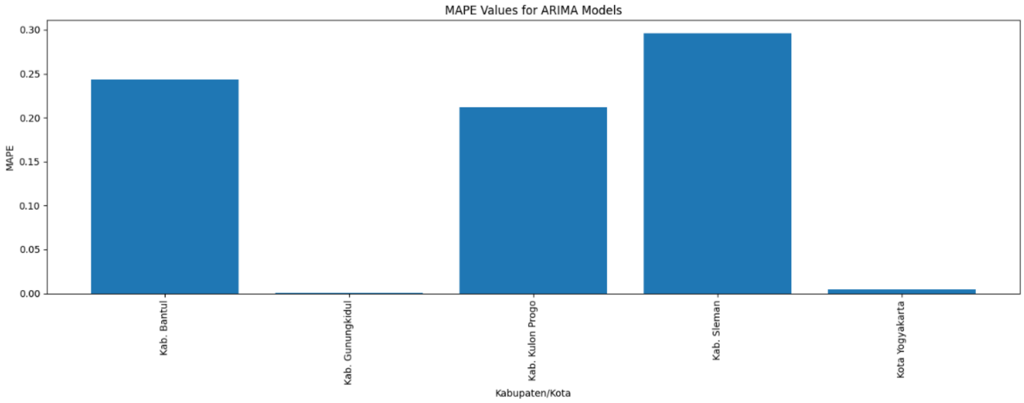


Figure 6. MAPE Value

The Figure 6., shows that the MAPE value by district is significant enough to conclude in measuring the error rate in percentage between the predicted and actual values, with a lower MAPE value in a district indicating a higher level of accuracy.

4.6 Deployment

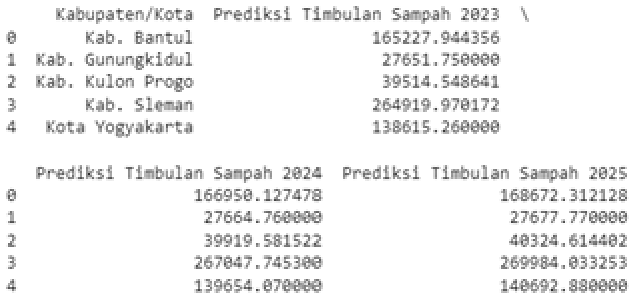
The final stage involves the forecasting modeling process by utilizing the Best Order (p, d, q) that has been previously identified. In this example, Yogyakarta City is used as an illustration to forecast the prediction of the Best Order (p, d, q) that has been known for the next five-year period. The results of the best order modeling of Yogyakarta City obtained the best order, namely (0,2,0) to show the predicted value of waste generation in the next five years. The amount of waste generation in 2021 is 136,537.64 tons with the lower limit of the confidence interval (mean_ci_lower) 136,537.17 tons and the upper limit of the confidence interval (mean_ci_upper) 136,538.10 tons, this shows an increase in waste generation over the next five years.

```
Kabupaten/Kota: Kota Yogyakarta
Best Order (p, d, q): (0, 2, 0)
Forecast for the next 5 years:
Timbulan Sampah Tahunan(ton)        mean   mean_se  mean_ci_lower
2021-01-01                     136537.64  0.236780  136537.175920
2022-01-01                     137576.45  0.529456  137575.412286
2023-01-01                     138615.26  0.885949  138613.523572
2024-01-01                     139654.07  1.296896  139651.528130
2025-01-01                     140692.88  1.756006  140689.438291

Timbulan Sampah Tahunan(ton)   mean_ci_upper
2021-01-01                     136538.104080
2022-01-01                     137577.487714
2023-01-01                     138616.996428
2024-01-01                     139656.611870
2025-01-01                     140696.321709
```

Figure 7. Prediction for The Next Five Years

Based on Fig 7., in Yogyakarta City, it is predicted that there will be an increase in the amount in 2022 of 137,576.45 tons, in 2023 of 138,615.26 tons, in 2024 of 139,654.07 and in 2025 of 140,692.88 tons. Similar predictions were also made for each regency or city in Yogyakarta Special Region Province in the dataset for the next five years. After predictions were made on the training dataset for the next five years. Then, the waste generation model for the years 2023 to 2025 was validated using actual data in the testing dataset. The prediction of waste generation for three years was done using the model trained on the previous dataset, and then trained on the actual data to obtain the prediction results for the expected year. The results of the waste generation prediction for each year 2023 to 2025 are as follows:



	Kabupaten/Kota	Prediksi Timbulan Sampah 2023	Prediksi Timbulan Sampah 2024	Prediksi Timbulan Sampah 2025
0	Kab. Bantul	165227.944356	166950.127478	168672.312128
1	Kab. Gunungkidul	27651.750000	27664.760000	27677.770000
2	Kab. Kulon Progo	39514.548641	39919.581522	40324.614402
3	Kab. Sleman	264919.970172	267047.745300	269984.033253
4	Kota Yogyakarta	138615.260000	139654.070000	140692.880000

Figure 8. Prediction for Each District with Actual Data

In the Figure 8, the predicted waste generation has increased from the last data year of 2022. Each district or city experienced an increase from 2023 to 2023. The district with the highest amount of waste generation is Sleman Regency which reached 264,919.97 tons, increased in 2024 to 267,047.74 tons and in 2025 rose to 269,984.03 tons.

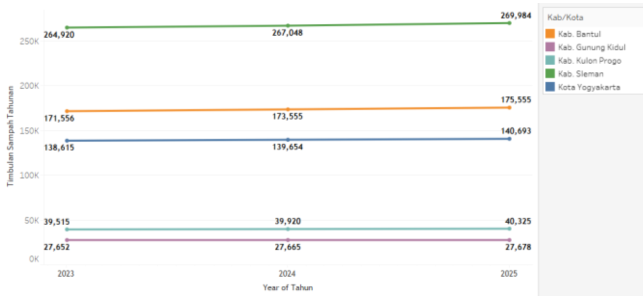


Figure 9. Visualization of Prediction Results

The Figure 9., shows a line graph projection with Tableau visualization, illustrating the increase in each district or city throughout 2023, 2024 and 2025. The predicted waste generation for each regency or city is increasing, as shown in the graph of Kulon Progo Regency increasing from 2022 to 39,514 tons in 2023, rising to 39,919 tons in 2024, and rising to 40,324 tons in 2025.

5. Conclusion

Waste generation in DIY Province for the period 2016-2022 as a whole continues to increase from year to year. The total average waste generation in DIY Province reached 614,098 tons over a period of 7 years. Sleman Regency is the regency with the largest contribution of waste generation among other regencies or cities, with an average generation of 256 thousand tons per year. Then followed by Bantul Regency with 157 thousand tons per year, Yogyakarta City with an average of 134 thousand tons per year, Kulon Progo Regency with 37 thousand tons per year and Gunung Kidul Regency with an average of 27 thousand tons per year. Overall, the districts/cities in DIY Province are predicted to experience an increase in the amount of waste generation from 2023 to 2025. The average waste generation in DIY Province is predicted to increase from a total of 638 thousand tons in 2022 to 642 thousand tons in 2023 and is predicted to continue to increase by 5 to 7 thousand tons per year. Analysis of the Root Mean Square Error (RMSE) value of the waste generation prediction model of Yogyakarta Special Region Province for each district/city shows that the accuracy of waste generation prediction varies among regions. The model with the smallest RMSE is Gunung Kidul Regency produces the best prediction, while Sleman Regency has the highest RMSE. Some of the prediction results with the ARIMA model still have high RMSE values, indicating challenges in forecasting waste generation in the future.

6. Declarations

6.1. Author Contributions

Conceptualization: RNA and RPA; Methodology: RPA; Software: RNA; Validation: RNA and RPA; Formal Analysis: RNA and RPA; Investigation: RNA; Resources: RPA; Data Curation: RPA; Writing Original Draft Preparation: RNA:

ISSN 2579-7069

Writing Review and Editing: RPA and RNA; Visualization: RNA. All authors have read and agreed to the published version of the manuscript.

6.2. Data Availability Statement

The data presented in this study are available on request from the corresponding author.

6.3. Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

6.4. Institutional Review Board Statement

Not applicable.

6.5. Informed Consent Statement

Not applicable.

6.6. Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

References

[1] S. Fitriani and A. Y. Astuti, "Pemetaan Jaringan Sampah Plastik Di Kota Yogyakarta," *J. PASTI*, vol. 15, no. 1, p. 73, 2021, doi: 10.22441/pasti.2021.v15i1.007.

[2] S. Zuhri and R. Nisa, "Peramalan Volume Sampah Menggunakan Pendekatan Arima Time Series," *J. Ind. Sci. Technol.*, vol. IV, no. 1, pp. 14–19, 2022.

[3] V. Arsanti and S. R. Giyarsih, "Pengelolaan Sampah oleh Masyarakat Perkotaan di Kota Yogyakarta," *J. Sains &Teknologi Lingkung.*, vol. 4, no. 1, pp. 55–66, 2012, doi: 10.20885/jstl.vol4.iss1.art6.

[4] E. Fitriastutik and M. Anityasari, "Forecasting Timbulan Sampah Kota Surabaya Menggunakan Time Series Analysis," *J. Tek. ITS*, vol. 9, no. 2, 2021, doi: 10.12962/j23373539.v9i2.56557.

[5] R. Ayu Wulandari and R. Gernowo, "Metode Autoregressive Integrared Moving Average (ARIMA) dan Metode Adaptive Neuro Fuzzy Inference System (ANFIS) dalam Analisis Curah Hujan," *Berk. Fis.*, vol. 22, no. 1, pp. 41–48, 2019.

[6] W. S. Rahayu, P. T. Juwono, and W. Soetopo, "Analisis Prediksi Debit Sungai Amprong Dengan Model Arima (Autoregressive Integrated Moving Average) Sebagai Dasar Penyusunan Pola Tata Tanam," *J. Tek. Pengair.*, vol. 10, no. 2, pp. 110–119, 2019, doi: 10.21776/ub.pengairan.2019.010.02.04.

[7] M. Rianto and R. Yunis, "Analisis Runtun Waktu Untuk Memprediksi Jumlah Mahasiswa Baru Dengan Model Random Forest," *Paradig. - J. Komput. dan Inform.*, vol. 23, no. 1, 2021, doi: 10.31294/p.v23i1.9781.

[8] S. Suhardono, I. Y. Septiariva, M. M. Sari, and I. W. K. Suryawan, "Kajian laju timbulan sampah di Yogyakarta selama pandemi Covid-19," *J. Pengelolaan Lingkung. Berkelanjutan (Journal Environ. Sustain. Manag.*, vol. 6, no. 3, pp. 169–179, 2022, doi: 10.36813/jplb.6.3.169-179.

[9] H. Hartati, "Penggunaan Metode Arima Dalam Meramal Pergerakan Inflasi," *J. Mat. Sains dan Teknol.*, vol. 18, no. 1, pp. 1–10, 2017, doi: 10.33830/jmst.v18i1.163.2017.

[10] A. D. C. D. S. Yunianto Setiawan, Searphin Nugroho, "Kecamatan Muara Badak Dihubungkan Dengan Tingkat," *Tek. Lingkung. Univer*, vol. 6, no. 1, pp. 53–57, 2022.

[11] A. Timbulan *et al.*, "Kraton Tepas Keprajuritan Yogyakarta Dan Analysis of Waste Generation and Composition in," 1994.

[12] C. Maisyarah, E. Haryatmi, R. Y. Fajriatifah, and Y. H. Puspita, "Prediksi Penyakit Diabetes menggunakan Algoritma Artificial Neural Network," *J. Data Sci. Inform.*, vol. 2, no. 1, pp. 46–52, 2022, [Online]. Available: http://publikasi.bigdatascience.id/index.php/jdsi

[13] N. Lundén, E. T. Bekar, A. Skoogh, and J. Bokrantz, "Domain knowledge in CRISP-DM: An application case in manufacturing," *IFAC-PapersOnLine*, vol. 56, no. 2, pp. 7603–7608, 2023. doi:10.1016/j.ifacol.2023.10.1156

[14] V. Plotnikova, M. Dumas, and F. P. Milani, “Applying the CRISP-DM data mining process in the financial services industry: Elicitation of adaptation requirements,” *Data and Knowledge Engineering,* vol. 139, no. 1, pp. 102013–102021, May 2022. doi:10.1016/j.datak.2022.102013

[15] A. Cheng, “Evaluating fintech industry’s risks: A preliminary analysis based on CRISP-DM Framework,” *Finance Research Letters,* vol. 55, no. 1, pp. 103966–103975, Jul. 2023. doi:10.1016/j.frl.2023.103966

[16] C. Schröer, F. Kruse, and J. M. Gómez, “A systematic literature review on applying CRISP-DM process model,” *Procedia Computer Science,* vol. 181, no. 1, pp. 526–534, 2021. doi:10.1016/j.procs.2021.01.199

[17] J. A. Solano, D. J. Lancheros Cuesta, S. F. Umaña Ibáñez, and J. R. Coronado-Hernández, “Predictive models assessment based on CRISP-DM methodology for students performance in Colombia - saber 11 test,” *Procedia Computer Science,* vol. 198, no. 1, pp. 512–517, 2022. doi:10.1016/j.procs.2021.12.278

[18] S. Huber, H. Wiemer, D. Schneider, and S. Ihlenfeldt, “DMME: Data Mining Methodology for engineering applications – a holistic extension to the CRISP-DM model,” *Procedia CIRP*, vol. 79, no. 1, pp. 403–408, 2019. doi:10.1016/j.procir.2019.02.106

[19] A. Kumar and P. S. V. S. S. Prasad, “Enhancing the scalability of fuzzy rough set approximate reduct computation through Fuzzy Min–max neural network and crisp discernibility relation formulation,” *Engineering Applications of Artificial Intelligence,* vol. 110, no. 1, pp. 104697–104705, Apr. 2022. doi:10.1016/j.engappai.2022.104697

[20] R. Duan, Y. Liu, J. Li, and S. Yan, “Mechanism for gel formation of pectin from mealy and crisp lotus rhizome induced by na+ and D-Glucono-D-Lactone,” *International Journal of Biological Macromolecules*, vol. 254, no. 1, pp. 127818–127826, Jan. 2024. doi:10.1016/j.ijbiomac.2023.127818