

Data Governance in Traffic Management Based on Average Daily Traffic (ADT) Data Prediction Using Python

¹Gohan Willy Christoper Sihite, ²Naila Syakirotul Rizkiyah

^{1,2}Infomation Systems, Telkom University Purwokerto, Banyumas, Indonesia

¹gohanw@student.telkomuniversity.ac.id, ²nailasya@student.telkomuniversity.ac.id

Submit: 25-03-2026, Revision: 25-05-2026, Accepted: 02-06-2026, Publication: 26-06-2026

Abstract— Data governance is crucial to ensure that traffic data is collected, managed, and used accurately to support quick, precise, and evidence-based decision-making in traffic management. The main challenge faced by many transportation agencies is the lack of an established data governance framework, which means that the utilization of Average Daily Traffic (ADT) data remains descriptive and does not yet support predictive planning. This situation results in traffic management being reactive and less effective in handling vehicle surges during critical periods. This study aims to implement a data governance framework in traffic management by developing comprehensive data management practices, including data acquisition, data quality assurance, and data-driven decision-making through ADT data prediction using the Python programming language. The approach applied is simple linear regression, applied to four years of historical ADT data to create a systematic and accountable prediction model, in accordance with data governance principles: accuracy, affordability, and policy relevance. The data were coded as numerical variables and analyzed to estimate future vehicle volumes clearly and replicably. The study's findings indicate that a Python-based prediction model, when integrated into the data governance structure, can provide more accurate, measurable, and policy-relevant traffic volume projections. This contributes to improving the quality of Data Governance, particularly in providing reliable and relevant information for strategic decisions. By incorporating this prediction system into the data governance framework, the relevant agencies are expected to be able to plan more proactive traffic management strategies, including vehicle flow regulation, road capacity optimization, and effective resource allocation based on structured understanding and strong data governance.

Keywords: Data Governance, Average Daily Traffic (ADT), Simple Linear Regression, Python, Traffic Management, Decision Support System.

Advances in information technology have brought about changes in many fields, among which is the field of transportation in particular traffic management [1]. The growing number of vehicles every year requires a traffic management system that is not just reactive, but also predictive based on data. In this context, Data Governance plays an important role in ensuring the available data can be managed, analyzed and utilized optimally in supporting effective and efficient decision making[2].

One important indicator in traffic analysis is the Average Daily Traffic (ADT), which is used to describe the level of vehicle density on a road segment during a certain period. However, the use of ADT data in many agencies is still limited to descriptive analysis, so it can not yet provide a predictive picture of the traffic conditions in the future. This leads to reactive and suboptimal decision making in traffic management, particularly in anticipating vehicle volume surges during certain periods, such as holidays or the mudik season[3].

To overcome the above problems, a prediction-based approach is needed that can process historical data into more valuable information[3]. A simple linear regression is one of the methods that can be used to identify the relationship between time variable and traffic volume. Using the Python programming language, the data processing process can be carried out more quickly, accurately and systematically so that it produces a prediction model that can be used as a basis for traffic management planning [4].

Based on this background, this research stems from the need for Data Governance in traffic management through ADT data prediction using a simple linear regression method based on Python[5]. It is hoped that the findings of this study can help improve the quality of decision-making, particularly by designing traffic management tactics that are more preventive, metric-based, and evidence-based to reduce the risk of congestion while increasing the efficiency of the transportation system.

1. INTRODUCTION

The effectiveness of the prediction model, however, depends not only on the selected analytical method but also on the quality and governance of the data being used. Inaccurate, incomplete, inconsistent, or poorly documented ADT data may produce unreliable prediction results and lead to inappropriate traffic management policies. Therefore, data governance principles, including data accuracy, consistency, accessibility, security, and accountability, must be integrated into the data processing stages. Through proper governance, traffic data can be treated as a strategic organizational asset rather than merely as administrative records collected periodically [6].

Previous studies have widely applied statistical and machine learning methods to predict traffic volume and congestion. Nevertheless, many studies tend to emphasize prediction accuracy without sufficiently discussing how the data are collected, validated, managed, and utilized within the decision-making process. In addition, the implementation of complex predictive algorithms may require substantial datasets, computational resources, and technical expertise that are not always available in public transportation agencies. Consequently, simple linear regression remains relevant as an interpretable, practical, and relatively easy-to-implement method for identifying traffic volume trends based on historical ADT data [7].

Accordingly, this research is expected to provide both methodological and practical contributions. Methodologically, the study demonstrates the integration of data governance principles with a Python-based simple linear regression model for predicting future ADT values. Practically, the resulting model can support transportation authorities in identifying potential increases in vehicle volume, determining priority road segments, and preparing preventive traffic engineering measures. Furthermore, the study may serve as an initial framework for developing a more comprehensive traffic analytics system by incorporating additional variables, such as population growth, land-use changes, weather conditions, public holidays, and major transportation events [8].

2. LITERATURE REVIEW

1. Average Daily Traffic (ADT)

The calculation of average daily traffic is the average traffic volume in a day that passes through one section of the road divided by the duration of observation (the duration of the vehicle survey), usually calculated throughout the year[1]. ADT is a standard term used in calculating 6 traffic loads on a road and is the basis in the

transportation planning process or in measuring pollution caused by traffic flow on a road. ADT is the result of dividing the number of vehicles obtained during observation with the duration of observation in road planning at the location.

$$ADT = \frac{\text{Traffic Volume During Observation}}{\text{During of Observation}}$$

This ADT data is quite thorough if observations are made at intervals of time intervals that are sufficient to describe traffic fluctuations during observation.

2. Traffic Management

Traffic management is the management and control of traffic flow by optimizing the use of existing infrastructure to provide convenience to traffic efficiently in the use of road space and smoothing the movement system[6]. This is related to the condition of the traffic flow and its supporting facilities at the present time and how to organize them to get the best performance.

Traffic management measures are generally implemented to balance traffic demand with the available capacity of the road network. These measures may include signal timing optimization, intersection improvement, lane management, parking control, traffic regulation, public transport prioritization, and the provision of clear road markings and signs. The selection of appropriate measures should be based on traffic volume, road capacity, travel speed, delay, queue length, accident records, and the characteristics of road users within the study area.

Effective traffic management also requires the integration of engineering, enforcement, education, and technological approaches. Traffic engineering is used to improve the physical and operational performance of roads, while law enforcement ensures that road users comply with traffic regulations. Public education helps increase awareness of safe and orderly driving behavior, whereas intelligent transportation systems can support real-time monitoring, traffic signal control, incident detection, and the dissemination of traffic information to road users.

The success of traffic management can be evaluated by comparing traffic performance before and after the implementation of a management strategy. Important indicators include reductions in travel time, delays, queue lengths, congestion levels, fuel consumption, emissions, and traffic accidents. Therefore, traffic management should not only focus on increasing vehicle movement but also consider road safety, accessibility, environmental

sustainability, and the mobility needs of pedestrians, cyclists, public transport users, and other road users

3. Simple Linear Regression

Linear regression is a static method used to analyze the relationship between independent variables (independent variables) and dependent variables (bound variables)[7]. In conclusion, linear regression is used to find out how changes in one variable affect other variables linearly. Recent research shows that linear regression is still an effective method and is widely used in analyzing data because it has a simple understanding and is quite accurate for data that has linear patterns.

In the application of linear regression, several assumptions need to be considered to ensure that the resulting model is valid and reliable. These assumptions include linearity, normality of residuals, homoscedasticity, independence of observations, and the absence of multicollinearity among independent variables. If these assumptions are fulfilled, the regression coefficients can provide an accurate explanation of the direction and magnitude of the relationship between variables. Therefore, assumption testing is an important stage before interpreting the results of a linear regression analysis.

Linear regression is commonly applied in various fields, including economics, business, engineering, health, and information technology. For example, this method can be used to predict sales based on promotional costs, estimate student performance based on study duration, or analyze the effect of system quality on user satisfaction. In addition to prediction, linear regression also helps researchers identify which independent variables have a significant influence on the dependent variable. Thus, linear regression can support evidence-based decision-making by transforming historical data into meaningful and measurable information.

4. Python

Python is a high-level, interpreted programming language that supports object-oriented paradigms and has dynamic semantics characteristics[8]. The language was designed to make the process of software development easier with flexible data structures, the dynamic typing system and dynamic binding mechanisms that allow writing more concise and adaptive code. One of the main advantages of Python is the simple and easy to understand syntax which increases the readability of the code and also facilitates the maintenance and development of the system[9]. Besides, Python provides support for using modules and packages, which allows developers to structure programs in a

modular manner and reuse previously written code [10]. Interpreter Python with its standard libraries is also available for free and can be run on a variety of platforms, making it one of the most flexible and easy-to-use programming languages in various fields including data processing, application development, and machine learning.

Python is also supported by a broad ecosystem of libraries and frameworks that can accelerate the development of various types of applications. Frameworks such as Django and Flask are commonly used in web application development, while libraries such as NumPy, Pandas, and Matplotlib support numerical computation, data analysis, and data visualization activities [11]. In the field of artificial intelligence, Python provides libraries such as TensorFlow, PyTorch, and Scikit-learn, which enable developers to build, train, and evaluate machine learning models efficiently. The availability of these tools reduces the need to develop functions from the beginning and allows the development process to be carried out more systematically and productively.

Another important characteristic of Python is its ability to integrate with other programming languages, databases, and external systems. Python can interact with languages such as C, C++, and Java, allowing developers to combine Python's simplicity with the processing performance of lower-level programming languages [12]. It also supports connectivity with various database management systems, including MySQL, PostgreSQL, SQLite, and MongoDB. This integration capability makes Python suitable for developing information systems that require data processing, application programming interfaces, automation, and communication between different software components.

Despite its advantages, Python also has several limitations that need to be considered in system development. As an interpreted language, Python generally has slower execution performance compared to compiled languages such as C or C++, particularly when processing computationally intensive tasks [13]. Its dynamic typing mechanism may also cause errors to be detected only during program execution if the development process is not supported by adequate testing. However, these limitations can be minimized through the use of appropriate libraries, structured coding practices, type annotations, and systematic testing methods. Therefore, Python remains a relevant programming language for developing flexible, maintainable, and scalable software applications.

5. Data Governance

Data Governance is a broad framework that oversees authority, control, and collaborative decision-making regarding the management of information assets within an entity[11]. This definition clarifies that Data Governance is not just about technology, but rather a comprehensive approach that unites policies, procedures, standards, and tasks in managing information as a whole [12].

The implementation of Data Governance requires a clear organizational structure that defines the authority, responsibilities, and accountability of every party involved in data management. This structure generally includes data owners, data stewards, data custodians, information technology units, and organizational leaders. Data owners are responsible for determining how data should be used, while data stewards ensure that data is managed according to established standards and business requirements. Meanwhile, data custodians focus on the technical storage, maintenance, and protection of data. A clear division of responsibilities can reduce overlapping authority and prevent uncertainty in decision-making related to organizational data [13].

Data Governance also plays an important role in maintaining data quality. Data used by an organization must be accurate, complete, consistent, relevant, valid, and available when required. Poor-quality data may result in inaccurate analysis, ineffective business processes, and inappropriate strategic decisions. Therefore, organizations need to establish data-quality standards, validation mechanisms, data-cleaning procedures, and regular monitoring processes. Through these mechanisms, Data Governance ensures that information assets can be trusted and effectively used to support both operational and strategic organizational activities [14].

In addition to data quality, Data Governance is closely related to information security, privacy, and regulatory compliance. Organizations must ensure that data can only be accessed, modified, distributed, and deleted by authorized parties. This requires the implementation of access controls, data classification, encryption, audit trails, backup procedures, and incident-response mechanisms. Data Governance provides the policies and decision-making framework that guide the implementation of these technical controls. Consequently, cybersecurity can be positioned as an important component of Data Governance because it protects information assets from unauthorized access, data leakage, manipulation, and other cyber threats [15].

Furthermore, Data Governance supports the management of the entire data life cycle, beginning with data creation and collection and continuing through storage, utilization, sharing, archiving, and destruction. Every stage of the data life cycle must be governed by clear standards to prevent unnecessary data accumulation, duplication, inconsistency, and misuse. Effective life-cycle management also allows organizations to determine how long certain data should be retained and when it should be securely destroyed. This approach not only improves operational efficiency but also reduces legal, financial, and reputational risks associated with inappropriate data management [16].

Ultimately, effective Data Governance strengthens the ability of an organization to make evidence-based decisions and derive value from its information assets. Reliable and well-managed data can support performance evaluation, business intelligence, risk analysis, service innovation, and strategic planning. However, the success of Data Governance depends not only on the availability of policies and technological systems but also on organizational commitment, leadership support, employee competence, and a culture of responsible data use. Therefore, Data Governance should be implemented as a continuous organizational process that is regularly evaluated and improved in accordance with changes in business needs, technology, regulations, and emerging risks [17].

The core elements of Data Governance consist of several closely connected parts.

1. Information policy is a set of rules that determine how information is collected, stored, utilized, and deleted within an organization [13].
2. Information manager refers to the determination of the functions and responsibilities of individuals or groups, including information owners, information managers, and information custodians, where each is responsible for the strategic, operational, and technical aspects of information management.
3. Information quality includes aspects of completeness, accuracy, timeliness, and consistency that must be met for information to be trusted as a basis for decision-making[14].
4. Information lifecycle management regulates the handling of information from the stages of acquisition, storage, processing, analysis, dissemination, to archiving and destruction of information.

The implementation of Data Governance has been proven to bring significant benefits in various organizational situations, including in the public sector[5]. Studies show that organizations that implement Data Governance in an organized manner can improve the quality of decision-making, reduce overlapping data, increase operational efficiency, and strengthen accountability and transparency in information management[15]. In the context of traffic management, Data Governance plays a crucial role in ensuring that traffic data such as Average Daily Traffic (ADT) derived from CCTV monitoring systems is managed in an organized and accountable manner. Without an adequate Data Governance framework, ADT data will only serve as past descriptive reports and cannot be optimally utilized to support predictive planning. Conversely, with the implementation of good Data Governance, ADT data can be managed through a structured data management cycle, starting from data acquisition, quality assurance, processing and analysis, to the presentation of predictive information, thereby producing relevant insights that can serve as the basis for formulating more proactive, measurable, and evidence-based traffic management policies.

3. RESEARCH METHODS

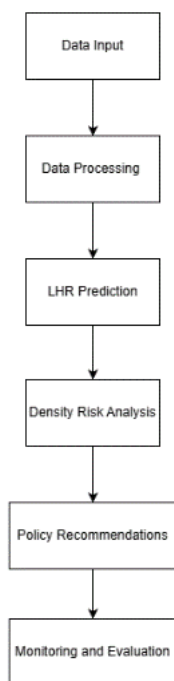


Fig. 1 Research Method

This study uses a quantitative approach with a predictive analysis method. This approach was chosen because the research focuses on processing Average Daily Traffic (ADT) numerical data to produce a prediction model that can be used as a basis for decision-making in traffic management.

A. Data Sources and Types

The data used in this study is secondary data in the form of ADT data obtained from CCTV traffic monitoring system. The data used is historical data in the last 4-year NATARU period. This data is quantitative and presented in the form of the number of vehicles per day.

B. Research Stages

1. Data Input
The initial stage is carried out by collecting ADT data from available sources, such as CCTV and related agency reports.
2. Data Processing
The data that has been collected is then processed through the stages of cleaning, grouping, and transformation into a ready-to-analyze format.
3. ADT Prediction
At this stage, analysis is carried out using simple linear regression to produce a vehicle volume prediction model.
4. Density Risk Analysis
The prediction results are used to analyze the potential level of traffic density and congestion risk.
5. Policy Recommendations
Based on the results of the analysis, policy recommendations were prepared in the form of a more effective and proactive traffic management strategy.
6. Monitoring and Evaluation
The final stage is monitoring and evaluating the resulting policies to ensure the effectiveness of implementation and as a basis for future improvements.

4. RESULTS AND DISCUSSION

This study is based on 4 years of continuous ADT data collected using traffic monitoring cameras. The data in the Process Using the python programming language were analyzed by a simple linear regression method.

The results of data processing show an increase in the volume of vehicles from year to year. A simple linear regression model built with the Python language predictive equation $y = ax + b$. The accuracy of the model is judged from the value of the determination coefficient R

Accuracy evaluation is carried out by calculating the value of the determination coefficient (R^2) which indicates how much variation the actual data can be explained by the model. The calculation results show an R^2 value that can adequately capture the growth pattern of vehicle volume. From the prediction results that have been produced, analysis is carried out to detect points and time periods that are most likely to experience congestion. The results of the analysis show that the volume of vehicles in the coming year will increase compared to the previous year.

These findings provide a solid basis for relevant agencies to prepare more proactive traffic management strategies, such as regulating vehicle flow, optimizing road capacity, and placing officers on target before peak congestion occurs.

In accordance with the perspective of Data Governance, the implementation of the python-based prediction model indicates that ADT data not only serves as a historical record but can be used as predictive information that can support evidence-based decision-making. This is in accordance with the ISACA principle which emphasizes the importance of balancing operational needs with the application of structured technology, starting from data collection, data analysis, to policy formulation based on predicted results. The use of a monitoring system that uses CCTV and is late integrated with a prediction model has proven to be effective in improving the quality of management in the context of traffic management.

Furthermore, the integration of ADT data, CCTV monitoring, and predictive analytics strengthens the reliability of traffic information management. A well-governed data system requires clear procedures for data collection, validation, storage, processing, and reporting. These procedures are important to ensure that the data used in the prediction model are accurate, consistent, complete, and traceable. Without adequate data quality controls, prediction results may contain bias and lead to inappropriate traffic management decisions. Therefore, periodic data verification and model evaluation should be conducted to maintain the accuracy and relevance of the prediction results.

From the perspective of IT Governance, the implementation of this system also requires clear

allocation of roles and responsibilities among the agencies involved. Traffic management authorities, information technology units, data analysts, and CCTV operators need to coordinate in managing the system and responding to the prediction results. Governance mechanisms should include access control, system security, data protection, maintenance schedules, and procedures for handling technical failures. These controls can reduce operational risks while ensuring that the traffic monitoring and prediction system remains available, secure, and capable of supporting public services.

In the future, the prediction model can be improved by incorporating additional variables, such as vehicle type, weather conditions, road characteristics, public events, holidays, accidents, and peak-hour patterns. The use of more advanced analytical methods, including multiple regression, time-series analysis, or machine-learning algorithms, may also improve prediction accuracy. Nevertheless, the simple linear regression model used in this study provides an understandable and practical initial framework for predicting traffic volume. Its implementation demonstrates that properly governed traffic data can be transformed into strategic information for planning, evaluating, and improving urban traffic management policies.

5. CONCLUSION

In this study, a simple linear regression vehicle volume prediction model has been successfully developed using the Python programming language, as well as utilizing Average Daily Traffic data collected for 4 years in the NATARU period. The resulting model can provide an estimate of vehicle volume in 2026 with better accuracy and measurement compared to the descriptive approach that has been applied by related agencies.

The application of this prediction model is expected to contribute to improving the quality of information technology management in traffic management. ADT data that was initially only descriptive can now be processed into relevant predictive information to support data-driven decision-making. Thus, relevant institutions can plan traffic management strategies more proactively, which include regulating vehicle flow, optimizing road capacity, and placing resources appropriately before peak congestion occurs.

For further research, it is suggested that further research explore more advanced prediction methods, such as polynomial regression or machine learning techniques, to improve model accuracy. In addition, expanding the coverage of the data by considering external factors such

as weather, national holidays, and road infrastructure conditions is also highly recommended.

REFERENCES

- [1] M. Damjanović, Ž. Stević, D. Stanimirović, I. Tanackov, and D. Marinković, "Impact Of The Number Of Vehicles On Traffic Safety: Multiphase Modeling," *Facta Universitatis, Series: Mechanical Engineering*, vol. 20, no. 1, pp. 177–197, Apr. 2022, doi: 10.22190/FUME220215012D.
- [2] D. Whittard, F. Ritchie, R. Musker, and M. Rose, "Measuring the value of data governance in agricultural investments: A case study," *Exp. Agric.*, vol. 58, Feb. 2022, doi: 10.1017/S0014479721000314.
- [3] A. A. A. Dabo, R. S. Mor, M. M. Albatal, M. Raji, and M. Aina, "Logistics Planning Strategies For Seasonal Traffic Management," *Logforum*, vol. 21, no. 2, pp. 167–181, Apr. 2025, doi: 10.17270/J.LOG.001150.
- [4] A. Pajaziti, F. Basholli, and Y. Zhaveli, "Identification and classification of fruits through robotic system by using artificial intelligence," *Engineering Applications*, vol. 2023, no. 2, pp. 154–163, 2023, [Online]. Available: <https://publish.mersin.edu.tr/index.php/enap>
- [5] H. M. Naguib, H. M. Kassem, and A. E.-H. M. A. Naem, "The impact of IT governance and data governance on financial and non-financial performance," *Future Business Journal*, vol. 10, no. 1, Dec. 2024, doi: 10.1186/s43093-024-00300-0.
- [6] Joshua Idowu Akerele, Abel Uzoka, Pascal Ugochukwu Ojukwu, and Olugbenga Jeremiah Olamijuwon, "Optimizing traffic management for public services during high-demand periods using cloud load balancers," *Computer Science & IT Research Journal*, vol. 5, no. 11, pp. 2594–2608, Nov. 2024, doi: 10.51594/csitrj.v5i11.1710.
- [7] H. Rajh-Weber, S. E. Huber, and M. Arendasy, "Using Heteroskedasticity-Consistent Standard Errors and the Bootstrap for Linear Regression Analysis Available in SPSS: A Tutorial," *Adv. Methods Pract. Psychol. Sci.*, vol. 9, no. 1, Jan. 2026, doi: 10.1177/25152459251408046.
- [8] O. Castro, P. Bruneau, J. S. Sottet, and D. Torregrossa, "Landscape of High-Performance Python to Develop Data Science and Machine Learning Applications," *ACM Comput. Surv.*, vol. 56, no. 3, Mar. 2024, doi: 10.1145/3617588.
- [9] V. Chaturvedi, "Modern Software Development with Java, Spring Boot, and Python: A Survey of Frameworks and Best Practices," *ESP Journal of Engineering & Technology Advancements*, vol. 3, no. 4, 2023, doi: 10.56472/25832646/JETA-V3I8P121.
- [10] S. Lehtola, "A call to arms: Making the case for more reusable libraries," Nov. 14, 2023, *American Institute of Physics Inc.* doi: 10.1063/5.0175165.
- [11] H. Y. Kim and J. S. Cho, "Data Governance Framework for Big Data Implementation with a Case of Korea," *Proceedings - 2017 IEEE 6th International Congress on Big Data, BigData Congress 2017*, pp. 384–391, 2017, doi: 10.1109/BigDataCongress.2017.56.
- [12] R. Setyadi and A. A. Rahman, "Enhancing IT Governance Based on Risk and Security Analysis in a Private School: A COBIT 2019 Approach," *International Journal On Informatics Visualization*, vol. 9, no. 5, pp. 2148–2156, Sep. 2025, doi: 10.62527/ijoiv.9.5.3414.
- [13] F. Shwedehe *et al.*, "The impact of intellectual property rights and the level of information sensitivity on information security in the United Arab Emirates," *Journal of Infrastructure, Policy and Development*, vol. 8, no. 8, 2024, doi: 10.24294/jipd.v8i8.6303.
- [14] H. Ghalavand, S. Shirshahi, A. Rahimi, Z. Zarrinabadi, and F. Amani, "Common data quality elements for health information systems: a systematic review," *BMC Med. Inform. Decis.*

Mak., vol. 24, no. 1, Dec. 2024, doi:
10.1186/s12911-024-02644-7.

- [15] A. Polzonetti and M. Sagratella, "Towards a data-driven enterprise: Effects on information, governance, infrastructures and security," *IEEE International Conference on Industrial Engineering and Engineering Management*, vol. 2017-Decem, pp. 1480–1484, 2018, doi: 10.1109/IEEM.2017.8290139.