
Application of Data Mining Using Multiple Linear Regression to Project Population Figures at the Central Bureau of Statistics of Ogan Ilir Regency

Meiky Alfarizi^{1*}

Abstract

The Central Bureau of Statistics (Badan Pusat Statistik BPS) is a government agency mandated to provide comprehensive, accurate, and up-to-date population statistics to build a reliable, effective, and efficient national statistical system that supports national development planning. However, the BPS office of Ogan Ilir Regency faces challenges in its current population-data collection system, which remains incomplete. This study applies a multiple linear regression analysis method a multivariate technique used to estimate the relationships between dependent and independent variables of metric or non-metric type. The findings demonstrate that data-mining techniques are effective tools for addressing data-related problems by identifying useful and relevant knowledge within large databases. The proposed projection information system is designed to meet population-data requirements at the regency or municipal level for both present and future needs.

Keywords

Data Mining; Projection Information System; Central Bureau of Statistics

Article History

Received 21 March 2025

Accepted 14 June 2025

How to Cite

Meiky, A, (2025). Application of Data Mining Using Multiple Linear Regression to Project Population Figures at the Central Bureau of Statistics of Ogan Ilir Regency, (JIKSI), 6(2), [54-58].

^{1*} Universitas Bina Darma, Indonesia, Corresponding email: [meiky.rizi05@gmail.com]

Introduction

In the contemporary era of globalization, rapid advancements in information technology have profoundly influenced every aspect of human activity. Technology has become an inseparable element of daily life, evident from the continuous emergence of innovations ranging from simple applications to transformative global solutions. Initially regarded as a remarkable novelty, the Internet has evolved into a fundamental medium for communication and interaction across social, economic, and governmental domains.

The integration of information and communication technology has revolutionized how people communicate and disseminate information. Historically, information exchange relied on paper-based or analog media such as letters, radio, and television, resulting in limited accessibility particularly across national borders. In contrast, modern Internet technologies enable information to spread rapidly and without geographic limitation, covering diverse fields including social sciences, economics, and politics. Within the social domain, technological systems are instrumental in storing and projecting large-scale population data across regions and countries.

The Central Bureau of Statistics (BPS) serves as a non-ministerial government institution reporting directly to the President of the Republic of Indonesia. Its mission is to provide comprehensive, accurate, and up-to-date population data that support evidence-based national development planning. The BPS office in Ogan Ilir Regency, South Sumatra, manages extensive datasets through periodic population surveys. Nevertheless, the agency has encountered difficulties in estimating annual population growth due to several technical constraints (Amin et al., 2019).

At present, the BPS Ogan Ilir primarily uses Microsoft Excel for population-data processing (Ramadhan, 2019). This manual approach hampers efficiency, complicates data analysis, and increases the risk of data loss. Personnel often experience challenges in locating specific information or deriving insights from the existing system, which lacks automation and analytical capability (Setyawati, 2018).

Data mining offers a viable solution to these challenges. It enables the discovery of meaningful patterns and relationships within large datasets and supports knowledge extraction for decision-making. The principal functions of data mining include estimation, association, prediction, classification, and clustering (Yusa & Sindu, 2015).

Accordingly, this study Application of Data Mining Using Multiple Linear Regression to Project Population Figures at the Central Bureau of Statistics of Ogan Ilir Regency aims to assist BPS Ogan Ilir in projecting population growth through the application of computational and statistical modeling techniques.

Methodology

2.1 Research Method

Linear regression is a statistical technique used to examine the relationship between one or more independent variables and a dependent variable. The independent variables (predictors) influence the dependent variable (outcome). Linear regression is applicable to data measured on interval or ratio scales (Yufensia, 2019).

Regression analysis quantifies the strength of association among variables, determines causal relationships, and enables prediction of dependent-variable values from known

predictors. Two main types of linear regression exist: simple linear regression (one predictor) and multiple linear regression (two or more predictors). In this study, multiple linear regression is employed within a data-mining framework for population projection (Lungan, 2006).

2.2 Tools and Data Processing

Data analysis was conducted using Microsoft Excel for preprocessing and MATLAB for simulation and prediction through the Backpropagation Neural Network (BPNN) algorithm.

2.3 System Development Approach

System Development Approach System development followed the Waterfall System Development Life Cycle (SDLC) model, consisting of five sequential stages:

1. Requirement Analysis
2. System Design
3. Program Coding
4. Testing
5. Maintenance

2.4 Backpropagation Neural Network (BPNN)

To enhance regression performance, a BPNN architecture was implemented. The principal steps included:

1. Input and Target Definition – Network inputs (X_1) represent historical population data of Ogan Ilir Regency, and targets (Y_k) denote projected population values for subsequent years.
2. Data Partitioning – Data were divided into training (80%) and testing (20%) subsets.
3. Normalization – Each variable was scaled within [0, 1]

2.5 Data Sources

Population data were obtained from the official BPS Ogan Ilir website (<https://oganilirkab.bps.go.id/>)

Results and Discussion

3.1 Data Preparation

The dataset comprised monthly population figures from 2011 to 2020. The maximum recorded value was 2,140,944 and the minimum 785,407. After normalization, data values ranged between 0.1 and 0.9. Training data encompassed the first five years (2011–2015), while testing data represented the subsequent five years (2016–2020). Each subset contained 48 records (12 per year).

3.2 System Interface Design

The designed system incorporated several user-interface (UI) components: Main Page: Displays menus for initiating and terminating data training and provides access to performance and regression visualization modules. Performance Page: Shows the Mean Squared Error (MSE) and epoch count, reflecting model-training accuracy. Regression Page: Presents scatterplots of predicted versus actual outputs to assess linear relationships.

Discussion

Training results indicated that the neural-network-enhanced regression achieved a correlation coefficient (R) of 0.83753, with optimal performance obtained at epoch 1000 and $MSE = 0.0011071$.

Testing yielded $R = 0.77084$, confirming adequate predictive reliability despite the limited training dataset. These findings suggest that the hybrid regression–neural-network

model effectively captures population-growth trends in Ogan Ilir Regency. Nonetheless, predictive precision could be improved by incorporating additional historical data and refining model parameters.

Overall, the model demonstrates potential as a decision-support tool for demographic analysis and planning at the regency level.

Conclusion and Recommendations

Based on the study results and analyses, the following conclusions are drawn:

1. The application of data-mining techniques using MATLAB successfully projected the population of Ogan Ilir Regency.
2. The developed system is capable of predicting population growth accurately within the examined region.
3. The resulting software can assist policymakers by providing quantitative projections to support strategic planning involving demographic factors.

Disclosure Statement

The authors declare no potential conflicts of interest with respect to the research, authorship, or publication of this article.

Acknowledgments

The authors express sincere appreciation to the Central Bureau of Statistics of Ogan Ilir Regency and to Universitas Bina Darma for their support and provision of population data that made this study possible.

References

- Afiana, F. N., Subarkah, P., & Hidayat, A. K. (2019). Comparative analysis of the TAM and UTAUT2 methods in measuring the success of SIMRS implementation at Wijaya Kusuma DKT Hospital Purwokerto. *Matrik: Journal of Management, Information Technology, and Computer Engineering*, 19(1), 17–26. <https://doi.org/10.30812/matrik.v19i1.432>
- Amin, M. F. A., Kunang, Y. N., & Purnamasari, S. D. (2019). Proceedings of the Bina Darma Conference on Computer Science 2019. Faculty of Computer Science, Universitas Bina Darma.
- Ramadhan, R. (2019). Use of Microsoft Excel in population data processing. *Bina Darma Journal of Computer Science*.
- Setyawati, L. (2018). Challenges of data management in local government statistical agencies. *Journal of Information Systems*.
- Yanto, R. (2020). Implementation of data mining for forecasting electricity demand in Lubuklinggau City. *Techno.Com*, 19(2), 197–206. <https://doi.org/10.33633/tc.v19i2.3447>
- Yufensia, S. (2019). Statistical modeling using multiple regression. Universitas Bina Darma.

Yusa, M., & Sindu, W. (2015). Data mining and clustering functions for knowledge discovery. Journal of Computational Research.

Biographical Notes

Meiky Alfarizi is a researcher at the Faculty of Computer Science, Universitas Bina Darma. His research interests include data mining, artificial intelligence, and population data analytics.

Linda Atika is a lecturer at the Faculty of Computer Science, Universitas Bina Darma. Her academic focus lies in software engineering, data analysis, and intelligent information systems.