

ABSTRAK

MODEL *DATA MINING* UNTUK MEMPREDIKSI PERTUMBUHAN PENDUDUK DI KABUPATEN SUBANG MENGGUNAKAN ALGORITMA *XGBOOST*

Oleh:

Nisairrizqy Nabilah Faizah

1197050096

Pertumbuhan jumlah penduduk merupakan salah satu permasalahan yang perlu diperhatikan dalam perencanaan pembangunan daerah karena perubahan jumlah penduduk berpengaruh terhadap kebutuhan pelayanan publik, infrastruktur, dan alokasi sumber daya. Ketersediaan data administrasi kependudukan yang terus diperbarui memberikan peluang untuk melakukan estimasi jumlah penduduk pada periode mendatang, namun pemanfaatan data tersebut sebagai dasar prediksi masih memerlukan pendekatan yang mampu mengidentifikasi hubungan antara berbagai indikator administrasi kependudukan dengan perubahan jumlah penduduk. Salah satu pendekatan yang dapat digunakan adalah data mining dengan algoritma Extreme Gradient Boosting (XGBoost), yang mampu memodelkan hubungan nonlinier antara variabel prediktor dan variabel target. Penelitian ini bertujuan membangun dan mengevaluasi model prediksi jumlah penduduk Kabupaten Subang menggunakan algoritma XGBoost dengan memanfaatkan indikator administrasi kependudukan sebagai variabel prediktor. Dataset yang digunakan merupakan data administrasi kependudukan Kabupaten Subang periode 2023–2025 yang diperoleh dari Dinas Kependudukan dan Pencatatan Sipil Kabupaten Subang. Proses penelitian dilakukan menggunakan tahapan CRISP-DM yang meliputi business understanding, data understanding, data preparation, modeling, evaluation, dan deployment/development. Model XGBoost kemudian dibandingkan dengan *Linear Regression* sebagai model pembanding menggunakan

metrik *Mean Absolute Error* (MAE), *Root Mean Squared Error* (RMSE), *Mean Absolute Percentage Error* (MAPE), dan *R² Score* serta divalidasi menggunakan *5-Fold Cross Validation*. Hasil pengujian pada data uji menunjukkan bahwa model XGBoost menghasilkan MAE sebesar 1.120,71 jiwa, RMSE sebesar 1.611,61 jiwa, MAPE sebesar 2,30%, dan *R² Score* sebesar 0,9972. Sementara itu, hasil *5-Fold Cross Validation* menghasilkan rata-rata MAE sebesar 2.723,32 jiwa, RMSE sebesar 5.500,84 jiwa, MAPE sebesar 3,86%, dan *R² Score* sebesar 0,9181. Hasil tersebut menunjukkan bahwa model XGBoost mampu menghasilkan prediksi jumlah penduduk dengan tingkat kesalahan relatif rendah dan performa yang baik pada beberapa pembagian data.



Kata kunci: Pertumbuhan Penduduk, Administrasi Kependudukan, *Data Mining*, XGBoost

ABSTRACT

DATA MINING MODEL TO PREDICT POPULATION GROWTH IN SUBANG USING XGBOOST ALGORITHM

By:

Nisairrizqy Nabilah Faizah

1197050096

Population growth is one of the issues that needs to be considered in regional development planning because changes in population affect the need for public services, infrastructure, and resource allocation. The availability of continuously updated population administration data provides an opportunity to estimate the population in the future period, but the use of this data as a basis for prediction still requires an approach that is able to identify the relationship between various population administration indicators and population changes. One approach that can be used is data mining with the Extreme Gradient Boosting (XGBoost) algorithm, which is able to model nonlinear relationships between predictor variables and target variables. This study aims to build and evaluate a population prediction model for Subang Regency using the XGBoost algorithm by utilizing population administration indicators as predictor variables. The dataset used is population administration data for Subang Regency for the period 2023–2025 obtained from the Subang Regency Population and Civil Registration Office. The research process was carried out using the CRISP-DM stages which include business understanding, data understanding, data preparation, modeling, evaluation, and deployment/development. The XGBoost model was then compared with Linear Regression as a comparison model using the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Mean Absolute Percentage Error (MAPE), and R^2 Score metrics and validated using 5-Fold Cross Validation. The test results on the test data showed that the XGBoost model produced an MAE of

1,120.71 people, an RMSE of 1,611.61 people, a MAPE of 2.30%, and an R^2 Score of 0.9972. Meanwhile, the results of 5-Fold Cross Validation produced an average MAE of 2,723.32 people, an RMSE of 5,500.84 people, a MAPE of 3.86%, and an R^2 Score of 0.9181. These results indicate that the XGBoost model is able to produce population predictions with a relatively low error rate and good performance on several data divisions.



Keywords: *Population Growth, Population Administration, Data Mining, XGBoost*