

ABSTRAK

Tingkat Pengangguran Terbuka (TPT) merupakan salah satu indikator penting dalam menilai kondisi ketenagakerjaan dan perekonomian suatu daerah. Tingginya tingkat pengangguran dapat menghambat pertumbuhan ekonomi dan menurunkan kesejahteraan masyarakat. Penelitian ini bertujuan untuk menganalisis faktor-faktor yang memengaruhi Tingkat Pengangguran Terbuka serta membangun model prediksi TPT pada kabupaten/kota di Provinsi Sumatera Utara menggunakan pendekatan Data Mining dengan algoritma Multiple Linear Regression. Data yang digunakan merupakan data sekunder yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Sumatera Utara periode 2019–2023, meliputi variabel jumlah penduduk, Produk Domestik Regional Bruto (PDRB), dan tingkat pendidikan dengan total 165 observasi. Tahapan penelitian meliputi pengumpulan data, pembersihan data, eksplorasi data, pembagian data training dan testing dengan rasio 80:20, pembangunan model, serta evaluasi menggunakan metrik MAE, RMSE, dan koefisien determinasi (R^2). Hasil penelitian menunjukkan bahwa tingkat pendidikan merupakan variabel yang berpengaruh signifikan terhadap TPT dengan indikasi terjadinya fenomena educated unemployment, sedangkan jumlah penduduk dan PDRB tidak berpengaruh signifikan secara parsial. Model yang dihasilkan memperoleh nilai R^2 sebesar 30,88%, MAE sebesar 1,76, dan RMSE sebesar 2,15, yang menunjukkan kemampuan prediksi yang cukup baik dan stabil. Hasil penelitian ini diharapkan dapat menjadi pendukung dalam pengambilan kebijakan ketenagakerjaan dan perencanaan pembangunan ekonomi daerah berbasis data di Provinsi Sumatera Utara.

ABSTRACT

The Open Unemployment Rate (OUR) is an important indicator for assessing employment conditions and regional economic performance. A high unemployment rate can hinder economic growth and reduce public welfare. This study aims to analyze the factors affecting the Open Unemployment Rate and develop a prediction model for regencies and municipalities in North Sumatra Province using a Data Mining approach with the Multiple Linear Regression algorithm. The study utilized secondary data obtained from the Central Statistics Agency (BPS) of North Sumatra for the period 2019–2023, consisting of population, Gross Regional Domestic Product (GRDP), and education level variables with a total of 165 observations. The research stages included data collection, data cleaning, exploratory analysis, data splitting into training and testing sets with an 80:20 ratio, model development, and evaluation using MAE, RMSE, and the coefficient of determination (R^2). The results indicate that education level is the only variable with a significant influence on the Open Unemployment Rate, reflecting the phenomenon of educated unemployment, while population and GRDP do not have significant partial effects. The developed model achieved an R^2 value of 30.88%, an MAE of 1.76, and an RMSE of 2.15, indicating satisfactory and stable predictive performance. The findings are expected to support employment policy formulation and data-driven regional economic development planning in North Sumatra Province.