

## ABSTRACT

### THE *k*-NEAREST NEIGHBOR (*k*-NN) REGRESSION METHOD IS USED TO PREDICT THE HUMAN DEVELOPMENT INDEX

By

Levia Febriyanti

The *k*-Nearest Neighbor (*k*-NN) method is one of the *Machine Learning* (ML) algorithms that can be used for classification and regression analysis. The *k*-NN method is considered a simple method to be applied in data analysis with multiple variable dimensions. In this research, the *k*-NN *regression* method was chosen with the aim of predicting the best *k* value and *Mean Absolute Percentage Error* (MAPE) on Human Development Index (HDI) in Indonesia for 2022. The data used in this research consisted of HDI, life expectancy index, mean years of schooling, expected years of schooling, and gross national index per capita from 2020-2021 of 34 provinces in Indonesia. The results of this research indicate that the optimal value of *k* is  $k = 2$  with an MAPE error of 0,9019%. The average HDI prediction in Indonesia for 2022 is 71,41.

**Key Words:** Regression analysis, multiple variable, *k*-Nearest Neighbor (*k*-NN) *Regression*, *Mean Absolute Percentage Error* (MAPE).

## ABSTRAK

### METODE *k*-NEAREST NEIGHBOR (*k*-NN) REGRESSION DALAM MEMPREDIKSI INDEKS PEMBANGUNAN MANUSIA

Oleh

Levia Febriyanti

Metode *k*-Nearest Neighbor (*k*-NN) merupakan salah satu algoritma *Machine Learning* (ML) yang dapat digunakan untuk klasifikasi dan analisis regresi. Metode *k*-NN dianggap sebagai metode yang sederhana untuk diterapkan dalam analisis data dengan dimensi peubah banyak. Pada penelitian ini metode *k*-NN *Regression* dipilih dengan tujuan untuk memprediksi nilai *k* dan *Mean Absolute Percentage Error* (MAPE) terbaik pada Indeks Pembangunan Manusia (IPM) di Indonesia tahun 2022. Data yang digunakan pada penelitian ini terdiri dari data IPM, angka harapan hidup, rata-rata lama sekolah, harapan lama sekolah, dan pengeluaran perkapita riil pertahun yang disesuaikan dari tahun 2020-2021 sebanyak 34 provinsi di Indonesia. Hasil penelitian ini menunjukkan bahwa nilai *k* yang optimal adalah  $k = 2$  dengan galat MAPE 0,9019% diperoleh hasil prediksi IPM di Indonesia pada tahun 2022 rata-rata sebesar 71,41.

**Kata Kunci:** Analisis Regresi, Peubah Banyak, *k*-Nearest Neighbor (*k*-NN) *Regression*, *Mean Absolute Percentage Error* (MAPE).