

ABSTRAK

Kesehatan merupakan derajat kemanusiaan yang paling mendasar dan tanpa kesehatan, manusia tidak bisa mendapatkan hak-hak yang lain. Oleh karena itu, negara wajib memenuhi dan menjamin hak kesehatan warga negaranya. Dalam memenuhi kewajibannya, pemerintah mencanangkan program Pembangunan Kesehatan pada tahun 2020 dengan Umur Harapan Hidup (UHH) sebagai salah satu indikator utama penilaian kinerjanya. Berdasarkan indikator utama tersebut, maka penting bagi pemerintah untuk mengetahui variabel-variabel yang mempengaruhi UHH. Selain itu, agar program Pembangunan Kesehatan berjalan lancar dan tepat sasaran, pemerintah perlu mengelompokkan kabupaten/kota di Indonesia berdasarkan tingkat kesehatan. Untuk memenuhi tujuan tersebut, maka penelitian ini menggunakan metode Regresi Linear Berganda untuk mengetahui variabel-variabel yang mempengaruhi UHH dan *Clustering* K-Means untuk mengelompokkan seluruh kabupaten/kota di Indonesia berdasarkan tingkat kesehatan serta Regresi Random Forest yang digunakan untuk perbandingan efektivitas model. Setelah dilakukan Regresi Linear Berganda, diketahui bahwa seluruh variabel indikator kesehatan, baik masing-masing maupun bersama-sama, berpengaruh dan signifikan terhadap UHH serta jumlah *cluster* optimal yang dihasilkan K-Means sebanyak 6. Meskipun begitu, hanya sebagian kecil nilai UHH yang dapat dijelaskan dengan variabel indikator kesehatan sehingga dilakukan kembali regresi baik keseluruhan maupun per cluster dengan metode Random Forest. Proporsi UHH yang dijelaskan oleh variabel indikator kesehatan meningkat tetapi hasilnya belum maksimal karena diduga adanya *noise* pada fitur-fitur kesehatan yang digunakan.

Kata Kunci: *Pembangunan Kesehatan, Umur Harapan Hidup (UHH), Regresi Linear Berganda, Clustering K-Means, Regresi Random Forest*

ABSTRACT

Health is a basic part of recognizing the degree of humanity and without it, a person will be unable to obtain his or her other rights. Therefore, the state has a responsibility to fulfill and protect the health rights of citizens. In fulfilling its obligations, the government launched the Health Development program in 2020 with Life Expectancy as one of the main indicators of its performance assessment. Based on these main indicators, it is important for the government to understand the variables that affect Life Expectancy. In addition, in order for the Health Development program to run smoothly and on target, the government needs to group districts/cities in Indonesia based on its health level. To meet these objectives, this study uses multiple linear regression to determine the variables that affect Life Expectancy and K-Means Clustering to group all districts/cities in Indonesia based on its health level and also Random Forest to compare the effectiveness of the models. Using Multiple Linear Regression, it was found that all health indicator variables, either individually or simultaneously, had a significant effect on Life Expectancy and the optimal number of clusters produced by K-Means was 6. However, only a small proportion of Life Expectancy could be explained by health indicator variables so the regression is continued for both overall and in each cluster using the Random Forest method. The proportion of Life Expectancy explained by the health indicator variable increased but the results has not reached the optimum value because it was suspected that there was noise in the health features used.

Keywords: Health Development, Life Expectancy, Multiple Linear Regression, K-Means Clustering, Random Forest Regression