

Prediksi Usia Biologis Menggunakan Metode Support Vector Regression dan Klemara-and-Doubal pada Data Pemeriksaan Medis = Biological Age Prediction Using the Support Vector Regression and Klemara-and-Doubal Methods on Medical Examination Data

Angelica Patricia Djaya Saputra, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920539071&lokasi=lokal>

Abstrak

Penuaan biologis mencerminkan kondisi kesehatan fisik yang sebenarnya karena menilai fungsi organ dan sistem tubuh yang sebenarnya pada setiap individu, berbeda dengan usia kronologis. Penelitian ini mengeksplorasi prediksi usia biologis menggunakan metode Support Vector Regression (SVR) dan Klemara-and-Doubal Method (KDM), yang berfokus pada pengaruh biomarker dan faktor eksternal pada proses penuaan. Pembangunan model memanfaatkan data pemeriksaan medis dari Kementerian Kesehatan Indonesia pada tahun 2011 dimana keterbaruan dari penelitian ini adalah melibatkan semua fitur yang berpengaruh terhadap usia biologis, termasuk faktor eksternal, tidak hanya biomarker saja. Kemudian, dilakukan pemanfaatan seluruh dataset tanpa membedakan subjek sehat dan tidak sehat. Pada dataset dilakukan data preprocessing agar dataset siap digunakan dengan melakukan filtering usia di atas 30 tahun, pemisahan dataset pria dan wanita, menghapus fitur yang tidak relevan, mengubah tipe data yang tidak sesuai, mengidentifikasi dan melakukan penanganan missing value serta outliers, dan melakukan encoding untuk data beripe kategorikal. Kemudian, dilakukan feature selection dengan menggunakan Spearman's rank Coefficient Corelation dan pembangunan model SVR dan KDM. Hasil penelitian menunjukkan bahwa terpilih 5 fitur untuk pria dan 6 fitur untuk wanita yang digunakan untuk membangun model SVR dan KDM. KDM menunjukkan performa evaluasi yang cukup baik dalam interpretasi variasi data dengan skor performa RMSE 1,39; R2 0,97; dan Adjusted R2 0,97 untuk pria dan RMSE 1,00; R2 0,99; dan Adjusted R2 0,99 untuk wanita. Metode ini lebih unggul daripada SVR yang cenderung menunjukkan performa yang kurang memuaskan dimana memiliki skor performa RMSE 6,36; R2 0,44; dan Adjusted R2 0,36 untuk pria dan RMSE 5,90; R2 0,57; dan Adjusted R2 0,53 untuk wanita. Berdasarkan hasil analisis dari berbagai teknik analisis yang dilakukan (analisis evaluasi performa, analisis hubungan usia kronologis dengan usia biologis, dan analisis evaluasi dengan melihat pola hasil estimasi) terlihat bahwa metode KDM lebih unggul dalam memprediksi usia biologis dibandingkan dengan SVR, terutama dalam hal konsistensi dan akurasi. Selain itu, analisis hubungan setiap fitur dengan usia biologis untuk tiap model menggambarkan pengaruh fitur-fitur tersebut terhadap fungsi organ tubuh seseorang.

.....The biological aging reflects the actual physical health condition as it assesses the real function of organs and body systems in each individual, different from chronological age. This research explores the prediction of biological age using the Support Vector Regression (SVR) method and the Klemara-and-Doubal Method (KDM), focusing on the influence of biomarkers and external factors on the aging process. The model development utilized medical examination data from the Indonesian Ministry of Health in 2011, where the novelty of this research is involving all features that affect biological age, including external factors, not just biomarkers. Then, the entire dataset was utilized without distinguishing between healthy and unhealthy subjects. In the dataset, data preprocessing was performed to make the dataset ready to use by filtering ages above 30 years, separating datasets for men and women, removing irrelevant features,

changing inappropriate data types, identifying and handling missing values and outliers, and encoding for categorical data. Subsequently, feature selection was conducted using Spearman's Rank Coefficient Correlation, and then the SVR and KDM models were built. The research results showed that 5 features for men and 6 features for women were selected to build the SVR and KDM models. KDM showed fairly good evaluation performance in interpreting data variations with performance scores of RMSE 1.39, R^2 0.97, and Adjusted R^2 0.97 for men and RMSE 1.00, R^2 0.99, and Adjusted R^2 0.99 for women. This method outperformed SVR, which tended to show less satisfactory performance with performance scores of RMSE 6.36, R^2 0.44, and Adjusted R^2 0.36 for men and RMSE 5.90, R^2 0.57, and Adjusted R^2 0.53 for women. Based on the analysis results from various techniques performed (performance evaluation analysis, analysis of the relationship between chronological age and biological age, and evaluation analysis by looking at the pattern of estimation results), it appears that the KDM method is superior in predicting biological age compared to SVR, especially in terms of consistency and accuracy. In addition, the analysis of the relationship of each feature with biological age for each model illustrates the influence of these features on the organ function of an individual.