

Analisis data indeks standar pencemaran udara di Jakarta tahun 2019 menggunakan imputasi dan clustering = Analysis of air pollution index in Jakarta in 2019 using imputation and clustering

Ahmad Syukron Alimudin, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20515783&lokasi=lokal>

Abstrak

Data Indeks Standar Pencemaran Udara di Jakarta tahun 2019 menunjukkan seberapa jauh kualitas udara yang dihasilkan tiap pengamatan harian. Variabel Indeks Standar Pencemaran Udara berdasarkan KEP-45/MENLH/10/1997 adalah partikulat berukuran 10 μm (PM10), Sulfur Dioksida (SO₂), Karbon Monoksida (CO), Oksidan dalam bentuk Ozon (O₃), dan Nitrogen Dioksida (NO₂). Unit observasi berdasarkan SPKU yang terdiri dari SPKU DKI1 Bundaran HI, SPKU DKI2 Kelapa Gading, SPKU DKI3 Jaga Karsa, SPKU DKI4 Lubang Buaya, dan SPKU DKI5 Kebon Jeruk. Adanya missing value pada data ISPU, maka dilakukan imputasi nol dan mean untuk mengganti missing value. Observasi harian ISPU tiap SPKU memiliki tiga tipe data, kemudian dikelompokkan menggunakan K-Means Clustering. Tiga tipe data tiap SPKU tersebut, meliputi data tanpa imputasi, data dengan imputasi nol, dan data dengan imputasi mean, kemudian diperoleh 5 cluster tiap masing-masing tipe data. Pada akhirnya, metode imputasi mean dapat diimplementasikan sebagai metode imputasi missing values pada data ISPU dengan proporsi missing values 3,83%, 4,10%, 10,68 %, 11,78 %, dan 15,06 %, sedangkan penerapan imputasi nol kurang cocok pada data ISPU, karena jaraknya terlalu jauh dari nilai ISPU terhitung serta dipengaruhi oleh missing value. Di sisi lain, hasil penelitian juga menunjukkan bahwa metode K-Means Clustering ini optimal dalam pengelompokan data berdasarkan Between SS/Total SS berkisar antara 67,5 %-85,2 %.

.....Air Pollution Standard Index data in Jakarta in 2019 shows how far each daily observation produces the air quality. Air Pollution Standard Index Variables based on KEP-45/MENLH/10/1997 are particulates measuring 10 m (PM10), Sulfur Dioxide (SO₂), Carbon Monoxide (CO), Oxidants in the form of Ozone (O₃), and Nitrogen Dioxide (NO₂). The observation unit based on the SPKU consists of SPKU DKI1 Bundaran HI, SPKU DKI2 Kelapa Gading, SPKU DKI3 Jaga Karsa, SPKU DKI4 Lubang Buaya, and SPKU DKI5 Kebon Jeruk. If there is a missing value in ISPU data, zero and mean imputation is carried out to replace the missing value. ISPU daily observations of each SPKU have three data types, then grouped using K-Means Clustering. The three data types for each SPKU, including data without imputation, data with zero imputation, and data with mean imputation, then obtained 5 clusters for each data type. In the end, the mean imputation method can be implemented as a method for imputing missing values on ISPU data with the proportion of missing values being 3.83%, 4.10%, 10.68%, 11.78%, and 15.06%, while the implementation of imputation is zero does not match the ISPU data, because the distance is too far from the calculated ISPU value and is influenced by the missing value. On the other hand, the results also show that the K-Means Clustering method is optimal in grouping data based on Between SS/Total SS ranging from 67.5%-85.2%.