

## ABSTRAK

Nama : Hady Satria  
NIM : 41518110117  
Pembimbing TA : Eliyani, Dr. Ir.  
Judul : Implementasi Streamlit untuk Mengklasifikasikan Tingkat Pencemaran Udara di DKI Jakarta dengan Penggunaan Algoritma Regression

Isu pencemaran udara cukup menjadi hal penting yang mengaitkannya dengan pertumbuhan hidup di perkotaan, isu ini biasa dikenal dengan polusi udara. Kondisi udara di Jakarta menunjukkan selama periode 2017-2020 umumnya udara dikategorikan tingkat sedang hingga tidak sehat. Hal ini diukur berdasarkan Indeks Standar Pencemaran Udara (ISPU) mengklasifikasikan kualitas udara di suatu daerah berdasarkan parameter zat-zat kimia serta akibat terhadap kesehatan. klasifikasi kualitas udara mulai dari kategori baik, sedang, tidak sehat, sangat tidak sehat, dan berbahaya, dapat diketahui dengan pemanfaatan algoritma tipe regression yaitu Logistic Regression (Regresi Logistik) dan Random Forest. Penelitian ini bermaksud untuk menghandalkan tipe algoritma regression dengan merancang dashboard untuk mengklasifikasikan tingkat pencemaran udara di DKI Jakarta dengan library Streamlit. Selain itu, difungsikan sebagai validasi terhadap algoritma yang telah di training terlebih dahulu mengenai tingkat evaluation dan keakuratannya. Dalam permasalahan klasifikasi indeks standar pencemar udara di DKI Jakarta menggunakan algoritma regression. Hasil *evaluation modelling*, untuk nilai sesuai berdasarkan proposi data *split* 80% : 20% dan 60% : 40% menunjukkan bahwa secara *accuracy* lebih unggul algoritma *Random Forest* sebesar 96,66% & 96,22%. Dengan nilai MEA, MSE, RMSE terkecil secara berurut 0,03866, 0,04920, 0,2218. untuk data *split* 80% : 20%, kemudian nilai 0,045, 0,0615, 0,2480, pada data *split* 60% : 40%. Untuk nilai evaluasi r-square lebih besar yaitu 0.89910 & 0.87777 yang menunjukkan model yang dikeluarkan oleh regresi lebih baik. Dilakukan validitas menggunakan streamlit untuk menguji dataset terbaru dengan hasil menunjukkan variabel *independent* yang diinput dapat diklasifikasikan dengan hasil modeling algoritma *regression*.

Kata kunci:

Algoritma Regression; Indeks Standar Pencemar Udara; Streamlit

## ABSTRACT

Name : Hady Satria  
Student Number : 41518110117  
Counsellor : Eliyani, Dr. Ir  
Title : Streamlit Implementation to Classify Air Pollution Levels in DKI Jakarta Using Regression Algorithms

The issue of air pollution is quite an important thing that relates to the growth of life in urban areas, this issue is commonly known as air pollution. Air conditions in Jakarta show that during the 2017-2020 period, the air is generally categorized as moderate to unhealthy. This is measured based on the Air Pollution Standard Index (ISPU) which classifies the air quality in an area based on the parameters of chemical substances and their effects on health. Air quality classification ranging from good, moderate, unhealthy, very unhealthy, and dangerous categories, can be identified by using the regression type algorithm, namely Logistic Regression and Random Forest. This study intends to rely on the type of regression algorithm by designing a dashboard to classify the level of air pollution in DKI Jakarta using the Streamlit library. In addition, it functions as a validation of the algorithm that has been previously trained regarding the evaluation level and its accuracy. In the problem of classifying the air pollutant standard index in DKI Jakarta using a regression algorithm. The results of the modeling evaluation, for the appropriate value based on the proportion of split data 80%: 20% and 60%: 40%, show that the Random Forest algorithm is superior to the accuracy of 96.66% & 96.22%. With the smallest MEA, MSE, RMSE values, respectively, 0.03866, 0.04920, 0.2218. for split data 80% : 20%, then the value is 0.045, 0.0615, 0.2480, on split data 60% : 40%. For the evaluation value of r-square is greater, namely 0.89910 & 0.87777 which indicates the model issued by the regression is better. Validity using streamlit to test the latest dataset with results showing the inputted independent variables can be classified with the results of the regression algorithm modeling.

Key words:

Regression Algorithm; Air Pollutant Standard Index; Streamlit