

**ABSTRAK**

Judul : PERBANDINGAN ALGORITMA KLASIFIKASI *DECISION TREE*, *NAÏVE BAYES*, DAN *K-NEAREST NEIGHBOR* DALAM MENENTUKAN KUALITAS UDARA DI DKI JAKARTA.  
Nama : Mohammad Barata Putra Gusti  
Program Studi : Teknik Informatika

DKI Jakarta termasuk ke dalam 20 besar kota berpolusi di dunia. Polusi ini diakibatkan terkontaminasinya udara bersih dengan zat berbahaya yang disebut polutan. Unsur polutan pada negara Asia menggunakan O<sub>3</sub>, PM<sub>10</sub>, PM<sub>2.5</sub> AQI, NO<sub>2</sub>/NO<sub>X</sub>, SO<sub>2</sub>, CO. Indeks standar pencemaran udara (ISPU) merupakan rentang nilai yang digunakan sebagai penentu klasifikasi kualitas udara. Tujuan penelitian ini, ingin melakukan prediksi kualitas udara di DKI Jakarta menggunakan *dataset* ISPU bersumber dari Jakarta Open Data dan algoritma *Decision Tree Classification and Regression Tree* (CART), *Naïve Bayes*, *K-Nearest Neighbor* (k-NN) dengan atribut PM<sub>10</sub>, CO, SO<sub>2</sub>, O<sub>3</sub>, NO<sub>2</sub> dan target/class BAIK, SEDANG, TIDAK SEHAT dan SANGAT TIDAK SEHAT. Metodologi yang digunakan yaitu *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Dalam *data preparation* dilakukan penanganan *missing values* menggunakan *k-Nearest Neighbor Imputation* skema 1NN dan *outlier* menggunakan *winsorizing*. *Stratified k-fold cross validation* diterapkan pada *training model* untuk menghindari *overfitting* dengan hasil akurasi *Decision Tree* sebesar 99.8%, *Naïve Bayes* 89% dan k-NN akurasi tertinggi 94.6% dan terendah 92%. Hasil evaluasi model menggunakan *confusion matrix* mendapatkan hasil akurasi *Decision Tree* yang sama dengan *training model* sebesar 99.8% dan error 0.2%, *Naïve Bayes* 91% akurasi 9% error dan *k-Nearest Neighbor* 95% akurasi 5% error. Kesimpulan dari penelitian yaitu model yang dihasilkan dari 3 (tiga) algoritma sangat baik dalam melakukan klasifikasi kualitas udara.

Kata kunci : CRISP-DM, ISPU, *Decision Tree*, *K-Nearest Neighbor*, *Naïve Bayes*

**ABSTRACT**

Title : *DECISION TREE CLASSIFICATION AND REGRESSION TREE CLASSIFICATION ALGORITHM IN DETERMINING AIR QUALITY IN DKI JAKARTA*  
Name : Mohammad Barata Putra Gusti  
Study Program : *Technical Infomation*

One of the world's top 20 cities with the highest pollution is DKI Jakarta. Contamination of clean air with harmful substances causes pollution, which is known as a pollutant. *Pollutant elements in Asian countries use O<sub>3</sub>, PM<sub>10</sub>, PM<sub>2.5</sub> AQI, NO<sub>2</sub>/NO<sub>X</sub>, SO<sub>2</sub>, and CO. The air pollution standard index (ISPU) is a range of values used to determine air quality classification. The purpose of this study is to predict air quality in DKI Jakarta using the ISPU dataset sourced from Jakarta Open Data and the Decision Tree Classification and Regression Tree (CART) algorithm, Naïve Bayes, K-Nearest Neighbor (k-NN) with attributes PM<sub>10</sub>, CO, SO<sub>2</sub>, O<sub>3</sub>, NO<sub>2</sub> and target/class GOOD, MEDIUM, UNHEALTHY and VERY UNHEALTHY. The methodology used is Cross-Industry Standard Process for Data Mining (CRISP-DM). In data preparation, handling missing values using k-Nearest Neighbor Imputation 1NN scheme and outliers using winsorizing. Stratified k-fold cross validation is applied to the training model to avoid overfitting with Decision Tree accuracy of 99.8%, Naïve Bayes at 89% and k-NN with the highest accuracy of 94.6% and the lowest 92%. The model evaluation results using the confusion matrix get the same Decision Tree accuracy as the training model of 99.8% and 0.2% error, Naive Bayes 91% accuracy 9% error and k-Nearest Neighbor 95% accuracy 5% error. The study concludes that the model generated from 3 (the three) algorithms is very good in classifying air quality.*

Keywords : *CRISP-DM, ISPU, Decision Tree, K-Nearest Neighbor, Naïve Bayes*