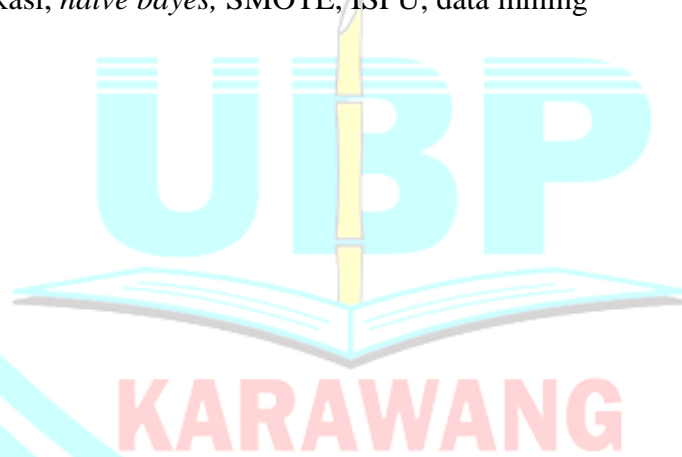


ABSTRAK

Pencemaran udara di DKI Jakarta menjadi sebuah isu penting dan berdampak negatif bagi kesehatan masyarakat. Penelitian ini menerapkan algoritma naive Bayes untuk mengklasifikasikan kualitas udara, dengan teknik SMOTE digunakan untuk mengatasi ketidakseimbangan data. Data yang dianalisis berasal dari data indeks pencemaran udara tahun 2022 hingga 2024, yang diambil dari lima stasiun pemantau udara di Jakarta. Proses analisis dilakukan mengikuti tahapan CRISP-DM, mulai dari memahami masalah hingga mengevaluasi model. Hasil penelitian menunjukkan bahwa SMOTE berhasil meningkatkan akurasi prediksi pada kelas yang lebih sedikit. Tanpa SMOTE, akurasi model mencapai 90% tetapi tampak bias terhadap kelas yang lebih sedikit, dengan nilai recall hanya 0,75 dan presisi 0,62. Sedangkan SMOTE, akurasi model menjadi 88%, dengan nilai presisi 0,86, recall 0,87, dan f1-score 0,87, yang menunjukkan hasil yang lebih seimbang di seluruh kelas.

Kata Kunci: klasifikasi, *naive bayes*, SMOTE, ISPU, data mining



ABSTRACT

Air pollution in DKI Jakarta is an important issue and has a negative impact on public health. This study applies the naive Bayes algorithm to classify air quality, Utilizing the SMOTE technique effectively addresses the issue of data imbalance. The data analyzed came from air pollution index data from 2022 to 2024, taken from five air monitoring stations in Jakarta. The analysis process was carried out following the CRISP-DM stages, starting from understanding the problem to evaluating the model. The results showed that SMOTE succeeded in increasing prediction accuracy in fewer classes. Without SMOTE, the model accuracy reached 90% but appeared biased towards fewer classes, with a recall value of only 0.75 and a precision of 0.62. While SMOTE, the model accuracy became 88%, with a precision value of 0.86, recall 0.87, and f1-score 0.87, which showed more balanced results across classes.

Keyword: *classification, naïve bayes, SMOTE, ISPU, data mining*

