

## ***ABSTRACT***

*Life in big cities cannot be separated from the problem of air pollution. Inhaled air will directly affect health such as respiratory disorders. Air quality in cities is also getting worse with increasing population, traffic, industrialization, and the energy we use. Although air pollution monitoring already exists in Indonesia, the places being monitored are still less spread out and only exist in a few points. In addition, the level of air pollution classification has no international standard.*

*The software proposed in this research is one of two final projects with the aim of making a monitoring and classification tool for air pollution. The algorithm used in the classification is Random Forest (RF), besides that there is data preprocessing first using Multiple Imputation by Chained Equations (MICE) for imputing missing data and using Synthetic Minority Oversampling Technique (SMOTE) to overcome imbalanced datasets.*

*The results of this study are to create software consisting of air pollution classification using machine learning, receive data from a multi-point power-saving microcontroller with a wireless mesh network, forward the data to the Antares server and the results of the classification can be viewed using the Android application. The machine learning model performance was measured using the geometric mean score ( $G - Mean$ ), specificity and sensitivity with the results of 99.08% for  $G - Mean$ , 99.6% for specificity, and 99.69% for sensitivity. The model that has been trained takes an average of  $17,025 \pm 2,928$  ms, while the time it takes to send data from the PC to the Antares server is  $81.61 \pm 8.88$  ms for the size of 232 Bytes.*

**Key Words:** *Monitoring, Machine Learning, Imbalance Data.*