

AIR QUALITY IN YOGYAKARTA DURING THE COVID-19 PANDEMIC

USING NAIVE BAYES AND SUPPORT VECTOR MACHINE

Air quality is essential for humans and the environment. Living things need clean air to minimize the probabilities of getting a disease. However during Covid-19, the virus has spread to various cities which polluted the air making people infected by the coronavirus. Yogyakarta is one of the biggest cities that has been affected by air pollution before and during the pandemic. Using Naive Bayes and SVM this study wanted to predict if the air quality is good or bad by comparing the different amounts of components in the air.

1. Variables

- Particulate Matter (PM10) as predictor variable (X variable)
- Sulfur Dioksida (SO2) as predictor variable (X variable)
- Karbon Monoksida (CO) as predictor variable (X variable)
- Ozon (O3) as predictor variable (X variable)
- Natrium Dioksida (NO2) as predictor variable (X variable)
- Category as response variable (Y variable)

2. Naive Bayes

Naive Bayes is a classification method using probability and statistical methods. The Naive Bayes method assumes that the attributes in the data have no correlation with each other, which makes them independent. The equation used in Naive Bayes is as follows:

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

Y = Particular class
X = Data of an undefined class
P(X|Y) = likelihood of a condition-dependent hypothesis
P(Y|X) = Probability of a class based on a hypothetical condition
P(Y) = Probability of Y
P(X) = Probability of X

3. Support Vector Machine

Support vector machine (SVM) is a supervised learning method in machine learning algorithms that use classification and regression. The basic form of SVM is a binary linear classification that wants to identify a boundary between two classes. SVM has a goal to solve problems on a global scale by processing data to find a hyperplane.

4. Comparison of models

Comparison	Training			Testing		
	Accuracy	Precision	Recall	Accuracy	Precision	Recall
Naive Bayes	0.97	1.00	0.84	0,96	1.00	0.81
		0.96	1.00		0.95	1.00
Support Vector Machine	1.00	1.00	0.98	0.95	0.94	0.81
		1.00	1.00		0.96	0.99

Conclusion

We can conclude that the best classification method to use on determining the predicted air quality in Yogyakarta is the Support Vector Machine model. The SVM model has a more accurate result than the Naive Bayes model using the same dataset. We can see the SVM model has almost a perfect f1-score for each category of air quality with an accuracy on data training is 1.00 and an accuracy on data testing is 0.95 which is higher than the Naive Bayes model.