Electronic Integrated Computer Algorithm Journal
Vol. 2, No. 2, April 2025, pp 105~124

ISSN: 3031-0350, DOI: 10.62123/enigma.v2i2.61

Quality Classification of Air Quality in Medan Industrial Area Using Naïve Bayes Method

Zhahrah Zhafirah[1], Al-Khowarizmi[1*]

[1] *Department of InformationTechnology, Universitas Muhammadiyah Sumatera Utara,* Indonesia.

DOI : 10.62123/enigma.v2i2.61

Received : March 09, 2025
Revised : March 21, 2025
Accepted : March 25, 2025

Keywords:
Classification, Naive Bayes, Data mining, Air Pollution Standard Index, Air

ABSTRACT

Advances in information technology have affected various aspects of life, including efforts to monitor air quality. Clean air is a basic human need, but technological developments and increased industry and the number of motorized vehicles have caused a decline in air quality. Air pollution has various negative impacts, including health problems and global warming. To help the community and government in monitoring air quality, this study implements a data mining method with a classification technique using the Naïve Bayes Algorithm. This method was chosen because of its effective ability to predict air quality based on historical data. This study uses data from the Air Pollution Standard Index (ISPU) parameters to build a classification model that can separate air quality categories, such as Good, Moderate, Unhealthy, Very Unhealthy, and Hazardous. The results of the study are expected to provide accurate information to the public about air quality in KIM, as well as assist the government in efforts to control air pollution.

1. INTRODUCTION

The progression of information technology has impacted every facet of existence, encompassing economics, politics, art, culture, and even the realm of education. An optimal environment constitutes the fundamental requirement for the sustenance of human life; thus, it is imperative that individuals have access to clean air [1]. In the present technological era, the availability of clean air is scarce due to extensive industrial advancements and a rising volume of cars. The rapid advancement of computer technology can enhance human productivity [2], exemplified by the creation of an air quality classification system that enables individuals to monitor the quality of the air they inhale daily. Although air is an invisible substance, the repercussions of air pollution can be perceived instantaneously when contamination occurs [3].

Air pollution refers to the presence of chemical, physical, or biological pollutants in the air that can adversely affect living organisms and the environment. The concerns induced by this pollution include health difficulties and a large increase in air temperature, promoting global warming. This commonly occurs in large cities or industrial locations that often discharge waste gasses from operations. In addition, the increasing number of motorized vehicles in urban locations [4].

2. LITERATURE REVIEW

2.1 Data Mining

The process of collecting important information from a collection of data is called data mining. Statistics, databases, machine learning, pattern recognition, artificial intelligence, and visualization all play a role in data mining. The data to be processed by data mining must follow the Knowledge Discovery in Databases (KDD) flow and have become 2 parts of data, namely training data and testing data. The Pareto principle is a principle that believes that 80% of a person's performance results are the result of 20% of the efforts that have been made. The division of training data is 80% of the data set and test data is 20% of the data set.

The findings of the experimental work conducted in this study showed that both the proposed weighted classifiers perform better than the traditional Naive Bayes, Support Vector Machines and Neural Network classifiers with respect to various performance metrics- accuracy, average precision, average recall, error rate and F1 score. Further, this study depicts that Covariance based Weighted Naive Bayes and Convergent Cross Mapping based Weighted Naive Bayes have an average accuracy of 83.6% and 82.12% respectively [5].

2.2 Naïve Bayes

Naïve Bayes is a widely used probabilistic classification algorithm that employs Bayes' theorem, assuming independence among predictors. Its simplicity and efficiency make it suitable for various applications, including text classification, spam filtering, and medical diagnosis [6]. The algorithm's strength lies in its ability to handle large datasets and deliver rapid predictions, making it an attractive choice for real-time applications, such as air quality monitoring.

*Corresponding Author Email: alkhowarizmi@umsu.ac.id

Naïve Bayes is a widely recognized algorithm for classification tasks due to its simplicity and effectiveness. It operates under the assumption of attribute independence, which allows for efficient computation of probabilities [7]. Despite the often-unrealistic nature of this assumption in real-world datasets, Naïve Bayes has demonstrated strong performance in various applications, including text classification, spam detection, and environmental monitoring.

The Naive Bayes algorithm is a classification method that utilizes probability and statistics to compare training data with testing data. This process involves several stages of equations to obtain the highest probability, which is then considered as the resulting information. The Naive Bayes algorithm can be applied to the classification of data that is continuous, such as numeric data, or categorical. In addition, this algorithm is able to model or directly calculate continuous data attributes. In the context of classifying air quality in the Medan Industrial Area (KIM), Naive Bayes is used to test Air Pollution Standard Index (ISPU) data with the aim of evaluating classification performance[8]. Naïve Bayes algorithm is effective for classification tasks, achieving satisfactory results despite its simplicity and the assumption of attribute independence. This supports the use of Naïve Bayes in air quality classification, where similar principles of categorization can be applied [9].

2.3 Classification

Classification is the process of finding a function that describes a class of data, with the aim of being able to estimate the class of an object or data whose label or value is unknown. To achieve this goal, the classification process forms a model or algorithm that is intended to be able to distinguish data into different classes based on certain functions. The model itself can be an "if-then" rule, a decision tree, or a mathematical formula [10]. Classification is a supervised learning technique in machine learning where the goal is to predict the categorical label of new observations based on past observations with known labels [11].

2.4 Confusion Matrix

Confusion matrix is a table matrix used as a performance calculation of a data model or algorithm. Each row of the matrix represents the actual class of the data, and each column represents the predicted class of the data (or vice versa). the confusion matrix is utilized to analyze the performance of classifiers in hypothetical experiments related to target detection, illustrating how well the classifier distinguishes between the presence and absence of a target [12].

2.5 Air

The presence of air has a very influential role in maintaining the survival of all living things, especially in the provision of oxygen. Oxygen is produced through the process of photosynthesis by plants and algae, which absorb carbon dioxide (CO2). For living things, oxygen functions as a substance needed in the respiration process. Every day, we need to breathe air as a necessity. Categorically, air can be divided into two categories, namely clean air and unclean air. One of innovative Bayesian approach for personalized decision-making regarding air quality, employing a hierarchical spatial-temporal model that incorporates detailed high-resolution data from a mobile air quality sensor, thereby providing Bayes-optimal journey decision support tailored to individual user preferences in urban environments [13].

2.6 Air Pollution

Air pollution or air pollution is part of environmental pollution which is currently increasing. Pollution is the entry of pollutants in the form of substances or gases into an environment so that it can reduce the quality of the environment. While what is said to be a pollutant is a substance or material whose content exceeds the limit and is at the wrong time and place, so that it is an environmental pollutant, for example: chemicals, dust, heat and noise. Air pollution or air pollution is the event of the entry, or mixing, of pollutants (hazardous elements) into the air layer (atmosphere) which can result in a decrease in air quality (environment). Air pollution can occur anywhere, both indoors such as homes, offices or closed rooms, or outdoors such as in cities, highways and so on. Usually, pollutants that pollute the air are in the form of gas and smoke. This pollution can come from the results of the process of incomplete fuel combustion, smoke from factory chimneys, power plants to motor vehicle smoke. The pollutants in the form of gas and smoke are the result of oxidation of various elements that make up the fuel, namely: CO2 (carbon dioxide), CO (carbon monoxide), SOx (sulfur oxide) and NOx (nitrogen oxide). Bayesian network (BN) analysis model for predicting air quality index and warning the air pollution risk at the city level. Further, a two-layer BN for analyzing influencing factors of various air pollutants is developed [14].

2.7 Python

Python is known for its simplicity and readability, making it accessible for both beginners and experienced programmers. This ease of use is beneficial when implementing complex algorithms like Naïve Bayes. The implementation of the Naïve Bayes classifier in Python demonstrated high accuracy (98% in the study) for classifying messages as spam or non-spam, showcasing Python's effectiveness for machine learning tasks [15].

2.8 Flowchart

Flowchart is a procedural stage of a program that is depicted using a graph, in general flowchart is applied to explain the processing flow. Flowchart contains a work step chart that shows the flow of a process that is depicted using symbols that are arranged systematically from the entire system [16].

3. RESULTS AND DISCUSSIONS

3.1 Data Collection

ISPU data in this study was compiled by the Medan Environmental Service. Overall, this dataset contains 100 data records with 10 attributes and 1 class. The data used is data from March to June. The attributes contained in this dataset include date, station, pm10, pm25, so2, co, o3, no2, max, critical, and category as in table 1 below:

Table 1. ISPU Data on March 2023

Date	PM_{10}	$PM_{2.5}$	SO_2	CO	O_3	NO_2	Max	Critical	Category	Station
1-03-2023	23	21	113	1947	73	140	1947	CO	Hazardous	KIM
2-03-2023	34	32	106	1231	49	82	1231	CO	Hazardous	KIM
3-03-2023	19	17	79	1162	44	141	1162	CO	Hazardous	KIM
4-03-2023	16	15	74	2335	44	141	2335	CO	Hazardous	KIM
5-03-2023	25	24	136	3362	101	79	3362	CO	Hazardous	KIM
6-03-2023	31	28	88	5275	116	133	5275	CO	Hazardous	KIM
7-03-2023	53	49	41		1	36	53	PM_{10}	Moderate	KIM
8-03-2023	53	49	44			39	53	PM_{10}	Moderate	KIM
9-03-2023	67	64	54	5203	227	121	5203	CO	Hazardous	KIM
10-03-2023	40	38	148	3603	139	115	3603	CO	Hazardous	KIM
11-03-2023	21	18	19			20	21	PM_{10}	Good	KIM
12-03-2023	29	26	131	3381	169	100	3381	CO	Hazardous	KIM
13-03-2023	31	28	170	5585	225	144	5585	CO	Hazardous	KIM
14-03-2023	40	37	141	3723	133	132	3723	CO	Hazardous	KIM
15-03-2023	36	33	175	6180		136	6180	CO	Hazardous	KIM
16-03-2023	36	33	175	6180		136	6180	CO	Hazardous	KIM
17-03-2023	41	38		7156		164	7156	CO	Hazardous	KIM
18-03-2023	37	34		7613		161	7613	CO	Hazardous	KIM
19-03-2023	41	38	177	6453		140	6453	CO	Hazardous	KIM
20-03-2023	45	42	170	6090	217	135	6090	CO	Hazardous	KIM
21-03-2023	59	56	124	2039	91	95	2039	CO	Hazardous	KIM
22-03-2023	34	33	107	260	30	83	260	CO	Hazardous	KIM
23-03-2023	17	15	112	689	45	164	689	CO	Hazardous	KIM
24-03-2023	19	18	176	6711		158	6711	CO	Hazardous	KIM
25-03-2023	23	22				192	192	NO_2	Unhealthy	KIM
26-03-2023	22	21	88	3511	183	71	3511	CO	Hazardous	KIM
27-03-2023	26	24	78	220	24	62	220	CO	Hazardous	KIM
28-03-2023	46	40	114	604	44	88	604	CO	Hazardous	KIM
29-03-2023	38	43	68	3069	111	55	3069	CO	Hazardous	KIM
30-03-2023	28	35	68	3706	139	54	3706	CO	Hazardous	KIM
31-03-2023	34	49	113	5875	228	91	5875	CO	Hazardous	KIM

Table 2. ISPU Data on April 2023

Tanggal	PM_{10}	$PM_{2.5}$	SO_2	CO	O_3	NO_2	Max	Critical	Category	Station
01-04-2023	52	53					53	PM10	Moderate	KIM
02-04-2023	43	48	120	5571	187	91	5571	CO	Hazardous	KIM
03-04-2023	33	34	94	5954	222	73	5954	CO	Hazardous	KIM
04-04-2023	45	43	123			96	123	SO2	Moderate	KIM
05-04-2023	26	25	140			114	140	SO2	Moderate	KIM
06-04-2023	39	36	31	1363	47	19	1363	CO	Hazardous	KIM
07-04-2023	35	32	33	701	19	59	701	CO	Hazardous	KIM
08-04-2023	26	24				2	26	PM10	Good	KIM
09-04-2023	48	43	98			63	98	SO2	Moderate	KIM
10-04-2023	38	38	63	3724	112	58	3724	CO	Hazardous	KIM
11-04-2023	58	57	74	3894	136	60	3894	CO	Hazardous	KIM
12-04-2023	47	41	83	2347	83	55	2347	CO	Hazardous	KIM
13-04-2023	48	39					48	PM10	Good	KIM
14-04-2023	40	32	113	4848	189	83	4848	CO	Hazardous	KIM
15-04-2023	46	26					46	PM10	Good	KIM

16-04-2023	47	34	143	5757	235	112	5757	CO	Hazardous	KIM
17-04-2023	47	32					47	PM10	Good	KIM
18-04-2023	42	22	60	294	8	102	294	CO	Very Unhealthy	KIM
19-04-2023	40	28	40	2850	30	26	2850	CO	Hazardous	KIM
20-04-2023	34	16	11	1118	28	27	1118	CO	Hazardous	KIM
21-04-2023	35	16	47	1217	40	129	1217	CO	Hazardous	KIM
22-04-2023	42	14	54	2830	85	34	2830	CO	Hazardous	KIM
23-04-2023	22	7	60	4151	87	48	4151	CO	Hazardous	KIM
24-04-2023	20	5	5	140	10	3	140	CO	Unhealthy	KIM
25-04-2023	28	6					28	PM10	Good	KIM
26-04-2023	44	6					44	PM10	Good	KIM
27-04-2023	38	5					38	PM10	Good	KIM
28-04-2023	28	4					28	PM10	Good	KIM
29-04-2023	30	3	74	4645	170	57	4645	CO	Hazardous	KIM
30-04-2023	34	3					34	PM10	Good	KIM

Table 3. ISPU Data on May 2023

Tanggal	**PM_{10}**	**$PM_{2.5}$**	**SO_2**	**CO**	**O_3**	**NO_2**	**Max**	**Critical**	**Category**	Station
01-05-2023	34	3					34	PM10	Good	KIM
02-05-2023	31	3					31	PM10	Good	KIM
03-05-2023	26	2					26	PM10	Good	KIM
04-05-2023	36	3					36	PM10	Good	KIM
05-05-2023	16	2					16	PM10	Good	KIM
06-05-2023	13	1					13	PM10	Good	KIM
07-05-2023	14	1					14	PM10	Good	KIM
08-05-2023	14	1					14	PM10	Good	KIM
09-05-2023	24	2					24	PM10	Good	KIM
10-05-2023	20	1					20	PM10	Good	KIM
11-05-2023	24	2					24	PM10	Good	KIM
12-05-2023	38	3	40	2807	104	38	2807	CO	Hazardous	KIM
13-05-2023	71	5	39	3453	67	34	3453	C0	Hazardous	KIM
14-05-2023	20	1					20	PM10	Good	KIM
15-05-2023	18	2	6	325	8	5	325	CO	Hazardous	KIM
16-05-2023	22	1					22	PM10	Good	KIM
17-05-2023	19	1	5	284	9	3	284	CO	Very Unhealthy	KIM
18-05-2023	14	1					14	PM10	Good	KIM
19-05-2023	15	1	151	6585		118	6585	CO	Hazardous	KIM
20-05-2023	22	2	67	2871	78	51	2871	CO	Hazardous	KIM
21-05-2023	12	1					12	PM10	Good	KIM
22-05-2023	20	1				166	166	NO2	Unhealthy	KIM
23-05-2023	17	1	16	1500	46	8	1500	CO	Hazardous	KIM
24-05-2023	24	2	95	5790	206	112	5790	CO	Hazardous	KIM
25-05-2023	30	2	74	6479	203	80	6479	CO	Hazardous	KIM
26-05-2023	13	1	60	5187	144	38	5187	CO	Hazardous	KIM
27-05-2023	14	1	81	4866	123	102	4866	CO	Hazardous	KIM
28-05-2023	10	1	65	2234	77	46	2234	CO	Hazardous	KIM
29-05-2023	18	1	143	1074	219	98	1074	CO	Hazardous	KIM
30-05-2023	19	1	16	185	41	7	185	CO	Unhealthy	KIM
31-05-2023	20	1	4	200	5	2	200	CO	Unhealthy	KIM

Table 4. ISPU Data on June 2023

Tanggal	**PM_{10}**	**$PM_{2.5}$**	**SO_2**	**CO**	**O_3**	**NO_2**	**Max**	**Critical**	**Category**	Station
01-06-2023	23	1	176			177	177	SO2	Unhealthy	KIM
02-06-2023	35	2	21	42	2	34	42	CO	Good	KIM
03-06-2023	38	2					38	PM10	Good	KIM
04-06-2023	14	1					14	PM10	Good	KIM
05-06-2023	18	1					18	PM10	Good	KIM
06-06-2023	16	1		84			84	CO	Moderate	KIM
07-06-2023	26	1		56			56	CO	Moderate	KIM
08-06-2023	33	2					33	PM10	Good	KIM
09-06-2023	35	2	57	5078	178	36	5078	CO	Hazardous	KIM
10-06-2023	23	1		1			23	PM10	Good	KIM
11-06-2023	22	1	1	34	1	1	34	CO	Good	KIM
12-06-2023	16	1					16	PM10	Good	KIM
13-06-2023	29	1					29	PM10	Good	KIM
14-06-2023	31	1					31	PM10	Good	KIM
15-06-2023	36	2					36	PM10	Good	KIM
16-06-2023	41	2		187			187	CO	Unhealthy	KIM
17-06-2023	36	2		100			100	CO	Moderate	KIM
18-06-2023	29	1					29	PM10	Good	KIM
19-06-2023	24	1	49	2308	78	64	2308	CO	Hazardous	KIM
20-06-2023	24	1	85	98		166	166	NO2	Unhealthy	KIM
21-06-2023	15	1	90	56		77	77	NO2	Moderate	KIM
22-06-2023	15	1	146	2971	107	156	2971	CO	Hazardous	KIM
23-06-2023	29	2	67	4768		27	4768	CO	Hazardous	KIM
24-06-2023	16	3	243	384	63	129	384	CO	Hazardous	KIM
25-06-2023	53	4	75	2453	58	37	2453	CO	Hazardous	KIM
26-06-2023	64	2	83	4814	365		4814	CO	Hazardous	KIM
27-06-2023	45	1	67	6874	45		6874	CO	Hazardous	KIM
28-06-2023	78	1	28	2871	736	21	2871	CO	Hazardous	KIM
29-06-2023	24	4	60	1987	233		1987	CO	Hazardous	KIM
30-06-2023	54	2	164	89		34	164	SO2	Unhealthy	KIM
31-06-2023	34	1	93	854	873	56	873	O3	Hazardous	KIM

3.2 Data Collection

Flowchart has a function to use, simplify a series of processes or procedures so that it can be easily understood and easily seen based on the sequence of steps of a process. The following is a flowchart of the whole used as a reference for research materials.

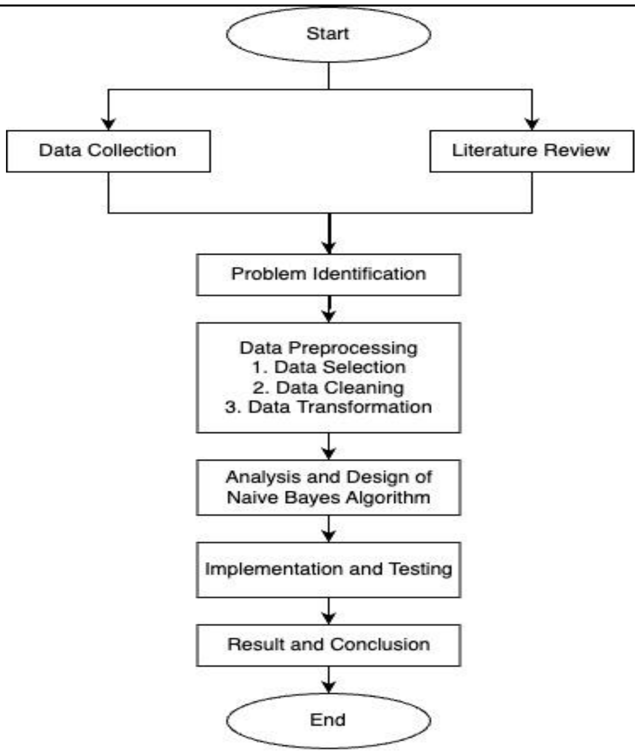

Figure 1. Classification System Flowchart

3.3 Discussion and Data Preparation

The data used in this study were obtained from the Medan Environmental Service data with a total of 100 data. Below is the dataset that will be the focus of this study. From the total data, 80% of this data will be allocated to train the model (training data) the remaining 20% will be allocated as data used as data used to test the model (test data).

	PM10	PM2.5	SO2	CO	O3	NO2	Max	Critical	Categori
0	23	21	113	1947	73.0	140	1947	CO	BERBAHAYA
1	34	32	106	1231	49.0	82	1231	CO	BERBAHAYA
2	19	17	79	1162	44.0	141	1162	CO	BERBAHAYA
3	16	15	74	2335	44.0	141	2335	CO	BERBAHAYA
4	25	24	136	3362	101.0	79	3362	CO	BERBAHAYA
...	...	...	...	...	...	...	...	...	...
95	24	1	85	98	0.0	166	166	NO2	TIDAK SEHAT
96	15	1	90	56	0.0	77	77	NO2	SEDANG
97	29	2	67	4768	0.0	27	4768	CO	BERBAHAYA
98	16	3	243	384	63.0	129	384	CO	BERBAHAYA
99	64	2	83	4814	365.0	0	4814	CO	BERBAHAYA

100 rows × 9 columns

Figure 2. Data Selection

After the processing stage, the next step is to change the object type data into data in the form of numbers. Figure 3 shows the process of changing the value of 'GOOD' to 0, 'HAZARDOUS' 1, 'VERY UNHEALTHY' 2, 'MODERATE' 3, 'UNHEALTHY' 4, then combined into a data frame and do not forget to change it to an integer data type.

```
[24] X = datalatih.drop(columns="Categori")
     y = datalatih.Categori
```

```
en=LabelEncoder()
datalatih['Categori']=en.fit_transform(datalatih['Categori'])
datalatih
```

	PM10	PM2.5	SO2	CO	O3	NO2	Max	Critical	Categori
0	23	21	113	1947	73.0	140	1947	CO	1
1	34	32	106	1231	49.0	82	1231	CO	1
2	19	17	79	1162	44.0	141	1162	CO	1
3	16	15	74	2335	44.0	141	2335	CO	1
4	25	24	136	3362	101.0	79	3362	CO	1
...	...	...	...	...	...	...	...	...	...
95	24	1	85	98	0.0	166	166	NO2	4
96	15	1	90	56	0.0	77	77	NO2	3
97	29	2	67	4768	0.0	27	4768	CO	1
98	16	3	243	384	63.0	129	384	CO	1
99	64	2	83	4814	365.0	0	4814	CO	1

100 rows × 9 columns

Figure 3. Category data transformation

Next, the important step is to define the independent variable (x) and the dependent variable (y). Figure 4 visualizes the process of declaring variables x and y.

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
X_train.shape, X_test.shape, y_train.shape, y_test.shape

((80, 8), (20, 8), (80,), (20,))
```

Figure 4. Variables X and Y

3.4 Training and Testing Data Sharing

It is a stage in machine learning to share training and testing data, this time the data is divided 75% for training data and the rest for testing data.

```
[30] pipeline.score(X_train, y_train)

0.9625

[31] pipeline.score(X_test, y_test)

0.8
```

Figure 5. Training and Testing Data

3.5 Implementation of the Naïve Bayes Method

Implementation of the naïve bayes method to classify air quality in the Medan industrial area which begins with defining the amount of data for each category, calculating the naïve bayes method, calculating the prior probability, calculating the likelihood, calculating the posterior probability. The amount of data based on the category is 100 with details:

a. GOOD = 22,
b. HAZARDOUS = 58,
c. MODERATE = 10,
d. UNHEALTHY = 8,
e. VERY UNHEALTHY = 2.

The sample data to be classified are PM10 = 20, PM2.5 = 18, SO2 = 100, CO = 1500, 03 = 50, NO2 = 100.

3.5.1 Good Air Quality Category

a. Defining Average and Standard Deviation

Table 5. Average and Deviation Standard of each pollutant for the Good Category

Pollutant Type	Average(μ)	Deviation Standard(σ)
PM_{10}	27,36	10,85
$PM_{2.5}$	7,82	11,71
SO_2	1,86	5,88
CO	3,50	11,24
O_3	0,14	0,47
NO_2	2,59	8,20

b. Calculating prior probability

P(GOOD) = 22/100 = 0.22

c. Calculating Likelihood

Likelihood calculation for PM10 pollutant:

$$\mu_{\text{PM10 | BAIK}} = 27.36, \quad \sigma_{\text{PM10 | BAIK}} = 10.85$$

$$P(\text{PM10} = 20|\text{BAIK}) = \frac{1}{\sqrt{2\pi(10.85)^2}} \times e^{-\frac{(20-27.36)^2}{2(10.85)^2}} \tag{1}$$

$$(20 - 27.36)^2 = (-7.36)^2 = 54.18$$

$$\text{Eksponen} = -\frac{54.18}{2 \times (10.85)^2} = -\frac{54.18}{235.62} \approx -0.23$$

$$e^{-0.23} \approx 0.794$$

$$\frac{1}{\sqrt{2\pi(10.85)^2}} = \frac{1}{\sqrt{2\pi \times 117.77}} \approx \frac{1}{27.24} \approx 0.0367$$

$$P(\text{PM10} = 20|\text{BAIK}) \approx 0.0367 \times 0.794 \approx 0.0291$$

Likelihood calculation for $PM_{2.5}$ pollutants:

$$\mu_{\text{PM2.5 | BAIK}} = 7.82, \quad \sigma_{\text{PM2.5 | BAIK}} = 11.71$$

$$P(\text{PM2.5} = 18|\text{BAIK}) = \frac{1}{\sqrt{2\pi(11.71)^2}} \times e^{-\frac{(18-7.82)^2}{2(11.71)^2}} \tag{2}$$

$$(18 - 7.82)^2 = 10.18^2 = 103.63$$

$$\text{Eksponen} = -\frac{103.63}{2 \times (11.71)^2} = -\frac{103.63}{274.29} \approx -0.378$$

$$e^{-0.378} \approx 0.685$$

$$\frac{1}{\sqrt{2\pi(11.71)^2}} = \frac{1}{\sqrt{2\pi \times 137.13}} \approx \frac{1}{29.62} \approx 0.0338$$

$$P(\text{PM2.5} = 18|\text{BAIK}) \approx 0.0338 \times 0.685 \approx 0.0231$$

Likelihood calculation for SO_2 pollutant:

$$\mu_{\text{SO2 | BAIK}} = 1.86, \quad \sigma_{\text{SO2 | BAIK}} = 5.88$$

$$P(\text{SO2} = 100|\text{BAIK}) = \frac{1}{\sqrt{2\pi(5.88)^2}} \times e^{-\frac{(100-1.86)^2}{2(5.88)^2}} \tag{3}$$

$$(100 - 1.86)^2 = 98.14^2 = 9627.32$$

$$\text{Eksponen} = -\frac{9627.32}{2 \times (5.88)^2} = -\frac{9627.32}{69.15} \approx -139.18$$

$$e^{-139.18} \approx 2.1 \times 10^{-61}$$

$$\frac{1}{\sqrt{2\pi(5.88)^2}} = \frac{1}{\sqrt{2\pi \times 34.57}} \approx \frac{1}{14.73} \approx 0.0679$$

$$P(\text{SO2} = 100|\text{BAIK}) \approx 0.0679 \times 2.1 \times 10^{-61} \approx 1.42 \times 10^{-62}$$

Likelihood calculation for CO pollutant:

$$\mu_{\text{CO}\,|\,\text{BAIK}} = 3.50, \quad \sigma_{\text{CO}\,|\,\text{BAIK}} = 11.24$$

$$P(\text{CO} = 1500|\text{BAIK}) = \frac{1}{\sqrt{2\pi(11.24)^2}} \times e^{-\frac{(1500-3.50)^2}{2(11.24)^2}} \tag{4}$$

$$(1500 - 3.50)^2 = 1496.50^2 = 2239502.25$$

$$\text{Eksponen} = -\frac{2239502.25}{2 \times (11.24)^2} = -\frac{2239502.25}{252.56} \approx -8864.11$$

$$e^{-8864.11} \approx 0 \quad \text{(Teramat sangat kecil hingga mendekati 0)}$$

$$\frac{1}{\sqrt{2\pi(11.24)^2}} = \frac{1}{\sqrt{2\pi \times 126.38}} \approx \frac{1}{27.88} \approx 0.0359$$

$$P(\text{CO} = 1500|\text{BAIK}) \approx 0 \quad \text{(Nilai yang sangat kecil mendekati 0)}$$

Likelihood calculation for O_3 pollutant:

$$\mu_{\text{O3}\,|\,\text{BAIK}} = 0.14, \quad \sigma_{\text{O3}\,|\,\text{BAIK}} = 0.47$$

$$P(\text{O3} = 50|\text{BAIK}) = \frac{1}{\sqrt{2\pi(0.47)^2}} \times e^{-\frac{(50-0.14)^2}{2(0.47)^2}} \tag{5}$$

$$(50 - 0.14)^2 = 49.86^2 = 2486.42$$

$$\text{Eksponen} = -\frac{2486.42}{2 \times (0.47)^2} = -\frac{2486.42}{0.4418} \approx -5629.17$$

$$e^{-5629.17} \approx 0 \quad \text{(Teramat sangat kecil hingga mendekati 0)}$$

$$\frac{1}{\sqrt{2\pi(0.47)^2}} = \frac{1}{\sqrt{2\pi \times 0.22}} \approx \frac{1}{0.83} \approx 1.20$$

$$P(\text{O3} = 50|\text{BAIK}) \approx 0 \quad \text{(Nilai yang sangat kecil mendekati 0)}$$

Likelihood calculation for NO_2 pollutant:

$$\mu_{\text{NO2}\,|\,\text{BAIK}} = 2.59, \quad \sigma_{\text{NO2}\,|\,\text{BAIK}} = 8.20$$

$$P(\text{NO2} = 100|\text{BAIK}) = \frac{1}{\sqrt{2\pi(8.20)^2}} \times e^{-\frac{(100-2.59)^2}{2(8.20)^2}} \tag{6}$$

$$(100-2.59)^2 = 97.41^2 = 9489.24$$

$$\text{Eksponen} = -\frac{9489.24}{2 \times (8.20)^2} = -\frac{9489.24}{134.48} \approx -70.56$$

$$e^{-70.56} \approx 2.54 \times 10^{-31}$$

$$\frac{1}{\sqrt{2\pi(8.20)^2}} = \frac{1}{\sqrt{2\pi \times 67.24}} \approx \frac{1}{20.61} \approx 0.0485$$

d. Calculating posterior probability
Posterior Probability Formula:
P(GOOD | X)=P(GOOD) × P(PM10=20 | GOOD) × P(PM2.5=18 | GOOD) × P(SO2=100 | GOOD) × P(CO=1500 | GOOD) × P(O3=50 | GOOD) × P(NO2=100 | GOOD)
The result is:
P(GOOD |X) ≈ 0.22 × 0.0291 × 0.0231 × 1.42 × 10^{-62} × 0 × 0 × 1.23×10^{-32}

3.5.2 Hazardous Air Quality Categories

a. Defining Average and Standard Deviation

Table 6. Average and Standard Deviation of each pollutant for the Hazardous Category

Pollutant Type	Average(μ)	Deviation Standard(σ)
PM_{10}	33,47	13,39
$PM_{2.5}$	23,26	17,34
SO_2	91,03	51,82
CO	3617,07	2096,57
O_3	104,18	82,57
NO_2	85,10	45,49

b. Calculating prior probability
P(HAZARDOUS) = 58/100 = 0.58
c. Calculating Likelihood
Likelihood calculation for PM_{10} pollutants:

$$\mu_{\text{PM10} \mid \text{BERBAHAYA}} = 33.47, \quad \sigma_{\text{PM10} \mid \text{BERBAHAYA}} = 13.39$$

$$P(\text{PM10} = 20|\text{BERBAHAYA}) = \frac{1}{\sqrt{2\pi(13.39)^2}} \times e^{-\frac{(20-33.47)^2}{2(13.39)^2}} \quad (7)$$

$$(20-33.47)^2 = (-13.47)^2 = 181.37$$

$$\text{Eksponen} = -\frac{181.37}{2 \times (13.39)^2} = -\frac{181.37}{358.60} \approx -0.506$$

$$e^{-0.506} \approx 0.603$$

$$\frac{1}{\sqrt{2\pi(13.39)^2}} = \frac{1}{\sqrt{841.87 \times 2\pi}} = \frac{1}{\sqrt{5290.99}} \approx \frac{1}{72.73} \approx 0.01375$$

$$P(\text{PM10} = 20|\text{BERBAHAYA}) \approx 0.01375 \times 0.603 \approx 0.00829$$

Likelihood calculation for $PM_{2.5}$ pollutants:

$$\mu_{\text{PM2.5} \mid \text{BERBAHAYA}} = 23.26, \quad \sigma_{\text{PM2.5} \mid \text{BERBAHAYA}} = 17.34$$

$$P(\text{PM2.5} = 18|\text{BERBAHAYA}) = \frac{1}{\sqrt{2\pi(17.34)^2}} \times e^{-\frac{(18-23.26)^2}{2(17.34)^2}} \quad (8)$$

$$(18-23.26)^2=(-5.26)^2=27.65$$

$$\text{Eksponen}=-\frac{27.65}{2\times(17.34)^2}=-\frac{27.65}{601.48}\approx-0.046$$

$$e^{-0.046}\approx 0.955$$

$$\frac{1}{\sqrt{2\pi(17.34)^2}}=\frac{1}{\sqrt{1899.44\times 2\pi}}=\frac{1}{\sqrt{11933.88}}\approx\frac{1}{109.24}\approx 0.00915$$

$$P(\text{PM2.5}=18|\text{BERBAHAYA})\approx 0.00915\times 0.955\approx 0.00874$$

Likelihood calculation for SO_2 pollutant:

$$\mu_{\text{SO2}\mid\text{BERBAHAYA}}=91.03,\quad \sigma_{\text{SO2}\mid\text{BERBAHAYA}}=51.82$$

$$P(\text{SO2}=100|\text{BERBAHAYA})=\frac{1}{\sqrt{2\pi(51.82)^2}}\times e^{-\frac{(100-91.03)^2}{2(51.82)^2}} \tag{9}$$

$$(100-91.03)^2=8.97^2=80.45$$

$$\text{Eksponen}=-\frac{80.45}{2\times(51.82)^2}=-\frac{80.45}{5378.22}\approx-0.015$$

$$e^{-0.015}\approx 0.985$$

$$\frac{1}{\sqrt{2\pi(51.82)^2}}=\frac{1}{\sqrt{5378.22\times 2\pi}}=\frac{1}{\sqrt{33796.09}}\approx\frac{1}{183.88}\approx 0.00544$$

$$P(\text{SO2}=100|\text{BERBAHAYA})\approx 0.00544\times 0.985\approx 0.00536$$

Likelihood calculation for CO pollutant:

$$\mu_{\text{CO}\mid\text{BERBAHAYA}}=3617.07,\quad \sigma_{\text{CO}\mid\text{BERBAHAYA}}=2096.57$$

$$P(\text{CO}=1500|\text{BERBAHAYA})=\frac{1}{\sqrt{2\pi(2096.57)^2}}\times e^{-\frac{(1500-3617.07)^2}{2(2096.57)^2}} \tag{10}$$

$$(1500-3617.07)^2=(-2117.07)^2=4481938.45$$

$$\text{Eksponen}=-\frac{4481938.45}{2\times(2096.57)^2}=-\frac{4481938.45}{8796900.81}\approx-0.509$$

$$e^{-0.509}\approx 0.601$$

$$\frac{1}{\sqrt{2\pi(2096.57)^2}}=\frac{1}{\sqrt{8796900.81\times 2\pi}}=\frac{1}{\sqrt{55269079.07}}\approx\frac{1}{7435.58}\approx 0.000134$$

$$P(\text{CO}=1500|\text{BERBAHAYA})\approx 0.0001345\times 0.601\approx 0.0000809$$

Likelihood calculation for O_3 pollutant:

$$\mu_{\text{O3}\mid\text{BERBAHAYA}}=104.18,\quad \sigma_{\text{O3}\mid\text{BERBAHAYA}}=82.57$$

$$P(\text{O3}=50|\text{BERBAHAYA})=\frac{1}{\sqrt{2\pi(82.57)^2}}\times e^{-\frac{(50-104.18)^2}{2(82.57)^2}} \tag{11}$$

$$(50-104.18)^2=(-54.18)^2=2935.56$$

$$\text{Eksponen}=-\frac{2935.56}{2\times(82.57)^2}=-\frac{2935.56}{13638.54}\approx-0.215$$

$$e^{-0.215}\approx 0.807$$

$$\frac{1}{\sqrt{2\pi(82.57)^2}}=\frac{1}{\sqrt{13638.54\times 2\pi}}=\frac{1}{\sqrt{85673.84}}\approx\frac{1}{292.73}\approx 0.00342$$

$$P(\text{O3}=50|\text{BERBAHAYA})\approx 0.00342\times 0.807\approx 0.00276$$

Likelihood calculation for NO_2 pollutant:

$$\mu_{\text{NO2 | BERBAHAYA}}=85.10,\quad \sigma_{\text{NO2 | BERBAHAYA}}=45.49$$

$$P(\text{NO2}=100|\text{BERBAHAYA})=\frac{1}{\sqrt{2\pi(45.49)^2}}\times e^{-\frac{(100-85.10)^2}{2(45.49)^2}} \tag{12}$$

$$(100-85.10)^2=(14.90)^2=222.01$$

$$\text{Eksponen}=-\frac{222.01}{2\times(45.49)^2}=-\frac{222.01}{4139.24}\approx-0.054$$

$$e^{-0.054}\approx 0.947$$

$$\frac{1}{\sqrt{2\pi(45.49)^2}}=\frac{1}{\sqrt{4139.24\times 2\pi}}=\frac{1}{\sqrt{26024.47}}\approx\frac{1}{161.27}\approx 0.0062$$

$$P(\text{NO2}=100|\text{BERBAHAYA})\approx 0.0062\times 0.947\approx 0.00587$$

d. Calculating posterior probability
Posterior probability:
P(HAZARDOUS|X) = P(HAZARDOUS) × P(PM10 = 20|HAZARDOUS) × P(PM2.5 = 18|HAZARDOUS) × P(SO2 = 100|HAZARDOUS) × P(CO = 1500|HAZARDOUS) × P(O3 = 50|HAZARDOUS) × P(NO2 = 100|HAZARDOUS)
Implementation of the posterior probability:
P(HAZARDOUS|X) ≈ 0.58 × 0.00829 × 0.00874 × 0.00536 × 0.0000809 × 0.00276 × 0.00587
Result:
P(HAZARDOUS|X) ≈ 0.58 × 7.08 × 10^{-14} ≈ 4.11 × 10^{-14}

3.5.3 Very Unhealthy Air Quality Category

a. Defining Average and Standard Deviation

Table 7. Average and Standard Deviation of each pollutant for the Very Unhealthy Category

Pollutant Type	Average(μ)	Deviation Standard(σ)
PM_{10}	30,50	16,62
$PM_{2.5}$	11,50	14,85
SO_2	32,50	38,89
CO	289,00	7,07
O_3	8,50	0,71
NO_2	52,50	70,00

b. Calculating prior probability
P(VERY UNHEALTHY) = 2/100 = 0.02

c. Calculating prior probabilities
Likelihood calculation for PM_{10} pollutants:

$$\mu_{\text{PM10}\,|\,\text{SANGAT TIDAK SEHAT}} = 30.50, \quad \sigma_{\text{PM10}\,|\,\text{SANGAT TIDAK SEHAT}} = 16.26$$

$$P(\text{PM10} = 20|\text{SANGAT TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi(16.26)^2}} \times e^{-\frac{(20-30.50)^2}{2(16.26)^2}} \tag{13}$$

$$(20 - 30.50)^2 = (-10.50)^2 = 110.25$$

$$\text{Eksponen} = -\frac{110.25}{2 \times (16.26)^2} = -\frac{110.25}{528.40} \approx -0.209$$

$$e^{-0.209} \approx 0.811$$

$$\frac{1}{\sqrt{2\pi(16.26)^2}} = \frac{1}{\sqrt{2\pi \times 264.40}} \approx \frac{1}{40.87} \approx 0.0245$$

$$P(\text{PM10} = 20|\text{SANGAT TIDAK SEHAT}) \approx 0.0245 \times 0.811 \approx 0.0199$$

Likelihood calculation for $PM_{2.5}$ pollutants:

$$\mu_{\text{PM2.5}\,|\,\text{SANGAT TIDAK SEHAT}} = 11.50, \quad \sigma_{\text{PM2.5}\,|\,\text{SANGAT TIDAK SEHAT}} = 14.85$$

$$P(\text{PM2.5} = 18|\text{SANGAT TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi(14.85)^2}} \times e^{-\frac{(18-11.50)^2}{2(14.85)^2}} \tag{14}$$

$$(18 - 11.50)^2 = 6.50^2 = 42.25$$

$$\text{Eksponen} = -\frac{42.25}{2 \times (14.85)^2} = -\frac{42.25}{440.30} \approx -0.096$$

$$e^{-0.096} \approx 0.908$$

$$\frac{1}{\sqrt{2\pi(14.85)^2}} = \frac{1}{\sqrt{2\pi \times 220.55}} \approx \frac{1}{37.32} \approx 0.0268$$

$$P(\text{PM2.5} = 18|\text{SANGAT TIDAK SEHAT}) \approx 0.0268 \times 0.908 \approx 0.0243$$

Likelihood calculation for SO_2 pollutant:

$$\mu_{\text{SO2}\,|\,\text{SANGAT TIDAK SEHAT}} = 32.50, \quad \sigma_{\text{SO2}\,|\,\text{SANGAT TIDAK SEHAT}} = 38.89$$

$$P(\text{SO2} = 100|\text{SANGAT TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi(38.89)^2}} \times e^{-\frac{(100-32.50)^2}{2(38.89)^2}} \tag{15}$$

$$(100 - 32.50)^2 = 67.50^2 = 4556.25$$

$$\text{Eksponen} = -\frac{4556.25}{2 \times (38.89)^2} = -\frac{4556.25}{3023.92} \approx -1.507$$

$$e^{-1.507} \approx 0.221$$

$$\frac{1}{\sqrt{2\pi(38.89)^2}} = \frac{1}{\sqrt{2\pi \times 1512.78}} \approx \frac{1}{97.68} \approx 0.0102$$

$$P(\text{SO2} = 100|\text{SANGAT TIDAK SEHAT}) \approx 0.0102 \times 0.221 \approx 0.00226$$

Likelihood calculation for CO pollutant:

$$\mu_{\text{CO}\,|\,\text{SANGAT TIDAK SEHAT}} = 289.00, \quad \sigma_{\text{CO}\,|\,\text{SANGAT TIDAK SEHAT}} = 7.07$$

$$P(\text{CO} = 1500|\text{SANGAT TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi(7.07)^2}} \times e^{-\frac{(1500-289.00)^2}{2(7.07)^2}} \tag{16}$$

$$(1500-289.00)^2 = 1211.00^2 = 1465841.00$$

$$\text{Eksponen} = -\frac{1465841.00}{2\times(7.07)^2} = -\frac{1465841.00}{99.97} \approx -14665.84$$

$$e^{-14665.84} \approx 0 \quad \text{(Teramat sangat kecil hingga mendekati 0)}$$

$$\frac{1}{\sqrt{2\pi(7.07)^2}} = \frac{1}{\sqrt{2\pi\times 49.98}} \approx \frac{1}{17.75} \approx 0.0563$$

$$P(\text{CO} = 1500|\text{SANGAT TIDAK SEHAT}) \approx 0$$

Likelihood calculation for O_3 pollutant:

$$\mu_{\text{O3 | SANGAT TIDAK SEHAT}} = 8.50, \quad \sigma_{\text{O3 | SANGAT TIDAK SEHAT}} = 0.71$$

$$P(\text{O3} = 50|\text{SANGAT TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi(0.71)^2}} \times e^{-\frac{(50-8.50)^2}{2(0.71)^2}} \tag{17}$$

$$(50-8.50)^2 = 41.50^2 = 1722.25$$

$$\text{Eksponen} = -\frac{1722.25}{2\times(0.71)^2} = -\frac{1722.25}{1.0082} \approx -1708.09$$

$$e^{-1708.09} \approx 0 \quad \text{(Teramat sangat kecil hingga mendekati 0)}$$

$$\frac{1}{\sqrt{2\pi(0.71)^2}} = \frac{1}{\sqrt{2\pi\times 0.5041}} \approx \frac{1}{1.78} \approx 0.5618$$

$$P(\text{O3} = 50|\text{SANGAT TIDAK SEHAT}) \approx 0$$

Likelihood calculation for NO_2 pollutant:

$$\mu_{\text{NO2 | SANGAT TIDAK SEHAT}} = 52.50, \quad \sigma_{\text{NO2 | SANGAT TIDAK SEHAT}} = 70.00$$

$$P(\text{NO2} = 100|\text{SANGAT TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi(70.00)^2}} \times e^{-\frac{(100-52.50)^2}{2(70.00)^2}} \tag{18}$$

$$(100-52.50)^2 = 47.50^2 = 2256.25$$

$$\text{Eksponen} = -\frac{2256.25}{2\times(70.00)^2} = -\frac{2256.25}{9800} \approx -0.230$$

$$e^{-0.230} \approx 0.794$$

$$\frac{1}{\sqrt{2\pi(70.00)^2}} = \frac{1}{\sqrt{2\pi\times 4900}} \approx \frac{1}{187.68} \approx 0.00533$$

$$P(\text{NO2} = 100|\text{SANGAT TIDAK SEHAT}) \approx 0.00533 \times 0.794 \approx 0.00423$$

d. Calculating posterior probability
Posterior probability:
P(SANGAT UNHEALTHY|X)=P(SANGAT UNHEALTHY)×P(PM10=20|SANGAT UNHEALTHY)×P(PM2.5=18|SANGAT UNHEALTHY)×P(SO2=100|SANGAT UNHEALTHY)×P(CO=1500|SANGAT UNHEALTHY)×P(O3=50|SANGAT UNHEALTHY)×P(NO2=100|SANGAT UNHEALTHY)
Result:
P(SANGAT UNHEALTHY|X)≈0.02×0.0199×0.0243×0.00226×0×0×0.00423

3.5.4 Moderate Air Quality Category

a. Defining Average and Standard Deviation

Table 8. Average and Standard Deviation of each pollutant for the Moderate Category

Pollutant Type	Average(μ)	Deviation Standard(σ)
PM_{10}	37,00	15,24
$PM_{2.5}$	26,70	23,12
SO_2	53,60	55,10
CO	29,60	40,23
O_3	0,10	0,32
NO_2	42,50	43,29

b. Calculating prior probability
P(GOOD) = 10/100 = 0.10
c. Calculating Likelihood
Likelihood calculation for PM_{10} pollutants:

$$\mu_{\text{PM10 | SEDANG}} = 37.00, \quad \sigma_{\text{PM10 | SEDANG}} = 15.24$$

$$P(\text{PM10} = 20|\text{SEDANG}) = \frac{1}{\sqrt{2\pi(15.24)^2}} \times e^{-\frac{(20-37.00)^2}{2(15.24)^2}} \tag{19}$$

$$(20 - 37.00)^2 = (-17.00)^2 = 289$$

$$\text{Eksponen} = -\frac{289}{2 \times (15.24)^2} = -\frac{289}{464.45} \approx -0.622$$

$$e^{-0.622} \approx 0.537$$

$$\frac{1}{\sqrt{2\pi(15.24)^2}} = \frac{1}{\sqrt{2\pi \times 232.26}} \approx \frac{1}{38.22} \approx 0.0262$$

$$P(\text{PM10} = 20|\text{SEDANG}) \approx 0.0262 \times 0.537 \approx 0.0141$$

Likelihood calculation for $PM_{2.5}$ pollutants:

$$\mu_{\text{PM2.5 | SEDANG}} = 26.70, \quad \sigma_{\text{PM2.5 | SEDANG}} = 23.12$$

$$P(\text{PM2.5} = 18|\text{SEDANG}) = \frac{1}{\sqrt{2\pi(23.12)^2}} \times e^{-\frac{(18-26.70)^2}{2(23.12)^2}} \tag{20}$$

$$(18 - 26.70)^2 = (-8.70)^2 = 75.69$$

$$\text{Eksponen} = -\frac{75.69}{2 \times (23.12)^2} = -\frac{75.69}{1067.36} \approx -0.071$$

$$e^{-0.071} \approx 0.932$$

$$\frac{1}{\sqrt{2\pi(23.12)^2}} = \frac{1}{\sqrt{2\pi \times 534.54}} \approx \frac{1}{58.14} \approx 0.0172$$

$$P(\text{PM2.5} = 18|\text{SEDANG}) \approx 0.0172 \times 0.932 \approx 0.0160$$

Likelihood calculation for SO_2 pollutant:

$$\mu_{\text{SO2 | SEDANG}} = 53.60, \quad \sigma_{\text{SO2 | SEDANG}} = 55.10$$

$$P(\text{SO2} = 100|\text{SEDANG}) = \frac{1}{\sqrt{2\pi(55.10)^2}} \times e^{-\frac{(100-53.60)^2}{2(55.10)^2}} \tag{21}$$

$$(100-53.60)^2 = 46.40^2 = 2153.66$$

$$\text{Eksponen} = -\frac{2153.66}{2\times(55.10)^2} = -\frac{2153.66}{6060.61} \approx -0.355$$

$$e^{-0.355} \approx 0.701$$

$$\frac{1}{\sqrt{2\pi(55.10)^2}} = \frac{1}{\sqrt{2\pi\times 3036.01}} \approx \frac{1}{138.78} \approx 0.0072$$

$$P(\text{SO2} = 100|\text{SEDANG}) \approx 0.0072 \times 0.701 \approx 0.00505$$

Likelihood calculation for CO pollutant:

$$\mu_{\text{CO}\,|\,\text{SEDANG}} = 29.60, \quad \sigma_{\text{CO}\,|\,\text{SEDANG}} = 40.23$$

$$P(\text{CO} = 1500|\text{SEDANG}) = \frac{1}{\sqrt{2\pi(40.23)^2}} \times e^{-\frac{(1500-29.60)^2}{2(40.23)^2}} \tag{22}$$

$$(1500-29.60)^2 = 1470.40^2 = 2169256.16$$

$$\text{Eksponen} = -\frac{2169256.16}{2\times(40.23)^2} = -\frac{2169256.16}{3244.39} \approx -668.72$$

$$e^{-668.72} \approx 0 \quad (\text{Teramat sangat kecil hingga mendekati 0})$$

$$\frac{1}{\sqrt{2\pi(40.23)^2}} = \frac{1}{\sqrt{2\pi\times 1618.47}} \approx \frac{1}{100.95} \approx 0.00991$$

$$P(\text{CO} = 1500|\text{SEDANG}) \approx 0 \quad (\text{Nilai yang sangat kecil mendekati 0})$$

Likelihood calculation for O_3 pollutant:

$$\mu_{\text{O3}\,|\,\text{SEDANG}} = 0.10, \quad \sigma_{\text{O3}\,|\,\text{SEDANG}} = 0.32$$

$$P(\text{O3} = 50|\text{SEDANG}) = \frac{1}{\sqrt{2\pi(0.32)^2}} \times e^{-\frac{(50-0.10)^2}{2(0.32)^2}} \tag{23}$$

$$(50-0.10)^2 = 49.90^2 = 2490.01$$

$$\text{Eksponen} = -\frac{2490.01}{2\times(0.32)^2} = -\frac{2490.01}{0.2048} \approx -12155.37$$

$$e^{-12155.37} \approx 0 \quad (\text{Teramat sangat kecil hingga mendekati 0})$$

$$\frac{1}{\sqrt{2\pi(0.32)^2}} = \frac{1}{\sqrt{2\pi\times 0.1024}} \approx \frac{1}{0.254} \approx 3.937$$

$$P(\text{O3} = 50|\text{SEDANG}) \approx 0 \quad (\text{Nilai yang sangat kecil mendekati 0})$$

Likelihood calculation for NO2 pollutant:

$$\mu_{\text{NO2}\,|\,\text{SEDANG}} = 42.50, \quad \sigma_{\text{NO2}\,|\,\text{SEDANG}} = 43.29$$

$$P(\text{NO2} = 100|\text{SEDANG}) = \frac{1}{\sqrt{2\pi(43.29)^2}} \times e^{-\frac{(100-42.50)^2}{2(43.29)^2}} \tag{24}$$

$$(100 - 42.50)^2 = 57.50^2 = 3306.25$$

$$\text{Eksponen} = -\frac{3306.25}{2 \times (43.29)^2} = -\frac{3306.25}{3745.47} \approx -0.883$$

$$e^{-0.883} \approx 0.414$$

$$\frac{1}{\sqrt{2\pi(43.29)^2}} = \frac{1}{\sqrt{2\pi \times 1873.99}} \approx \frac{1}{108.67} \approx 0.0092$$

$$P(\text{NO2} = 100|\text{SEDANG}) \approx 0.0092 \times 0.414 \approx 0.00381$$

d. Calculating posterior probability
   Posterior probability formula:
   P(SEDANG|X)=P(SEDANG)×P(PM10=20|SEDANG)×P(PM2.5=18|SEDANG)×P(SO2=100|SEDANG)×P(CO=1500|SEDANG)×P(O3=50|SEDANG)×P(NO2=100|SEDANG)
   Result:
   P(SEDANG|X)=0.10×0.0141×0.0160×0.00505×0×0×0.00381

3.5.5 Unhealthy Air Quality Category

a. Defining Average and Standard Deviation

Table 9. Average and Standard Deviation of each pollutant for the Unhealthy Category

Pollutant Type	Average(μ)	Deviation Standard(σ)
PM_{10}	23,75	7,21
$PM_{2.5}$	4,25	7,30
SO_2	35,75	63,55
CO	101,25	89,75
O_3	7,00	14,21
NO_2	89,13	92,44

b. Calculating prior probability
   P(UNHEALTHY) = 8/100 = 0.08
c. Calculating Likelihood
   Likelihood calculation for PM_{10} pollutants:

$$\mu_{\text{PM10 | TIDAK SEHAT}} = 23.75, \quad \sigma_{\text{PM10 | TIDAK SEHAT}} = 7.21$$

$$P(\text{PM10} = 20|\text{TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi(7.21)^2}} \times e^{-\frac{(20-23.75)^2}{2(7.21)^2}} \tag{25}$$

$$(20 - 23.75)^2 = (-3.75)^2 = 14.06$$

$$\text{Eksponen} = -\frac{14.06}{2 \times (7.21)^2} = -\frac{14.06}{104.05} \approx -0.135$$

$$e^{-0.135} \approx 0.874$$

$$\frac{1}{\sqrt{2\pi(7.21)^2}} = \frac{1}{\sqrt{2\pi \times 52.00}} \approx \frac{1}{18.09} \approx 0.0553$$

$$P(\text{PM10} = 20|\text{TIDAK SEHAT}) \approx 0.0553 \times 0.874 \approx 0.0483$$

Likelihood calculation for $PM_{2.5}$ pollutants:

$$\mu_{\text{PM2.5 | TIDAK SEHAT}} = 4.25, \quad \sigma_{\text{PM2.5 | TIDAK SEHAT}} = 7.30$$

$$P(\text{PM2.5} = 18|\text{TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi(7.30)^2}} \times e^{-\frac{(18-4.25)^2}{2(7.30)^2}} \tag{26}$$

$$(18 - 4.25)^2 = 13.75^2 = 189.06$$

$$\text{Eksponen} = -\frac{189.06}{2 \times (7.30)^2} = -\frac{189.06}{106.73} \approx -1.772$$

$$e^{-1.772} \approx 0.170$$

$$\frac{1}{\sqrt{2\pi(7.30)^2}} = \frac{1}{\sqrt{2\pi \times 53.29}} \approx \frac{1}{18.34} \approx 0.0545$$

$$P(\text{PM2.5} = 18|\text{TIDAK SEHAT}) \approx 0.0545 \times 0.170 \approx 0.00927$$

Likelihood calculation for SO_2 pollutant:

$$\mu_{\text{SO2 | TIDAK SEHAT}} = 35.75, \quad \sigma_{\text{SO2 | TIDAK SEHAT}} = 63.55$$

$$P(\text{SO2} = 100|\text{TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi(63.55)^2}} \times e^{-\frac{(100-35.75)^2}{2(63.55)^2}} \tag{27}$$

$$(100 - 35.75)^2 = 64.25^2 = 4128.06$$

$$\text{Eksponen} = -\frac{4128.06}{2 \times (63.55)^2} = -\frac{4128.06}{8061.78} \approx -0.512$$

$$e^{-0.512} \approx 0.599$$

$$\frac{1}{\sqrt{2\pi(63.55)^2}} = \frac{1}{\sqrt{2\pi \times 4038.60}} \approx \frac{1}{159.33} \approx 0.00627$$

$$P(\text{SO2} = 100|\text{TIDAK SEHAT}) \approx 0.00627 \times 0.599 \approx 0.00375$$

Likelihood calculation for CO pollutant:

$$\mu_{\text{CO | TIDAK SEHAT}} = 101.25, \quad \sigma_{\text{CO | TIDAK SEHAT}} = 89.75$$

$$P(\text{CO} = 1500|\text{TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi(89.75)^2}} \times e^{-\frac{(1500-101.25)^2}{2(89.75)^2}} \tag{28}$$

$$(1500 - 101.25)^2 = 1398.75^2 = 1956506.56$$

$$\text{Eksponen} = -\frac{1956506.56}{2 \times (89.75)^2} = -\frac{1956506.56}{16100.06} \approx -121.54$$

$$e^{-121.54} \approx 0 \quad (\text{Teramat sangat kecil hingga mendekati 0})$$

$$\frac{1}{\sqrt{2\pi(89.75)^2}} = \frac{1}{\sqrt{2\pi \times 8056.56}} \approx \frac{1}{224.88} \approx 0.00445$$

$$P(\text{CO} = 1500|\text{TIDAK SEHAT}) \approx 0$$

Likelihood calculation for O_3 pollutant:

$$\mu_{\text{O3 | TIDAK SEHAT}} = 7.00, \quad \sigma_{\text{O3 | TIDAK SEHAT}} = 14.21$$

$$P(\text{O3} = 50|\text{TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi(14.21)^2}} \times e^{-\frac{(50-7.00)^2}{2(14.21)^2}} \tag{29}$$

$$(50 - 7.00)^2 = 43.00^2 = 1849.00$$

$$\text{Eksponen} = -\frac{1849.00}{2 \times (14.21)^2} = -\frac{1849.00}{403.91} \approx -4.576$$

$$e^{-4.576} \approx 0.0103$$

$$\frac{1}{\sqrt{2\pi(14.21)^2}} = \frac{1}{\sqrt{2\pi \times 201.92}} \approx \frac{1}{35.67} \approx 0.028$$

$$P(\text{O3} = 50|\text{TIDAK SEHAT}) \approx 0.028 \times 0.0103 \approx 0.00029$$

Likelihood calculation for NO_2 pollutant:

$$\mu_{\text{NO2 | TIDAK SEHAT}} = 89.13, \quad \sigma_{\text{NO2 | TIDAK SEHAT}} = 92.44$$

$$P(\text{NO2} = 100|\text{TIDAK SEHAT}) = \frac{1}{\sqrt{2\pi(92.44)^2}} \times e^{-\frac{(100-89.13)^2}{2(92.44)^2}} \tag{30}$$

$$(100 - 89.13)^2 = 10.87^2 = 118.12$$

$$\text{Eksponen} = -\frac{118.12}{2 \times (92.44)^2} = -\frac{118.12}{17088.96} \approx -0.00691$$

$$e^{-0.00691} \approx 0.9931$$

$$\frac{1}{\sqrt{2\pi(92.44)^2}} = \frac{1}{\sqrt{2\pi \times 8545.44}} \approx \frac{1}{231.34} \approx 0.00432$$

$$P(\text{NO2} = 100|\text{TIDAK SEHAT}) \approx 0.00432 \times 0.9931 \approx 0.00429$$

d. Calculating posterior probability
   Posterior probability:
   P(UNHEALTHY|X)=P(UNHEALTHY)×P(PM10=20|UNHEALTHY)×P(PM2.5=18|UNHEALTHY)×P(SO2=100|UNHEALTHY)×P(CO=1500|UNHEALTHY)×P(O3=50|UNHEALTHY)×P(NO2=100|UNHEALTHY)
   Result:
   P(UNHEALTHY|X)≈0.08×0.0483×0.00927×0.00375×0×0.00029×0.00429

3.6 Confusion Matrix

At this stage, the performance testing of the implemented algorithm is carried out by utilizing the confusion matrix. Figure 6 shows the results of the confusion matrix for the “naïve bayes” algorithm.

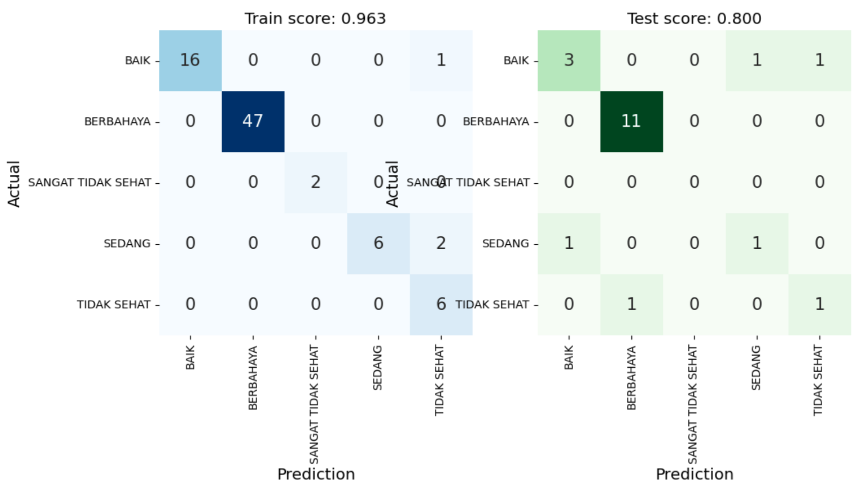

Figure 6. Confusion Matrix

4. CONCLUSION

This study successfully built an efficient and effective system to classify air quality in the Medan Industrial Area (KIM). This system uses 100 available data to provide accurate classification results, which can be used as a reference in monitoring and managing air quality in the area. The implementation of this system proves that technology can be an important tool in supporting environmental conservation efforts, especially in industrial areas. The results of the study show that the Naïve Bayes method is effective in classifying air quality in the Medan Industrial Area (KIM). This method is able to process data well and provide predictions with accurate good moderate, unhealthy, very unhealthy, hazardous variables related to air quality categories. The accuracy achieved by this method shows that Naïve Bayes is a reliable and reliable tool for similar applications in other fields that require data-based classification.

REFERENCES

[1] A. Pozzer, S. C. Anenberg, S. Dey, A. Haines, J. Lelieveld, and S. Chowdhury, "Mortality attributable to ambient air pollution: A review of global estimates," *Geohealth*, vol. 7, no. 1, p. e2022GH000711, 2023.

[2] V. Sima, I. G. Gheorghe, J. Subić, and D. Nancu, "Influences of the industry 4.0 revolution on the human capital development and consumer behavior: A systematic review," *Sustainability*, vol. 12, no. 10, p. 4035, 2020.

[3] A. Lahal, "Rancang bangun alat monitoring polusi udara berbasis Arduino," (Doctoral dissertation, Prodi Teknik Informatika)., 2021.

[4] M. O. P. Kurniawan, "Perbandingan Metode Random Forest Dan K-Nearest Neighbor Dalam Klasifikasi Kualitas Udara Di Yogyakarta," (Doctoral dissertation, Universitas Atma Jaya Yogyakarta), 2024.

[5] J. K. Sethi and M. Mittal, "Efficient weighted naive bayes classifiers to predict air quality index," *Earth Sci Inform*, vol. 15, no. 1, pp. 541–552, 2022, doi: 10.1007/s12145-021-00755-7.

[6] I. Wickramasinghe and H. Kalutarage, "Naive Bayes: applications, variations and vulnerabilities: a review of literature with code snippets for implementation," *Soft comput*, vol. 25, no. 3, pp. 2277–2293, Feb. 2021, doi: 10.1007/s00500-020-05297-6.

[7] S. Chen, G. I. Webb, L. Liu, and X. Ma, "A novel selective naïve Bayes algorithm," *Knowl Based Syst*, vol. 192, Mar. 2020, doi: 10.1016/j.knosys.2019.105361.

[8] P. Alusvigayana, A. S. Yuwono, M. Yani, and S. Syarwan, "Evaluation of the Air Pollutant Standard Index (ISPU) parameter concentration limits in industrial estates on Java Island," *Jurnal Pengelolaan Sumberdaya Alam dan Lingkungan (Journal of Natural Resources and Environmental Management)*, vol. 13, no. 4, pp. 537–548, 2023.

[9] A. S. Chan, F. Fachrizal, and A. R. Lubis, "Outcome Prediction Using Naïve Bayes Algorithm in The Selection of Role Hero Mobile Legend," *J Phys Conf Ser*, vol. 1566, no. 1, p. 012041, Jun. 2020, doi: 10.1088/1742-6596/1566/1/012041.

[10] A. Nugroho, I. Asror, and Y. F. A. Wibowo, "Klasifikasi Tingkat Kualitas Udara DKI Jakarta Berdasarkan Open Government Data Menggunakan Algoritma Random Forest," *eProceedings of Engineering*, vol. 10, no. 2, 2023.

[11] A. R. Lubis, M. K. M. Nasution, O. S. Sitompul, and E. M. Zamzami, "The feature extraction for classifying words on social media with the Naïve Bayes algorithm," *IAES International Journal of Artificial Intelligence*, vol. 11, no. 3, p. 1041, 2022.

[12] P. M. Shankar, "Pedagogy of Bayes' rule, confusion matrix, transition matrix, and receiver operating characteristics," *Computer Applications in Engineering Education*, vol. 27, no. 2, pp. 510–518, Mar. 2019, doi: 10.1002/cae.22093.

[13] L. C. Dawkins, D. B. Williamson, K. L. Mengersen, L. Morawska, R. Jayaratne, and G. Shaddick, "Where Is the Clean Air? A Bayesian Decision Framework for Personalised Cyclist Route Selection Using R-INLA," *Bayesian Anal*, vol. 16, no. 1, pp. 61–91, Mar. 2021, doi: 10.1214/19-BA1193.

[14] X. Xie, J. Zuo, B. Xie, T. A. Dooling, and S. Mohanarajah, "Bayesian network reasoning and machine learning with multiple data features: air pollution risk monitoring and early warning," *Natural Hazards*, vol. 107, no. 3, pp. 2555–2572, 2021, doi: 10.1007/s11069-021-04504-3.

[15] A. S. Thanuja Nishadi, "Text Analysis: Naïve Bayes Algorithm using Python JupyterLab," *International Journal of Scientific and Research Publications (IJSRP)*, vol. 9, no. 11, p. p9515, Nov. 2019, doi: 10.29322/IJSRP.9.11.2019.p9515.

[16] W. Weng, "Flowchart," in *A Beginner's Guide to Informatics and Artificial Intelligence: A Short Course Towards Practical Problem Solving*, W. Weng, Ed., Singapore: Springer Nature Singapore, 2024, pp. 13–20. doi: 10.1007/978-981-97-1477-3_2.