Jurnal Media Teknik Elektro dan Komputer
Vol. 01, No.01 (January – June, 2024)
E-ISSN: XXXX-XXXX

Weather Forecast Using Naive Bayes Classifier Method

Joehari Azhar[1], Widya Syaharani[2]

[1]Universitas Islam Negeri Sumatera Utara; joeazhar12@gmail.com
[2]Universitas Islam Negeri Sumatera Utara; widiasyaharani53@gmail.com

ABSTRACT

Weather forecasting is one of the important areas in human life. Planning, transportation, agriculture, and tourism are just a few of the uses for weather forecasts. Weather forecasting is the process of predicting future weather conditions. Various methods exist for weather forecasting, including manual and computer-based methods. The manual calculation process for weather prediction still lacks accuracy, so researchers conducted a study to develop a simple system that can produce more accurate weather predictions. The method used in this study is a Naïve Bayes classifier by using training data as data for an event from previously known facts or reality. The final test results, conducted on a simple system and compared with manual calculations, demonstrated a higher level of accuracy.

Keywords: Forecast, Weather, Naïve Bayes, Classification

Corresponding Author:
Joehari Azhar
Universitas Islam Negeri Sumatera Utara; joeazhar12@gmail.com

1. INTRODUCTION

The development of information technology has changed the way humans work and live. Originally limited to calculations, computers have rapidly evolved to perform a wide range of tasks previously unattainable by humans. This is thanks to the development of computer software and hardware, as well as increasingly sophisticated algorithms. Artificial intelligence (AI) is a branch of computer science that studies how to make computers think and act like humans. AI has the potential to solve complex problems, including weather forecasting. The Naïve Bayes method is one of the AI methods that uses the Bayes theorem to predict future weather conditions. In the Naïve Bayes method, weather conditions are represented as classes. These classes can include sunny, cloudy, rainy, and others. We use historical weather data to calculate the probability of each class. We then use these probabilities to predict future weather conditions.

Weather forecasting is one of the most important fields in human life. Planning, transportation, agriculture, and tourism are just a few of the uses for weather forecasting. Generally, we use statistical methods for weather forecasting. The Naive Bayes method is a commonly used statistical method for weather forecasting. According to (Prasetya, 2020), weather forecasting is the process of predicting future weather conditions. Various methods exist for weather forecasting, including manual and computer-based approaches. Statistical methods utilize data to draw conclusions. We can use this method for various purposes, including weather forecasting. This research aims to analyze the accuracy

of weather forecasts using the Naïve Bayes algorithm. The Naïve Bayes algorithm is a probability classification method that uses the Bayes theorem. This method presupposes the independence of each attribute of the target variable and offers several benefits, including its ease of implementation, adaptability to diverse weather data types, and ability to yield precise outcomes. We can use the Naive Bayes algorithm for weather forecasting, which predicts weather conditions based on historical weather data. Historical weather data can take various forms such as temperature, humidity, air pressure, among others.

2. LITERATURE REVIEW

2.1. Weather

Weather refers to the conditions of the air at a specific time and location within a brief period. The factors that influence it are air temperature, wind, and humidity. Data mining is the process of obtaining information from enormous data sets using retrieval techniques and algorithms from the fields of statistics, machine learning, and database management systems. Analyzing data from various perspectives, data mining generates important information that can either reduce costs, increase profits, or both. Technically, data mining can be defined as the process of finding patterns or correlations between hundreds or thousands of fields in a huge database (Nur Kirana et al., 2024).

2.2. Prediction

Prediction is the result of predicting, forecasting, or estimating future values using past data. Prediction is a component of planning and decision-making and shows what will happen in certain situations (Yani et al., 2022). We can use scientific or subjective methods to make predictions. For example, weather predictions are always based on the latest data and information, including satellite observations (Wijaya et al., 2021).

2.3. Naïve Bayes classifier

Naive Bayes is a simple probabilistic classifier that calculates a set of probabilities by summing the frequencies and combinations of values from a given dataset (Hayuningtyas, 2019)(Yuliyana & Sinaga, 2019). The algorithm uses Bayes' theorem and assumes all attributes are independent or not interdependent given the value of the class variable. The English scientist Thomas Bayes proposed Naive Bayes as a classification with probability and statistical methods, specifically predicting future opportunities based on previous experience (Roy & Hendriyawan, 2019). Bayesian classifiers have a high level of speed and accuracy when applied to large databases (Muzaki & Witanti, 2021).A Naive Bayes classifier is a simple classifier that calculates probabilities by summing frequencies and values from existing data. Equation (1) provides the Naive Bayes Classifier formula for classification.

$$(Y|X) = \frac{(Y) \prod i=1\ q\ (Xi\ |Y)}{P(Xi\ )} \quad (1)$$

Description:

X	: $\{X_1, X_2, X_3, \dots, X_q\}$ as many as q attributes or q
$(Y\|X)$	: Probability of data with vector X in class Y
(Y)	: Initial probability of class Y (a priori probability)
$(X\|Y)$	: Posterior probability
$\prod i=1\ q\ (Xi\ \|Y)$	: Independent probability of class Y in vector X
i	: Observation i

In the case of weather classification, these features may include temperature, humidity, and wind speed, while the target class could be weather conditions such as sunny, cloudy, or rainy (Sandfreni et al., 2021).

1. A comprehensive aspect
   The literature review should cover a variety of relevant sources, including scientific journals, books, articles, and others related to the research topic.
2. Analytical
   Beyond simply summarizing the information, the literature reviewer should also be able to evaluate and critique the existing literature, highlighting the strengths, weaknesses, and relevance of each source.
3. Structured and organized.
   In writing, a literature review must have a clear structure, organizing the sources discussed according to the theme or sub-topic.

3. METHODS

The research method for weather prediction using the Naive Bayes method can involve steps such as the following:

1. Data Collection: Collect historical weather data that includes various variables such as temperature, air humidity, wind speed, and weather conditions (e.g., sunny, cloudy, rainy).
2. Data Preparation: Performing data pre-processing such as cleaning invalid or missing data, converting categorical data into a format suitable for analysis, and dividing the data into training sets and testing sets.
3. Naive Bayes Model Training: Using the training set to train the Naive Bayes model. The model will learn from historical data to understand patterns related to the prediction variable (e.g., weather conditions) based on other variables (e.g., temperature, humidity, and wind speed) (Rizqi & Kusumaningsih, 2022).
4. Model Evaluation: Evaluation of model performance using a separate test set. Commonly used evaluation metrics for classifiers such as Naive Bayes include accuracy, precision, recall, and F1-score.
5. Prediction: Once the model has been evaluated and deemed adequate, use the model to make weather predictions based on the new data provided.
6. Analyze Results: Finally, analyze the prediction results to understand the extent to which the Naive Bayes model can be used to accurately predict the weather based on the given data.

4. RESULTS AND DISCUSSION

The first step in this study's Naive Bayes calculation method is to prepare the training data. In this study, the author used training data in five categories: weather, temperature, humidity, windy, and play, as shown in Table 1.

Table 1. Data Training

No	Weather	Temperature	Humidity	Windy	Main
1	Sunny	Hot	High	False	No
2	Sunny	Hot	High	Correct	No
3	Cloudy	Hot	High	False	Yes
4	Rain	Cool	High	False	Yes
5	Rain	Cold	Normal	False	Yes
6	Rain	Cold	Normal	Correct	Yes
7	Cloudy	Cold	Normal	Correct	Yes
8	Sunny	Cool	High	False	No

No	Weather	Temperature	Humidity	Windy	Main
9	Sunny	Cool	Normal	False	Yes
10	Rain	Cool	Normal	False	Yes

The dataset used as training data above can be used to predict an event from previously known facts or reality.

4.1. Naïve Bayes Algorithm Calculation Method

To find out the results of the 11th main data set, the implementation steps can be seen in Table 2.

Table 2. Data Testing

No	Weather	Temperature	Humidity	Windy	Main
11	Rain	Cold	High	Correct	???

The stages of the calculation process using the naive Bayes algorithm are:

1. Determine class probability
2. Determine category probability
3. Testing the results of naive Bayes classification 4.

4.2. Calculating Probability

We use probability as a value to gauge the likelihood of a random event occurring. Table 3 displays the attributes in weather forecasting research.

Table 3. Category Probabilities

Category/Attribute	Subset	Yes	No
Weather	Sunny	0,14	1
	Cloudy	0,28	0
	Rain	0,57	0
Temperature	Hot	0,14	0,66
	Cool	0,28	0,33
	Cold	0,57	0
Humidity	High	0,28	1
	Normal	0,71	0
Windy	False	0,71	0,66
	Correct	0,28	0,33

a. Probability Class

The probability class can be seen in Table 4.

Table 4. Probabilitas Class

Ya	7 /10	0,7
Tidak	3 /10	0,3

b. Category Probability

How to determine the category probability is by using Equation (2)

$$PH|X = \frac{P(H|X)P(H)}{P(X)} \quad (2)$$

4.3. *Testing the naive Bayes classification results*

Table 5 displays the classification results of the naive Bayes classification.

Table 5. Naive Bayes Classification Testing Results

No	Yes	No	Fact	Classification	Prediction
1	0,002727536	0,13068	Tidak	Tidak	Suitable
2	0,001075648	0,06534	Tidak	Tidak	Suitable
3	0,005455072	0	Ya	Ya	Suitable
4	0,022209936	0	Ya	Ya	Suitable
5	0,114647463	0	Ya	Ya	Suitable
6	0,045213084	0	Ya	Ya	Suitable
7	0,022209936	0	Ya	Ya	Suitable
8	0,005455072	0,06534	Tidak	Tidak	Suitable
9	0,028159026	0	Ya	Ya	Suitable
10	0,056318052	0	Ya	Ya	Suitable

From the results of the calculation and testing stages using the naive bayes algorithm above, the main result obtained on the 11th data point is yes.

4.4. *Use Case Diagram*

A use case diagram is an initial description of the interface of a system (Fitriana, 2020) (Kusnadi et al., 2019). The use case diagram explains how the system operates. A use case diagram is a component of the UML. According to (Sujono et al., 2019) and (Rifki et al., 2023)., UML (Unified Modeling Language) is a visual modeling language that visualizes, specifies, builds, and documents a software development system.

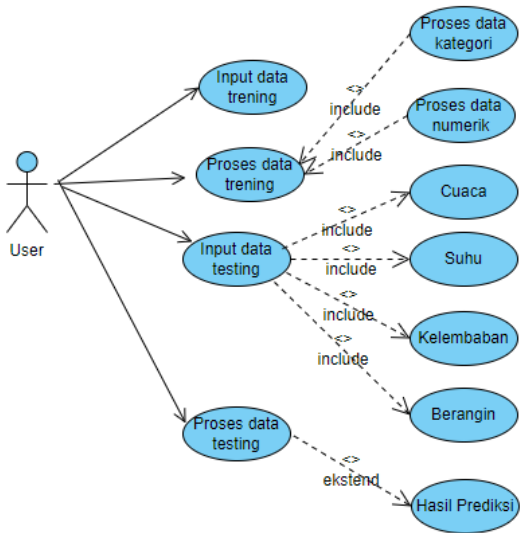

Figure 1. Use Case Diagram of the System

4.5. *System Testing Implementation*

When creating a weather forecast system, several factors need to be taken into account, which will serve as rules within the Naive Bayes algorithm. The following are some of the rules applied to the process of making a weather forecast system.

1. We assume that the data contains attributes such as weather, temperature, humidity, and windy conditions.
2. Each attribute has several possible values and the probability of occurrence of those values in the context of '*Yes*' or '*No*' categories.
3. To identify the more likely category, predictions are based on a comparison of 'Yes' and 'No' probabilities.

The Naive Bayes method is a classification algorithm based on Bayes' theorem with the assumption that the features used for classification are independent of each other (Salmi & Rustam, 2019). The Naive Bayes method goes through the following stages to create a weather forecast system. Table 6 shows the general steps to apply the Naive Bayes algorithm method to the given weather data.

Table 6. Stages of Naïve Bayes Algorithm Implementation

Steps	Description
Initial Probability Initialization	The system stores the initial probabilities for the 'Yes' and 'No' categories.
Feature Probability Initialization	The system records the likelihood of each feature, such as weather, temperature, humidity, and windiness, for both 'Yes' and 'No' classifications, using a dictionary format.
Prediction Process	The system uses weather, temperature, humidity, and wind as inputs to make predictions. The Naive Bayes formula determines the probability of each feature and the initial category probabilities ('Yes' and 'No').
Category Probability Calculation	We use the initial probabilities of 'Yes' and 'No' for each category as the starting point. To obtain the final probabilities of 'Yes' and 'No', multiply the feature probabilities for each category.
Determination of Classification Result	We are comparing the probabilities of 'Yes' and 'No' that were previously calculated. If the probability of 'Yes' is higher than the probability of 'No', the classification result is 'Yes', and vice versa.

In summary, the stages of the Naive Bayes algorithm in the system involve initializing the probabilities for each feature and category, calculating the probabilities for the categories given the input, and determining the categories based on the resulting probabilities. Figure 2 describes the following simple system for weather prediction using the Naive Bayes method/algorithm.

```
>>>
= RESTART: C:\Users\widia\Downloads\pemodelankel5.py
Masukkan cuaca (Cerah/Berawan/Hujan): Hujan
Masukkan suhu (Panas/Sejuk/Dingin): Dingin
Masukkan kelembapan (Tinggi/Normal): Tinggi
Masukkan kondisi berangin (Benar/Salah): Benar
Hasil Klasifikasi Main: Ya
Probabilitas Main: 0.01865889212827988
>>>
```

Figure 2. System Testing Results

From the test results above, the calculation of the prediction of the probability of playing using the Naïve Bayes method/algorithm applied in the system produces more accurate data, where the output displays the results of the calculation of the number of predictions of the probability of playing sought.

4. CONCLUSION

This research has successfully applied the Naive Bayes Classifier method to predict weather. The Naive Bayes Classifier method works by assuming that the features used for classification are independent of each other. This assumption can cause a decrease in accuracy if the features are not truly independent. To improve accuracy, future research can use larger and more diverse weather datasets. Additionally, researchers can conduct research to optimize the Naive Bayes Classifier method's performance, such as through the use of ensemble methods. This recommendation can assist other researchers in further exploring the Naive Bayes Classifier method for weather forecasting.

REFERENCES

Fitriana, G. F. (2020). Pengujian Aplikasi Pengenalan Tulisan Tangan menggunakan Model Behaviour Use case. *Jatisi (Jurnal Teknik Informatika Dan Sistem Informasi)*, *7*(2), 200–213.

Hayuningtyas, R. Y. (2019). Penerapan Algoritma Naïve Bayes untuk Rekomendasi Pakaian Wanita. *Jurnal Informatika*, *6*(1), 18–22.

Kusnadi, I. T., Supiandi, A., Kusnadi, W., & Riniawati, R. (2019). Pengembangan Sistem Inventori Perusahaan Menggunakan Metode Usecase Driven. *Jurnal Teknologi Dan Informasi*, *9*(1), 24–36.

Muzaki, A., & Witanti, A. (2021). Sentiment analysis of the community in the twitter to the 2020 election in pandemic covid-19 by method naive bayes classifier. *Jurnal Teknik Informatika (Jutif)*, *2*(2), 101–107.

Nur Kirana, A., Nurhakim, B., Eka Permana, S., Prihartono, W., & Dwilestari, G. (2024). Implementasi Algoritma Naive Bayes Untuk Memprediksi Cuaca Menggunakan Rapidminer. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(2), 1637–1642. https://doi.org/10.36040/jati.v8i2.8967

Prasetya, R. (2020). PENERAPAN TEKNIK DATA MINING DENGAN ALGORITMA CLASSIFICATION TREE UNTUK PREDIKSI HUJAN. *Jurnal Widya Climago, 2*(2).

Rifki, M. I., Raditya, M. E., & Hasugian, A. H. (2023). Text Data Security Application Using a Mobile-Based Base64 Algorithm. *Instal: Jurnal Komputer*, *15*(02), 224–235.

Rizqi, A. A., & Kusumaningsih, D. (2022). Klasifikasi Curah Hujan di Kota Bogor Provinsi Jawa Barat dengan Menggunakan Metode Naive Bayes. *Jurnal Seminar Nasional Mahasiswa Fakultas Teknologi Informasi (SENAFTI)*, *September*, 542–550.

Roy, A., & Hendriyawan, A. M. (2019). *Implementasi Metode Naive Bayes Classifier Untuk Perkitaan Cuaca*. 12.

Salmi, N., & Rustam, Z. (2019). Naïve Bayes classifier models for predicting the colon cancer. *IOP Conference Series: Materials Science and Engineering*, *546*(5), 52068.

Sandfreni, S., Ulum, M. B., & Azizah, A. H. (2021). Analisis Perancangan Sistem Informasi Pusat Studi Pada Fakultas Ilmu Komputer Universitas Esa Unggul. *Sebatik*, *25*(2), 345–356. https://doi.org/10.46984/sebatik.v25i2.1587

Sujono, S., Mayasari, M. S., & Koloniawan, K. (2019). Prototipe Aplikasi Simpan Pinjam Pada Koperasi Darma Karya Pangkalpinang Babel. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, *8*(1), 68–73. https://doi.org/10.32736/sisfokom.v8i1.609

Wijaya, E., Tarigan, F. A., & Michael. (2021). Aplikasi Prediksi Penentuan Kelancaran Pembayaran KoperasiDengan Algoritma C5.0. *Jurnal Times: Technology Informatics & Computer System*, *X*(1), 31–38.

Yani, V. I., Aradea, A., & Mubarok, H. (2022). Optimasi Prakiraan Cuaca Menggunakan Metode

Ensemble pada Naïve Bayes dan C4. 5. *Jurnal Teknik Informatika Dan Sistem Informasi*, *8*(3), 607–619.

Yuliyana, Y., & Sinaga, A. S. R. M. (2019). Sistem Pakar Diagnosa Penyakit Gigi Menggunakan Metode Naive Bayes. *Fountain of Informatics Journal*, *4*(1), 19–23.