International Journal of Psychosocial Rehabilitation, Conference Special Issue
ISSN: 1475-7192

Improving the Accuracy of Weather Forecasting Prediction on Data Cloud by Using Naive Bayes Classifier

[1]Rizky Maulana, [2]Opim Salim Sitompul , [3]Herman Mawengkang

ABSTRACT--*Weather Forecasting is a problem that can be predicted by using research. In this research, we use Naive Bayes Classifier to predict the weather on Data Cloud that we take from meteoblue.com and we tested it in Aceh, which are Bireuen and Gayo. The results shown that the accuracy of the prediction is about 89,47%.*

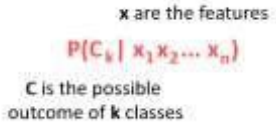
Keywords—*Accuracy, Weather, Forecasting ,Prediction, Naive Bayes Classifier*

I. INTRODUCTION

Weather prediction is an activity to predict the condition of the Earth's atmosphere for the future with or without the help of technology. Humans have tried to make an accurate weather prediction, from the time of the ancient until now with technology and computer assistance. The main focus on using computer assistance is to improve a better accuracy of the prediction.

II. LITERATURE REVIEW

Naive Bayes Classifier is a classifier model that using a probability and statistic that introduced by Thomas Bayes. The model behind Naive Bayes Classifier has something to do with probability distributions. The aim is to maximize the probability of the target class given the x features.



So the x features can be the income, age and LTI. The C possible outcomes can be 0 or 1 when dealing with the credit scoring problem. 0 if the client defaulted and 1 if the client is able to pay back the debt. We can use Bayestheorem to decompose the conditional probability.

[1] *Department of Computer Science, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia*

[2] *Department of Computer Science, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia*

[3] *Department of Computer Science, Faculty of Computer Science and Information Technology, Universitas Sumatera Utara, Medan, Indonesia*

Received: 02 Jan 2020 | Revised: 12 Feb 2020 | Accepted: 17 Mar 2020

$$P(C_k \mid x_1 x_2 \dots x_n) = \frac{p(C_k)\, p(x \mid C_k)}{p(x)}$$

the same for all c_k values

And here comes the naive assumption that the features are independent of each other. Two random variables are independent if the realization of one does not affect the probability distribution of the other. Hence, we can decompose the joint distribution.

$$P(x_1 x_2 \dots x_n \mid C_k) = P(x_1 \mid C_k)\, P(x_2 \mid C_k) \dots P(x_n \mid C_k)$$

$$P(C_k \mid x_1 x_2 \dots x_n) \sim p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k)$$

You may ask what happened to the p(x) denominator? Because it is the same for all c values, we can omit it. Finally, Naive Bayes Classifier picks the class with the highest probability. So the problem reduces to a maximum finding problem (the dominator does not affect this value).

$$C = \arg\max_{c_k} p(C_k) \prod_{i=1}^{n} p(x_i \mid C_k)$$

III. RESEARCH METHOD

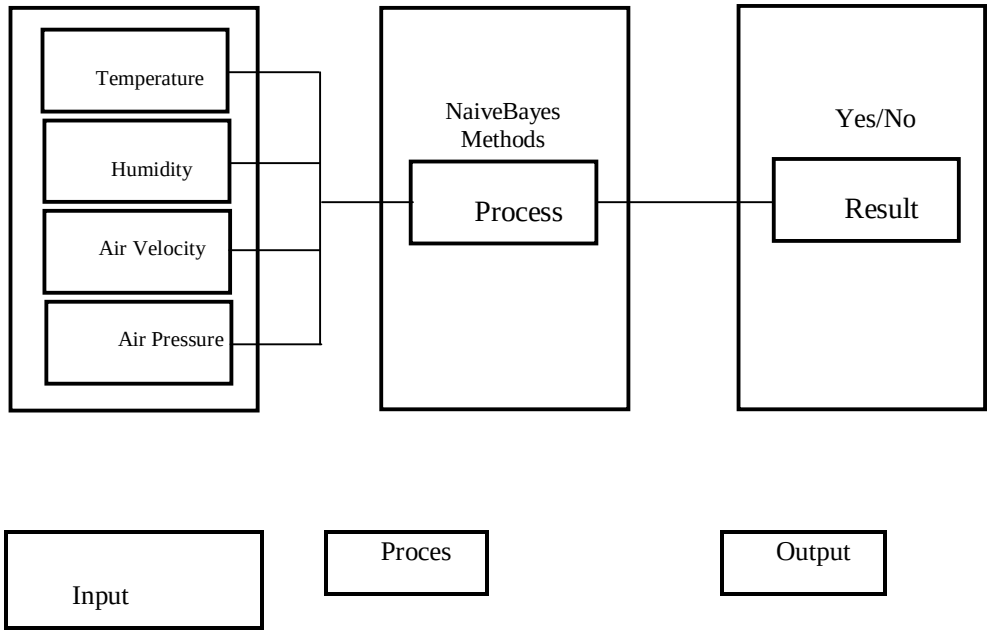


Figure 1: Proposed Method

Based on Figure 1, the stages of research carried out in determining the weather conditions is inputting the process from taking data in four weather variables in the form of temperature, humidity, air velocity and air pressure, then collected to be processed using the Naïve Bayes algorithm. After the temperature at the input, then carry out the stages of the process of collecting temperature data, air humidity, air velocity and air pressure that

are useful for the training data, to categorize data on the temperature scale by cold / medium / hot, comfortable / humid humidity, fast / standard air velocity and normal / no air pressure, and then the calculate the probability the occurrence of each class.

IV. RESULTS AND DISCUSSIONS

In this section contains the result that obtained in this research, to analyze the comparation of the prediction with the actual data, we provide the detail as follows Table 1.

Time	Temp		RH		Velocity		Pressure	
	Predict	Actual	Predict	Actual	Predict	Actual	Predict	Actual
Earlier	26,17	26,65	92,69	92,50	1,08	1,55	1011,20	1011,05
Morning	30,14	29,64	69,37	71,80	1,47	1,00	1012,18	1012,20
Day	30,17	28,99	71,06	77,20	2,02	1,50	1009,72	1009,70
Night	27,25	27,12	87,37	90,75	0,88	0,90	1012,55	1012,30

Table 1: Comparation of Prediction

4.1 Naive Bayes Calculation

The steps of manual calculation for the data classification process using the Naive Bayess Classifier method are as follows:

1. Calculate the Probabiliies

$P(Class|Yes) = \frac{16}{38} = 0.421$

$P(Class|No) = \frac{22}{38} = 0.578$

2. Create The Frequency Table from the training data

Frequency Table		Condition	
		Yes	**No**
Temp	Cold	5	8
	Normal	5	8
	Hot	6	6

Frequency Table		Condition	
		Yes	**No**
Humidity	Dry	3	9
	Normal	9	5
	Humid	4	8

Frequency Table		Condition	
		Yes	**No**
Air Velocity	Normal	11	8

	Fast	5	14

Frequency Table		Condition	
		Yes	No
Air Pressure	Normal	13	6
	No	3	16

Table 2: Frequency Table

V. CALCULATE THE PROBABILITIES

In this section we calculate the probabilities on each City using Naive Bayes Classifier which the results shown in figures below.

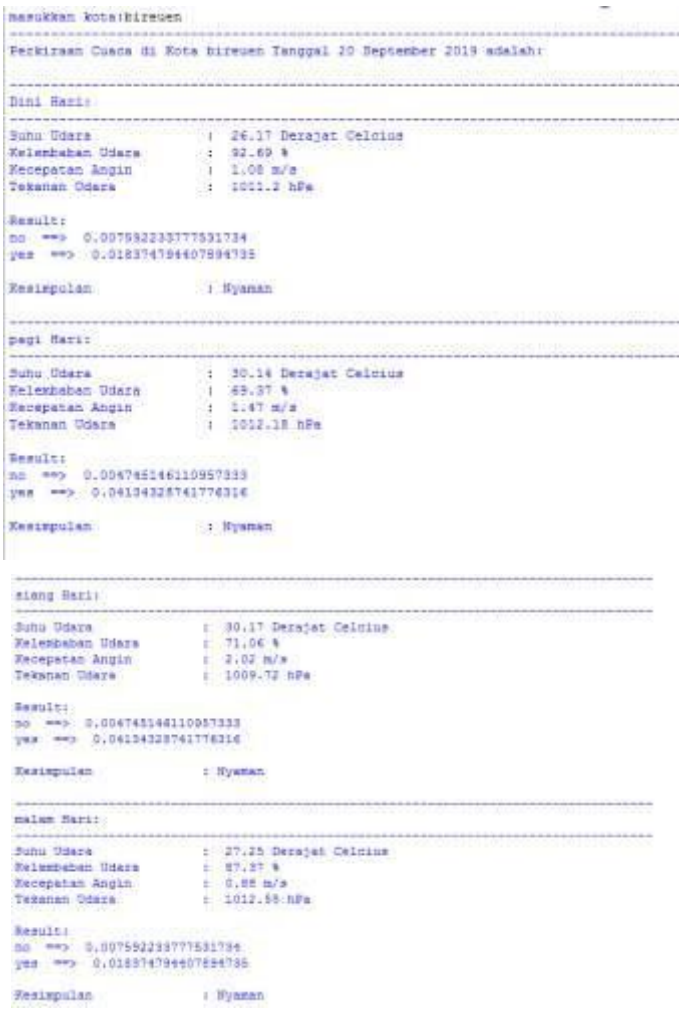


Figure 2: Results test on Bireuen City

The results of the weather prediction process using the naïve bayes method shown that the weather clusters between the comfortable or the uncomfortable based on real weather data as air temperature, humidity, air velocity

and air pressure obtained from the cloud server (meteoblue.com). The data is first entered into the class for each weather parameter, then the p(class) value is calculated, which is getting the result in Bireuen City is comfortable.

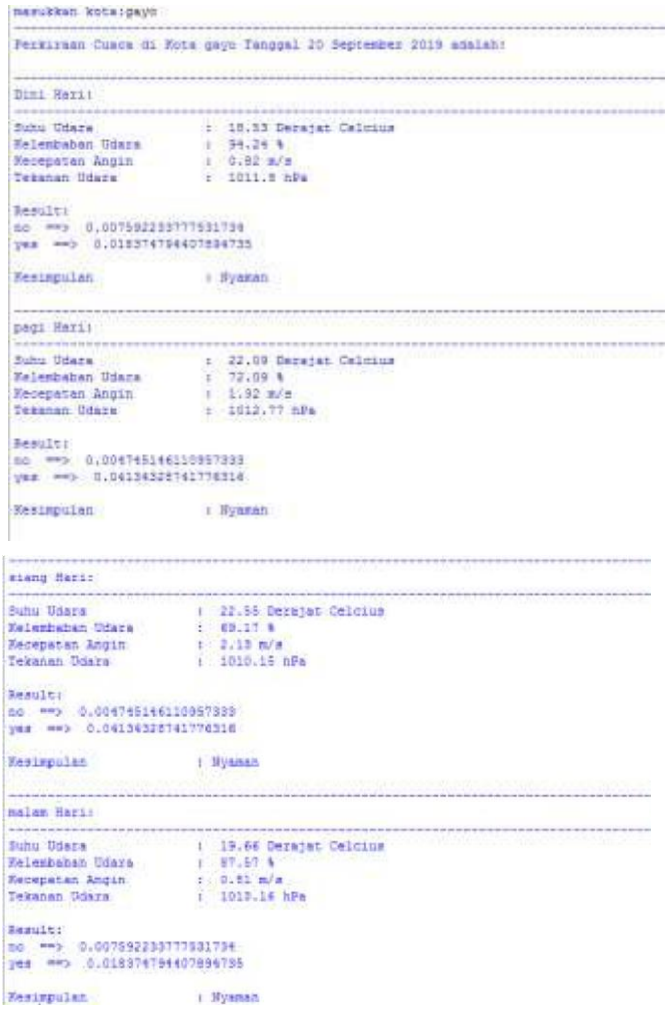


Figure 3: Results test on Gayo City

The results of the weather prediction process using the naïve bayes method shown that the weather clusters between the comfortable or the uncomfortable based on real weather data as air temperature, humidity, air velocity and air pressure obtained from the cloud server (meteoblue.com). The data is first entered into the class for each weather parameter, then the p(class) value is calculated, which is getting the result in Gayo City is comfortable. The manual calculations for the accuracy as follows:

$$Accuracy = \frac{21 + 13}{38} = 0{,}8947 = 89{,}47\ \%$$

VI. CONCLUSIONS

The accuracy value for the weather prediction using Naive Bayes Classifier is 89,47% which means this algorithm works properly on predicting the weather forecast. This study suggest that the accuracy of the method can significantly improve by combining the Naive Bayes Classifier with another conjungtion methods.

REFERENCES

1. Abusharha, A. A., & Pearce, E. I. (2013). The effect of low humidity on the human tear film. Cornea, 32(4), 429-434
2. Ashton, K. (2009). That 'internet of things' thing. RFID journal, 22(7), 97-114 3. Botthof, A., & Hartmann, E. A. (2015). Zukunft der Arbeit in Industrie 4.0: Springer Vieweg Berlin
3. De Freitas, C. (2015). Weather and place-based human behavior: recreational preferences and sensitivity. International journal of biometeorology, 59(1), 55-63
4. Ngai, E. W., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. Decision support systems, 50(3), 559-569
5. Padhi, N. (2017). Setting up a Smart Factory (Industry 4.0)-A Practical Approach. MESA International, 13-17
6. Sangkertadi, S. R., & Tungka, A. (2008). Thermal Comfort Comparison of Traditional Architecture and Modern Style Housing in North Sulawesi Indonesia. Proceeding of 9th SENVAR+ 2nd ISESEE, Selangor, Malaysia, 1-3.