

PAPER • OPEN ACCESS

Predicting Rainfall Intensity using Naïve Bayes and Information Gain Methods (Case Study: Sleman Regency)

To cite this article: I G W Sena *et al* 2020 *J. Phys.: Conf. Ser.* **1577** 012011

View the [article online](#) for updates and enhancements.



IOP | ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection—download the first chapter of every title for free.

Editorial Board Members

Subramaniam Ananthakrishnan

Pavel Belov

Charles Cavalcante Casimiro

Sergio Colafrancesco

Mérouane Debbah

Lars Jacob Foged

Debatosh Guha

Sébastien Lalléchère

Jean-Daniel Lan Sun Luk

Dominique Lesselier

André de Lustrac

James McLean

Eric Mokole

Vikass Monebhurrin

Shailendra Oree

Lionel Pichon

Blaise Ravelo

Tapan Sarkar

Russell Taylor

Bernard Veyret



Table of contents

Volume 1577

2020

◀ Previous issue Next issue ▶

The 2nd 2019 ICERA: International Conference on Electronics Representation and Algorithm "Innovation and Transformation for Best Practices in Global Community" 12-13 December 2019, Yogyakarta, Indonesia

Accepted papers received: 16 June 2020

Published online: 15 July 2020

Open all abstracts

Preface

OPEN ACCESS 011001

Preface

+ Open abstract  View article  PDF

OPEN ACCESS 011002

Sponsors

+ Open abstract  View article  PDF

OPEN ACCESS 011003

Peer review statement

+ Open abstract  View article  PDF

Computational Methods

OPEN ACCESS 012001

Decision Support System for Determining Floating Dock Locations in Maluku Islands Using AHP-TOPSIS

M Tukan, B J Camerling, M T Afifudin and Hozairi

+ Open abstract  View article  PDF

This site uses cookies. By continuing to use this site you agree to our use of cookies. To find out more, see our Privacy and Cookies policy.

012001 

Deep Contextual of Document Using Deep LSTM Meet Matrix Factorization to Handle Sparse Data: Proposed Model

Hanafi, N Suryana and ASH Basari

[+ Open abstract](#) [View article](#) [PDF](#)

OPEN ACCESS

012003

Deep learning-based object detection and geographic coordinate estimation system for GeoTiff imagery

B M Pratama, D Gunawan and R A G Gultom

[+ Open abstract](#) [View article](#) [PDF](#)

OPEN ACCESS

012004

Electroencephalography based Emotion Recognition using Fisher's Linear Discriminant Analysis on Support Vector Machine

I N Yulita, D Novita, A Sholahuddin and Emilliano

[+ Open abstract](#) [View article](#) [PDF](#)

OPEN ACCESS

012005

Fuzzy Logic System Implementation with Mamdani Method in Computer-Based Intelligence Quotient Test to Determining the Type of Intelligence Dimension

G N Putri, B Dirgantoro, P Aulia and C Setianingsih

[+ Open abstract](#) [View article](#) [PDF](#)

OPEN ACCESS

012006

Generate Contextual Insight of Product Review Using Deep LSTM and Word Embedding

Hanafi, N Suryana and ASH Basari

[+ Open abstract](#) [View article](#) [PDF](#)

OPEN ACCESS

012007

Implementation of Support Systems for Determination of Amphibious Vehicle Landing in Disaster Emergency Response Using Fuzzy Takagi Sugeno

Abdurahman, G Harsono, Y Prihanto and R A G Gultom

[+ Open abstract](#) [View article](#) [PDF](#)

OPEN ACCESS

012008

Information System of Agricultural Commodities Mapping Based on Machine Learning

D Nur, C Riyanti and M Olivya

[+ Open abstract](#) [View article](#) [PDF](#)

This site uses cookies. By continuing to use this site you agree to our use of cookies. To find out more, see our Privacy and Cookies policy.



OPEN ACCESS

012009

Least-Square Support Vector Machine (LS-SVM) Parameters Optimization using Hybrid Cuckoo Search and Harmony Search Algorithm for Pre Collision Warning on Driver Assistance System (DAS)

A Sumarudin, A Puspaningrum and A Suheryadi

[+ Open abstract](#) [View article](#) [PDF](#)

OPEN ACCESS

012010

Non-periodic Noisy Signals Denoising Using Adaptive Neuro-Fuzzy Inference System (ANFIS)

I Santoso, A Warsito, T Prakoso, A Sofwan, A A Zahra, Y Christyono and M A Riyadi

[+ Open abstract](#) [View article](#) [PDF](#)

OPEN ACCESS

012011

Predicting Rainfall Intensity using Naïve Bayes and Information Gain Methods (Case Study: Sleman Regency)

I G W Sena, J W Dillak, P Leunupun and A J Santoso

[+ Open abstract](#) [View article](#) [PDF](#)

OPEN ACCESS

012012

The Medical Facilities Selection Based on Location-Based Services Application Using SAW and TOPSIS Algorithm

M Z Rohman, Irwansyah and W E Sari

[+ Open abstract](#) [View article](#) [PDF](#)

OPEN ACCESS

012013

TOPSIS Method for Decision Support Systems in Determining the Interests of Medical Student

E Riswanto, D R Melany, B S Wiratama and Syafrianto

[+ Open abstract](#) [View article](#) [PDF](#)

Computer Graphics and Vision

OPEN ACCESS

012014

A New Image Segmentation of Leptomeningeal Metastasis in Leukemia Patients

H Z Iilmadina, A M Arymurthy and Rosalina

[+ Open abstract](#) [View article](#) [PDF](#)

This site uses cookies. By continuing to use this site you agree to our use of cookies. To find out more, see our Privacy and Cookies policy.

012014

An Analysis and Comparison Performance of DNA and Chaotic Method Combination for Image Encryption

V Saputra, D R I M Setiadi and E H Rachmawanto

[+ Open abstract](#) [View article](#) [PDF](#)

OPEN ACCESS

012016

Detecting Ictal and Interictal Condition of EEG Signal using Higuchi Fractal Dimension and Support Vector Machine

I Wijayanto, S Hadiyoso, S Aulia and B S Atmojo

[+ Open abstract](#) [View article](#) [PDF](#)

OPEN ACCESS

012017

Detecting the Burned Area in Volcanic Region by Using Multitemporal Landsat-8 OLI (Case Study: Mt. Sumbing, Central Java)

I Prasasti, D Triyono and Suwarsono

[+ Open abstract](#) [View article](#) [PDF](#)

OPEN ACCESS

012018

Finite State Machines for Building Believable Non-Playable Character in the Game of Khalid ibn Al-Walid

K Fathoni, R Y Hakkun and H A T Nurhadi

[+ Open abstract](#) [View article](#) [PDF](#)

OPEN ACCESS

012019

Real Time Video Analytics Based on Deep Learning and Big Data for Smart Station

F Hidayat, F Hamami, I A Dahlan, S H Supangkat, A Fadillah and A Hidayatuloh

[+ Open abstract](#) [View article](#) [PDF](#)

OPEN ACCESS

012020

Shallot Quality Classification using HSV Color Models and Size Identification based on Naive Bayes Classifier

A Susanto, Z H Dewantoro, C A Sari, D R I M Setiadi, E H Rachmawanto and I U W Mulyono

[+ Open abstract](#) [View article](#) [PDF](#)

Modeling and Simulation

OPEN ACCESS

012021

A Fuzzy Logic-based Control System for Microwave Ovens

This site uses cookies. By continuing to use this site you agree to our use of cookies. To find out more, see our Privacy and Cookies policy.



[+ Open abstract](#) [View article](#) [PDF](#)

OPEN ACCESS

012022

Automatic Guided Vehicle that detects *Dactylopius Opuntiae* in Cactus Pears

N J Luwes, J C Ogochukwu and B P Rskotsoane

[+ Open abstract](#) [View article](#) [PDF](#)

OPEN ACCESS

012023

Automatic Question Generator System Conceptual Model for Mathematic and Geometry Parallel Question Replication

A A B Prasetyanto, T B Adji and I Hidayah

[+ Open abstract](#) [View article](#) [PDF](#)

OPEN ACCESS

012024

Collaborative Decision-Making for Human-Technology Interaction - A Case Study Using an Automated Water Bottling Plant

J Coetzer, R B Kuriakose and H J Vermaak

[+ Open abstract](#) [View article](#) [PDF](#)

OPEN ACCESS

012025

Development of Radar-based Sensor System for Smart Level Crossing Technology

M Rosyidi, N Irawati, S Nugroho, S Bismantoko, T Widodo, A Harvono and U Chasanah

[+ Open abstract](#) [View article](#) [PDF](#)

OPEN ACCESS

012026

Distributed Classifier for SDGs Topics in Online News using RabbitMQ Message Broker

A Nugroho, Widyawan and S S Kusumawardani

[+ Open abstract](#) [View article](#) [PDF](#)

OPEN ACCESS

012027

Generation of Synthetic Continuous Numerical Data Using Generative Adversarial Networks

A H Aziira, N A Setiawan and I Soesanti

[+ Open abstract](#) [View article](#) [PDF](#)

OPEN ACCESS

012028

Identification the Maturity Level of *Carica Papaya* Using the K-Nearest Neighbor

I B Suban, A Paramarthy, M Fortwonatus and A J Santoso

[+ Open abstract](#) [View article](#) [PDF](#)

This site uses cookies. By continuing to use this site you agree to our use of cookies. To find out more, see our [Privacy and Cookies policy](#)



-
- OPEN ACCESS** 012029
Interference Mitigation using Adaptive Beamforming with RLS Algorithm for Coexistence between 5G and Fixed Satellite Services in C-Band
C B Muhammad and K Anwar
[+](#) [Open abstract](#) [View article](#) [PDF](#)
-
- OPEN ACCESS** 012030
Navigation system for an automatic guided vehicle
E M Ngandu, N Luwes and K Kusakana
[+](#) [Open abstract](#) [View article](#) [PDF](#)
-
- OPEN ACCESS** 012031
Networked Control System in Quadrotor Altitude Control with Time Delay Compensation
R Panuntun, O Wahyunggoro, S Herdjunto, A R Rafsanjani and N Setiawan
[+](#) [Open abstract](#) [View article](#) [PDF](#)
-
- OPEN ACCESS** 012032
Securing Text Messages using the Beaufort-Vigenere Hybrid Method
E Sugiarto, D R I M Setiadi, A Fahmi, E H Rachmawanto, C A Sari, Md K Sarker and B Widjanto
[+](#) [Open abstract](#) [View article](#) [PDF](#)
-
- OPEN ACCESS** 012033
Self Adaptive and Simulated Annealing Hyper-Heuristics Approach for Post-Enrollment Course Timetabling
H M Kartika and M Ahmad
[+](#) [Open abstract](#) [View article](#) [PDF](#)
-
- OPEN ACCESS** 012034
System Identification of Two-Wheeled Robot Dynamics Using Neural Networks
N Uddin
[+](#) [Open abstract](#) [View article](#) [PDF](#)
-
- OPEN ACCESS** 012035
Wireless communication design of internet of things based on FPGA and WiFi Module
F W Wibowo
[+](#) [Open abstract](#) [View article](#) [PDF](#)


This site uses cookies. By continuing to use this site you agree to our use of cookies. To find out more, see our Privacy and Cookies policy.



Scientific Analysis and Workflow

-
- OPEN ACCESS** 012036
A Cost-Aware Strategy for Deadline Constrained Scientific Workflows
S Manam, K Moessner and S Vural
[+](#) Open abstract [View article](#) [PDF](#)
-
- OPEN ACCESS** 012037
A survey of technical efficiency in crane systems using POET structure
SF Phiri, K Kusakana and BP Numbi
[+](#) Open abstract [View article](#) [PDF](#)
-
- OPEN ACCESS** 012038
CFD Analysis of Airflow Through Prism Obstacles Inside Solar Air Heater Channel
L Diana, A G Safitra, D Ichسانی and S Nugroho
[+](#) Open abstract [View article](#) [PDF](#)
-
- OPEN ACCESS** 012039
Comparison and Analysis of Stator Plate Holder on Yokeless and Segmented Armature Machine
A F Desanti, D A Asfani, M N Yuniarto and Y U Nugraha
[+](#) Open abstract [View article](#) [PDF](#)
-
- OPEN ACCESS** 012040
Comparison of EMD, VMD and EEMD Methods in Respiration Wave Extraction Based on PPG Waves
S Hadiyoso, E M Dewi and I Wijayanto
[+](#) Open abstract [View article](#) [PDF](#)
-
- OPEN ACCESS** 012041
Design of Phishing Simulation Dashboard Using Analytic Data Concepts
R Septiana and R K Julian
[+](#) Open abstract [View article](#) [PDF](#)
-
- OPEN ACCESS** 012042
Development of Web-based Research and Community Service Database at Universitas Negeri Surabaya
Warju, S E Cahyaningrum, Nurkholis, L Saksono, S R Nudin and S R Ariyanto
This site uses cookies. By continuing to use this site you agree to our use of cookies. To find out more, see [our Open Access Cookies Policy](#).
[+](#) Open abstract [View article](#) [PDF](#)



-
- OPEN ACCESS** 012043
Energy Management Methodology for Sustainable Water Development and Servicing, Considering the POET Based Concept
P B Ngancha, K Kusakana and E D Markus
[+](#) Open abstract [View article](#) [PDF](#)
-
- OPEN ACCESS** 012044
Evaluation on EMG Electrode Reduction in Recognizing the Pattern of Hand Gesture by Using SVM Method
H A Winarno, A I Poernama, I Soesanti and H A Nugroho
[+](#) Open abstract [View article](#) [PDF](#)
-
- OPEN ACCESS** 012045
IoT Water Monitor Implementation Strategy
G A Gericke and R B Kuriakose
[+](#) Open abstract [View article](#) [PDF](#)
-
- OPEN ACCESS** 012046
IT Governance Audit and Determination of Work Priorities Using Analytical Hierarchy Process: Case Study the Government of North Maluku, Indonesia
A Arief, D Natsir, A Khairan and D I Sensuse
[+](#) Open abstract [View article](#) [PDF](#)
-
- OPEN ACCESS** 012047
Machine to Machine Communication Protocol for SMART Manufacturing Units
G A Gericke, R B Kuriakose, H J Vermaak and Ole Madsen
[+](#) Open abstract [View article](#) [PDF](#)
-
- OPEN ACCESS** 012048
Naïve Bayes Method to Determine Learning Specialization for New Students
M Wahyu, A A Munaji, R A Halim and A J Santoso
[+](#) Open abstract [View article](#) [PDF](#)
-
- OPEN ACCESS** 012049
Optimal energy management of Ice thermal energy storage-based air conditioning system for commercial buildings in real-time – A review based on POET framework
O Y Odufuwa, K Kusakana and B P Numbi
[+](#) Open abstract [View article](#) [PDF](#)
-
- This site uses cookies. By continuing to use this site you agree to our use of cookies. To find out more, see our Privacy and Cookies policy. 

-
- OPEN ACCESS** 012050
PAMELA-CL: Partition Membership Based on Lazy Classifier for Neuromarketing
I N Yulita, A Sholahuddin, Emilliano and D Novita
[+ Open abstract](#) [View article](#) [PDF](#)
-
- OPEN ACCESS** 012051
Performance Analysis of Algorithms on Different Types of Health Related Datasets
N N Khanom, F Nihar, S S Hassan and L Islam
[+ Open abstract](#) [View article](#) [PDF](#)
-
- OPEN ACCESS** 012052
POET Structured energy management and efficiency improvement of a grid-integrated electric vehicle energy Charging Stations
L Bokopane, K Kusakana and H Vermaak
[+ Open abstract](#) [View article](#) [PDF](#)
-
- OPEN ACCESS** 012053
Prototype of Integrated Livestock Recording Application with Animal Identification and Certification System in Kebumen
U Subagyo and D Ardiansyah
[+ Open abstract](#) [View article](#) [PDF](#)
-
- OPEN ACCESS** 012054
Reverse Engineering Website Navigation Using an Information Architecture Approach (Case Study: Kanal Pengetahuan Universitas Gadjah Mada)
M Fikri, S S Kusumawardani and R Ferdiana
[+ Open abstract](#) [View article](#) [PDF](#)
-
- OPEN ACCESS** 012055
Surface Deformation due to the 2017-2018 Agung Volcano Eruption from Interferometric Synthetic Aperture Radar (InSAR) Sentinel-1 TOPS
Suwarsono, D Triyono, M R Khomarudin and Rokhmatuloh
[+ Open abstract](#) [View article](#) [PDF](#)
-
- OPEN ACCESS** 012056
The Comparison of Cloud Migration Effort on Platform as a Service
R B Suryawan, R Ferdiana and Widyawan
[+ Open abstract](#) [View article](#) [PDF](#)

This site uses cookies. By continuing to use this site you agree to our use of cookies. To find out more, see our Privacy and Cookies policy.



JOURNAL LINKS

[Journal home](#)

[Journal Scope](#)

[Information for organizers](#)

[Information for authors](#)

[Contact us](#)

[Reprint services from Curran Associates](#)

This site uses cookies. By continuing to use this site you agree to our use of cookies. To find out more, see our Privacy and Cookies policy.



Predicting Rainfall Intensity using Naïve Bayes and Information Gain Methods (Case Study: Sleman Regency)

I G W Sena¹, J W Dillak², P Leunupun³ and A J Santoso⁴

^{1,2,3,4}Universitas Atma Jaya Yogyakarta, Jl. Babarsari No.44, Janti, Caturtunggal, Depok, Sleman, Daerah Istimewa Yogyakarta 55281

E-mail: albjoko@staff.uajy.ac.id

Abstract. Climate change, which has an impact on environmental problems in tropical countries, is still a severe problem, and efforts to prevent and manage it are continuously pursued. Indonesia, as a tropical country with topographical conditions and strategic geographical position, causes Indonesia to have different weather and climate patterns. Climatologically there are significant differences between the rainy season and the dry season. Both these seasons can bring blessings but also disasters if not appropriately handled, flooding in the rainy season and drought in the dry season. High rainfall can cause floods and landslides, whereas if using excess water in the rainy season can be a solution for water shortages in the dry season. The purpose of this study is to predict the rainfall intensity with the Naïve Bayes method and what parameters are considered the most dominant causes of heavy rainfall using the information gain method. The source of the data in this study came from BMKG data, which was during the rainy season between October to March from 2016 - 2019 in the Sleman Regency. The results of this study showed that the Naïve Bayes method could be used to predict rainfall intensity in Sleman Regency. Also, the most influential parameter on rainfall intensity is the average temperature with an information gain value of 0.047811028.

1. Introduction

Various conditions arising from climate change are global warming, which affects work productivity and impacts on economic decline [1], increased global emissions, greenhouse effect, warmer sea, rising sea level. These conditions cause damage to coral reefs and coastal settlements, high strength of tropical cyclones and rainstorms, groundwater that cannot be consumed because it becomes saltier [2].

As a tropical country with a rainy and dry season, Indonesia also feels the impact of climate change that is happening. Rising land-sea temperature gradients, affecting rainfall patterns that also impact human life, drought, floods, health, water and food supplies, changes in ecosystems, and infrastructure [3]. One way to overcome the disaster is through crisis planning and control, by predicting Iranian rainfall applying rainwater harvest management that has been proven to control the drought crisis and save water needs [4].

Rainfall intensity prediction can be useful information for planning flood and drought risk management, also could be useful for preparing outdoor activities. The purpose of this study is to determine the prediction of rainfall intensity and determine which parameters are the most dominant causes of rainfall intensity using Naïve Bayes method and Information Gain in Sleman Regency, Daerah Istimewa Yogyakarta (DIY) Province, Indonesia. Researchers used the Naïve Bayes method because, this method is considered more accurate to be used in terms of comparison and prediction [5].



2. Literature Review

Data Mining (DM) is one of the fields in computer science that demands the extraction of hidden information from various data storage, data marts, and repositories [6]. DM allows researchers to find beneficial information from available data. This discovery is significant in business because it can be used as a key for making decisions [7]. There are several methods to do DM, such as decision tree, neural networks, Naïve Bayes, K-nearest neighbor, etc. This research will use the Naïve Bayes method to predict the student's performance.

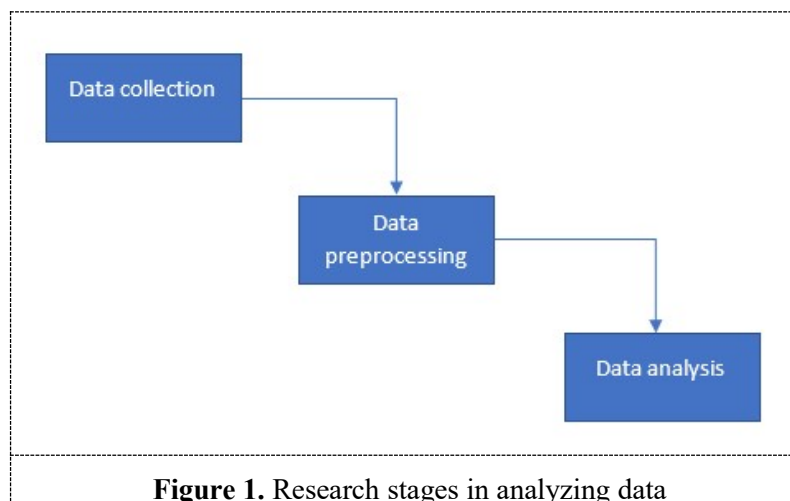
The Naïve Bayes method has a simple algorithm but can produce better results than other methods, mainly when used at high input dimensions [8]. Research by T. Devasia et al. [5] using the Naive Bayes method to predict student performance in Amrita Vishwa Vidyapeetham University, Mysuru. In this study, there were 700 responses with 19 attributes. The results of this study indicate that, in terms of accuracy and comparison, the Naive Bayes method algorithm is more accurate compared to other methods such as Neural Network, Decision Tree, etc.

Another research to [9] predicted rainfall intensity in Malang City using Naïve Bayes and Laplace Estimator method. The result of this research proved that Naïve Bayes and Laplace Estimator method could be used to predict rainfall time series for seven days with a percentage accuracy of 97.74%, 97.74% for precision, 100% for sensitivity and error rate of 2.26%.

Conceptually, the studies above are not much different from this research, which is to make predictions using Naïve Bayes. However, this study focuses on the prediction of rainfall intensity in Sleman Regency by using four parameters, namely average temperature, average humidity, duration of sun exposure, and average wind speed. This study will also determine which parameters most influence the intensity of rainfall using Information Gain, which is not done in the studies above.

3. Material and Methods

This section will explain some of the stages of researchers in the process of analyzing data. This stage is divided into data collection, data pre-processing, data analysis.



3.1. Data collection

Data sources in this study were obtained from the Indonesian Meteorology, Climatology and Geophysics Agency, or *Badan Meteorologi Klimatologi dan Geofisika* (BMKG) Indonesia, which was downloaded on the official BMKG website (<http://dataonline.bmkg.go.id/>). Data that have been downloaded are rainfall data, as well as data for each parameter used to predict rainfall, include average temperature, average humidity, duration of sun exposure, and average wind speed in Sleman Regency, Province of DIY, Indonesia. Data collected are only the data during the rainy season in Indonesia, namely in the range of October - March with a total of 577 data, starting from October

(2016-2018), November (2016-2018), December (2016-2018), January (2017-2019), February (2017-2019), and March (2017-2019).

3.2. Data pre-processing

After the data is collected, the next step is data pre-processing. Pre-processing data is a necessary step for preparing the dataset before applying classification techniques [7]. At this step, all the data is inputted in Microsoft Excel, then checking whether there is an error or invalid data to be used. The unused data is the rainfall data with a value of 0, as well as data with a value of 8888 (unmeasured data) and 9999 (no data or no measurement). From a total of 577 data, after passing the pre-processing step, there were 329 valid data and 248 invalid data.

At this step, classification and weighting of class values are also carried out for each parameter. The researcher classified each parameter into three classes (**Table 1**). Weighting class values are based on the minimum and maximum values of each parameter, based on the data that has been pre-processed before.

Table 1. Classification and weighting of class values for each parameter

Parameter	Class	Class Value
Average temperature (°C)	Low	≤ 25
	Medium	$> 25 - \leq 26.5$
	High	> 26.5
Average humidity (%)	Fairy Damp	≤ 84
	Damp	$> 84 - \leq 91$
	Very Damp	> 91
Sun exposure duration (hour)	Briefly	< 3.3
	Medium	$\geq 3.3 - < 6.6$
	Long	≥ 6.6
Average wind speed (m/s)	Slow	< 1
	Medium	$\geq 1 - < 2$
	Fast	≥ 2

While for the classification of rainfall, researchers are based on the standard categories of rainfall from BMKG, namely, (1) low categories ranging from 1-5 mm/hour; or 5-20 mm/day, (2) medium categories ranging from 5-10 mm/hour; or 20-50 mm/day, (3) heavy categories ranging from 10-20 mm/hour; or 50-100 mm/day, and (4) very heavy categories > 20 mm/hour; or > 100 mm/day. However, in this study, we divide rainfall into only three categories or classes, because based on the data that we have got, there is no value above 100 mm/day.

Table 2. Rainfall Classification

Class	Class Value (mm/day)
Low	≤ 20
Medium	$> 20 - \leq 50$
High	> 50

3.3. Data Analysis

Researchers use the Naïve Bayes method to predict rainfall, and Information Gain to determine the most influential parameters in predicting rainfall.

3.3.1. Naïve Bayes

The method invented in the 18th century by Thomas Bayes used the Bayes theorem. The theorem is used based on conditional probabilities based on the past for future probabilities [10]. The Naïve Bayes equation is shown in formula (1).

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

with, $P(H|X)$ = the probability that the hypothesis H is true for condition X; $P(X|H)$ = the probability that proof X is true for hypothesis H; $P(H)$ = the previous probability; $P(X)$ = the observed probability; X = proof (data sample); and H = hypothesis.

3.3.2. Information Gain

To find the effectiveness of each parameter, researchers used information gain as a method in this analysis to determine which parameters are most influential for high rainfall [10]. Information gain for parameter A is shown in formula (2).

$$Gain(S,A) = Entropy(S) - \sum_{v \in values(A)} \frac{|S_v|}{|S|} Entropy(S_v) \quad (2)$$

with, A = parameter; V = the declared value for parameter A; Values(A) = a set of values of A; $|S_v|$ = the number of samples for value v; $|S|$ = the sum of all data samples; and Entropy = entropy for samples that have a value of v

For entropy itself has its own calculations. Entropy is formulated in equation (3).

$$Entropy(S) = \sum_i^c - p_i \log_2 p_i \quad (3)$$

with, c = the number of values in the target parameter; and p_i = the ratio between the number of samples in class i.

4. Result

4.1. Predicting Rainfall Intensity with Naïve Bayes

There are two types of data used as test material, such as data-training and data-testing. Data-training is the data result in the pre-processing step with 329 data. Meanwhile, data-testing is a combination of class of each parameter, that is not contained in the data-training. This research used five parameter class combinations for data-testing. The result of data-training testing can be seen in **Table 3**.

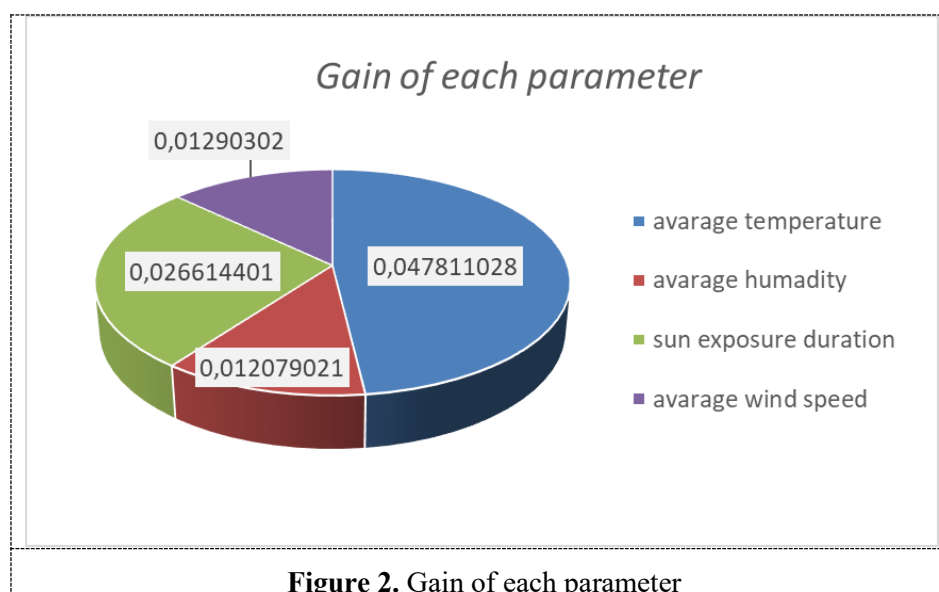
Table 3. Data-training testing to predict rainfall intensity

No	Parameters				Entropy of Rainfall Intensity		
	Average temperature	Average humidity	Sun exposure duration	Average wind speed	Low	Medium	High
1	High	Very Damp	Long	Slow	0.00151	0.00018	0.00009
2	Low	Damp	Long	Slow	0.00156	0.00073	0.00123
3	Medium	Fairy Damp	Briefly	Slow	0.01139	0.00196	0.00030
4	High	Very Damp	Briefly	Slow	0.00582	0.00095	0.00010
5	Low	Damp	Briefly	Medium	0.00816	0.00590	0.00267

Prediction result in Table 3 shows that, (1) the highest entropy score of rainfall intensity is in low class which is mean that for condition number 1 the rainfall intensity is low; (2) the highest entropy score of rainfall intensity is in low class which is mean that for condition number 2 the rainfall intensity is low; (3) the highest entropy score of rainfall intensity is in low class which is mean that for condition number 3 the rainfall intensity is low; (4) the highest entropy score of rainfall intensity is in low class which is mean that for condition number 4 the rainfall intensity is low; and (5) the highest entropy score of rainfall intensity is in low class which is mean that for condition number 5 the rainfall intensity is low.

4.2. Effectiveness of parameters with Information Gain

In this section, the results of the calculation of the most influential parameters will be displayed using the information gain method. The first step is to search for the Entropy (S) value and look for the entropy value of each parameter. The next step is to find the gain value of each parameter. The following results of the calculation of each gain value will be displayed using a pie chart in **Figure 2**.

**Figure 2.** Gain of each parameter

Based on the formula in Chapter 3 (Material and Method), the value of Entropy(S) is 1.06393. Then, this value is used to calculate information gain. Based on calculation using Information Gain, the most influential parameter in predicting rainfall is the average temperature with the amount of gain value is

0,047811028. The second one is sun exposure duration with 0,026614401, followed by average humidity and average wind speed with 0,01290302 and 0,012079021.

5. Conclusion

Based on the results of this study, it can be concluded that the Naïve Bayes method can be used to predict rainfall intensity in Sleman Regency. However, keep in mind, the standard class value for each parameter is different in each region. Also, based on the four parameters analyzed using Information Gain, it can be concluded that the most influential parameter on rainfall intensity is the average temperature with an entropy value of 0.047811028.

6. References

- [1] T. Kjellstrom, "Impact of Climate Conditions on Occupational Health and Related Economic Losses: A New Feature of Global and Urban Health in the Context of Climate Change," *Asia-Pacific J. Public Heal.*, vol. 28, pp. 28S-37S, 2014.
- [2] T. Weir, L. Dovey, and D. Orcherton, "Social and cultural issues raised by climate change in Pacific Island countries: an overview," *Reg. Environ. Chang.*, vol. 17, no. 4, pp. 1017–1028, 2017.
- [3] R. Chadwick, P. Good, G. Martin, and D. P. Rowell, "Large rainfall changes consistently projected over substantial areas of tropical land," *Nat. Clim. Chang.*, vol. 6, no. 2, pp. 177–181, 2016.
- [4] M. Ravanshadnia, M. Ghanbari, and M. H. Fardani, "Planning Crisis Management of Water Resources in Tehran in the Next Twenty Five Years," *J. Appl. Environ. Biol. Sci.*, vol. 5, no. 10, pp. 50–59, 2015.
- [5] T. Devasia, T. P. Vinushree, and V. Hegde, "Prediction of students performance using Educational Data Mining," *Proc. 2016 Int. Conf. Data Min. Adv. Comput. SAPIENCE 2016*, pp. 91–95, 2016.
- [6] S. Hussain, N. A. Dahan, F. M. Ba-Alwib, and N. Ribata, "Educational data mining and analysis of students' academic performance using WEKA," *Indones. J. Electr. Eng. Comput. Sci.*, vol. 9, no. 2, pp. 447–459, 2018.
- [7] A. Mueen, B. Zafar, and U. Manzoor, "Modeling and Predicting Students' Academic Performance Using Data Mining Techniques," *Int. J. Mod. Educ. Comput. Sci.*, vol. 8, no. 11, pp. 36–42, 2016.
- [8] C. Anuradha and T. Velmurugan, "Feature Selection Techniques To Analyse Student Academic Performance Using Naïve Bayes Classifier," *3rd Int. Conf. Small Mediu. Bus.*, no. February, pp. 345–350, 2016.
- [9] M. Muthmainnah, M. Ashar, I. M. Wirawan, and T. Widiyaningtyas, "Time Series Forecast for Rainfall Intensity in Malang City with Naive Bayes Methodology," *3rd Int. Conf. Sustain. Inf. Eng. Technol. SIET 2018 - Proc.*, pp. 137–141, 2019.
- [10] Suyatno, *DATA MINING UNTUK KLASIFIKASI DAN KLASTERISASI DATA*. Penerbit INFORMATIKA, 2017.