KHATULISTIWA SMART : JOURNAL OF ARTIFICIAL INTELLIGENCE
e-ISSN : 3089-9710
p-ISSN : XXXX-XXXX

Khatulistiwa Smart : Journal of Artificial Intelligence
Vol. 2 No. 1 (2025) : 39-47
Available online at : https://journal.literasikhatulistiwa.org/index.php/smart

Classification of Forest Fire-Prone Areas Using the K-Nearest Neighbor Algorithm: A Case Study of Baluran National Park

M Noer Fadli Hidayat [1]
[1] Universitas Negri Malang, Indonesia
m.noer.2305349@students.um.ac.id [1]

Doi

Received: 01 juni 2025 *Accepted: 16 juni 2025* *Published: 30 juni 2025*

Abstract Forest and land fires are one of the most recurring and destructive natural disasters in Indonesia, particularly during the dry season when rainfall is significantly low. Among the most affected areas in East Java is Baluran National Park, a region highly vulnerable due to its dominant savanna ecosystem. In response to the urgent need for effective forest fire risk prediction, this study aims to classify fire-prone areas using weather-related data and machine learning techniques. The research focuses on the application of the K-Nearest Neighbor (K-NN) algorithm to predict fire risk levels based on meteorological parameters. The dataset used in this study was obtained from Visualcrossing.com and consists of 211 weather records with 32 explanatory variables, such as maximum and minimum temperature, wind speed, sea-level pressure, solar radiation, and solar energy, along with one target variable representing fire risk level (categorized as High, Medium, or Low). The research method involves several stages: data preprocessing (handling missing values and converting nominal data into numeric), transformation (splitting into training and testing sets using the Percentage Split technique), and classification using K-NN implemented on Google Colab. The K-NN algorithm was configured with K = 3, using Euclidean Distance as the distance metric. The classification process produced a high accuracy rate of 98%, indicating the robustness and effectiveness of K-NN in classifying forest fire risks based on weather data. This model was further validated by comparing predicted outputs against actual values in the testing dataset, showing high consistency. The results suggest that the K-NN algorithm is highly applicable for environmental classification problems and can support decision-making systems in early warning and disaster mitigation efforts. This study contributes to the growing field of data-driven disaster risk management and highlights the

potential of machine learning in enhancing environmental monitoring systems.

Keywords *Forest Fires; Baluran; K-Nearest Neighbor; Classification; East Java*

1. INTRODUCTION

Forest and land fires are among the most severe and recurring natural disasters in Indonesia. This phenomenon occurs across almost all regions of the country and typically intensifies during the dry season due to extremely low rainfall. Such disasters result in substantial losses, not only reducing forest and land cover and threatening biodiversity, but also endangering the health and safety of nearby communities through exposure to dense smoke, which can lead to respiratory problems and even life-threatening conditions (Mahdi, 2022; Pratiwi et al., 2021). According to the Ministry of Environment and Forestry (KLHK), the total area of forest and land fires in Indonesia in 2021 reached 358,867 hectares, an increase of 20.85% compared to the 296,942 hectares recorded in 2020. East Java contributed the fifth largest burned area at 15,458 hectares, following Papua and three other provinces. The trend of forest fires in Indonesia fluctuates from year to year, with 2019 being the worst year in the 2016–2021 period, recording 1,649,258 hectares burned—an increase of 311% compared to the previous year (Mahdi, 2022; Primajaya, Sari, & Khusaeri, 2020).

One of the regions in East Java frequently affected by forest fires is Baluran National Park, located in Banyuputih, Situbondo Regency. Often dubbed the "African Savanna of Java," this park consists predominantly of savanna, in addition to forest, mountains, and coastal areas. These characteristics, along with dry vegetation and climate variability, make it particularly susceptible to fires. In 2020, Baluran experienced two separate fire incidents that destroyed approximately 5.2 hectares of forest area (Detiknews, as cited in Mahdi, 2022).

Weather conditions play a significant role in influencing the likelihood of forest fires. Critical variables include maximum and minimum temperature, wind speed, surface pressure, solar radiation, and solar energy (Noviansyah, Rismawan, & Midyanti, 2018). Identifying these environmental factors and their patterns can help in predicting and minimizing fire risks in vulnerable areas. One effective way to achieve this is through the application of data mining, a process used to convert large volumes of raw data into meaningful insights (Twin, 2021). This is especially useful when analyzing data from Automatic Weather Stations (AWS), which provide continuous environmental monitoring that can be processed to identify potential fire zones (Noviansyah et al., 2018).

A commonly used method in data mining for classification tasks is the K-Nearest Neighbor (K-NN) algorithm. K-NN is a supervised learning approach where new data instances are classified based on the majority class among their K nearest neighbors (Rahayu, Prianto, & Novia, 2021). This technique is popular due to its simplicity and high accuracy in real-world applications, particularly when the relationship between variables is non-linear or complex. Studies have shown the effectiveness of K-NN in classifying forest fire risk zones based on environmental variables. For instance, in Kubu Raya and West Kalimantan, researchers successfully implemented K-NN to classify fire risk levels using meteorological data, achieving significant accuracy (Rudiyan, Dzulkifli, & Munazar, 2022; Karo et al., 2022).

Comparative studies have also demonstrated that K-NN performs favorably when compared with other algorithms such as Naïve Bayes, especially in environmental and disaster-related prediction tasks (Pratiwi et al., 2021; Rahayu et al., 2021).

Furthermore, other research has highlighted how classification algorithms like C4.5, K-Means, and Naïve Bayes can be utilized to improve forecasting and prioritization efforts in various domains, including health services and forest management (Primajaya et al., 2020; W. I. Rahayu et al., 2021). The growing integration of machine learning techniques in environmental monitoring reflects a broader trend toward the digitization of risk management and early warning systems (A. Twin, 2021). In line with this trend, this study aims to classify fire-prone areas based on weather data—such as temperature, wind speed, and solar radiation—using the K-Nearest Neighbor algorithm. This research contributes to disaster risk mitigation efforts, particularly in fire-vulnerable regions like East Java, by offering a predictive model that can assist stakeholders in early intervention and resource allocation.

2. RESEARCH METHODS

In general, this research is divided into two scopes: literature study and model development. The literature study aims to strengthen the research problem and serve as the theoretical foundation of the study. The scope of the literature review includes journals, articles, and books that discuss the application of the K-Nearest Neighbor algorithm, the influence of distance measurement methods in K-Nearest Neighbor, and classification models for forest and land fires. The method consists of several stages, including selection, preprocessing, transformation, data mining, and evaluation. These stages are part of the Knowledge Discovery in Databases (KDD) process, as illustrated in Figure 1.

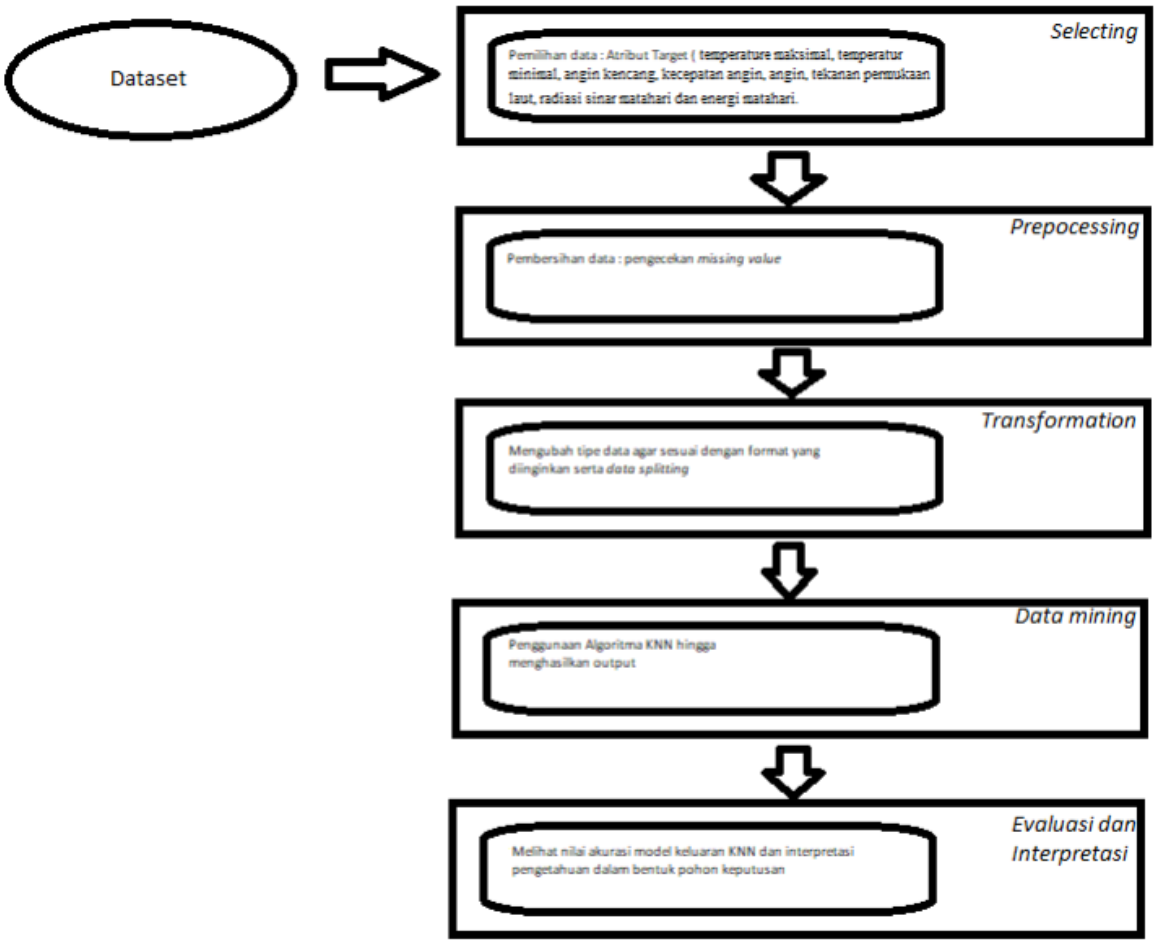


Figure 1. Stages of Knowladge Discovery Methodology in Database

2.1. Selecting

There are 32 explanatory attributes and 1 target attribute, referred to as independent and dependent variables, respectively. The explanatory variables collected include maximum temperature, minimum temperature, strong wind, wind speed, wind direction, sea-level pressure, solar radiation, and solar energy. The target variable is labeled as "Icon." Weather condition data were obtained from visualcrossing.com.

2.2. Preprocessing

In this stage, the collected data undergo a missing value check. If any missing values are found, they are handled by removing the corresponding data entries.

2.3. Transformation

This stage involves converting data types from nominal or object types into numerical formats, such as integers or floats. Additionally, the dataset is split into two subsets: training data and testing data. The Percentage Split technique is used for this data division.

2.4. Data Mining

Google Colab is used to process the data using the K-Nearest Neighbor algorithm. Google Colab provides various libraries that can be utilized to implement the K-Nearest Neighbor algorithm effectively.

2.5. Evaluation and Interpretation

Evaluation is performed by inputting the testing data into the model generated by the K-Nearest Neighbor algorithm. The main evaluation indicator is the model's accuracy. The higher the accuracy value, the better the predictive quality of the model. To calculate the distance between two points in the KNN algorithm, the Euclidean Distance formula is used.

$$dis\ (x_1, x_2) \sqrt{\sum_{i=0}^{n} (x_{1i} - x_{2i})^2}$$

Formula 1 Dimensional Space

$$dis = \sqrt{\sum_{i=0}^{n} (x_{1i} - x_{2i})^2 + (x_{1i} - x_{2i})^2 + \cdots}$$

Multi Dimensional Space Formula

3. RESULTS AND DISCUSSION

3.1. Data

The dataset was obtained from the website https://www.visualcrossing.com. The collected data is in raw form and must undergo preprocessing before it can be used for prediction purposes. The dataset used in this study is shown in the following figure:

	name	datetime	tempmax	tempmin	temp	feelslikemax	feelslikemin	feelslike	dew	humidity	precip	precipprob	precipcover	preciptype	snow	snowdepth	windgust	windspeed	winddir	sealevelpress	cloudcover	visibility	solarradiation	solarenergy	uvindex	severerisk	sunrise	sunset	moonphase	conditions	description	icon	stations
0	7.8329165,114.38	2022-01-01	29,2	23,6	25,3	34,2	23,6	25,8	23,1	87,8	37	100	12,5	rain				16,6	206,4	1008,7	87,7	7,3	104,9	9	4		2022-01-01	2022-01-01	0,99	Rain, Partia	Partly cloud	rain	96987099999,remote
1	7.8329165,114.38	2022-01-02	29,2	23,5	25	32,8	23,5	25,3	22,8	88,2	3,2	100	16,67	rain				10,8	234,3	1008,7	88,6	7,3	111,1	9,6	4		2022-01-02	2022-01-02	1	Rain, Partia	Partly cloud	rain	96987099999,remote
2	7.8329165,114.38	2022-01-03	30,9	23	25,2	34,5	23	25,7	22	83,5	9,5	100	66,67	rain				11,5	211	1008,6	84,2	7,6	156,3	13,4	5		2022-01-03	2022-01-03	0	Rain, Partia	Partly cloud	rain	96987099999,remote
3	7.8329165,114.38	2022-01-04	30,6	22,7	25,4	34,6	22,7	26,1	21,9	81,5	1,9	100	66,67	rain				13	210,5	1009,2	86,8	8,3	163,4	14,1	6		2022-01-04	2022-01-04	0,02	Rain, Partia	Partly cloud	rain	96987099999,remote
4	7.8329165,114.38	2022-01-05	29,4	22,8	25,2	33,1	22,8	25,7	21,8	82	7,2	100	8,33	rain				14,8	222,5	1009,5	79	7,5	157,7	13,7	5		2022-01-05	2022-01-05	0,05	Rain, Partia	Partly cloud	rain	96987099999,remote
5	7.8329165,114.38	2022-01-06	30,4	22,3	25,6	33,6	22,3	26,3	21,1	77,1	0,8	100	4,17	rain				13	232,4	1009,7	70,5	7,4	314,7	27,3	10		2022-01-06	2022-01-06	0,09	Rain, Partia	Partly cloud	rain	96987099999,remote
6	7.8329165,114.38	2022-01-07	30,2	22,3	24,8	34,7	22,3	25,1	21,4	81,9	4	100	8,33	rain				9,4	233	1010	83,1	7,4	244,3	21,1	8		2022-01-07	2022-01-07	0,14	Rain, Partia	Partly cloud	rain	96987099999,remote
7	7.8329165,114.38	2022-01-08	30,7	22,7	25,5	34,7	22,7	26,3	21,8	81,2	4,4	100	66,67	rain				9,4	165	1009,2	79,7	7,6	201,5	17,6	7		2022-01-08	2022-01-08	0,19	Rain, Partia	Partly cloud	rain	96987099999,remote
8	7.8329165,114.38	2022-01-09	31,9	23,1	25,3	37	23,1	26,1	22,5	85,1	85	100	12,5	rain				22,3	250,9	1008,4	82	7,3	180,2	15,8	6		2022-01-09	2022-01-09	0,24	Rain, Partia	Partly cloud	rain	96987099999,remote
9	7.8329165,114.38	2022-01-10	29,4	23,2	25,6	34,5	23,2	26,5	22,8	85,1	6	100	8,33	rain	0	0	3,6	18,4	235,9	1008,6	81,4	12,4	247,9	21,4	9		10-2022-01-10	2022-01-10	0,28	Rain, Partia	Partly cloud	rain	96987099999,remote
10	7.8329165,114.38	2022-01-11	30,2	22,5	25,4	36,3	22,5	26,2	22,8	85,8	8	100	8,33	rain	0	0	33,8	18,4	225,8	1010	93	15	127,9	11,1	6		30-2022-01-1	2022-01-11	0,33	Rain, Overc	Cloudy skies	rain	96987099999,remote
11	7.8329165,114.38	2022-01-12	30,6	22,8	25,6	35	22,8	26,7	22,6	84,3	0,2	100	4,17	rain	0	0	38,9	18,4	249,1	1010	87	16,7	237,6	20,5	8		30-2022-01-1	2022-01-12	0,38	Rain, Partia	Partly cloud	rain	96987099999,remote
12	7.8329165,114.38	2022-01-13	31,7	22,9	26,1	36,5	22,9	27,4	22,7	82,5	10,6	100	8,33	rain	0	0	24,1	18,4	238,8	1009,8	95	18,1	260,2	22,4	8		30-2022-01-1	2022-01-13	0,41	Rain, Overc	Cloudy skies	rain	96987099999,remote
13	7.8329165,114.38	2022-01-14	31,6	23,2	26,5	37,4	23,2	28,2	22,9	81,1	5,6	100	33,33	rain	0	0	23,8	14,4	237,4	1008,5	85,7	16,2	265,8	23,1	8		30-2022-01-1	2022-01-14	0,45	Rain, Partia	Partly cloud	rain	96987099999,remote
14	7.8329165,114.38	2022-01-15	30,9	23,6	25,7	36,5	23,6	26,5	23,2	86,5	28	100	12,5	rain	0	0	21,2	15,1	254,2	1008,6	94,2	18,2	135,2	11,9	5		30-2022-01-1	2022-01-15	0,47	Rain, Overc	Cloudy skies	rain	96987099999,remote
15	7.8329165,114.38	2022-01-16	31,4	22,9	26,7	37,4	22,9	28,5	23,1	81,7	3,6	100	37,5	rain	0	0	22,3	20,5	259,5	1008,5	92,7	18,3	256,6	22,4	7		60-2022-01-1	2022-01-16	0,49	Rain, Overc	Cloudy skies	rain	96987099999,remote
16	7.8329165,114.38	2022-01-17	32,7	24,2	26,2	37,8	24,2	27,3	23	83,1	7,8	100	12,5	rain	0	0	18	16,6	255	1008,3	95,5	17,8	141,4	12,1	5		30-2022-01-1	2022-01-17	0,5	Rain, Overc	Cloudy skies	rain	96987099999,remote
17	7.8329165,114.38	2022-01-18	28,7	23	24,3	32,5	23	24,7	22,6	90,2	41,4	100	8,33	rain	0	0	16,9	11,2	236,8	1008,5	89,8	15,8	52,8	4,4	3		10-2022-01-1	2022-01-18	0,5	Rain, Partia	Partly cloud	rain	96987099999,remote
18	7.8329165,114.38	2022-01-19	30,1	22,9	25,1	34,3	22,9	25,8	22,7	86,9	37	100	16,67	rain	0	0	15,5	9,4	248,8	1007,8	87,6	13,7	239,9	20,7	7		30-2022-01-1	2022-01-19	0,52	Rain, Partia	Partly cloud	rain	96987099999,remote
19	7.8329165,114.38	2022-01-20	30,2	22,8	25	34	22,8	25,7	22,6	86,9	0,6	100	4,17	rain	0	0	18	11,2	252,8	1008,6	95,6	11,8	112,1	10,2	5		10-2022-01-2	2022-01-20	0,53	Rain, Overc	Cloudy skies	rain	96987099999,remote
20	7.8329165,114.38	2022-01-21	30,4	23	25,7	34,5	23	26,6	22,1	81,1	0,2	100	4,17	rain	0	0	22,3	14,8	231,6	1008,7	95,3	17,3	192,4	16,8	6		30-2022-01-2	2022-01-21	0,56	Rain, Overc	Cloudy skies	rain	96987099999,remote
21	7.8329165,114.38	2022-01-22	30,7	22,9	26,2	36,4	22,9	27,6	22,5	80,4	2,4	100	16,67	rain	0	0	17,3	18,4	222,4	1008,1	89,6	17,8	259,5	22,5	8		60-2022-01-2	2022-01-22	0,6	Rain, Partia	Partly cloud	rain	96987099999,remote
22	7.8329165,114.38	2022-01-23	33,7	23,6	27	37,9	23,6	28,7	22,5	77,5	0,7	100	4,17	rain	0	0	16,6	20,5	255,4	1007,3	56,7	18,3	310,2	26,9	9		30-2022-01-2	2022-01-23	0,65	Rain, Partia	Partly cloud	rain	96987099999,remote
23	7.8329165,114.38	2022-01-24	32,9	23,4	27,2	38,5	23,4	29,1	22,4	75,8	1,1	100	16,67	rain	0	0	18	18,4	234	1005,8	42,8	19	308,3	26,8	9		30-2022-01-2	2022-01-24	0,7	Rain, Partia	Partly cloud	rain	96987099999,remote
24	7.8329165,114.38	2022-01-25	32,7	23,9	27,3	38,1	23,9	29,2	22,6	76,4	0,1	100	4,17	rain	0	0	28,4	20,5	246,6	1005,7	56,7	17,4	323,7	28,1	9		30-2022-01-2	2022-01-25	0,75	Rain, Partia	Partly cloud	rain	96987099999,remote
25	7.8329165,114.38	2022-01-26	33,1	23,9	26,9	38	23,9	29	23,4	82,1	16,2	100	37,5	rain	0	0	28,4	20,9	248,5	1006,4	96,5	17,5	265,3	22,8	8		30-2022-01-2	2022-01-26	0,81	Rain, Overc	Cloudy skies	rain	96987099999,remote
26	7.8329165,114.38	2022-01-27	30,7	24,1	26,4	36,4	24,1	28,2	23,4	83,6	36	100	4,17	rain	0	0	19,8	13,7	245,7	1007,7	94,2	14,9	222,4	19,4	7		60-2022-01-2	2022-01-27	0,86	Rain, Overc	Cloudy skies	rain	96987099999,remote
27	7.8329165,114.38	2022-01-28	30,7	23,7	25,7	36,7	23,7	26,7	22,7	83,8	22,2	100	12,5	rain	0	0	14	13	197,3	1009	91,9	17,6	194,6	17	7		30-2022-01-2	2022-01-28	0,91	Rain, Overc	Cloudy skies	rain	96987099999,remote
28	7.8329165,114.38	2022-01-29	31,7	23,4	26,3	37,3	23,4	27,7	22,8	81,5	1,1	100	20,83	rain	0	0	10,8	13	214,3	1009,1	98,1	18,5	195,2	16,9	7		60-2022-01-2	2022-01-29	0,95	Rain, Overc	Cloudy skies	rain	96987099999,remote
29	7.8329165,114.38	2022-01-30	31,2	24	26,1	37,5	24	27,4	23,4	85,4	0,8	100	4,17	rain	0	0	9	18,4	202,2	1008,8	94,9	18	183,1	15,9	6		60-2022-01-3	2022-01-30	0,98	Rain, Overc	Cloudy skies	rain	96987099999,remote
30	7.8329165,114.38	2022-01-31	30,9	23,6	26,4	37,1	23,6	27,9	22,9	81,9	4	100	16,67	rain	0	0	14	13	225,9	1008	94,9	18,3	239,4	20,9	7		60-2022-01-3	2022-01-31	1	Rain, Overc	Cloudy skies	rain	96987099999,remote
31	7.8329165,114.38	2022-02-01	30,7	23,4	26,4	34,4	23,4	27,7	22,6	80,5	2	100	4,17	rain	0	0	19,1	14,8	248,1	1008	78,5	18	295,6	25,8	9		30-2022-02-0	2022-02-01	0	Rain, Partia	Partly cloud	rain	96987099999,remote
32	7.8329165,114.38	2022-02-02	31,2	23,5	27	37,8	23,5	29	22,5	77,3	3,6	100	29,17	rain	0	0	18,7	20,5	246,3	1008,5	67,6	17,2	313,1	26,9	9		30-2022-02-0	2022-02-02	0,01	Rain, Partia	Partly cloud	rain	96987099999,remote
33	7.8329165,114.38	2022-02-03	32,7	23	26,4	37,3	23	27,8	22,4	79,8	5	100	8,33	rain	0	0	10,4	13	194,4	1008,6	61,6	16,8	293,2	25,3	9		30-2022-02-0	2022-02-03	0,03	Rain, Partia	Partly cloud	rain	96987099999,remote
34	7.8329165,114.38	2022-02-04	31,7	22,7	26,2	36	22,7	27,6	22,2	79,6	4,1	100	25	rain	0	0	13,3	12,6	153,9	1007,2	62,9	16,4	241,7	21	8		30-2022-02-0	2022-02-04	0,07	Rain, Partia	Partly cloud	rain	96987099999,remote
35	7.8329165,114.38	2022-02-05	29,9	22,7	25,8	34,3	22,7	27	22,7	83,4	10	100	8,33	rain	0	0	15,8	13	256,9	1007	74,2	16,7	257,1	22,4	7		30-2022-02-0	2022-02-05	0,11	Rain, Partia	Partly cloud	rain	96987099999,remote
36	7.8329165,114.38	2022-02-06	30,7	23,4	26,2	36,8	23,4	27,8	22,9	82,7	0,2	100	4,17	rain	0	0	15,8	22,3	235,1	1007,5	48,8	17,6	288,4	24,9	8		30-2022-02-0	2022-02-06	0,15	Rain, Partia	Partly cloud	rain	96987099999,remote
37	7.8329165,114.38	2022-02-07	32,7	23,5	26,6	39,9	23,5	28,2	22,9	80,8	0,4	100	8,33	rain	0	0	37,1	22,3	257,2	1007,9	78,3	19	276,1	23,9	8		30-2022-02-0	2022-02-07	0,2	Rain, Partia	Partly cloud	rain	96987099999,remote
38	7.8329165,114.38	2022-02-08	32,2	23,8	26,7	38,3	23,8	28,4	23,1	81,6	5	100	4,17	rain	0	0	24,5	22,3	241,2	1008,8	86,3	14,5	278,1	24,3	8		60-2022-02-0	2022-02-08	0,25	Rain, Partia	Partly cloud	rain	96987099999,remote
39	7.8329165,114.38	2022-02-09	31,4	22,9	25,9	36,6	22,9	27	22,9	84,3	11,2	100	12,5	rain	0	0	22,3	14,8	184,2	1008,1	76,7	17,3	179,8	15,6	7		30-2022-02-0	2022-02-09	0,29	Rain, Partia	Partly cloud	rain	96987099999,remote
40	7.8329165,114.38	2022-02-10	30,2	22,8	25,6	35,4	22,8	26,6	22,4	83,2	7	100	4,17	rain	0	0	15,1	22,3	272,6	1007,4	93,5	17,4	254,1	21,8	7		30-2022-02-1	2022-02-10	0,34	Rain, Overc	Cloudy skies	rain	96987099999,remote
41	7.8329165,114.38	2022-02-11	29,4	22,8	25,8	33,3	22,8	26,9	22,5	82,6	4,3	100	25	rain	0	0	11,9	11,2	184,2	1007,4	95,4	17,5	268,1	23,2	8		30-2022-02-1	2022-02-11	0,38	Rain, Overc	Cloudy skies	rain	96987099999,remote
42	7.8329165,114.38	2022-02-12	29,9	23,5	25,2	34,8	23,5	25,9	23,2	88,8	7,5	100	45,83	rain	0	0	18,4	22,3	206,1	1006,7	95,6	12,1	76,4	7,1	3		30-2022-02-1	2022-02-12	0,42	Rain, Overc	Cloudy skies	rain	96987099999,remote
43	7.8329165,114.38	2022-02-13	31,7	23,1	25,7	36	23,1	26,8	23	85,5	6,5	100	37,5	rain	0	0	10,8	13	196,2	1006,9	93,3	14,6	148	13,5	4		30-2022-02-1	2022-02-13	0,45	Rain, Overc	Cloudy skies	rain	96987099999,remote
44	7.8329165,114.38	2022-02-14	28,7	23	25,4	32,7	23	26	22,9	86,5	14,6	100	12,5	rain	0	0	7,2	9,4	162,3	1008,3	86,3	14,1	233,7	20,2	8		30-2022-02-1	2022-02-14	0,48	Rain, Overc	Cloudy skies	rain	96987099999,remote
45	7.8329165,114.38	2022-02-15	31,2	22,8	26,2	36,4	22,8	28	23,1	83,7	2	100	8,33	rain	0	0	10,4	20,5	160,9	1007,9	60,4	17,2	304,4	26,3	8		30-2022-02-1	2022-02-15	0,49	Rain, Partia	Partly cloud	rain	96987099999,remote
46	7.8329165,114.38	2022-02-16	31,4	22,4	26,5	36,1	22,4	28,1	22,6	80	2,4	100	29,17	rain	0	0	18	22,3	154,7	1006,6	58,6	17,6	306,5	26,5	9		30-2022-02-1	2022-02-16	0,5	Rain, Partia	Partly cloud	rain	96987099999,remote
47	7.8329165,114.38	2022-02-17	29,9	23,5	26,4	36,1	23,5	27,9	22,7	80,6	3,8	100	16,67	rain	0	0	18	31,7	115,3	1006,8	63,6	18,5	291,1	25,1	8		30-2022-02-1	2022-02-17	0,51	Rain, Partia	Partly cloud	rain	96987099999,remote
48	7.8329165,114.38	2022-02-18	30,6	22,6	26,1	35,5	22,6	27,4	22,2	80	7,1	100	25	rain	0	0	28,4	29,5	159	1008,6	69,4	15,2	299,6	26,2	9		30-2022-02-1	2022-02-18	0,52	Rain, Partia	Partly cloud	rain	96987099999,remote
49	7.8329165,114.38	2022-02-19	32,6	22,4	26,4	37,9	22,4	27,7	21,6	75,7	1	100	20,83	rain	0	0	19,4	12,2	257,6	1008,6	78,4	19	267,9	23,3	8		30-2022-02-1	2022-02-19	0,54	Rain, Partia	Partly cloud	rain	96987099999,remote
50	7.8329165,114.38	2022-02-20	31,6	23,1	26,2	38	23,1	27,6	22,5	80,8	5,3	100	29,17	rain	0	0	14	14,8	244,2	1008,2	86,9	17,1	199,7	17,4	7		60-2022-02-2	2022-02-20	0,58	Rain, Partia	Partly cloud	rain	96987099999,remote
51	7.8329165,114.38	2022-02-21	29,9	22,9	25,7	33,9	22,9	26,5	22,1	81,1	7	100	4,17	rain	0	0	11,2	13	155,7	1008,2	92,5	17,5	232,4	20,2	7		60-2022-02-2	2022-02-21	0,62	Rain, Overc	Cloudy skies	rain	96987099999,remote
52	7.8329165,114.38	2022-02-22	33,1	23,2	26	38,2	23,2	27,2	21,6	77,5	261	100	4,17	rain	0	0	25,6	13	181,9	1008,5	85,4	16,7	249	21,5	8		30-2022-02-2	2022-02-22	0,67	Rain, Partia	Partly cloud	rain	96987099999,remote
53	7.8329165,114.38	2022-02-23	29,6	22,2	25,6	32,8	22,2	26,4	21,5	78,8	3	100	4,17	rain	0	0	11,5	27,7	139,2	1009,3	75,7	17	285,8	24,9	8		30-2022-02-2	2022-02-23	0,73	Rain, Partia	Partly cloud	rain	96987099999,remote
54	7.8329165,114.38	2022-02-24	30,4	22,6	25,5	34,8	22,6	26,6	21,7	80,1	1,6	100	8,33	rain	0	0	14,4	14,8	247,9	1009,7	78,9	17,7	304	26,2	8		30-2022-02-2	2022-02-24	0,79	Rain, Partia	Partly cloud	rain	96987099999,remote

Figure 2. Raw Data

3.2. System Design Process

The design process in this study is carried out based on the system workflow, which is illustrated in the following figure:

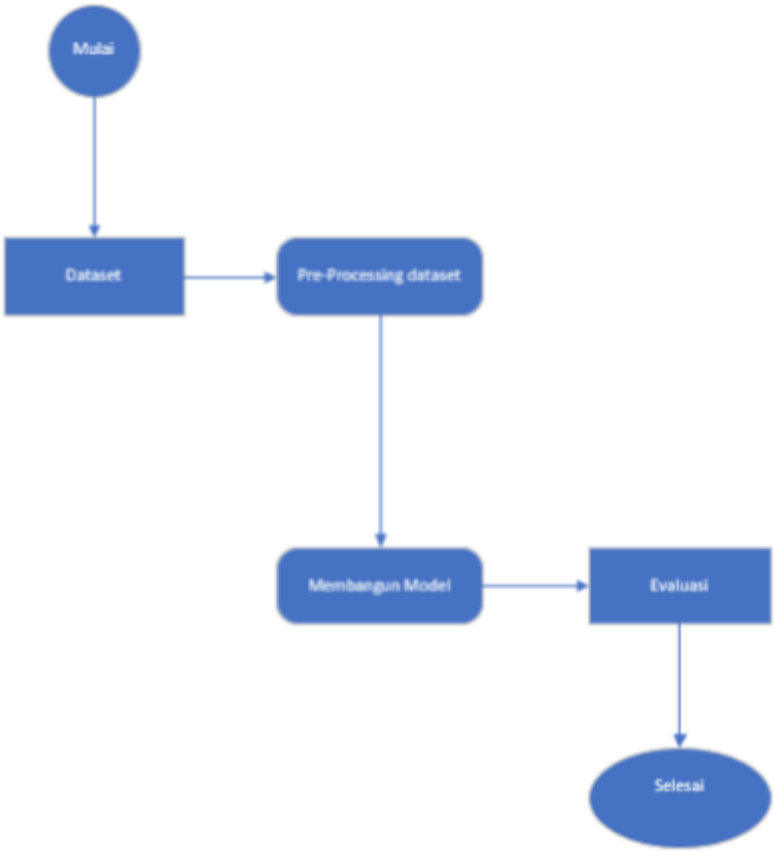


Figure 3. Process Flow Diagram

3.2.1. Data Input

The first step in the process is the input of tabulated weather data. The source code used in this process is shown in the following figure:

```
[1]
from sklearn.metrics import accuracy_score
from sklearn.model_selection import train_test_split

from scipy.stats import mode

[2] dataset = pd.read_csv('/content/drive/MyDrive/Colab Notebooks/bahan/dataset.csv')
dataset
```

```
dataset.head()
```

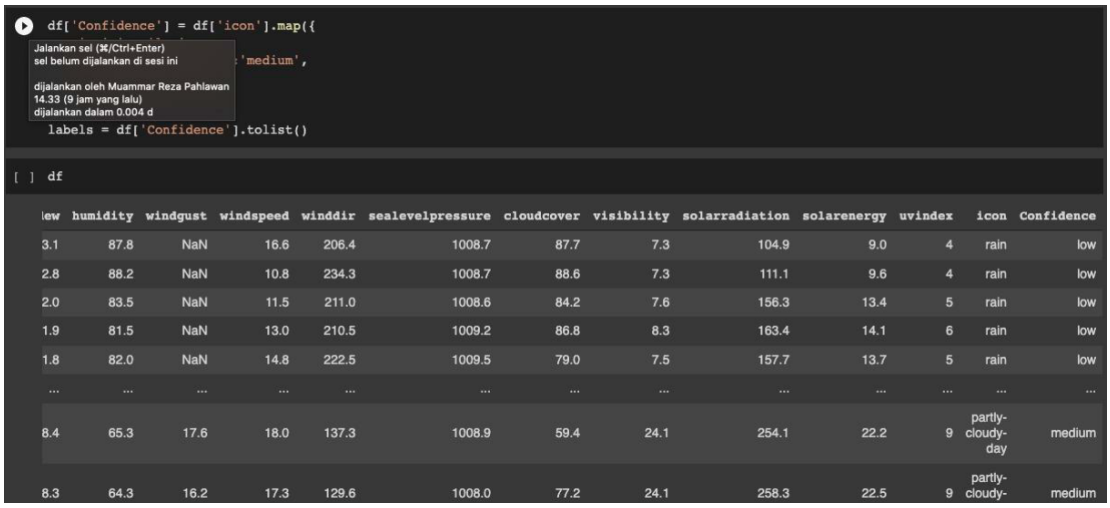
tempmin	temp	feelslikemax	feelslikemin	feelslike	dew	humidity	...	solarenergy	uvindex	severerisk	sunrise	sunset	moonphase
23.6	25.3	34.2	23.6	25.8	23.1	87.8	...	9.0	4	NaN	2022-01-01T05:08:49	2022-01-01T17:42:56	0.99
23.5	25.0	32.8	23.5	25.3	22.8	88.2	...	9.6	4	NaN	2022-01-02T05:09:21	2022-01-02T17:43:21	1.00
23.0	25.2	34.5	23.0	25.7	22.0	83.5	...	13.4	5	NaN	2022-01-03T05:09:52	2022-01-03T17:43:44	0.00

Figure 4. Data input process

3.2.2. S

The data preprocessing stage is carried out after the input of raw data to ensure it is suitable for use in training and testing phases. This step involves several

important activities that are critical for preparing the dataset for machine learning analysis. One of the primary tasks is converting nominal data into numerical formats, which allows the data to be processed computationally by algorithms that require numerical input. In addition, a class attribute is introduced to represent the classification category for forest fire occurrences. Several textual attributes—such as condition, situation, description, and icon—are also transformed into numeric values. This transformation ensures consistency and compatibility across all features in the dataset. These preprocessing efforts are fundamental to the success of the classification model, particularly because the K-Nearest Neighbor algorithm relies heavily on numerical distance calculations to make predictions. The entire process of data transformation is visually represented in the following figure.



```
df['Confidence'] = df['icon'].map({
    'medium',
labels = df['Confidence'].tolist()
```

```
[ ] df
```

lew	humidity	windgust	windspeed	winddir	sealevelpressure	cloudcover	visibility	solarradiation	solarenergy	uvindex	icon	Confidence
3.1	87.8	NaN	16.6	206.4	1008.7	87.7	7.3	104.9	9.0	4	rain	low
2.8	88.2	NaN	10.8	234.3	1008.7	88.6	7.3	111.1	9.6	4	rain	low
2.0	83.5	NaN	11.5	211.0	1008.6	84.2	7.6	156.3	13.4	5	rain	low
1.9	81.5	NaN	13.0	210.5	1009.2	86.8	8.3	163.4	14.1	6	rain	low
1.8	82.0	NaN	14.8	222.5	1009.5	79.0	7.5	157.7	13.7	5	rain	low
...	...	...	...	...	...	...	...	...	...	...	...	...
8.4	65.3	17.6	18.0	137.3	1008.9	59.4	24.1	254.1	22.2	9	partly-cloudy-day	medium
8.3	64.3	16.2	17.3	129.6	1008.0	77.2	24.1	258.3	22.5	9	partly-cloudy-	medium

Figure 5. Confidence Attribute Addition Data

windgust	windspeed	winddir	sealevelpressure	cloudcover	visibility	solarradiation	solarenergy	uvindex	icon	Confidence	Class
NaN	16.6	206.4	1008.7	87.7	7.3	104.9	9.0	4	rain	low	0
NaN	10.8	234.3	1008.7	88.6	7.3	111.1	9.6	4	rain	low	0
NaN	11.5	211.0	1008.6	84.2	7.6	156.3	13.4	5	rain	low	0
NaN	13.0	210.5	1009.2	86.8	8.3	163.4	14.1	6	rain	low	0
NaN	14.8	222.5	1009.5	79.0	7.5	157.7	13.7	5	rain	low	0
...	...	...	...	...	...	...	...	...	...	...	...
17.6	18.0	137.3	1008.9	59.4	24.1	254.1	22.2	9	partly-cloudy-day	medium	1
16.2	17.3	129.6	1008.0	77.2	24.1	258.3	22.5	9	partly-cloudy-day	medium	1
19.4	20.5	132.2	1006.9	66.6	24.1	252.9	22.2	9	partly-cloudy-day	medium	1
23.8	20.5	128.8	1007.0	76.0	24.1	235.8	20.1	8	partly-cloudy-day	medium	1
24.1	19.8	126.0	1008.0	67.3	24.1	260.0	22.8	9	partly-cloudy-day	medium	1

Figure 6. Class Addition

3.2.3. K-Nearest Neighbor Classification

In this stage, the classification process is carried out using the K-Nearest Neighbor (K-NN) algorithm. The prepared dataset, which has undergone preprocessing, is now ready to be processed by the classification model. The K-NN algorithm works by identifying the K closest data points (neighbors) to a given input instance based on a chosen distance metric commonly Euclidean distance and assigning the class label that is most common among those neighbors. This method enables the model to classify new or unseen data based on similarities to existing labeled data in the training set. The implementation of this classification process using the K-NN algorithm is shown in the following figure.

Figure 7. Flow Chart K-Nearest Neighbor

3.2.4. Classification Results

The classification process was conducted using a training dataset consisting of 211 records, with three target classes: High, Medium, and Low. The data was split using a 70:30 ratio for training and testing purposes. The model was evaluated using 30% of the total dataset as testing data, and the results demonstrated a high level of accuracy. With the value of K set to 3, the K-Nearest Neighbor algorithm achieved a classification accuracy of 98%, indicating the model's strong ability to correctly predict forest fire risk categories based on the input features.

3.2.5. Validation of Results

To validate the model, the predicted results were compared against the actual class values. This validation process ensures the reliability of the model by measuring its ability to correctly classify new data based on patterns learned during training. The high correspondence between predicted and actual results further supports the robustness and effectiveness of the classification model.

4. CONCLUSIONS

The classification process, as defined in the selected methodology, was implemented using the K-Nearest Neighbor (K-NN) algorithm, which serves as a proximity-based search

technique. This method classifies new data points based on the majority class among their nearest neighbors in the training dataset. For this study, a total of 211 weather data records were used, comprising various environmental attributes relevant to forest fire prediction. The dataset was divided using a 70:30 ratio, with 70% used for training the model and the remaining 30% for testing. The testing phase aimed to evaluate the model's ability to correctly classify unseen data based on patterns learned from the training phase. The value of K was set to 3, meaning that each prediction was made based on the three closest data points in the feature space. This configuration yielded a classification accuracy of 98%, demonstrating the model's strong performance in distinguishing between the three fire risk categories: High, Medium, and Low. The high level of accuracy indicates that the selected features—such as temperature, wind speed, solar radiation, and pressure—are highly relevant and contribute significantly to the classification outcomes. Additionally, the result suggests that the K-NN algorithm is well-suited for this type of environmental classification problem, especially when the dataset is well-preprocessed and properly structured. This accuracy score reflects the potential of K-NN in supporting early warning systems for forest fire risks and assisting decision-makers in disaster mitigation planning.

5. REFERENCES

Karo, I. M., Khosuri, A., Septory, J. S. I., & Supandi, D. P. (2022). Pengaruh metode pengukuran jarak pada algoritma k-NN untuk klasifikasi kebakaran hutan dan lahan. Jurnal Media Informatika Budidarma, 6(2), 1174–1182.

Mahdi, M. I. (2022, April 22). Luas kebakaran hutan dan lahan Indonesia meningkat pada 2021. Varia – Data Indonesia.

Noviansyah, M. R., Rismawan, T., & Midyanti, D. M. (2018). Penerapan data mining menggunakan metode K-Nearest Neighbor untuk klasifikasi indeks cuaca kebakaran berdasarkan data AWS (Automatic Weather Station) (Studi kasus: Kabupaten Kubu Raya). Jurnal Coding, Sistem Komputer Untan, 6(2), 48–56.

Pratiwi, T. A., Irsyad, M., Kurniawan, R., Agustian, S., & Negara, B. S. (2021). Klasifikasi kebakaran hutan dan lahan menggunakan algoritma Naïve Bayes di Kabupaten Pelalawan. CESS (Journal of Computer Engineering System and Science), 6(2), 139–148.

Primajaya, A., Sari, B. N., & Khusaeri, A. (2020). Prediksi potensi kebakaran hutan dengan algoritma klasifikasi C4.5 (Studi kasus Provinsi Kalimantan Barat). JEPIN (Jurnal Edukasi dan Penelitian Informatika), 6(2), 188–192.

Rahayu, W. I., Prianto, C., & Novia, E. A. (2021). Perbandingan algoritma K-Means dan Naïve Bayes untuk memprediksi prioritas pembayaran tagihan rumah sakit berdasarkan tingkat kepentingan pada PT. Pertamina (Persero). Jurnal Teknik Informatika, 14(1), 1–8.

Rudiyan, A., Dzulkifli, A. E., & Munazar, K. (2022). Klasifikasi kebakaran hutan menggunakan metode K-Nearest Neighbor: Studi kasus hutan Provinsi Kalimantan Barat. JTIM: Jurnal Teknologi Informasi dan Multimedia, 4(2), 195–202.

Twin, A. (2021, September 17). Business marketing essentials: Data mining. Investopedia.