

PREDIKSI CURAH HUJAN MENGGUNAKAN ALGORITMA NAÏVE BAYES DAN K-MEANS DI WILAYAH MALANG

Diterima Redaksi: 30 September 2023; Revisi Akhir: 1 November 2023; Diterbitkan Online: 30 November 2023

M Shochibul Burhan¹⁾, Mohamad Nur Cholish²⁾, Achmad Latifudin³⁾, Mahendra Widikara Tri Nugroho⁴⁾

^{1, 3, 4)} Aplikasi Komputer Fakultas Pertanian Peternakan, Universitas Muhammadiyah Malang

²⁾ Teknik Elektro, Sekolah Tinggi Teknologi Industri Turen

e-mail: shochibul@umm.ac.id¹⁾, cholish8918@gmail.com²⁾, pangeranindonesia999@gmail.com³⁾, mahendra@gmail.com³⁾

Abstrak: Banyak manfaat dari curah hujan untuk kehidupan manusia, antara lain untuk pertanian, penurunan suhu permukaan bumi, metabolisme tubuh dan lain sebagainya. Selain memiliki manfaat yang besar, curah hujan juga dapat membawa dampak negative seperti banjir, wabah penyakit dan bencana alam lainnya jika tidak ditangani dengan benar. Curah hujan akan mampu ditangani dengan baik jika mampu diprediksikan waktu turunnya, sehingga mitigasi secara dini mampu dilakukan. Dengan metode prediksi menggunakan Algoritme Naïve Bayes turunnya hujan akan mampu diprediksi dengan menghitung dari data curah hujan sebelumnya. Data latih dan data uji diambil dari BMKG Karangploso Malang, dari data tersebut akan di lakukan preprocessing terlebih dahulu dengan Algoritma K-Means untuk mengelompokan data latih menjadi tiga variable yaitu Rendah, Sedang, Tinggi. Hasil dari metode ini menghasilkan akurasi sebesar 80% dalam memprediksi curah hujan. Metode Naïve Bayes dan K-Means mampu memprediksi curah hujan secara akurat, dimana dengan nilai tersebut berasal dari hasil data uji wilayah Malang. Nilai yang salah terdapat 20% dimana tergolong kecil dan nilai tersebut cukup akurat.

Kata Kunci— *Prediksi Hujan, Naïve Bayes, K-Means*

Abstract: There are many benefits of rainfall for human life, including for agriculture, reducing the earth's surface temperature, body metabolism and so on. Apart from having great benefits, rainfall can also have negative impacts such as floods, disease outbreaks and other disasters if not handled properly. Rainfall will be able to be handled if it can be predicted when it will fall, so that early mitigation can be carried out. With the prediction method using the Naïve Bayes algorithm, rainfall can be predicted by calculating from previous rainfall data. The training data and test data were taken from BMKG Karangploso Malang. Preprocessing of this data will be carried out first with the K-Means algorithm to group the training data into three variables, Low, Medium and High. The results of this method produce an accuracy of 80% in predicting rainfall. The Naïve Bayes and K-Means methods are able to predict rainfall accurately, with these values coming from the results of test data for the Malang area. There are 20% incorrect values, which is relatively small and this value is quite accurate.

Keywords— *Rain Forecasting, Naïve Bayes, K-Means*

I. PENDAHULUAN

Curah hujan adalah jumlah air yang jatuh di tanah datar selama periode tertentu yang diukur dengan satuan tinggi (mm) [1]. Dengan kemajuan teknologi curah hujan dapat diperkirakan sedini mungkin agar meminimalisir dampak negatif yang ditimbulkan. Peramalan curah hujan sangat bermanfaat membantu masyarakat dalam kehidupan sehari-hari seperti operasional penerbangan, sistem pertanian, sistem penampungan air, kegiatan konstruksi dan transportasi [2]. Didalam penelitian ini digunakan Algoritma Naïve Bayes untuk meramalkan curah hujan pada suatu daerah, dengan menggunakan data latih berupa Temperatur, Kelembaban, Lama Penyinaran Matahari, Kecepatan Angin dan Curah Hujan selama satu bulan. Sebelum data digunakan data latih terlebih dahulu dilakukan *preprocessing* data menggunakan Algoritma K-Means untuk menentukan tingkatatan disetiap variable data dengan menentukan tingkatatan Rendah, Sedang, Tinggi.

Pada penelitian ini digunakan data dari hasil pengamatan stasiun Badan Meteorologi Klimatologi dan Geofisika (BMKG) Karangploso, selama periode tahun 2022. Data yang mempengaruhi terjadinya curah hujan yang berkaitan antar satu dengan lainnya yaitu Data Temperatur, Kelembaban, Lama Penyinaran Matahari, Kecepatan Angin dan Curah Hujan [4]. Dari data ini kemudian di hitung *probabilitasnya* menggunakan terhadap curah hujan menggunakan Algoritma *Naïve Bayes*.

Metode *Naïve Bayes* merupakan pengklasifikasian dengan metode probabilitas dan statistic yang ditemukan oleh ilmuwan asal inggris Thomas Bayes. *Naïve Bayes* adalah sebuah metode menghitung peluang dengan cara mencari nilai peluang (*probabilitas*) disetiap keanggotaan (*variable*) kelas tertentu. Probabilitas pada teori *Naïve Bayes* bersifat independent tidak dipengaruhi oleh kehadiran fitur lainnya. [5]. Sebagai contoh, buah dapat dianggap sebagai apel jika merah, bulat, dan sekitar 3 inci diameter. Semua properti ini secara independen berkontribusi pada probabilitas bahwa buah ini adalah apel dan itulah sebabnya dikenal sebagai 'naif'. *Naïve Bayes* mudah untuk membangun dan sangat berguna untuk set data yang sangat besar. Seiring dengan kesederhanaan, *Naive Bayes* dikenal dapat mengungguli metode klasifikasi yang sangat canggih sekalipun [5].

Beberapa penelitian yang menggunakan *Naïve Bayes* antara lain *Naïve Bayes* untuk memprediksi keadaan Genetik tanaman jagung yang dilakukan, Hubert dkk. Mampu menghasilkan akurasi 85% [5]. Kemudian Yudana dkk, tahun 2021 menggunakan *Naïve Bayes* untuk memprediksi kandungan nitrogen pada tanah, dan menghasilkan akurasi 85% [6]. *Naïve Bayes* juga mampu melakukan prediksi kepadatan lalu lintas, pada penelititn yang dilakukan Chen dkk. Mampu menghasilkan keakuratan hingga 93% [7].

Sebelum dilakukan training data data terlebih dahulu akan dilakukan preprocessing data untuk menentukan sebuah variable data berada pada kondisi Rendah, Sedang atau Tinggi. Dalam penelitian ini digunakan Algoritma K-Means, Salah satu algoritma clustering yang populer saat ini adalah K-Means. Algoritma K-Means terbukti cukup andal dan banyak digunakan dalam komunitas peneliti data mining dan berbagai macam aplikasi kecil hingga menengah karena kemudahan implementasinya. K-Means merupakan salah satu metode data clustering non hirarki yang digunakan untuk mempartisi data yang ada kedalam bentuk dua atau lebih cluster/kelompok. Metode K-Means mempartisi data kedalam cluster sehingga data yang memiliki karakteristik sama dikelompokkan kedalam satu cluster yang sama [8][9]. Dalam melakukan pengelompokan algoritma K-Means mengelompokan data ke dalam beberapa kluster k yang telah ditentukan, lalu menentukan titik pusat setiap kluster dan mengelompokan data sesuai dengan kemiripan pada masing-masing *kluster* [10].

II. TINJAUAN PUSTAKA

Tinjauan pustaka yang digunakan dalam penelitian ini yaitu teori *Naïve Bayes* dan *Probabilitas Variabel* yang digunakan untuk menghitung peluang dari data latih. Algoritma *K-means* digunakan untuk *preprocessing* data dan *Root Means Square Error* (RMSE) digunakan untuk menghitung tingkat akurasi dari prediksi yang dihasilkan.

Naïve Bayes pada penelitian ini digunakan sebagai perhitungan peluang (*probabilitas*), yaitu peluang hujan atau tidak hujan. *Naïve Bayes* memiliki kelebihan yaitu cepat dalam perhitungan dan memiliki akurasi yang tinggi [11]. *K-means* pada penelitian ini digunakan untuk menentukan data kedalam 3 bagian Rendah, Sedang, Tinggi. Selain itu berperan juga dalam *preprocessing data* agar data sebelum di proses oleh *Naïve Bayes* data tidak ada yang menyimpang dari angka standard atau data yang bersih [13]. Berdasarkan nilai titik tengah (*centroid*) yang telah dilakukan perhitungan. *K-Means* memiliki keunggulan pada *fleksibilitas* dapat diterapkan pada beberapa perhitungan [12]. Dan *RMSE* (*Root Means Square Error*) digunakan untuk mengevaluasi keakuratan prediksi yang dihasilkan dari *Naïve Bayes*. *RMSE* merupakan nilai rata-rata kuadrat dari jumlah kesalahan pada sebuah prediksi [14].

A. *Naïve Bayes*

Naïve Bayes merupakan metode probabilitas pengkalisifikasian sederhana berdasarkan teorima bayes, pengkalisifikasian dilakukan melalui data latih yang diperoleh dari masa lampau. Teori Bayes di formulasikan pada persamaan 1.

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

- $P(C|X)$ = Probabilitas hipotesisi C berdasarkan kondisi X
 C = Variabel Kelas (Kelas Spesifik)
 X = Variabel Kondisi
 $P(X|C)$ = Probabilitas X berdasarkan kondisi pada hipotesis C
 $P(C)$ = Probabilitas hipotesis C
 $P(X)$ = Probabilitas X

Pada persamaan 1 adalah rumus pada *Naïve bayes*, pada penelitian ini C adalah variabel dari Temperatur, Lama Penyinaran Matahari dan Kecepatan Angin. Dan X adalah kondisi Hujan atau Tidak hujan.

B. Probabilitas Variabel

Setiap variable dihitung probabilitasnya meliputi Probabilitas Temperatur $P(Temperatur)$, Probabilitas Kelembaban $P(Kelembaban)$, Probabilitas Lama Penyinaran Matahari $P(LamaPenyinaran)$, Kecepatan Angin $P(Kecepatan Angin)$. Seperti pada rumus berikut:

$$P(Temperatur|Hujan) = \frac{\sum(Temperatur|Hujan)}{\sum(Hujan)} \quad (2)$$

$$P(Temperatur|TidakHujan) = \frac{\sum(Temperatur|TidakHujan)}{\sum(TidakHujan)} \quad (3)$$

Kelembaban pada data latih diklasifikasikan menjadi 2 klasifikasi yaitu kelembaban yang statusnya hujan dan kelembaban yang statusnya tidak hujan.

$$P(LamaPenyinaran|Hujan) = \frac{\sum(LamaPenyinaran|Hujan)}{\sum(Hujan)} \quad (4)$$

$$P(LamaPenyinaran|TidakHujan) = \frac{\sum(LamaPenyinaran|TidakHujan)}{\sum(TidakHujan)} \quad (5)$$

Lama Penyinaran pada data latih diklasifikasikan menjadi 2 klasifikasi yaitu lama penyinaran yang statusnya hujan dan lama penyinaran yang statusnya tidak hujan.

$$P(KecepatanAngin|Hujan) = \frac{\sum(KecepatanAngin|Hujan)}{\sum(Hujan)} \quad (6)$$

$$P(KecepatanAngin|TidakHujan) = \frac{\sum(KecepatanAngin|TidakHujan)}{\sum(TidakHujan)} \quad (7)$$

Kecepatan Angin pada data latih diklasifikasikan menjadi 2 klasifikasi yaitu lama kecepatan angin yang statusnya hujan dan kecepatan angin yang statusnya tidak hujan.

C. K-Means

K-Means merupakan Algoritma untuk pengelompokan data (*cluster*) . Dimana setiap cluster akan terus mengalami perubahan, sampai data anggotanya sesuai dengan titik cluster [15]. Langkah-langkah *clustering* dalam *K-Means* sebagai berikut:

1. Menentukan jumlah cluster
2. Menentukan jumlah centroid

Dalam menentukan centroid untuk tahap awal dilakukan secara random, jika sudah masuk ke dalam tahap iterasi maka digunakan rumus 8.

$$\bar{V}_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} x_{kj} \quad (8)$$

dimana:

- V_{ij} = Centroid / rata-rata *cluster* ke i untuk variable ke j
- N_i = Jumlah data yang menjadi anggota *cluster* ke I
- i, k = Indeks dari *cluster*
- j = Indeks dari *variable*
- x_{kj} = Nilai data ke k yang ada didalam cluster tersebut untuk variable ke j

Setelah ditentukan nilai *cluster*, setiap data akan dimasukan ke dalam *cluster* sesuai dengan jarak terdekat (data dikurangi nilai centroid). Setelah semua *variabel* masuk ke dalam *cluster* kemudian dilakukan penyesuaian dengan merubah centroid menggunakan Rumus 8. Dengan cara menghitung jumlah data *variable* dibagi dengan total jumlah data, maka hasil dari perhitungan tersebut dijadikan *cluster* terbaru.

3. Menghitung jarak antara titik centroid dengan titik tiap objek dengan metode *Euclidian Distance* pada rumus 9

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (9)$$

dimana:

- D_e = *Euclidean Distance*
- I = Banyaknya Objek
- (x,y) = Kordinat Objek
- (s,t) = Kordinat Centroid

Rumus 9 adalah menghitung data yang terdekat dengan titik *centroid* dengan cara mengurangi data dengan *centroid*.

4. Pengelompokan objek

Untuk menentukan anggota cluster adalah dengan memperhitungkan jarak minimum objek. Nilai yang diperoleh dalam keanggotaan data pada distance matriks adalah 0 atau 1, dimana nilai 1 untuk data yang dialokasikan ke cluster dan nilai 0 untuk data yang dialokasikan ke cluster yang lain.

5. Kembali ke tahap 2, lakukan perulangan hingga nilai centroid yang dihasilkan tetap dan anggota cluster tidak berpindah ke cluster lain.

D. RMSE

Perhitungan untuk mengetahui persentase keakuratan sebuah data salah satunya dengan rumus *RMSE (Root Mean Square Error)*. Ditunjukkan pada Rumus 10.

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (10)$$

dimana:

y_i = Prediksi

x_i = Nilai Hasil Permalan

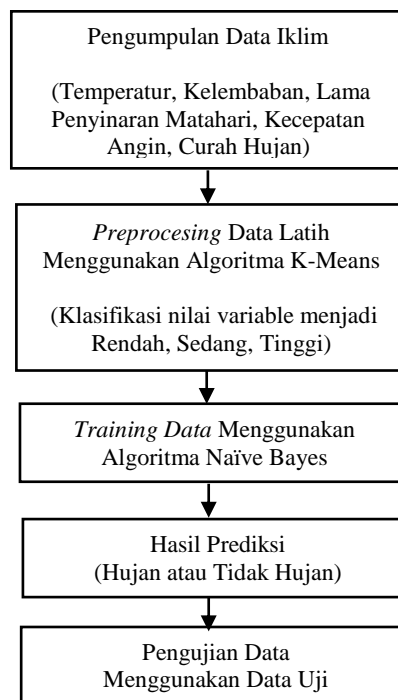
n = Banyaknya Data

Σ = Summation (Jumlah Keseluruhan Nilai)

Rumus 10 digunakan untuk menghitung tingkat prediksi sebuah data, dengan cara mengurangi nilai prediksi dengan data sebenarnya.

III. METODE PENELITIAN

Pada proses prediksi turunnya hujan dalam penelitian ini menggunakan beberapa metode dan tahapan ditunjukkan pada Gambar 1. Berikut:



Gambar 1. Tahapan Penelitian

Tahapan pertama adalah tahapan pengumpulan Data Iklim, yang bersumber dari data resmi Badan Metereologi Klimatologi Dan Geofisika (BMKG) Indonesia, Wilayah Malang Raya. Data diambil dari tanggal 1 juni 2022 sampai 30 Juni 2022. Data yang digunakan meliputi Data Temperatur, Kelembaban, Lama Penyinaran Matahari, Kecepatan Angin Dan Curah Hujan.

Tahapan Kedua adalah preprocessing data, Sebelum data masuk ke data latih, terlebih dahulu data akan dikelompokkan menjadi 3 kelompok untuk memudahkan dalam proses perhitungan *Probabilitas* di *Naïve Bayes*. Tiga kelompok adalah Rendah, Sedang, Tinggi. Dengan penentuannya menggunakan Algoritma *K-Means*.

Tahapan Ketiga adalah *data training* yaitu melatih Algoritma *Naïve Bayes* untuk mendapatkan probabilitas di setiap *variable* Temperatur, Kelembaban, Lama Penyinaran Matahari, Kecepatan Angin.

Tahapan Keempat adalah hasil prediksi diperoleh dari hasil perhitungan *Naïve Bayes* terhadap data uji, dengan nilai keluaran Hujan atau Tidak Hujan.

Tahapan Kelima pengujian data, setelah melakukan prediksi turunnya hujan, selanjutnya adalah menguji data apakah sesuai dengan kejadian sesungguhnya. Data uji diambil dari tanggal 21 – 30 Juni 2022.

IV. HASIL DAN PEMBAHASAN

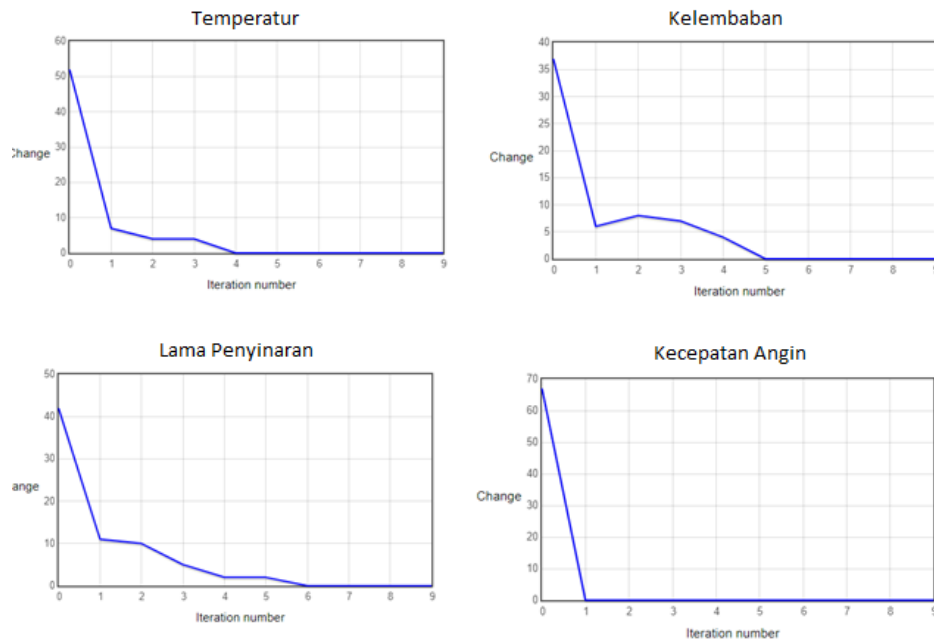
Data latih yang diperoleh adalah data di daerah Malang Raya dari tanggal 1 juni 2022 – 30 juni 2022 data yang diperoleh pada Tabel 1 berikut:

Tabel 1. Data Latih Wilayah Malang

Tanggal	Temperatur	Kelembaban	Curah Hujan	Lama Penyinaran Matahari	Kecepatan Angin
01-06-2022	24.7	89	55.5	7	1
02-06-2022	25.1	85	1.8	0.5	1
03-06-2022	25.5	82	0	5.8	1
04-06-2022	24.8	88	0	6	0
05-06-2022	25.6	86	20.5	5	1
06-06-2022	25.7	82	0	5	1
07-06-2022	26	81	0	4.3	1
08-06-2022	26.2	85	0	7.7	1
09-06-2022	26.2	85	0	5.6	1
10-06-2022	25.6	84	0	3.8	0
11-06-2022	25.8	86	0	4.8	1
12-06-2022	25.6	86	28	5.6	1
13-06-2022	24.8	84	1.5	5	1
14-06-2022	25.4	86	0.7	3.5	0
15-06-2022	26.6	82	23.5	5.7	1
16-06-2022	25.2	90	0	6.5	1
17-06-2022	25.5	86	1.8	3.6	0
18-06-2022	25.3	83	0.5	5.5	0
19-06-2022	25.2	86	2	6.6	1
20-06-2022	25.3	83	2.2	3.5	1
21-06-2022	25.2	79	0	6.8	2
22-06-2022	25.4	76	0	8.8	1
23-06-2022	25.8	81	0	9	2
24-06-2022	25.6	79	0	6.9	2
25-06-2022	25.7	76	0	7.7	2
26-06-2022	25.1	84	0	7	2
27-06-2022	25.2	81	1.5	3.5	2
28-06-2022	24.7	77	0	3.4	2
29-06-2022	24.8	75	0	9.5	2
30-06-2022	25.2	80	0	9.5	2

Pada Tabel.1 pada *variable* curah hujan jika data 0 maka tidak hujan, jika data lebih dari 0 maka turun hujan.

Data hujan kita *preprocessing* menjadi bentuk sederhana dengan tiga parameter yaitu Rendah, Sedang, Tinggi. Kita gunakan Algoritma K-Means untuk mengklasifikasikan setiap parameter. Perhitungan K-Means ditunjukkan pada Grafik Gambar 2.



Gambar 2. Grafik perhitungan centroid pada variable

Dari Gambar 1 Didapat data centroid seperti pada Tabel 2

Tabel 2. Hasil Centroid Setiap Variable

Parameter	Centroid Temperatur	Centroid Kelembaban	Centroid Lama Penyinaran Matahari	Centroid Kecepatan Angin
Tinggi	25.867	86.000	7.990	2.000
Sedang	25.277	81.182	5.592	1.000
Rendah	24.760	76.000	3.657	0.000

Dari hasil *centroid* kemudian dilakukan penentuan setiap *variable* untuk masuk pada data Tinggi, Sedang, atau Rendah. Dengan cara mencari nilai terdekat pada data latih dengan *centroid*. Dihasilkan sebuah data yang telah dilakukan *preprocessing* pada Tabel 3.

Tabel 3. Data Latih Wilayah Malang Setelah Preprocessing

Tanggal	Temperatur	Kelembaban	Curah Hujan	Lama Penyinaran Matahari	Kecepatan Angin
01-06-2022	Sedang	Tinggi	Hujan	Tinggi	Rendah
02-06-2022	Tinggi	Sedang	Hujan	Rendah	Sedang
03-06-2022	Tinggi	Sedang	Tidak Hujan	Tinggi	Rendah
04-06-2022	Tinggi	Sedang	Tidak Hujan	Tinggi	Rendah
05-06-2022	Tinggi	Sedang	Tidak Hujan	Sedang	Rendah
06-06-2022	Sedang	Tinggi	Hujan	Tinggi	Rendah
07-06-2022	Tinggi	Sedang	Hujan	Sedang	Sedang
08-06-2022	Sedang	Tinggi	Tidak Hujan	Tinggi	Sedang
09-06-2022	Sedang	Tinggi	Hujan	Sedang	Rendah
10-06-2022	Tinggi	Tinggi	Hujan	Rendah	Sedang
11-06-2022	Tinggi	Tinggi	Tidak Hujan	Sedang	Sedang
12-06-2022	Tinggi	Sedang	Hujan	Tinggi	Sedang
13-06-2022	Tinggi	Tinggi	Hujan	Tinggi	Rendah
14-06-2022	Rendah	Tinggi	Hujan	Sedang	Rendah
15-06-2022	Tinggi	Tinggi	Hujan	Rendah	Sedang
16-06-2022	Sedang	Tinggi	Hujan	Sedang	Sedang
17-06-2022	Sedang	Tinggi	Hujan	Sedang	Rendah
18-06-2022	Tinggi	Sedang	Hujan	Rendah	Rendah
19-06-2022	Tinggi	Sedang	Tidak Hujan	Tinggi	Rendah
20-06-2022	Tinggi	Sedang	Hujan	Tinggi	Sedang

21-06-2022	Tinggi	Rendah	Tidak Hujan	Tinggi	Sedang
22-06-2022	Tinggi	Rendah	Tidak Hujan	Tinggi	Sedang
23-06-2022	Tinggi	Sedang	Tidak Hujan	Tinggi	Sedang
24-06-2022	Tinggi	Sedang	Tidak Hujan	Sedang	Rendah
25-06-2022	Tinggi	Sedang	Tidak Hujan	Tinggi	Sedang
26-06-2022	Tinggi	Sedang	Tidak Hujan	Sedang	Rendah
27-06-2022	Tinggi	Sedang	Hujan	Sedang	Sedang
28-06-2022	Sedang	Sedang	Hujan	Tinggi	Sedang
29-06-2022	Tinggi	Rendah	Tidak Hujan	Tinggi	Rendah
30-06-2022	Sedang	Sedang	Tidak Hujan	Tinggi	Rendah

Dari Tabel 3. Dilakukan perhitungan menggunakan Algoritma Naïve Bayes menggunakan data uji dari tanggal 21 – 30 Juni 2022 didapatkan hasil pada Tabel 4.

$$RMAE = \frac{|10-2|}{10} = 80$$

Dari hasil pengujian data cuaca pada bulan Juni 2022. Didapatkan hasil yang sangat baik keakuratannya yaitu sebesar 80 % data sesuai dengan keadaan yang sebenarnya.

V. KESIMPULAN

Setelah dengan perhitungan dengan tepat maka didapatkan hasil dengan menggunakan Algoritma K-Means dan Naïve Bayes mampu memprediksi turunnya hujan sebesar 80 % dari keaslian di lapangan. Dengan data yang tidak sesuai yaitu hanya 2 atau dua kesalahan prediksi dengan metode tersebut, Maka metode naïve bayes sebagai perhitungan probabilitas hujan dalam suatu daerah teruji.

DAFTAR PUSTAKA

- [1] Kurniawan Agusta "Evaluasi Pengukuran Curah Hujan Antara Hasil Pengukuran Permukaan (AWS, HELLMAN, OBS) dan Hasil Estimasi (Citra Satelit =GSMaP) Di Stasiun Klimatologi Mlati Tahun 2018" Jurnal Geografi, Edukasi dan Lingkungan (JGEL) Vol. 4, No. 1, Januari 2020
- [2] Marthin Luter Laila, Yudi Setiawan "Perbandingan Hasil Klasifikasi Curah Hujan menggunakan Metode SVM dan NBC" Jurnal Statiska industri dan Komputasi, vol. 05, pp. 51-56, 2020.
- [3] Pemerintah Kota Malang, "Geografis Kota Malang" <http://malangkota.go.id/sekilas-malang/geografis>. Diakses 27 Juli 2022
- [4] BMKG Jawa Timur "Stasiun Pengamatan Iklim Karangploso" <https://staklim-jatim.bmkg.go.id> Diakses 1 Januari 2022
- [5] P. Hubert, R. Sudaryono Phoenix, D. Suhartono, Classifying promotion images using optical character recognition and Naïve Bayes classifier, *Procedia. Comput. Sci.* 179 (2020) (2021) 498–506
- [6] Anton Yuda, Dedi Sulisty, Ilham Mufandi, GIS-based and Naïve Bayes for nitrogen soil mapping in Lendah, Indonesia, *Sensing and Bio Sensing Research.* 33 (2021) 100435
- [7] Chen Hongh, Hu Shonghua, Hua Rui, Zhao Xiuju "Improved naïve Bayes classification algorithm for traffic risk management" *EURASIP Journal on Advances in Signal Processing* (2021) 2021:30
- [8] Amril Mutoi Siregar, Tukino, Sultan Faisal "Klasifikasi untuk Prediksi Cuaca Menggunakan Esemble Learning" *Jurnal Pengkajian dan Penerapan Teknik Informatika*, vol. 13, 2020
- [9] Hidayat Miwan Kurniawan, "Implementation of K-Means and K-Medoids in Grouping Potential Areas of Chicken Meat Production" *Jurnal Teknologi Industri Pertanian* 32 (3): 239-247, Desember 2022.
- [10] Mulyo Iksan Adityo, Heikal Jerry " Customer Clustering Using The K-Means Clustering Algorithm in Shopping Mall in Indonesia" *Management Analysis Journal* 12 (4) (2022)
- [11] Putro H Febrianto, Vulandari Retno Tri, Saptomo L Wawan "Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan" *Jurnal TIKomSiN*, Vol. 8, No. 2, Oktober 2020

- [12] Pambudi Tities Wahyu, Witanti Arita " Penerapan Algoritma K-Means Clustering untuk Menganalisis Penjualan pada Toko Ayu Collection Barbasis Web" Jurnal Informatika Universitas Pamulang Vol. 6, No. 3, September 2021 (645-650)
- [13] Pradana Dany, Sugiharti Endang "Implementation Data Mining with Naïve Bayes Classifier Method and Leplace Smoothing to Predict Students Learning Results" Recursive Journal of Informatics, Vol. 1, March 2023
- [14] Sanjaya Indra Fadil, Heksaputra Dadang "Prediksi Rerata Harga Beras Tingkat Grosir Indonesia dengan Long Short Term Memory" Jurnal Teknik Informatika dan Sistem Informasi Vol. 7, No. 2, Agustus 2020, Hal. 163-174
- [15] Sarasvananda Ida Bagus Gede, Wardoyo Retantyo, Sari Anny Kartika "The K-Means Clustering Algorithm With Semantic Similarity To Estimate The Cost of Hospitalization" IJCCS (Indonesian Journal of Computing and Cybernetics Systems) Vol.13, No.4, October 2019,pp. 313~322