

Prediksi Perubahan Iklim Untuk Pertumbuhan Tanaman Jeruk Keprok Menggunakan Naïve Bayes

Ahmad Chusyairi ¹⁾, Toto Haryanto ²⁾, Rachmad Nur Hayat ³⁾

^{1,2)}Fakultas Informatika, Universitas Bina Insani

Jl. Siliwangi No.6, Rawa Panjang, Kota Bekasi, Jawa Barat, Indonesia

³⁾Departement of Computer Science, IPB University

Jl. Kampus IPB, Jl. Raya Dramaga, Babakan, Kec. Dramaga, Kabupaten Bogor, Jawa Barat, Indonesia

E-Mail : ahmadchusyairi@binainsani.ac.id ¹⁾; totoharyanto@apps.ipb.ac.id ²⁾; rachmadnurhayat@outlook.com ³⁾;

ABSTRAK

Iklim merupakan suatu pola data yang sangat berpengaruh pada pertumbuhan tanaman di masa mendatang, Tanaman jeruk keprok merupakan bahan baku untuk kebutuhan pangan masyarakat, sehingga diperlukan strategi untuk menjaga stabilitasi produksi melalui berbagai strategi terutama memaksimalkan data yang mudah di akses di internet. Iklim memiliki komponen yang dapat mempengaruhi pertumbuhan suatu tanaman diantaranya temperatur udara, kelembapan udara, lama penyinaran, dan curah hujan. Dalam rangka memaksimalkan penelitian ini, metode teorema bayes digunakan untuk dapat melakukan klasifikasi data didasarkan pada nilai probabilitas iklim baik atau buruk pada pertumbuhan tanaman, diharapkan naïve bayes dengan akurasi sebesar 85%, presisi sebesar 83%, recall sebesar 100%, dan *F1 Score* sebesar 90% dapat membantu upaya menjaga stabilitasi pertumbuhan tanaman dengan memanfaatkan data iklim yang tersedia.

Kata Kunci – iklim, naïve bayes, prediksi, tanaman, teorema bayes

1. PENDAHULUAN

Salah satu komponen utama dalam membangun daya tahan suatu perekonomian adalah pangan, terlebih saat terjadinya goncangan, gangguan, dan krisis. Tanaman pangan dan hortikultura merupakan bahan baku untuk kebutuhan pangan masyarakat. Oleh karena itu, jaminan ketersediannya menjadi suatu keniscayaan. Strategi untuk menyiasatinya, antara lain (a) tetap menjaga stabilisasi produksi dengan memberikan bantuan pasokan input produksi dan bantuan alat-alat Bertani; (b) membuat pemetaan jadwal tanam dan panen untuk mempermudah jalur distribusi dari wilayah sentra produksi ke wilayah konsumen; (c) mempersiapkan armada panen dan system distribusi produk pangan; (d) dengan berkurangnya daya beli masyarakat, maka peran Satgas Pangan dibutuhkan dalam menstabilkan harga dan jaminan serta kemudahan untuk mendapatkan pangan.

Berbagai jenis tanaman pangan dan pertumbuhannya yang dibudidayakan pada suatu Kawasan memiliki keadaan iklim yang menjadi pengaruh signifikan dalam proses Bertani. Produktivitas tanaman menjadi naik turun diakibatkan kondisi iklim yang terjadi pada saat-saat tertentu. Dikarenakan pengaruh iklim tersebut, pertumbuhan tanaman perlu diketahui oleh petani yaitu dengan mengetahui kecocokan tanaman pada kondisi iklim di Kawasan tersebut.

Klimatologi adalah cabang ilmu yang mempelajari tentang iklim, didefinisikan sebagai pola cuaca yang dirata-ratakan selama periode waktu tertentu untuk memperoleh pola yang konsisten dari kondisi atmosfer yang diharapkan (Ankit, 2018). Dengan memperoleh data rata-rata cuaca untuk jangka waktu yang lama dan tidak terbatas memungkinkan untuk memprediksi pola iklim suatu daerah. Kondisi iklim terkandung dalam beberapa unsur dan

komponen yang diantaranya yaitu suhu udara (C°), angin, kelembaban (%), penguapan, curah hujan (mm), serta lama penyinaran matahari.

Sebagai negara tropis dengan musim hujan dan kemarau, Indonesia juga merasakan dampak dari perubahan iklim yang terjadi (Sena, Dillak, Leunupun, & Santoso, 2020), maka dari itu sebagai upaya untuk mengantisipasi kerugian terhadap pertumbuhan tanaman di masa mendatang, Stasiun Klimatologi Jawa Timur yang terletak pada Lintang - 7.90080, bujur 112.59790 memperoleh data laporan harian iklim yang dapat peneliti ambil dari website BMKG dengan rentang waktu mulai dari Januari 2020 hingga September 2022, format dari data adalah excel, sehingga dapat mudah dilakukan pengolahan untuk keperluan analisis dan prediksi perubahan iklim melalui metode data mining dan algoritma Naïve Bayes.

2. TINJAUAN PUSAKA

Perubahan iklim, seperti yang didefinisikan oleh UNFCCC adalah 'perubahan yang dikaitkan secara langsung atau tidak langsung dengan aktivitas manusia yang mengubah komposisi atmosfer global dan selain variabilitas iklim alami yang diamati selama periode waktu yang sebanding'. Perubahan iklim diproyeksikan menyebabkan peningkatan suhu, variasi curah hujan, dan peningkatan frekuensi kejadian ekstrim seperti panas, gelombang dingin, hari-hari beku, kekeringan, banjir, dan lain-lain. Berbagai proses tanaman seperti pertumbuhan vegetatif, pembungaan, pembuahan dan kualitas buah sangat rentan terhadap perubahan iklim. Dua parameter utama perubahan iklim yang memiliki implikasi luas pada tanaman adalah pola curah hujan yang lebih tidak menentu dan suhu tinggi yang tidak dapat diprediksi yang akibatnya dapat menurunkan produktivitas tanaman. Kekeringan mengurangi

pembentukan buah dan meningkatkan keretakan buah pada delima dan lengkeng. Peningkatan suhu mempengaruhi fotosintesis secara langsung, menyebabkan perubahan gula, asam organik, kandungan flavonoid, kekencangan, dan aktivitas antioksidan (Haokip, Shankar, & Lalrinnggheta, 2020).

A. Data Mining

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstrak dan mengidentifikasi informasi yang berguna dan pengetahuan terkait dari database besar. Istilah data mining memiliki esensi sebagai suatu disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menggali pengetahuan dari data atau informasi yang kita miliki. Data mining, sering juga disebut sebagai *Knowledge Discovery in Database* (KDD). Penemuan Pengetahuan dalam Basis Data adalah aktivitas yang mencakup pengumpulan, penggunaan data historis untuk menemukan keteraturan, pola, atau hubungan dalam kumpulan data yang besar (Ali, Khairan, Tempola, & Fuad, 2021).

B. Confusion Matrix

Dalam pengenalan pola dan disiplin lain dari *machine learning*, jumlah elemen diagonal dari matriks konfusi banyak digunakan untuk mengukur keberhasilan klasifikasi berdasarkan algoritma atau pengamatan manusia (Düntsche & Gediga, 2019). Maka dari itu Confusion Matrix digunakan untuk menggambarkan performa dari sebuah model atau algoritma secara spesifik. Setiap baris dari matrix tersebut, merepresentasikan kelas aktual dari data, dan setiap kolom merepresentasikan kelas prediksi dari data (atau sebaliknya) (Saputro & Sari, 2019). Bentuk dari *confusion matrix* dapat dilihat pada Tabel 1.

Table 1. Bentuk Confusion Table

	Prediksi Positif	Prediksi Negatif
Aktual Positif	True Positif (TN)	False Negatif (FN)
Aktual Negatif	False Positif (FP)	True Negatif (TN)

Keterangan pada setiap isian dari confusion matrix dijelaskan sebagai berikut :

1) *True Positif* = Memberi jumlah banyak data yang diprediksi sebagai positif dan ternyata aktual menyatakan demikian

2) *True Negatif* = Memberi jumlah banyak data yang diprediksi sebagai negative dan ternyata aktual menyatakan demikian

3) *False Positif* = Memberi jumlah banyak data yang salah diprediksi sebagai positif yang seharusnya negatif

4) *False Negatif* = Memberi jumlah banyak data yang salah diprediksi sebagai negative yang seharusnya positif

Dengan menggunakan angka-angka tersebut, dapat diperoleh skor yang sangat berguna untuk membantu mengukur performa model melalui teknik perhitungan yang diantaranya :

Akurasi, yaitu seberapa baik model benar melakukan klasifikasi, persamaan 1 menunjukkan cara memperoleh skor akurasi.

$$\frac{TP + TN}{Total} \tag{1}$$

Precision, yaitu skor yang mengukur seberapa sering model dengan benar memprediksi positif, untuk mendapatkan skor ini dapat diperoleh pada persamaan 2

$$\frac{TP}{Total} \tag{2}$$

Recall, merupakan skor untuk mengukur seberapa sering model memprediksi positif ketika aktualnya juga positif, skor ini dapat diperoleh melalui persamaan 3.

$$\frac{TP}{FN + TP} \tag{3}$$

F1-Score, merupakan rata-rata harmonic dari *Precision* dan *Recall*. *F1-Score* dapat diperoleh melalui persamaan 4.

$$2 * \frac{precision * recall}{precision + recall} \tag{4}$$

3. METODE PENELITIAN

Metodologi yang digunakan pada penelitian ini memanfaatkan algoritma yang disebut dengan Algoritma Naïve Bayes (NB). Naïve Bayes merupakan salah satu dari seluruh algoritma dengan teknik klasifikasi. Dalam sejarahnya, Naïve Bayes adalah klasifikasi dengan metode yang mengandalkan nilai probabilitas dan statistik yang diusulkan oleh ilmuwan Inggris bernama Thomas Bayes dengan memprediksi peluang masa depan berdasarkan data pengalaman sebelumnya, sehingga kini lebih dikenal dengan teorema Bayes. Teorema tersebut digabungkan dengan Naïve yang mana merupakan asumsi kondisi diantara atribut independen. Klasifikasi Naïve Bayes mengasumsikan bahwa suatu sifat ada atau tidak pasti dari suatu kelas tidak ada hubungannya dengan karakteristik kelas tersebut (Ali, Khairan, Tempola, & Fuad, 2021). Probabilitas bayes merupakan salah satu cara untuk mengatasi ketidak pastian data dengan memanfaatkan data sampel yang diperoleh dari populasi juga memperhitungkan suatu distribusi awal yang disebut menggambarkan pengetahuan awal tentang parameter sebelum pengamatan dilakukan dan dinyatakan dalam suatu distribusi (Zunaidi, Pane, & Nasuyha, 2021).

Teorema Bayes yang hanya dibatasi oleh dua buah kejadian dapat diperluas untuk kejadian *n* buah sebagaimana yang ditunjukkan pada rumus berikut:

$$P(B_n|A) = \frac{P(B_n \cap A)}{P(A)} \tag{5}$$

$$P(A|B_n) = \frac{P(A \cap B_n)}{P(B_n)} \quad (6)$$

Dengan syarat $P(A)$ dan $P(B_n) \neq 0$

Persamaan diatas dapat dinyatakan menjadi sebuah persamaan yang lengkap dengan menyamakan pembilang pada kedua persamaan tersebut, sehingga diperoleh hubungan antara probabilitas kejadian bersyarat antara A dengan himpunan B dan sebaliknya. Berikut adalah persamaan teorema bayes:

$$P(B_n|A) = \frac{P(A|B_n)P(B_n)}{P(A)} \quad (7)$$

Berdasarkan hubungan probabilitas A dengan probabilitas kejadian bersyarat dapat dinyatakan sebagai berikut :

$$P(A) = \sum_n^N P(A|B_n)P(B_n) \quad (8)$$

Persamaan berikut merupakan gabungan persamaan (7) dengan persamaan (8) sehingga terlihat lebih kompleks:

$$P(B_n|A) = \frac{P(A|B_n)P(B_n)}{P(A|B_1)P(B_1)+\dots+P(A|B_N)P(B_N)} \quad (10)$$

Sehingga, dapat disimpulkan bahwa dalam menangani data atribut yang tidak relevan, Naive Bayes sangat efisien, handal dan efektif efektif dan handal. Sehingga kemungkinan besar dapat mengatasi dataset rumit dalam jumlah banyak baik dengan atribut yang bersifat variabel maupun kontinu.

4. HASIL DAN PEMBAHASAN

A. Tanaman

Pada penelitian ini, penulis melakukan standarisasi terhadap tanaman jeruk berjenis keprok yang mana menjadi fokus pada penelitian ini karena buah jeruk lebih cenderung memiliki potensi pertumbuhan baik akan perubahan iklim yang terjadi pada daerah tersebut (Putri Br Sitepu, Fajriani, & Sulisty, 2022), didukung dengan masa panen yang bisa dilakukan 5 (lima) kali dalam dua belas bulan pada daerah tersebut (Balai Penelitian Tanaman Jeruk dan Subtropika, 2019). Buah jeruk merupakan salah satu jenis buah-buahan yang paling banyak digemari oleh masyarakat kita. Dapat disadari bahwa komoditas buah jeruk memang dapat meningkatkan taraf hidup masyarakat, terutama jenis komoditas jeruk keprok yang mempunyai nilai ekonomis tinggi, tahan agak lama dan mudah menyimpannya (Rahimah, Neni, & Rasmilah, 2020). Dengan menjadikan tanaman jeruk keprok tersebut menjadi acuan, maka untuk mengklasifikasi bahwa iklim

baik maupun buruk akan menggunakan data iklim sebagaimana yang ditampilkan tabel 1.

Tabel 2. Acuan Standarisasi Tanaman Terhadap Iklim

Properti	Nilai Rentang
Temperatur Udara	25-30 C
Kelembapan Udara	50-85%
Lama Penyinaran	6-8 Jam
Curah Hujan	2.6-8mm/hari

Sumber: Balai Penelitian Tanaman Jeruk dan Subtropika (Balai Penelitian Tanaman Jeruk dan Subtropika, 2019).

B. Dataset BMKG



Gambar 1 Lokasi Stasiun Klimatologi Jawa Timur

Dataset BMKG yang didapat secara online (BMKG, n.d.) dari Stasiun Klimatologi Jawa Timur berlokasi pada lintang -7.90080 dan bujur 112.59790 berformat excel (*csv comma delimited*) memiliki dimensi 982 baris data iklim yang tercatat sejak Januari 2020 hingga awal September 2022 dan 11 kolom (features) terdiri atas tanggal yaitu kolom 1, temperatur minimum yang dinotasikan sebagai "tn" yaitu kolom 2(°C), temperature maksimum yang dinotasikan sebagai "tx" yaitu kolom 3 (°C), temperatur rata-rata yang dinotasikan sebagai "tavg" yaitu kolom 4 (°C), kelembapan rata-rata yang dinotasikan sebagai "rh_avg" yaitu kolom 5 (%), curah hujan yang dinotasikan sebagai "rr" yaitu kolom 6 (mm), lamanya penyinaran sinar matahari yang dinotasikan sebagai "ss" yaitu kolom 7 (jam), kecepatan angin maksimum yang dinotasikan sebagai ff_x yaitu kolom 8 (m/s), arah angin saat kecepatan maksimum yang dinotasikan sebagai ddd_x yaitu kolom 9 (°), kecepatan angin rata-rata yang dinotasikan sebagai ff_avg yaitu kolom 10 (m/s), dan arah angin terbanyak yang dinotasikan sebagai ddd_car yaitu kolom 11 (°), berikut pada tabel 2 ditampilkan isian dataset yang dimaksud pada penjelasan sebelumnya.

Dataset masih mengandung isian yang menyebabkan dataset tidak dapat diproses secara komputasi dengan baik (disebut sebagai data yang masih mentah atau *raw data*), maka dari itu perlu dilakukan tahap *preprocessing data* sebagaimana yang akan dijelaskan pada pembahasan selanjutnya

Tabel 3. Dataset Iklim BMKG

No	1	2	3	4	5	6	7	8	9	10
1	01/01/2020	20.4	28.2	23.7	87	43.3	4.4	3.0	170	1
2	02/01/2020	21.2	28.8	23.8	90	2.6	3.9	4	150	1
3	03/01/2020	21.4	28.8	24.0	86	1.9	4.6	4	160	1
4	04/01/2020	21.6	28.6	24.2	86	2.1	3	4	290	1
5	05/01/2020	22	28.3	24.1	88	1	1.3	3	140	2
...										
977	04/09/2022	19.8	29	23.7	78	0	9	3	170	2
978	05/09/2022	19.4	29.3	23.8	78	16.4	5.7	5	90	2
979	06/09/2022	19	29.4	23.6	81	0	2.6	4	140	1
980	07/09/2022	19	28	24.1	83	0	5.2	4	200	2
981	08/09/2022	19.3	30	23.5	81	11.4	2.9	5	110	2

C. Pemrosesan Dataset

Data yang didapat dari BMKG tidak dapat diproses langsung untuk keperluan analisa dan masih dalam bentuk *raw data*, maka dari itu diperlukan penyederhanaan data agar lebih mudah saat melakukan proses komputasi pada data. Tahap ini berikut disebut sebagai preprocessing yang mana pada tahap ini dibagi menjadi 4, yaitu.

1) *Data Cleaning*: Dataset iklim memiliki beberapa data dengan kriteria yang memenuhi syarat untuk dihapus seperti data masih ada yang tidak ada isianannya diantara *feature*, dan juga terdapat data dengan nilai konstan seperti 8888 memiliki arti data tidak terukur sedangkan 9999 memiliki arti tidak ada data (tidak dilakukan pengukuran).

2) *Feature Selection*: Sebagai data yang nantinya akan diintegrasikan dengan data standarisasi tanaman, beberapa *feature* pada dataset tidak diperlukan, selain itu manfaat *feature selection* akan membantu peneliti melakukan pendekatan pada suatu data dengan cara yang lebih sistematis dan ramah dengan *machine learning*. Pemilihan *feature* menggunakan pendekatan berbasis relevansi terhadap data eksternal, sehingga *feature* yang ada pada data eksternal-lah yang akan dijadikan acuan untuk nantinya diterapkan pada dataset BMKG.

3) *Discretization*: Diantara *feature* mengandung nilai yang bersifat kontinu, maka dari itu nilai-nilai tersebut perlu diubah menjadi nilai-nilai diskrit, yaitu mengelompokannya sebagai buruk dan baik. Penentuan nilai-nilai diskrit tersebut didasari atas standarisasi pertumbuhan tanaman (*Heksaputra, Naimah, Azani, & Iswari, 2013*), sebagai contoh dari Tabel 1, jeruk keprok akan tumbuh dengan baik dengan syarat temperature udara dari 25 hingga 30°C, kelembapan udara 50% hingga 85%, lama

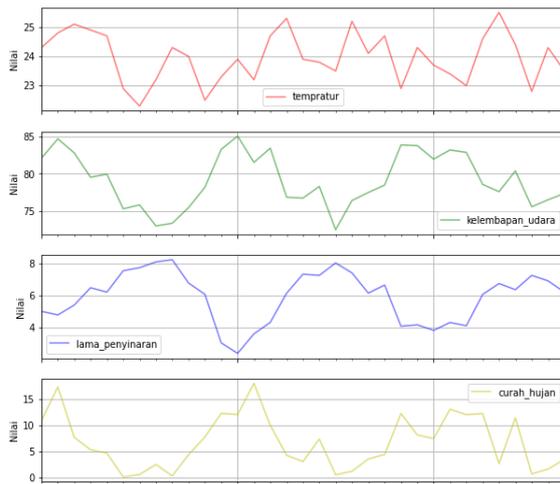
penyinaran 6 hingga 8 jam, dan curah hujan dari 2.6 hingga 8 mm/hari nya.

Data yang dimiliki masih dalam bentuk per-hari, sehingga perlu diubah menjadi menjadi data per-bulan secara *timeseries* dengan menggunakan fungsi *resample*, sehingga didapat data sebagaimana yang ditunjukan oleh tabel 3 yaitu kolom 1 (satu) untuk menyatakan tahun dan bulan, kolom 2 (dua) untuk menyatakan *mean* (rerata) suhu per-bulan, kolom 3 (tiga) untuk menyatakan *mean* kelembapan udara selama per-bulan, kolom 4 (empat) untuk menyatakan *mean* lama penyinaran perbulan, dan kolom 5 yang menyatakan *mean* curah hujan selama sebulan sehingga bentuk dataset menjadi seperti berikut:

Tabel 4. Dataset Hasil Preprocessing

No	1	2	3	4	5
1	2020-01	24.66	82.13	5.0	10.8
2	2020-02	24.22	84.72	4.8	17.32
3	2020-03	24.53	82.83	5.4	7.7
4	2020-04	24.92	79.55	6.5	5.31
5	2020-05	24.71	79.96	6.2	4.67
6	2020-06	23.71	75.30	7.5	0.10
7	2020-07	23	75.83	7.7	0.58
...					
27	2022-03	24.30	82.89	4.1	12
28	2022-04	24.42	78.58	6	12.22
29	2022-05	24.93	77.60	6.7	2.6
30	2022-06	23.58	80.40	6.3	11.41
31	2022-07	23.32	75.58	7.2	0.70
32	2022-08	23.47	76.51	6.9	1.61
33	2022-09	24.05	77.37	6.2	3.47

Pada Gambar 1, ditampilkan visual yang menggambarkan keadaan iklim dari bulan ke bulan, dapat dilihat bahwa tempratur memiliki rerata tempratur yang antara 23 hingga 25°C, kelembapan udara antara 72 hingga 85%, rata-rata lama penyinaran perhari selama perbulannya terjadi 2.4 hingga 8.2 jam, dan curah hujan dari 0 hingga 18mm/hari-nya, dapat dilihat juga bahwa awal tahun hingga menuju pertengahan tahun terjadi perubahan iklim yang cukup signifikan pada sekitaran bulan januari yang cenderung musim penghujan. Dapat terlihat bahwa frekuensi hujan yang menurun sekitar bulan juli diiringi dengan rata-rata lama penyinaran yang semakin naik durasinya, serta temprturnya yang stabil.



Gambar 2 Perubahan Iklim Bulan ke Bulan

Sebagaimana yang disebutkan sebelumnya, selanjutnya dilakukan proses *discretization* pada dataset dengan menjadikan syarat iklim tanaman jeruk sebagai cakupan iklim yang baik, maka dari itu pemberian kolom “peluang pertumbuhan” dilakukan dengan cara menggolongkan terdahulu tiap data terutama pada nilai masing-masing *feature* supaya dinotasikan sebagai buruk ataupun baik, sebagai contoh temperatur pada bulan Januari tahun 2022 dikatakan baik jika rerata suhu perbulannya masuk pada kriteria iklim tanaman jeruk antara 25 hingga 30°C.

Tabel 5 *Discretized Data*

No	1	2	3	4	5
1	2020-01	Buruk	Baik	Buruk	Buruk
2	2020-02	Buruk	Baik	Buruk	Buruk
3	2020-03	Baik	Baik	Baik	Buruk
4	2020-04	Buruk	Baik	Baik	Baik
5	2020-05	Buruk	Baik	Baik	Baik
6	2020-06	Buruk	Baik	Buruk	Baik
7	2020-07	Buruk	Baik	Buruk	Baik
...					
27	2022-03	Buruk	Baik	Buruk	Buruk
28	2022-04	Buruk	Baik	Buruk	Baik
29	2022-05	Baik	Baik	Baik	Baik
30	2022-06	Buruk	Baik	Buruk	Baik
31	2022-07	Buruk	Baik	Buruk	Baik
32	2022-08	Buruk	Baik	Buruk	Baik
33	2022-09	Buruk	Baik	Baik	Baik

Selanjutnya untuk mendapatkan nilai probabilitas baik dan buruk dengan membagi total kemunculan suatu nilai yang dijadikan target dibagi total *feature* yang terkandung. Adapun yang menjadi ketentuan pada penelitian kali ini adalah nilai probabilitas baik (persamaan 6), jika probabilitas baik kurang dari 0,5, maka peluang pertumbuhan pada bulan tersebut dikatakan buruk atau dinotasikan sebagai angka 0 (nol), selain dengan ketentuan tersebut, maka dapat dikatakan bahwa pada bulan tersebut peluang pertumbuhannya baik yang akan dinotasikan sebagai angka 1 (satu). Presentase ditunjukkan pada tabel V

yang dinotasikan sebagai kolom 6 dan hasil klasifikasi peluang pertumbuhan yang dinotasikan sebagai kolom 7.

$$p = \frac{\text{banyak baik yang muncul}}{\text{total feature}} = \frac{2}{4} = 0.50 \quad (6)$$

Hasil dari klasifikasi peluang baik hingga buruk pertumbuhan tanaman jeruk keprok berdasarkan iklim Januari 2020 hingga September 2022 terdapat pada tabel berikut:

Tabel 6 *Discretized Data*

	1	2	3	4	5	6	7
2020-01	Buruk	Baik	Buruk	Buruk	0.25	0	
2020-02	Buruk	Baik	Buruk	Buruk	0.25	0	
2020-03	Baik	Baik	Baik	Buruk	0.75	1	
2020-04	Buruk	Baik	Baik	Baik	0.75	1	
2020-05	Buruk	Baik	Baik	Baik	0.75	1	
2020-06	Buruk	Baik	Buruk	Baik	0.50	1	
2020-07	Buruk	Baik	Buruk	Baik	0.50	1	
...							
2022-03	Buruk	Baik	Buruk	Buruk	0.25	0	
2022-04	Buruk	Baik	Buruk	Baik	0.50	1	
2022-05	Baik	Baik	Baik	Baik	1	1	
2022-06	Buruk	Baik	Buruk	Baik	0.50	1	
2022-07	Buruk	Baik	Buruk	Baik	0.50	1	
2022-08	Buruk	Baik	Buruk	Baik	0.50	1	
2022-09	Buruk	Baik	Baik	Baik	0.75	1	

Selanjutnya, menaruh kolom peluang pertumbuhan (kolom 7) untuk digabungkan ke dataset iklim bulanan dengan nama kolom 6, sehingga bentuk dari dataset yang akan digunakan oleh model Naïve Bayes adalah sebagai berikut:

Tabel 7 *Discretized Data yang Di merge*

No	1	2	3	4	5	6
1	2020-01	24.66	82.13	5.0	10.8	0
2	2020-02	24.22	84.72	4.8	17.32	0
3	2020-03	24.53	82.83	5.4	7.7	1
4	2020-04	24.92	79.55	6.5	5.31	1
5	2020-05	24.71	79.96	6.2	4.67	1
6	2020-06	23.71	75.30	7.5	0.10	1
7	2020-07	23	75.83	7.7	0.58	1
...						
27	2022-03	24.30	82.89	4.1	12	0
28	2022-04	24.42	78.58	6	12.22	1
29	2022-05	24.93	77.60	6.7	2.6	1
30	2022-06	23.58	80.40	6.3	11.41	1
31	2022-07	23.32	75.58	7.2	0.70	1
32	2022-08	23.47	76.51	6.9	1.61	1
33	2022-09	24.05	77.37	6.2	3.47	1

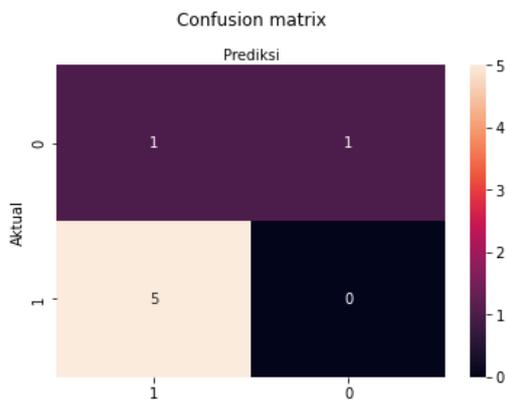
Dengan bentuk data sebagaimana yang ditampilkan pada Tabel VI, maka model dapat mudah membedakan data berbasis dua kelas yaitu angka 0 (nol) untuk menyatakan peluang pertumbuhan buruk dan angka 1 (satu) untuk menyatakan peluang pertumbuhan yang baik.

D. Data Training dan Testing

Selanjutnya dataset dibagi menjadi data training dan data uji, Adapun dilakukannya pembagian ini untuk memberikan porsi untuk model mempelajari data dan mengambil beberapa data yang tidak akan model pelajari sebagai bahan evaluasi model mendatang, porsi pembagian untuk data training dan data testing yaitu 80% untuk diketahui oleh model atau disebut sebagai data training, dan 20% untuk data yang akan menjadi bahan uji model. Perlu diketahui bahwa penskalaan data tidak diterapkan yang mana biasa dilakukan setelah pembagian, berdasarkan pengujian yang dilakukan oleh Ambarwari et al. (2022) (Ambarwari, Adrian, & Herdiyeni, 2020) Naïve Bayes tetap memiliki performa yang stabil Ketika dataset tidak menerapkan normalisasi data min-max maupun standarisasi (*zero-mean*) (Ahsan, Mahmud, Saha, Gupta, & Siddique, 2021).

E. Evaluasi Model

Evaluasi dilakukan untuk mengetahui sejauh apa model dapat mempelajari data, berdasarkan hasil *confusion matrix*, didapat bahwa model dapat memprediksi menggunakan 20% dari dataset yang dimiliki dengan benar sebanyak 6 data dan kesalahan prediksi yang dihasilkan sejumlah 1 data, sehingga didapat dengan dengan tingkat skor akurasi sebesar 0.85, presisi sebesar 0.83, *recall* sebesar 1.0 dan *F1 Score* sebesar 0.90.



Gambar 3 *Confusion Matrix* Model

F. Uji Prediksi

Berdasarkan pada data training, misalkan diketahui data untuk bulan oktober tahun 2022 adalah sebagai berikut:

TABEL VIII
ISIAN UNTUK MENGUJI MODEL

Properti	Nilai	Unit
Tempratur	24	°C
Kelembapan	75	%
Lama Penyinaran	6	Jam
Curah Hujan	4	mm/hari

Berdasarkan standarisasi iklim tanaman jeruk keprok, maka Naïve Bayes mampu memprediksi berdasarkan data iklim yang sudah berlalu

menggunakan fungsi *predict* yang parameternya adalah *array* nilai keadaan baru. Berdasarkan uji coba ini, model dapat menghasilkan keluaran berupa kelas 1 (satu) yang artinya iklim pada bulan Oktober pada daerah malang adalah baik bagi tanaman jeruk keprok sebagaimana yang ditampilkan pada gambar 4. Probabilitas untuk kelas baik dan buruk juga ditampilkan menggunakan fungsi *predict_proba* yang mana probabilitas peluang pertumbuhan baik pada bulan Oktober sebesar 0.92 atau jika ingin mendapatkan nilai presentase yaitu dengan mengkalikan nilai tersebut dengan 100 maka menghasilkan 92.5% peluang baik, begitu juga dengan peluang buruk dengan nilai 0.7 yang jika dikalikan 100 maka diketahui bahwa peluang pertumbuhan jeruk keprok untuk buruk memiliki tingkat probabilitas sebesar 13%.

```

tempratur = 24
kelembapan = 75
lama_penyinaran = 6
curah_hujan = 4

new_data = np.array([[tempratur, kelembapan, lama_penyinaran, curah_hujan]])

result_classes = ['BURUK', 'BAIK']
y_new_pred = gnb.predict(new_data)
y_new_pred_proba = gnb.predict_proba(new_data)

print('Tempratur Udara =', new_data[0][0], 'derajat')
print('Kelembapan Udara =', new_data[0][1], '%')
print('Lama Penyinaran perhari =', new_data[0][2], 'jam')
print('Curah Hujan =', new_data[0][3], 'mm/hari')
print('Prediksi Peluang Pertumbuhan: ')
print(result_classes[0], 'dengan probabilitas', y_new_pred_proba[0][0] * 100, '%')
print(result_classes[1], 'dengan probabilitas', y_new_pred_proba[0][1] * 100, '%')

```

Tempratur Udara = 24 derajat
Kelembapan Udara = 75 %
Lama Penyinaran perhari = 6 jam
Curah Hujan = 4 mm/hari
Prediksi Peluang Pertumbuhan:
BURUK dengan probabilitas 7.438889547004312 %
BAIK dengan probabilitas 92.56191045299568 %

Gambar 4 Contoh Hasil Percobaan Uji Prediksi Naive Bayes

5. KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan, maka didapat bahwa prediksi perubahan iklim mendatang terhadap pertumbuhan tanaman dapat dimungkinkan dengan memiliki nilai acuan cakupan nilai iklim yang ter-orientasi pada suatu tanaman, dalam penelitian ini digunakan data iklim jeruk keprok yang menjadi acuan. Dengan memiliki acuan, maka dapat diketahui probabilitas baik atau buruknya sebuah nilai iklim terhadap tanaman jeruk keprok, sehingga dapat diketahui peluang pertumbuhan baik maupun buruk suatu tanaman berdasarkan iklim di daerah dimana dataset didapat yaitu daerah Malang.

Hasil probabilitas selanjutnya digunakan pada data yang masih belum diketahui kelasnya, sehingga didapat model Naïve Bayes yang dapat menangani data iklim yang disinkronisasikan dengan data iklim tanaman jeruk keprok, didapat Naïve Bayes dengan akurasi sebesar 85%, presisi sebesar 83%, recall sebesar 100%, dan *F1 Score* sebesar 90%, serta dapat memberikan keluaran prediksi iklim baik dengan contoh data mendatang yaitu tempratur

senilai 24°C, kelembapan senilai 75%, lama penyinaran yaitu 6 jam dan curah hujan 4mm/hari nya. Dengan skor Naïve Bayes tersebut diharapkan dapat mencukupi syarat upaya untuk menjaga stabilitas pertumbuhan tanaman terutama jeruk keprok yang ada pada daerah yang menjadi subjek penelitian.

6. DAFTAR PUSTAKA

- Ahsan, M. M., Mahmud, M., Saha, P. K., Gupta, K. D., & Siddique, Z. (2021). Effect if Data Scaling Methods on Machine Learning Algorithms and Model Performance. *MDPI Technologies Journal*, *X*(52), 1-17.
- Ali, A., Khairan, A., Tempola, F., & Fuad, A. (2021). Application of Naive Bayes to Predict the Potential of Rain in Ternate City. *International Conference on Science and Technology (ICST 2021)*, 328, 5.
- Ambarwari, A., Adrian, Q. J., & Herdiyeni, Y. (2020). Analisis Pengaruh Data Scaling Terhadap Performa Algoritma Machine Learning untuk Identifikasi Tanaman. *RESTI (Rekayasa Sistem dan Teknologi Informasi)*, *4*(1), 117-122.
- Ankit, M. (2018). Climatology-Definition, Nature and Scope: A Review. *International Journal of Research in Electronics and Computer Engineering*, *6*(4), 1118-1120.
- Balai Penelitian Tanaman Jeruk dan Subtropika. (2019, Januari 9). Retrieved Oktober 4, 2022, from <http://balitjestro.litbang.pertanian.go.id>
- BMKG. (n.d.). *Database Online Pusat*. Retrieved Oktober 1, 2022, from <https://dataonline.bmkg.go.id/>
- Düntsche, I., & Gediga, G. (2019). Confusion Matrices and Rough Set Data Analysis. *IOP Conferensation Series: Journal of Physics*, 1-6.
- Haokip, S. W., Shankar, K., & Lalrinnggheta, J. (2020). Climate Change and Its Impact on Fruit Crops. *Journal of Pharmacognosy and Phytochemistry*, *9*(1), 435-548.
- Heksaputra, D., Naimah, Z., Azani, Y., & Iswari, L. (2013). Penentuan Pengaruh Iklim Terhadap Pertumbuhan Tanaman dengan Naive Bayes. *Seminar Nasional Aplikasi Teknologi Informasi*, 34-39.
- Putri Br Sitepu, D., Fajriani, S., & Sulistyono, R. (2022). Impact of Climate Change On Productivity Citrus (Citrus Sp.). *Plantropica: Journal of Agricultural Science*, *7*(2), 18-23.
- Rahimah, E. N., Neni, N., & Rasmi, I. (2020). Budidaya Tanaman Jeruk Keprok Oleh Kelompok Tani Untuk Meningkatkan Kondisi Ekonomi di Desa Sindangsari Kecamatan Paseh Kabupaten Bandung. *Jurnal Geosarea*, *III*(2), 1-9.
- Saputro, I. W., & Sari, B. W. (2019). Uji Performa Algoritma Naive Bayes untuk Prediksi Masa Studi Siswa. *Citec Journal*, *VI*(1), 1-11.
- Sena, I. G., Dillak, J. W., Leunupun, P., & Santoso, A. J. (2020). Predicting Rainfall Intensity using Naive Bayes and Information Gain Methods (Case Study: Sleman Regency). *Journal of Physics: Conference Series*, 1-6.
- Zunaidi, M., Pane, U. F., & Nasyuha, A. H. (2021, Oktober). Analisis Teorema Bayes Dalam Mendiagnosa Penyakit Tanaman Pisang. *Jurnal Media Informatika Budidarma*, *5*(4), 1302-1308.