

Model Prediksi Multikomoditas (Padi, Jagung, dan Umbi-Umbian) Berbasis Faktor Cuaca Menggunakan Algoritma Naïve Bayes

Aurellia Ainun Nisa^{1,*}, M Riko Anshori Prasetya¹, Nurhaeni¹, Subandi²

¹ Fakultas Sains dan Teknologi, Program Studi Sistem Informasi, Universitas Sari Mulia, Banjarmasin, Indonesia

² Program Studi Sistem Informasi Kota Cerdas, Politeknik Negeri Banjarmasin, Banjarmasin, Indonesia

Email: ^{1,*}aurellnisa25@gmail.com, ¹riko.anshori@gmail.com, ²nurhaeni@unism.ac.id, ²subandi@poliban.ac.id

Email Penulis Korespondensi: aurellnisa25@email.com

Submitted: 27/09/2025; Accepted: 06/12/2025; Published: 08/12/2025

Abstrak—Sektor pertanian memiliki peran strategis dalam menopang ketahanan pangan dan ekonomi nasional Indonesia. Produktivitas pertanian di Provinsi Kalimantan Selatan, khususnya Kabupaten Barito Kuala, sangat dipengaruhi oleh dinamika iklim seperti perubahan suhu, curah hujan, dan kelembaban udara. Penelitian ini mengembangkan model prediksi hasil panen tanaman pangan (padi, jagung, dan umbi-umbian) berbasis faktor cuaca menggunakan algoritma *Naïve Bayes* sebagai sistem pendukung keputusan (*Decision Support System*) untuk mitigasi risiko gagal panen. Data penelitian diperoleh dari Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) serta Badan Pusat Statistik (BPS) Kabupaten Barito Kuala periode 2018–2023, mencakup variabel suhu udara, curah hujan, kelembaban, lama penyinaran, dan hasil panen. Tahapan penelitian meliputi *data preprocessing* (pembersihan, imputasi nilai hilang, augmentasi, dan pelabelan), pemodelan *machine learning*, serta evaluasi performa menggunakan metrik akurasi dan *weighted F1-score*. Hasil pengujian menunjukkan bahwa model *Naïve Bayes* mampu mengklasifikasikan kategori hasil panen (tinggi, sedang, rendah) dengan tingkat akurasi sebesar 90% dan nilai *weighted F1-score* sebesar 0,89. Nilai tersebut menunjukkan performa yang konsisten dan stabil pada berbagai variasi data cuaca. Keunggulan utama penelitian ini terletak pada integrasi data cuaca lokal dengan model *machine learning* ringan yang efisien dan mudah diimplementasikan pada sistem prediksi pertanian daerah. Hasil penelitian ini memberikan kontribusi nyata terhadap penguatan ketahanan pangan serta pengelolaan risiko pertanian berbasis data di wilayah tropis lembab seperti Kalimantan Selatan.

Kata Kunci: Pertanian; Prediksi Hasil Panen; Faktor Cuaca; Algoritma Naïve Bayes; Kalimantan Selatan

Abstract—The agricultural sector plays a strategic role in supporting Indonesia's food security and national economy. Agricultural productivity in South Kalimantan Province, particularly in Barito Kuala Regency, is strongly influenced by climatic dynamics such as temperature variation, rainfall, and humidity. This study develops a crop yield prediction model for major food commodities (rice, corn, and tubers) based on weather factors using the *Naïve Bayes* algorithm as a Decision Support System (DSS) to mitigate crop failure risks. The research data were obtained from the Meteorology, Climatology, and Geophysics Agency (BMKG) and the Central Bureau of Statistics (BPS) of Barito Kuala Regency for the period 2018–2023, covering temperature, rainfall, humidity, sunlight duration, and yield production. The research stages include data preprocessing (cleaning, missing value imputation, augmentation, and labeling), *machine learning* modeling, and performance evaluation using accuracy and *weighted F1-score* metrics. The experimental results show that the *Naïve Bayes* model can classify crop yield categories (high, medium, low) with an accuracy of 90% and a *weighted F1-score* of 0.89. These results demonstrate stable and consistent performance across various climatic conditions. The main advantage of this study lies in the integration of local weather data with a lightweight *machine learning* model that is computationally efficient and easily implemented in regional agricultural prediction systems. This research provides a tangible contribution to strengthening food security and data-driven agricultural risk management in tropical humid regions such as South Kalimantan.

Keywords: Agricultural; Crop Yield Prediction; Weather Factors; Naïve Bayes; South Kalimantan

1. PENDAHULUAN

Sektor pertanian merupakan salah satu pilar fundamental dalam struktur ekonomi nasional Indonesia yang memegang peran strategis tidak hanya dalam penyediaan bahan pangan bagi 270 juta lebih penduduk, tetapi juga sebagai penyerap tenaga kerja dalam jumlah besar [1]. Berdasarkan data Badan Pusat Statistik (BPS), sektor ini menjadi sumber mata pencaharian bagi sekitar 88,89 juta jiwa, atau hampir sepertiga dari total angkatan kerja Indonesia. Komoditas utama seperti padi, jagung, dan umbi-umbian tersebar luas di berbagai wilayah dan menjadi penopang ketahanan pangan nasional yang stabil [2].

Meskipun memiliki peran yang sangat vital, sektor pertanian nasional masih dihadapkan pada beragam tantangan struktural maupun non-struktural yang mempengaruhi produktivitas. Salah satu tantangan paling signifikan adalah tingginya kerentanan terhadap dinamika iklim. Fenomena perubahan iklim global yang semakin nyata, disertai kejadian cuaca ekstrem seperti kekeringan berkepanjangan, curah hujan berlebih, serta pergeseran pola musim, telah memberikan dampak langsung terhadap siklus tanam, kesuburan lahan, dan keberhasilan panen [3][4]. Dampak tersebut tidak hanya menurunkan hasil produksi, tetapi juga mengakibatkan kerugian ekonomi, menurunkan daya saing produk pertanian, serta mempengaruhi kesejahteraan petani [5].

Kerentanan ini semakin diperburuk oleh ketidakpastian cuaca yang tinggi [4]. Dalam konteks produksi pangan, ketidakpastian ini berimplikasi pada tingginya risiko gagal panen, yang pada gilirannya berpotensi mengganggu stabilitas pasokan dan harga pangan di tingkat nasional. Jika dibiarkan, kondisi ini dapat berujung pada ancaman serius terhadap ketahanan pangan nasional, terutama di tengah dinamika pasar global yang fluktuatif.

Di tingkat regional, Provinsi Kalimantan Selatan memiliki potensi lahan pertanian yang cukup luas, namun distribusi produktivitas antarwilayah menunjukkan disparitas yang cukup mencolok. Kabupaten Barito Kuala, yang

dikenal sebagai salah satu sentra pertanian di wilayah ini, justru mencatatkan hasil produksi yang jauh di bawah beberapa wilayah lain. Berdasarkan data BPS periode 2018–2023, produksi padi di Kabupaten Barito Kuala tercatat sebesar 1.346,32 ton, sedangkan Kota Banjarbaru yang memiliki luas lahan lebih kecil mampu mencapai 85.456,32 ton.

Perbedaan tersebut mengindikasikan adanya pengaruh signifikan dari faktor-faktor lokal, termasuk variabel iklim seperti curah hujan, suhu udara, dan kelembaban tanah. Faktor-faktor ini memainkan peran penting dalam menentukan produktivitas tanaman pangan, terutama padi yang sangat sensitif terhadap kondisi iklim. Sayangnya, hingga saat ini pemanfaatan data meteorologis secara sistematis untuk mendukung perencanaan pertanian di wilayah ini masih sangat terbatas. Padahal, integrasi data iklim ke dalam proses pengambilan keputusan berpotensi besar untuk meminimalkan risiko produksi akibat cuaca yang tidak menentu.

Perkembangan teknologi informasi dan analitik data telah membuka peluang baru bagi sektor pertanian dalam mengantisipasi risiko iklim dan meningkatkan produktivitas. Salah satu pendekatan yang menjanjikan adalah penerapan data mining untuk mengidentifikasi pola hubungan antara variabel iklim dan hasil panen.

Algoritma Naïve Bayes, yang merupakan metode klasifikasi berbasis probabilitas, telah lama digunakan secara luas di berbagai bidang untuk menangani dataset berskala besar dan kompleks [6]. Meskipun beberapa penelitian terdahulu menunjukkan bahwa algoritma lain seperti Random Forest atau Support Vector Machine (SVM) sering kali menghasilkan akurasi yang lebih tinggi, pemilihan Naïve Bayes dalam penelitian ini didasarkan pada pertimbangan metodologis dan praktis. Naïve Bayes memiliki kemampuan dalam menangani asumsi independensi fitur yang kuat suatu karakteristik yang umum ditemukan pada variabel iklim seperti suhu, curah hujan, dan kelembaban serta menawarkan kesederhanaan komputasi yang memungkinkan penerapan cepat dalam sistem pendukung keputusan (*decision support system*) di tingkat petani. Pendekatan ini penting terutama dalam konteks implementasi di daerah yang memiliki keterbatasan infrastruktur digital dan sumber daya komputasi.

Dengan mengintegrasikan data historis cuaca (suhu, curah hujan, kelembaban) dan produksi pertanian [7], model prediksi berbasis Naïve Bayes dapat dikembangkan sebagai sistem pendukung keputusan (*decision support system*) yang membantu petani dan pembuat kebijakan dalam merencanakan pola tanam secara lebih adaptif terhadap perubahan iklim [8][9].

Sejumlah penelitian terdahulu telah mengkaji potensi penerapan Naïve Bayes, Dalam penelitian yang dilakukan oleh Chandra Pharaba pada tahun 2020, pemanfaatan metode Naïve Bayes yang dikombinasikan dengan algoritma KNN dan SVM diterapkan untuk memprediksi hasil panen padi. Hasil penelitian tersebut menunjukkan bahwa akurasi tertinggi yang dicapai adalah sebesar 97,53% [10]. Selanjutnya pada penelitian Ginantra 2021 berhasil mengimplementasikan metode Naïve Bayes untuk memprediksi hasil panen kelapa sawit dengan tingkat akurasi mencapai 100%, yang mengindikasikan efektivitas metode ini dalam mengolah data pertanian [11]. Penelitian yang dilakukan oleh L. Febriyanti dan H. Zakaria (2023) juga melaporkan keberhasilan penerapan Naïve Bayes dalam memprediksi produktivitas kacang tanah berbasis variabel iklim [12]. Hasil serupa diperoleh oleh Heksaputra (2020) yang mengonfirmasi efektivitas metode ini dalam memprediksi hasil panen tanaman padi [13]. Sementara itu, A. Chusyairi (2023) memodelkan pertumbuhan jeruk keprok dengan tingkat akurasi prediksi sebesar 85%, menggunakan variabel perubahan iklim sebagai prediktor utama [14].

Meskipun memberikan hasil yang menjanjikan, mayoritas penelitian sebelumnya masih memiliki sejumlah keterbatasan yang membuka peluang untuk dilakukan penelitian lanjutan. Pertama, sebagian besar penelitian hanya berfokus pada satu jenis tanaman, sehingga model prediksi yang dihasilkan belum sepenuhnya teruji untuk berbagai komoditas pertanian lain yang memiliki karakteristik berbeda. Kedua, lokasi penelitian umumnya dilakukan di luar wilayah Kalimantan, padahal Kalimantan, khususnya Kalimantan Selatan, memiliki kondisi iklim tropis lembab, pola curah hujan, dan ekosistem yang unik sehingga memerlukan pendekatan yang lebih kontekstual. Dari beberapa kesimpulan tersebut, dapat dipahami bahwa belum banyak penelitian yang secara komprehensif mengintegrasikan hasil prediksi ke dalam kerangka perencanaan pertanian adaptif yang sesuai dengan karakteristik agroklimat di Kalimantan Selatan.

Oleh karena itu, penelitian ini berupaya mengisi kekosongan tersebut dengan menghadirkan model prediksi hasil panen berbasis faktor cuaca menggunakan algoritma Naïve Bayes, yang tidak hanya menekankan pada tingkat akurasi prediksi, tetapi juga diarahkan untuk mendukung pengambilan keputusan strategis dalam perencanaan waktu tanam serta pengelolaan risiko pertanian di wilayah Kalimantan Selatan. Berdasarkan latar belakang tersebut, penelitian ini memiliki tujuan utama untuk menganalisis pengaruh variabel iklim terhadap hasil panen tanaman pangan utama padi, jagung, dan umbi-umbian—di Provinsi Kalimantan Selatan. Fokus penelitian diarahkan pada penerapan algoritma Naïve Bayes sebagai model prediktif untuk mengestimasi potensi hasil panen berdasarkan pola data historis cuaca dan produksi yang tersedia. Dengan demikian, penelitian ini tidak hanya berupaya mengidentifikasi hubungan kausal maupun korelasional antara variabel iklim dan produktivitas pertanian, tetapi juga mengembangkan sistem prediksi berbasis data historis yang adaptif, akurat, serta dapat diimplementasikan secara praktis oleh petani maupun pengambil kebijakan sebagai alat bantu dalam perencanaan produksi, mitigasi risiko gagal panen, dan perumusan strategi ketahanan pangan di tingkat regional.

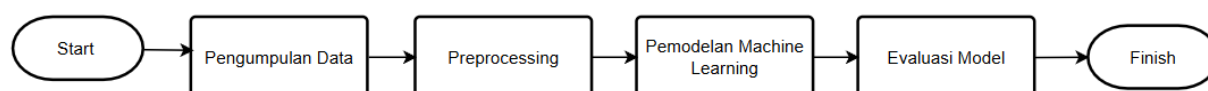
Kebaruan (*novelty*) dari penelitian ini terletak pada fokus wilayah kajian di Kalimantan Selatan yang hingga kini masih jarang mendapat perhatian dalam studi serupa, serta penggunaan pendekatan integratif antara teknik machine learning dan kebutuhan praktis perencanaan pertanian lokal. Harapannya, model prediksi yang dikembangkan dapat meningkatkan akurasi estimasi hasil panen, mengurangi risiko gagal panen akibat anomali cuaca,

dan pada akhirnya memperkuat ketahanan serta keberlanjutan sistem pertanian di tengah tantangan perubahan iklim di Indonesia.

2. METODOLOGI PENELITIAN

Mengacu studi-studi yang telah menggunakan model machine learning, metode penelitian ini terdiri dari beberapa tahapan utama, dimulai dari pengumpulan data cuaca dan data hasil panen tanaman pangan Kabupaten Barito Kuala yang mencakup Suhu, Kelembaban, Lama sinar matahari, Curah hujan, Jenis tanaman, Hasil panen. Terdapat beberapa tahapan Preprocessing Data meliputi menghapus karakter khusus pada data, imputasi nilai hilang, Data augmentasi, Pelabelan data, kemudian dilakukan Pemodelan Machine Learning, dan Evaluasi menggunakan algoritma naïve bayes. Pada tahap evaluasi, menggunakan metric accuracy untuk mengetahui seberapa berpengaruh cuaca terhadap hasil panen tanaman pangan.

Untuk memberikan gambaran yang lebih komprehensif mengenai proses penelitian yang dilakukan, berikut ini disajikan alur metodologi penelitian secara visual. Gambar 1 menjelaskan empat tahapan utama yang menjadi kerangka kerja dalam pengembangan mode, dimulai dari proses Pengumpulan Data, Preprocessing Data, Pemodelan Machine Learning, Evaluasi Model.



Gambar 1. Kerangka Kerja

2.1 Pengumpulan Data

Dataset ini mencerminkan kondisi nyata dalam data cuaca dan hasil panen tanaman pangan di Kalsel. Penelitian ini menggunakan dataset dalam format file excel yang diperoleh dari situs bmkg.go.id dan BPS Kabupaten Barito Kuala. Dataset ini mencakup informasi penting terkait Suhu, Kelembaban, Curah hujan, Lama sinar matahari, Jenis tanaman, Hasil panen.

2.2 Preprocessing

Pada tahap preprocessing, data yang telah dikumpulkan akan diproses untuk dipersiapkan ke tahap pemodelan algoritma naïve bayes. Hal ini bertujuan untuk memastikan bahwa data yang digunakan dalam pemodelan memiliki kualitas dan konsistensi yang memadai sehingga dapat menghasilkan prediksi yang akurat dan andal. Proses ini mencakup serangkaian langkah sistematis untuk mengidentifikasi, mengoreksi, dan menghapus ketidaksesuaian maupun ketidaklengkapan pada data mentah, baik yang bersumber dari pengukuran variabel iklim (suhu, curah hujan, kelembaban) maupun data historis hasil panen [15].

2.2.1 Data Cleaning

Tahap Cleaning yaitu tahap pertama dalam proses preprocessing, pada tahap ini digunakan untuk menghapus baris, menghapus spasi dan karakter khusus lainnya. Proses cleaning menggunakan beberapa langkah untuk membersihkan data dari elemen-elemen yang tidak diinginkan. menghapus baris, spasi, dan menghapus karakter khusus [16].

2.2.2 Imputasi Data

Imputasi data merupakan tahap kedua dalam preprocessing pada tahap ini dilakukan pengisian nilai yang hilang (missing values) dalam suatu dataset dengan nilai pengganti memungkinkan analisis data yang lebih akurat dan komprehensif. Proses ini sangat penting karena data yang hilang dapat mengganggu analisis [17].

2.2.3 Data Augmentasi

Data augmentation merupakan tahap ketiga dalam proses preprocessing. Pada tahap ini, kategori curah hujan dan kelembaban diubah ke dalam tiga tingkatan, yaitu Rendah, Normal, dan Tinggi. Selanjutnya, dilakukan proses augmentasi dengan teknik oversampling untuk menangani masalah ketidakseimbangan kelas (class imbalance) pada dataset dengan memperbanyak sampel dari kelas minoritas. Oversampling bertujuan memperbaiki representasi kelas minoritas dan mengurangi bias pada model, namun metode ini *dapat meningkatkan risiko overfitting* jika dilakukan dengan penggunaan sederhana. Oleh karena itu, disarankan untuk menerapkan teknik augmentasi yang lebih canggih (misalnya SMOTE atau varian berbasis sintesis), serta mengombinasikannya dengan strategi mitigasi seperti validasi silang (cross-validation), regularisasi, atau prosedur undersampling pada kelas mayoritas untuk menjaga generalisasi model [18].

2.2.4 Labelling

Labelling merupakan tahap keempat dalam preprocessing data yang berfungsi memberikan label pada data curah hujan. Tahap ini krusial dalam mentransformasi data mentah menjadi informasi terstruktur agar dapat dipahami dan

diproses lebih lanjut oleh model machine learning. Dalam konteks machine learning, label menjadi komponen esensial terutama pada tugas klasifikasi dan prediksi, karena tanpa label model tidak memiliki acuan untuk mempelajari pola data. Dengan demikian, labelling bukan sekadar pelengkap, melainkan fondasi penting yang menentukan efektivitas dan akurasi model dalam menghasilkan output [19].

2.3 Pemodelan Machine Learning

Setelah tahap *preprocessing* data selesai dilakukan, langkah selanjutnya adalah membangun model prediksi hasil panen menggunakan algoritma Naïve Bayes. Pemilihan algoritma ini didasarkan pada sejumlah keunggulan, yaitu: kemudahan perhitungan, efisiensi waktu, serta performa baik pada data dengan jumlah atribut relatif banyak. Keunggulan lainnya adalah kapabilitasnya dalam menangani dataset besar serta menyediakan interpretasi probabilistik yang mudah dipahami dalam konteks prediksi berbasis faktor cuaca [20] [21].

Secara matematis, Naïve Bayes menggunakan Teorema Bayes dengan asumsi independensi antarvariabel prediktor. Walaupun variabel cuaca seperti curah hujan, kelembaban, suhu udara, dan intensitas cahaya matahari memiliki potensi keterkaitan, asumsi independensi ini tetap memungkinkan pembentukan model prediktif dengan akurasi memuaskan. Penggabungan antara probabilitas awal (*prior probability*) dan bukti baru (*likelihood*) menghasilkan probabilitas akhir (*posterior probability*), yang menyediakan output probabilistik, bukan deterministic.

Penelitian sebelumnya telah menunjukkan efektivitas Naïve Bayes dalam memprediksi hasil panen berbasis data iklim. Misalnya, [22] menerapkan Naïve Bayes untuk memprediksi hasil panen tanaman pangan menggunakan data suhu, kelembaban, curah hujan, dan fotoperiodisitas—dengan akurasi sekitar 72,78 % dan RMSE sebesar 0,438. Selain itu, [23] mengaplikasikan Naïve Bayes untuk memprediksi hasil panen jagung, cabai, dan tomat berdasarkan data suhu, kelembaban, dan curah hujan; klasifikasi hasil panen dikategorikan sebagai rendah, sedang, dan tinggi.

Dalam penelitian ini, variabel cuaca yang digunakan sebagai atribut prediktor mencakup curah hujan, rata-rata suhu udara, tingkat kelembaban, serta intensitas cahaya matahari, sementara variabel target berupa kategori hasil panen (tinggi, sedang, rendah). Model Naïve Bayes dilatih menggunakan dataset historis hasil panen tanaman pangan di Kalimantan Selatan yang telah diselaraskan dengan data iklim pada periode yang sama. Dengan pendekatan ini, algoritma diharapkan menghasilkan sistem prediksi hasil panen berbasis faktor cuaca yang efektif, mendukung pengambilan keputusan strategis di sektor pertanian lokal.

Aspek “Naïve” berasal dari asumsi bahwa setiap atribut dianggap independen. Meskipun kenyataannya sering terjadi korelasi antar variabel, asumsi ini menyederhanakan komputasi probabilitas bersyarat sehingga algoritma tetap efisien meskipun jumlah atribut besar. Sifatnya yang ringan secara komputasi menjadikan Naïve Bayes cepat serta mampu memproses dataset besar tanpa mengorbankan kualitas prediksi.

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

Rumus tersebut menggambarkan hubungan antara dua kejadian, yaitu H dan X, dalam konteks teori probabilitas bersyarat. Simbol H dan X masing-masing merepresentasikan suatu kejadian yang dapat saling memengaruhi. Nilai $P(H|X)$ menunjukkan peluang terjadinya H jika X telah terjadi, sedangkan $P(X|H)$ menyatakan peluang X terjadi apabila H benar. Adapun $P(H)$ dan $P(X)$ merupakan probabilitas bebas yang tidak bergantung satu sama lain. Hubungan ini menjadi dasar dari Teorema Bayes, yang digunakan untuk memperbarui probabilitas suatu hipotesis (H) berdasarkan bukti atau data pengamatan baru (X). Pendekatan ini banyak digunakan dalam berbagai bidang seperti sistem pakar, prediksi, serta klasifikasi berbasis pembelajaran mesin.

2.4 Evaluasi Model

Evaluasi model merupakan tahapan yang sangat krusial dalam penelitian berbasis prediksi, karena melalui tahap ini dapat diketahui sejauh mana tingkat kinerja algoritma yang digunakan dalam menghasilkan prediksi yang akurat, konsisten, serta dapat diandalkan. Proses evaluasi tidak hanya bertujuan untuk menilai kemampuan model pada data latih, tetapi juga untuk memastikan bahwa model tetap memiliki performa yang stabil ketika diuji pada data baru yang sebelumnya tidak dikenali. Dengan demikian, evaluasi berfungsi sebagai instrumen validasi terhadap kualitas data, metode, serta kehandalan model yang dibangun dalam sistem prediksi [24].

Dalam konteks penelitian ini, yang berfokus pada prediksi hasil panen tanaman pangan berbasis faktor cuaca menggunakan algoritma Naïve Bayes di Kalimantan Selatan, tahap evaluasi memiliki signifikansi yang lebih besar. Hal ini disebabkan karena hasil panen merupakan indikator penting bagi ketahanan pangan, sehingga kesalahan prediksi berpotensi memberikan dampak pada proses perencanaan produksi, distribusi, hingga kebijakan pertanian di tingkat daerah. Oleh sebab itu, evaluasi dilakukan dengan tujuan menilai tingkat kesesuaian antara hasil prediksi sistem dengan data aktual hasil panen yang diperoleh dari catatan historis. Semakin tinggi tingkat kesesuaian prediksi, semakin tinggi pula tingkat kepercayaan yang dapat diberikan terhadap model dalam mendukung pengambilan keputusan strategis di sektor pertanian.

Proses evaluasi dalam penelitian ini menggunakan metrik akurasi (*accuracy*), yaitu rasio antara jumlah prediksi yang benar dengan jumlah keseluruhan data uji yang digunakan. Metrik ini dipilih karena mampu memberikan gambaran yang sederhana namun representatif mengenai kemampuan algoritma Naïve Bayes dalam mengklasifikasikan kategori hasil panen ke dalam tiga kelas, yaitu *tinggi*, *sedang*, dan *rendah*, berdasarkan variabel cuaca seperti curah hujan, suhu udara, kelembaban, serta intensitas cahaya matahari. Melalui pendekatan evaluasi ini,

penelitian diharapkan dapat memberikan kontribusi nyata dalam pengembangan sistem prediksi pertanian yang adaptif terhadap dinamika iklim di Kalimantan Selatan. Secara teknis, prosedur evaluasi model dilakukan melalui tahapan sebagai berikut:

- Membagi dataset menjadi data latih (training set) dan data uji (testing set) dengan perbandingan 80:20.
- Melatih model menggunakan data latih untuk menghitung probabilitas bersyarat dari setiap atribut cuaca terhadap kategori hasil panen.
- Menguji model pada data uji untuk menilai kinerja algoritma dalam menghasilkan prediksi.
- Mengevaluasi performa model menggunakan metrik akurasi, guna mengukur sejauh mana model mampu memberikan hasil prediksi yang konsisten dan dapat diandalkan.

2.4 Accuracy

Akurasi merupakan salah satu indikator utama dalam evaluasi kinerja model klasifikasi yang digunakan untuk menilai sejauh mana ketepatan hasil prediksi model terhadap data aktual yang dimiliki. Metrik ini sering dipilih karena mampu memberikan gambaran sederhana namun cukup representatif mengenai performa keseluruhan dari sebuah algoritma klasifikasi. Secara teknis, akurasi dihitung dengan membandingkan jumlah prediksi yang dikategorikan benar oleh model dengan total keseluruhan data uji yang digunakan dalam proses pengujian. Semakin besar proporsi prediksi yang benar, maka semakin tinggi pula nilai akurasi yang dihasilkan oleh model tersebut.

Dalam konteks penelitian ini, akurasi memiliki peranan penting karena berfungsi sebagai tolok ukur untuk menilai kemampuan algoritma Naïve Bayes dalam melakukan prediksi hasil panen tanaman pangan berbasis faktor cuaca di Kalimantan Selatan. Faktor cuaca yang digunakan sebagai variabel prediktor, seperti curah hujan, kelembaban, suhu udara, dan intensitas cahaya matahari, memiliki pengaruh langsung terhadap tingkat produktivitas pertanian. Oleh karena itu, akurasi tidak hanya menggambarkan ketepatan komputasional, tetapi juga mencerminkan relevansi model dalam merepresentasikan fenomena nyata yang terjadi di lapangan.

Nilai akurasi yang tinggi menunjukkan bahwa algoritma Naïve Bayes mampu mengidentifikasi pola-pola yang signifikan antara faktor cuaca dan kategori hasil panen, sehingga prediksi yang dihasilkan dapat mendekati kondisi aktual. Hal ini penting terutama bagi para pengambil keputusan di sektor pertanian, karena prediksi yang akurat dapat menjadi dasar dalam merumuskan strategi perencanaan tanam, pengelolaan sumber daya, hingga mitigasi risiko gagal panen akibat kondisi cuaca ekstrem. Dengan demikian, akurasi bukan hanya sekadar angka statistik, melainkan indikator kunci yang menentukan seberapa andal model dapat digunakan dalam mendukung sistem prediksi hasil panen berbasis data di wilayah Kalimantan Selatan. Perhitungan akurasi dapat diformulasikan sebagai berikut:

$$\text{Akurasi} = \frac{\text{Jumlah Prediksi Benar}}{\text{Jumlah Seluruh Data Uji}} \times 100 \quad (2)$$

Nilai akurasi menggambarkan seberapa baik model mampu mengenali pola data dengan tepat. Semakin tinggi nilai akurasi, maka semakin baik kinerja model dalam melakukan klasifikasi atau prediksi terhadap data baru.

Penggunaan akurasi sebagai ukuran kinerja pada penelitian ini penting karena memberikan gambaran kuantitatif mengenai ketepatan model Naïve Bayes dalam mengklasifikasikan hasil panen berdasarkan variabel cuaca. Melalui akurasi, efektivitas model dapat dinilai secara objektif terhadap data historis, sehingga hasil prediksi yang konsisten dapat mendukung pengambilan keputusan di sektor pertanian. Akurasi yang tinggi menunjukkan kemampuan model dalam meminimalisasi kesalahan klasifikasi, sekaligus berperan dalam mitigasi risiko kegagalan panen akibat variabilitas cuaca di Kalimantan Selatan. Dengan demikian, akurasi tidak hanya menjadi tolok ukur keberhasilan teknis, tetapi juga landasan dalam penyusunan strategi pertanian yang adaptif dan berkelanjutan.

Namun, penggunaan akurasi tidak selalu cukup dalam kondisi data yang tidak seimbang (imbalanced), seperti pada kasus penelitian ini yang melibatkan klasifikasi tiga kelas hasil panen: *Tinggi*, *Sedang*, dan *Rendah*. Pada situasi seperti ini, model dapat menghasilkan nilai akurasi tinggi meskipun gagal memprediksi kelas minoritas (misalnya kategori *Panen Rendah*). Untuk mengatasi keterbatasan tersebut, penelitian ini juga menggunakan metrik evaluasi tambahan, yaitu F1-Score.

F1-Score merupakan harmonisasi antara *Precision* dan *Recall*, yang memberikan penilaian lebih seimbang terutama ketika distribusi kelas tidak merata. Perhitungan metrik tambahan dilakukan menggunakan formula umum berikut:

$$F1_{\text{Weighted}} = \frac{\sum_{i=1}^n (\text{support}_i \times F1_i)}{\sum_{i=1}^n \text{support}_i} \quad (3)$$

Rumus ini menghasilkan nilai F1-Score keseluruhan yang memperhitungkan proporsi data di setiap kelas, sehingga lebih representatif untuk model yang menangani data tidak seimbang. Semakin tinggi nilai F1 Weighted, maka semakin baik pula kemampuan model dalam memberikan prediksi yang konsisten dan seimbang di semua kelas. Oleh karena itu, metrik ini sering digunakan bersamaan dengan akurasi untuk memberikan gambaran yang lebih komprehensif terhadap performa model klasifikasi, seperti pada algoritma Naïve Bayes dalam penelitian prediksi hasil panen berbasis data cuaca.

Dengan penerapan metrik akurasi dan *weighted F1-score*, evaluasi kinerja model menjadi lebih komprehensif dan mampu menggambarkan sejauh mana algoritma Naïve Bayes dapat melakukan klasifikasi hasil panen secara adil pada kondisi data yang tidak seimbang.

3. HASIL DAN PEMBAHASAN

Bagian hasil dan pembahasan ini menyajikan temuan penelitian yang diperoleh dari proses pengolahan dan analisis data, sekaligus menguraikan interpretasi terhadap hasil tersebut. Penyajian hasil tidak hanya menampilkan keluaran model, tetapi juga dihubungkan dengan teori, penelitian terdahulu, serta kondisi nyata di lapangan. Dengan demikian, pembahasan yang disusun diharapkan mampu memberikan pemahaman komprehensif mengenai hubungan antara faktor cuaca dengan hasil panen tanaman pangan di Kabupaten Barito Kuala.

Analisis dilakukan dengan memanfaatkan data sekunder yang telah melalui beberapa tahapan agar sesuai untuk diolah menggunakan metode *machine learning*. Selanjutnya, hasil analisis dibandingkan dengan literatur yang relevan untuk memastikan validitas temuan dan mengidentifikasi kontribusi penelitian ini terhadap pengembangan pengetahuan di bidang pertanian berbasis data.

3.1 Hasil

Tabel 1. Data Cleaning

Tanggal	Tn	Tx	Tavg	RH avg	RR	ss	ff x	ddd x	ff avg	ddd car
1/11/2018	25	32.2	26.8	84	3.2	6.4	4	210	2	N
2/11/2018	24	35.2	28	81	3.8	1.4	4		2	E
3/11/2018	24	35.2	27.9	80	10	6	5	230	1	C
4/11/2018	24	35.2	28.7	73		8	5	250	2	S
5/11/2018	24	36.2	29.2	70	0.2	9	5	120	2	SE

Tabel 1 dari total 3.100 entri data yang terkumpul pada tahap awal, hanya 2.529 data yang dinyatakan valid dan dapat digunakan untuk tahap analisis berikutnya. Hal ini mengindikasikan bahwa terdapat 571 entri data yang tereliminasi akibat mengandung ketidaksesuaian format, duplikasi, maupun inkonsistensi nilai. Proses ini menjadi penting karena keberadaan data yang tidak valid dapat memengaruhi performa algoritma Naïve Bayes, sehingga pembersihan data menjadi prasyarat untuk memperoleh hasil prediksi yang lebih akurat dan reliabel.

Tabel 2. Data Imputasi

Tanggal	Tn	Tx	Tavg	RH avg	RR	ss	ff x	ddd x	ff avg	ddd car
1/11/2018	25	32.2	26.8	84	3.2	6.4	4	210	2	N
2/11/2018	24	35.2	28	81	3.8	1.4	4	260	2	E
3/11/2018	24	35.2	27.9	80	10	6	5	230	1	C
4/11/2018	24	35.2	28.7	73	11	8	5	250	2	S
5/11/2018	24	36.2	29.2	70	0.2	9	5	120	2	SE

Tabel 2 proses imputasi dilakukan dengan cara mengisi nilai kosong pada variabel tertentu menggunakan metode penggantian nilai yang dianggap relevan. Hasil dari tahapan ini menunjukkan adanya perubahan jumlah data, di mana dataset yang semula berjumlah 2.529 entri bertambah menjadi 2.620 entri setelah proses imputasi selesai dilakukan.

Tabel 3. Data Augmentasi

Tanggal	Tavg (Temperatur)	RH avg (Kelembaban)	RR (Curah Hujan)	ss (Lama sinar matahari)
01-01-2018	Normal	Tinggi	Rendah	Rendah
02-01-2018	Normal	Tinggi	Rendah	Tinggi
03-01-2018	Normal	Tinggi	Normal	Rendah
04-01-2018	Normal	Tinggi	Rendah	Rendah
05-01-2018	Normal	Tinggi	Normal	Normal

Tabel 3 Proses kategorisasi ini bertujuan untuk menyederhanakan data kontinu menjadi data diskrit sehingga lebih sesuai dengan kebutuhan pemodelan menggunakan algoritma Naïve Bayes. Dengan adanya pengelompokan ke dalam tiga kategori utama, variabel cuaca yang semula memiliki rentang nilai yang beragam dapat direpresentasikan secara lebih sederhana tanpa kehilangan informasi penting yang relevan terhadap hasil panen. Selain itu, pendekatan ini juga memudahkan interpretasi hasil, khususnya dalam menjelaskan hubungan antara kondisi cuaca dengan tingkat produktivitas tanaman pangan.

3.1.1 Penerapan Naïve Bayes

Pada uji coba pertama (tanpa augmentasi), hasil menunjukkan bahwa akurasi model mencapai 73% dengan pembagian data *training* dan *testing* sebesar 80:20, serta 72% dengan pembagian 70:30. Nilai akurasi ini masih tergolong rendah karena distribusi data antar kelas hasil panen (*Tinggi*, *Sedang*, *Rendah*) tidak seimbang, menyebabkan model lebih sering memprediksi kelas mayoritas (*bias toward majority class*).

Sebagai langkah perbaikan, diterapkan teknik augmentasi data menggunakan Synthetic Minority Oversampling Technique (SMOTE) untuk mengatasi masalah *imbalanced data*. SMOTE bekerja dengan

membangkitkan data sintesis pada kelas minoritas melalui interpolasi antara sampel yang berdekatan, sehingga distribusi antar kelas menjadi lebih seimbang tanpa menambah kelas baru secara acak. Setelah augmentasi, jumlah data meningkat dari 2.529 menjadi 3.024 entri, dan proporsi antar kelas menjadi lebih merata.

Hasil pengujian setelah penerapan SMOTE menunjukkan peningkatan performa model yang signifikan. Dengan pembagian data 80:20, model menghasilkan:

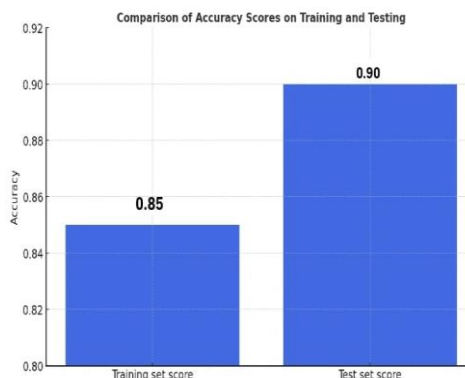
```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import OrdinalEncoder
from sklearn.naive_bayes import CategoricalNB
from sklearn.metrics import accuracy_score, f1_score, classification_report
```

Gambar 2 Pemodelan Naïve Bayes

Gambar 2 kode digunakan untuk membangun sistem prediksi hasil panen tanaman pangan berbasis faktor cuaca dengan tahapan:

- Mengimpor dan menyiapkan data dari sumber eksternal (BMKG dan BPS) menggunakan *Pandas* dan *NumPy*.
- Melakukan preprocessing (konversi kategori dan pembagian data) agar data siap digunakan oleh model *Naïve Bayes*.
- Melatih model dan mengevaluasi performanya menggunakan metrik seperti *accuracy* dan *F1-score*.

Dalam konteks penelitian pertanian di Kalimantan Selatan, kode ini berperan penting dalam mendukung sistem prediksi hasil panen berbasis cuaca secara ilmiah, efisien, dan adaptif terhadap variabilitas iklim.



Gambar 3 Hasil Akurasi

Gambar 3 memperlihatkan perbandingan tingkat akurasi antara data latih (*training set*) dan data uji (*testing set*) pada model klasifikasi hasil panen menggunakan algoritma Naïve Bayes. Berdasarkan hasil pengujian, diperoleh nilai akurasi sebesar 0,85 (85%) pada data latih dan 0,90 (90%) pada data uji.

Tabel 4 Accuray & F1-Score

Evaluation Matrix	Score
Accuracy	90%
F1-score	0,89%

Tabel 4 menunjukkan bahwa model Naïve Bayes memperoleh akurasi sebesar 90% dan F1-score sebesar 0,89. Nilai akurasi yang tinggi mencerminkan kemampuan model dalam mengklasifikasikan hasil panen berdasarkan faktor cuaca dengan tingkat kesalahan yang rendah. Sementara itu, nilai F1-score yang mendekati akurasi menunjukkan keseimbangan antara *precision* dan *recall*, menandakan bahwa model mampu mengenali baik kelas mayoritas maupun minoritas secara proporsional. Hasil ini mengindikasikan bahwa algoritma Naïve Bayes memiliki performa prediktif yang stabil dan layak digunakan dalam sistem pendukung keputusan pertanian berbasis data cuaca di Kalimantan Selatan.

Nilai Weighted F1-Score memberikan ukuran yang lebih representatif terhadap kinerja model karena mempertimbangkan ketidakseimbangan jumlah data di setiap kelas. F1-Score dihitung berdasarkan rata-rata harmonis antara Precision dan Recall untuk setiap kelas, kemudian dibobot berdasarkan proporsi jumlah sampel pada masing-masing kelas (*support*).

Dengan demikian, hasil uji coba menegaskan bahwa penerapan SMOTE secara efektif memperbaiki distribusi kelas dan meningkatkan tidak hanya akurasi, tetapi juga stabilitas prediksi antar kelas. Hal ini menunjukkan bahwa algoritma Naïve Bayes memiliki kemampuan generalisasi yang baik ketika data latih berada dalam distribusi yang seimbang, menjadikannya metode yang efisien untuk prediksi hasil panen berbasis faktor cuaca di wilayah Kalimantan Selatan.

3.1.2 Evaluasi

Berdasarkan hasil evaluasi kinerja model prediksi menggunakan algoritma Naïve Bayes, diperoleh tingkat akurasi sebesar 90% dan hasil F1-Score sebesar 0.89. Capaian ini menunjukkan bahwa model memiliki performa yang baik dalam melakukan prediksi hasil panen tanaman pangan. Tingkat akurasi tersebut mencerminkan kemampuan model dalam menggeneralisasi pola hubungan antara variabel cuaca—seperti curah hujan, suhu udara, dan kelembaban dengan produktivitas pertanian. Dengan demikian, dapat disimpulkan bahwa algoritma Naïve Bayes memiliki potensi signifikan dalam mendukung proses prediksi hasil pertanian berbasis data meteorologi.

Hasil evaluasi ini juga mengindikasikan bahwa informasi cuaca dapat dimanfaatkan secara efektif untuk memperkirakan hasil panen. Pemanfaatan data meteorologi sebagai variabel prediktor terbukti mampu meningkatkan efisiensi dalam pengambilan keputusan, misalnya terkait perencanaan musim tanam, pengelolaan sumber daya, serta strategi mitigasi risiko gagal panen akibat anomali iklim. Pencapaian akurasi 90% menegaskan bahwa pendekatan ini dapat menjadi salah satu solusi berbasis teknologi yang relevan untuk mendukung sektor pertanian, khususnya dalam menghadapi tantangan ketidakpastian iklim di masa depan.

Meskipun demikian, penelitian ini masih memiliki keterbatasan, terutama karena hanya memanfaatkan faktor cuaca sebagai variabel utama dalam pemodelan. Faktor lain seperti kesuburan tanah, jenis varietas tanaman, penggunaan pupuk, serta praktik budidaya pertanian juga berpotensi memberikan kontribusi terhadap peningkatan akurasi prediksi. Oleh karena itu, penelitian lebih lanjut disarankan untuk mengintegrasikan variabel-variabel tersebut agar model prediksi dapat menghasilkan estimasi yang lebih komprehensif dan akurat.

3.2 Pembahasan

Penelitian ini menggunakan data hasil panen empat komoditas utama padi, jagung, ubi jalar, dan ubi kayu serta data cuaca dari BPS dan BMKG periode 2018–2023. Setelah dilakukan SMOTE untuk menyeimbangkan distribusi kelas, jumlah data meningkat dari 2.529 menjadi 3.024 record. Teknik ini memperbaiki kinerja model dengan mengurangi bias terhadap kelas mayoritas.

Model Naïve Bayes menunjukkan performa tinggi dengan akurasi 90% dan Weighted F1-Score 0.89, menunjukkan keseimbangan yang baik antara *precision* dan *recall*. Analisis faktor menunjukkan curah hujan sebagai variabel paling dominan dalam menentukan hasil panen, khususnya di Kabupaten Barito Kuala. Dominasi ini dipengaruhi oleh karakteristik lahan polder dan sawah pasang surut, di mana kelebihan atau kekurangan air akibat variabilitas curah hujan secara langsung memengaruhi produktivitas tanaman.

Meskipun variabel cuaca seperti suhu, kelembaban, dan penyinaran saling berkorelasi, Naïve Bayes tetap efektif karena pendekatan probabilistiknya mampu memanfaatkan pola distribusi antar variabel tanpa memerlukan asumsi independensi yang ketat. Hal ini memungkinkan model menghasilkan prediksi stabil dan akurat pada kondisi lingkungan yang kompleks.

Secara praktis, model ini dapat mendukung pengambilan keputusan pertanian berbasis data cuaca, seperti penentuan waktu tanam, antisipasi gagal panen, dan perencanaan ketahanan pangan di wilayah pasang surut Kalimantan Selatan. Temuan ini menegaskan relevansi metode probabilistik sederhana dalam mengolah data agroklimat yang dinamis dan tidak selalu seimbang.

3.3 Hasil analisis

Curah hujan terbukti menjadi faktor iklim paling dominan dalam memengaruhi hasil panen di Kalimantan Selatan. Kekurangan air pada fase kritis menurunkan jumlah anakan dan menghambat pengisian biji, sedangkan curah hujan berlebih meningkatkan risiko genangan dan serangan hama. Variabilitas tahunan juga memperbesar ketidakpastian produksi, sehingga curah hujan perlu diposisikan sebagai faktor strategis dalam manajemen pertanian dan adaptasi perubahan iklim.

Suhu udara dan kelembaban memberikan pengaruh tambahan terhadap produktivitas. Suhu optimal 25–30°C mendukung fotosintesis, sedangkan suhu ekstrem menimbulkan stres fisiologis. Kelembaban tinggi berpotensi meningkatkan prevalensi penyakit, sejalan dengan temuan terdahulu. Oleh karena itu, variabel suhu dan kelembaban tetap relevan dalam integrasi model prediksi berbasis Naïve Bayes.

Lama penyinaran matahari meski tidak dominan, berperan dalam menunjang fotosintesis dan pertumbuhan vegetatif. Variasi antar musim dapat memengaruhi stabilitas produksi, sehingga variabel ini tetap penting diperhatikan sebagai faktor pelengkap dalam analisis produktivitas jangka panjang.

Tantangan muncul pada klasifikasi hasil panen kategori sedang, di mana data sering tumpang tindih (*overlapping distribution*). Kondisi ini menurunkan akurasi Naïve Bayes karena asumsi independensi variabel sulit dipenuhi. Pendekatan hibrid, misalnya mengkombinasikan Naïve Bayes dengan ensemble learning, berpotensi meningkatkan akurasi klasifikasi dan konsistensi prediksi.

Secara praktis, hasil penelitian ini bermanfaat bagi petani dan pemerintah daerah. Prediksi hasil panen dapat digunakan untuk menentukan waktu tanam, mengantisipasi risiko gagal panen, serta mendukung sistem peringatan dini. Bagi pemangku kebijakan, model ini dapat menjadi dasar pemetaan potensi produksi, distribusi pangan, dan mitigasi dampak perubahan iklim. Temuan ini menguatkan efektivitas Naïve Bayes dalam mengolah data pertanian yang kompleks, sekaligus membuka peluang implementasi nyata dalam memperkuat ketahanan pangan daerah.

3.4 Rekomendasi

Berdasarkan hasil analisis yang telah dilakukan, didapatkan sejumlah rekomendasi strategis yang bertujuan untuk meningkatkan produktivitas tanaman pangan di Kalimantan Selatan sebagai berikut:

- a. **Memanfaatkan Informasi Cuaca Secara Berkala**
Petani disarankan untuk secara rutin memantau prakiraan cuaca dari BMKG atau aplikasi cuaca untuk merencanakan kegiatan bertani seperti penanaman, pemupukan, dan panen. Dikarenakan cuaca ekstrem seperti hujan lebat atau suhu tinggi yang berkepanjangan terbukti berdampak negatif terhadap hasil panen.
- b. **Menyesuaikan Waktu Tanam dengan Pola Cuaca**
Berdasarkan hasil analisis, penyesuaian waktu tanam yang selaras dengan periode cuaca yang relatif stabil cenderung menghasilkan panen yang lebih optimal. Strategi berupa percepatan atau penundaan waktu tanam dapat diterapkan sebagai upaya mitigasi terhadap kondisi cuaca yang kurang mendukung. Dengan demikian, penyesuaian waktu tanam yang tepat berpotensi meningkatkan produktivitas hasil panen secara signifikan.
- c. **Peningkatan Edukasi dan Literasi Data Pertanian**
Kolaborasi antara petani, dan Dinas pertanian Kabupaten Barito Kuala untuk melakukan pelatihan atau penyuluhan tentang pentingnya data cuaca dan cara memanfaatkannya dalam kegiatan pertanian sangat disarankan.
- d. **Diversifikasi Tanaman Berdasarkan Cuaca**
Petani disarankan untuk mempertimbangkan alternatif jenis tanaman pangan yang memiliki toleransi lebih tinggi terhadap kondisi iklim apabila kondisi cuaca tidak mendukung untuk budidaya tanaman pangan seperti padi, jagung, ubi kayu, dan ubi jalar. Beberapa tanaman yang relatif lebih adaptif terhadap perubahan cuaca antara lain sawi, kangkung, bayam, terong, semangka, labu, dan melon ditunjukkan pada Tabel 5.

Tabel 5. Diversifikasi Tanaman

Jenis Tanaman	Waktu Panen	Pola Cuaca Tanaman
Sawi	Setiap Minggu	Curah hujan normal dan kelembaban tinggi
Kangkung	Setiap Minggu	Curah hujan rendah dan lama sinar matahari tinggi
Bayam	Setiap Minggu	Lama sinar matahari tinggi
Terong	Setiap Minggu	Curah hujan rendah lama sinar matahari tinggi
Semangka	3 Bulan	Lama sinar matahari tinggi
Labu	3 Bulan	Curah hujan tinggi
Melon	3 Bulan	Curah hujan rendah, kelembaban rendah

3.5. Perbandingan dengan Penelitian Terdahulu

Perbandingan nilai akurasi dengan penelitian-penelitian tersebut dapat dilihat pada Gambar 2



Gambar 4. Perbandingan dengan Peneliti Terdahulu

Sejumlah penelitian sebelumnya menunjukkan efektivitas algoritma Naïve Bayes dalam memprediksi hasil panen berbasis faktor iklim. Chandraprabha dan Dhanaraj (2020) berhasil memperoleh akurasi 97,53% pada prediksi produktivitas padi, menegaskan kemampuan algoritma ini dalam memodelkan hubungan antara variabel iklim dan hasil panen. Chusyairi et al. (2023) membuktikan bahwa kualitas dan jumlah data berpengaruh signifikan, di mana akurasi prediksi pertumbuhan jeruk keprok meningkat dari 65% menjadi 85% setelah augmentasi data. Heksaputra et al. (2020) juga mengonfirmasi hal serupa melalui penelitian pada kacang tanah, dengan akurasi yang naik dari 55% menjadi 72% pasca augmentasi.

Berbeda dengan studi terdahulu, penelitian ini memfokuskan pada prediksi hasil panen tanaman pangan di Kalimantan Selatan dengan atribut prediktor berupa variabel cuaca. Eksperimen menunjukkan akurasi awal 70% yang meningkat menjadi 90% setelah dilakukan oversampling, seiring pertambahan data dari 2.620 menjadi 3.024.

Peningkatan tersebut membuktikan kontribusi signifikan teknik augmentasi terhadap performa model, sekaligus memperlihatkan keunggulan hasil penelitian ini dibandingkan studi-studi sebelumnya.

Secara keseluruhan, temuan ini menegaskan bahwa Naïve Bayes, meskipun sederhana, mampu memberikan prediksi yang efektif dan efisien, khususnya pada data berskala besar. Keterbatasannya dalam menangani distribusi data tidak seimbang dapat diminimalisasi melalui teknik augmentasi. Oleh karena itu, integrasi antara data iklim dan pembelajaran mesin dipandang memiliki prospek strategis dalam mendukung ketahanan pangan, khususnya di wilayah tropis seperti Kalimantan Selatan.

4. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan bahwa faktor cuaca memiliki pengaruh signifikan terhadap produktivitas hasil panen tanaman pangan di Kalimantan Selatan, khususnya pada komoditas utama seperti padi, jagung, ubi jalar, dan ubi kayu. Variabel suhu, curah hujan, kelembapan, dan lama penyinaran matahari terbukti memberikan kontribusi penting dalam menentukan tingkat produktivitas panen. Penerapan algoritma Naïve Bayes menunjukkan performa prediksi yang baik, dengan akurasi sebesar 90%, dan weighted F1-Score sebesar 0.89. Kombinasi metrik ini menegaskan bahwa model tidak hanya mampu mengenali pola data mayoritas, tetapi juga cukup andal dalam memprediksi kelas minoritas, sehingga hasilnya lebih representatif terhadap kondisi nyata di lapangan. Dengan demikian, pendekatan probabilistik ini dapat dijadikan dasar dalam pengambilan keputusan pertanian berbasis data, seperti perencanaan pola tanam dan strategi adaptasi terhadap variabilitas iklim. Temuan ini memperkuat bukti bahwa pendekatan machine learning khususnya metode berbasis probabilitas efektif digunakan dalam menganalisis sistem pertanian yang kompleks dan penuh ketidakpastian, seperti wilayah pasang surut Barito Kuala, di mana fluktuasi curah hujan menjadi faktor dominan yang memengaruhi produktivitas lahan. Namun, penelitian ini memiliki keterbatasan. Rentang data yang digunakan hanya mencakup periode 2018–2023, sehingga belum sepenuhnya mencerminkan variasi iklim jangka panjang. Selain itu, variabel yang digunakan masih terbatas pada faktor cuaca, tanpa mempertimbangkan aspek agronomis dan sosial-ekonomi seperti jenis tanah, pola irigasi, serta praktik budidaya. Oleh karena itu, penelitian selanjutnya disarankan untuk memperluas cakupan temporal dan variabel, serta menguji algoritma perbandingan seperti SVM, Random Forest, atau Deep Learning agar sistem prediksi hasil panen menjadi lebih adaptif, komprehensif, dan akurat.

REFERENCES

- [1] A. Istiana, M. Husaini, dan L. Anjardiani, “Analisis Penyerapan Tenaga Kerja Sektor Pertanian Di Provinsi Kalimantan Selatan,” *Front. Agribisnis*, vol. 7, no. 1, hal. 24, 2023, doi: 10.20527/frontbiz.v7i1.8275.
- [2] J. Juswadi, P. Sumarna, dan N. S. Mulyati, “Produksi Jagung dan Umbi-Umbian, dan Peranannya dalam Perekonomian Indonesia,” *Paspalum J. Ilm. Pertan.*, vol. 8, no. 1, hal. 22, 2020, doi: 10.35138/paspalum.v8i1.149.
- [3] G. S. Malhi, M. Kaur, dan P. Kaushik, “Impacto del cambio climático en la agricultura y sus estrategias de mitigación: una revisión,” *Sustain.*, vol. 13, no. 3, hal. 1–21, 2021.
- [4] V. uli Sihombing, Ulidesi Siadari, Ratna Sari, dan Muhtar Ardansah Munthe, “The Impact Of Climate Change On Productivity and Food Security In Indonesia,” *J. Agri Socio Econ. Bus.*, vol. 5, no. 02, hal. 191–202, 2023, doi: 10.31186/jaseb.5.2.191-202.
- [5] Sevina Yushinta Anjani, Bagus Setiawan, dan Sofi Ayu Nur Martasari, “Dampak Perubahan Iklim Terhadap Ketahanan Pangan Di Indonesia,” *J. Pendidik. Dan Ilmu Sos.*, vol. 2, no. 3, hal. 46–55, 2024, doi: 10.54066/jupendis.v2i3.1850.
- [6] A. Yudhana, D. Sulisty, dan I. Mufandi, “GIS-based and Naïve Bayes for nitrogen soil mapping in Lendah, Indonesia,” *Sens. Bio-Sensing Res.*, vol. 33, hal. 100435, 2021, doi: 10.1016/j.sbsr.2021.100435.
- [7] S. M. Shawon, F. B. Ema, A. K. Mahi, F. L. Niha, dan H. T. Zubair, “Crop Yield Prediction Using Machine Learning: An extensive and systematic literature review,” *Smart Agric. Technol.*, vol. 10, no. September 2024, hal. 100718, 2025, doi: 10.1016/j.atech.2024.100718.
- [8] T. C. Olayinka, A. O. Adetunmbi, O. O. Obe, E. O. Ibam, dan A. S. Olayinka, “A Data-Driven Machine Learning Approach Toward An Improved Maize Crop Production,” *Franklin Open*, hal. 100334, 2025, doi: 10.1016/j.fraope.2025.100334.
- [9] O. Peretz, M. Koren, dan O. Koren, “Naive Bayes classifier – An Ensemble Procedure for Recall and Precision Enrichment,” *Eng. Appl. Artif. Intell.*, vol. 136, no. PB, hal. 108972, 2024, doi: 10.1016/j.engappai.2024.108972.
- [10] M. Chandraprabha dan R. K. Dhanaraj, “Machine learning based Pedantic Analysis of Predictive Algorithms in Crop Yield Management,” *Proc. 4th Int. Conf. Electron. Commun. Aerosp. Technol. ICECA 2020*, no. 3, hal. 1340–1345, 2020, doi: 10.1109/ICECA49313.2020.9297544.
- [11] N. L. W. S. R. Ginantra *et al.*, *Data Mining dan Penerapan Algoritma*. 2021.
- [12] L. Febriyanti dan H. Zakaria, “Implementasi Data Mining Untuk Memprediksi Produktivitas Pada Tanaman Kacang Tanah Menggunakan Metode Naive Bayes (Studi Kasus : Perkebunan Kacang Tanah Di Kota Bogor),” *Log. J. Ilmu Komput. dan Pendidik.*, vol. 1, no. 2, hal. 105–118, 2023, [Daring]. Tersedia pada: <https://journal.mediapublikasi.id/index.php/logic/article/view/2353>
- [13] Heksaputra D, Azani Y, dan Naimah Z, “Penentuan Pengaruh Iklim Terhadap Pertumbuhan Tanaman dengan Naïve Bayes Lizda Iswari,” *Semin. Nas. Apl. Teknol. Inf.*, hal. 34–39, 2020.
- [14] A. Chusyairi, T. Haryanto, dan R. N. Hayat, “Prediksi Perubahan Iklim Untuk Pertumbuhan Tanaman Jeruk Keprok Menggunakan Naïve Bayes,” *Inform. Mulawarman J. Ilm. Ilmu Komput.*, vol. 18, no. 1, hal. 23, 2023, doi: 10.30872/jim.v18i1.9352.
- [15] M. Y. Shams, S. A. Gamel, dan F. M. Talaat, “Enhancing Crop Recommendation Systems with Explainable Artificial



- Intelligence: A Study on Agricultural Decision-Making,” *Neural Comput. Appl.*, vol. 36, no. 11, hal. 5695–5714, 2024, doi: 10.1007/s00521-023-09391-2.
- [16] L. Yusuf dan R. Mualifa, “Development of a Data Cleaning System for Consumer Master Data using Sorted Neighborhood and N-Gram Methods,” *ELKOMIKA J. Tek. Energi Elektr. Tek. Telekomun. Tek. Elektron.*, vol. 13, no. 1, hal. 57, 2025, doi: 10.26760/elkomika.v13i1.57.
- [17] M. Ardiansyah, “Peningkatan Kualitas Statistik Resmi Produktivitas Padi melalui Imputasi Data Non-respons Menggunakan Model Aditif Geospasial,” *J. Apl. Stat. Komputasi Stat.*, vol. 15, no. 2, hal. 43–52, 2023, doi: 10.34123/jurnalask.v15i2.443.
- [18] A. Nur Azizah, M. Falach Asy’ari, I. Wisma Dwi Prastya, dan D. Purwitasari, “Easy Data Augmentation untuk Data yang Imbalance pada Konsultasi Kesehatan Daring,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 5, hal. 1095–1104, 2023, doi: 10.25126/jtiik.2023107082.
- [19] P. Ayuningtyas, S. Khomsah, dan S. Sudioanto, “Pelabelan Sentimen Berbasis Semi-Supervised Learning menggunakan Algoritma LSTM dan GRU,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 9, no. 3, hal. 217–229, 2024, doi: 10.14421/jiska.2024.9.3.217-229.
- [20] M. Vishwakarma dan N. Kesswani, “A new two-phase intrusion detection system with Naïve Bayes machine learning for data classification and elliptic envelop method for anomaly detection,” *Decis. Anal. J.*, vol. 7, no. January, hal. 100233, 2023, doi: 10.1016/j.dajour.2023.100233.
- [21] H. Yoshikawa, “Can naive Bayes classifier predict infection in a close contact of COVID-19? A comparative test for predictability of the predictive model and healthcare workers in Japan: Infection Prediction in a Close Contact of COVID-19,” *J. Infect. Chemother.*, vol. 28, no. 6, hal. 774–779, 2022, doi: 10.1016/j.jiac.2022.02.017.
- [22] M. K. Senapaty, A. Ray, dan N. Padhy, “A Decision Support System for Crop Recommendation Using Machine Learning Classification Algorithms,” *Agric.*, vol. 14, no. 8, 2024, doi: 10.3390/agriculture14081256.
- [23] M. A. Jabed dan M. A. Azmi Murad, “Crop Yield Prediction in Agriculture: A Comprehensive Review of Machine Learning and Deep Learning Approaches, with Insights For Future Research and Sustainability,” *Heliyon*, vol. 10, no. 24, hal. e40836, 2024, doi: 10.1016/j.heliyon.2024.e40836.
- [24] H. Chen, S. Hu, R. Hua, dan X. Zhao, “Improved Naive Bayes Classification Algorithm For Traffic Risk Management,” *EURASIP J. Adv. Signal Process.*, vol. 2021, no. 1, 2021, doi: 10.1186/s13634-021-00742-6.