

**PERBANDINGAN ALGORITMA *CATBOOST* DAN *EXTRA TREES*
CLASSIFIER UNTUK PREDIKSI TINGKAT KEBERHASILAN
PANEN PADI BERDASARKAN FAKTOR CUACA
DAN KONDISI TANAH**

SKRIPSI

DISUSUN OLEH

RIO FADLY

2209020135



UMSU

Unggul | Cerdas | Terpercaya

**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

MEDAN

2026

**PERBANDINGAN ALGORITMA *CATBOOST* DAN *EXTRA TREES*
CLASSIFIER UNTUK PREDIKSI TINGKAT KEBERHASILAN
PANEN PADI BERDASARKAN FAKTOR CUACA
DAN KONDISI TANAH**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana
Komputer (S.Kom) dalam Program Studi Teknologi Informasi pada Fakultas
Ilmu Komputer dan Teknologi Informasi, Universitas Muhammadiyah
Sumatera Utara**

RIO FADLY

2209020135

**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

MEDAN

2026

LEMBAR PENGESAHAN

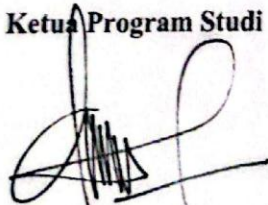
Judul Skripsi : PERBANDINGAN ALGORITMA CATBOOST DAN
EXTRA TREES CLASSIFIER UNTUK PREDIKSI
TINGKAT KEBERHASILAN PANEN PADI
BERDASARKAN FAKTOR CUACA DAN KONDISI
TANAH
Nama Mahasiswa : RIO FADLY
NPM : 2209020135
Program Studi : TEKNOLOGI INFORMASI

Menyetujui
Komisi Pembimbing



(Mulkan Azhari, S.Kom., M.Kom.)
NIDN. 0108129402

Ketua Program Studi



(Fatma Sari Hutagalung, S.Kom., M.Kom.)
NIDN. 0117019301

Dekan



(Dr. Al-Khowafizmi, S.Kom., M.Kom.)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

**PERBANDINGAN ALGORITMA *CATBOOST* DAN *EXTRA TREES*
CLASSIFIER UNTUK PREDIKSI TINGKAT KEBERHASILAN
PANEN PADI BERDASARKAN FAKTOR CUACA
DAN KONDISI TANAH**

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, April 2026

Yang membuat pernyataan



Rio Fadly

NPM. 2209020135

**PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama : Rio Fadly
NPM : 2209020135
Program Studi : Teknologi Informasi
Karya Ilmiah : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bedas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

**PERBANDINGAN ALGORITMA *CATBOOST* DAN *EXTRA TREES*
CLASSIFIER UNTUK PREDIKSI TINGKAT KEBERHASILAN
PANEN PADI BERDASARKAN FAKTOR CUACA
DAN KONDISI TANAH**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, 10 April 2026

Yang membuat pernyataan



Rio Fadly

NPM. 2209020135

DATA PRIBADI

Nama Lengkap : Rio Fadly
Tempat dan Tanggal Lahir : Tualang, 20 November 2004
Alamat Rumah : Jl. Kabupaten No 19.E, Perbaungan,
Simpang Tiga Pekan
Telepon/Faks/HP : 083191590800
E-mail : riofadly534@gmail.com
Instansi Tempat Kerja : -
Alamat Kantor : -

DATA PENDIDIKAN

SD : SDN 101929 Perbaungan TAMAT: 2016
SMP : Setia Budi Abadi Perbaungan TAMAT: 2019
SMA : SMAN 1 Perbaungan TAMAT: 2022

KATA PENGANTAR



Alhamdulillah, dengan mengucapkan puji syukur saya panjatkan kehadiran Allah SWT, yang telah melimpahkan Rahmat dan karunia-Nya, sehingga pada akhirnya penulis dapat menyelesaikan skripsi yang berjudul **“Perbandingan Algoritma *CatBoost* Dan *Extra Trees Classifier* Untuk Prediksi Tingkat Keberhasilan Panen Padi Berdasarkan Faktor Cuaca Dan Kondisi Tanah”**.

Penulisan skripsi ini dibuat sebagai salah satu syarat kelulusan program Sarjana Komputer dari Program Studi Teknologi Informasi Fakultas Ilmu Komputer Dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara. Penulis menyadari bahwa tanpa bimbingan dan dorongan dari semua pihak, maka penulisan skripsi ini tidak berjalan lancar. Oleh karena itu Penulis tentunya berterima kasih kepada berbagai pihak dalam dukungan serta doa dalam penyelesaian skripsi.

Penulis juga mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Akrim, M.Pd., Rektor Universitas Muhammadiyah Sumatera Utara (UMSU)
2. Bapak Dr. Al-Khowarizmi, S.Kom.,M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Ibu Dr.Firahmi Rizky, S.Kom.,M.Kom Wakil Dekan I Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
4. Bapak Basri, S.Si.,M.Kom Wakil Dekan III Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.

5. Ibu Fatma Hutagalung, S.Kom.,M.kom Ketua Program Studi Teknologi Informasi Universitas Muhammadiyah Sumatera Utara.
6. Bapak Okvi Nugroho, S.Kom.,M.Kom Sekretaris Program Studi Teknologi Informasi Universitas Muhammadiyah Sumatera Utara.
7. Bapak Mulkan Azhari, S.Kom,M.Kom Dosen pembimbing skripsi yang telah menyediakan waktu, tenaga dan pikiran untuk mengarahkan penulis dalam penyusunan skripsi ini. Terima kasih atas ilmu dan pengetahuan yang telah diberikan.
8. Seluruh dosen Fakultas Ilmu Komputer Prodi Teknologi Informasi Universitas Muhammadiyah Sumatera Utara yang telah memberikan ilmu dan pengalamannya serta mendidik penulis selama masa kuliah.
9. Teristimewa untuk Ibu Tercinta, Ibunda Siti Asiah yang sudah menjadi ibu sekaligus ayah buat penulis. Terima kasih telah membesarkan dan mendidik penulis dari lahir hingga saat ini.
10. Rekan-rekan mahasiswa/mahasiswi Teknologi Informasi Angkatan 2022 yang telah memberikan kenangan serta pengalaman.
11. Seluruh pihak yang tidak dapat disebutkan satu persatu namun telah memberikan dukungan, semangat dan masukkan kepada penulis.

Akhir kata, penulis menyadari skripsi ini masih jauh dari kata sempurna, hal ini dikarenakan keterbatasan pengetahuan, pengalaman, maupun kemampuan yang dimiliki penulis. Oleh karena itu, penulis mengharapkan kepada semua pihak agar dapat memberikan saran dan kritik yang bersifat membangun dari para pembaca sebagai bahan masukan bagi penulis di masa yang akan datang. Penulis berharap semoga skripsi ini dapat bermanfaat bagi para pembaca. Serta kepada

semua pihak yang telah memberikan bantuan dalam penulisan ini, penulis ucapkan banyak terima kasih dan semoga Allah SWT membalas kebaikan kalian. *Aamiin yarabbal'alamin.*

**PERBANDINGAN ALGORITMA *CATBOOST* DAN *EXTRA TREES*
CLASSIFIER UNTUK PREDIKSI TINGKAT KEBERHASILAN
PANEN PADI BERDASARKAN FAKTOR CUACA
DAN KONDISI TANAH**

ABSTRAK

Produksi padi merupakan salah satu faktor penting dalam menjaga ketahanan pangan, khususnya di Indonesia yang menjadikan beras sebagai bahan pangan utama. Namun, hasil panen padi sering dipengaruhi oleh berbagai faktor lingkungan seperti curah hujan, suhu, kelembapan, pH tanah, serta kandungan unsur hara tanah seperti nitrogen, fosfor, dan kalium. Ketidakpastian kondisi lingkungan tersebut menyebabkan petani kesulitan dalam memprediksi potensi hasil panen secara akurat. Oleh karena itu, diperlukan suatu sistem yang mampu membantu memprediksi hasil panen padi berdasarkan kondisi lingkungan dan tanah secara cepat dan akurat. Penelitian ini bertujuan untuk membangun sistem prediksi hasil panen padi berbasis *machine learning* dengan menggunakan algoritma *CatBoost* dan *ExtraTrees*, serta mengimplementasikan *Explainable Artificial Intelligence (XAI)* menggunakan metode *SHAP (SHapley Additive Explanations)* untuk memberikan interpretasi terhadap hasil prediksi model. Dataset yang digunakan terdiri dari 1500 data yang memuat variabel curah hujan, suhu, kelembapan, pH tanah, nitrogen, fosfor, dan kalium dengan tiga kelas hasil panen yaitu rendah, sedang, dan tinggi. Data kemudian diproses menggunakan metode *train-test split* dengan proporsi 80% data pelatihan dan 20% data pengujian. Hasil penelitian menunjukkan bahwa algoritma *CatBoost* memiliki performa yang lebih baik dibandingkan *ExtraTrees* dengan nilai akurasi yang lebih tinggi. Berdasarkan evaluasi menggunakan *ROC Curve*, model *CatBoost* memperoleh nilai AUC sebesar 0,99 pada kelas rendah, 0,98 pada kelas sedang, dan 0,99 pada kelas tinggi, yang termasuk dalam kategori *excellent classification*. Sementara itu, model *ExtraTrees* memperoleh nilai AUC sebesar 0,96 pada kelas rendah, 0,81 pada kelas sedang, dan 0,93 pada kelas tinggi. Analisis menggunakan metode *SHAP* menunjukkan bahwa faktor yang paling berpengaruh terhadap hasil prediksi panen adalah nitrogen, pH tanah, dan kelembapan. Sistem yang dibangun juga

diimplementasikan dalam bentuk aplikasi web berbasis Laravel yang dapat digunakan untuk melakukan prediksi dan memberikan analisis kondisi lahan secara otomatis. Dengan adanya sistem ini diharapkan dapat membantu petani atau pengguna dalam mengetahui potensi hasil panen padi serta memahami faktor-faktor lingkungan yang mempengaruhi hasil panen sehingga dapat menjadi alat bantu pengambilan keputusan (*decision support system*) dalam pengelolaan lahan pertanian.

Kata Kunci: *Machine Learning*, Prediksi Panen Padi, *CatBoost*, *ExtraTrees*, *Explainable AI*, *SHAP*.

**COMPARISON OF CATBOOST AND EXTRA TREES CLASSIFIER
ALGORITHMS FOR PREDICTING RICE HARVEST SUCCESS
LEVEL BASED ON WEATHER FACTORS
AND SOIL CONDITIONS.**

ABSTRACT

Rice production is one of the important factors in maintaining food security, especially in Indonesia where rice is the main staple food. However, rice yield is often influenced by various environmental factors such as rainfall, temperature, humidity, soil pH, and soil nutrients including nitrogen, phosphorus, and potassium. The uncertainty of these environmental conditions makes it difficult for farmers to accurately predict potential crop yields. Therefore, a system is needed to help predict rice yield based on environmental and soil conditions quickly and accurately. This study aims to develop a rice yield prediction system based on machine learning using the CatBoost and ExtraTrees algorithms, and to implement Explainable Artificial Intelligence (XAI) using the SHAP (SHapley Additive Explanations) method to provide interpretation of the model predictions. The dataset used in this study consists of 1500 records containing variables of rainfall, temperature, humidity, soil pH, nitrogen, phosphorus, and potassium with three yield classes, namely low, medium, and high. The dataset was processed using the train-test split method with a proportion of 80% training data and 20% testing data. The results show that the CatBoost algorithm performs better than ExtraTrees with higher accuracy performance. Based on evaluation using the ROC Curve, the CatBoost model achieved AUC values of 0.99 for the low class, 0.98 for the medium class, and 0.99 for the high class, which are categorized as excellent classification performance. Meanwhile, the ExtraTrees model achieved AUC values of 0.96 for the low class, 0.81 for the medium class, and 0.93 for the high class. The SHAP analysis indicates that the most influential factors affecting the prediction results are nitrogen, soil pH, and humidity. The developed system is also implemented in a web-based application using the Laravel framework, which allows users to perform predictions and obtain automatic analysis of land conditions. This system is expected to assist farmers or users in identifying potential rice yields and

understanding environmental factors that influence crop production. Therefore, it can serve as a decision support system in agricultural land management.

Keywords: Machine Learning, Rice Yield Prediction, CatBoost, ExtraTrees, Explainable AI, SHAP.

DAFTAR ISI

LEMBAR PENGESAHAN	iii
PENYATAAN ORISINALITAS	iv
PENYATAAN PERSETUJUAN PUBLIKASI	v
RIWAYAT HIDUP	vi
KATA PENGANTAR.....	vii
ABSTRAK	x
ABSTRACT.....	xii
DAFTAR ISI	xiv
DAFTAR TABEL.....	xvi
DAFTAR GAMBAR	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	5
BAB II LANDASAN TEORI	8
2.1 Studi Terdahulu.....	8
2.2 Tinjauan Umum Penelitian Penelitian Hasil Panen.....	11
2.2.1 Perkembangan Teknologi Pertanian.....	13
2.2.2 Permasalahan Prediksi Keberhasilan Panen Padi	14
2.3 Konsep Dasar <i>Machine Learning</i>	15
2.3.1 Pengertian <i>Machine Learning</i>	16
2.3.2 Jenis-jenis <i>Machine Learning</i>	16
2.3.3 Penerapan <i>Machine Learning</i> Di Pertanian	18
2.4 Prediksi Dan Klasifikasi	22
2.4.1 Dasar Prediksi Dengan <i>Machine Learning</i>	23
2.4.2 Dasar Klasifikasi Dalam <i>Machine Learning</i>	24
2.4.3 Peran <i>Machine Learning</i> Di Pertanian.....	24
2.5 Algoritma <i>CatBoost</i>	26
2.5.1 Konsep Dasar dan Mekanisme Kerja <i>CatBoost</i>	26

2.5.2 Karakteristik Algoritma <i>CatBoost</i>	29
2.5.3 Keunggulan Algoritma <i>CatBoost</i>	30
2.6 Algoritma Extra Tress Classifier	30
2.6.1 Konsep Dasar dan Mekanisme kerja <i>Extra Tress Classifier</i>	31
2.6.2 Karakteristik Algoritma <i>Extra Tress Classifier</i>	32
2.6.3 Keunggulan Algoritma <i>Extra Tress Classifier</i>	33
2.7 Faktor Lingkungan Mempengaruhi Panen Padi	33
2.7.1 Faktor Cuaca	35
2.7.2 Faktor Tanah	36
BAB III METODOLOGI PENELITIAN.....	38
3.1 Desain Dan Jenis Penelitian	38
3.2 Objek, Dan Waktu Penelitian	39
3.3 Data Penelitian.....	42
3.3.1 Variabel Penelitian	43
3.3.2 Skema Kelas Tingkat Keberhasilan Panen	43
3.4 Tahapan Penelitian.....	44
3.5 Prapemrosesan Data	45
3.6 Pembagian Data Dan Strategi Validasi	46
3.7 Pembangunan Model Dan Penalaan Parameter.....	47
3.7.1 Model <i>CatBoost</i>	47
3.7.2 Model <i>Extra Tress Classifier</i>	48
3.8 Metode Evaluasi Model.....	48
3.9 Interpretasi Model dan Analisis Hasil	49
3.10 Perangkat Lunak Dan Implementasi.....	50
BAB IV HASIL DAN PEMBAHASAN.....	53
4.1 Implementasi Sistem.....	53
4.2 Implementasi Dataset	54
4.3 Hasil <i>Preprocessing</i> Data	56
4.4 Pembagian Dataset	57
4.5 Hasil Evaluasi Model.....	58
4.6 Confusion Matrix dan Kurva ROC <i>CatBoost</i>	58
4.7 Confusion Matrix dan Kurva ROC Extra Trees	64

4.8 Feature Importance	71
4.8.1 Analisis <i>Feature Importance</i> Model <i>CatBoost</i>	71
4.8.2 Analisis <i>Feature Importance</i> Model <i>Extra Trees</i>	73
4.9 SHAP Beeswarm Plot <i>CatBoost</i>	74
4.9.1 Analisis <i>SHAP CatBoost</i> (Kelas Rendah).....	75
4.9.2 Analisis <i>SHAP CatBoost</i> (Kelas Sedang)	76
4.9.3 Analisis <i>SHAP CatBoost</i> (Kelas Tinggi).....	77
4.10 SHAP Beeswarm Plot <i>Extra Trees Classifier</i>	78
4.10.1 Analisis <i>SHAP Extra Trees</i> (Kelas Rendah)	79
4.10.2 Analisis <i>SHAP Extra Trees</i> (Kelas Sedang).....	80
4.10.3 Analisis <i>SHAP Extra Trees</i> (Kelas Tinggi).....	81
4.11 Kesimpulan Perbandingan Model	83
4.12 Penerapan Algoritma Terpilih	84
4.13 Pengujian <i>Prototype</i> Perangkat Lunak(<i>Software</i>)	90
4.14 Lingkungan Pengujian	92
4.15 Aspek Penelitian Lanjut	93
BAB V PENUTUP.....	94
5.1 Kesimpulan	94
5.2 Saran	95
DAFTAR PUSTAKA	97

DAFTAR TABEL

Tabel 2.1 Studi Terdahulu	8
Tabel 3.1 Ruang Lingkup Penelitian	40
Tabel 3.2 Waktu Penelitian.....	41
Tabel 3.3 Rekapitulasi Data Sekunder	42
Tabel 3.4 Peran Library Perangkat Lunak.....	50
Tabel 4.4 Struktur Dataset.....	55
Tabel 4.5 Hasil Perhitungan <i>CatBoost</i>	63
Tabel 4.6 Hasil Perhitungan <i>Extra Trees Classifier</i>	70
Tabel 4.7 Tingkat Kepentingan Fitur <i>CatBoost</i>	72
Tabel 4.8 Tingkat Kepentingan Fitur <i>Extra Trees Classifier</i>	73
Tabel 4.9 Karakteristik <i>Interpretasi</i>	83
Tabel 4.10 Pengujian Kotak Hitam	90
Tabel 4.11 Pengujian Data <i>Upload</i>	91
Tabel 4.12 Pengujian <i>Dashboard</i>	92

DAFTAR GAMBAR

Gambar 2.1 Teknik Pengumpulan Data	19
Gambar 2.2 Kerangka Konsep	21
Gambar 2.3 Hasil Prediksi Dan Rekomendasi	23
Gambar 2.4 Cara Kerja <i>CatBoost</i>	28
Gambar 2.5 Cara Kerja <i>Extra Tress Classifier</i>	32
Gambar 3.1 <i>supervised machine learning</i> :	39
Gambar 3.2 Tahapan Penelitian	44
Gambar 3.3 Tahapan Prapemrosesan	46
Gambar 3.4 Proses Model <i>CatBoost</i>	47
Gambar 3.5 Proses Model <i>Extra Tress Classifier</i>	48
Gambar 3.6 Model <i>SHAP</i>	49
Gambar 3.7 Model Representasi Diagram	51
Gambar 4.1 Implementasi Sistem	54
Gambar 4.2 <i>Confusion Matrix Training CatBoost</i>	59
Gambar 4.3 <i>Confusion Matrix Testing CatBoost</i>	61
Gambar 4.4 Nilai AUC dengan Algoritma <i>CatBoost</i>	64
Gambar 4.5 <i>Confusion Matrix Training Extra Trees Classifier</i>	65
Gambar 4.6 <i>Confusion Matrix Testing Extra Trees Classifier</i>	67
Gambar 4.7 Nilai AUC dengan Algoritma <i>Extra Trees Classifier</i>	70
Gambar 4.8 <i>Feature Importance CatBoost</i>	71
Gambar 4.9 <i>Feature Importance Extra Trees Classifier</i>	73
Gambar 4.10 <i>SHAP CatBoost</i> (Kelas Rendah)	75
Gambar 4.11 <i>SHAP CatBoost</i> (Kelas Sedang).....	76
Gambar 4.12 <i>SHAP CatBoost</i> (Kelas Tinggi)	77
Gambar 4.13 <i>SHAP Extra Trees Classifier</i> (Kelas Rendah)	79
Gambar 4.14 <i>SHAP Extra Trees Classifier</i> (Kelas Sedang).....	80
Gambar 4.15 <i>SHAP Extra Trees Classifier</i> (Kelas Tinggi).....	81
Gambar 4.16 <i>Activity Diagram Login</i>	84
Gambar 4.17 <i>Use Case Login</i>	85
Gambar 4.18 <i>Class Diagram</i>	85
Gambar 4.19 <i>Flow Chart Prototype</i> Prediksi Keberhasilan Panen.....	86
Gambar 4.20 <i>Form Login</i>	87
Gambar 4.21 Tampilan Prediksi Panen	88
Gambar 4.22 <i>Form Upload Data</i>	88
Gambar 4.23 Halaman <i>Dashboard</i>	89
Gambar 4.24 Halaman Riwayat	89
Gambar 4.25 Grafik Akurasi	90

BAB I

PENDAHULUAN

1.1 Latar Belakang

Sektor pertanian merupakan salah satu pilar penting penggerak perekonomian Indonesia, sejalan dengan statusnya sebagai negara agraris. Di antara berbagai komoditas pertanian, padi menempati posisi yang sangat strategis karena menjadi sumber pangan utama bagi lebih dari 280 juta penduduk Indonesia (Nizami et al., 2025). Namun dalam beberapa dekade terakhir, sektor pertanian menghadapi tantangan yang semakin kompleks, khususnya terkait perubahan iklim, degradasi kondisi tanah, serta keterbatasan akurasi dalam memprediksi hasil panen. Curah hujan yang tidak menentu, peningkatan suhu udara, perubahan pola musim, serta kondisi tanah yang beragam sering menyebabkan produktivitas padi di berbagai daerah. Situasi ini menuntut adanya pendekatan baru yang dapat membantu petani dan instansi pemerintah dalam mencerminkan tingkat keberhasilan panen secara lebih akurat dan berbasis data.

Selama ini, prediksi keberhasilan panen padi umumnya masih bergantung pada metode manual atau pendekatan statistik sederhana yang tidak mampu menangkap hubungan kompleks antar variabel lingkungan. Padahal, suksesnya panen padi dipengaruhi oleh banyak faktor seperti curah hujan, suhu udara, kelembapan, intensitas penyelarasan, pH tanah, kandungan unsur hara (N, P, K), tekstur tanah, jenis varietas, serta pola pemupukan. Ketika variabel-variabel tersebut saling berinteraksi, diperlukan metode pengolahan data yang lebih canggih agar prediksi dapat dilakukan secara lebih tepat dan konsisten. *Machine learning* adalah proses untuk mendapatkan pengetahuan dari data. Bidang ini merupakan gabungan

antara statistik, kecerdasan buatan, dan ilmu komputer. Pembelajaran mesin juga sering disebut sebagai analitik prediktif atau pembelajaran statistik (Shawon et al., 2025). Dengan memanfaatkan data sekunder, *machine learning* dapat menghasilkan model yang mampu memberikan prediksi panen yang akurat. Namun, sebagian besar penelitian terdahulu di Indonesia masih menggunakan algoritma umum seperti *Linear Regression*, *Decision Tree*, *Random Forest*, atau *Support Vector Machine (SVM)*. Penggunaan algoritma modern seperti *CatBoost* dan *Extra Trees Classifier* masih sangat jarang diterapkan, terutama pada bidang pertanian berdasarkan faktor cuaca dan kondisi tanah.

CatBoost merupakan algoritma boosting modern yang dikembangkan oleh Yandex, dengan kemampuan unggul dalam menangani fitur kategorikal dan data non-linear tanpa memerlukan proses pengkodean yang kompleks. Algoritma ini mampu memberikan performa tinggi pada dataset dengan variabel campuran (numerik dan kategori), sehingga sangat cocok digunakan untuk data pertanian yang umumnya heterogen. Sementara itu, *Extra Trees Classifier* merupakan varian dari algoritma ensemble berbasis pohon yang memiliki tingkat pengacakan lebih tinggi dibandingkan *Random Forest*. Dengan pengacakan pada pemilihan ambang batas dan fitur, algoritma ini dapat memberikan generalisasi yang lebih baik pada dataset besar.

Algoritma kedua ini menawarkan keunggulan namun belum banyak diteliti secara langsung dalam konteks prediksi keberhasilan panen padi. Oleh karena itu, penelitian ini penting dilakukan untuk mengetahui sejauh mana kinerja kedua algoritma tersebut dalam memprediksi produktivitas padi berdasarkan faktor cuaca dan kondisi tanah. Perbandingan kinerja kedua model ini diharapkan mampu

memberikan wawasan baru mengenai algoritma yang paling tepat untuk digunakan dalam sistem prediksi pertanian modern.

Selain itu, hasil penelitian ini berpotensi memberikan manfaat praktis dalam dunia pertanian Indonesia. Prediksi tingkat keberhasilan panen yang lebih akurat dapat membantu petani, penyuluh pertanian, dan dinas pertanian dalam merencanakan jadwal tanam, strategi pemupukan, manajemen irigasi, serta estimasi stok pangan daerah. Dengan demikian, penelitian ini mendukung pengembangan pertanian presisi (pertanian presisi) yang menjadi solusi utama dalam menahan tantangan perubahan iklim.

Berdasarkan latar belakang tersebut, penelitian ini fokus pada penerapan dan perbandingan dua algoritma pembelajaran mesin yang jarang digunakan *CatBoost* dan *Extra Trees Classifier* untuk memprediksi tingkat keberhasilan panen padi berdasarkan faktor cuaca dan kondisi tanah dengan dataset yang besar. Tujuan akhirnya adalah menghasilkan model prediksi yang akurat, stabil, dan mampu mendukung pengambilan keputusan yang efektif di sektor pertanian berbasis teknologi informasi.

1.2 Rumusan Masalah

Berdasarkan latar belakang permasalahan di atas, maka pokok permasalahan yang akan dibahas di penelitian ini adalah sebagai berikut:

1. Prediksi keberhasilan panen padi belum optimal karena dipengaruhi banyak faktor (cuaca dan kondisi tanah) yang bersifat dinamis dan non-linear.
2. Metode konvensional dan algoritma machine learning umum belum mampu menghasilkan prediksi yang akurat dan stabil, terutama pada data pertanian yang besar dan beragam (heterogen).

3. Belum diketahui secara jelas perbandingan performa algoritma modern seperti CatBoost dan Extra Trees dalam memprediksi tingkat keberhasilan panen padi.
4. Kurangnya kemampuan model dalam mengidentifikasi variabel paling berpengaruh (seperti curah hujan, suhu, pH, nitrogen, dll) yang menentukan hasil panen.
5. Hasil prediksi belum sepenuhnya dapat digunakan sebagai dasar pengambilan keputusan praktis di bidang pertanian.

1.3 Batasan Masalah

Adapun Batasan masalah dari penelitian ini adalah sebagai berikut:

1. Penelitian hanya menggunakan variabel: curah hujan, suhu, kelembapan, pH tanah, nitrogen, fosfor, dan kalium.
2. Faktor lain seperti hama, varietas benih, irigasi, dan pola pemupukan tidak dianalisis.
3. Data yang digunakan merupakan dataset sekunder/simulasi, bukan hasil pengambilan data lapangan langsung.
4. Penelitian tidak melakukan prediksi jangka panjang (forecasting).
5. Dataset merepresentasikan kondisi pertanian padi di Sumatera Utara tahun 2024–2025 dalam bentuk data simulasi.
6. Algoritma yang dibandingkan hanya CatBoost dan Extra Trees Classifier.
7. Evaluasi model terbatas pada metrik: akurasi, presisi, recall, dan F1-score
8. Output model berupa klasifikasi (rendah, sedang, tinggi), bukan prediksi jumlah hasil panen (regresi).
9. Penelitian tidak mencakup implementasi real-time atau integrasi dengan IoT

10. Fokus penelitian hanya pada pengembangan dan evaluasi model berbasis data historis.

1.4 Tujuan Penelitian

1. Membangun model prediksi tingkat keberhasilan panen padi berbasis machine learning menggunakan variabel cuaca dan kondisi tanah.
2. Membandingkan kinerja algoritma CatBoost dan Extra Trees Classifier berdasarkan metrik accuracy, precision, recall, dan F1-score.
3. Mengevaluasi performa model pada data uji, stabilitas model, serta waktu komputasi.
4. Mengidentifikasi faktor lingkungan yang paling berpengaruh terhadap keberhasilan panen padi.
5. Menghasilkan model yang dapat dijadikan referensi dalam pengambilan keputusan di bidang pertanian.

1.5 Manfaat Penelitian

1. Manfaat Institusi

Penelitian ini diharapkan memberikan manfaat bagi universitas sebagai bentuk kontribusi akademik dalam pengembangan penelitian di bidang informasi teknologi, khususnya pada penerapan *machine learning* untuk sektor pertanian. Hasil penelitian ini dapat memperkaya koleksi karya ilmiah universitas dan menjadi referensi bagi mahasiswa maupun dosen dalam pengembangan penelitian sejenis. Selain itu, penelitian ini juga dapat mendukung peningkatan kualitas program studi akademik serta mendorong kolaborasi lintas disiplin antara informasi teknologi dan bidang pertanian.

2. Manfaat Keilmuan

Penelitian ini memberikan kontribusi terhadap pengembangan ilmu pengetahuan di bidang *machine learning* dengan menghadirkan kajian analisis dua algoritma pembelajaran ansambel yang relatif jarang digunakan, yaitu *CatBoost* dan *Extra Trees Classifier*. Penelitian ini memperluas wawasan ilmiah mengenai penerapan algoritma pembelajaran mesin modern pada data pertanian, khususnya dalam prediksi tingkat keberhasilan panen padi berdasarkan faktor cuaca dan kondisi tanah. Hasil penelitian diharapkan dapat menjadi rujukan dalam penelitian selanjutnya serta menyajikan literatur ilmiah terkait pertanian berbasis data.

3. Manfaat Penulis

Penelitian ini memberikan manfaat berupa peningkatan pemahaman dan keterampilan dalam mengolah data berukuran besar, menerapkan pembelajaran mesin algoritma, serta menyebarkan kinerja model prediktif. Penulis memperoleh pengalaman langsung dalam melakukan penelitian ilmiah secara sistematis, mulai dari pengumpulan data, pra-pemrosesan, pemodelan, hingga interpretasi hasil. Pengalaman ini diharapkan dapat menjadi bekal akademik dan profesional bagi penulis di bidang teknologi informasi dan data *science*.

4. Manfaat Sosial

Penelitian ini diharapkan dapat dirasakan oleh masyarakat luas, khususnya pelaku di sektor pertanian. Model prediksi yang dihasilkan dapat memberikan gambaran awal mengenai tingkat keberhasilan panen padi sehingga dapat dimanfaatkan sebagai bahan pertimbangan dalam perencanaan tanam dan pengelolaan lahan. Dengan adanya berbasis data prediksi, risiko pemanasan hasil panen akibat perubahan cuaca dan kondisi tanah dapat diminimalkan.

Secara tidak langsung, penelitian ini juga mendukung upaya peningkatan ketahanan pangan dan mendorong pemanfaatan teknologi informasi sebagai solusi inovatif.

BAB II

LANDASAN TEORI

2.1 Studi Terdahulu

Tabel 2.1 Studi Terdahulu

No	Peneliti (Tahun)	Fokus Penelitian	Metodologi Utama	Limitasi (<i>Research Gap</i>)
1	Shawon dkk. (2024)	Identifikasi teknik & fitur dominan prediksi hasil panen	<i>Pembelajaran Mesin (SLR)</i>	Menunjukkan algoritma terbaik & tren ML saat ini
2	Razavi dkk. (2024)	<i>CatBoost</i> unggul di berbagai metrik	<i>CatBoost, GBT</i>	Menegaskan penggunaan <i>CatBoost</i>
3	Bahtiar (2023)	Model linier mencapai R^2 tinggi	<i>Regresi Linier</i>	Referensi metode regresi dalam prediksi
4	Manurung (2025)	MSE rendah & R^2 tinggi	<i>Regresor Hutan Acak</i>	Dasar penggunaan ML pada padi lokal
5	Bayutama & Saiful (2025)	RF sedikit unggul, wawasan ansambel	<i>Random Forest & XGBoost</i>	Konteks ML komparatif
6	Quille-Mamani dkk. (2025)	Variabel spektral & tekstural penting	<i>MLR, SVR & RF</i>	Menampilkan pendekatan ML modern

Berdasarkan tinjauan literatur pada Tabel 2.1, dapat diketahui bahwa penelitian mengenai prediksi hasil panen menggunakan pendekatan pembelajaran mesin (*machine learning*) telah mengalami perkembangan yang signifikan dalam beberapa tahun terakhir. Penelitian-penelitian tersebut tidak hanya fokus pada akurasi prediksi, tetapi juga pada pemilihan algoritma, fitur dominan, serta

pendekatan metodologis yang sesuai dengan karakteristik data pertanian yang kompleks.

Penelitian yang dilakukan oleh (Shawon et al., 2025) merupakan kajian literatur sistematis (*Systematic Literature Review/SLR*) yang membahas berbagai teknik dan algoritma pembelajaran mesin yang digunakan dalam prediksi hasil panen. Penelitian ini mengidentifikasi bahwa algoritma berbasis *ensemble* dan *boosting* merupakan pendekatan yang paling banyak digunakan dan memberikan kinerja yang baik dalam memodelkan hubungan antara variabel lingkungan dan hasil panen. Juga menyoroti bahwa fitur cuaca dan kondisi lingkungan merupakan faktor dominan dalam proses prediksi. Hasil penelitian ini memberikan gambaran tren penelitian terkini serta menjadi landasan teoritis yang kuat dalam pemilihan pendekatan *machine learning* pada penelitian ini.

Menggunakan metode *regresi linier* untuk memprediksi hasil panen padi dan menunjukkan bahwa pendekatan statistik klasik masih memiliki kemampuan yang cukup baik, ditunjukkan dengan pencapaian nilai koefisien *determinasi* (R^2) yang tinggi. Meskipun metode regresi linier memiliki keterbatasan dalam menangani hubungan nonlinier, penelitian ini penting sebagai referensi pembandingan terhadap metode pembelajaran mesin yang lebih kompleks. Dengan adanya penelitian ini, dapat terlihat perbedaan karakteristik dan kinerja antara metode statistik konvensional dan metode pembelajaran mesin (Diyanti et al., 2023).

Secara khusus meneliti penggunaan algoritma *CatBoost* dalam prediksi hasil panen dan membandingkannya dengan beberapa algoritma machine learning lainnya. Hasil penelitian menunjukkan bahwa *CatBoost* mampu memberikan kinerja yang unggul pada berbagai metrik evaluasi, terutama dalam menangani data

dengan karakteristik kompleks dan heterogen. Keunggulan ini disebabkan oleh kemampuan *CatBoost* dalam menangani fitur kategorikal dan mengurangi bias prediksi. Temuan tersebut menegaskan bahwa *CatBoost* merupakan algoritma yang potensial untuk digunakan dalam penelitian prediksi hasil panen, sehingga relevan dengan penelitian ini yang menjadikan *CatBoost* sebagai salah satu algoritma utama (Razavi et al., 2024).

Menerapkan algoritma *Random Forest Regressor* untuk memprediksi produksi padi pada skala lokal. Hasil penelitian menunjukkan bahwa model *Random Forest* mampu menghasilkan nilai *Mean Squared Error (MSE)* yang rendah serta nilai R^2 yang tinggi, yang mengindikasikan kemampuan model dalam memodelkan hubungan nonlinier antara variabel lingkungan dan hasil panen. Penelitian ini menunjukkan bahwa algoritma ensemble berbasis pohon keputusan sangat sesuai untuk data pertanian yang memiliki variabilitas tinggi. Temuan ini memperkuat penggunaan *machine learning* pada konteks pertanian lokal dan relevan dengan penelitian ini sebagai referensi metodologis (Manurung et al., 2025).

Melakukan penelitian komparatif antara algoritma *Random Forest* dan *XGBoost* dalam prediksi hasil panen. Hasil penelitian menunjukkan bahwa *Random Forest* sedikit lebih unggul dibandingkan *XGBoost*, meskipun kedua algoritma tersebut sama-sama memberikan performa yang kompetitif. Penelitian ini memberikan wawasan penting mengenai karakteristik algoritma serta menunjukkan bahwa pemilihan algoritma yang tepat sangat bergantung pada karakteristik dataset yang digunakan. Pendekatan komparatif yang dilakukan dalam penelitian ini

sejalan dengan tujuan penelitian ini, yaitu membandingkan kinerja beberapa algoritma pembelajaran mesin (Narendra Bayutama Wibisono, 2025).

Pentingnya pemilihan fitur dalam prediksi hasil panen padi dengan memanfaatkan variabel spektral dan tekstural. Penelitian ini menggunakan beberapa metode pembelajaran mesin, seperti *Regresi Linier Berganda*, *Regresi Vektor Dukungan*, dan *Hutan Acak*, untuk memberikan pengaruh fitur-fitur tersebut terhadap hasil prediksi. Hasil penelitian menunjukkan bahwa pemanfaatan fitur yang tepat dapat meningkatkan akurasi model secara signifikan. Penelitian ini mencerminkan pendekatan pembelajaran mesin modern dan menegaskan bahwa kualitas fitur memiliki peran penting selain algoritma pemilihan (Quille-Mamani et al., 2025).

Berdasarkan uraian penelitian terdahulu tersebut, dapat disimpulkan bahwa sebagian besar penelitian sebelumnya fokus pada penggunaan satu algoritma atau perbandingan algoritma yang berbeda dalam konteks prediksi hasil panen. Namun, penelitian yang secara khusus membandingkan algoritma *CatBoost* dengan *Extra Trees Classifier* dalam memprediksi tingkat keberhasilan panen padi berdasarkan faktor cuaca dan kondisi tanah masih relatif terbatas. Oleh karena itu, penelitian ini diharapkan dapat mengisi celah penelitian (*research gap*) tersebut dan memberikan kontribusi ilmiah berupa rekomendasi algoritma yang paling sesuai untuk prediksi keberhasilan panen padi.

2.2 Tinjauan Umum Penelitian Di Bidang Prediksi Hasil Panen

Tinjauan umum penelitian di bidang prediksi hasil panen bertujuan untuk memberikan gambaran menyeluruh mengenai arah, perkembangan, dan pendekatan ilmiah yang digunakan dalam mengkaji permasalahan ketidakpastian hasil

pertanian. Prediksi hasil panen menjadi isu penting karena berperan strategis dalam mendukung ketahanan pangan, perencanaan distribusi hasil pertanian, serta pengambilan kebijakan di sektor agrikultur. Ketepatan prediksi hasil panen dapat membantu berbagai pihak, mulai dari petani, pemerintah, hingga pelaku industri pertanian, dalam merancang strategi produksi dan mitigasi risiko secara lebih efektif.

Dalam praktiknya, prediksi hasil panen dipengaruhi oleh berbagai faktor yang bersifat kompleks dan saling berkaitan. Faktor lingkungan seperti cuaca dan kondisi tanah berinteraksi dengan faktor biologis tanaman serta faktor teknis budidaya. Kompleksitas hubungan antar faktor tersebut menyebabkan hasil panen sulit diprediksi secara akurat apabila hanya mengandalkan pendekatan konvensional atau pengalaman empiris semata. Oleh karena itu, penelitian di bidang prediksi hasil panen terus berkembang untuk mencari metode yang mampu merepresentasikan kondisi nyata di lapangan secara lebih komprehensif (Azhari & Nasution, 2025).

Seiring dengan perkembangan teknologi informasi, pendekatan berbasis data mulai banyak digunakan dalam penelitian prediksi hasil panen. Data historis yang mencakup variabel cuaca, kondisi tanah, dan hasil panen dimanfaatkan untuk membangun model prediksi yang lebih sistematis. Pendekatan ini memungkinkan peneliti untuk mengidentifikasi pola dan kecenderungan yang tidak mudah diamati secara langsung. Dengan memanfaatkan data dalam jumlah besar dan beragam, penelitian prediksi hasil panen dapat menghasilkan model yang lebih adaptif terhadap perubahan kondisi lingkungan.

Penelitian-penelitian terkini menunjukkan bahwa penggunaan metode komputasi cerdas, khususnya machine learning, memberikan kontribusi signifikan

dalam meningkatkan akurasi prediksi hasil panen (Shawon et al., 2025). *Machine learning* memiliki kemampuan untuk mempelajari hubungan nonlinier antar variabel dan menyesuaikan model berdasarkan karakteristik data yang digunakan. Oleh karena itu, tinjauan umum ini menjadi landasan penting dalam memahami posisi penelitian yang dilakukan, sekaligus mengarahkan pembahasan pada subbab selanjutnya yang mengkaji perkembangan teknologi, permasalahan prediksi, serta pendekatan algoritmik yang digunakan dalam penelitian ini.

2.2.1 Perkembangan Teknologi Di Bidang Pertanian

Perkembangan teknologi informasi dan komputasi telah membawa perubahan yang signifikan dalam sektor pertanian (Shawon et al., 2025). Pertanian yang sebelumnya sangat bergantung pada pengalaman empiris dan intuisi petani kini mulai beralih ke pendekatan berbasis data dan teknologi digital. Pemanfaatan data cuaca, kondisi tanah, serta histori hasil panen memungkinkan proses perencanaan pertanian dilakukan secara lebih terukur dan sistematis.

Kemajuan teknologi sensor, sistem informasi geografis, serta ketersediaan data historis dalam jumlah besar mendorong penggunaan pendekatan analitik untuk memahami dinamika pertanian. Teknologi ini memungkinkan pengumpulan data secara kontinu dan akurat, sehingga dapat digunakan sebagai dasar dalam pengambilan keputusan. Dalam konteks ini, teknologi komputasi cerdas seperti *machine learning* menjadi salah satu solusi yang potensial untuk mengolah data pertanian yang kompleks (Azhari & Rahman, 2022).

Penerapan teknologi berbasis data diharapkan mampu meningkatkan efisiensi dan produktivitas sektor pertanian. Dengan adanya sistem prediksi yang andal, petani dan pemangku kepentingan dapat melakukan perencanaan tanam,

pengelolaan sumber daya, serta mitigasi risiko gagal panen secara lebih optimal. Selain itu, pemanfaatan teknologi komputasi cerdas dalam pertanian juga membuka peluang integrasi antara data historis dan data real-time. Integrasi ini memungkinkan proses pengambilan keputusan dilakukan secara adaptif sesuai dengan kondisi lapangan yang terus berubah. Dengan dukungan analisis berbasis *machine learning*, informasi yang dihasilkan tidak hanya bersifat deskriptif, tetapi juga prediktif, sehingga dapat membantu meningkatkan ketahanan dan keberlanjutan sistem pertanian di masa depan.

2.2.2 Permasalahan Prediksi Keberhasilan Panen Padi

Permasalahan prediksi keberhasilan panen padi merupakan isu yang kompleks karena melibatkan banyak variabel yang saling berinteraksi dan bersifat dinamis. Faktor cuaca seperti curah hujan, suhu udara, dan kelembapan berubah secara temporal, sementara kondisi tanah dapat berbeda antar wilayah meskipun berada dalam satu daerah administrasi yang sama. Kompleksitas ini menyebabkan pendekatan prediksi konvensional yang berbasis pengalaman atau metode statistik sederhana sering kali tidak mampu memberikan hasil yang akurat dan konsisten.

Selain faktor lingkungan, keberhasilan panen padi juga dipengaruhi oleh ketidakpastian eksternal, seperti perubahan iklim dan anomali cuaca ekstrem. Kondisi tersebut dapat menyebabkan pergeseran pola tanam dan meningkatkan risiko gagal panen. Oleh karena itu, diperlukan pendekatan prediksi yang mampu mempelajari pola nonlinier dari data. Pendekatan *machine learning* dinilai relevan untuk mengatasi permasalahan tersebut karena memiliki kemampuan dalam mengolah data berskala besar dan kompleks. Beberapa penelitian menunjukkan bahwa model *machine learning* mampu menangkap hubungan nonlinier antara

faktor cuaca, kondisi tanah, dan hasil panen dengan tingkat akurasi yang lebih baik dibandingkan metode konvensional (Yunis et al., 2024).

2.3 Konsep Dasar *Machine Learning*

Konsep dasar *machine learning* menjadi landasan utama dalam penelitian ini karena pendekatan yang digunakan untuk memprediksi tingkat keberhasilan panen padi sepenuhnya bergantung pada kemampuan model dalam mempelajari pola dari data historis. *Machine learning* merupakan cabang dari kecerdasan buatan yang berfokus pada pengembangan algoritma yang memungkinkan sistem komputer untuk belajar dari data dan meningkatkan kinerjanya seiring dengan bertambahnya pengalaman, tanpa harus diprogram secara eksplisit untuk setiap kemungkinan kondisi.

Konsep dasar *machine learning* menjadi landasan utama dalam penelitian ini karena pendekatan yang digunakan untuk memprediksi tingkat keberhasilan panen padi sepenuhnya bergantung pada kemampuan model dalam mempelajari pola dari data historis. *Machine learning* merupakan cabang dari kecerdasan buatan yang berfokus pada pengembangan algoritma yang memungkinkan sistem komputer untuk belajar dari data dan meningkatkan kinerjanya seiring dengan bertambahnya pengalaman, tanpa harus diprogram secara eksplisit untuk setiap kemungkinan kondisi.

Dalam konteks penelitian ilmiah, *machine learning* dipandang sebagai pendekatan yang efektif untuk menangani permasalahan dengan karakteristik data yang kompleks, berdimensi tinggi, dan memiliki hubungan nonlinier antar variabel. Data pertanian, khususnya data cuaca dan kondisi tanah, memiliki karakteristik tersebut sehingga pendekatan *machine learning* dinilai lebih sesuai dibandingkan

metode statistik konvensional. Beberapa penelitian mutakhir menunjukkan bahwa machine learning mampu memberikan akurasi prediksi yang lebih baik dalam bidang pertanian karena kemampuannya memodelkan hubungan yang kompleks antar variabel (Shawon et al., 2025).

2.3.1 Pengertian *Machine Learning*

Machine learning dapat didefinisikan sebagai suatu metode komputasi yang memungkinkan sistem belajar secara otomatis dari data dengan tujuan untuk melakukan prediksi atau pengambilan keputusan. Proses pembelajaran dalam *machine learning* dilakukan dengan cara mengidentifikasi pola, keterkaitan, serta struktur tersembunyi dalam data historis, kemudian memanfaatkan pola tersebut untuk memprediksi data baru.

Pendekatan ini berbeda dengan pemrograman tradisional yang mengandalkan aturan eksplisit. Dalam *machine learning*, model dibangun berdasarkan data sehingga kualitas dan karakteristik data sangat memengaruhi performa model yang dihasilkan. Oleh karena itu, pemilihan algoritma dan pengolahan data menjadi faktor penting dalam penerapan *machine learning*.

Dalam penelitian-penelitian terkini di bidang pertanian, *machine learning* dipahami sebagai pendekatan prediktif yang mampu memodelkan hubungan kompleks antara faktor lingkungan dan hasil panen. Selain itu, menegaskan bahwa *machine learning* berperan penting dalam meningkatkan ketahanan sistem pertanian terhadap perubahan iklim melalui prediksi hasil panen yang lebih akurat dan adaptif (Islam et al., 2024).

2.3.2 Jenis-Jenis *Machine Learning*

Secara umum, machine learning diklasifikasikan ke dalam tiga jenis utama

yaitu *supervised learning*, *unsupervised learning*, dan *reinforcement learning*. *Supervised learning* merupakan metode pembelajaran yang menggunakan data berlabel, di mana setiap data input memiliki target atau keluaran yang jelas. Pendekatan ini banyak digunakan dalam permasalahan prediksi karena model dilatih untuk memetakan hubungan antara variabel input dan output.

Unsupervised learning merupakan metode pembelajaran tanpa label yang bertujuan untuk menemukan pola atau struktur tersembunyi dalam data, seperti pengelompokan dan reduksi dimensi. Sementara itu, *reinforcement learning* merupakan metode pembelajaran yang berfokus pada interaksi antara agen dan lingkungan melalui mekanisme umpan balik berupa penghargaan dan hukuman. Ketiga pendekatan ini memiliki karakteristik dan tujuan yang berbeda, sehingga pemilihannya harus disesuaikan dengan jenis permasalahan yang dihadapi.

Dalam konteks prediksi hasil panen dan klasifikasi tingkat keberhasilan panen, *supervised learning* merupakan pendekatan yang paling relevan karena data yang digunakan umumnya memiliki label yang jelas. menunjukkan bahwa sebagian besar penelitian prediksi hasil panen menggunakan *supervised learning* karena kemampuannya menghasilkan model yang stabil dan mudah dievaluasi (Islam et al., 2024).

Dalam penelitian ini, pendekatan *supervised learning* digunakan karena data yang tersedia memiliki label berupa tingkat keberhasilan panen padi. Penggunaan *supervised learning* memungkinkan model untuk secara langsung mempelajari hubungan antara faktor cuaca, kondisi tanah, dan hasil panen, sehingga hasil prediksi yang dihasilkan dapat dievaluasi secara kuantitatif dan objektif. Secara umum, machine learning diklasifikasikan ke dalam tiga jenis utama, yaitu

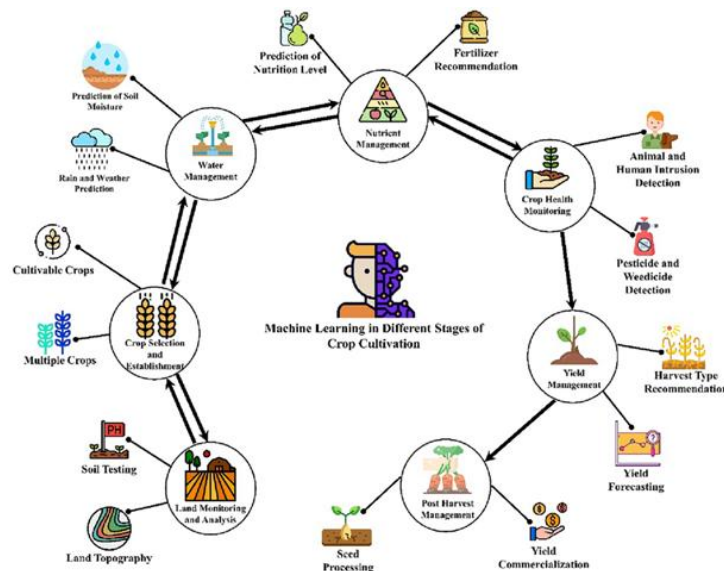
supervised learning, *unsupervised learning*, dan *reinforcement learning*. *Supervised learning* merupakan metode pembelajaran yang menggunakan data berlabel, di mana setiap data input memiliki target atau keluaran yang jelas. Pendekatan ini banyak digunakan dalam permasalahan prediksi karena model dilatih untuk memetakan hubungan antara variabel input dan output.

Unsupervised learning merupakan metode pembelajaran tanpa label yang bertujuan untuk menemukan pola atau struktur tersembunyi dalam data, seperti pengelompokan atau reduksi dimensi. Sementara itu, *reinforcement learning* merupakan metode pembelajaran berbasis interaksi dengan lingkungan melalui mekanisme *reward dan punishment*. Dalam penelitian ini, pendekatan supervised learning digunakan karena data yang tersedia memiliki label berupa tingkat keberhasilan panen padi, sehingga sesuai untuk diterapkan pada algoritma klasifikasi.

2.3.3 Penerapan *Machine Learning* Di Pertanian

Tahap awal dalam kerangka konsep tersebut adalah pengumpulan data (data acquisition) . Data yang dikumpulkan umumnya berasal dari berbagai sumber, seperti data cuaca (curah hujan, suhu udara, dan kelembaban), data kondisi tanah (pH tanah, kandungan unsur hara, tekstur tanah), serta data historis hasil panen. Data ini merepresentasikan kondisi lingkungan dan agronomi yang mempengaruhi pertumbuhan tanaman. kualitas dan kelengkapan data lingkungan menjadi faktor krusial dalam keberhasilan penerapan pembelajaran mesin pada sektor pertanian karena data tersebut berfungsi sebagai dasar model pembelajaran prediktif (Islam et al., 2024).

Setelah data terkumpul, tahapan berikutnya adalah pra-pemrosesan data (*data preprocessing*). Pada tahap ini, data dibersihkan dari nilai yang hilang (*missing value*), data pencilan (*outliers*), serta dilakukan transformasi dan normalisasi agar data siap digunakan oleh algoritma *machine learning*. Proses *preprocessing* bertujuan untuk meningkatkan kualitas data sehingga model yang dihasilkan memiliki kinerja yang lebih stabil dan akurat. menegaskan bahwa tahapan *preprocessing* memiliki pengaruh yang signifikan terhadap kinerja model prediksi hasil panen, terutama ketika data bersifat heterogen dan berasal dari berbagai sumber (Mugemangango et al., 2024).



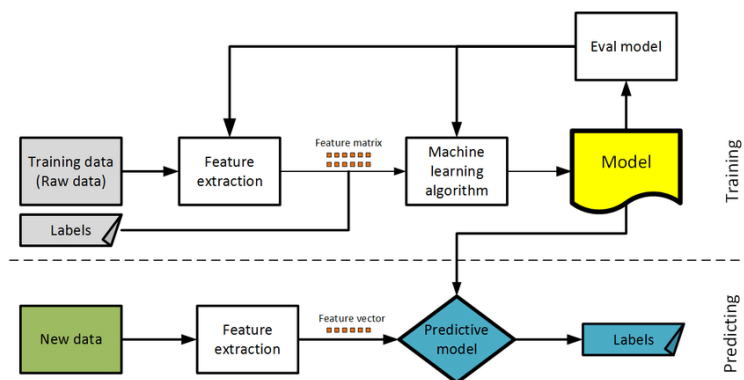
Gambar 2.1 Teknik Pengumpulan Data

Tahap selanjutnya dalam kerangka konsep tersebut adalah ekstraksi dan pemilihan fitur (fitur ekstraksi dan seleksi fitur). Pada tahap ini, variabel-variabel yang relevan dipilih untuk mewakili karakteristik data secara optimal. Contoh fitur yang digunakan antara lain curah hujan, suhu, kandungan *nitrogen (N)*, *fosfor (P)*, *kalium (K)*, serta tekstur tanah. Pemilihan fitur yang tepat dapat membantu model pembelajaran mesin dalam memahami hubungan kompleks antara kondisi

lingkungan dan hasil panen. menyatakan bahwa pemilihan fitur yang relevan dapat meningkatkan akurasi prediksi dan mengurangi kompleksitas model (Jhaharia et al., 2025).

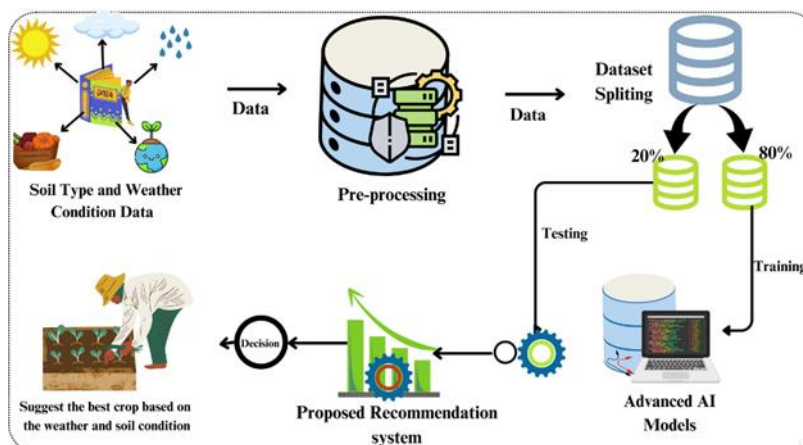
Setelah fitur ditentukan, data kemudian diproses menggunakan *algoritma machine learning*. Pada tahap ini, model dibangun dengan memanfaatkan data latih (training data) untuk mempelajari pola hubungan antara variabel input dan output. Berbagai algoritma pembelajaran mesin dapat digunakan, seperti *Random Forest, Support Vector Machine, Gradient Boosting, CatBoost, dan Extra Trees*. Model yang telah dibangun kemudian diuji menggunakan uji data (testing data) untuk memunculkan kemampuannya dalam melakukan prediksi terhadap data yang belum pernah dilihat sebelumnya. menjelaskan bahwa algoritma berbasis ensemble sering kali memberikan kinerja yang lebih baik dalam prediksi hasil panen karena mampu menangani hubungan nonlinier antar variabel (Islam et al., 2024).

Tahap berikutnya adalah evaluasi model, yang bertujuan untuk mengukur kinerja model pembelajaran mesin. Evaluasi dilakukan menggunakan metrik tertentu, seperti akurasi, presisi, *recall*, dan *F1-score* untuk kasus klasifikasi, atau *Mean Squared Error (MSE)* dan R^2 untuk kasus regresi. Evaluasi ini penting untuk memastikan bahwa model yang digunakan memiliki tingkat kebisingan yang memadai sebelum diterapkan dalam sistem nyata. Model evaluasi yang komprehensif juga membantu dalam memilih model yang tidak hanya akurat tetapi juga stabil dan dapat diinterpretasikan (Petropoulos et al., 2025).



Gambar 2.2 Kerangka Konsep

Tahap akhir dalam kerangka konsep adalah hasil prediksi dan rekomendasi. Output dari model *machine learning* dapat berupa prediksi tingkat keberhasilan panen, estimasi hasil produksi, atau rekomendasi tindakan pertanian tertentu, seperti penyesuaian waktu tanam atau penggunaan input agronomi. Hasil ini diharapkan dapat membantu petani dan pemangku kepentingan dalam mengambil keputusan yang lebih tepat dan berbasis data. Penerapan kerangka ini sejalan dengan konsep pertanian presisi, di mana teknologi digunakan untuk meningkatkan efisiensi dan kehausan sistem pertanian (Saha et al., 2025).



Gambar 2.3 Hasil Prediksi dan rekomendasi

secara keseluruhan merepresentasikan alur penerapan *machine learning* dalam pertanian yang terstruktur dan sistematis, mulai dari pengumpulan data

hingga prediksi yang dihasilkannya. Kerangka ini menjadi dasar konsep dalam penelitian ini untuk menerapkan dan membandingkan pembelajaran mesin algoritma dalam memprediksi tingkat keberhasilan panen padi berdasarkan faktor cuaca dan kondisi tanah.

2.4 Prediksi Dan Klasifikasi

Prediksi dan klasifikasi merupakan dua konsep fundamental dalam *machine learning* yang sangat relevan dalam konteks pemodelan prediktif pada data pertanian. Prediksi dalam pembelajaran mesin merujuk pada kemampuan sebuah model untuk meramalkan atau memprediksi nilai keluaran (output) berdasarkan data input yang diberikan. Sedangkan klasifikasi adalah sebuah bentuk prediksi di mana hasil keluaran berupa kelas atau kategori, bukan nilai numerik kontinu. Dalam penelitian terkait keberhasilan panen padi, klasifikasi yang dipilih karena target yang ingin dicapai adalah tingkat keberhasilan panen dalam beberapa kelas, misalnya “rendah”, “sedang”, dan “tinggi”. Konsep ini sangat penting karena pendekatan klasifikasi memungkinkan peneliti dan praktisi pertanian untuk membuat keputusan berdasarkan kategori risiko dan tingkat prestasi tanaman, yang kemudian dapat digunakan dalam strategi manajemen pertanian secara lebih efektif.

Implementasi prediksi menggunakan pembelajaran mesin tidak hanya sederhana dengan menerapkan algoritma, tetapi juga melalui proses belajar dari data historis agar pola-pola hubungan antar variabel dapat dikenal. Misalnya, model dapat mempelajari hubungan antara variabel input seperti suhu udara, curah hujan, kandungan unsur hara tanah, serta pola tanam yang dilakukan petani, kemudian memetakan hubungan tersebut ke dalam hasil panen yang dirinci. Menurut penelitian yang berfokus pada estimasi hasil panen di berbagai lingkungan

pertanian, pembelajaran mesin mampu memberikan representasi hubungan kompleks antar variabel yang tidak mudah ditangkap metode statistik tradisional (Jabed & Azmi Murad, 2024).

2.4.1 Dasar Prediksi Dengan *Machine Learning*

Prediksi dengan *machine learning* memanfaatkan teknik statistik dan algoritma komputasi untuk menggeneralisasi pola dari data pelatihan ke data baru yang belum pernah dilihat sebelumnya. Pendekatan ini berbeda dengan metode deterministik karena hasilnya tidak hanya berdasarkan formula baku, melainkan berdasarkan kemampuan model untuk “belajar” pola data yang tersedia.

Dalam banyak penelitian yang fokus pada prediksi hasil panen, berbagai model seperti *Random Forest*, *Gradient Boosting*, *Support Vector Machine (SVM)*, dan *deep learning* digunakan untuk memodelkan hubungan kompleks antara variabel input dan output hasil panen. Implementasi model prediktif ini memungkinkan sistem pertanian untuk mengantisipasi hasil produksi dengan mempertimbangkan dinamika faktor lingkungan dan internal tanaman (Jabed & Azmi Murad, 2024).

Selain itu, penelitian berbasis penginderaan jauh dan citra satelit juga menunjukkan bahwa prediksi hasil panen dapat diperkuat dengan data spasial dan temporal yang lebih luas. Analisis sistematis terhadap penelitian-penelitian estimasi hasil berdasarkan data *UAV (Unmanned Aerial Vehicle)* yang dikombinasikan dengan algoritma pembelajaran mesin menunjukkan bahwa pendekatan ini efektif dalam memetakan fase pertumbuhan tanaman dan menghasilkan prediksi yang lebih akurat (Yuan et al., 2024).

2.4.2 Dasar Klasifikasi Dalam *Machine Learning*

Klasifikasi adalah cabang dari pembelajaran mesin di mana model dibor untuk menentukan kelas atau kategori sebuah input berdasarkan label data yang tersedia. Dalam kasus prediksi keberhasilan panen padi, klasifikasi memiliki peran penting karena output yang dihasilkan berbentuk kelas tingkat keberhasilan panen, bukan nilai numerik. Klasifikasi algoritma seperti *Decision Tree*, *Random Forest*, *k-Nearest Neighbor (k-NN)*, dan *SVM* sering digunakan karena mekanisme pengelompokan yang kuat dan kemampuan mereka menangani dataset dengan banyak variabel. Penggunaan klasifikasi dalam konteks pertanian juga terbukti efektif dalam tugas rekomendasi jenis tanaman terbaik berdasarkan kondisi lingkungan lokal (Gopi & Karthikeyan, 2023).

Sebuah penelitian lain yang mengeksplorasi arsitektur klasifikasi dalam konteks pertanian menegaskan bahwa klasifikasi model optimasi tidak hanya meningkatkan akurasi prediksi, tetapi juga membantu dalam pengambilan keputusan operasional di lapangan, seperti penentuan pola irigasi dan rekomendasi nutrisi tanaman. Hal ini menunjukkan bahwa klasifikasi memiliki manfaat praktis pada sistem pertanian yang bersifat berbasis keputusan sistem pendukung Keputusan (Badshah et al., 2024).

2.4.3 Peran *Machine Learning* Dalam Prediksi Dan Klasifikasi Di Pertanian

Dalam beberapa tahun terakhir, perkembangan teknik *machine learning* telah memperluas kemampuan prediksi dan klasifikasi di sektor pertanian dengan integrasi data yang luas, termasuk data iklim, tanah, citra satelit, dan faktor agronomi lainnya. Di antaranya, pendekatan pembelajaran ensemble seperti *Random Forest* dan *Gradient Boosting* sering digunakan karena kemampuannya

dalam menangani hubungan nonlinier antar variabel serta resistensinya terhadap overfitting. Algoritma ini tidak hanya mampu memprediksi hasil panen secara numerik, tetapi juga dikombinasikan dengan metode klasifikasi untuk menghasilkan kategori hasil yang dapat dengan mudah diinterpretasikan oleh pengguna akhir (Azhari et al., 2021).

Secara empiris, prediksi yang dilakukan menggunakan pembelajaran mesin membantu pengambilan keputusan strategi, seperti menentukan waktu tanam yang optimal atau mengestimasi risiko gagal panen. Sistem-sistem prediksi ini juga umumnya dilengkapi dengan metrik evaluasi seperti akurasi, *presisi*, *recall*, dan *F1-score* untuk klasifikasi, serta R^2 atau *Mean Absolute Error (MAE)* untuk prediksi numerik, sehingga hasil model dapat diukur secara tujuan sebelum diterapkan dalam praktik pertanian.

Rumus nya Adalah :

$$Accuracy = \frac{Jumlah\ Prediksi\ Benar}{Jumlah\ Seluruh\ Data} \quad (2.1)$$

$$Precision = \frac{TP}{TP+FP} \quad (2.2)$$

$$Recall = \frac{TP}{TP+FN} \quad (2.3)$$

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.4)$$

Keterangan Variabel:

1. TP : True Positive
2. TN : True Negative
3. FP : False Positive
4. FN : False Negative
5. TP_i, FP_i, FN_i : nilai pada kelas ke- i

6. n_i : jumlah data kelas ke- i
7. N : total data
8. k : jumlah kelas

2.5 Algoritma *CatBoost*

Algoritma *CatBoost* merupakan salah satu metode ensemble learning berbasis gradient boosting yang dirancang untuk meningkatkan kinerja prediksi pada data dengan karakteristik kompleks dan heterogen. Dalam beberapa tahun terakhir, *CatBoost* semakin banyak digunakan dalam penelitian prediksi di berbagai domain, termasuk pertanian, karena kemampuannya dalam menangani hubungan nonlinier antar variabel serta stabilitas performanya pada dataset yang mengandung kombinasi fitur numerik dan kategorikal.

Pada konteks prediksi keberhasilan panen padi, *CatBoost* dinilai relevan karena data yang digunakan umumnya berasal dari berbagai sumber, seperti data cuaca, kondisi tanah, dan praktik budidaya, yang memiliki karakteristik beragam. Beberapa penelitian terkini menunjukkan bahwa algoritma boosting, termasuk *CatBoost*, mampu memberikan performa yang unggul dibandingkan metode konvensional dalam memodelkan data pertanian yang kompleks (Islam et al., 2024).

2.5.1 Konsep Dasar Dan Mekanisme Kerja *CatBoost*

CatBoost bekerja dengan membangun model secara bertahap melalui proses *boosting*, di mana setiap pohon keputusan baru dilatih untuk memperbaiki kesalahan prediksi dari model sebelumnya. Proses ini dilakukan dengan meminimalkan fungsi kerugian (*loss function*) secara iteratif sehingga performa

model meningkat secara bertahap. Pendekatan ini memungkinkan *CatBoost* untuk mempelajari pola yang kompleks dan nonlinier dalam data.

Salah satu karakteristik utama *CatBoost* adalah kemampuannya dalam menangani fitur kategorikal secara internal tanpa memerlukan proses *encoding* manual yang kompleks. Mekanisme ini membantu mengurangi potensi bias dan kehilangan informasi yang sering terjadi pada teknik *encoding* konvensional. Studi yang membandingkan berbagai algoritma *gradient boosting* dalam prediksi hasil panen menunjukkan bahwa pendekatan seperti *CatBoost* memberikan kestabilan model yang lebih baik ketika data memiliki keragaman fitur yang tinggi (Mugemangango et al., 2024).

Selain itu, *CatBoost* menerapkan strategi pelatihan yang dirancang untuk meningkatkan kemampuan generalisasi model, khususnya pada data yang memiliki distribusi tidak seimbang atau mengandung *noise*. Melalui pengaturan struktur pohon dan pembaruan model yang terkontrol pada setiap iterasi, *CatBoost* mampu menekan risiko *overfitting* tanpa mengorbankan akurasi. Dalam konteks data pertanian yang dipengaruhi oleh variasi cuaca dan kondisi tanah, mekanisme ini memungkinkan *CatBoost* menghasilkan prediksi yang lebih stabil dan konsisten dibandingkan beberapa metode *boosting* lainnya.

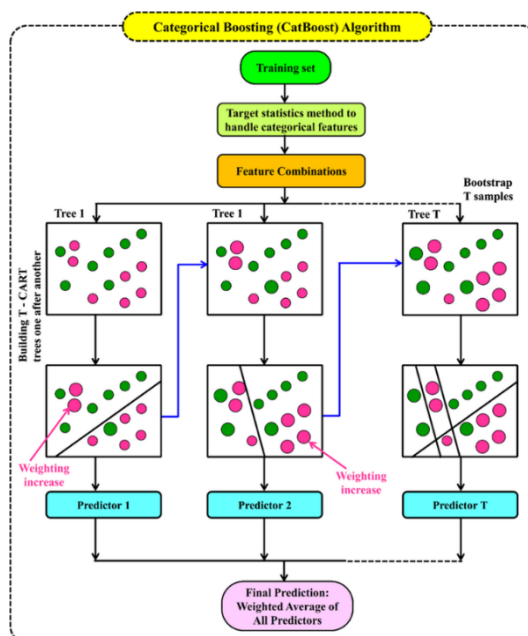
Penelitian terkini di bidang pertanian menunjukkan bahwa algoritma *gradient boosting* dengan mekanisme pembelajaran yang terkontrol, seperti *CatBoost*, memiliki keunggulan dalam memodelkan hubungan nonlinier antar variabel lingkungan. (Islam et al., 2024) melaporkan bahwa pendekatan *boosting* yang stabil mampu meningkatkan akurasi prediksi hasil panen pada data iklim dan tanah yang kompleks serta bervariasi. Temuan tersebut memperkuat alasan pemilihan *CatBoost*

sebagai algoritma yang sesuai untuk prediksi tingkat keberhasilan panen padi dalam penelitian ini.

$$\text{Rumus CatBoost(Gradient Boosting): } F_m(x) = F_{m-1}(x) + \alpha \cdot h_m(x)$$

Keterangan :

1. $F_m(x)$: model pada iterasi ke- m
2. $F_{m-1}(x)$: model sebelumnya
3. $h_m(x)$: pohon keputusan ke- m
4. α : *learning rate*



Gambar 2.4 Cara Kerja CatBoost

Gambar tersebut menggambarkan cara kerja algoritma *CatBoost* sebagai metode pembelajaran ensemble berbasis peningkatan gradien. Pada ilustrasi tersebut terlihat bahwa model *CatBoost* dibangun dari sekelompok pohon keputusan yang disusun. Diagram alur *CatBoost* juga menunjukkan proses pembelajaran yang dimulai dari input data, kemudian melalui tahapan pelatihan bertahap (*boosting iterations*), hingga menghasilkan model akhir dengan tingkat

akurasi yang lebih tinggi. Karakteristik utama *CatBoost* yang tergambar dalam ilustrasi tersebut ada. Dalam konteks prediksi keberhasilan panen padi, visualisasi ini membantu menjelaskan bagaimana *CatBoost* memanfaatkan data cuaca dan kondisi tanah untuk membangun model prediktif yang kuat. Setiap keputusan pohon berkontribusi dalam menangkap pola tertentu dari data lingkungan, sehingga hasil akhir berupa prediksi tingkat keberhasilan panen dapat dihasilkan secara lebih akurat dan konsisten.

2.5.2 Karakteristik Algoritma *CatBoost*

Karakteristik utama *CatBoost* terletak pada pendekatan pembelajaran yang terkontrol dan sistematis. Algoritma ini membangun pohon keputusan secara berurutan dengan kedalaman yang relatif terbatas untuk mencegah terjadinya *overfitting*. Selain itu, *CatBoost* dirancang untuk bekerja secara efisien pada dataset berukuran sedang hingga besar dengan tingkat variabilitas data yang tinggi.

Dalam penelitian pertanian, karakteristik ini sangat penting karena data cuaca dan kondisi tanah sering kali bersifat fluktuatif dan tidak linier. Penelitian terbaru menunjukkan bahwa *CatBoost* mampu mempertahankan kinerja prediksi yang stabil meskipun dihadapkan pada variasi data lingkungan yang signifikan (Petropoulos et al., 2025).

Keunggulan utama *CatBoost* dibandingkan algoritma lain terletak pada kemampuannya menghasilkan prediksi yang akurat dengan kebutuhan *feature engineering* yang relatif lebih rendah. Selain itu, *CatBoost* memiliki mekanisme internal untuk mengurangi *overfitting* dan meningkatkan generalisasi model. Keunggulan ini sangat relevan dalam penelitian prediksi keberhasilan panen padi yang menggunakan data historis dengan variasi kondisi lingkungan.

2.5.3 Keunggulan Algoritma *CatBoost*

Keunggulan utama *CatBoost* dibandingkan algoritma lain terletak pada kemampuannya menghasilkan prediksi yang akurat dengan kebutuhan *feature engineering* yang relatif lebih rendah. Selain itu, *CatBoost* memiliki mekanisme internal untuk mengurangi overfitting dan meningkatkan generalisasi model. Keunggulan ini sangat relevan dalam penelitian prediksi keberhasilan panen padi yang menggunakan data historis dengan variasi kondisi lingkungan.

Beberapa penelitian aplikasi *machine learning* di bidang pertanian melaporkan bahwa algoritma berbasis *boosting*, termasuk *CatBoost*, memberikan akurasi yang lebih tinggi dibandingkan algoritma berbasis pohon keputusan tunggal atau metode statistik klasik (Saha et al., 2025). Dengan demikian, *CatBoost* dipilih dalam penelitian ini sebagai salah satu algoritma utama untuk dibandingkan dengan *Extra Trees Classifier* dalam memprediksi tingkat keberhasilan panen padi.

2.6 Algoritma *Extra Trees Classifier*

Algoritma *Extra Trees Classifier* atau *Extremely Randomized Trees* merupakan salah satu metode *ensemble learning* berbasis pohon keputusan yang dikembangkan sebagai pengembangan dari algoritma *Random Forest*. *Extra Trees Classifier* membangun sejumlah besar pohon keputusan secara paralel dan menggabungkan hasil prediksinya untuk menghasilkan keputusan akhir. Pendekatan ensemble ini bertujuan untuk meningkatkan akurasi prediksi serta mengurangi risiko overfitting yang sering terjadi pada model pohon keputusan tunggal.

Dalam konteks prediksi keberhasilan panen padi, *Extra Trees Classifier* relevan digunakan karena mampu menangani data berdimensi tinggi yang berasal

dari berbagai variabel lingkungan, seperti faktor cuaca dan kondisi tanah. Penelitian terkini menunjukkan bahwa algoritma berbasis pohon keputusan dan ensemble memiliki performa yang baik dalam memodelkan hubungan nonlinier antar variabel lingkungan pada sektor pertanian (Mugemangango et al., 2024).

2.6.1 Konsep Dasar Dan Mekanisme Kerja *Extra Trees Classifier*

Extra Trees Classifier bekerja dengan membangun banyak pohon keputusan menggunakan tingkat randomisasi yang lebih tinggi dibandingkan *Random Forest*. Pada setiap pohon, pemilihan fitur dan penentuan titik pemisah (split) dilakukan secara acak tanpa mencari split terbaik secara mendalam. Pendekatan ini menghasilkan keragaman yang tinggi antar pohon, sehingga meningkatkan kemampuan generalisasi model.

Proses pelatihan *Extra Trees Classifier* dilakukan secara paralel, di mana setiap pohon keputusan dilatih menggunakan seluruh dataset pelatihan. Hasil prediksi dari masing-masing pohon kemudian digabungkan melalui mekanisme voting untuk menentukan kelas akhir. Mekanisme ini memungkinkan *Extra Trees Classifier* menghasilkan prediksi yang stabil meskipun data mengandung noise atau variasi yang tinggi.

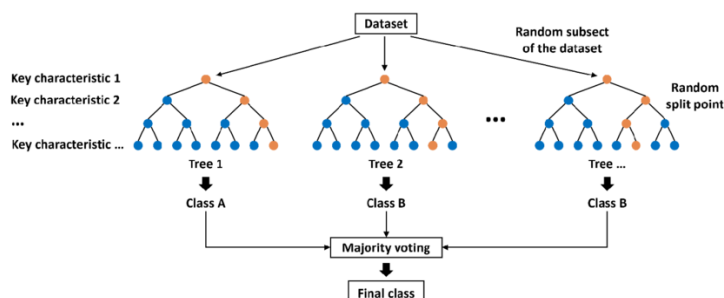
Selain efisiensi komputasi dan kemampuan menangani data berdimensi tinggi, penelitian lain juga menunjukkan bahwa *Extra Trees Regressor* mampu memberikan kinerja prediksi yang sangat baik dalam kasus prediksi hasil panen tanaman. Menemukan bahwa model *Extra Trees* menghasilkan skor R^2 yang lebih tinggi serta MAE dan $RMSE$ yang lebih rendah dibandingkan model lain dalam memprediksi produksi tanaman dengan berbagai variabel lingkungan (cuaca, kondisi tanah) sebagai input. Temuan ini menunjukkan bahwa pendekatan

pengacakan tinggi pada pembentukan keputusan pohon dalam *Extra Trees* dapat menangkap keterkaitan kompleks antar variabel secara efektif, sehingga model ini dapat menjadi algoritma yang berguna sebagai pembanding dalam penelitian ini (Nikhil et al., 2024).

$$\text{Rumus } Extra\ Trees\ (Voting\ Ensemble): \hat{y} = mode\{h_1(x), h_2(x), \dots, h_n(x)\}$$

Keterangan :

1. \hat{y} : hasil prediksi akhir
2. $h_i(x)$: prediksi dari pohon keputusan ke- i
3. n : jumlah pohon keputusan



Gambar 2.5 Cara Kerja *Extra Trees Classifier*

2.6.2 Karakteristik Algoritma *Extra Classifier*

Karakteristik utama *Extra Trees Classifier* terletak pada tingkat randomisasi yang tinggi dalam proses pembentukan pohon keputusan. Randomisasi ini mencakup pemilihan fitur dan nilai threshold secara acak, yang bertujuan untuk meningkatkan keberagaman model dalam ensemble. Selain itu, *Extra Trees Classifier* memiliki kecepatan pelatihan yang relatif tinggi karena tidak melakukan pencarian split optimal secara mendalam.

Dalam penelitian pertanian, karakteristik ini menjadi keunggulan karena data lingkungan seperti curah hujan dan suhu udara sering kali bersifat fluktuatif. Studi terbaru menunjukkan bahwa *Extra Trees Classifier* mampu mempertahankan

performa yang stabil pada dataset lingkungan yang kompleks dan tidak terstruktur dengan baik (Islam et al., 2024).

2.6.3 Keunggulan Algoritma *Extra Classifier*

Keunggulan utama *Extra Trees Classifier* adalah efisiensi komputasi dan kemampuannya dalam menangani dataset berukuran besar dengan tingkat variabilitas yang tinggi. Tingkat randomisasi yang tinggi pada pemilihan fitur dan penentuan titik pemisah (split) membantu mengurangi korelasi antar pohon keputusan sehingga risiko *overfitting* dapat ditekan secara signifikan. Pendekatan ini menjadikan *Extra Trees Classifier* sangat sesuai untuk data lingkungan yang bersifat fluktuatif, seperti data cuaca dan kondisi tanah. Selain itu, *Extra Trees Classifier* mampu memberikan performa prediksi yang kompetitif dengan waktu pelatihan yang relatif lebih singkat dibandingkan metode *boosting*. *Extra Trees*, memiliki kinerja yang stabil dalam memodelkan hubungan nonlinier antara faktor lingkungan dan hasil panen padi. Stabilitas ini penting dalam konteks pertanian karena data sering mengandung noise dan pencilan (Mugemangango et al., 2024).

2.7. Faktor Lingkungan Yang Mempengaruhi Panen Padi

Faktor lingkungan memberikan pengaruh yang sangat besar terhadap pertumbuhan dan hasil panen padi karena tanaman padi sangat *sensitif* terhadap kondisi fisik dan kimia lingkungan tempat tumbuhnya. Komponen utama yang sering dianalisis dalam penelitian terkait prediksi hasil panen meliputi variabel cuaca seperti curah hujan, suhu udara, kelembapan, dan intensitas radiasi matahari. Variasi pada komponen-komponen ini dapat mempengaruhi proses fisiologis tanaman, termasuk fotosintesis, transpirasi, dan pembentukan biji padi. Penelitian prediksi hasil panen yang mempertimbangkan variabel cuaca dan karakteristik

tanah menunjukkan bahwa penggabungan kedua jenis variabel tersebut memberikan hasil model yang lebih akurat dan representatif terhadap kenyataan lapangan (Himayanta & Wardhani, 2025).

Selain variabel cuaca, kondisi tanah juga merupakan penentu penting dalam keberhasilan produksi padi karena tanah menyediakan unsur hara, udara, dan dukungan fisik bagi tanaman. Parameter tanah seperti pH tanah, kandungan *nitrogen, fosfor, dan kalium*, serta tekstur tanah sangat mempengaruhi kemampuan tanaman menyerap nutrisi secara efektif. Kondisi tanah yang suboptimal seperti pH yang ekstrem atau rendahnya kandungan unsur hara dapat menghambat proses pertumbuhan dan akhirnya menurunkan hasil panen padi meskipun kondisi cuaca mendukung. Penelitian yang menggunakan pendekatan *machine learning* dalam prediksi hasil panen telah menunjukkan bahwa memasukkan variabel tanah bersama variabel cuaca secara signifikan meningkatkan kinerja model prediksi dibandingkan penggunaan satu jenis variabel saja (Himayanta & Wardhani, 2025).

Integrasi faktor cuaca dan kondisi tanah ke dalam model prediksi berbasis *machine learning* memungkinkan pemodelan hubungan nonlinier yang kompleks antara lingkungan dan hasil panen. Pendekatan ini sangat penting terutama di tengah tantangan perubahan iklim yang menyebabkan cuaca ekstrem dan perubahan pola curah hujan. Oleh karena itu, variabel lingkungan seperti cuaca dan tanah sebaiknya diintegrasikan ke dalam model pembelajaran mesin untuk meningkatkan akurasi prediksi keberhasilan panen yang lebih realistis dan aplikatif.

2.7.1 Faktor Cuaca

Faktor cuaca merupakan aspek lingkungan yang sangat dominan dalam mempengaruhi pertumbuhan dan hasil panen padi karena variabel-variabel iklim

secara langsung mempengaruhi proses fisiologis tanaman seperti fotosintesis, transpirasi, pembentukan bunga, dan pengisian gabah. Variabel cuaca yang umum digunakan dalam model prediksi hasil panen antara lain curah hujan, suhu udara maksimum dan minimum, kelembaban, serta intensitas radiasi matahari.

variabel-variabel iklim ini dapat digunakan dalam model *machine learning* untuk memperkirakan perubahan hasil panen di masa depan berdasarkan proyeksi iklim historis dan perubahan pola iklim. Studi tersebut mengintegrasikan data curah hujan, suhu, dan hasil panen untuk memancarkan dampak perubahan iklim terhadap produksi padi dengan menggunakan beberapa teknik pemodelan, termasuk jaringan saraf *feed-forward*, dan membuktikan bahwa cuaca memiliki kontribusi yang signifikan terhadap variasi hasil panen di kawasan pertanian (Tasneem et al., 2025).

Selain itu, variabilitas cuaca seperti perubahan pola curah hujan dan suhu ekstrem sering kali menimbulkan tekanan fisiologis pada tanaman padi. Curah hujan yang tidak seimbang baik terlalu sedikit maupun berlebihan dapat menyebabkan stress udara, gangguan hidrologi tanah, serta risiko banjir atau kekeringan yang berdampak pada penurunan produktivitas tanaman. Penelitian lain menunjukkan bahwa suhu udara yang meningkat di atas ambang optimal tanaman padi dapat mempercepat fase pemasakan sehingga mengurangi *biomassa* dan jumlah gabah yang terbentuk, yang pada akhirnya menurunkan hasil panen (Purwaningrum et al., 2025).

Lebih jauh lagi, penelitian empiris mengenai dampak iklim terhadap produksi padi telah mencatat bahwa kombinasi variabel cuaca dapat merusak kondisi pertanian jika tidak diantisipasi dengan strategi adaptasi yang tepat. Misalnya, curah hujan yang menurun dikombinasikan dengan suhu udara tinggi akan

meningkatkan kemungkinan tanaman mengalami stres musim panas dan kekeringan, sedangkan curah hujan tinggi yang tidak diimbangi oleh drainase yang baik dapat menyebabkan oksigen tanah rendah dan gangguan perakaran. Hasil kajian ini menekankan pentingnya memasukkan faktor cuaca secara holistik dalam model prediksi hasil panen agar mampu mencerminkan interaksi kompleks antara variabel lingkungan dan produktivitas tanaman padi secara realistis (Masahid et al., 2025).

2.7.2 Faktor Tanah

kondisi tanah mempunyai peranan yang sangat penting dalam menentukan keberhasilan panen padi karena tanah berfungsi sebagai media tumbuh utama sekaligus penyedia unsur hara dan udara bagi tanaman. Parameter tanah seperti pH dan kandungan unsur hara makro (*nitrogen, fosfor, dan kalium*) secara langsung mempengaruhi kemampuan tanaman dalam menyerap nutrisi yang dibutuhkan selama fase pertumbuhan.

Selain unsur kimia tanah, faktor fisik seperti tekstur dan kelembaban tanah juga berpengaruh signifikan terhadap perkembangan sistem perakaran tanaman padi. Tanah dengan tekstur yang sesuai mampu menahan udara secara optimal sekaligus menyediakan aerasi yang cukup bagi akar tanaman. Studi berbasis *machine learning* dalam prediksi hasil panen menunjukkan bahwa parameter tanah fisik dan kimia yang dikombinasikan dapat meningkatkan akurasi model prediksi dibandingkan penggunaan variabel cuaca saja (Mugemangango et al., 2024).

Dalam penelitian prediksi keberhasilan panen padi berbasis *machine learning*, kondisi tanah tidak hanya dipandang sebagai variabel statis, tetapi juga sebagai faktor yang berinteraksi dengan dinamika iklim dan indikator vegetasi.

Pendekatan multi-sumber yang menggabungkan variabel iklim, penginderaan jauh, dan properti tanah dilaporkan mampu meningkatkan ketepatan pemodelan karena model dapat menangkap hubungan nonlinier dan efek gabungan antar faktor lingkungan. Studi terkini yang memanfaatkan data multi-sumber untuk pemodelan hasil padi menegaskan bahwa properti tanah merupakan salah satu komponen penting yang perlu dipertimbangkan bersama variabel iklim dalam membangun prediksi yang lebih akurat dan aplikatif (Saha et al., 2025).

BAB III

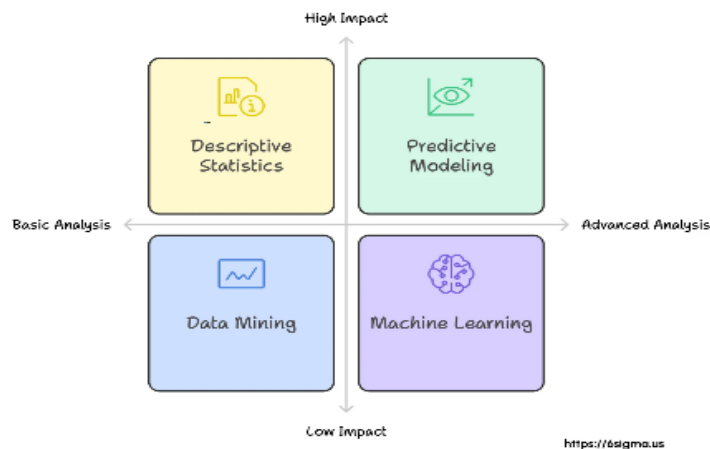
METODOLOGI PENELITIAN

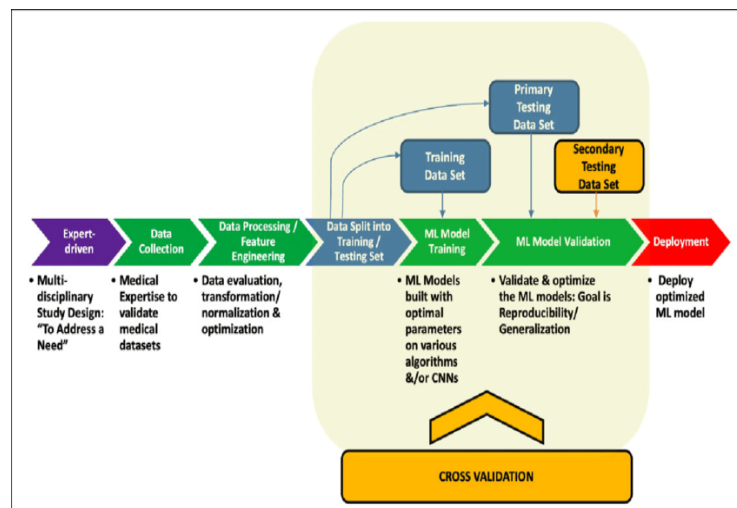
3.1 Desain Dan Jenis Penelitian

Penelitian ini merupakan penelitian kuantitatif dengan pendekatan *data-driven* menggunakan *supervised machine learning* untuk memprediksi tingkat keberhasilan panen padi berdasarkan faktor cuaca dan kondisi tanah. Desain penelitian diarahkan pada pembangunan dan pengujian model klasifikasi, kemudian membandingkan performa dua algoritma *ensemble* yang berbeda pendekatan, yaitu *CatBoost* dan *Extra Trees Classifier*. Pendekatan ini dipilih karena prediksi hasil panen dipengaruhi hubungan nonlinier antar faktor lingkungan, sehingga membutuhkan model yang mampu mempelajari pola kompleks dari data historis. Selain itu, validasi dan evaluasi model dilakukan secara sistematis agar hasil penelitian dapat dipertanggung jawabkan secara ilmiah dan relevan untuk konteks pertanian berbasis data (Filippi et al., 2025).



Quantitative Data Analysis





Gambar 3.1 *supervised machine learning*

Gambar menunjukkan desain penelitian kuantitatif dengan pendekatan *supervised machine learning* yang digunakan dalam penelitian ini. Penelitian dirancang untuk membangun dan membandingkan dua algoritma pembelajaran ansambel, yaitu *CatBoost* dan *Extra Trees Classifier*, dalam memprediksi tingkat keberhasilan panen padi. Desain ini menekankan pada penggunaan data historis sebagai dasar model pembelajaran serta model evaluasi menggunakan metrik klasifikasi. Pendekatan komparatif dipilih untuk mengidentifikasi algoritma yang memiliki kinerja terbaik dalam menangani data cuaca dan kondisi tanah yang bersifat kompleks dan nonlinier.

3.2 Objek Dan Waktu Penelitian

Objek penelitian ini adalah prediksi tingkat keberhasilan panen padi (variabel target) menggunakan variabel lingkungan sebagai fitur, yang mencakup faktor cuaca dan kondisi tanah. Penelitian berfokus pada data historis yang merepresentasikan kondisi lapangan pada suatu wilayah/area budidaya padi yang tercakup di dalam dataset. Waktu penelitian mencakup tahapan pengumpulan data,

	Machine Learning												
6	Evaluasi model(ROC, Confusion Matrix)						✓						
7	Implementasi sstem web							✓					
8	Pengujian Sistem								✓	✓			
9	Penyusunan laporan skripsi										✓		
10	Revisi Skripsi												✓

3.3 Data Penelitian

Data yang digunakan pada penelitian ini berupa data sekunder yang bersifat historis, yaitu data yang telah tersedia dan dikumpulkan oleh lembaga atau sumber resmi sebelumnya. Dataset penelitian disusun dalam bentuk data simulasi yang merepresentasikan kondisi cuaca dan tanah pada periode 2024–2025. Rentang waktu ini dipilih untuk menggambarkan kondisi pertanian dalam beberapa musim tanam terakhir secara realistis. Dataset digunakan untuk melatih dan menguji model machine learning tanpa melibatkan pengambilan data lapangan secara langsung. Penggunaan data sekunder dipilih karena data lingkungan dan pertanian membutuhkan rentang waktu historis yang panjang agar pola hubungan antara faktor cuaca, kondisi tanah, dan keberhasilan panen dapat dianalisis secara akurat. Dataset memuat variabel cuaca (misalnya curah hujan, suhu, kelembapan) dan variabel kondisi tanah (misalnya pH, unsur hara, atau indikator terkait), serta label tingkat keberhasilan panen. Penggunaan data historis memungkinkan model

mempelajari keterkaitan antara kondisi lingkungan dan keberhasilan panen pada periode sebelumnya untuk menghasilkan prediksi pada kondisi baru.

Tabel 3.3 Rekapitulasi Data Sekunder

No	Jenis Data	Variabel	Sumber Resmi
1	Cuaca	Curah Hujan,Suhu,Kelmbapan	BMKG
2	Kondisi Tanah	pH,N,P,K	Kementerian Pertanian RI
3	Hasil Panen	Produksi dan produktivitas padi	BPS
4	Data Pendukung	Statistik Global	FAO

Menampilkan struktur dataset yang digunakan dalam penelitian ini. Dataset terdiri atas variabel independen berupa faktor cuaca (seperti curah hujan, suhu udara, dan kelembapan) serta kondisi tanah (seperti pH dan unsur hara), dan variabel dependen berupa tingkat keberhasilan panen padi. Struktur ini menunjukkan hubungan antara fitur masukan dan label target yang akan dipelajari oleh model klasifikasi. Penyajian struktur data secara visual memudahkan pemahaman mengenai komponen dataset yang digunakan dalam pengujian model.

3.3.1 Variabel Penelitian

Variabel dalam penelitian ini dibagi menjadi variabel independen (fitur) dan variabel dependen (label).

1. Variabel independen meliputi faktor cuaca dan kondisi tanah yang digunakan sebagai masukan model. Variabel-variabel ini dipilih karena literatur prediksi hasil panen menunjukkan bahwa kombinasi indikator iklim dan tanah secara konsisten menjadi faktor penentu utama dalam pemodelan hasil pertanian(Jabed & Azmi Murad, 2024).

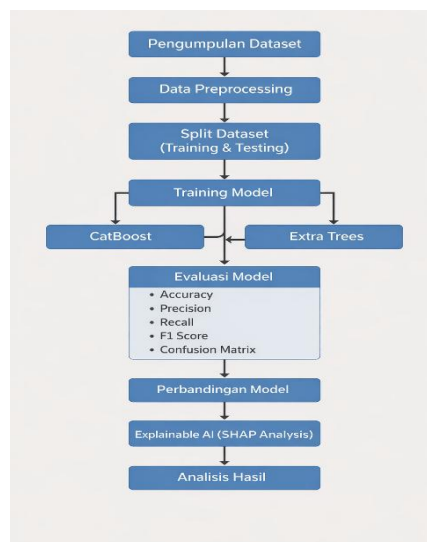
2. Variabel dependen adalah tingkat keberhasilan panen padi yang didefinisikan sebagai kelas (misalnya rendah/sedang/tinggi atau berhasil/tidakberhasil) sesuai kebutuhan penelitian.

3.3.2 Skema Kelas Tingkat Keberhasilan Panen

Penentuan skema kelas dilakukan untuk memastikan label target dapat dipelajari oleh model klasifikasi. Skema kelas dapat disusun berdasarkan aturan tertentu (misalnya ambang batas hasil/produktivitas atau indikator keberhasilan panen pada dataset). Penetapan kelas yang jelas membantu evaluasi model secara objektif melalui metrik klasifikasi.

3.4 Tahapan Penelitian

Tahapan penelitian disusun secara berurutan agar proses pembangunan model dapat direplikasi. Secara umum, tahapan penelitian mencakup: (1) pengumpulan dataset; (2) pemahaman data (*data understanding*) dan eksplorasi awal; (3) pembagian data latih dan data uji; (4) Training Data *CatBoost* dan *Extra Trees Classifier*; (5) Evaluasi Model; (6) Perbandingan Model; (7) SHAP Analysis; (8) analisis perbandingan dan interpretasi hasil.



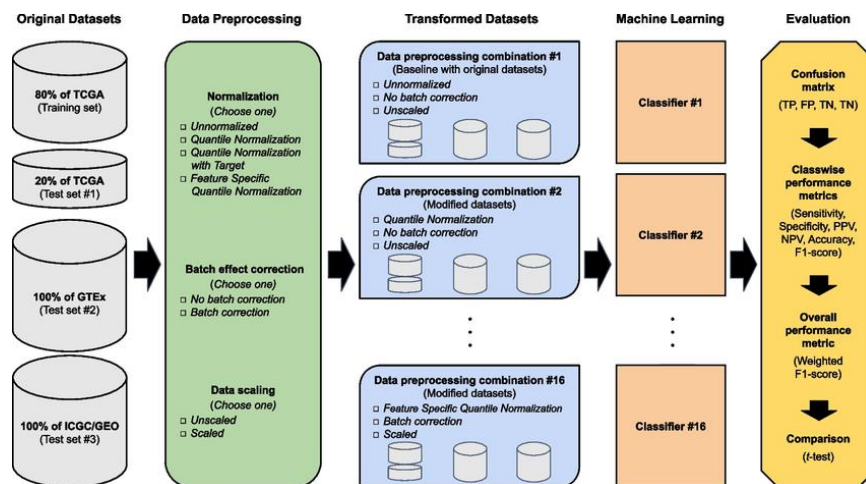
Gambar 3.2 Tahapan Penelitian

Menunjukkan tahapan penelitian yang dilakukan secara sistematis, dimulai dari pengumpulan data, eksplorasi dan prapemrosesan data, pembagian data latih dan data uji, pelatihan model *CatBoost* dan *Extra Trees Classifier*, hingga evaluasi dan analisis hasil. Setiap tahapan saling berkaitan dan dirancang untuk memastikan bahwa proses penelitian dapat direplikasi serta hasil yang diperoleh bersifat objektif. Visualisasi tahapan ini memberikan gambaran menyeluruh mengenai alur kerja penelitian dan menjadi pedoman pelaksanaan penelitian dari awal hingga akhir.

3.5 Prapemrosesan Data

Prapemrosesan dilakukan untuk meningkatkan kualitas data sehingga model yang dibangun lebih stabil dan akurat. Tahapan prapemrosesan pada penelitian ini meliputi:

1. Pembersihan data, seperti menghapus duplikasi dan memperbaiki format/tipe data.
2. Penanganan nilai hilang (missing values) menggunakan strategi yang sesuai karakter data.
3. Penanganan data pencilan (outlier) jika ditemukan nilai ekstrem yang tidak wajar.
4. *Transformasi fitur* (jika diperlukan), misalnya normalisasi/standarisasi untuk fitur tertentu.
5. *Encoding* data kategorikal bila terdapat fitur kategorikal, dengan mempertimbangkan karakteristik algoritma. *CatBoost* dapat menangani fitur kategorikal secara internal, sedangkan *Extra Trees* umumnya memerlukan representasi numerik (Elbeltagi et al., 2025).



Gambar 3.3 Tahapan Prapemrosesan

Tahapan prapemrosesan data yang dilakukan sebelum pelatihan model. Tahapan ini mencakup pembersihan data, penanganan nilai hilang, penanganan outlier, serta transformasi data agar sesuai dengan kebutuhan algoritma. Proses prapemrosesan sangat penting karena kualitas data sangat menentukan performa model *machine learning*, khususnya pada data lingkungan yang cenderung bervariasi dan mengandung noise (Filippi et al., 2025).

3.6 Pembagian Data Dan Strategi Validasi

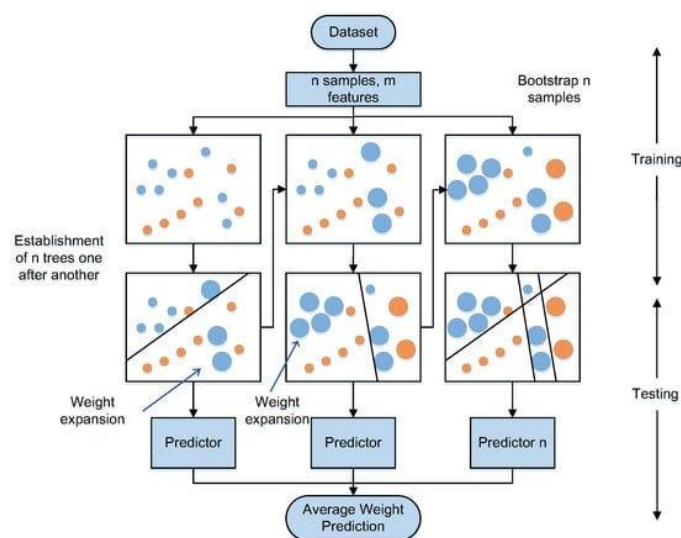
Dataset dibagi menjadi data latih (*training set*) dan data uji (*testing set*) untuk menilai kemampuan generalisasi model. Pembagian dilakukan dengan rasio tertentu (misalnya 80:20 atau 70:30) dan, bila memungkinkan, menggunakan stratified split agar proporsi kelas pada data latih dan uji tetap seimbang. Selain pembagian latih–uji, penelitian ini menggunakan strategi validasi seperti k-fold cross-validation untuk memastikan evaluasi lebih robust. Pemilihan strategi validasi yang diperhatikan karena penelitian hasil panen berbasis data iklim menunjukkan bahwa strategi *cross validation* dapat memengaruhi hasil performa dan interpretasi model (Sweet et al., 2023).

3.7 Pembangunan Model Dan Penalaan Parameter

Penalaan parameter dilakukan untuk memperoleh konfigurasi model terbaik. Metode penalaan dapat berupa *grid search* atau *random search* yang dikombinasikan dengan *cross-validation*. Penalaan dilakukan pada kedua algoritma dengan ruang parameter yang sebanding agar perbandingan lebih objektif. Praktik penalaan ini sejalan dengan rekomendasi metodologis dalam pemodelan prediksi hasil panen, di mana pemilihan konfigurasi dan strategi validasi memengaruhi kualitas prediksi (Filippi et al., 2025).

3.7.1 Model *CatBoost*

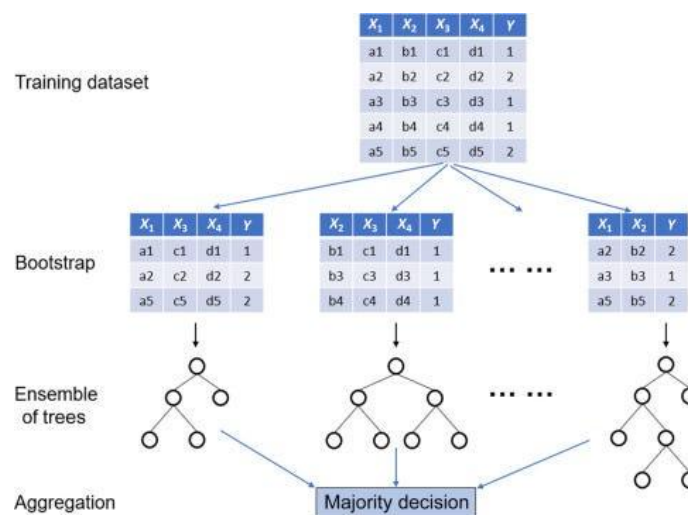
Model *CatBoost* dibangun sebagai metode *gradient boosting* berbasis pohon keputusan. Model dilatih menggunakan data latih untuk mempelajari hubungan antara fitur lingkungan dan label keberhasilan panen. *CatBoost* relevan digunakan pada data pertanian karena mampu memodelkan hubungan nonlinier dan berpotensi stabil ketika menghadapi *noise* pada data lingkungan (Razavi et al., 2024). Parameter penting yang diperhatikan pada pelatihan antara lain learning rate, jumlah iterasi/trees, kedalaman pohon, serta regularisasi.



Gambar 3.4 Proses Model *CatBoost*

3.7.2 Model Extra Trees Classifier

Model *Extra Trees Classifier* dibangun dengan membangkitkan banyak pohon keputusan secara paralel dengan randomisasi tinggi pada pemilihan fitur dan titik split. Model ini cenderung efisien secara komputasi dan kuat untuk fitur multivariat. Dalam aplikasi prediksi hasil pertanian, pendekatan *Extra Trees* dilaporkan efektif untuk mengurangi varians model dan menghindari overfitting pada data kompleks (Nuraini et al., 2025).



Gambar 3.5 Model *Extra Trees Classifier*

3.8 Metode Evaluasi Model

Evaluasi model dilakukan menggunakan metrik klasifikasi untuk menilai performa prediksi tingkat keberhasilan panen. Metrik yang digunakan meliputi:

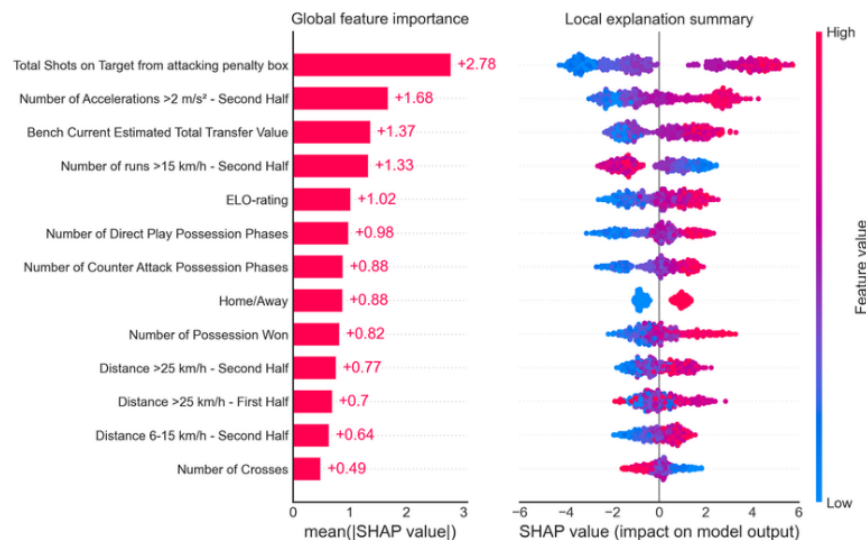
1. *Confusion matrix* untuk melihat distribusi benar–salah pada setiap kelas.
2. *Accuracy* untuk mengukur ketepatan prediksi keseluruhan.
3. *Precision* untuk mengukur ketepatan prediksi pada kelas tertentu.
4. *Recall* untuk mengukur kemampuan model menangkap kelas yang benar.

5. *F1-score* sebagai ukuran keseimbangan antara precision dan recall.

Metrik tersebut umum digunakan pada studi prediksi pertanian karena memberikan gambaran performa model secara lebih menyeluruh, terutama ketika data kelas tidak seimbang atau ketika kesalahan pada kelas tertentu lebih kritis (Jabed & Azmi Murad, 2024).

3.9 Interpretasi Model Dan Analisis Hasil

Selain mengevaluasi performa, penelitian ini juga melakukan analisis interpretasi untuk memahami faktor apa yang paling memengaruhi prediksi. Interpretabilitas penting dalam penerapan machine learning di pertanian karena membantu pemangku kepentingan memahami alasan di balik prediksi model. Pendekatan explainable AI seperti SHAP banyak digunakan untuk menjelaskan kontribusi fitur (misalnya curah hujan, suhu, pH tanah, unsur hara) terhadap output model (Yenkikar et al., 2025). Analisis ini mendukung penyusunan rekomendasi yang lebih informatif.



Gambar 3.6 Model *SHAP*

menunjukkan ilustrasi interpretasi model menggunakan pendekatan

feature importance atau SHAP. Pendekatan ini digunakan untuk mengetahui kontribusi masing-masing variabel terhadap hasil prediksi. Interpretabilitas model sangat penting dalam penerapan machine learning di bidang pertanian karena membantu pemangku kepentingan memahami faktor lingkungan yang paling berpengaruh terhadap keberhasilan panen (Yenkikar et al., 2025).

3.10 Perangkat Lunak Dan Implementasi

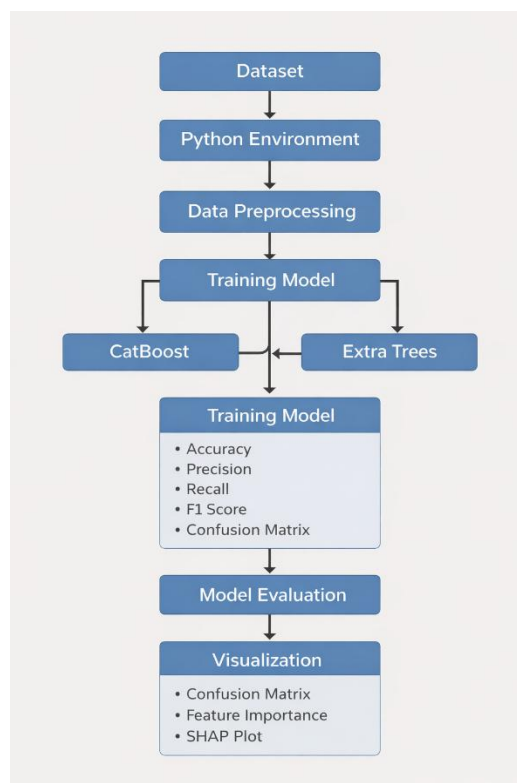
Implementasi penelitian dilakukan menggunakan perangkat lunak pemrograman dan pustaka machine learning yang sesuai (misalnya *Python* dan ekosistemnya). Proses meliputi pemuatan dataset, prapemrosesan, pelatihan model, penalaan parameter, evaluasi, serta visualisasi hasil evaluasi dan interpretasi fitur. Penggunaan perangkat lunak yang terdokumentasi dengan baik mendukung replikasi penelitian.

Tabel 3.4 Peran Library Perangkat Lunak

No	Library	Peran Utama
1	<i>Python 3.11</i>	Bahasa Pemograman Utama
2	<i>Pandas</i>	Manajemen dan Manipulasi Data
3	<i>Numpy</i>	Operasi Numerik Dan Komputasi
4	<i>CatBoost</i>	Implementasi algoritma CatBoost
5	<i>Scikit-Learn</i>	<i>Extra Trees</i> , Evaluasi dan Validasi
6	<i>Matplotlib, Seaborn</i>	Visualisasi

Implementasi penelitian ini menggunakan dua komponen utama yaitu sistem machine learning berbasis Python untuk proses pemodelan dan sistem aplikasi web berbasis Laravel sebagai antarmuka pengguna (dashboard). Sistem web digunakan

untuk menampilkan hasil prediksi, visualisasi grafik, serta analisis model seperti confusion matrix, feature importance, dan SHAP plot. Sistem dilakukan dengan beberapa tahapan sebagai berikut: (1). Membuat dataset menggunakan library pandas; (2). Melakukan *Preprocessing* data seperti pembersihan data dan transformasi fitur; (3). Membagi dataset menjadi *training* data dan *testing* data; (4). Melatih model menggunakan algoritma *Catboost* dan *Extra Trees Classifier*; (5). Menguji model menggunakan data testing; (6). Menghitung metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*; (7). Membuat visualisasi hasil evaluasi menggunakan *confusion matrix* dan grafik *feature importance*; (8). Menggunakan metode SHAP untuk menjelaskan kontribusi setiap fitur terhadap hasil prediksi.



Gambar 3.7 Representasi Diagram

Penelitian ini menghasilkan prototipe sistem prediksi keberhasilan panen padi berbasis machine learning yang diimplementasikan menggunakan bahasa pemrograman Python. Sistem ini memanfaatkan berbagai pustaka pemrosesan data dan machine learning seperti Pandas, NumPy, CatBoost, dan Scikit-learn untuk melakukan pengolahan data, pembangunan model, evaluasi performa, serta visualisasi hasil analisis. Prototipe sistem ini dirancang untuk menerima data faktor cuaca dan kondisi tanah sebagai input, kemudian menghasilkan prediksi tingkat keberhasilan panen padi sebagai output yang dapat digunakan sebagai sistem pendukung pengambilan keputusan di bidang pertanian.

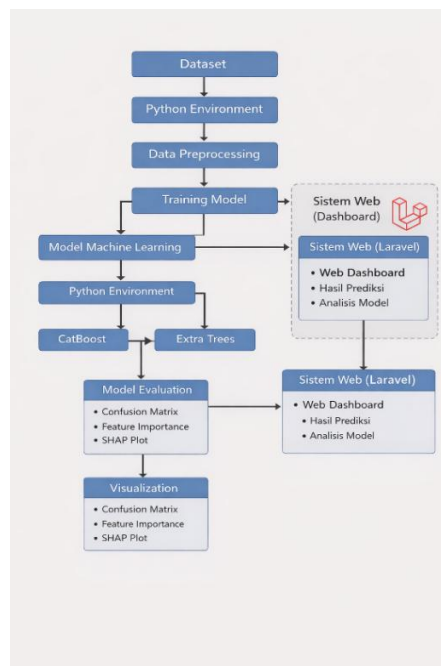
BAB IV

HASIL DAN PEMBAHASAN

4.1 Implementasi Sistem

Implementasi sistem pada penelitian ini dilakukan dengan mengintegrasikan model *machine learning* yang dibangun menggunakan bahasa pemrograman *Python* dengan sistem antarmuka berbasis web dashboard. Model *machine learning* digunakan untuk memproses dataset yang terdiri dari variabel lingkungan seperti curah hujan, suhu udara, kelembapan, pH tanah, nitrogen, fosfor, dan kalium.

Sistem yang dikembangkan bertujuan untuk memprediksi tingkat keberhasilan panen padi berdasarkan faktor lingkungan tersebut. Output sistem berupa hasil klasifikasi tingkat keberhasilan panen yang dibagi menjadi tiga kategori yaitu rendah, sedang, dan tinggi. Implementasi sistem dilakukan melalui beberapa tahapan utama, yaitu: (1). Membuat dataset menggunakan *library pandas*; (2). Melakukan *preprocessing* data untuk meningkatkan kualitas data; (3). Membagi dataset menjadi data latih dan data uji; (4). Melatih model menggunakan algoritma *CatBoost* dan *Extra Trees Classifier*; (5). Mengevaluasi performa model menggunakan metrik klasifikasi; (6). Menampilkan hasil analisis dalam bentuk visualisasi grafik dan dashboard.



Gambar 4.1 Implementasi Sistem

Penggunaan sistem berbasis web memungkinkan hasil analisis model machine learning ditampilkan secara lebih interaktif dan mudah dipahami oleh pengguna. Menurut penelitian yang dilakukan oleh (Shawon et al., 2025), penerapan machine learning dalam bidang pertanian dapat meningkatkan akurasi prediksi hasil panen dengan memanfaatkan data lingkungan secara lebih efektif.

4.2 Implementasi Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari data lingkungan dan kondisi tanah yang mempengaruhi pertumbuhan tanaman padi. Dataset tersebut memuat beberapa variabel yang digunakan sebagai fitur input model machine learning serta satu variabel target yang menunjukkan tingkat keberhasilan panen padi. Variabel input yang digunakan meliputi curah hujan, suhu udara, kelembapan udara, pH tanah, kandungan nitrogen, kandungan fosfor, dan kandungan kalium. Variabel target dalam dataset ini adalah kelas keberhasilan panen padi yang dikategorikan menjadi tiga kelas yaitu rendah, sedang, dan tinggi.

Contoh struktur dataset ditunjukkan pada tabel berikut.

Tabel 4.4 Struktur Dataset

N O	curah hujan(mm)	suhu(^o C)	kelembapa n(%)	ph tanah	nitrogen (N)	fosfor (p)	kalium (K)	kelas_pa nen
1	182	30.4	89	5.74	20.3	14.1	21.3	Rendah
2	259	31.8	63	6.01	38.3	22.8	22.9	Sedang
3	172	30.7	84	6.7	33.2	26.6	15.9	Rendah
4	94	30.9	88	5.15	22.1	28	40.3	Sedang
5	186	27.3	86	5.46	38	18.2	43.4	Tinggi
6	151	28.4	57	6.4	30.2	22	42.9	Sedang
7	268	26	57	5.18	34.1	33.4	19.5	Rendah
8	100	25.6	82	5.14	48.2	38.2	46.6	Tinggi
9	182	28	69	5.65	40.9	35.9	38.5	Tinggi
10	201	28.8	77	5.17	25	26.9	46.4	Tinggi
11	290	26.7	65	6.05	58.5	22.3	30.7	Tinggi
12	294	28.6	86	5.41	45.4	20	30.3	Tinggi
13	154	31.1	87	6.61	50.6	13.7	42.5	Sedang
14	282	28.5	78	6.41	23.2	13.2	18.3	Sedang
15	167	29.8	88	6.26	21.9	30.7	33.2	Sedang
16	196	30.4	82	6.11	20.8	17.8	44.3	Sedang
17	179	31.9	61	6.55	27.6	17.6	44.1	Rendah
18	183	28.8	58	5.4	36.6	39.9	48.5	Sedang
19	231	30.5	55	6.57	57.4	15.6	39.4	Sedang
20	210	31.7	55	6.42	38.3	14.4	22.3	Rendah
21	229	31.6	74	6.35	31.6	13	15.7	Sedang
22	132	25.1	73	6.01	42.5	27.4	31.8	Tinggi

23	81	27.3	83	5.44	24.5	28.8	33	Sedang
24	167	26.6	64	5.76	44.4	21	23.4	Sedang
25	237	24.7	72	6.43	42.7	26.1	19.4	Tinggi
26	117	29.1	82	5.06	51.7	33.7	35.5	Sedang
27	209	29.9	89	6.41	59.5	34.5	41.7	Tinggi
28	271	30.8	85	5.43	30	38.7	49.8	Sedang
29	267	25	81	5.26	40.4	34.7	41.5	Tinggi
30	100	31	65	5.21	56.4	26.9	42.8	Sedang
31	240	29.1	59	5.05	51.4	15.4	36.6	Sedang
32	283	29.6	62	5.94	39.5	13	49.5	Rendah
33	137	31.3	88	5.39	57.8	11.6	37.4	Sedang
34	101	29	66	6.78	22.3	30	45.1	Sedang
35	168	26.7	55	6.03	57.2	34.2	48.1	Tinggi
36	128	30.6	85	5.25	22	25.3	47.5	Sedang
37	298	26.9	58	6.24	41.9	11.6	28.3	Sedang
38	138	24.3	59	6.62	55.6	26.1	17.8	Rendah
39	249	30.6	84	6.34	30.5	38.1	21	Sedang
40	299	26.8	78	5.14	38.9	27.9	17.1	Sedang
41	267	30.2	57	6.61	33	34.3	18.7	Rendah
42	287	26.9	65	6.43	58.1	34	21.5	Sedang

Penggunaan dataset dengan variabel lingkungan dan kondisi tanah bertujuan untuk menangkap hubungan antara faktor-faktor alam yang mempengaruhi pertumbuhan tanaman padi dengan tingkat keberhasilan panen. Data tersebut kemudian digunakan sebagai dasar dalam proses pembelajaran model machine learning sehingga model dapat mengenali pola hubungan antara variabel input dan

hasil panen. Dataset yang digunakan dalam penelitian ini terdiri dari sekitar 1500 data yang diperoleh dari sumber data sekunder dan telah melalui proses validasi sebelum digunakan dalam penelitian.

Menurut penelitian yang dilakukan oleh (Himayanta & Wardhani, 2025), penggunaan dataset lingkungan yang mencakup faktor cuaca dan kondisi tanah dapat meningkatkan performa model machine learning dalam memprediksi hasil panen tanaman karena variabel-variabel tersebut memiliki hubungan yang kuat dengan pertumbuhan tanaman.

4.3 Hasil preprocessing data

Tahap *preprocessing* data merupakan salah satu tahapan penting dalam proses pengembangan model machine learning karena kualitas data yang digunakan akan sangat mempengaruhi performa model yang dihasilkan. Pada tahap ini dilakukan beberapa proses pengolahan data untuk memastikan bahwa dataset yang digunakan berada dalam kondisi yang optimal sebelum digunakan dalam proses pelatihan model.

Namun pada penelitian ini dataset yang digunakan merupakan dataset rule-based yang telah disiapkan sebelumnya sehingga tidak ditemukan data duplikat, nilai hilang, maupun *outlier* yang signifikan. Oleh karena itu tahapan preprocessing dilakukan secara minimal sebelum proses pelatihan model *machine learning*.

Tahap *preprocessing* juga bertujuan untuk memastikan bahwa setiap variabel dalam dataset memiliki format yang konsisten sehingga dapat diproses dengan baik oleh algoritma *machine learning*. Menurut (Dinh et al., 2025), preprocessing data merupakan tahap yang sangat penting dalam pipeline machine

learning karena data yang tidak bersih dapat menyebabkan model menghasilkan prediksi yang tidak akurat.

4.4 Pembagian Dataset

Pembagian dataset ini bertujuan untuk mengevaluasi kemampuan model dalam melakukan prediksi terhadap data baru yang belum pernah digunakan dalam proses pelatihan model. Dalam penelitian ini, pembagian dataset dilakukan menggunakan metode train-test split dengan rasio 80:20 yaitu data training 1200 dan data testing 300, di mana 80 persen dari dataset digunakan sebagai data latih dan 20 persen sisanya digunakan sebagai data uji.

Data latih digunakan untuk melatih model machine learning agar dapat mengenali pola hubungan antara variabel input dan variabel target. Sedangkan data uji digunakan untuk mengukur performa model setelah proses pelatihan selesai dilakukan. Dengan menggunakan metode pembagian dataset ini, performa model dapat diuji secara objektif terhadap data yang tidak digunakan selama proses pelatihan.

Menurut (van Klompenburg et al., 2020), pembagian dataset menjadi data latih dan data uji merupakan langkah penting dalam pengembangan model machine learning untuk memastikan bahwa model memiliki kemampuan generalisasi yang baik terhadap data baru.

4.5 Hasil Evaluasi Model

Evaluasi model dilakukan untuk mengukur performa algoritma machine learning dalam melakukan klasifikasi tingkat keberhasilan panen padi. Dalam penelitian ini, evaluasi model dilakukan menggunakan beberapa metrik evaluasi yang umum digunakan dalam tugas klasifikasi yaitu *accuracy*, *precision*, *recall*, dan

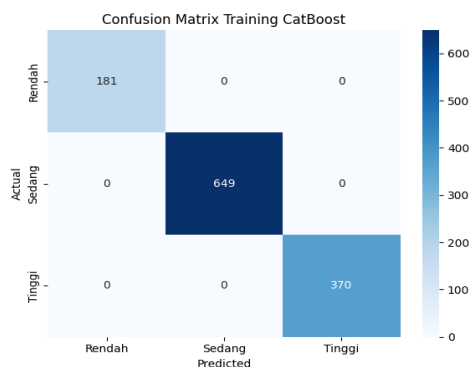
F1-score. Metrik-metrik tersebut digunakan untuk mengukur sejauh mana model mampu melakukan prediksi secara benar terhadap data uji.

Accuracy digunakan untuk mengukur persentase prediksi yang benar terhadap seluruh data uji. *Precision* digunakan untuk mengukur tingkat ketepatan model dalam memprediksi kelas tertentu. *Recall* digunakan untuk mengukur kemampuan model dalam mendeteksi seluruh data yang termasuk dalam suatu kelas. Sedangkan *F1-score* merupakan kombinasi antara *precision* dan *recall* yang memberikan gambaran keseimbangan antara kedua metrik tersebut.

Hasil evaluasi model kemudian dibandingkan untuk mengetahui algoritma mana yang memiliki performa terbaik dalam memprediksi tingkat keberhasilan panen padi. Menurut (Foody, 2023), penggunaan beberapa metrik evaluasi dalam model klasifikasi sangat penting karena setiap metrik memberikan perspektif yang berbeda dalam menilai performa model.

4.6 *Confusion Matrix* dan Kurva *ROC CatBoost*

Pengujian dilakukan dengan confusion matrix yang terdiri dari *accuracy*, *precision*, *recall* dan *F1-score* dilakukan dengan dataset sebanyak 1.500 data yang diolah dengan menggunakan *CatBoost*. Pengujian *Confusion Matrix* untuk dataset yang diolah menggunakan *CatBoost* dapat dilihat pada gambar di bawah ini. Pengujian menggunakan data testing yang berjumlah 1.200 data.



Gambar 4.2 *Confusion Matrix Training CatBoost*

Nilai *accuracy Confusion matrix* tersebut adalah sebagai berikut:

$$Accuracy = \frac{\text{Jumlah Prediksi Benar}}{\text{Jumlah Seluruh Data}} \quad (2.1)$$

$$Accuracy = \frac{1200}{1200}$$

$$Accuracy = 1.00$$

$$Accuracy = 100\%$$

Nilai *precision* dari *confusion matrix* tersebut Adalah sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \quad (2.2)$$

1. Kelas Rendah

$$Precision = \frac{181}{181 + 0}$$

$$Precision = 1.00$$

2. Kelas Sedang

$$Precision = \frac{649}{649 + 0}$$

$$Precision = 1.00$$

3. Kelas Tinggi

$$Precision = \frac{370}{370 + 0}$$

$$Precision = 1.00$$

Nilai *Recall* dari *confusion matrix* tersebut adalah sebagai berikut:

$$Recall = \frac{TP}{TP+FN} \quad (2.3)$$

1. Kelas Rendah

$$Recall = \frac{181}{181 + 0} = 1.00$$

2. Kelas Sedang

$$Recall = \frac{649}{649 + 0} = 1.00$$

3. Kelas Tinggi

$$Precision = \frac{370}{370 + 0} = 1.00$$

Nilai *F1-Score* dari *confusion matrix* tersebut adalah sebagai berikut:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.4)$$

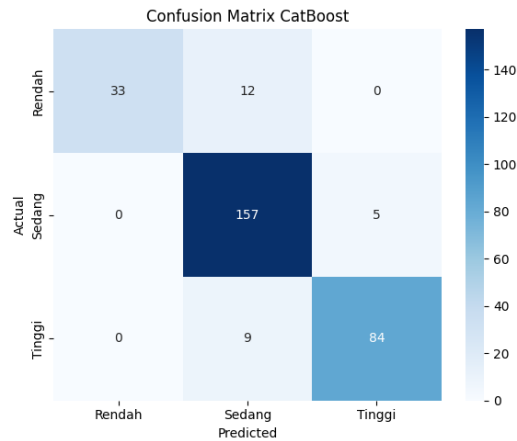
Karena $Precision = Recall = 1$

$$F1 = 2 \times \frac{1 \times 1}{1 + 1} = 1.00$$

Setelah melakukan pengujian menggunakan data training yang berjumlah 1.200 data, menunjukkan performa yang sangat tinggi dengan nilai akurasi mencapai 100%. Hal ini ditunjukkan dari *confusion matrix* dimana seluruh data pada masing-masing kelas berhasil diprediksi dengan benar. Nilai *precision*, *recall*, dan *F1-score* pada ketiga kelas yaitu rendah, sedang, dan tinggi juga mencapai nilai 1.00. Hal ini menunjukkan bahwa model mampu mempelajari pola data training secara sempurna.

Selanjutnya melakukan pengujian menggunakan data *testing* yang berjumlah 300 data, pada tahap pengujian ini nilai akurasi akan mengalami

penurunan, untuk lebih jelas akan saya paparkan gambar dan rumus pengerjaannya yang dapat dilihat di bawah ini.



Gambar 4.3 *Confusion Matrix Testing CatBoost*

Nilai *accuracy* *Confusion matrix* tersebut adalah sebagai berikut:

$$Accuracy = \frac{\text{Jumlah Prediksi Benar}}{\text{Jumlah Seluruh Data}} \quad (2.1)$$

$$\text{Prediksi benar (diagonal matrix)} = 33 + 157 + 84 = 274$$

$$Accuracy = \frac{274}{300} = 0.9133$$

$$Accuracy = 91.33\%$$

Nilai *Precision* *Confusion matrix* tersebut adalah sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \quad (2.2)$$

1. Kelas Rendah

$$Precision = \frac{33}{33 + 12}$$

$$Precision = \frac{33}{45} = 0.7333$$

$$Precision = 73.33\%$$

2. Kelas Sedang

$$Precision = \frac{157}{157 + 5}$$

$$Precision = \frac{157}{162} = 0.9691$$

$$Precision = 96.91\%$$

3. Kelas Tinggi

$$Precision = \frac{84}{84 + 9}$$

$$Precision = \frac{84}{93} = 0.9032$$

$$Precision = 90.32\%$$

Nilai *Recall Confusion matrix* tersebut adalah sebagai berikut:

$$Recall = \frac{TP}{TP+FN} \quad (2.3)$$

1. Kelas Rendah

$$Recall = \frac{33}{33 + 0} = 1$$

$$Recall = 100\%$$

2. Kelas Sedang

$$Recall = \frac{157}{157 + 21}$$

$$Recall = \frac{157}{178} = 0.8820$$

$$Recall = 88.20\%$$

3. Kelas Tinggi

$$Recall = \frac{84}{84 + 5}$$

$$Recall = \frac{84}{89} = 0.9438$$

$$Recall = 94.38\%$$

Nilai *F1-Score* dari *confusion matrix* tersebut adalah sebagai berikut:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.4)$$

1. Kelas Rendah

$$F1 = 2 \times \frac{0.7333 \times 1}{0.7333 + 1} = 0.8461$$

$$F1 = 84.61\%$$

2. Kelas sedang

$$F1 = 2 \times \frac{0.9691 \times 0.8820}{0.9691 + 0.8820} = 0.9235$$

$$F1 = 92.35\%$$

3. Kelas Tinggi

$$F1 = 2 \times \frac{0.9032 \times 0.9438}{0.9032 + 0.9438} = 0.9230$$

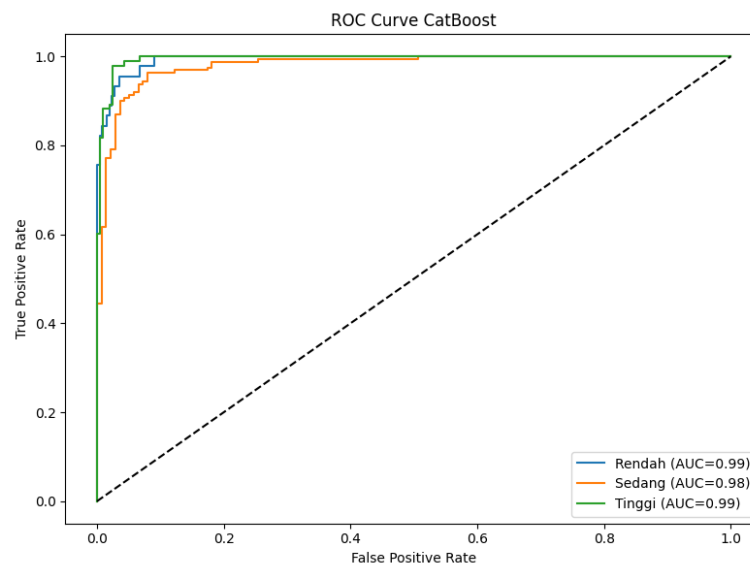
$$F1 = 92.30\%$$

Tabel 4.5 Hasil Perhitungan *CatBoost*

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>
Rendah	73.33%	100%	84.61%
Sedang	96.91%	88.20%	92.35%
Tinggi	90.32%	94.38%	92.30%

Accuracy Keseluruhan: 91.33%

Kurva ROC untuk algoritma *CatBoost* sebagai berikut:

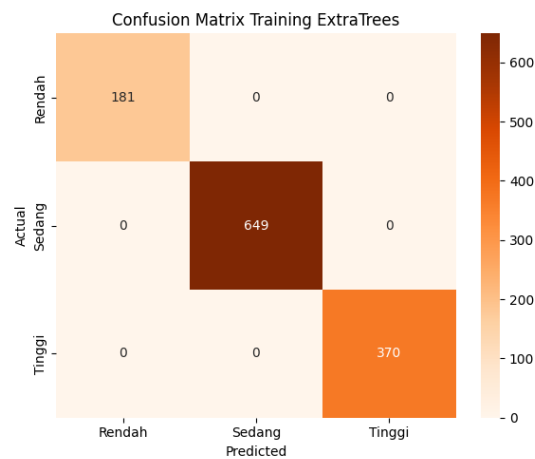


Gambar 4.4 Nilai AUC dengan algoritma *CatBoost*

Berdasarkan hasil *confusion matrix* yang diperoleh dari model *CatBoost*, diperoleh nilai prediksi benar sebanyak 274 dari total 300 data uji. Dengan menggunakan rumus *accuracy*, diperoleh nilai akurasi sebesar 91,33%. Nilai *precision*, *recall*, dan *F1-score* dihitung untuk masing-masing kelas menggunakan rumus evaluasi klasifikasi. Hasil perhitungan menunjukkan bahwa model mampu melakukan klasifikasi dengan tingkat akurasi yang tinggi terutama pada kelas panen sedang dan tinggi.

4.7 *Confusion Matrix* dan Kurva *ROC Extra Trees Classifier*

Pengujian dilakukan dengan *confusion matrix* yang terdiri dari *accuracy*, *precision*, *recall* dan *F1-score* dilakukan dengan dataset sebanyak 1.500 data yang diolah dengan menggunakan *Extra Trees Classifier*. Pengujian *Confusion Matrix* untuk dataset yang diolah menggunakan *Extra Trees Classifier* dapat dilihat pada gambar di bawah ini. Pengujian menggunakan data testing yang berjumlah 1.200 data.



Gambar 4.5 *Confusion Matrix Training Extra Trees Classifier*

Nilai *accuracy* *Confusion matrix* tersebut adalah sebagai berikut:

$$Accuracy = \frac{Jumlah\ Prediksi\ Benar}{Jumlah\ Seluruh\ Data} \quad (2.1)$$

$$Accuracy = \frac{1200}{1200}$$

$$Accuracy = 1.00$$

$$Accuracy = 100\%$$

Nilai *precision* dari *confusion matrix* tersebut Adalah sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \quad (2.2)$$

1. Kelas Rendah

$$Precision = \frac{181}{181 + 0}$$

$$Precision = 1.00$$

2. Kelas Sedang

$$Precision = \frac{649}{649 + 0}$$

$$Precision = 1.00$$

3. Kelas Tinggi

$$Precision = \frac{370}{370 + 0}$$

$$Precision = 1.00$$

Nilai *Recall* dari *confusion matrix* tersebut adalah sebagai berikut:

$$Recall = \frac{TP}{TP+FN} \quad (2.3)$$

1. Kelas Rendah

$$Recall = \frac{181}{181 + 0} = 1.00$$

2. Kelas Sedang

$$Recall = \frac{649}{649 + 0} = 1.00$$

3. Kelas Tinggi

$$Recall = \frac{370}{370 + 0} = 1.00$$

Nilai *F1-Score* dari *confusion matrix* tersebut adalah sebagai berikut:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.4)$$

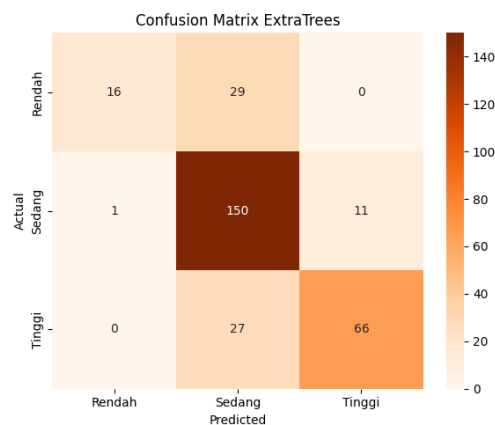
Karena $Precision = Recall = 1$

$$F1 = 2 \times \frac{1 \times 1}{1 + 1} = 1.00$$

Setelah melakukan pengujian menggunakan data training yang berjumlah 1.200 data, menunjukkan performa yang sangat tinggi dengan nilai akurasi mencapai 100%. Hal ini ditunjukkan dari *confusion matrix* dimana seluruh data pada masing-masing kelas berhasil diprediksi dengan benar. Nilai *precision*, *recall*, dan *F1-score* pada ketiga kelas yaitu rendah, sedang, dan tinggi juga mencapai nilai

1.00. Hal ini menunjukkan bahwa model mampu mempelajari pola data training secara sempurna.

Selanjutnya melakukan pengujian menggunakan data *testing* yang berjumlah 300 data, pada tahap pengujian ini nilai akurasi akan mengalami penurunan, untuk lebih jelas akan saya paparkan gambar dan rumus pengerjaannya yang dapat dilihat di bawah ini.



Gambar 4.6 *Confusion Matrix Testing Extra Trees Classifier*

Nilai *accuracy* *Confusion matrix* tersebut adalah sebagai berikut:

$$Accuracy = \frac{\text{Jumlah Prediksi Benar}}{\text{Jumlah Seluruh Data}} \quad (2.1)$$

Prediksi benar (diagonal matrix) = 16 + 150 + 66 = 232

$$Accuracy = \frac{232}{300} = 0.7733$$

$$Accuracy = 77.33\%$$

Nilai *Precision* *Confusion matrix* tersebut adalah sebagai berikut:

$$Precision = \frac{TP}{TP+FP} \quad (2.2)$$

1. Kelas Rendah

$$Precision = \frac{16}{16 + 1}$$

$$Precision = \frac{16}{17} = 0.9412$$

$$Precision = 94.12\%$$

2. Kelas Sedang

$$Precision = \frac{150}{150 + 56}$$

$$Precision = \frac{150}{206} = 0.7281$$

$$Precision = 72.81\%$$

3. Kelas Tinggi

$$Precision = \frac{66}{66 + 11}$$

$$Precision = \frac{66}{77} = 0.8571$$

$$Precision = 85.71\%$$

Nilai *Recall Confusion matrix* tersebut adalah sebagai berikut:

$$Recall = \frac{TP}{TP+FN} \quad (2.3)$$

1. Kelas Rendah

$$Recall = \frac{16}{16 + 29}$$

$$Recall = \frac{16}{45} = 0.3556$$

$$Recall = 35.56\%$$

2. Kelas Sedang

$$Recall = \frac{150}{150 + 12}$$

$$Recall = \frac{150}{162} = 0.9259$$

$$Recall = 92.59\%$$

3. Kelas Tinggi

$$Recall = \frac{66}{66 + 27}$$

$$Recall = \frac{66}{93} = 0.7097$$

$$Recall = 70.97\%$$

Nilai *F1-Score* dari *confusion matrix* tersebut adalah sebagai berikut:

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (2.4)$$

1. Kelas Rendah

$$F1 = 2 \times \frac{0.9412 \times 0.3556}{0.9412 + 0.3556} = 0.516$$

$$F1 = 51.6\%$$

2. Kelas sedang

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$F1 = 2 \times \frac{0.7281 \times 0.9259}{0.7281 + 0.9259} = 0.815$$

$$F1 = 81.5\%$$

3. Kelas Tinggi

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

$$F1 = 2 \times \frac{0.8571 \times 0.7097}{0.8571 + 0.7097} = 0.776$$

$$F1 = 77.6\%$$

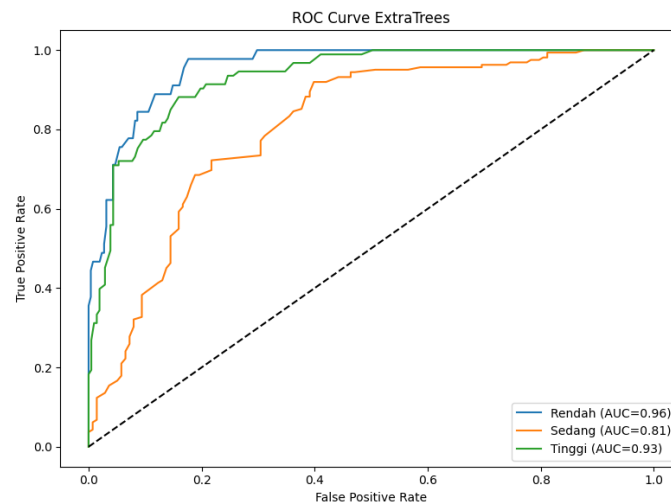
Tabel 4.6 Hasil Perhitungan *Extra Trees Classifier*

Kelas	<i>Precision</i>	<i>Recall</i>	<i>F1 Score</i>
Rendah	94.12%	35.56%	51.6%

Sedang	72.81%	92.59%	81.5%
Tinggi	85.71%	70.97%	77.6%

Accuracy Keseluruhan : 77.33%

Kurva ROC untuk algoritma *Extra Trees Classifier* sebagai berikut:



Gambar 4.7 Nilai AUC dengan algoritma *Extra Trees Classifier*

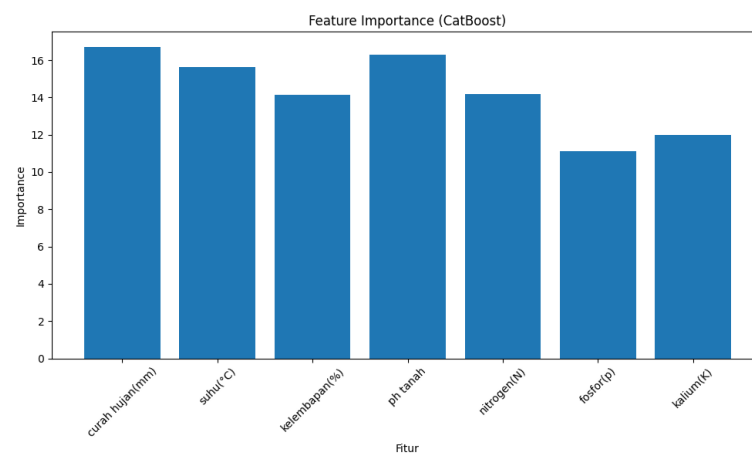
Berdasarkan hasil *confusion matrix* pada model *ExtraTreesClassifier* diperoleh jumlah prediksi benar sebanyak 232 dari total 300 data uji. Dengan menggunakan rumus *accuracy* diperoleh nilai akurasi sebesar 77,33%. Nilai *precision*, *recall*, dan *F1-score* kemudian dihitung untuk masing-masing kelas. Hasil perhitungan menunjukkan bahwa model memiliki performa yang cukup baik pada kelas panen sedang dengan nilai *recall* sebesar 92,59%, namun performa pada kelas panen rendah relatif lebih rendah dengan nilai *recall* sebesar 35,56%.

4.8 *Feature Importance*

Feature importance merupakan metode yang digunakan untuk mengukur seberapa besar kontribusi setiap fitur terhadap hasil prediksi model *machine learning*. Metode ini banyak digunakan dalam *Explainable Artificial Intelligence*

(XAI) untuk memahami bagaimana model mengambil keputusan. Secara umum, *feature importance* mengukur pengaruh suatu variabel terhadap perubahan nilai prediksi model. Semakin besar nilai importance suatu fitur, maka semakin besar pula kontribusi fitur tersebut dalam menentukan hasil prediksi model (Fang et al., 2020). Pada model berbasis *tree ensemble* seperti *ExtraTrees* dan *CatBoost*, nilai *importance* dihitung dari seberapa besar fitur tersebut digunakan dalam proses pemisahan node pada pohon keputusan dan seberapa besar pengurangan impurity yang dihasilkan.

4.8.1 Analisis *Feature Importance* Pada Model *CatBoost*



Gambar 4.8 *Feature Importance CatBoost*

Grafik pertama menunjukkan *feature importance* dari algoritma *CatBoost* terhadap variabel lingkungan yang digunakan dalam penelitian ini. Berdasarkan grafik tersebut, tingkat kepentingan fitur dapat diurutkan sebagai berikut:

Tabel 4.7 Tingkat Kepentingan Fitur *CatBoost*

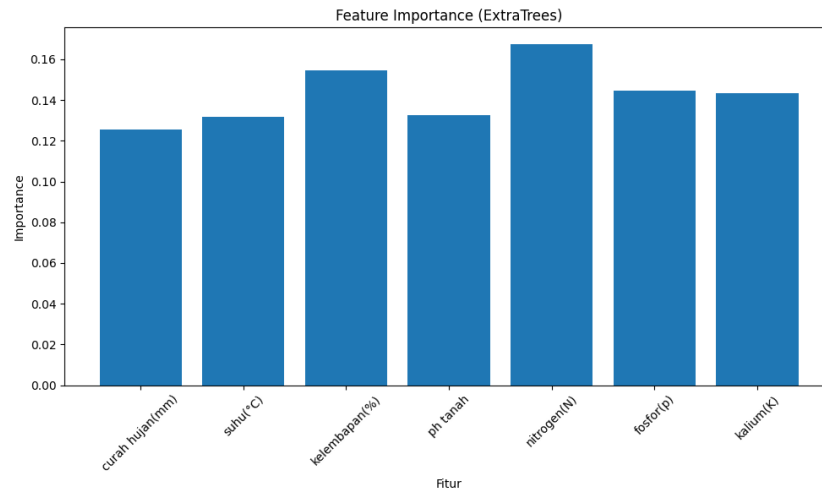
Fitur	Importance
Curah hujan	16.7
pH tanah	16.2

Suhu	15.6
Nitrogen	14.2
Kelembapan	14.1
Kalium	12.0
Fosfor	11.1

Dari grafik tersebut dapat diketahui bahwa curah hujan dan pH tanah merupakan faktor yang paling dominan dalam menentukan prediksi tingkat keberhasilan panen padi pada model *CatBoost*. Hal ini menunjukkan bahwa kondisi lingkungan terutama air dan tingkat keasaman tanah memiliki pengaruh besar terhadap produktivitas tanaman padi.

Penelitian lain yang menggunakan *CatBoost* dalam bidang pertanian juga menunjukkan bahwa variabel lingkungan seperti suhu, curah hujan, dan kondisi tanah memiliki kontribusi terbesar terhadap performa model prediksi pertanian. Dalam model boosting seperti *CatBoost*, proses pembelajaran dilakukan secara bertahap dengan menambahkan pohon keputusan secara berurutan untuk memperbaiki kesalahan model sebelumnya, sehingga fitur yang paling berpengaruh akan mendapatkan bobot kontribusi yang lebih besar.

4.8.2 Analisis *Feature Importance* Pada Model *Extra Trees Classifier*



Gambar 4.9 *Feature Importance Extra Trees Classifier*

Grafik kedua menunjukkan *feature importance* yang dihasilkan oleh model *ExtraTreesClassifier*. Berdasarkan grafik tersebut, tingkat kepentingan fitur dapat diurutkan sebagai berikut:

Tabel 4.8 Tingkat Kepentingan Fitur *Extra Trees Classifier*

Fitur	Importance
Nitrogen	0.167
Kelembapan	0.155
Fosfor	0.145
Kalium	0.144
pH tanah	0.133
Suhu	0.132
Curah hujan	0.126

Pada model *ExtraTrees*, fitur yang paling berpengaruh adalah kandungan nitrogen dalam tanah. Nitrogen merupakan unsur hara utama yang berperan dalam

pertumbuhan vegetatif tanaman padi, terutama dalam pembentukan daun dan batang tanaman. Oleh karena itu, tingginya nilai importance pada nitrogen menunjukkan bahwa ketersediaan unsur hara dalam tanah sangat berpengaruh terhadap hasil panen.

Algoritma *ExtraTrees* bekerja dengan membangun banyak pohon keputusan secara acak (*randomized trees*) dan kemudian menggabungkan hasil prediksi dari semua pohon tersebut. Metode ini menghitung feature importance berdasarkan penurunan impurity yang dihasilkan setiap fitur saat digunakan dalam proses pemisahan data pada node pohon keputusan

Secara umum dapat disimpulkan bahwa factor yang paling berpengaruh terhadap hasil panen dalam penelitian ini adalah: curah hujan, pH tanah, nitrogen, suhu dan kelembapan. Hal ini menunjukkan bahwa faktor iklim dan kondisi tanah merupakan variabel utama yang menentukan keberhasilan produksi padi.

4.9 SHAP Beeswarm Plot CatBoost

Metode *SHAP* (*SHapley Additive Explanations*) digunakan untuk menjelaskan bagaimana setiap fitur mempengaruhi prediksi model machine learning. *SHAP* berasal dari teori permainan *Shapley Value*, yang menghitung kontribusi setiap fitur terhadap hasil prediksi model (Das & Rad, 2020).

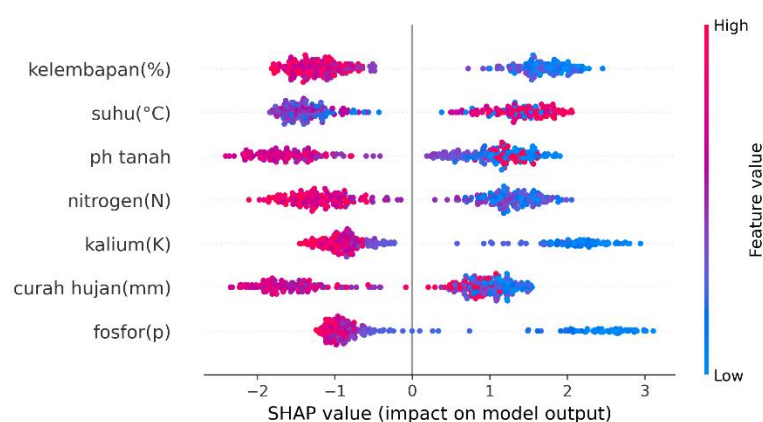
Pada SHAP Beeswarm Plot:

1. Sumbu X (*SHAP value*) menunjukkan pengaruh fitur terhadap output model.
2. Nilai *SHAP* positif berarti fitur tersebut meningkatkan kemungkinan suatu kelas.

3. Nilai *SHAP* negatif berarti fitur tersebut menurunkan kemungkinan kelas tersebut.
4. Warna merah menunjukkan nilai fitur tinggi.
5. Warna biru menunjukkan nilai fitur rendah.
6. Setiap titik mewakili satu sampel data.

Semakin jauh titik dari 0, semakin besar pengaruh fitur terhadap keputusan model.

4.9.1 Analisis *SHAP CatBoost* (Kelas Rendah)



Gambar 4.10 *SHAP CatBoost* (Kelas Rendah)

Fitur yang paling dominan dari distribusi SHAP adalah: Kelembapan, Fosfor, Nitrogen dan Suhu.

Kelembapan

Nilai kelembapan tinggi (merah) berada di sisi SHAP negatif, yang berarti kelembapan tinggi mengurangi kemungkinan panen rendah. Sebaliknya, kelembapan rendah (biru) berada di sisi positif sehingga meningkatkan peluang panen rendah. Hal ini menunjukkan bahwa kondisi lingkungan yang kering dapat menurunkan produktivitas tanaman.

Fosfor

Nilai fosfor rendah banyak berada pada sisi *SHAP positif*, artinya ketika kandungan fosfor rendah maka kemungkinan panen rendah meningkat. Fosfor berperan dalam pembentukan akar dan metabolisme tanaman sehingga kekurangannya dapat menurunkan produksi.

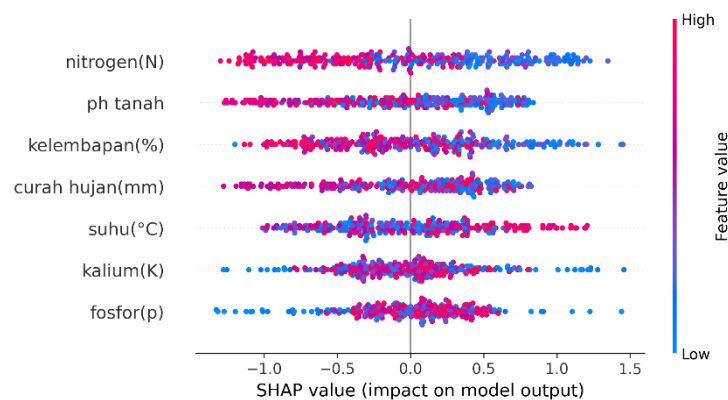
Nitrogen

Nilai nitrogen rendah meningkatkan kemungkinan kelas rendah. Nitrogen merupakan unsur penting dalam pembentukan daun dan proses fotosintesis.

Suhu

Suhu tinggi (merah) cenderung meningkatkan kemungkinan panen rendah. Hal ini menunjukkan bahwa suhu yang terlalu tinggi dapat menghambat pertumbuhan tanaman.

4.9.2 Analisis *SHAP CatBoost* (Kelas Sedang)



Gambar 4.11 *SHAP CatBoost* (Kelas Sedang)

Fitur yang paling dominan dari distribusi SHAP adalah: Nitrogen, pH tanah, suhu dan kelembapan.

Nitrogen

Nitrogen dengan nilai sedang memberikan kontribusi positif terhadap kelas sedang. Artinya kondisi nitrogen yang cukup membantu tanaman tumbuh secara normal.

pH tanah

pH tanah yang seimbang memberikan kontribusi terhadap produktivitas sedang. Tanah yang terlalu asam atau terlalu basa dapat menghambat penyerapan nutrisi.

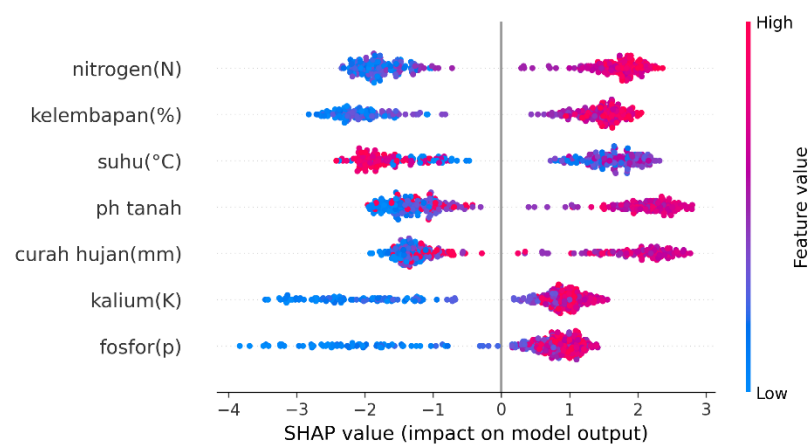
Suhu

Suhu optimal mendukung pertumbuhan tanaman yang stabil sehingga menghasilkan panen sedang.

Kelembapan

Kelembapan sedang membantu menjaga keseimbangan air tanaman sehingga produktivitas tidak terlalu rendah maupun terlalu tinggi.

4.9.3 Analisis *SHAP CatBoost* (Kelas Tinggi)



Gambar 4.12 *SHAP CatBoost* (Kelas Tinggi)

Fitur yang paling dominan dari distribusi SHAP adalah: Nitrogen, kelembapan, pH tanah dan curah hujan.

Nitrogen

Nitrogen tinggi (merah) berada pada sisi *SHAP positif*, yang berarti semakin tinggi nitrogen maka semakin besar kemungkinan panen tinggi. Nitrogen meningkatkan pertumbuhan daun dan proses fotosintesis.

Kelembapan

Kelembapan tinggi juga meningkatkan kemungkinan panen tinggi karena tanaman mendapatkan cukup air untuk pertumbuhan.

pH tanah

pH tanah yang optimal meningkatkan kemampuan tanaman menyerap unsur hara dari tanah.

Curah Hujan

Curah hujan yang cukup memberikan suplai air yang diperlukan tanaman untuk pertumbuhan maksimal.

Berdasarkan ketiga grafik *SHAP Beeswarm Plot* dapat disimpulkan bahwa:

1. Nitrogen (N) merupakan fitur paling berpengaruh dalam menentukan hasil panen.
2. Kelembapan (%) berperan penting dalam kondisi lingkungan pertumbuhan tanaman.
3. pH tanah mempengaruhi kemampuan tanaman menyerap nutrisi.
4. Curah hujan dan suhu berperan sebagai faktor lingkungan tambahan.

Model *CatBoost* mampu mengidentifikasi hubungan kompleks antara unsur hara tanah, kondisi lingkungan, dan tingkat hasil panen padi.

4.10 SHAP Beeswarm Plot Extra Trees Classifier

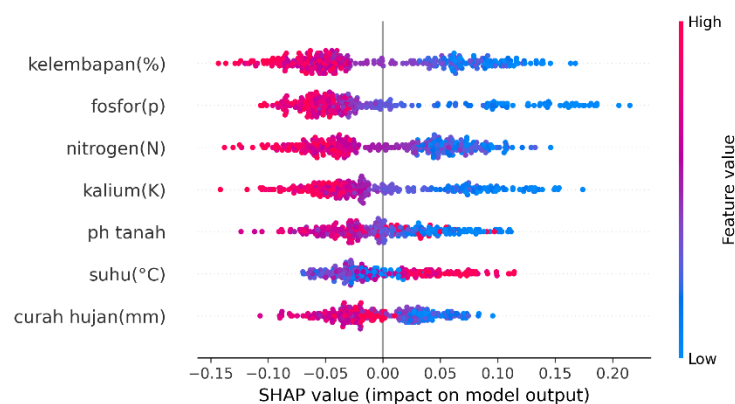
Metode *SHAP* (*SHapley Additive Explanations*) digunakan untuk menjelaskan bagaimana setiap fitur mempengaruhi prediksi model machine learning. *SHAP* berasal dari teori permainan *Shapley Value*, yang menghitung kontribusi setiap fitur terhadap hasil prediksi model (Das & Rad, 2020).

Pada SHAP Beeswarm Plot:

1. Sumbu X (*SHAP value*) menunjukkan pengaruh fitur terhadap output model.
2. Nilai *SHAP* positif berarti fitur tersebut meningkatkan kemungkinan suatu kelas.
3. Nilai *SHAP* negatif berarti fitur tersebut menurunkan kemungkinan kelas tersebut.
4. Warna merah menunjukkan nilai fitur tinggi.
5. Warna biru menunjukkan nilai fitur rendah.
6. Setiap titik mewakili satu sampel data.

Semakin jauh titik dari 0, semakin besar pengaruh fitur terhadap keputusan model.

4.10.1 Analisis SHAP Extra Trees (Kelas Rendah)



Gambar 4.13 SHAP Extra Trees(Kelas Rendah)

Fitur yang paling dominan dari distribusi *SHAP* adalah: Kelembapan, Fosfor, Nitrogen dan Kalium.

Kelembapan

Nilai kelembapan tinggi (merah) banyak berada di sisi negatif *SHAP value*. Artinya, semakin tinggi kelembapan maka kemungkinan panen rendah menurun. Sebaliknya kelembapan rendah (biru) meningkatkan peluang panen rendah. Ini menunjukkan bahwa kondisi lingkungan yang kering berpotensi menyebabkan hasil panen lebih rendah.

Fospor

Nilai fosfor tinggi juga berada di sisi negatif. Artinya semakin tinggi kandungan fosfor tanah, maka kemungkinan panen rendah menurun. Ini sesuai dengan teori agronomi bahwa fosfor berperan penting dalam pertumbuhan akar tanaman.

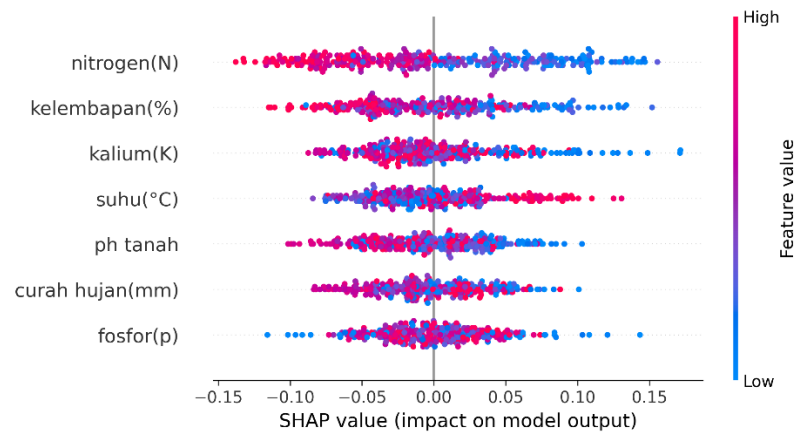
Nitrogen

Nilai nitrogen rendah cenderung meningkatkan prediksi kelas rendah. Nitrogen merupakan unsur penting dalam pertumbuhan vegetatif tanaman. Jika nitrogen rendah, produksi tanaman akan menurun.

Kalium

Kalium rendah juga meningkatkan kemungkinan panen rendah. Kalium berperan dalam ketahanan tanaman terhadap stres lingkungan.

4.10.2 Analisis *SHAP Extra Tress* (Kelas Sedang)



Gambar 4.14 *SHAP Extra Tress* (Kelas Sedang)

Fitur yang paling dominan dari distribusi SHAP adalah: Nitrogen, Kelembapan, Kalium, Suhu.

Nitrogen

Distribusi titik menunjukkan bahwa nitrogen dengan nilai sedang memberikan kontribusi positif terhadap prediksi kelas sedang. Artinya kondisi nitrogen yang seimbang membantu tanaman menghasilkan panen dengan produktivitas sedang.

Kelembapan

Kelembapan sedang meningkatkan kemungkinan panen sedang. Hal ini menunjukkan bahwa kondisi lingkungan yang tidak terlalu kering maupun terlalu basah menghasilkan pertumbuhan tanaman yang stabil.

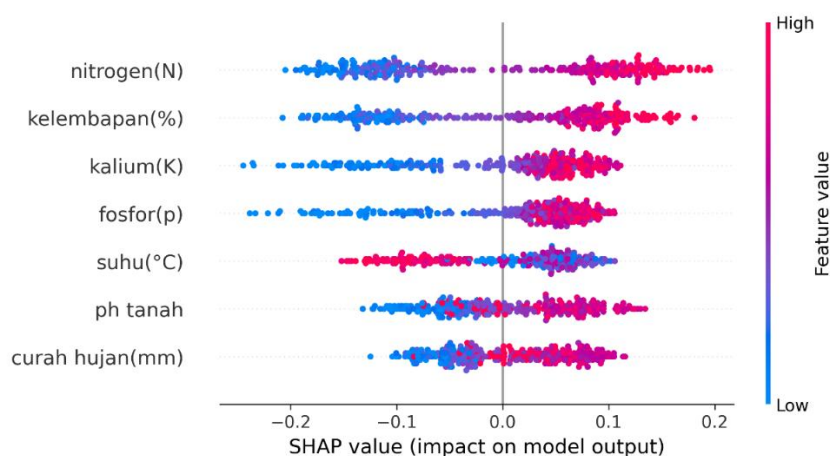
Kalium

Kalium juga memberikan kontribusi terhadap kestabilan pertumbuhan tanaman. Distribusi *SHAP* menunjukkan bahwa kalium dengan nilai sedang cenderung menghasilkan panen sedang.

Suhu

Suhu memiliki pengaruh moderat. Suhu yang terlalu tinggi atau terlalu rendah dapat menurunkan produktivitas tanaman.

4.10.3 Analisis *SHAP Extra Trees* (Kelas Tinggi)



Gambar 4.15 *SHAP Extra Trees* (Kelas Tinggi)

Fitur yang paling dominan dari distribusi SHAP adalah: Nitrogen, kelembapan, Kalium, Fospor.

Nitrogen

Nilai nitrogen tinggi (merah) berada pada *SHAP positif*. Artinya semakin tinggi nitrogen, semakin besar kemungkinan panen tinggi. Hal ini sesuai dengan teori bahwa nitrogen meningkatkan pertumbuhan daun dan fotosintesis tanaman.

Kelembapan

Kelembapan tinggi juga meningkatkan prediksi panen tinggi. Lingkungan yang cukup lembab membantu proses fotosintesis dan pertumbuhan tanaman.

Kalium

Kalium tinggi memberikan kontribusi positif terhadap panen tinggi. Kalium meningkatkan kualitas hasil tanaman dan ketahanan terhadap penyakit.

Fosfor

Fosfor juga memberikan kontribusi positif terhadap peningkatan produksi tanaman. Fosfor membantu pembentukan akar dan perkembangan tanaman.

Berdasarkan ketiga grafik *SHAP Beeswarm Plot*, dapat disimpulkan bahwa:

1. Nitrogen (N) merupakan fitur paling penting dalam menentukan tingkat hasil panen.
2. Kelembapan (%) memiliki pengaruh besar terhadap kondisi lingkungan pertumbuhan tanaman.
3. Kalium (K) dan Fosfor (P) berperan dalam meningkatkan kualitas dan produktivitas tanaman.
4. Faktor lingkungan seperti suhu dan curah hujan memiliki pengaruh tambahan tetapi tidak sebesar unsur hara tanah.

Model *ExtraTrees* mampu mengidentifikasi hubungan kompleks antara kondisi lingkungan, unsur hara tanah, dan tingkat hasil panen padi.

4.11 Kesimpulan Perbandingan Model

Berdasarkan hasil analisis *SHAP Beeswarm Plot*, dapat disimpulkan bahwa kedua model berhasil mengidentifikasi faktor-faktor penting yang mempengaruhi tingkat hasil panen padi, terutama unsur hara tanah seperti nitrogen, fosfor, dan kalium, serta kondisi lingkungan seperti kelembapan dan suhu.

Namun demikian, model *CatBoost* menunjukkan interpretasi yang lebih stabil dan lebih jelas dalam menggambarkan kontribusi fitur terhadap prediksi model, sedangkan model *ExtraTrees* memberikan distribusi kontribusi fitur yang lebih bervariasi karena sifat randomisasi pada algoritma ensemble-nya.

Tabel 4.9 Karakteristik Interpretasi

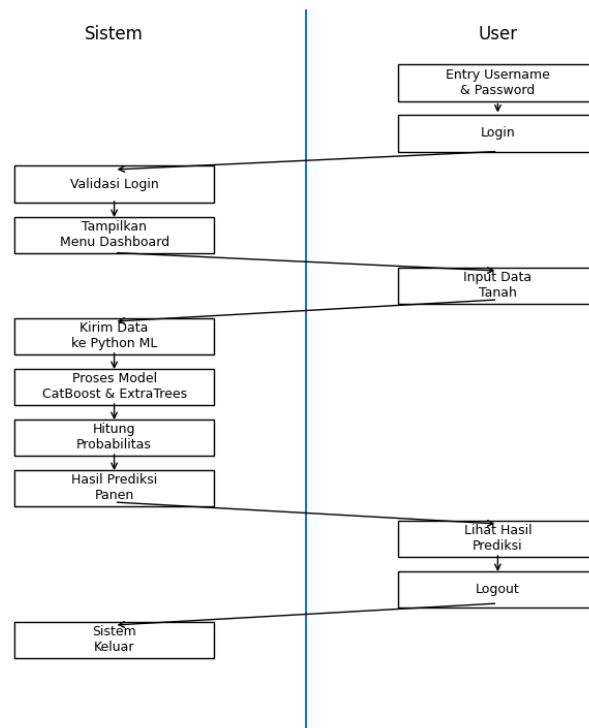
Aspek	<i>CatBoost</i>	<i>ExtraTrees</i>
Konsistensi interpretasi	Tinggi	Sedang
Stabilitas <i>SHAP value</i>	Lebih Stabil	Lebih Menyebar
Kemampuan menangkap interaksi fitur	Sangat baik	baik
Kompleksitas model	Tinggi	Sedang

Dari perbandingan tersebut dapat disimpulkan bahwa *CatBoost* memiliki kemampuan interpretasi yang lebih stabil, sedangkan *ExtraTrees* memberikan interpretasi yang lebih eksploratif karena variasi pohon keputusan yang lebih tinggi. Dengan demikian, dalam konteks interpretasi model menggunakan Explainable AI, *CatBoost* cenderung lebih mudah diinterpretasikan dibandingkan *ExtraTrees*, meskipun kedua model tetap memberikan informasi yang konsisten mengenai faktor-faktor yang mempengaruhi produktivitas panen padi.

4.12 Penerapan Algoritma Terpilih

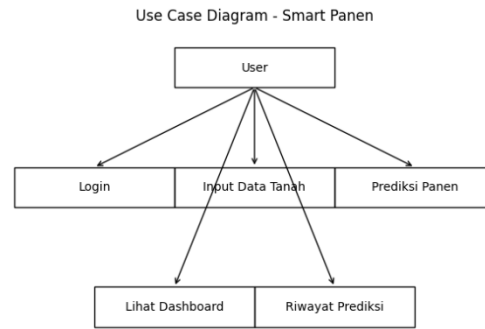
Dari hasil evaluasi dan validasi di atas dapat diketahui bahwa metode *CatBoost* memiliki Tingkat akurasi dan performa yang baik, sehingga *rule* yang dihasilkan oleh metode *CatBoost* dapat dijadikan sebagai *rule* untuk pembuatan *prototype* yang dapat memudahkan petani dalam memprediksi hasil panen padi.

Berikut adalah alur dari Activity Diagram, Use Case, Class Diagram dan Flow Chart yang diimplementasikan ke dalam prototype Prediksi tingkat keberhasilan panen.

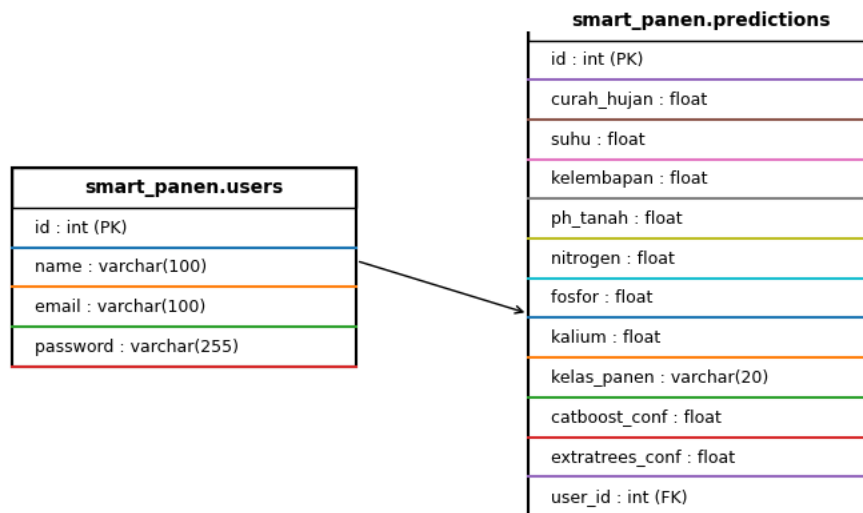


Gambar 4.16 Activity Diagram Login

User membuka web dengan membuat akun, setelah itu akan login pada sistem dan akan menampilkan menu dashboard, disitu ada terdapat pilihan untuk mulai prediksi setelah itu user akan mengisi data yang di cantumkan di sistem setelah selesai user menunggu hasil prediksi ,setelah itu akan muncul grafik yang menentukan prediksi rendah, sedang atau tinggi. Setelah selesai user dapat *log out* .Dibawah ini adalah use case dari menu login dan class diagram.

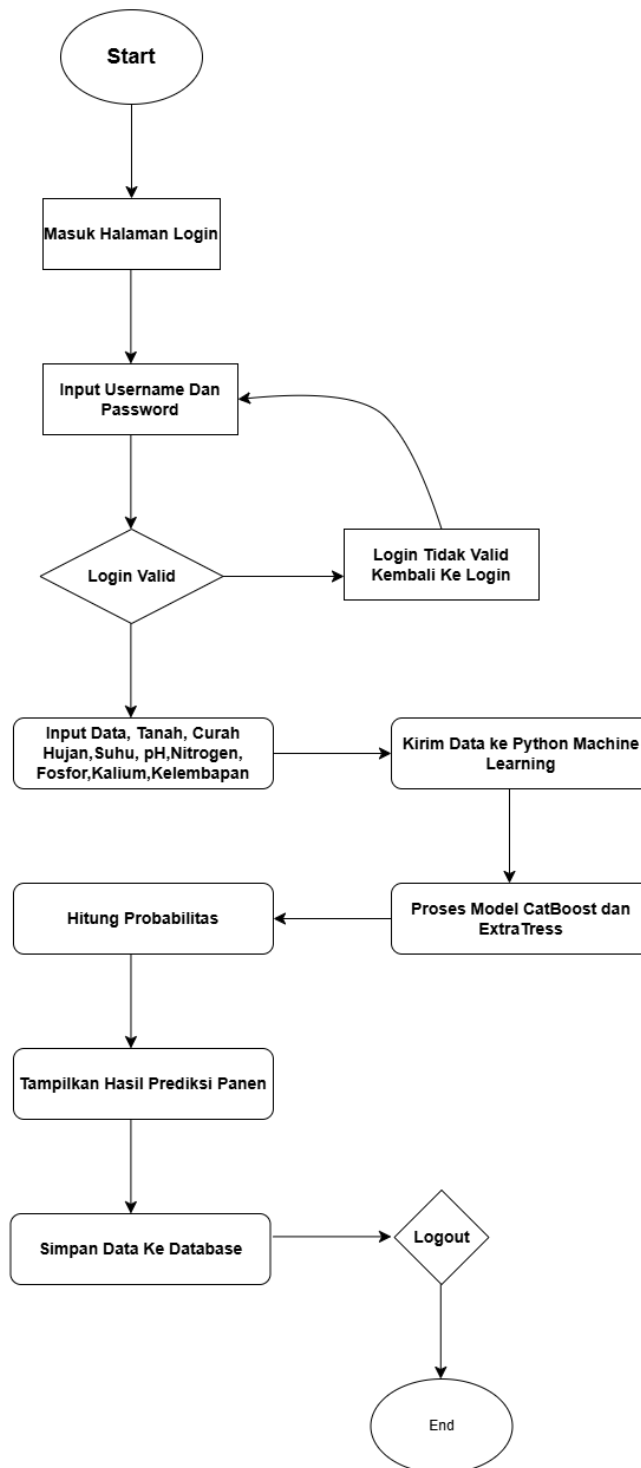


Gambar 4.17 Use case Login



Gambar 4.18 Class Diagram

Prototype Sistem Prediksi Hasil Panen Padi

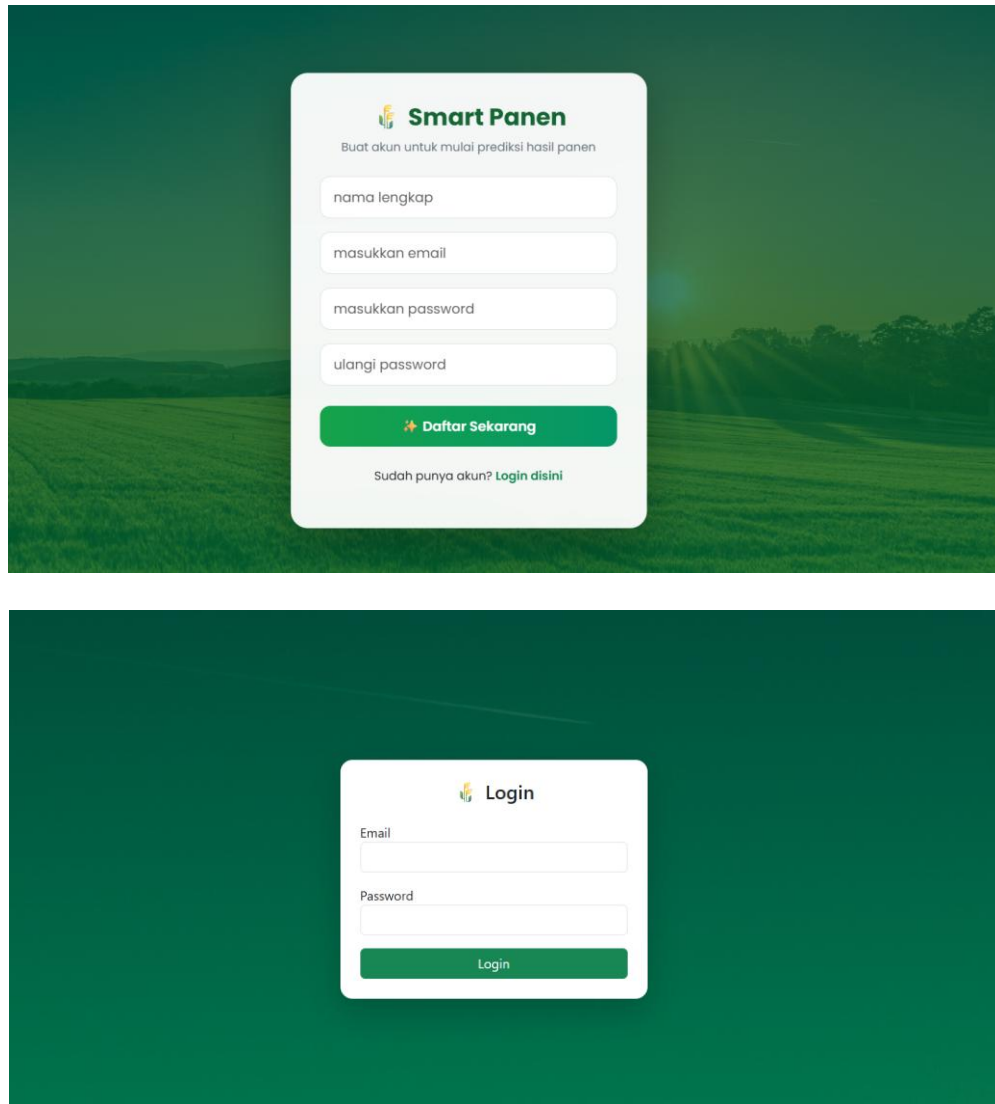


Gambar 4.19 *Flow chart Prototype* Prediksi Keberhasilan Pnaen

Prototipe yang digunakan dalam penelitian ini dibuat berbasis web dengan menggunakan *PHP* dan database menggunakan *Mysql*. *Tampilan untuk Form* utama

Graphical User Interface (GUI) *Prototype* prediksi kelayakan pemberian kredit dapat dilihat pada gambar dibawah.

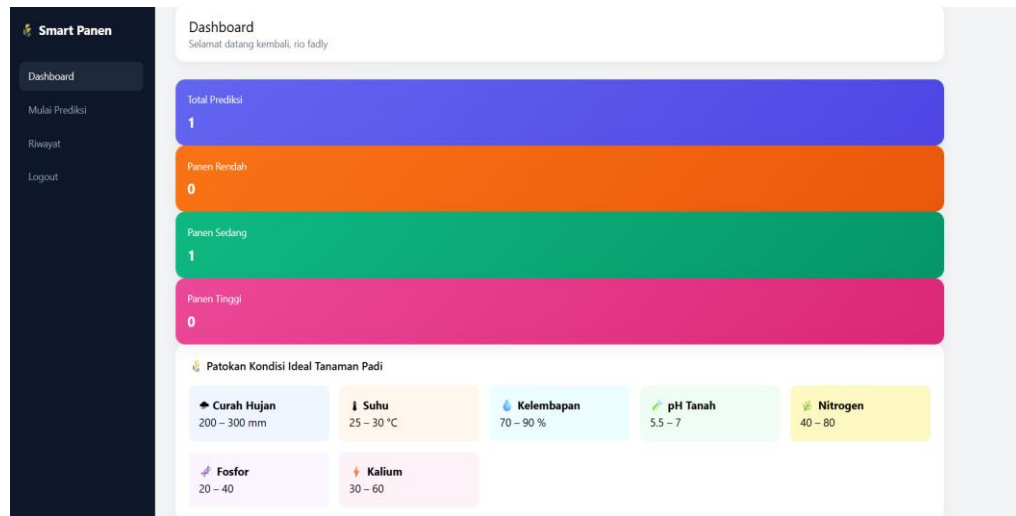
1. Tampilan *Login*



The image displays two screenshots of the 'Smart Panen' application interface. The top screenshot shows the registration form, which includes the title 'Smart Panen', a subtitle 'Buat akun untuk mulai prediksi hasil panen', and four input fields: 'nama lengkap', 'masukkan email', 'masukkan password', and 'ulangi password'. A green button labeled 'Daftar Sekarang' is positioned below the fields, and a link 'Sudah punya akun? Login disini' is at the bottom. The bottom screenshot shows the login form, titled 'Login', with two input fields: 'Email' and 'Password', and a green button labeled 'Login'. Both forms are centered on a dark green background with a blurred image of a field.

Gambar 4.20 *Form Login*

2. Tampilan Halaman Utama



Gambar 4.21 tampilan Prediksi Panen

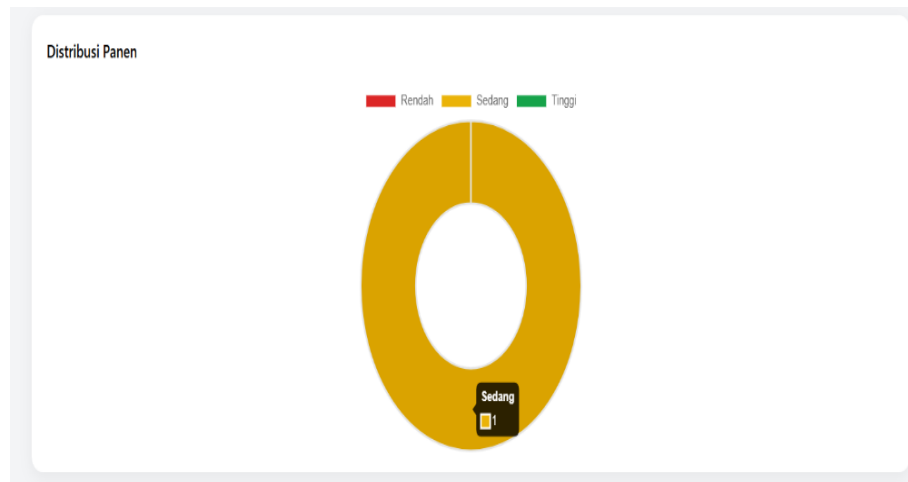
3. Tampilan Halaman *Upload* Prediksi

The screenshot shows the 'Upload Prediction' form. It features a central white panel with input fields for: pH Tanah (6), Nitrogen (29), Fosfor (40), Kalium (35), and Kelembapan (90). A green 'Proses Prediksi' button is located below these fields. The result is displayed in a green box: 'PANEN TINGGI' with 'CatBoost : 99.7%' and 'ExtraTrees : 66.0%'. At the bottom, there are two status indicators: 'Analisis Kondisi Lahan' and a warning 'Nitrogen tanah terlalu rendah'.

Gambar 4.22 Form Upload Data

Pada menu *Form upload*, *user* dapat mengisi menu-menu yang telah disediakan untuk menguji data. Selanjutnya akan ditampilkan hasil prediksi berapa persen kemungkinan akan keakuratan hasil panen.

4. Tampilan *Dashboard*



Gambar 4.23 Halaman *Dashboard*

Pada Menu *Dashboard*, user dapat melihat grafik dari data-data yang telah dipilih, dengan melihat di halaman dashboard grafik akan muncul setelah hasil prediksi selesai dimulai.

5. Tampilan menu Riwayat

Riwayat Prediksi Panen

Tampilkan: data

Cari:

No	Nama	Kelas Panen	Tanggal	Aksi
1	-	Sedang	08 Mar 2026 08:43	Hapus

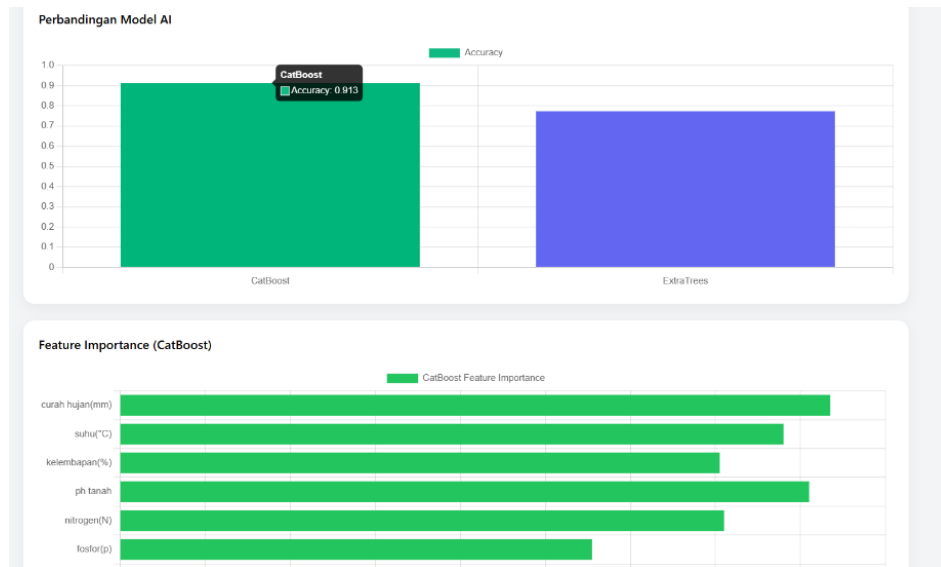
Menampilkan 1 sampai 1 dari 1 data

1

Gambar 4.24 Halaman Riwayat

Di menu Riwayat user dapat melihat hasil panen, tanggal prediksi dan dapat menghapus jika terjadi kesalahan dalam data yang di isi.

6. Tampilan *Dashboard Grafik*



Gambar 4.25 *Grafik Akurasi*

Di sini user dapat melihat indikator yang menjadi titik tumpu prediksi dan nilai ke akuratan prediksi dan bisa mengetahui variabel yang paling berpengaruh untuk hasil panen berhasil atau tidak

4.13 Pengujian *Prototype Perangkat Lunak (software)*

Pengujian dilakukan dengan tujuan untuk mengetahui apakah aplikasi dibangun sesuai dengan fungsional yang diharapkan.

Tabel 4.10 Pengujian Kotak Hitam

Kelas Uji	Butir Uji	Jenis Pengujian
<i>File Upload</i>	Masukkan Data	<i>Black Box</i>
	Upload Data	<i>Black Box</i>
<i>Dashboard</i>	Lihat <i>Grafik</i>	<i>Black Box</i>

1. Pengujian Data *Upload*

Pada pengujian data upload terbagi atas dua bagian yaitu, masukkan data dan prediksi data.

Tabel 4.11 Pengujian Data Upload

Kasus dan Hasil Uji (Data Sesuai)			
Data Masukan	Yang Diharapkan	Pengamatan	Kesimpulan
Data yang akan di <i>upload</i>	Sistem Mengeluarkan Hasil Prediksi	Dapat Melakukan Hasil Prediksi	Diterima
Klik Upload Prediksi	Hasil berhasil diimport oleh sistem	Dapat mengeluarkan Hasil Prediksi	Diterima
Kasus dan Hasil uji(data salah)			
Data Masukkan	Yang Diharapkan	Pengamatan	Kesimpulan
Data tidak diisi semua	Tidak bisa melakukan prediksi	Data tidak tersimpan dan menampilkan kesalahan , sesuai yang diharapkan	Diterima

2. Pengujian *Dashboard*

Pada pengujian *Dashboard* user hanya meng-klik *Dashboard*

Tabel 4.12 Pengujian *Dashboard*

Kasus dan Hasil Uji (Data Sesuai)			
Data Masukan	Yang Diharapkan	Pengamatan	Kesimpulan
Data yang telah di upload	Sistem Mengeluarkan Hasil Prediksi	Dapat Melakukan Hasil Prediksi	Diterima
Klik Riwayat	Dapat menampilkan detail prediksi	Dapat Menampilkan sesuai yang diharapkan	Diterima

Berdasarkan Pengujian *Black Box* yang sudah dilakukan, menjelaskan bahwa aplikasi yang dibangun sudah berjalan dengan baik sesuai dengan yang diharapkan.

4.14 Lingkungan Pengujian

Lingkungan pengujian memberikan gambaran tentang spesifikasi *hardware* dan *software* yang digunakan oleh pengguna dalam proses pengujian sistem, baik pengujian *validasi* maupun kualitas. *Spesifikasi* tersebut diperoleh dalam proses observasi berdasarkan aspek sistem. Berikut ringkasan singkat spesifikasi dari perangkat keras dan perangkat lunak yang digunakan oleh pengguna dalam proses pengujian.

1. *Hardware*
 - a. 12th Gen Intel(R) Core(TM) i5-1235U (1.30 GHz)
 - b. 8.00 GB (7.68 GB usable)
 - c. 64-bit operating system, x64-based processor
2. *Software*
 - a. *Laragon*
 - b. *PHP*
 - c. *Mysql*
 - d. *Web Browser(google Chrome)*

Perangkat lunak yang digunakan untuk membangun aplikasi prototype aplikasi prediksi Tingkat keberhasilan panen adalah menggunakan *Software Open Source* yaitu *PHP*, *Laragon* dan *Mysql* untuk database. Ditinjau dari segi kehandalan *software* tersebut sudah teruji dan sudah banyak digunakan untuk membangun sistem tanpa terkendala masalah lisensi karena bersifat *open source*.

4.15 Aspek Penelitian Lanjut

Penulis menyadari keterbatasan dalam penelitian saat ini, diharapkan pada penelitian selanjutnya yang sejenis dapat dipertimbangkan hal-hal berikut :

1. Mengkombinasikan lebih banyak metode dalam Analisa data dan penyelesaian masalah, sehingga didapat sebuah sistem yang lebih efektif dan efisien dalam pengolahan ataupun penyajian informasi.
2. Penelitian ini dapat dikembangkan dengan algoritma klasifikasi yang lain yang terdapat di *Machine Learning*.

BAB V

PENUTUP

5.1 Kesimpulan

Dari pengukuran kinerja dengan melakukan komparasi dua algoritma yang telah dilakukan berdasarkan jumlah data maka dapat disimpulkan bahwa algoritma *CatBoost* memiliki kemampuan dalam akurasi keputusan untuk menentukan tingkat keberhasilan panen padi dengan indikator rendah, sedang, dan tinggi. Kedua algoritma dapat digunakan dalam menentukan tingkat keberhasilan panen padi. Kedua algoritma ini dikomparasi kemudian diuji akurasinya. Tingkat akurasi tertinggi lah yang digunakan dalam menentukan tingkat keberhasilan panen padi. Adapun hasil penelitian dari percobaan yang telah dilakukan dapat disimpulkan bahwa:

1. Algoritma *CatBoost* memiliki tingkat akurasi tertinggi yaitu 91,33% sedangkan untuk *Extra Trees Classifier* sebesar 77,33% maka selisih sebesar 14,00%. Hasil evaluasi model menunjukkan bahwa algoritma *CatBoost* memiliki performa yang lebih baik dibandingkan dengan *ExtraTrees* dalam melakukan klasifikasi hasil panen padi. Berdasarkan analisis ROC Curve, model *CatBoost* memperoleh nilai AUC sebesar 0.99 pada kelas rendah, 0.98 pada kelas sedang, dan 0.99 pada kelas tinggi, yang termasuk dalam kategori excellent classification performance. Sementara itu, model *ExtraTrees* memperoleh nilai AUC sebesar 0.96 pada kelas rendah, 0.81 pada kelas sedang, dan 0.93 pada kelas tinggi.

2. *Rule* yang dihasilkan algoritma *CatBoost* diterapkan dalam *prototype* prediksi keberhasilan panen padi dengan hasil akurasi pengujian yang sangat stabil. Berdasarkan akurasi yang dihasilkan *prototype* menunjukkan bahwa metode dan *prototype* yang diterapkan sudah baik dalam memprediksi tingkat keberhasilan panen padi.

3. Implementasi metode *Explainable AI* menggunakan *SHAP (SHapley Additive Explanations)* mampu memberikan interpretasi terhadap model *machine learning* dengan menunjukkan kontribusi masing-masing variabel terhadap hasil prediksi. Hasil analisis *SHAP* menunjukkan bahwa variabel curah hujan, suhu, kelembapan, dan *nitrogen* merupakan faktor yang paling berpengaruh dalam menentukan hasil prediksi panen padi.

Dari hasil penelitian ini diharapkan algoritma terpilih yaitu algoritma *CatBoost* dalam memprediksi keputusan tingkat keberhasilan panen padi dengan cepat, sehingga membantu pencapaian kinerja petani.

5.2 Saran

Agar penelitian ini bisa ditingkatkan, berikut adalah saran-saran yang diusulkan:

1. Pada penelitian ini variabel yang digunakan masih terbatas pada faktor lingkungan dan unsur hara tanah seperti curah hujan, suhu, kelembapan, pH tanah, nitrogen, fosfor, dan kalium. Penelitian selanjutnya disarankan untuk menambahkan variabel lain yang juga berpengaruh terhadap produktivitas padi, seperti: Jenis varietas padi, jenis tanah, Intensitas sinar matahari, luas lahan, penggunaan pupuk dan pestisida dan kondisi irigasi. Dengan penambahan variabel

tersebut diharapkan model dapat menghasilkan prediksi yang lebih akurat dan representatif terhadap kondisi nyata di lapangan.

2. Sistem yang dikembangkan pada penelitian ini masih berbasis web. Penelitian selanjutnya dapat mengembangkan sistem menjadi aplikasi mobile atau mengintegrasikannya dengan teknologi *Internet of Things (IoT)* seperti sensor tanah dan sensor cuaca. Dengan demikian data lingkungan dapat diperoleh secara real-time, sehingga sistem dapat memberikan rekomendasi prediksi yang lebih cepat dan akurat kepada petani.

3. Penelitian ini menggunakan algoritma CatBoost dan ExtraTrees sebagai metode klasifikasi. Untuk penelitian selanjutnya disarankan melakukan perbandingan dengan algoritma lain seperti: *Random Forest, XGBoost, LightGBM, Neurol Network*. Dengan melakukan perbandingan lebih banyak algoritma, diharapkan dapat ditemukan model yang memiliki kinerja prediksi yang lebih optimal.

DAFTAR PUSTAKA

- Azhari, M., & Nasution, P. (2025). *Application of Data Mining to Determine the Performance of Family Planning Field Officers (PLKB) Using the C4 . 5 Algorithm.*
- Azhari, M., & Rahman, M. (2022). Analisis Perbandingan Algoritma WP Dan TOPSIS Dalam Menentukan Kandidat Peserta Lomba Kompetensi Siswa. *It (Informatic Technique) Journal*, 10(1), 42. <https://doi.org/10.22303/it.10.1.2022.42-55>
- Azhari, M., Situmorang, Z., & Rosnelly, R. (2021). Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes. *Jurnal Media Informatika Budidarma*, 5(2), 640. <https://doi.org/10.30865/mib.v5i2.2937>
- Badshah, A., Yousef Alkazemi, B., Din, F., Zamli, K. Z., & Haris, M. (2024). Crop Classification and Yield Prediction Using Robust Machine Learning Models for Agricultural Sustainability. *IEEE Access*, 12(November), 162799–162813. <https://doi.org/10.1109/ACCESS.2024.3486653>
- Das, A., & Rad, P. (2020). *Opportunities and Challenges in Explainable Artificial Intelligence (XAI): A Survey.* 1–24. <http://arxiv.org/abs/2006.11371>
- Dinh, T., Wong, H., Lisik, D., Koren, M., Tran, D., Yu, P. S., & Torres-Sospedra, J. (2025). Data clustering: a fundamental method in data science and management. *Data Science and Management.* <https://doi.org/10.1016/j.dsm.2025.08.001>
- Diyanti, Martanto, & Bahtiar, A. (2023). Jurnal Informatika Terpadu Prediksi Hasil Panen Padi Tahun 2023 Menggunakan Metode Regresi Linear Di Kabupaten Indramayu. *Jurnal Informatika Terpadu*, 9(1), 18–23. <https://journal.nurulfikri.ac.id/index.php/JIT>
- Elbeltagi, A., Srivastava, A., Cao, X., Bilali, A. El, Raza, A., Khadke, L., & Salem, A. (2025). An interpretable machine learning approach based on SHAP, Sobol and LIME values for precise estimation of daily soybean crop coefficients. *Scientific Reports*, 15(1), 1–20. <https://doi.org/10.1038/s41598-025-20386-y>
- Fang, F., Ventre, C., Li, L., Kanthan, L., Wu, F., & Basios, M. (2020). *Better Model Selection with a new Definition of Feature Importance.* <http://arxiv.org/abs/2009.07708>
- Filippi, P., Han, S. Y., & Bishop, T. F. A. (2025). *Addressing the Common Issues in Published Studies.* 1–19.
- Foody, G. M. (2023). Challenges in the real world use of classification accuracy metrics: From recall and precision to the Matthews correlation coefficient. *PLoS ONE*, 18(10 October), 1–27. <https://doi.org/10.1371/journal.pone.0291908>
- Gopi, S. R., & Karthikeyan, M. (2023). Effectiveness of Crop Recommendation and Yield Prediction using Hybrid Moth Flame Optimization with Machine Learning. *Engineering, Technology and Applied Science Research*, 13(4), 11360–11365. <https://doi.org/10.48084/etasr.6092>
- Himayanta, K. L., & Wardhani, D. F. (2025). Prediksi Hasil Panen Padi Dengan Machine Learning. *Jurnal Kelitbangan*, 13(1), 1–14.
- Islam, M. M., Alharthi, M., Alkadi, R. S., Islam, R., & Masum, A. K. M. (2024). Crop yield prediction through machine learning: A path towards sustainable

- agriculture and climate resilience in Saudi Arabia. *AIMS Agriculture and Food*, 9(4), 980–1003. <https://doi.org/10.3934/AGRFOOD.2024053>
- Jabed, M. A., & Azmi Murad, M. A. (2024). Crop yield prediction in agriculture: A comprehensive review of machine learning and deep learning approaches, with insights for future research and sustainability. *Heliyon*, 10(24). <https://doi.org/10.1016/j.heliyon.2024.e40836>
- Jhajharia, K., Sharma, N. V., & Mathur, P. (2025). *A Machine Learning Model for Crop Yield Prediction Using Remote Sensing Data*. 6(November 2024), 577–590. <https://doi.org/10.47857/irjms.2025.v06i02.03182>
- Manurung, D., Zealtiel, B., & Lubis, A. H. (2025). Prediksi Produksi Tanaman Padi di Indonesia dengan Menggunakan Algoritma Random Forest Regressor. *Journal of Computing and Informatics Research*, 4(3), 345. <https://doi.org/10.47065/comforch.v4i3.2125>
- Masahid, M., Dawud, M. Y., & Abryandoko, E. W. (2025). Pengaruh Fluktuasi Unsur Iklim Tahunan terhadap Produksi Padi: Studi Empiris Berdasarkan Data Historis. *Agrikultura*, 36(2), 216–227. <https://doi.org/10.24198/agrikultura.v36i2.63758>
- Mugemangango, C., Nzabanita, J., Muhoza, D. N., & Cahill, N. D. (2024). Comparative Analysis of Machine Learning Models for Predicting Rice Yield: Insights from Agricultural Inputs and Practices in Rwanda. *Research on World Agricultural Economy*, 5(4), 350–366. <https://doi.org/10.36956/rwae.v5i4.1247>
- Narendra Bayutama Wibisono, S. S. (2025). Crop Yield Prediction Using Random Forest Algorithm and Xgboost Machine Learning Model Narendra. *International Journal of Research and Innovation in Social Science (IJRISS)*, VII(2454), 1175–1189. <https://doi.org/10.47772/IJRISS>
- Nikhil, U. V., Pandiyan, A. M., Raja, S. P., & Stamenkovic, Z. (2024). Machine Learning-Based Crop Yield Prediction in South India: Performance Analysis of Various Models. *Computers*, 13(6). <https://doi.org/10.3390/computers13060137>
- Nizami, T., Mustaqim, M. A., & Ariannor, W. (2025). Analisis Kinerja Model Machine learning dalam Prediksi Gagal Panen Gabah. *Progresif: Jurnal Ilmiah Komputer*, 21(1), 184. <https://doi.org/10.35889/progresif.v21i1.2501>
- Nuraini, D., Violina, D., Anamisa, D. R., Khotimah, B. K., Jauhari, A., & Mufarroha, F. A. (2025). Prediksi Hasil Panen Padi dengan Metode Multiple Linear Regression dan Particle Swarm Optimization untuk Meningkatkan Produksi Padi di Madura. *JUSIFOR: Jurnal Sistem Informasi Dan Informatika*, 4(1), 1–8. <https://doi.org/10.70609/jusifor.v4i1.5857>
- Petropoulos, T., Benos, L., Berruto, R., Miserendino, G., Marinoudi, V., Busato, P., Zisis, C., & Bochtis, D. (2025). *Interpretable Machine Learning for Legume Yield Prediction Using Satellite Remote Sensing Data*. 1–18.
- Purwaningrum, Y., Asbur, Y., Atmaja Nasution, S., & Nuh, M. (2025). PENGARUH PERUBAHAN IKLIM TERHADAP PRODUKTIVITAS TANAMAN PADI GOGO (*Oryza sativa* L.). *AGRILAND Jurnal Ilmu Pertanian*, 13(1), 19–23.
- Quille-Mamani, J., Ramos-Fernández, L., Huanuqueño-Murillo, J., Quispe-Tito, D., Cruz-Villacorta, L., Pino-Vargas, E., Flores del Pino, L., Heros-Aguilar, E., & Ángel Ruiz, L. (2025). Rice Yield Prediction Using Spectral and

- Textural Indices Derived from UAV Imagery and Machine Learning Models in Lambayeque, Peru. *Remote Sensing*, 17(4). <https://doi.org/10.3390/rs17040632>
- Razavi, M. A., Nejadhashemi, A. P., Majidi, B., Razavi, H. S., Kpodo, J., Eeswaran, R., Ciampitti, I., & Prasad, P. V. V. (2024). Enhancing crop yield prediction in Senegal using advanced machine learning techniques and synthetic data. *Artificial Intelligence in Agriculture*, 14, 99–114. <https://doi.org/10.1016/j.aiaa.2024.11.005>
- Saha, S., Kucher, O. D., Utkina, A. O., & Rebouh, N. Y. (2025). *Precision agriculture for improving crop yield predictions : a literature review*. July, 1–11. <https://doi.org/10.3389/fagro.2025.1566201>
- Shawon, S. M., Ema, F. B., Mahi, A. K., Niha, F. L., & Zubair, H. T. (2025). Crop yield prediction using machine learning: An extensive and systematic literature review. *Smart Agricultural Technology*, 10(September 2024), 100718. <https://doi.org/10.1016/j.atech.2024.100718>
- Sweet, L., Müller, C., Anand, M., & Zscheischler, J. (2023). Cross-Validation Strategy Impacts the Performance and Interpretation of Machine Learning Models. *Artificial Intelligence for the Earth Systems*, 2(4), 1–14. <https://doi.org/10.1175/aies-d-23-0026.1>
- Tasneem, K. T., Shahzad, M. U., Rashid, J., Othman, K. M., Zafar, T., & Faheem, M. (2025). Predicting rice yield and impact of climate change on rice production using machine learning models. *Theoretical and Applied Climatology*, 156(12). <https://doi.org/10.1007/s00704-025-05912-2>
- van Klompenburg, T., Kassahun, A., & Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177(January), 105709. <https://doi.org/10.1016/j.compag.2020.105709>
- Yenkikar, A., Mishra, V. P., Bali, M., & Ara, T. (2025). An explainable AI-based hybrid machine learning model for interpretability and enhanced crop yield prediction. *MethodsX*, 15(April), 103442. <https://doi.org/10.1016/j.mex.2025.103442>
- Yuan, J., Zhang, Y., Zheng, Z., Yao, W., Wang, W., & Guo, L. (2024). Grain Crop Yield Prediction Using Machine Learning Based on UAV Remote Sensing: A Systematic Literature Review. *Drones*, 8(10). <https://doi.org/10.3390/drones8100559>
- Yunis, R., Sudarto, & Adiputra Pardosi, I. (2024). Enhancing Rice Production Prediction: A Comparative Machine Learning Analysis of Climate Variables. *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, 13(1), 91–104. <https://doi.org/10.23887/janapati.v13i1.71527>