



**PREDIKSI PH FERMENTASI KOPI BERBASIS *ENSEMBLE*
LEARNING DAN SISTEM *IOT* PEMANTAUAN SUHU**

*diajukan untuk memenuhi sebagian persyaratan memperoleh gelar Sarjana pada
program studi Informatika*

SKRIPSI

Oleh

**Alian Zahid Alaziz
212410103047**

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS JEMBER
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
JEMBER
2026**

PERSEMBAHAN

Skripsi ini saya persembahkan untuk:

1. Allah SWT yang senantiasa memberikan rahmat, hidayah, kesehatan, kekuatan, serta kemudahan dalam menyelesaikan penelitian dan skripsi ini;
2. Ibu tercinta, Siti Khoiriyah, kedua kakak, Rifdah Rosyidah dan Azizah Laila Muslimah, serta keluarga besar yang selalu memberikan doa, dukungan, motivasi, dan semangat dalam setiap langkah perjalanan saya hingga dapat menyelesaikan skripsi ini;
3. Para tenaga pendidik, khususnya dosen pembimbing saya yaitu Ibu Nelly Oktavia Adiwijaya S.Si., MT. Serta dosen penguji saya Bapak Yudha Alif Auliya S.Kom., M.Kom., dan Ibu Maliatul Fitriyasari M.Sc., yang telah memberikan saran, masukan, dan motivasi dalam penyusunan skripsi ini;
4. Saya sendiri Alian Zahid Alaziz yang selalu berusaha, berdoa, bertahan, dan tidak menyerah selama proses penyusunan skripsi ini;
5. Seluruh pihak yang tidak dapat disebutkan satu per satu, yang telah memberikan bantuan, doa, dukungan, dan motivasi selama proses penelitian maupun penyusunan skripsi ini;
6. Almamater tercinta Program Studi Informatika Fakultas Ilmu Komputer Universitas Jember.

MOTTO

"Maka sesungguhnya bersama kesulitan ada kemudahan. Sesungguhnya bersama kesulitan ada kemudahan."

فَإِنَّ مَعَ الْعُسْرِ يُسْرًا ﴿٥﴾ إِنَّ مَعَ الْعُسْرِ يُسْرًا ﴿٦﴾

(QS. Al-Insyirah: 5–6)

PERNYATAAN ORISINALITAS

Saya yang bertanda tangan di bawah ini :

Nama : Alian Zahid Alaziz

NIM : 212410103047

Menyatakan dengan sesungguhnya bahwa skripsi yang berjudul: "*Prediksi pH Fermentasi Kopi Berbasis Ensemble Learning dan Sistem IoT Pemantauan Suhu*" adalah benar-benar hasil karya sendiri, kecuali jika dalam pengutipan substansi disebutkan sumbernya, dan belum pernah diajukan pada institusi manapun, serta bukan karya jiplakan. Saya bertanggung jawab atas keabsahan dan kebenaran isinya sesuai dengan sikap ilmiah yang harus dijunjung tinggi.

Demikian pernyataan ini saya buat dengan sebenarnya, tanpa adanya tekanan dan paksaan dari pihak manapun serta bersedia mendapat sanksi akademik jika ternyata di kemudian hari pernyataan ini tidak benar.

Jember, 6 Juli 2026

Yang menyatakan,



Alian Zahid Alaziz

NIM 212410103047

HALAMAN PERSETUJUAN

Skripsi berjudul "*Prediksi pH Fermentasi Kopi Berbasis Ensemble Learning dan Sistem IoT Pemantauan Suhu*" telah diuji dan disetujui oleh Fakultas Ilmu Komputer Universitas Jember pada:

Hari : Jum'at

Tanggal : 17 Juli 2026

Tempat : Fakultas Ilmu Komputer Universitas Jember

Pembimbing

1. Pembimbing Utama

Nama : Nelly Oktavia Adiwijaya S.Si.,MT.

NIP : 198410242009122008

Tanda Tangan


(.....)

Penguji

1. Penguji Utama

Nama : Yudha Alif Auliya S.Kom., M.Kom.

NIP : 199206302022031009


(.....)

2. Penguji Anggota 1

Nama : Maliatul Fitriyasari M.Sc.

NIP : 199503152023212038


(.....)

ABSTRAK

Fermentasi kopi merupakan salah satu tahapan penting yang memengaruhi kualitas akhir biji kopi. Salah satu parameter utama yang menentukan keberhasilan fermentasi adalah nilai pH. Namun, pengukuran pH belum dapat dilakukan secara kontinu karena memerlukan peralatan khusus serta proses pengukuran yang kurang praktis untuk diterapkan di lapangan. Penelitian ini bertujuan membangun model prediksi pH fermentasi kopi berbasis *Ensemble Learning* dengan memanfaatkan data suhu yang diperoleh dari sistem *Internet of Things (IoT)*. Sistem *IoT* menggunakan mikrokontroler ESP32 dan sensor suhu DS18B20 untuk melakukan pemantauan suhu secara *real-time* selama proses fermentasi, sedangkan data pH digunakan sebagai variabel target dalam proses pemodelan.

Dataset penelitian terdiri atas 1.452 data hasil pengamatan fermentasi kopi yang diperoleh dari PTPN I Regional 5 Kebun Kalisat Jampit. Tahap pra pemrosesan data meliputi penanganan data hilang dan penanganan *outlier*. Selanjutnya dilakukan pembagian dataset dengan skenario 80:20, 70:30, dan 60:40. Pendekatan *Ensemble Learning* diterapkan menggunakan dua algoritma, yaitu *Random Forest* yang menggabungkan banyak *decision tree* melalui mekanisme *bagging* dan *Extreme Gradient Boosting (XGBoost)* yang membangun model secara bertahap melalui mekanisme *boosting*. Kinerja model dievaluasi menggunakan koefisien determinasi (R^2) dan *Root Mean Square Error (RMSE)*

Hasil Penelitian menunjukkan bahwa model *Random Forest* memberikan performa terbaik pada skenario pembagian dataset 60:40 menggunakan parameter (Ahmed, 2024), dengan nilai R^2 sebesar 0.9097 dan *RMSE* sebesar 0.1516. Sementara itu, penggunaan *Grid Search* berhasil meningkatkan performa *XGBoost* hingga mencapai nilai R^2 sebesar 0.9040 dan *RMSE* sebesar 0.1563 pada skenario pembagian dataset yang sama.

Kata Kunci: Fermentasi kopi, *Ensemble Learning*, *Internet of Things (IoT)*.

ABSTRACT

Coffee fermentation is one of the key stages that influence the final quality of coffee beans. One of the primary parameters determining the success of fermentation is the pH value. However, continuous pH measurement is not yet feasible because it requires specialized equipment and a measurement process that is impractical to implement in the field. This study aims to develop a coffee fermentation pH prediction model based on Ensemble Learning by utilizing temperature data obtained from an Internet of Things (IoT) system. The IoT system uses an ESP32 microcontroller and a DS18B20 temperature sensor to monitor temperature in real-time during the fermentation process, while pH data is used as the target variable in the modeling process.

The research dataset consists of 1,452 observations of coffee fermentation obtained from PTPN I Regional 5, Kalisat Jampit Plantation. The data preprocessing stage includes handling missing data and outliers. Subsequently, the

dataset was split into training and testing sets using 80:20, 70:30, and 60:40 splits. The Ensemble Learning approach was applied using two algorithms: Random Forest, which combines multiple decision trees through the bagging mechanism, and Extreme Gradient Boosting (XGBoost), which builds models incrementally through the boosting mechanism. Model performance was evaluated using the coefficient of determination (R^2) and Root Mean Square Error (RMSE).

The research results show that the Random Forest model delivered the best performance in a 60:40 dataset split scenario using the parameters specified by Ahmed (2024), with an R^2 value of 0.9097 and an RMSE of 0.1516. Meanwhile, the use of Grid Search successfully improved XGBoost's performance, achieving an R^2 value of 0.9040 and an RMSE of 0.1563 under the same dataset split scenario.

Keywords: Coffee fermentation, Ensemble Learning, Internet of Things (IoT).

RINGKASAN

“*Prediksi pH Fermentasi Kopi Berbasis Ensemble Learning dan Sistem Iot Pemantauan Suhu*”; Alian Zahid Alaziz, 212410103047; 45 halaman; Program Studi Informatika Fakultas Ilmu Komputer Universitas Jember.

Fermentasi kopi merupakan salah satu tahapan penting yang menentukan kualitas akhir biji kopi karena memengaruhi karakteristik cita rasa, aroma, dan tingkat keasaman. Selama proses fermentasi, aktivitas mikroorganisme mengakibatkan perubahan kondisi lingkungan, termasuk suhu dan pH. Suhu dapat dipantau secara kontinu menggunakan sistem *Internet of Things (IoT)*, sedangkan pengukuran pH belum dapat dilakukan secara kontinu karena memerlukan peralatan khusus serta proses pengukuran yang kurang praktis untuk diterapkan di lapangan. Kondisi tersebut menyebabkan perubahan pH selama fermentasi tidak dapat dipantau secara *real-time*, sehingga pengendalian proses fermentasi menjadi kurang optimal. Oleh karena itu, diperlukan suatu model yang mampu memprediksi nilai pH berdasarkan parameter yang dapat dipantau secara kontinu, yaitu suhu dan waktu fermentasi.

Penelitian ini bertujuan mengembangkan model prediksi pH fermentasi kopi menggunakan metode *Ensemble Learning* dengan memanfaatkan data suhu yang diperoleh dari sistem *IoT*. Sistem pemantauan dikembangkan menggunakan mikrokontroler ESP32 dan sensor suhu DS18B20 untuk memperoleh data suhu secara *real-time* selama proses fermentasi. Dataset penelitian diperoleh dari proses fermentasi kopi di PTPN I Regional 5 Kebun Kalisat Jampit dengan total 1.452 data hasil pengamatan. Variabel masukan yang digunakan dalam pemodelan adalah suhu dan waktu fermentasi, sedangkan variabel keluaran berupa nilai pH.

Hasil penelitian menunjukkan bahwa model *Random Forest* memberikan performa terbaik pada skenario pembagian dataset 60:40 menggunakan parameter (Ahmed, 2024), dengan nilai R^2 sebesar 0.9097 dan $RMSE$ sebesar 0.1516. Sementara itu, penggunaan *Grid Search* berhasil meningkatkan performa *XGBoost* hingga mencapai nilai R^2 sebesar 0.9040 dan $RMSE$ sebesar 0.1563 pada skenario pembagian dataset yang sama. Dengan demikian penerapan metode *Ensemble Learning* yang dipadukan dengan sistem *IoT* mampu menghasilkan model prediksi pH dengan tingkat akurasi yang baik.

PRAKATA

Puji syukur ke hadirat Allah SWT., atas segala rahmat dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “*Prediksi pH Fermentasi Kopi Berbasis Ensemble Learning dan Sistem Iot Pemantauan Suhu*”. Skripsi ini disusun untuk salah satu syarat menyelesaikan pendidikan Strata Satu (S1) pada program studi Informatika Fakultas Ilmu Komputer Universitas Jember. Dalam penulisan skripsi ini, penulis mendapatkan banyak bantuan, dukungan, dan motivasi dari berbagai pihak. Maka dari itu, penulis menyampaikan terima kasih yang tulus dan tak terhingga kepada:

1. Prof. Drs. Antonius Cahya Prihandoko, M.App.Sc., Ph.D. selaku dekan Fakultas Ilmu Komputer;
2. Dosen pembimbing penulis yaitu Ibu Nelly Oktavia Adiwijaya S.Si., MT. Serta dosen penguji penulis Bapak Yudha Alif Auliya S.Kom., M.Kom., dan Ibu Maliatul Fitriyasari M.Sc., yang telah memberikan saran, masukan, dan motivasi dalam penyusunan skripsi penulis;
3. Bapak Dr. Dwiretno Istiyadi Swasono, ST., M.Kom., selaku kepala Laboratorium Pertanian Cerdas beserta anggota yang telah memfasilitasi dan membantu dalam penyelesaian skripsi penulis;
4. Gama Wisnu Fajarianto S.Kom., M.Kom. selaku dosen pembimbing akademik penulis;
5. Seluruh dosen dan staf karyawan Fakultas Ilmu Komputer Universitas Jember yang telah memberikan ilmu dan dukungan penuh dalam penyelesaian skripsi penulis;
6. Ibu tercinta, Siti Khoiriyah, kedua kakak, Rifdah Rosyidah dan Azizah Laila Muslimah, serta keluarga besar yang selalu memberikan doa, dukungan, motivasi, dan semangat dalam setiap langkah perjalanan penulis hingga dapat menyelesaikan skripsi penulis;

7. Sahabat, teman-teman, dan seluruh pihak yang tidak dapat penulis sebutkan satu per satu yang telah memberikan bantuan, doa, dukungan, dan semangat selama proses penyusunan skripsi penulis.

Penulis menyadari bahwa tugas akhir ini masih jauh dari kata sempurna, oleh sebab itu diharapkan adanya masukan yang membangun sehingga skripsi ini dapat bermanfaat bagi banyak pihak.

Jember, 6 Juli 2026

Penulis

DAFTAR ISI

PERSEMBAHAN.....	ii
MOTTO	iii
PERNYATAAN ORISINALITAS.....	iv
HALAMAN PERSETUJUAN	v
ABSTRAK	vi
RINGKASAN	viii
PRAKATA.....	ix
DAFTAR ISI.....	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR.....	xiv
BAB 1. PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Penelitian	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
BAB 2. TINJAUAN PUSTAKA.....	6
2.1 Penelitian Terdahulu	6
2.2 Fermentasi Kopi	8
2.3 Parameter pH sebagai Indikator Mutu Fermentasi.....	10
2.4 Pemantauan Suhu Menggunakan Sistem <i>IoT</i>	10
2.5 <i>Ensemble Learning</i>	11
2.7 <i>Extreme Gradient Boosting (XGBoost)</i>	13
2.8 Evaluasi	14
2.8.1 <i>R-Squared (R^2)</i>	14
2.8.2 <i>Root Mean Square Error (RMSE)</i>	14
BAB 3. METODOLOGI PENELITIAN.....	15
3.1 Jenis Penelitian.....	15
3.2 Objek Penelitian	15
3.3 Tempat dan Waktu Penelitian	15
3.4 Tahapan Penelitian	15

3.4.1	Pengumpulan Data	16
3.4.2	Pra-pemrosesan Data.....	16
3.4.3	Pembagian Dataset	17
3.4.4	Implementasi Model.....	17
3.4.5	Pengujian Model	17
BAB 4.	Hasil dan pembahasan.....	18
4.1	Pengumpulan Data	18
4.2	Pra-pemrosesan Data.....	20
4.2.1	Pemeriksaan Data Kosong (<i>Missing Value</i>).....	20
4.2.2	Pemeriksaan Data Pencilan (<i>Outlier</i>).....	21
4.3	Pembagian Dataset	22
4.4	Implementasi Model.....	22
4.4.1	<i>Random Forest</i>	22
4.4.2	<i>XGBoost</i>	23
4.5	Pengujian Model	23
4.5.1	<i>Random Forest</i>	23
4.5.2	<i>XGBoost</i>	25
BAB 5.	KESIMPULAN DAN SARAN.....	29
5.1	KESIMPULAN.....	29
5.2	SARAN	29
	Daftar Pustaka.....	31

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu.....	7
Tabel 3.1 Skenario pengujian model.....	17
Tabel 4.1 Statistik deskriptif data penelitian.....	20
Tabel 4.2 Pemeriksaan <i>missing value</i>	21
Tabel 4.3 Hasil pengujian model <i>Random Forest</i>	23
Tabel 4.4 Hasil pengujian model <i>XGBoost</i>	26

DAFTAR GAMBAR

Gambar 2.1 Arsitektur <i>Ensemble Learning</i>	12
Gambar 3.1 Tahapan penelitian	15
Gambar 3.2 Rancangan alat pengumpulan data	16
Gambar 4.1 Pengambilan data fermentasi kopi	18
Gambar 4.2 Titik pengamatan pada kolam fermentasi	19
Gambar 4.3 <i>Plot</i> sebaran data suhu dan pH	19
Gambar 4.4 <i>Boxplot</i> suhu dan pH	22
Gambar 4.5 <i>Scatter plot</i> prediksi dan aktual <i>Random Forest</i>	25
Gambar 4.6 <i>Scatter plot</i> prediksi dan aktual <i>XGBoost</i>	28

BAB 1. PENDAHULUAN

1.1 Latar Belakang

Indonesia termasuk salah satu produsen kopi utama dunia dengan menempati peringkat keempat berdasarkan volume produksi, setelah Brasil, Vietnam dan Kolombia (Indonesia Investment, 2024). Fermentasi adalah operasi pascapanen yang paling penting yang memengaruhi kualitas biji termasuk kadar air, indeks fermentasi dan pH (Widyaningrum et al., 2022). Proses fermentasi dengan suhu terkontrol 27°C menghasilkan skor sensorik tertinggi dibandingkan suhu 18°C, karena suhu menjaga kestabilan metabolisme mikroba dan pembentukan asam organik (Costa et al., 2025). Sementara itu, pH ideal berada pada kisaran 3.7-4.0 dimana proses dekomposisi lendir kopi berjalan optimal tanpa menghasilkan cita rasa berlebih (Polanía Rivera et al., 2024). Proses ini menghasilkan asam organik, alkohol dan senyawa volatil yang berpengaruh pada cita rasa, aroma dan keasaman kopi (Rocha et al., 2024). Di berbagai wilayah penghasil kopi, pemantauan proses fermentasi masih dilakukan dengan cara manual. Pemantauan suhu hanya dilakukan sesekali, sedangkan pH umumnya tidak diukur secara kontinu karena memerlukan alat laboratorium dan proses pengukuran yang tidak praktis di lapangan. Keterbatasan ini membuat pengendalian proses fermentasi menjadi sulit, sehingga dapat mengurangi konsistensi mutu hasil olahan. Padahal, otomatisasi pemantauan dapat membantu petani untuk menjaga kondisi fermentasi tetap stabil serta mengurangi risiko penurunan kualitas. Perkembangan teknologi *Internet of Things (IoT)* memberikan peluang untuk menciptakan sistem pemantauan fermentasi yang dapat melakukan pemantauan suhu secara berkelanjutan (*real-time*) dan memberikan notifikasi saat suhu terdeteksi berada di luar kisaran ideal (Adeleke et al., 2023). Beberapa penelitian rancang bangun *IoT* telah dilakukan, tetapi umumnya masih terbatas pada pemantauan dasar tanpa adanya integrasi analitik cerdas berbasis data (Vinanjana, 2023).

Di sisi lain, pemanfaatan *Machine learning (ML)* dalam sektor agrikultur menunjukkan perkembangan yang pesat. *ML* telah diterapkan untuk klasifikasi

kualitas kopi, prediksi karakter sensorik, hingga deteksi cacat berbasis citra (Dhanya et al., 2022; Motta et al., 2024). Namun, penerapan *ML* untuk memodelkan dinamika fermentasi terutama dalam prediksi pH berdasarkan parameter proses seperti suhu dan waktu masih sangat terbatas. Padahal, pH adalah indikator penting yang sulit diukur secara kontinu di lapangan, sehingga prediksi pH berbasis data *real-time* menjadi solusi yang sangat relevan bagi petani.

Terdapat beberapa penelitian terdahulu yang membahas penerapan *Machine learning* dalam prediksi parameter fermentasi dan kualitas pangan. Penelitian (Tirado-Kulieva et al., 2024) menggunakan *Near-Infrared Spectroscopy (NIRS)* dan *Partial Least Squares Regression (PLSR)* dengan teknik seleksi fitur β coefficients, *Variable Importance Projection (VIP)*, dan *Covering Array Feature Selection (CAFS)* untuk memprediksi perubahan pH dan *Total Soluble Solids (TSS)* selama fermentasi kopi. Performa terbaik ditunjukkan oleh metode *CAFS*. Untuk prediksi pH pada varietas *Caturra*, metode tersebut menghasilkan nilai R^2 sebesar 0.903 dengan *RMSE* sebesar 0.096. Adapun pada prediksi *TSS* varietas *Typica*, diperoleh nilai R^2 sebesar 0.922 dan *RMSE* sebesar 0.688. Penelitian berikutnya oleh (Saetae, 2025) mengembangkan *framework Machine learning* berbasis *hybrid Long Short-Term Memory (LSTM)* dan *Support Vector Regression (SVR)* untuk memprediksi pertumbuhan mikroba dan proses *acidification* pada fermentasi yogurt di suhu industri. Hasil penelitian menunjukkan bahwa model *hybrid LSTM* memberikan performa terbaik pada suhu 37°C dengan nilai R^2 sebesar 0.9982 dan *RMSE* sebesar 0.0074 untuk *Lactobacillus delbrueckii subsp. bulgaricus*. Selain itu, model *SVR* menunjukkan performa terbaik dalam prediksi *acidification* dengan nilai R^2 mencapai 0.9999 dan *RMSE* sebesar 0.0094 pada suhu 45°C. Sementara itu, penelitian (Adeleke et al., 2025) menggunakan pendekatan *Internet of Things (IoT)* dan *Machine learning* berbasis *Convolutional Neural Network (CNN)*, *Feedforward Neural Network (FNN)*, dan *Random Forest* untuk memprediksi *Total Titratable Acidity (TTA)* pada fermentasi amasi. Berdasarkan hasil pengujian, model *CNN* memberikan kemampuan prediksi yang paling optimal dengan memperoleh nilai R^2 sebesar 0.9475. Di sisi lain, penelitian (Anker et al., 2024) menggunakan beberapa algoritma *Machine learning Regression* seperti *Random*

Forest Regressor (RFR), *Gradient Boosting Regression (GBR)*, *Extreme Gradient Boosting (XGB)*, *Support Vector Regression (SVR)*, dan *Artificial Neural Network Regression (NNR)* untuk memprediksi *aroma partitioning* pada matriks produk susu. Berdasarkan hasil pengujian, *Random Forest Regressor* menjadi model dengan kinerja terbaik pada suhu 37°C dengan R^2 sebesar 0.99, sedangkan *XGBoost* unggul pada suhu 7°C dengan R^2 sebesar 0.88.

Meskipun banyak penelitian telah mengembangkan sistem *IoT* untuk pemantauan proses fermentasi serta memanfaatkan *ML* dalam analisis kualitas kopi, belum banyak penelitian yang secara khusus mengintegrasikan pemantauan suhu berbasis *IoT* dengan model prediksi pH fermentasi dengan pendekatan *Ensemble Learning*. Sebagian besar sistem *IoT* yang ada masih berfokus pada rancang bangun perangkat tanpa analitik prediktif, sementara penelitian *ML* di bidang kopi umumnya terbatas pada klasifikasi kualitas atau karakteristik sensorik, bukan memodelkan dinamika fermentasi secara *real-time*. Selain itu, pH yang merupakan parameter penting dalam fermentasi jarang diprediksi secara kontinu, penggunaan alat pengukur fisik-elektrokimia yang digunakan untuk mengukur pH secara *real-time* rentan mengalami kerusakan dan pergeseran kalibrasi akibat terpapar lingkungan fermentasi kopi yang pekat dan asam. Oleh karena itu, teknik pemodelan prediktif berbasis parameter suhu dan waktu fermentasi menjadi solusi yang efisien dan ekonomis. Selain menekan biaya perawatan perangkat keras, model ini mampu memberikan prediksi tren tingkat pH secara akurat guna mencegah fermentasi berlebihan yang dapat merusak cita rasa biji kopi.

Berdasarkan gap tersebut, penelitian ini berfokus pada pengembangan sistem *IoT* untuk pemantauan suhu fermentasi secara *real-time* serta model prediksi pH berbasis *Ensemble Learning*. Integrasi kedua komponen ini diharapkan dapat membantu petani menjaga kestabilan proses fermentasi, meningkatkan konsistensi mutu hasil olahan, serta mendorong penerapan teknologi berbasis data dalam praktik pascapanen kopi di Indonesia. Penelitian ini juga berpotensi untuk memberikan kontribusi signifikan dalam inovasi digitalisasi proses fermentasi yang lebih efisien dan dapat dengan mudah diterapkan oleh petani kecil.

1.2 Rumusan Masalah

Berdasarkan uraian yang terdapat pada latar belakang, rumusan masalah pada penelitian ini adalah sebagai berikut:

1. Bagaimana membangun model *Ensemble Learning* yang mampu memprediksi nilai pH selama fermentasi kopi berdasarkan data suhu dan waktu fermentasi yang diperoleh melalui sistem pemantauan berbasis *IoT*?
2. Bagaimana performa model *Ensemble Learning*, yaitu *Random Forest* dan *XGBoost*, dalam memprediksi nilai pH fermentasi kopi berdasarkan data suhu dan waktu fermentasi yang diperoleh melalui sistem pemantauan berbasis *IoT*?

1.3 Batasan Penelitian

1. Ruang lingkup penelitian hanya pada tahap proses fermentasi basah kopi arabika.
2. Variabel yang diteliti berupa pH, suhu dan waktu.
3. Model prediksi dibuat dengan metode *Random Forest* dan *XGBoost*.

1.4 Tujuan Penelitian

Adapun tujuan yang ingin dicapai dari penelitian ini adalah sebagai berikut:

1. Membangun model *Ensemble Learning* yang mampu memprediksi nilai pH selama proses fermentasi biji kopi berdasarkan data suhu dan waktu fermentasi yang diperoleh melalui sistem pemantauan berbasis *IoT*.
2. Menganalisis performa model *Ensemble Learning*, yaitu *Random Forest* dan *XGBoost*, dalam memprediksi nilai pH fermentasi kopi berdasarkan data suhu dan waktu fermentasi yang diperoleh melalui sistem pemantauan berbasis *IoT*.

1.5 Manfaat Penelitian

Adapun manfaat yang diharapkan dapat diperoleh dari penelitian ini adalah sebagai berikut:

1. Bagi industri.
Hasil penelitian ini diharapkan dapat digunakan untuk pemantauan fermentasi secara *real-time* dan otomatis.
2. Bagi peneliti
Menjadikan penelitian ini sebagai referensi untuk penelitian khususnya dalam memprediksi pH fermentasi berbasis metode *Ensemble Learning*.
3. Bagi penulis
Menambah wawasan dan pengalaman tentang proses fermentasi dengan berbasis metode *Ensemble Learning*.

BAB 2. TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian tentang prediksi pH fermentasi sudah dilakukan dengan beberapa metode *Machine learning*, contohnya penelitian oleh (Tirado-Kulieva et al., 2024) memanfaatkan *Near-Infrared Spectroscopy (NIRS)* dan *Partial Least Squares Regression (PLSR)* untuk memprediksi perubahan pH dan *Total Soluble Solids (TSS)* selama fermentasi kopi. Dataset terdiri dari 2100 profil spektrum pada rentang 1100–2100 nm, ditambah data pH dan *TSS* hasil pengukuran laboratorium. Profil diekstraksi dari tiga varietas kopi (*Typica*, *Caturra*, dan *Catimor*), mencakup tujuh waktu fermentasi. Penelitian ini menggunakan teknik seleksi fitur seperti β coefficients, *Variable Importance Projection (VIP)*, dan *Covering Array Feature Selection (CAFS)*, dengan hasil terbaik diperoleh metode *CAFS* yang menghasilkan nilai R^2 sebesar 0.903 dan *RMSE* sebesar 0.096 pada varietas *Caturra* untuk prediksi pH, serta R^2 sebesar 0.922 dan *RMSE* sebesar 0.688 pada varietas *Typica* untuk prediksi *TSS*. Berikutnya penelitian oleh (Saetae, 2025) berkontribusi dalam pengembangan *framework Machine learning* berbasis *hybrid Long Short-Term Memory (LSTM)* dan *Support Vector Regression (SVR)* untuk memprediksi pertumbuhan mikroba dan proses *acidification* pada fermentasi yogurt di suhu industri. Model *hybrid LSTM* digunakan untuk memprediksi pertumbuhan *Streptococcus thermophilus* dan *Lactobacillus delbrueckii subsp. bulgaricus*, sedangkan model *SVR* digunakan untuk memprediksi perubahan pH selama fermentasi. Hasil penelitian menunjukkan bahwa model *hybrid LSTM* memberikan performa terbaik pada suhu 37°C dengan nilai R^2 sebesar 0.9982 dan *RMSE* sebesar 0.0074 untuk *Lactobacillus delbrueckii subsp. bulgaricus*. Selain itu, model *SVR* menunjukkan performa terbaik dalam prediksi *acidification* dengan nilai R^2 mencapai 0.9999 dan *RMSE* sebesar 0.0094 pada suhu 45°C. Selanjutnya penelitian oleh (Adeleke et al., 2025) berkontribusi dalam pengembangan sistem fermentasi cerdas berbasis *Internet of Things (IoT)* dan *Machine learning* untuk mengontrol tingkat keasaman fermentasi amasi dengan model *Convolutional Neural Network*

(*CNN*), *Feedforward Neural Network (FNN)*, dan *Random Forest*. Hasil penelitian menunjukkan bahwa model *CNN* memperoleh performa terbaik dengan nilai R^2 sebesar 0.9475 dalam prediksi *Total Titratable Acidity (TTA)*. Penelitian berikutnya oleh (Anker et al., 2024) berkontribusi dalam penerapan *Machine learning regression* untuk memprediksi *aroma partitioning* pada matriks produk susu dengan membandingkan beberapa algoritma seperti *Linear Regression (LR)*, *Random Forest Regressor (RFR)*, *Gradient Boosting Regression (GBR)*, *XGBoost (XGB)*, *Support Vector Regression (SVR)*, dan *Artificial Neural Network Regression (NNR)*. Hasil penelitian menunjukkan bahwa model *Random Forest Regressor* memberikan performa terbaik pada suhu 37°C dengan R^2 sebesar 0.99, sedangkan *XGBoost* unggul pada suhu 7°C dengan R^2 sebesar 0.88. Uraian lengkap penelitian yang telah dijelaskan sebelumnya disajikan pada Tabel 2.1.

Tabel 2.1 Penelitian Terdahulu.

NO	Tujuan	Input	Metode Dan Hasil
1	Prediksi perubahan pH fermentasi kopi (Tirado-Kulieva et al. 2024)	Dataset terdiri dari 2100 profil spektrum pada rentang 1100–2100 nm, ditambah data pH dan TSS hasil pengukuran laboratorium. Profil diekstraksi dari tiga varietas kopi, mencakup tujuh waktu fermentasi	Regresi: <i>PLSR</i> . Hasil: - <i>CAFS + PLSR</i> (pH): $R^2 = 0.825\text{--}0.903$. $RMSE = 0.096\text{--}0.158$ - <i>CAFS + PLSR</i> (TSS): $R^2 = 0.865\text{--}0.922$. $RMSE = 0.688\text{--}1.059$
2	Memprediksi pertumbuhan mikroba dan <i>acidification</i> pada fermentasi yogurt di suhu industri Saetae (2025)	Data pertumbuhan <i>Streptococcus thermophilus</i> , <i>Lactobacillus bulgaricus</i> , pH, dan suhu fermentasi (37°C, 41°C, 45°C)	Regresi: <i>Hybrid LSTM, SVR</i> Hasil: - <i>Hybrid LSTM</i> (Growth) $R^2 = 0.9982$, $RMSE = 0.0074$ pada 37°C. - <i>SVR</i> (pH) $R^2 = 0.9999$, $RMSE = 0.0094$ pada 45°C
3	Mengembangkan sistem <i>IoT</i> berbasis <i>ML</i> untuk prediksi serta kontrol tingkat keasaman fermentasi amasi secara <i>real-time</i>	Dataset fermentasi amasi sebanyak 549.483 data yang terdiri dari pH, <i>electrical conductivity (EC)</i> , suhu, status inokulasi, status pasteurisasi, waktu	Regresi/ <i>Deep Learning</i> : <i>CNN, Feedforward Neural Network (FNN), Random Forest</i> . Hasil: - <i>CNN</i> : $R^2 = 0.9475$. - <i>FNN</i> : $R^2 = 0.9374$ - <i>Random Forest</i> : $R^2 = 0.9242$

NO	Tujuan	Input	Metode Dan Hasil
	Adeleke et al. (2025)	fermentasi, dan nilai TTA dari 7 sampel fermentasi pada suhu 25°C, 27°C, dan 32°C	
4	Memprediksi nilai Kmg (koefisien partisi aroma) pada produk dairy (Anker et al., 2024)	Data sifat fisikokimia aroma: <i>log P</i> , <i>molecular weight</i> , <i>specific gravity</i> , suhu, komposisi matriks dairy	Regresi: <i>Linear Regression</i> , <i>Random Forest</i> , <i>Gradient Boosting</i> , <i>XGBoost</i> , <i>SVR</i> , <i>Neural Network</i> . Hasil: - <i>Linear Regression</i> : $R^2 = 0.865$ - <i>Random Forest</i> : $R^2 = 0.99$ - <i>Gradient Boosting</i> : $R^2 = 0.958$ - <i>XGBoost</i> : $R^2 = 0.880$ - <i>SVR</i> : $R^2 = 0.943$ - <i>Neural Network</i> : $R^2 = 0.924$

2.2 Fermentasi Kopi

Fermentasi merupakan proses biokimia yang melibatkan aktivitas mikroorganisme seperti *Saccharomyces cerevisiae*, *Lactobacillus plantarum*, dan *Acetobacter spp.* untuk menguraikan senyawa kompleks dalam lendir buah kopi (*mucilage*) menjadi senyawa yang lebih sederhana. Proses ini juga berperan penting dalam pembentukan aroma, rasa, dan karakteristik sensorik kopi (De Bruyn et al., 2017). Selama fermentasi, bakteri dan ragi memetabolisme gula yang terdapat dalam lendir kopi, yang menyebabkan produksi asam dan penurunan pH (Cardoso et al., 2021). Penurunan pH ini dapat menghasilkan rasa yang lebih *citrus* dan asam, yang umumnya diinginkan dalam kopi. Tingkat pH yang terlalu rendah dapat menghasilkan rasa yang terlalu asam, sehingga pH menjadi titik kontrol kritis dalam produksi kopi spesialti (Elhalis et al., 2023). Selama fermentasi berlangsung, nilai pH menjadi indikator penting karena menunjukkan aktivitas mikroba dan tingkat pembentukan asam organik. Penurunan pH yang stabil menandakan fermentasi berjalan dengan optimal, sedangkan pH yang terlalu rendah dapat menyebabkan *over-fermentation* yang berdampak pada kualitas kopi (Silva et al., 2024).

Suhu dalam proses fermentasi dapat memengaruhi dinamika fermentasi kopi, terutama aktivitas mikroba, perubahan komposisi kimia dan kualitas sensorik

akhir biji kopi. Suhu merupakan faktor kunci yang menentukan laju dan arah metabolisme mikroorganisme seperti *Saccharomyces cerevisiae* dan *Lactobacillus plantarum*, yang berperan dalam konversi gula menjadi asam organik dan alkohol. Fermentasi pada suhu lebih tinggi, seperti 27°C, mempercepat degradasi sukrosa dan meningkatkan pembentukan senyawa volatil yang berkontribusi pada aroma kompleks dan skor *cupping* yang lebih tinggi dalam waktu lebih singkat (48 jam), sedangkan suhu lebih rendah seperti 18°C memerlukan durasi lebih lama untuk mencapai kualitas optimal (Costa et al., 2025). Pengendalian suhu juga terbukti menstabilkan komunitas mikroba dan mencegah fluktuasi suhu internal bioreaktor yang dapat mengganggu kinetika fermentasi. Waktu fermentasi merupakan faktor penting dalam pengolahan kopi, yang secara signifikan mempengaruhi kualitas dan karakteristik produk akhir. Durasi fermentasi sangat bervariasi dan tergantung pada suhu di mana proses dilakukan. Misalnya, fermentasi spontan mulai dari 12 hingga 19 jam. Fermentasi terkontrol pada 15 °C membutuhkan durasi yang lebih lama, sekitar 36 hingga 48 jam, sedangkan pada 30 °C memakan waktu sekitar 16 hingga 22 jam (Peñuela-Martínez et al., 2023). Aktivitas mikroorganisme selama proses fermentasi berperan dalam perubahan biokimia substrat lendir, yang menghasilkan senyawa yang meningkatkan aroma dan rasa kopi (Haile & Kang, 2019).

Fermentasi kopi menggunakan metode basah biasanya dilakukan secara spontan, dimana berbagai reaksi terjadi tergantung pada kondisi prosesnya yang menyebabkan hasil bervariasi (Poltronieri & Rossi, 2016). Dalam proses fermentasi basah, penghilangan mesokarp atau *mucilage* melalui fermentasi memiliki potensi besar untuk meningkatkan kualitas kopi karena terbentuknya senyawa yang dapat memperbaiki aroma dan rasa. Senyawa-senyawa ini terbentuk dari transformasi biokimia substrat *mucilage* yang dihasilkan oleh mikroorganisme yang ada (Haile & Kang, 2019).

Pengendalian suhu dapat memperpanjang proses fermentasi meskipun pengendalian ini dilakukan pada suhu yang lebih tinggi daripada suhu lingkungan. Selain itu, pengendalian suhu juga mempengaruhi perilaku parameter lainnya seperti pH, sehingga kondisi fermentasi yang terkontrol menunjukkan nilai

mendekati 4.00 dibandingkan dengan nilai kurang dari 3.50 pada fermentasi spontan (Peñuela-Martínez et al., 2023).

2.3 Parameter pH sebagai Indikator Mutu Fermentasi

pH memainkan peran penting dalam fermentasi kopi, bertindak sebagai indikator kunci dari proses dan memengaruhi transformasi kimia yang pada akhirnya membentuk profil sensorik kopi. Penurunan nilai pH adalah tanda pasti dari proses fermentasi aktif. Saat mikroorganisme memetabolisme gula dan senyawa lain dalam lendir kopi, mereka menghasilkan asam organik, yang menyebabkan penurunan pada pH (Rocha et al., 2024).

Selama proses fermentasi, kopi arabika memiliki nilai pH akhir berkisar antara 3.7 hingga 4.0. Penurunan pH selama fermentasi terutama disebabkan oleh pembentukan dan disosiasi asam, seperti asetikasi laktat, asetat, format, dan alkohol, dengan asam laktat sangat berpengaruh (Polanía Rivera et al., 2024). Skor kualitas cangkir tertinggi adalah 85 poin *SCA* dan sesuai dengan fermentasi kopi yang dikontrol suhunya pada ketinggian tertinggi (Peñuela-Martínez et al., 2023).

2.4 Pemantauan Suhu Menggunakan Sistem *IoT*

Internet of Things (IoT) mengacu pada konsep integrasi perangkat fisik dengan jaringan internet yang memungkinkan proses komunikasi serta pertukaran informasi berlangsung secara otomatis tanpa keterlibatan manusia secara langsung (Ashton, 2009; (Atzori et al., 2010). *IoT* memiliki potensi besar di berbagai bidang, seperti pertanian, kesehatan, industri dan lainnya (Khanna & Kaur, 2019).

Sensor suhu yang umum digunakan dalam sistem *IoT* untuk pemantauan fermentasi antara lain DS18B20, DHT22 dan SHT31. Sensor DS18B20 banyak digunakan karena akurat, tahan air, dan mudah diintegrasikan dengan mikrokontroler. DHT22 memiliki keunggulan dapat mengukur suhu dan kelembaban sekaligus, sehingga cocok untuk fermentasi yang sensitif terhadap kelembaban. Untuk tingkat presisi tinggi, SHT31 menjadi pilihan ideal karena mampu memberikan hasil dengan ketelitian $\pm 0.3^{\circ}\text{C}$.

Arsitektur sistem *IoT* pada proses fermentasi umumnya terdiri dari empat lapisan utama, yaitu *node* sensor, *gateway*, *server/cloud*, dan *dashboard*. Pada tahap *node* sensor, perangkat seperti ESP32 mengumpulkan data suhu dan pH dari lingkungan fermentasi menggunakan sensor seperti DS18B20. Data kemudian dikirim ke *gateway* melalui protokol komunikasi seperti *WiFi*, *ZigBee*, atau *LoRa*, sebelum diteruskan ke *server/cloud* seperti *ThingsBoard*, *Ubidots*, atau *AWS IoT* untuk disimpan dan dianalisis. Hasil pemantauan akhirnya ditampilkan pada *dashboard* yang dapat diakses melalui komputer atau ponsel secara *real-time*, menggunakan platform seperti *Blynk*, sehingga operator dapat memantau proses fermentasi dari mana saja dan kapan saja.

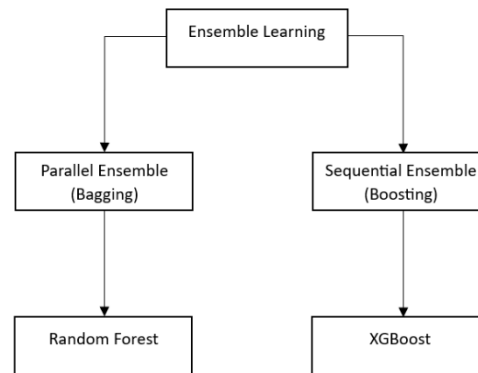
Penerapan *IoT* dalam proses fermentasi memiliki banyak kelebihan, antara lain kemampuan *monitoring* dan kontrol *real-time*, efisiensi tenaga kerja, serta peningkatan konsistensi kualitas produk. Dengan sistem *IoT*, operator dapat mendeteksi perubahan suhu, pH, dan kelembaban secara langsung dan melakukan tindakan korektif sebelum kualitas produk menurun. Namun demikian, *IoT* juga memiliki beberapa keterbatasan seperti ketergantungan pada konektivitas jaringan, biaya awal instalasi sensor dan server yang cukup tinggi, serta kebutuhan kalibrasi rutin agar data tetap akurat. Selain itu, keamanan data dan privasi juga menjadi aspek penting yang perlu diperhatikan dalam sistem berbasis *cloud*.

IoT telah banyak diterapkan dalam berbagai proses fermentasi di sektor pangan dan agroindustri. Pada fermentasi tempe, sensor DHT22 digunakan untuk memantau suhu dan kelembaban selama proses pengolahan kedelai agar dapat dikendalikan secara *real-time* (Maghfira et al., 2025). Sementara pada fermentasi kopi, sensor suhu DS18B20 dan sensor pH digunakan untuk memantau suhu dan pH selama proses fermentasi kopi (Vinanjana, 2023).

2.5 *Ensemble Learning*

Ensemble Learning adalah pendekatan dalam *ML* yang menggabungkan beberapa model (*base learners*) untuk menghasilkan prediksi yang lebih akurat dan stabil dibandingkan model tunggal (Zhou, 2012). Berdasarkan mekanisme pembentukan model, *Ensemble Learning* diklasifikasikan menjadi tiga kategori

utama, yaitu *bagging*, *boosting*, dan *stacking* (Mienye & Sun, 2022). Arsitektur umum *Ensemble Learning* berdasarkan mekanisme pembentukan model ditunjukkan pada Gambar 2.1



Gambar 2.1 Arsitektur *Ensemble Learning*

Berdasarkan gambar 2.1, *bagging* adalah metode *ensemble* yang menggabungkan beberapa model (*base learners*) yang dilatih secara independen pada *subset* data yang diambil secara acak dengan pengembalian (*bootstrap sampling*). Setiap model menghasilkan prediksi, lalu hasilnya digabungkan melalui *voting* (untuk klasifikasi) atau rata-rata (untuk regresi). Contoh model *bagging* yaitu *Random Forest* dan Extra Trees (*Extremely Randomized Trees*). *Boosting* membangun model secara berurutan (*sequential*). Setiap model berikutnya berusaha memperbaiki kesalahan (*residual*) dari model sebelumnya. Bobot data yang salah diprediksi akan dinaikkan agar model berikutnya lebih fokus pada titik sulit tersebut. Prediksi akhir diperoleh dari kombinasi berbobot semua model. Contoh model *boosting* yaitu *Gradient Boosting*, *XGBoost* dan *LightGBM*. Sementara itu, *stacking* menggabungkan beberapa model dari jenis berbeda seperti *Random Forest*, *XGBoost*, *SVM*, *Neural Network*. Hasil prediksi dari masing-masing model dasar kemudian digunakan sebagai masukan bagi sebuah *meta-learner* yang bertugas mempelajari kombinasi terbaik untuk menghasilkan prediksi akhir.

2.6 *Random Forest*

Random Forest adalah metode *Ensemble Learning* berbasis *bagging* yang dikembangkan oleh Breiman untuk meningkatkan akurasi model klasifikasi dan

regresi. *Random Forest* membangun banyak pohon keputusan pada *subset* data yang berbeda dan menggabungkan hasil prediksi dari masing-masing pohon untuk menghasilkan nilai rata-rata yang akurat (Breiman, 2001). Rumus *Random Forest*

$$\hat{y} = \frac{1}{N} \sum_{i=1}^N f_i(x) \quad (1)$$

Keterangan:

$\hat{y}$: prediksi akhir

N : jumlah pohon

$f_i(x)$: prediksi dari pohon ke- i untuk *input* x

2.7 *Extreme Gradient Boosting (XGBoost)*

Extreme Gradient Boosting (XGBoost) adalah versi penyempurnaan dari algoritma *Gradient Boosting Machine (GBM)* yang dikembangkan oleh Chen dan Guestrin. *XGBoost* menggunakan prinsip *boosting* dimana model dilatih secara bertahap, dengan setiap model baru berfokus memperbaiki kesalahan dari model sebelumnya. Keunggulan utama *XGBoost* terletak pada kemampuan regularisasi, *missing value handling*, serta efisiensi komputasi yang tinggi melalui optimasi paralel (Chen & Guestrin, 2016). Rumus *XGBoost*

$$\hat{y}_i = \sum_{k=1}^K f_k(x_i), f_k \in \mathcal{F} \quad (2)$$

dengan $\mathcal{F}$ adalah himpunan semua pohon regresi.

Fungsi objektif *XGBoost*:

$$Obj(\theta) = \sum_{i=1}^n l(y_i, \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad (3)$$

dengan

$$\Omega(f_k) = \gamma T + \frac{1}{2} \lambda \sum_{j=1}^T w_j^2 \quad (4)$$

Keterangan:

$(\hat{y}_i, y_i)$: *training loss*

$\Omega(f_k)$: kompleksitas model pohon

f_k : pohon regresi

T : jumlah daun pada pohon

w_j : bobot pada *node* daun pohon regresi

γ dan λ : parameter regularisasi

2.8 Evaluasi

Evaluasi perlu dilakukan untuk mengetahui performa model prediksi.

2.8.1 *R-Squared (R^2)*

Kemampuan model prediktif untuk menjelaskan variasi pada data aktual diukur dengan Koefisien Determinasi (*R-Squared*). R^2 memiliki rentang nilai antara 0 hingga 1. Nilai yang mendekati 1 menunjukkan bahwa model dapat menjelaskan sebagian besar variasi pada data target, sedangkan nilai yang mendekati 0 menunjukkan bahwa kemampuan penjelasan yang rendah.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (6)$$

Keterangan:

$\sum_{i=1}^n (y_i - \hat{y}_i)^2$: galat prediksi

$\sum_{i=1}^n (y_i - \bar{y})^2$: variabilitas total data

y_i : nilai aktual dari data

$\hat{y}_i$: nilai prediksi dari model

$\bar{y}$: rata-rata dari nilai aktual y

2.8.2 *Root Mean Square Error (RMSE)*

Besarnya selisih rata-rata antara nilai aktual dan nilai prediksi model diukur dengan *Root Mean Square Error (RMSE)*, yang memiliki satuan yang sama dengan variabel target.

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (5)$$

Keterangan:

n : jumlah data (sampel)

y_i : nilai aktual pada data ke- i

$\hat{y}_i$: nilai prediksi dari model pada data ke- i

$(y_i - \hat{y}_i)^2$: selisih kuadrat antara nilai aktual dan prediksi

BAB 3. METODOLOGI PENELITIAN

3.1 Jenis Penelitian

Jenis penelitian yang digunakan pada penelitian ini yaitu penelitian kuantitatif. Penelitian ini menggunakan metode *Random Forest* dan *XGBoost* bertujuan untuk memberikan solusi praktis dalam pengendalian mutu fermentasi biji kopi.

3.2 Objek Penelitian

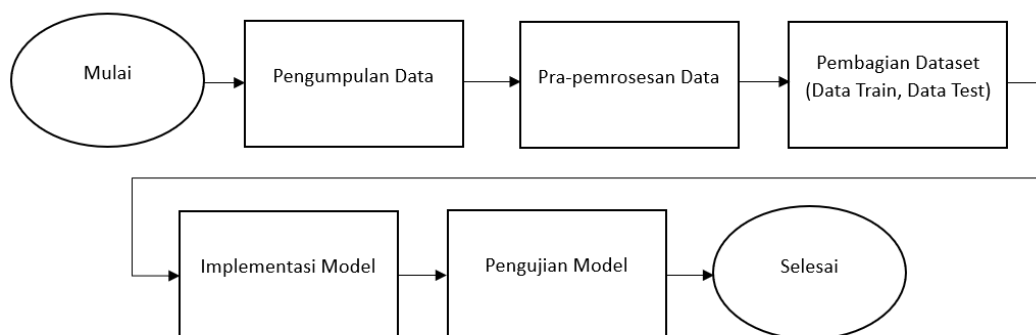
Objek penelitian ini menggunakan fermentasi biji kopi arabika yang didapat dari PT Perkebunan Nusantara I (PTPN I) Regional 5, Kebun Kalisat Jampit, Bondowoso, Jawa Timur

3.3 Tempat dan Waktu Penelitian

Penelitian ini dilakukan di PTPN I Regional 5 dan Fakultas Ilmu Komputer Universitas Jember. Waktu penelitian dilakukan pada bulan Juni 2025.

3.4 Tahapan Penelitian

Tahapan dalam penelitian ini dapat dilihat pada gambar 3.1



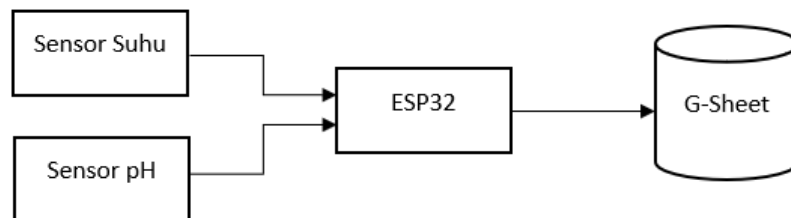
Gambar 3.1 Tahapan penelitian

Alur tahapan penelitian ini dimulai dengan proses pengumpulan data, yang kemudian dilanjutkan dengan pra-pemrosesan data. Setelah itu, dilakukan

pembagian dataset dan dilanjutkan dengan implementasi model. Seluruh model kemudian akan diuji pada tahap pengujian model untuk menarik kesimpulan.

3.4.1 Pengumpulan Data

Pengumpulan data dilakukan di PTPN I Regional 5 Kebun Kalisat Jampit menggunakan *IoT* yang terdiri dari ESP32 sebagai mikrokontroler, sensor suhu DS18B20 dan sensor pH. Proses pengumpulan data dilakukan dengan menempatkan sensor pada kolam fermentasi dengan 11 titik berbeda secara bergantian. Pada setiap titik pengumpulan data dilakukan dengan interval waktu setiap 5 detik selama ± 1 menit melalui mikrokontroler yang terhubung dengan Google Sheets, sehingga pada setiap titik diperoleh 12 data dengan total 1452 data. Pengumpulan data tersebut dilakukan secara berkala setiap ± 3 jam, dengan total durasi selama ± 30 jam. Rancangan alat dapat dilihat pada Gambar 3.2.



Gambar 3.2 Rancangan alat pengumpulan data

3.4.2 Pra-pemrosesan Data

Data dari hasil pengumpulan masih berupa data mentah yang berpotensi mengandung *noise*, *missing value*, *redundancy* dan *outlier*, sehingga diperlukan tahapan *preprocessing* sebelum digunakan dalam pemodelan. Proses *preprocessing* meliputi proses imputasi data dengan mengisi nilai 0 dan nilai kosong dengan median masing-masing barisnya. Selanjutnya dilakukan penanganan *outlier* menggunakan metode *Interquartile Range (IQR)*. Data yang berada di luar batas bawah dan batas atas dianggap sebagai *outlier*, kemudian dihapus agar tidak memengaruhi performa model prediksi. Tahapan *preprocessing* ini dilakukan untuk meningkatkan kualitas data sehingga proses pelatihan model *Random Forest* dan *XGBoost* dapat menghasilkan prediksi yang lebih akurat dan stabil.

3.4.3 Pembagian Dataset

Pembagian dataset akan dilakukan dengan beberapa uji coba, yakni dengan perbandingan *train:test* sebesar 80:20, 70:30 dan 60:40. Tujuan dari percobaan ini untuk mencari ukuran dataset yang memberikan hasil terbaik pada model.

3.4.4 Implementasi Model

Pada tahap implementasi langkah yang diambil adalah melakukan prediksi menggunakan *Ensemble Learning*, yaitu *Random Forest* dan *XGBoost*. Setiap model akan dilatih dengan 3 proporsi dataset, yakni 80:20, 70:30 dan 60:40. *Hyperparameter* yang akan diterapkan dalam model *Random Forest* dan *XGBoost* dapat dilihat pada Tabel 3.1.

Tabel 3.1 Skenario pengujian model

Metode	Hyperparameter	Referensi
<i>Random Forest</i>	<i>n_estimators</i> = 500	(Ahmed, 2024)
	<i>max_depth</i> = 20	
	<i>min_samples_split</i> = 2	
<i>XGBoost</i>	<i>n_estimators</i> = 500	(Ahmed, 2024)
	<i>max_depth</i> = 5	
	<i>learning_rate</i> = 0.1	
	<i>min_child_weight</i> = 2	
	<i>subsample</i> = 0.8	
	<i>colsample_bytree</i> = 1	

3.4.5 Pengujian Model

Setelah implementasi model, pengujian model dilakukan untuk mengukur sejauh mana model *Ensemble Learning* mampu memprediksi nilai pH fermentasi kopi secara akurat berdasarkan data suhu dan waktu hasil pemantauan sistem *IoT*. Model akan diuji menggunakan beberapa metrik evaluasi, yaitu *Coefficient of Determination (R^2)* dan *Root Mean Square Error (RMSE)*. Penggunaan lebih dari satu metrik bertujuan untuk memberikan gambaran performa model yang lebih komprehensif.

BAB 4. Hasil dan pembahasan

4.1 Pengumpulan Data

Pengumpulan data dilakukan di PTPN I Regional 5 Kebun Kalisat Jampit dengan memanfaatkan teknologi *Internet of Things (IoT)* untuk memantau kondisi fermentasi kopi secara otomatis. Sistem *IoT* yang digunakan terdiri dari ESP32 sebagai mikrokontroler, sensor suhu DS18B20 untuk mengukur suhu fermentasi, serta sensor pH untuk mengukur tingkat keasaman selama proses fermentasi berlangsung. Untuk pengambilan sampel pada kolam fermentasi dapat dilihat pada Gambar 4.1.



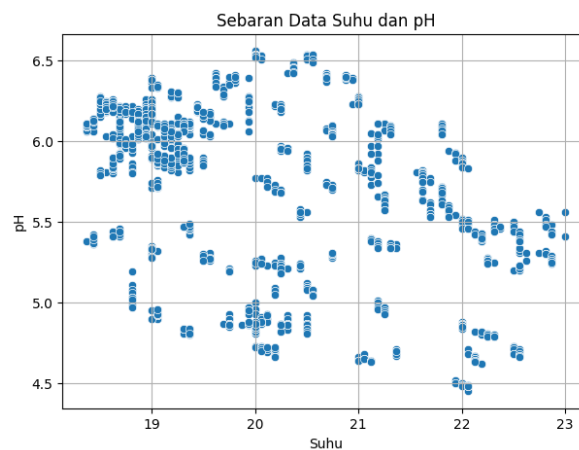
Gambar 4.1 Pengambilan data fermentasi kopi

Pengambilan data dilakukan pada kolam fermentasi kopi yang sedang digunakan dalam proses fermentasi basah. Pengukuran dilakukan pada 11 titik berbeda yang tersebar di dalam kolam fermentasi. Sensor dipindahkan secara bergantian pada setiap titik pengamatan untuk mendapatkan variasi data dari fermentasi dalam kolam. Pada setiap titik, pengukuran dilakukan setiap 5 detik selama kurang lebih 1 menit sehingga diperoleh sebanyak 12 data pengamatan pada masing-masing titik. Titik pengamatan pada kolam fermentasi bisa dilihat pada Gambar 4.2.



Gambar 4.2 Titik pengamatan pada kolam fermentasi

Proses pengambilan data dilakukan secara berkala dengan interval sekitar 3 jam selama proses fermentasi berlangsung. Pengamatan dilakukan selama kurang lebih 30 jam sehingga perubahan suhu dan pH yang terjadi pada berbagai fase fermentasi dapat terdokumentasi dengan baik. Data yang diperoleh kemudian dikirim secara otomatis oleh ESP32 ke Google Sheets sebagai media penyimpanan. Berdasarkan proses pengumpulan data tersebut, diperoleh total 1.452 data yang terdiri dari data suhu, pH dan waktu fermentasi. Sebelum digunakan dalam proses pengumpulan data, sensor pH terlebih dahulu dikalibrasi untuk memastikan hasil pengukuran yang diperoleh memiliki tingkat akurasi yang baik. Sebaran data suhu dan pH hasil pengumpulan data ditampilkan pada Gambar 4.3.



Gambar 4.3 *Plot* sebaran data suhu dan pH

Untuk mengetahui karakteristik awal data penelitian, dilakukan analisis statistik deskriptif. Variabel suhu memiliki nilai minimum sebesar 18.37°C, nilai maksimum 23°C, dengan rata-rata 20°C dan standar deviasi 1.218°C. Sementara itu, variabel pH memiliki nilai minimum 4.45, nilai maksimum 6.56, dengan rata-rata 5.7 dan standar deviasi 0.518. Statistik deskriptif tersebut menunjukkan karakteristik awal data yang akan digunakan pada tahapan pra-pemrosesan, meliputi penanganan data kosong dan *outlier* sebelum dilakukan pembentukan model. Ringkasan analisis statistik deskriptif data penelitian disajikan pada Tabel 4.1.

Tabel 4.1 Statistik deskriptif data penelitian

	Jumlah Data	Mean	Std. Deviation	Min	Q1	Median	Q3	Max
Suhu	1452	20.054	1.218	18.37	19.0	19.75	21.0	23.0
pH	1452	5.703	0.518	4.45	5.31	5.86	6.11	6.56
Waktu Fermentasi	1452	15	9.490	0.00	6.00	15.00	24.00	30.00

4.2 Pra-pemrosesan Data

Tahap pra-pemrosesan data dilakukan untuk meningkatkan kualitas dataset sebelum digunakan dalam proses pelatihan model *Ensemble Learning*. Data hasil pengumpulan dari sistem *IoT* tidak dapat langsung digunakan jika masih terdapat kemungkinan adanya data kosong (*missing value*) dan data pencilan (*outlier*) yang dapat memengaruhi kinerja model. Oleh karena itu, dilakukan pemeriksaan terhadap data untuk mengidentifikasi adanya *missing value* maupun *outlier* yang memerlukan penanganan.

4.2.1 Pemeriksaan Data Kosong (*Missing Value*)

Pemeriksaan *missing value* dilakukan terhadap seluruh variabel yang digunakan dalam penelitian, yaitu suhu, pH dan waktu fermentasi. Tahapan ini bertujuan untuk memastikan bahwa seluruh data hasil pengukuran menggunakan sistem *Internet of Things (IoT)* telah terekam secara lengkap sehingga tidak terdapat nilai yang hilang yang berpotensi memengaruhi proses pembelajaran model.

Hasil pemeriksaan menunjukkan bahwa seluruh data pada setiap variabel telah tercatat secara lengkap tanpa ditemukan nilai yang kosong. Dengan demikian, tidak diperlukan proses penanganan *missing value*, seperti penghapusan data maupun imputasi, karena seluruh data telah memenuhi syarat untuk digunakan pada tahapan selanjutnya. Hasil pemeriksaan data kosong dapat dilihat pada Tabel 4.2.

Tabel 4.2 Pemeriksaan *missing value*

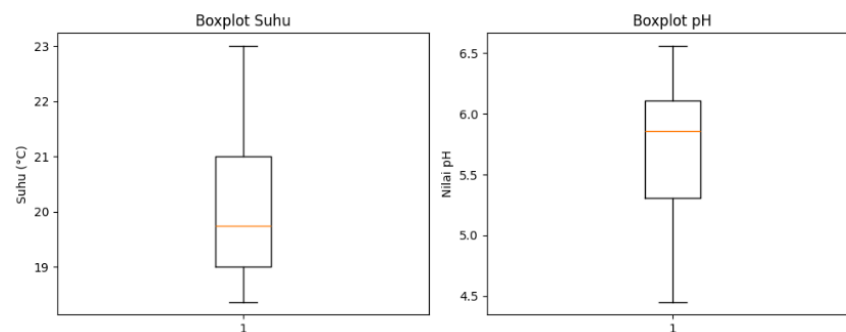
Pemeriksaan <i>Missing Value</i>	
Suhu	0
pH	0
Waktu Fermentasi	0

4.2.2 Pemeriksaan Data Pencilan (*Outlier*)

Setelah dilakukan pemeriksaan data kosong, tahapan berikutnya adalah identifikasi data pencilan (*outlier*). Pemeriksaan ini bertujuan untuk mengetahui apakah terdapat data yang menyimpang secara signifikan dari pola distribusi data sehingga berpotensi memengaruhi kinerja model prediksi. Pada penelitian ini, identifikasi *outlier* dilakukan menggunakan metode *Interquartile Range (IQR)* dengan menentukan batas bawah (*lower bound*) dan batas atas (*upper bound*) dari masing-masing variabel.

Berdasarkan hasil identifikasi menggunakan metode *IQR*, tidak ditemukan data yang berada di luar batas bawah maupun batas atas pada variabel suhu dan pH. Hal ini menunjukkan bahwa seluruh data masih berada dalam rentang distribusi yang wajar dan merepresentasikan kondisi fermentasi kopi selama proses pengambilan data berlangsung. Oleh karena itu, tidak dilakukan proses penghapusan data, sehingga seluruh data tetap digunakan pada tahapan pembagian dataset dan implementasi model *Random Forest* serta *XGBoost*. Visualisasi hasil identifikasi *outlier* menggunakan *boxplot* ditunjukkan pada Gambar 4.4.

Berdasarkan gambar tersebut, tidak terdapat titik data yang berada di luar *whisker boxplot* sehingga dapat disimpulkan bahwa dataset tidak mengandung *outlier*.



Gambar 4.4 *Boxplot* suhu dan pH

4.3 Pembagian Dataset

Setelah tahap pra-pemrosesan data, metode *train_test_split* dari pustaka *scikit-learn* digunakan untuk membagi kumpulan data menjadi set pelatihan dan set pengujian. Tiga rasio pembagian digunakan dalam penelitian ini 80:20, 70:30, dan 60:40. Rasio-rasio ini dipilih untuk membandingkan pengaruh ukuran data pelatihan dan data pengujian terhadap kinerja model. Nilai parameter *test_size* yang digunakan masing-masing sebesar 0.2, 0.3 dan 0.4, sesuai dengan proporsi data pengujian pada setiap skenario.

4.4 Implementasi Model

Setelah proses pembagian dataset, tahap selanjutnya adalah implementasi model. Implementasi model dilakukan dengan melatih model menggunakan data latih (*training data*) yang diperoleh dari setiap skenario pembagian dataset.

4.4.1 *Random Forest*

Random Forest merupakan algoritma *Ensemble Learning* yang bekerja dengan membangun sejumlah pohon keputusan (*decision tree*) dan menggabungkan hasil prediksi dari seluruh pohon tersebut untuk menghasilkan prediksi akhir. Parameter yang digunakan pada model *Random Forest* ditunjukkan

pada Tabel 3.1. Selain itu, parameter selain dari Tabel 3.1 juga akan dicoba untuk melatih model.

4.4.2 *XGBoost*

XGBoost (Extreme Gradient Boosting) merupakan algoritma *boosting* yang bekerja dengan membangun model secara bertahap untuk memperbaiki kesalahan prediksi pada model sebelumnya. Parameter yang digunakan pada model *XGBoost* ditunjukkan pada Tabel 3.1. Selain itu, parameter selain dari Tabel 3.1 juga akan dicoba untuk melatih model.

4.5 Pengujian Model

Tahap terakhir dalam penelitian ini adalah pengujian model yang bertujuan untuk mengetahui performa model dalam memprediksi nilai pH selama proses fermentasi kopi.

4.5.1 *Random Forest*

Selanjutnya, model *Random Forest* yang telah melalui proses pelatihan diuji menggunakan tiga skenario pembagian dataset, yaitu 80:20, 70:30, dan 60:40, untuk melihat performa model pada masing-masing kondisi. Untuk lebih lengkapnya dapat dilihat pada Tabel 4.3.

Tabel 4.3 Hasil pengujian model *Random Forest*

No	Parameter	Train			Test			Splitting
		R^2	RMSE	Time (s)	R^2	RMSE	Time (s)	
1	$n_estimators = 500$							
	$max_depth = 20$	0.9449	0.1221	0.98	0.9027	0.1575	0.18	80:20
	$min_samples_split = 2$							
2	$n_estimators = 400$							
	$max_depth = 20$	0.9449	0.1222	0.87	0.9032	0.1572	0.16	80:20
	$min_samples_split = 5$							
3	$n_estimators = 500$	0.9502	0.1176	0.99	0.8964	0.1586	0.18	70:30

No	Parameter	Train			Test			Splitting
		R^2	RMSE	Time (s)	R^2	RMSE	Time (s)	
	$max_depth = 20$ $min_samples_split = 2$							
	$n_estimators = 400$							
4	$max_depth = 20$ $min_samples_split = 5$	0.9502	0.1177	0.84	0.8966	0.1584	0.16	70:30
	$n_estimators = 500$							
5	$max_depth = 20$ $min_samples_split = 2$	0.9492	0.1185	1.96	0.9097	0.1516	0.18	60:40
	$n_estimators = 400$							
6	$max_depth = 20$ $min_samples_split = 5$	0.9490	0.1187	1.51	0.9094	0.1518	0.15	60:40

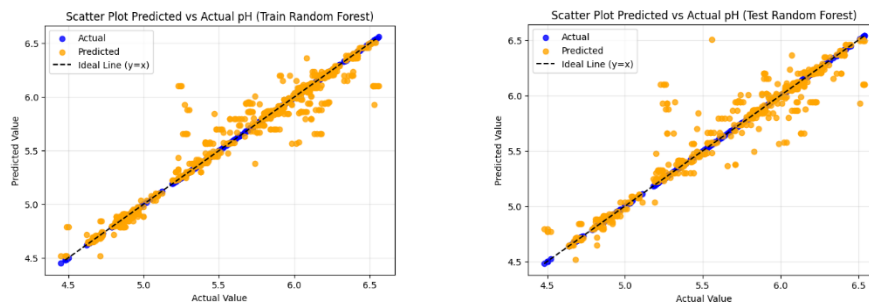
Berdasarkan hasil pengujian menggunakan parameter pada Tabel 3.1, model *Random Forest* menunjukkan performa yang baik pada seluruh skenario pembagian data. Nilai R^2 data uji yang diperoleh berturut-turut sebesar 0,9027 pada skenario 80:20, 0,8964 pada skenario 70:30, dan 0,9097 pada skenario 60:40. Sementara itu, nilai *RMSE* data uji masing-masing sebesar 0,1575, 0,1586, dan 0,1516. Hasil tersebut menunjukkan bahwa model mampu menjelaskan lebih dari 89% variasi data pH dengan tingkat kesalahan prediksi yang relatif rendah.

Setelah dilakukan pengujian menggunakan *Grid Search*, performa model *Random Forest* mengalami perubahan yang tidak terlalu signifikan. Nilai R^2 data uji pada skenario 80:20 meningkat menjadi 0,9032, pada skenario 70:30 menjadi 0,8966, sedangkan pada skenario 60:40 sedikit menurun menjadi 0,9094. Nilai *RMSE* yang diperoleh juga menunjukkan hasil yang relatif serupa dengan hasil pengujian sebelumnya, yaitu sebesar 0.1572, 0.1584, dan 0.1518.

Secara keseluruhan, skenario pembagian data 60:40 menghasilkan performa terbaik pada model *Random Forest* dengan nilai R^2 tertinggi dan *RMSE* terendah. Hasil ini menunjukkan bahwa proporsi data latih sebesar 60% sudah cukup untuk mempelajari pola hubungan antara waktu fermentasi, suhu, dan pH pada dataset

penelitian ini, sedangkan data uji sebesar 40% tetap mampu memberikan evaluasi yang baik terhadap kemampuan model dalam melakukan prediksi.

Pada tahap ini dilakukan perbandingan antara nilai pH aktual dan nilai pH prediksi menggunakan model *Random Forest* pada data *training* dan *testing*. Visualisasi hasil prediksi ditunjukkan pada gambar 4.5.



Gambar 4.5 *Scatter plot* prediksi dan aktual *Random Forest*

Berdasarkan gambar 4.5, sebagian besar titik pada data *training* berada dekat dengan garis ideal ($x = y$), yang menunjukkan bahwa model *Random Forest* mampu mempelajari hubungan antara suhu, waktu fermentasi, dan nilai pH dengan baik. Pada data *testing*, sebagian besar titik juga masih berada di sekitar garis ideal meskipun terdapat sedikit penyimpangan pada beberapa titik. Hasil ini sejalan dengan nilai R^2 sebesar 0,9097 dan $RMSE$ sebesar 0,1516 pada skenario pembagian data 60:40, sehingga menunjukkan bahwa model *Random Forest* mampu menghasilkan prediksi pH dengan tingkat akurasi yang baik.

4.5.2 *XGBoost*

Selanjutnya, model *XGBoost* yang telah melalui proses pelatihan diuji menggunakan tiga skenario pembagian dataset, yaitu 80:20, 70:30, dan 60:40, untuk melihat performa model pada masing-masing kondisi. Untuk lebih lengkapnya dapat dilihat pada Tabel 4.4.

Tabel 4.4 Hasil pengujian model *XGBoost*

No	Parameter	Train		Test			Splitting	
		R^2	RMSE	Time (s)	R^2	RMSE		Time (s)
1	$n_estimators = 500$	0.8628	0.1928	0.14	0.8257	0.2109	0.02	80:20
	$max_depth = 5$							
	$learning_rate = 0.1$							
	$min_child_weight = 2$							
	$subsample = 0.8$							
	$colsample_bytree = 1$							
2	$n_estimators = 400$	0.9444	0.1227	0.10	0.8989	0.1606	0.01	80:20
	$max_depth = 4$							
	$learning_rate = 0.2$							
	$min_child_weight = 3$							
	$subsample = 1.0$							
	$colsample_bytree = 1$							
3	$n_estimators = 500$	0.8718	0.1889	0.13	0.8066	0.2167	0.02	70:30
	$max_depth = 5$							
	$learning_rate = 0.1$							
	$min_child_weight = 2$							
	$subsample = 0.8$							
	$colsample_bytree = 1$							
4	$n_estimators = 400$	0.9498	0.1182	0.10	0.8931	0.1611	0.02	70:30
	$max_depth = 4$							
	$learning_rate = 0.2$							
	$min_child_weight = 3$							
	$subsample = 1.0$							
	$colsample_bytree = 1$							
5	$n_estimators = 500$	0.8758	0.1854	0.23	0.8048	0.2229	0.02	60:40
	$max_depth = 5$							
	$learning_rate = 0.1$							
	$min_child_weight = 2$							
	$subsample = 0.8$							
	$colsample_bytree = 1$							
6	$n_estimators = 400$	0.9489	0.1188	0.21	0.9040	0.1563	0.02	60:40
	$max_depth = 4$							

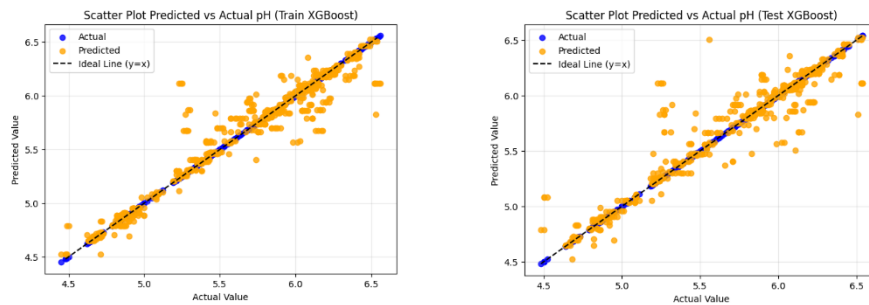
No	Parameter	Train			Test			Splitting
		R^2	RMSE	Time (s)	R^2	RMSE	Time (s)	
	$learning_rate = 0.2$							
	$min_child_weight = 3$							
	$subsample = 1.0$							
	$colsample_bytree = 1$							

Berdasarkan hasil pengujian menggunakan parameter pada Tabel 3.1, model *XGBoost* menghasilkan nilai R^2 data uji sebesar 0,8257 pada skenario 80:20, 0,8066 pada skenario 70:30, dan 0,8048 pada skenario 60:40. Nilai *RMSE* data uji yang diperoleh masing-masing sebesar 0,2109, 0,2167, dan 0,2229. Hasil ini menunjukkan bahwa performa *XGBoost* masih berada di bawah *Random Forest* ketika menggunakan parameter awal dari (Ahmed, 2024).

Setelah dilakukan pengujian menggunakan *Grid Search*, performa model *XGBoost* mengalami peningkatan yang cukup signifikan. Nilai R^2 data uji meningkat menjadi 0,8989 pada skenario 80:20, 0,8931 pada skenario 70:30, dan 0,9040 pada skenario 60:40. Selain itu, nilai *RMSE* data uji menurun menjadi 0,1606, 0,1611, dan 0,1563 pada masing-masing skenario pembagian data.

Meskipun demikian, performa terbaik *XGBoost* setelah optimasi masih sedikit berada di bawah *Random Forest*. Hal ini menunjukkan bahwa pada dataset penelitian yang berjumlah 1.452 data, *Random Forest* cenderung lebih mampu melakukan generalisasi dibandingkan *XGBoost*. Selain itu, seperti pada *Random Forest*, pembagian data 60:40 juga memberikan hasil terbaik pada model *XGBoost* dengan nilai R^2 tertinggi dan *RMSE* terendah.

Pada tahap ini juga dilakukan perbandingan antara nilai pH aktual dan nilai pH prediksi menggunakan model *XGBoost* pada data *training* dan *testing*. Visualisasi hasil prediksi ditunjukkan pada gambar 4.6.



Gambar 4.6 *Scatter plot* prediksi dan aktual *XGBoost*

Berdasarkan Gambar 4.6, sebagian besar titik pada data training berada di sekitar garis ideal ($y = x$), yang menunjukkan bahwa model *XGBoost* mampu mempelajari pola hubungan antara variabel masukan dan nilai pH. Pada data *testing*, sebagian besar titik juga berada di sekitar garis ideal meskipun masih terdapat beberapa titik yang memiliki selisih prediksi terhadap nilai aktual. Setelah dilakukan optimasi menggunakan *Grid Search*, model *XGBoost* mampu menghasilkan prediksi yang lebih mendekati nilai aktual dengan nilai R^2 sebesar 0,9058 dan $RMSE$ sebesar 0,1549 pada skenario pembagian data 60:40. Meskipun performanya meningkat secara signifikan, hasil prediksi *Random Forest* masih lebih baik dibandingkan *XGBoost*.

Hasil pengujian menunjukkan bahwa pendekatan *Random Forest* bekerja lebih baik daripada metode *XGBoost*, sebagaimana dibuktikan oleh nilai R^2 yang lebih tinggi dan nilai $RMSE$ yang lebih rendah. Hal ini disebabkan oleh karakteristik dataset yang digunakan dalam penelitian, yang terdiri dari 1.452 data dan dua variabel prediktor yaitu suhu dan waktu fermentasi. Dalam hal ini, pendekatan *Random Forest* mengurangi varians dengan menggabungkan beberapa pohon keputusan melalui proses *bagging* yang meningkatkan kinerja generalisasi dan menghasilkan model yang lebih stabil. Di sisi lain, *XGBoost* menggunakan teknik *boosting* yang secara bertahap membangun pohon keputusan, yang umumnya memberikan manfaat lebih besar untuk dataset yang lebih kompleks atau memiliki banyak data dan atribut. Oleh karena itu, model *Random Forest* lebih cocok untuk memprediksi nilai pH selama proses fermentasi kopi, sebagaimana ditunjukkan oleh hasil evaluasi model terbaik dalam penelitian ini.

BAB 5. KESIMPULAN DAN SARAN

5.1 KESIMPULAN

1. Berdasarkan hasil penelitian, model *Ensemble Learning* untuk memprediksi nilai pH selama fermentasi kopi berdasarkan data suhu dan waktu fermentasi yang diperoleh melalui sistem pemantauan berbasis *IoT* berhasil dibangun menggunakan algoritma *Random Forest* dan *XGBoost*.
2. Berdasarkan hasil evaluasi model, *Random Forest* memberikan performa terbaik pada skenario pembagian dataset 60:40 menggunakan parameter (Ahmed, 2024), dengan nilai R^2 sebesar 0.9097 dan *RMSE* sebesar 0.1516. Sementara itu, penggunaan *Grid Search* berhasil meningkatkan performa *XGBoost* hingga mencapai nilai R^2 sebesar 0.9040 dan *RMSE* sebesar 0.1563 pada skenario pembagian dataset yang sama. Berdasarkan hasil evaluasi tersebut, *Random Forest* merupakan model yang paling optimal untuk memprediksi pH fermentasi kopi pada dataset penelitian ini.

5.2 SARAN

Berdasarkan hasil penelitian yang telah dilakukan, beberapa saran yang dapat diberikan untuk pengembangan penelitian selanjutnya adalah sebagai berikut:

1. Menambah jumlah data fermentasi kopi sehingga model dapat mempelajari pola perubahan pH secara lebih baik dan menghasilkan prediksi yang lebih akurat.
2. Menambahkan variabel lain yang berpotensi memengaruhi proses fermentasi kopi, seperti kelembapan, total padatan terlarut (*Total Dissolved Solids/TDS*), kadar gula, atau jenis mikroorganisme yang terlibat dalam fermentasi.
3. Menguji metode machine learning lainnya, seperti *Support Vector Regression (SVR)*, *Artificial Neural Network (ANN)*, atau metode *ensemble* lainnya dan atau dengan pendekatan *deep learning*.

4. Mengembangkan sistem agar tidak hanya melakukan *monitoring*, tetapi juga mampu melakukan prediksi secara *real-time* dan menampilkan hasil prediksi langsung pada *dashboard monitoring*.
5. Melakukan optimasi *hyperparameter* yang lebih luas menggunakan metode seperti *Random Search*, *Bayesian Optimization*, atau *Genetic Algorithm* untuk memperoleh konfigurasi model yang lebih optimal.

Daftar Pustaka

- Adeleke, I., Adebo, O. A., & Nwulu, N. (2025). Leveraging IoT and machine learning for smart fermentation of amasi: A predictive framework for acidity control. *Journal of Agriculture and Food Research*, 24, 102409. <https://doi.org/10.1016/j.jafr.2025.102409>
- Adeleke, I., Nwulu, N., & Adebo, O. A. (2023). INTERNET OF THINGS (IoT) in the food fermentation process: A bibliometric review. *Journal of Food Process Engineering*, 46(5), e14321. <https://doi.org/10.1111/jfpe.14321>
- Ahmed, N. S. (2024). Machine Learning Models for Pavement Structural Condition Prediction: A Comparative Study of Random Forest (RF) and eXtreme Gradient Boosting (XGBoost). *Open Journal of Civil Engineering*, 14(04), 570–586. <https://doi.org/10.4236/ojce.2024.144031>
- Anker, M., Borsum, C., Zhang, Y., Zhang, Y., & Krupitzer, C. (2024). Using a Machine Learning Regression Approach to Predict the Aroma Partitioning in Dairy Matrices. *Processes*, 12(2), 266. <https://doi.org/10.3390/pr12020266>
- Ashton, K. (2009). That "Internet of Things" Thing. *RFID Journal*, 22(7), 97–114.
- Atzori, L., Iera, A., & Morabito, G. (2010). The Internet of Things: A survey. *Computer Networks*, 54(15), 2787–2805. <https://doi.org/10.1016/j.comnet.2010.05.010>
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32. <https://doi.org/10.1023/A:1010933404324>
- Cardoso, W. S., Agnoletti, B. Z., De Freitas, R., De Abreu Pinheiro, F., & Pereira, L. L. (2021). Biochemical Aspects of Coffee Fermentation. In L. Louzada Pereira & T. Rizzo Moreira (Eds.), *Quality Determinants In Coffee Production* (pp. 149–208). Springer International Publishing. https://doi.org/10.1007/978-3-030-54437-9_4
- Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. <https://doi.org/10.1145/2939672.2939785>
- Costa, G. X. R., Bertarini, P. L. L., Silva, L. C. F., Amaral, L. R. D., Gomes, M. D. S., De Oliveira, L. M., & Santos, L. D. (2025). The impact of ambient temperature regulation on arabica coffee fermentation. *Applied Food Research*, 5(1), 101028. <https://doi.org/10.1016/j.afres.2025.101028>
- De Bruyn, F., Zhang, S. J., Pothakos, V., Torres, J., Lambot, C., Moroni, A. V., Callanan, M., Sybesma, W., Weckx, S., & De Vuyst, L. (2017). Exploring the Impacts of Postharvest Processing on the Microbiota and Metabolite Profiles during Green Coffee Bean Production. *Applied and Environmental Microbiology*, 83(1), e02398-16. <https://doi.org/10.1128/AEM.02398-16>
- Dhanya, V. G., Subeesh, A., Kushwaha, N. L., Vishwakarma, D. K., Nagesh Kumar, T., Ritika, G., & Singh, A. N. (2022). Deep learning based computer vision approaches for smart agricultural applications. *Artificial Intelligence in Agriculture*, 6, 211–229. <https://doi.org/10.1016/j.aiia.2022.09.007>

- Elhalis, H., Cox, J., & Zhao, J. (2023). Coffee fermentation: Expedition from traditional to controlled process and perspectives for industrialization. *Applied Food Research*, 3(1), 100253. <https://doi.org/10.1016/j.afres.2022.100253>
- Haile, M., & Kang, W. H. (2019). The Role of Microbes in Coffee Fermentation and Their Impact on Coffee Quality. *Journal of Food Quality*, 2019, 1–6. <https://doi.org/10.1155/2019/4836709>
- Indonesia Investment (2024). Coffee: World Coffee Market.
- Khanna, A., & Kaur, S. (2019). Evolution of Internet of Things (IoT) and its significant impact in the field of Precision Agriculture. *Computers and Electronics in Agriculture*, 157, 218–231. <https://doi.org/10.1016/j.compag.2018.12.039>
- Maghfira, S. P., Suprianto, B., Rakhmawati, L., & Firmansyah, R. (2025). Rancang Bangun Sistem Kontrol Suhu dan Kelembapan Berbasis IOT dengan Fuzzy Logic untuk Optimasi Proses Fermentasi pada Pengolahan Tempe. *JURNAL TEKNIK ELEKTRO*, 14(3), 257–263. <https://doi.org/10.26740/jte.v14n3.p257-263>
- Mienye, I. D., & Sun, Y. (2022). A Survey of Ensemble Learning: Concepts, Algorithms, Applications, and Prospects. *IEEE Access*, 10, 99129–99149. <https://doi.org/10.1109/ACCESS.2022.3207287>
- Motta, I. V. C., Vuillermé, N., Pham, H.-H., & De Figueiredo, F. A. P. (2024). Machine learning techniques for coffee classification: A comprehensive review of scientific research. *Artificial Intelligence Review*, 58(1), 15. <https://doi.org/10.1007/s10462-024-11004-w>
- Peñuela-Martínez, A. E., Moreno-Riascos, S., & Medina-Rivera, R. (2023). Influence of Temperature-Controlled Fermentation on the Quality of Mild Coffee (*Coffea arabica* L.) Cultivated at Different Elevations. *Agriculture*, 13(6), 1132. <https://doi.org/10.3390/agriculture13061132>
- Polanía Rivera, A. M., López Silva, J., Torres-Valenzuela, L. S., & Plaza Dorado, J. L. (2024). Development of Starter Inoculum for Controlled Arabica Coffee Fermentation Using Coffee By-Products (Pulp and Mucilage Broth), Yeast, and Lactic Acid Bacteria. *Fermentation*, 10(10), 516. <https://doi.org/10.3390/fermentation10100516>
- Poltronieri, P., & Rossi, F. (2016). Challenges in Specialty Coffee Processing and Quality Assurance. *Challenges*, 7(2), 19. <https://doi.org/10.3390/challe7020019>
- Rocha, R. A. R., Cruz, M. A. D. D., Silva, L. C. F., Costa, G. X. R., Amaral, L. R., Bertarini, P. L. L., Gomes, M. S., & Santos, L. D. (2024). Evaluation of Arabica Coffee Fermentation Using Machine Learning. *Foods*, 13(3), 454. <https://doi.org/10.3390/foods13030454>
- Saetae, D. (2025). Machine learning-based prediction of microbial growth and acidification in yogurt fermentation at industrial temperatures. *LWT*, 231, 118326. <https://doi.org/10.1016/j.lwt.2025.118326>
- Silva, L. C. F., Pereira, P. V. R., Cruz, M. A. D. D., Costa, G. X. R., Rocha, R. A. R., Bertarini, P. L. L., Amaral, L. R. D., Gomes, M. S., & Santos, L. D. (2024). Enhancing Sensory Quality of Coffee: The Impact of Fermentation

- Techniques on *Coffea arabica* cv. Catiguá MG2. *Foods*, 13(5), 653. <https://doi.org/10.3390/foods13050653>
- Tirado-Kulieva, V., Quijano-Jara, C., Avila-George, H., & Castro, W. (2024). Predicting the evolution of pH and total soluble solids during coffee fermentation using near-infrared spectroscopy coupled with chemometrics. *Current Research in Food Science*, 9, 100788. <https://doi.org/10.1016/j.crfs.2024.100788>
- Vinanjana, S. N. (n.d.). *PROGRAM STUDI TEKNOLOGI INDUSTRI PERTANIAN FAKULTAS TEKNOLOGI PERTANIAN UNIVERSITAS JEMBER 2023*.
- Widyaningrum, W., Yohanes Aris Purwanto, Slamet Widodo, Supijatno, & Iriani, E. S. (2022). Portable/Handheld NIR sebagai Teknologi Evaluasi Mutu Bahan Pertanian secara Non-Destruktif. *Jurnal Keteknikaan Pertanian*, 10(1), 59–68. <https://doi.org/10.19028/jtep.010.1.59-68>
- Zhou, Z.-H. (2012). *Ensemble Methods: Foundations and Algorithms* (0 ed.). Chapman and Hall/CRC. <https://doi.org/10.1201/b12207>