JIRK
Journal of Innovation Research and Knowledge
Vol.4, No.5, Oktober 2024

ANALISIS PREDIKTIF MENGGUNAKAN MACHINE LEARNING UNTUK MENANGGULANGI MASALAH REJECT PRODUK PADA PROSES PRODUKSI PT. XYZ

Oleh
Adevian Fairuz Pratama[1], Dhebys Suryani[2], Firgi Sotya Izzudin[3], Adzikirani[4]
[1,2,3,4]Politeknik Negeri Malang
E-mail: [1]adevianfairuz@polinema.ac.id, [2]dhebys.suryani@polinema.ac.id, [3]zzuuddiinn28@gmail.com, [4]adzikirani@polinema.ac.id

Article History:
Received: 08-09-2024
Revised: 22-09-2024
Accepted: 11-10-2024

Keywords:
Reject, Keamanan Dan Kualitas, Machine Learning, XGBoost, MAPE

Abstract: *PT. XYZ adalah perusahaan makanan dan minuman di Indonesia, terkenal dengan minuman isotoniknya yang menggantikan cairan tubuh saat aktivitas fisik.Produk ini populer di kalangan konsumen lokal dan internasional karena formulanya yang inovatif. PT. XYZ sangat menekankan kualitas, sehingga produk yang tidak memenuhi standar ketat, seperti tekanan botol, akan di-reject untuk memastikan keamanan dan kualitas. Penelitian ini menggunakan metode machine learning, khususnya XGBoost, untuk analisis prediktif. Data diambil dari mesin Liquid Nitrogen dan Bottle Pressure Detector, mencakup pressure bottle, temperatur tangki, sensor liquid nitrogen, dan lainnya pada tahun 2023. Data diolahdan dianalisis untuk memastikan kesesuaian dengan kebutuhan PT. XYZ dan kecocokan dengan metode XGBoost. Business Understanding mengidentifikasi tujuan bisnis dan dampak deteksi tekanan botol. Data Understanding dilakukan dengan Exploratory Data Analysis (EDA). Data dibagi menjadi 80% untuk training dan 20% untuk testing. Data Preparation melibatkan preprocessing untuk membersihkan duplikasi dan missing value. Modeling menggunakan algoritma XGBoost dengan parameter learning_rate sebesar 0.5, gamma sebesar 2,max_depth sebesar 6, n_estimators sebesar 100, colsample_bytree sebesar 0.4, subsample sebesar 0.7, reg_lambda sebesar 3, dan min_child_weight sebesar 1 terbukti terbaik. Evaluasi menunjukkan nilai MAPE 6.88% dan akurasi 93.12%. Prediksi menunjukkan jumlah reject sekitar 23 pada 1 Januari 2024 pukul 07:00, dan sekitar 22 pukul 08:00. Rekomendasi untuk penelitian selanjutnya adalah menggunakan lebih banyak data dan menggabungkan XGBoost dengan model machine learning lainnya*

PENDAHULUAN

PT. XYZ adalah sebuah perusahaan yang beroperasi di sektor minuman dan makanan di Indonesia. Perusahaan ini terkenal dengan produk unggulannya, yaitu produk minuman isotonik yang dirancang khusus untuk menggantikan cairan tubuh yang hilang selama aktivitas fisik intens.

Produk dari PT. XYZ ini telah menjadi salah satu pilihan populer di kalangan konsumen di Indonesia dan di beberapa pasar internasional. Minuman ini dikenal karena formulanya yang inovatif dan kemampuannya untuk memberikan hidrasi optimal bagi para konsumennya.

Produk yang tidak memnuhi standar mutu yang tinggi dalam setiap tahapan produksinya, sehingga tidak semua produk yang diproduksi dapat mencapai tahap penjualan. Komitmen terhadap kualitas ini memastikan bahwa hanya produk-produk yang memenuhi standar ketat yang disebarluaskan ke pasar, sehingga konsumen dapat yakin akan keunggulan dan keamanan setiap produk dari PT. XYZ.Setiap produk yang tidak memenuhi standar mutu yang ditetapkan oleh PT.

XYZ akan secara tegas di-*reject*. Dalam proses pengawasan kualitas, terdapat berbagai jenis *reject* yang dapat terjadi, dan salah satu contohnya adalah *reject* yang terkait dengan masalah tekanan pada botol. Tindakan ini dilakukan guna memastikan bahwa setiap produk yang sampai ke konsumen adalah produk yang memenuhi tingkat kualitas tertinggi, termasuk dalam hal keamanan kemasan seperti tekanan pada botol.

Reject terhadap masalah tekanan pada botol ini menjadi langkah penting dalam menjaga keamanan dan kualitas produk. Hal ini dilakukan guna mencegah potensi risiko atau masalah yang dapat timbul pada saat penggunaan produk oleh konsumen. Dengan adanya pengawasan mutu yang ketat, PT. XYZ berkomitmen untuk memberikan produk berkualitas tinggi yang aman dan memenuhi harapan konsumen. Proses pengecekan ini menjadi salah satu bentuk dedikasi perusahaan terhadap kepuasan pelanggan dan integritas produk yang dihasilkan.

Penelitian ini secara khusus difokuskan pada analisis prediktif dengan penerapan metode machine learning, dengan algoritma *XGBoost* sebagai pilihan utama. Metode ini telah terbukti efektif dalam berbagai aplikasi, termasuk dalam prediksi dan klasifikasi data. Melalui pemahaman mendalam terhadap konsep-konsep dasar machine learning dan *XGBoost*, penelitian ini bertujuan untuk mengembangkan model prediktif yang akurat dan handal.

Langkah-langkah eksperimental akan mencakup Business Understanding, Data Understanding, Data Preparation, Modeling *XGBoost*, dan Evaluation Model. Analisis hasil yang mendalam juga akan dilakukan untuk mengevaluasi performa model, termasuk pengoptimalan parameter yang sesuai dengan karakteristik data yang digunakan. Penelitian ini tidak hanya bertujuan untuk menghasilkan model prediktif yang baik, tetapi juga untuk memberikan wawasan tentang faktor-faktor yang paling berpengaruh terhadap hasil prediksi menggunakan *XGBoost*.

Melalui penelitian ini, diharapkan dapat ditemukan solusi yang dapat diterapkan secara praktis dalam berbagai konteks, sekaligus memberikan kontribusi pada pemahaman yang lebih mendalam mengenai potensi dan batasan dari metode machine learning *XGBoost* dalam analisis prediktif. Selain itu, hasil dari penelitian ini diharapkan dapat memberikan manfaat nyata bagi pemangku kepentingan yang terlibat dalam pengambilan keputusan berbasis data.

ISSN 2798-3471 (Cetak)
ISSN 2798-3641 (Online)

LANDASAN TEORI

Pada penelitian sebelumnya mengenai cara mengetahui penyebab *reject* padaPT. XYZ. Peneliti menggunakan metode *Fault Tree Analysis (FTA)* untuk mengetahui penyebab *reject* pada produk dan dapat memberikan solusi pencegahan*reject*. Terdapat 5 jenis *reject* pada PT. Ustegra yaitu *reject* melembung dengan peluang terjadi sebesar 1,137%, *reject* kotoran bahan dengan peluang terjadi sebesar 1,18%, *reject* ambles dengan peluang terjadi sebesar 1,16%, *reject* hardness dengan peluang terjadi sebesar 1,54%, *reject* PU dengan peluang terjadi sebesar 1,152%. Akar masalah penyebab *reject* tersebut terjadi diantaranya operator terburu-buru, operator kelelahan, mesin yang memerlukan perawatan, deadline kerja produksi yang terlalu singkat, operator malas, sarana prasarana kebersihan lingkungan yang kurang lengkap, dan operator yang bergurau saat melakukan pekerjaannya. Adapun rekomendasi untuk mengurangi dan mencegah kelima *reject* tersebut terjadi kembali, diantaranya melakukan perawatan mesin secara berkala, memberikan estimasi waktu pengerjaan yang lebih lama kepada konsumen dari estimasi pengerjaan yang sebenarnya, memberikan reward bagi operator yang bisamemenuhi target tanpa melakukan kesalahan (Benedictus V.B.P, 2021).

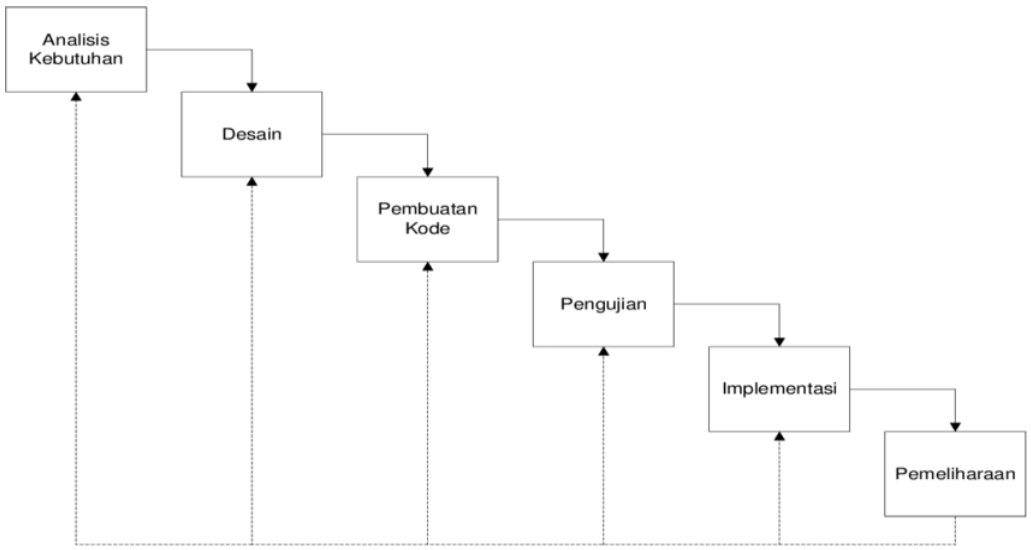
XGBoost juga pernah dilakukann untuk meramalkan kecelakaan lalu lintas menurut akibatnya. Dengan data dari BPS Provinsi Bali dengan periode tahun 1996 sampai dengan tahun 2019 dalam rentang waktu tahunan. Hasil dari penelitian ini, menunjukkan model *XGBoost* memiliki performa yang baik pada dua kategori yaitu kategori jumlah orang meninggal akibat kecelakaan dengan nilai RMSE 4.92 dan jumlah orang yang mengalami luka berat dengan nilai RMSE 4.11. Nilai RMSE model *XGBoost* untuk kategori jumlah kejadian kecelakaan lalu lintas yaitu sebesar 21.69 dan kategori orang yang mengalami luka ringan akibat kecelakaan yaitu sebesar 77.24 (Pandika Pinata, Sukarsa, & Rusjayanthi, 2020)

Penelitian sebelumnya tentang Prediksi Rating Aplikasi Playstore Menggunakan XGBoost, peneliti melakukan perbandingan dengan menggunakan beberapa model yaitu Decision Tree, K-Nearest Neighbors, Gradient Boosting, Random Forest, LightGBM, dan XGBoost. Dari hasil penelitian, menunjukkan bahwa algoritma XGBoost berkinerja lebih baik untuk prediksi rating aplikasi dan menghasilkan tingkat akurasi mencapai 77,5 % (Dwika Ananda Agustina Pertiwi & Muslim, 2020). Pada penelitian lainnya, mengenai Pengukuran Kadar Hemoglobin Non- Invasif Berbasis Android Menggunakan Algoritma Extreme Gradient Boosting. Peneliti menggunakan model algoritma XGBoost untuk mengukur kadar hemoglobin non-invasif. Dari hasil penelitian tersebut, didapatkanakurasi model algoritma XGBoost mencapai 94,91 % dengan nilai RMSE yaitu 0,80, yang berarti nilai RMSE mendekati 0 yang menunjukkan tingkat kesalahan yang rendah (Sri Dewi Sartika Syarifuddin & RendyMunad, 2023)

METODE PENELITIAN

Model Waterfall merupakan salah satu model SDLC yang sering digunakan dalam pengembangan sistem informasi atau perangkat lunak. Model inimenggunakan pendekatan sistematis dan berurutan. Tahapan dalam model ini dimulai dari tahap perencanaan hingga tahap pengelolaan (maintenance) dan dilakukan secara bertahap (Wahid, 2020). Model waterfall sering dianggap cocok untuk penelitian karena pendekatannya yang terstruktur

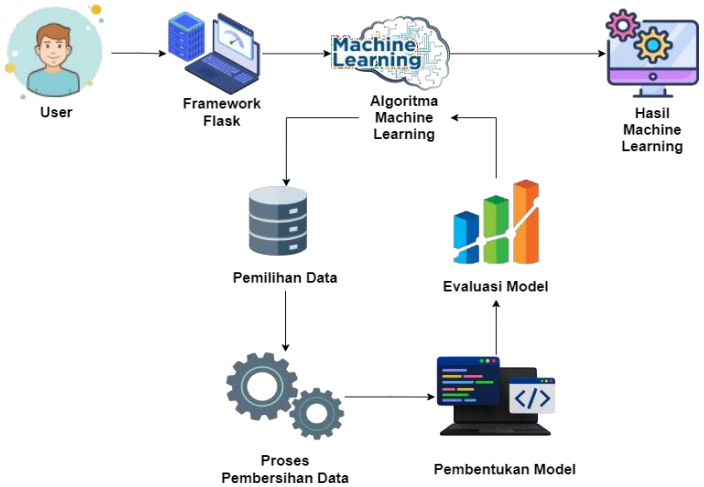
dan berurutan. Dalam model ini, setiap fasepengembangan perangkat lunak dilakukan secara berurutan dan tidak dapat dimulai sebelum fase sebelumnya selesai. Ini membuatnya cocok untuk penelitiankarena memungkinkan peneliti untuk memahami dan mengontrol setiap langkah dengan cermat. Dapat dilihat pada Gambar 1



Gambar 1. Model Waterfall

Desain Sistem

Website sistem prediksi nilai pressure ini nantinya akan memiliki beberapa fitur sederhana, untuk mengetahui proses system ini berjalan bisa dilihat pada Gambar 2:



Gambar 2 Sistem Diagram

Diagram sistem di atas adalah sistem pembelajaran mesin yang menggunakan framework Flask. Sistem ini terdiri dari tiga komponen utama, yaitu:

a. Data

Data merupakan komponen penting dalam sistem pembelajaran mesin. Datayang digunakan dalam sistem ini harus bersih dan berkualitas. Data dapat diperoleh dari berbagai sumber, seperti internet, database, atau sensor.

ISSN 2798-3471 (Cetak)
ISSN 2798-3641 (Online)

b. Algoritma

Algoritma pembelajaran mesin digunakan untuk mengolah data dan menghasilkan model pembelajaran. Model pembelajaran ini selanjutnya digunakan untuk melakukan prediksi atau klasifikasi.

c. Hasil

Hasil sistem pembelajaran mesin adalah prediksi atau klasifikasi yangdihasilkan oleh model pembelajaran. Hasil ini dapat digunakan untuk berbagai keperluan, seperti pengambilan keputusan, rekomendasi, atau pengendalian.

Berdasarkan gambar 3.4, proses pembelajaran mesin dalam sistem ini terdiri dari empat langkah, yaitu:

1. Pemilihan data

Pemilihan data merupakan langkah awal dalam proses pembelajaran mesin.Data yang dipilih harus sesuai dengan tujuan pembelajaran mesin.

2. Pembersihan data

Pembersihan data bertujuan untuk menghilangkan noise atau data yang tidak relevan dari data yang dipilih.

3. Pembentukan model

Pembentukan model merupakan langkah utama dalam proses pembelajaranmesin. Algoritma pembelajaran mesin digunakan untuk mengolah data danmenghasilkan model pembelajaran.

4. Evaluasi model

Evaluasi model bertujuan untuk mengukur kinerja model pembelajaran. Evaluasi model dilakukan dengan menggunakan data uji yang tidak digunakan dalam pembentukan model.

Secara umum, desain sistem ini dapat digunakan untuk berbagai jenis tugas pembelajaran mesin. Namun, untuk mendapatkan hasil yang optimal, sistem ini perlu dirancang dengan mempertimbangkan karakteristik data dan tujuan pembelajaran mesin.

Deskripsi Sistem

Sistem ini dirancang untuk memprediksi kemungkinan terjadinya produk *reject* sebelum produk tersebut dihasilkan, sehingga langkah-langkah pencegahan dapat dilakukan untuk mengurangi jumlah produk yang tidakmemenuhi standar kualitas.

Sistem prediktif ini akan mengumpulkan data historis dari proses produksi, termasuk data mengenai nilai temperature mesin dan parameterproses lainnya. Data tersebut akan digunakan untuk melatih model machine learning yang mampu mengidentifikasi pola dan faktor-faktor yang berkontribusi terhadap produk *reject*.

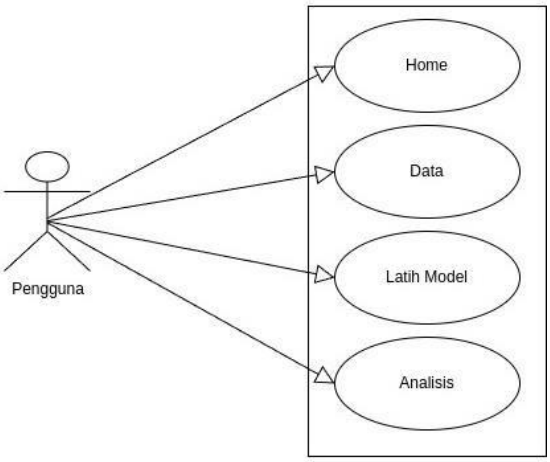
Analisis Pengguna

Pengguna adalah salah satu bagian yang penting dari suatu sistem. Karakteristik pengguna menjadi dasar sebuah perancangan antarmuka/interfacesistem. Dalam sistem ini pengguna yang terlibat hanyalah user. Dapat dilihat pada Tabel 1.

Tabel 1. Kebutuhan Pengguan

No	Kategori Pengguna	Keterangan
1	User	• Dalam system ini user melaukan prediksi jumlah *reject* debgan mengupload data produksi dengan format csv • Darihasil prediksi user melihat visualisasi jumlah *reject* beserta dengan nilai MAPE yang dihasilkan

Use Case Diagram

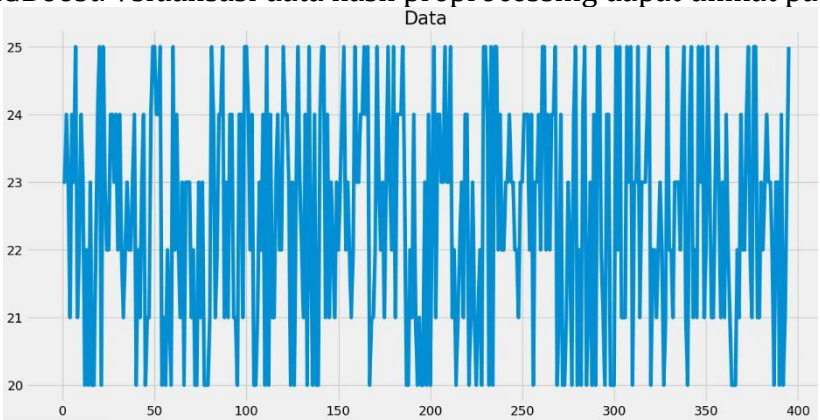


Gambar 3. Use Case Diagram

HASIL DAN PEMBAHASAN

Implementaasi Model XGBoost

Data yang telah dipreprocessing selanjutnya akan digunakan untuk pemodelan menggunakan XGBoost. Vsiualisasi data hasil preprocessing dapat dilihat pada Gambar 5.3.



Gambar 4. Grafik Data Jumlah *Reject* Produk

ISSN 2798-3471 (Cetak)
ISSN 2798-3641 (Online)

Tahap selanjutnya adalah merubah data menjadi data matrix atau data frame dengan bantuan function exprs agar dapat dianalisis lebih lanjut dengan model XGBoost. Data yang sudah berbentuk data frame atau data matrix merupakan syarat untuk mengolah data sebelum melakukan pemodelan, penulis harus melakukan pembagian data training dan data testing. Pembagian data training dan data testing dapat dilihat pada Tabel 2.

Tabel 1. Ratio Pembagian Data

	Traning	Testing
Ratio	80%	20%
Jumlah	316	80

Dimana data training merupakan kumpulan dataset yang akan digunakan untuk membentuk model klasifikasi dalam memprediksi kelas data baru dan data testing merupakan kumpulan dataset baru yang akan digunakan untuk mengukur sistem dapat melakukan klasifikasi dengan benar. Penulis menggunakan ratio untuk data latih dan data uji sebesar 80% : 20%. dengan jumlah sample sebanyak 396 dari data. Jumlah masing-masing pembagian data training dan testing dijelaskan pada table 5.1. Visualisasi hasil pembagian data dapat dilihat pada Gambar 5.4.

Gambar 5. Grafik Pembagian Data Jumlah *Reject* Produk

Tahap selanjutnya yaitu melakukan permodelan menggunakan XGBoost. Untuk tahap pemodelan XGBoost dibutuhkan parameter seperti learning_rate (eta), gamma, sub_sample, n_estimator, min_child_weight, max_depth, colsample_bytree, dan lambda. Parameter learning_rate (eta) berfungsi untuk mengontrol seberapa cepat atau lambat model belajar, gamma menentukan minimal loss reduction yang diperlukan untuk membuat split tambahan di pohon keputusan,sub_sample berfungsi sebagai proporsi dari sampel data yang digunakan untuk melatih setiap pohon. Nilai yang lebih rendah dapat mengurangi overfitting tetapi dapat meningkatkan varians. n_estimators adalah jumlah pohon keputusan yang ingin dibangun. Menentukan seberapa banyak iterasi atau pohon yang akan dibuat dalam proses pelatihan. min_child_weight adalah minimum jumlah total bobot (nilai Hessian) dari semua

contoh yang dibutuhkan untuk membagi suatu node. Membantu dalam mengontrol overfitting. max_depth adalah maksimum kedalaman pohon yang diizinkan selama pelatihan. Memungkinkan model untuk menyesuaikan dengan pola yang lebih kompleks, tetapi juga meningkatkan risiko overfitting. colsample_bytree adalah proporsi dari fitur yang diambil secara acak untuk membangun setiap pohon. Membantu dalam meningkatkan keberagaman model. Terakhir adalah lambda adalah parameter regularisasi L2 (Ridge Regression) yang mengendalikan ukuran bobot untuk menghindari overfitting. Untuk nilai parameter yang digunakan dalam penilitian ini dapat dilihat pada tabel5.2.

Tabel 2 Parameter Model

Parameter	Nilai
learning_rate	0.1
gamma	5
lambda	3
sub_sample	0.7
n_estimators	100
min_child_weight	1
max_depth	6
colsample_bytree	0.4

Langkah selanjutnya proses perhitungan model XGBoost dilakukan menggunakan bahasa pemrograman Python. Hasil perhitungan ditunjukkan pada gambar 6

```
XGBRegressor
XGBRegressor(base_score=None, booster=None, callbacks=None,
             colsample_bylevel=None, colsample_bynode=None,
             colsample_bytree=0.4, device=None, early_stopping_rounds=None,
             enable_categorical=False, eval_metric='rmse', feature_types=None,
             gamma=2, grow_policy=None, importance_type=None,
             interaction_constraints=None, learning_rate=0.5, max_bin=None,
             max_cat_threshold=None, max_cat_to_onehot=None,
             max_delta_step=None, max_depth=6, max_leaves=None,
             min_child_weight=1, missing=nan, monotone_constraints=None,
             multi_strategy=None, n_estimators=100, n_jobs=None,
             num_parallel_tree=None, random_state=None, ...)
```

Gambar 6 Model *XGBoost*

Pengujian

Pengujian Nilai MAPE

A. Nilai Mape Berdasarkan Rasio Data

Pengujian keakuratan pada prediksi menggunakan XGBoost di sistem ini dilakukan dengan menggunakan nilai MAPE (Mean Absolute Percentage Error). Semakin kecil nilai MAPE yang dihasilkan maka akan semakin bagus atau akurat pemodelan prediksi yang dihasilkan.Pengujian nilai MAPE dilakukan dengan menggunakan perhitungan pada Google Collab. Pada pengujian nilai MAPE sebelumnya dilakukan uji coba split data yaitu membagi data menjadi dua bagian data pelatihan (data training) dan data uji (data testing). Proses ini bertujuan untuk mengevaluasi performa model XGBoost untuk meraih prediksi yangakurat.

ISSN 2798-3471 (Cetak)
ISSN 2798-3641 (Online)

Proses ini bertujuan untuk mengevaluasi performa model XGBoost dalam berbagai konteks pembagian data, sehingga dapat diidentifikasi rasio yang paling optimal untuk meraih prediksi yang akurat. Hasil skenario uji coba berbagai rasio dapat dilihat pada tabel 5.4.

Tabel 2 Uji Coba MAPE Berdasarkan Rasio Data Berdasarkan hasil pengujian yang ditampilkan pada Tabel 5.4,

pembagian data dengan rasio 80:20 menunjukkan performa terbaik dibandingkan dengan rasio pembagian lainnya. Hal ini dapat dilihat dari nilai Mean Absolute Percentage Error (MAPE) yang paling rendah yaitu7.19% serta nilai akurasi yang paling tinggi mencapai 92.81%. Dibandingkan dengan rasio lainnya, seperti 70:30, 60:40, dan 50:50, rasio 80:20 menunjukkan efisiensi yang lebih baik dalam memprediksi data dengan kesalahan yang lebih kecil dan keakuratan yang lebih tinggi. Dengan demikian, dapat disimpulkan bahwa pembagian data dengan rasio80:20 merupakan pilihan terbaik untuk mencapai hasil yang optimal dalam konteks pengujian ini.

B. Nilai Mape Berdasarkan Kombinasi *Tuning Hyperparameter*

Dari hasil pengujian rasio pembagian data, dicoba untuk dilakukan pengujian dengan menambahkan *Tuning Hyperparameter.* Hal ini dilakukan untuk meningkatkan akurasi dari permodelan, sehingga permodelan dapat memprediksi dengan lebih akurat. Pengujian selanjutnya, merupakan hasil pengujian dari pengaruh penggunaan *hyperparameter* terhadap prediksi permodelan. Terdapat 8 parameteryang diperkirakan dapat meningkatkan kinerja model dalam memprediksimenggunakan metode XGBoost. *Tuning Hyperparameter* yang dilakukan pada 8 parameter ini menggunakan metode *randomized search CV*. *randomized search CV* memiliki kemungkinan lebih tinggi untuk menemukan kombinasi yang baik. Ini karena tidak semua parameter memiliki pentingan yang sama, dan *randomized search CV* dapat menemukan kombinasi yang optimal dengan lebih cepat. Hasil nilai hyperparameter terbaik sebagai berikut:

Tabel 3 Hasil Nilai Terbaik Tuning Hyperparameter

Hyperparameter	*Randomized Search Values*	Nilai *Hyperparameter* Terbaik	MAPE	Accuracy
learning_rate	0.01, 0.1, 0.2, 0.5	0.5	6.88%	93.12 %
gamma	1, 2, 3, 4, 5	2		
max_depth	3, 4, 5, 6, 7	6		
n_estimators	100, 200, 300, 400, 500, 600	100		
colsample_bytree	0.2, 0.4, 0.5, 0.6, 0.7	0.4		
subsample	0.2, 0.4, 0.5, 0.6, 0.7	0.7		
reg_lambda	1, 2, 3, 4, 5	3		
min_child_weight	1, 3, 5, 7	1		

Pada Tabel 3telah ditujukkan hasil dari pengujian dengan *Tuning Hyperparameter*, didapatkan nilai terbaik dari 8 parameter yaitu*learning_rate* sebesar 0.5, *gamma* sebesar 2, *max_depth* sebesar 6, *n_estimators* sebesar 100, *colsample_bytree* sebesar 0.4, *subsample* sebesar 0.7, *reg_lambda* sebesar 3, dan *min_child_weight* sebesar 1membuat nilai MAPE semakin rendah dari 7.19% menjadi mendapatkannilai MAPE sebesar 6.88% dan masuk kedalam kriteria MAPE baik. Kombinasi *Tuning Hyperparameter* ini meningkatkan nilai akurasi prediksi dari 92.81% menjadi sebesar 93.12%. Kombinasi inilah yang nantinya akan ditambahkan pada pemodelan untuk melakukan prediksi jumlah reject.

Hasil Pengujian Fungsionalitas

Kesimpulan yang didapatkan dari hasil pengujian sistem prediksi jumlah *reject* produk pada penelitian ini fitur sudah sesuai dengan fungsinya masing-masing.

Hasil Pengujian Nilai MAPE

Hasil pengujian untuk prediksi jumlah *reject* produk yang telah dilakukan menggunakan nilai MAPE, dapat dibedakan menjadi beberapa pengujian dengan variasi rasio data training dan testing yang berbeda, dapat ditarik kesimpulan bahwaefektivitas model ini bervariasi seiring dengan pembagian data yang digunakan. Proses evaluasi ini menunjukkan bahwa pemilihan proporsi data training dan testing memainkan peran krusial dalam menentukkan sejauh mana model mampu menggeneralisasi pola dari suatu data yang diberikan. Kinerja model *XGBoost* yang didapat tidak bersifat konstan dan dapat dipengaruhi oleh komposisi dataset yang digunakan untuk melatih dan menguji model. Oleh karena itu penting untuk menganalisis lebih lanjut mengenai strategi pembagian data yang optimal untuk meningkatkan performa prediktif model. Dari hasil pengujian dengan kombinasi *Tuning Hyperparameter* menghasilkan nilai terbaik dari beberapa parameter, yaitu *learning_rate* sebesar 0.5, *gamma* sebesar 2, *max_depth* sebesar 6, *n_estimators* sebesar 100, *colsample_bytree* sebesar 0.4, *subsample* sebesar 0.7, *reg_lambda* sebesar 3, dan *min_child_weight* sebesar 1. Kombinasi *Tuning Hyperparameter* iniberhasil menurunkan nilai MAPE dan meningkatkan akurasi prediksi. Hasil eksperimen menyoroti pentingnya pemahaman yang mendalam terhadap karakteristik dataset dan pemilihan metode validasi yang sesuai. Oleh sebab itu, perlu dipertimbangkan dengan cermat bagaimana pembagian data training dan testing agar dapat menghasilkan model *XGBoost* yang dapat diandalkan dan responsif terhadap variasi pola dalam data waktu. Dari hasil pengujian, didapat model terbaik seperti pada Tabel 4.

Tabel 4 Hasil Skenario Uji Coba Rasio dan *Tuning Hyperparameter* Terbaik

No	Skenario Uji	*Hyperparameter*	Nilai	MAPE	Akurasi
1	Menggunakan rasio pembagian data 80:20	*Learnning_rate*	0.5	6.88%	93.12%
		gamma	2		
		lambda	3		
		n_estimators	100		
		min_child_weight	1		
		max_depth	6		
		colsample_bytree	0.4		

ISSN 2798-3471 (Cetak)
ISSN 2798-3641 (Online)

Dari hasil skenario uji dengan rasio data training dan testing dengan kombinasi *Tuning Hyperparameter* pada model *XGBoost*, ditemukan bahwa performa model bervariasi tergantung pada pembagian data yang digunakan. Padaskenario uji untuk data *reject* produk, model terbaik yang dihasilkan adalah denganmenggunakan pembagian data 80% sebagai data training dan 20% sebagai data testing, dengan parameter *learning_rate* sebesar 0.5, *gamma* sebesar 2, *max_depth*sebesar 6, *n_estimators* sebesar 100, *colsample_bytree* sebesar 0.4, *subsample* sebesar 0.7, *reg_lambda* sebesar 3, dan *min_child_weight* sebesar 1. Hasilnya menunjukkan MAPE sebesar 6.88% dan akurasi 93.12%.

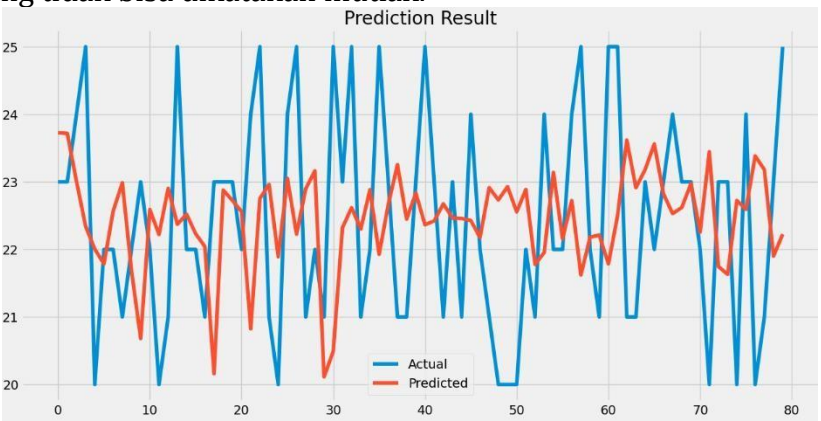
Untuk memahami kinerja model secara lebih mendalam, perbandingan antara data aktual dan prediksi perlu dianalisis. Evaluasi ini penting untuk mengevaluasi sejauh mana model dapat mereplikasi dan memprediksi data dengan akurasi yang optimal.

Tabel 5 Hasil Perbandingan Nilai Aktual dan Prediksi

No	Tanggal	Data Aktual	Prediksi	MAPE
1.	2023-11-03 11:00:00	23	21	
2.	2023-11-04 07:00:00	23	22	
3.	2023-11-04 09:00:00	24	25	
4.	2023-11-04 10:00:00	25	22	
5.	2023-11-04 11:00:00	20	23	
6.	2023-11-06 01:00:00	22	22	
7.	2023-11-06 10:00:00	22	21	
8.	2023-11-07 08:00:00	21	23	6.88%
9.	2023-11-07 09:00:00	22	23	
10.	2023-11-08 03:00:00	23	23	
11.	2023-11-11 09:00:00	22	21	
12.	2023-11-11 11:00:00	20	23	
13.	2023-11-12 05:00:00	21	22	
14.	2023-11-12 00:00:00	25	21	
15.	2023-11-13 09:00:00	22	23	

16.	2023-11-13 11:00:00	22	22	
17.	2023-11-14 00:00:00	21	22	
18.	2023-11-15 07:00:00	23	21	
19.	2023-11-16 02:00:00	23	22	
20.	2023-11-16 09:00:00	23	22	
21.	2023-11-16 11:00:00	22	24	
22.	2023-11-20 03:00:00	24	20	
23.	2023-11-20 05:00:00	25	22	
24.	2023-11-03 11:00:00	23	21	

Pada Tabel 5 terlihat bahwa hasil data prediksi yang diperoleh dari model dengan parameter *learning_rate* sebesar 0.5, *gamma* sebesar 2, *max_depth*sebesar 6, *n_estimators* sebesar 100, *colsample_bytree* sebesar 0.4, *subsample* sebesar 0.7, *reg_lambda* sebesar 3, dan *min_child_weight* sebesar 1 cukup bagusyang memiliki nilai MAPE sebesar 6.88%, dengan melalui beberapa proses metode *XGBoost* model tersebut memeuhi kriteria setiap prosesnya dibandingkandengan model yang lain. Jika dilihat dari data aktual dan data prediksi , terdapat data prediksi yang memiliki nilai lebih tinggi daripada data aktualnya, ada juga yang memiliki nilai lebih kecil dari pada nilai aktual. Hal tersebut cukup wajar mengingat bahwa ini adalah data prediksi yang tidak bisa dikatakan mutlak.



Gambar 7 Grafik Hasil Prediksi

Pada gambar 7 tersebut bisa dilihat prediksi yang dilakukan oleh metode *XGBoost* untuk memprediksikan jumlah *reject* produk dalam 2 jam kedepan. Untuk hasil prediksi yang dilakukan menghasilkan nilai sebesar 22.092 dibulatkan menjadi 22 pada tanggal 2 Januari 2024 pukul 00:00 dan tanggal 2 Januari 2024

ISSN 2798-3471 (Cetak)
ISSN 2798-3641 (Online)

pukul 01:00 sebesar 21.911 dibulatkan menjadi 22. Dalam dua jamkedepan dapat diprediksikan atau diramalkan terjadi kenaikan untuk jumlah*reject*.

KESIMPULAN

Berdasarkan dari rumusan masalah dan hasil penelitian yang dilakukan, makadapat disimpulkan bahwa :

1. Penerapan metode *XGBoost* untuk memprediksi jumlah *reject* produk pada data mesin produksi PT XYZ dilakukan melalui serangkaian langkah mulaidari preprocessing data, mengidentifikasi model, estimasi parameter model, pemilihan dan pelatihan model hingga evaluasi model. Berdasarkan proses yang telah dilakukan menunjukkan bahwa model dengan parameter *learning_rate* sebesar 0.5, *gamma* sebesar 2, *max_depth* sebesar 6, *n_estimators* sebesar 100, *colsample_bytree* sebesar 0.4, *subsample* sebesar 0.7, *reg_lambda* sebesar 3, dan *min_child_weight* sebesar 1 adalahyang terbaik. Model ini konsisten memenuhi kriteria mulai dari uji parameter hingga evaluasi model. Hasil ini menunjukkan bahwa pembagian data dengan rasio tertentu dapat memengaruhi performa model*XGBoost.*
2. Metode XGBoost terbukti dapat memberikan prediksi yang cukup akurat dalam memproyeksikan jumlah reject produk pada proses produksi PT XYZ. Dalam penelitian ini, XGBoost dengan parameter *learning_rate* sebesar 0.5, *gamma* sebesar 2, *max_depth* sebesar 6, *n_estimators* sebesar 100, *colsample_bytree* sebesar 0.4, *subsample* sebesar 0.7, *reg_lambda* sebesar 3, dan *min_child_weight* sebesar 1 menghasilkan nilai MAPE 6.88% dengan akurasi sebesar 93,12%. Prediksi untuk tanggal 2 Januari 2024 pukul 00:00 menunjukkan jumlah reject diperkirakan sekitar 22, sedangkan untuk tanggal 2 Januari 2024 pukul 01:00 diperkirakan sekitar 22

DAFTAR PUSTAKA

[1] Afikah Agustiningsih, Y. F., and I. A. Kautsar, "Classification of Vacational High School Graduates' Ability in Industry using Extreme Gradient Boosting (XGBoost), Random Forest And Logistic Regression: Klasifikasi Kemampuan Lulusan SMK di Industri Menggunakan Extreme Gradient Boosting (XGBoost), Random," *Jurnal Teknik Informatika (JUTIF)*, pp. 977-985, 2023.

[2] Ahmedbahaaaldin Ibrahem Ahmed Osman, A.-S., "Extreme gradient boosting (Xgboost) model to predict the groundwater levels in Selangor Malaysia," *Ain Shams Engineering Journal*, pp. 1545-1556, 2021.

[3] Althof Thabibi, R. S., "PERBANDINGAN MODEL MULTIPLE LINEAR REGRESSION DAN DECISION TREE REGRESSION (STUDI KASUS: PREDIKSI HARGA SAHAM TELKOM, INDOSAT, DAN XL)," *Jurnal Ilmiah Teknologi dan Rekayasa*, pp. 78-92, 2023.

[4] Benedictus V.B.P, R. R., "ANALISA PENCEGAHAN REJECT PADA PRODUKSI ROLL KARET DENGAN FAULT TREE ANALYSIS (FTA) DENGAN REKOMENDASI PERBAIKAN FAILURE MODE AND EFFECTS ANALYSIS (FMEA) DI PT. USTEGRA (USAHA TEHNIK GRAFIKA)," *Jurnal Manajemen Industri dan Teknologi*, pp. 25-37, 2021.

[5] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *ACM SIGKDD*

International Conference on Knowledge Discovery and Data Mining, pp. 785-794, 2016.
[6] D. Chicco, M. Warrens, and G. Jurman, "The coefficient of determination R-squared is more informative than SMAPE, MAE, MAPE, MSE, and RMSE in regression analysis evaluation," *PeerJ Computer Science*, 2021.
[7] Dwika Ananda Agustina Pertiwi, T. M., and M. A. Muslim, "Prediksi Rating Aplikasi Playstore Menggunakan," *Seminar Nasional Ilmu Komputer (SNIK 2020)*, 2020.
[8] T. Fawcett, "An introduction to ROC analysis," *Pattern Recognition Letters*, vol. 27, pp. 861–874, 2006.
[9] J. Friedman, "GREEDY FUNCTION APPROXIMATION: A GRADIENT BOOSTING MACHINE," *Annals of Statistics*, pp. 1189-1232, 2001.
[10] A. A. Hania, "Mengenal Artificial Intelligence, Machine," *Jurnal Teknologi Indonesia*, 2017.
[11] I. Hanif, "Implementing Extreme Gradient Boosting (XGBoost) Classifier to Improve Customer Churn Prediction," *Proceedings of the 1st International Conference on Statistics and Analytic*, 2019.
[12] I. Goodfellow and Y. B. Bengio, "Deep Learning," MIT Press, 2016.
[13] J. Bergstra and Y. B. Bengio, "Random Search for Hyper-Parameter Optimization," *Journal of Machine Learning Research*, vol. 13, pp. 281-305, 2012.
[14] B. Jange, "Prediksi Harga Saham Bank BCA Menggunakan XGBoost," *Journal of Economics and Accounting*, pp. 231-237, 2022.
[15] I. M. Karo Karo, "Implementasi Metode XGBoost dan Feature Importance untuk Klasifikasi pada Kebakaran Hutan dan Lahan," *Journal of Software Engineering, Information and Communication Technology*, pp. 11-18, 2020.
[16] R. Kohavi, "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection," *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pp. 1137-1143, 1995.
[17] M. Ulfah Siregar, P. H., "Housing Price Prediction Using a Hybrid Genetic Algorithm with Extreme Gradient Boosting," *The 2022 International Conference on Computer, Control, Informatics and Its Applications (IC3INA)*, 2022.
[18] Pandika Pinata, N. N., M. Sukarsa, and N. D. Rusjayanthi, "Prediksi Kecelakaan Lalu Lintas di Bali dengan XGBoost pada Python," *JURNAL ILMIAH MERPATI*, 2020.
[19] S. Dewi Sartika Syarifuddin, A. K., and S. RendyMunad, "Sistem Informasi Pengukuran Kadar Hemoglobin Non-Invasif Berbasis Android Menggunakan Algoritma Extreme Gradient Boosting," *Komputika Jurnal Sistem Komputer*, pp. 13-23, 2023.
[20] A. A. Wahid, "Analisis Metode Waterfall Untuk Pengembangan Sistem Informasi," *Jurnal Ilmu-ilmu Informatika dan Manajemen STMIK*, 2020

ISSN 2798-3471 (Cetak)
ISSN 2798-3641 (Online)