e-ISSN 2549-7472
Edumatic: Jurnal Pendidikan Informatika
Vol. 9 No. 1, April, 2025, **Hal. 51-59**
DOI: 10.29408/edumatic.v9i1.28988

Peningkatan Akurasi Prediksi Curah Hujan menggunakan Gradient Boosting dan CatBoost dengan Pendekatan Voting Classifier

Dina Fudhlatina [1,*] , Fikri Budiman [1]

[1] Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Indonesia
[*] Correspondence: 111202113626@mhs.dinus.ac.id

Received: 23 December 2025 | Revised: 30 December 2024 | Accepted: 20 Januari 2025 | Published: 9 April 2025

Abstrak

Prediksi curah hujan yang akurat sangat penting untuk pertanian, mitigasi bencana, dan pengelolaan sumber daya air, terutama dalam menghadapi dampak perubahan iklim. Penelitian ini bertujuan untuk peningkatan akurasi prediksi curah hujan menggunakan *gradient boosting* dan *CatBoost* dengan pendekatan *voting classifier*. Data yang digunakan pada penelitian ini berjumlah 1.461 berdasarkan data cuaca dari BMKG Kota Semarang (2020-2023). Data dianalisis menggunakan algoritma *gradient boosting* dan *CatBoost* dalam kerangka *Voting Classifier*. Fitur *input* meliputi suhu (Tn, Tx, Tavg), kelembapan (RH_avg), curah hujan (RR), durasi sinar matahari (ss), kecepatan angin (ff_x, ff_avg), dan arah angin (ddd_x). Teknik *GridSearchCV* digunakan untuk optimasi *hyperparameter*. Model memprediksi berdasarkan kategori intensitas curah hujan seperti: tanpa hujan, hujan ringan, hujan sedang, hujan lebat, dan hujan ekstrem. Hasil menunjukkan model dengan optimasi dan pendekatan *ensemble* mencapai *accuracy* 87,89%, *precision* 0,88, *recall* 0,88, *f1-score* 0,88, dan *Cohen's Kappa* 0,8486. Sedangkan *gradient boosting* dan *CatBoost* secara individu menghasilkan *accuracy* 75,99% dan 85,68%. Dengan fitur *input* data, model memprediksi kategori hujan ekstrem yang sesuai dengan data aktual. Penelitian ini memberikan kontribusi penting pada pengembangan sistem peringatan dini cuaca, mitigasi bencana, dan pengelolaan iklim.

Kata kunci: *catboost*; *gradient boosting*; optimasi *hyperparameter;* prediksi curah hujan; *voting lassifier*

Abstract

Accurate rainfall prediction is essential for agriculture, disaster mitigation, and water resource management, especially in the face of climate change impacts. This research aims to improve the accuracy of rainfall prediction using Gradient Boosting and CatBoost with a voting classifier approach. The data used in this study amounted to 1,461 based on weather data from BMKG Semarang City (2020-2023). The data was analyzed using the Gradient Boosting and CatBoost algorithms with a voting classifier framework. The input features include temperature (Tn, Tx, Tavg), humidity (RH_avg), rainfall (RR), length of irradiation (ss), wind speed (ff_x, ff_avg), and wind direction (ddd_x). The GridSearchCV technique was used for hyperparameter optimization. The model predicts based on rainfall intensity categories, namely no rain, light rain, moderate rain, heavy rain, and extreme rain. The results showed that the model with optimization and ensemble approach achieved 87.89% accuracy, 0.88 precision, 0.88 recall, 0.88 f1-score, and 0.8486 cohen's kappa. Meanwhile, gradient boosting and CatBoost individually produced 75.99% and 85.68% accuracy. With these data input features, the model is able to predict extreme rainfall categories that match the actual data. This research is an important contribution to the development of early weather warning systems, disaster mitigation, and climate management.

Keywords: *catboost; gradient boosting; hyperparameter optimation; rainfall prediction; voting lassifier*

PENDAHULUAN

Curah hujan merupakan salah satu faktor penting dalam pemantauan kondisi cuaca dan iklim yang sangat mempengaruhi berbagai aspek kehidupan manusia, terutama dalam konteks perubahan iklim dan peningkatan frekuensi fenomena cuaca ekstrem (Afifah et al., 2024; Mabruroh & Wiyanto, 2023; Suwarman et al., 2022). Perubahan iklim global telah menyebabkan pola hujan menjadi semakin tidak menentu dan sulit diprediksi (Purify et al., 2024). Ketidakpastian ini sering kali memicu bencana alam, seperti banjir dan tanah longsor, yang berdampak signifikan pada lingkungan dan kehidupan manusia (Azhari et al., 2023; Sari et al., 2024). Indonesia menghadapi kerugian ekonomi kurang lebih Rp 10 triliun akibat banjir yang disebabkan oleh curah hujan ekstrem (Kusuma et al., 2022). Salah satu penyebab utama adalah rendahnya akurasi sistem prediksi curah hujan yang tersedia, khususnya dalam mengolah kompleksitas data iklim yang melibatkan berbagai variabel, seperti kelembapan, suhu, dan pola angin.

Mengatasi masalah tersebut, penelitian ini menerapkan kombinasi algoritma *gradient boosting* dan *CatBoost* dalam kerangka *voting classifier*. Hasil mengenai adanya sistem prediksi curah hujan yang akurat dapat membantu berbagai sektor meliputi pertanian, perhutanan, perencanaan perkotaan, penanganan bencana, dan juga pengelolaan sumber daya air (Hayu et al., 2024; Runtulalo & Manongga, 2024).

Gradient boosting dan *CatBoost* merupakan algoritma *machine learning* yang telah banyak digunakan dalam prediksi curah hujan berkat kemampuannya dalam menangani dataset yang kompleks (Jasman et al., 2022; Pahlevi et al., 2024). *Gradient boosting* dikenal efektif meningkatkan kinerja model melalui pendekatan *ensemble* yang iteratif (Ananda et al., 2024; Azhari & Hidajat, 2024; Hastuti & Budiman, 2024; Putri & Arianto, 2024; Sari et al., 2020). Sementara *CatBoost* unggul dalam mengelola fitur kategorikal dan mencegah *overfitting* melalui *ordered boosting* (Irfannandhy et al., 2024). Berbagai metode *machine learning* telah diterapkan untuk menangani kompleksitas data iklim, khususnya dalam prediksi curah hujan.

Istianto et al. (2024) menggunakan algoritma *CatBoost* untuk memprediksi curah hujan dengan dataset "*Rain in Australia*," yang mencakup 23 atribut dan lebih dari 145.000 data observasi. Penelitian ini menonjolkan kemampuan *CatBoost* dalam menangani fitur kategorikal, mencegah *overfitting*, dan memproses data besar. Sehingga menghasilkan akurasi hingga 94,22% setelah melalui tahapan pembersihan data, penyeimbangan kelas, seleksi fitur, dan encoding. Meskipun *CatBoost* terbukti sangat efektif, penelitian ini tidak menerapkan algoritma tersebut dalam kerangka *ensemble*. Selain itu, penelitian ini berfokus pada klasifikasi curah hujan secara biner tanpa mengelompokkan intensitas curah hujan ke dalam kategori spesifik seperti tanpa hujan, hujan ringan, hujan sedang, hujan lebat, dan hujan ekstrem.

Penelitian lain, seperti Usman & Sudibyo (2022) menggunakan metode *gradient boosting*, *random forest*, dan *regresi logistik* dengan sembilan variabel yang memengaruhi curah hujan. Meskipun metode ini mampu memberikan hasil akurasi yang baik dalam pengklasifikasian curah hujan secara biner, namun penelitian tersebut tidak memanfaatkan penggabungan model dalam kerangka ensemble untuk meningkatkan generalisasi dan stabilitas prediksi, terutama pada data meteorologi yang memiliki kompleksitas tinggi. Selain itu, minimnya penelitian yang mengelompokkan intensitas curah hujan ke dalam beberapa kategori spesifik, yang seharusnya lebih relevan untuk aplikasi praktis seperti mitigasi bencana.

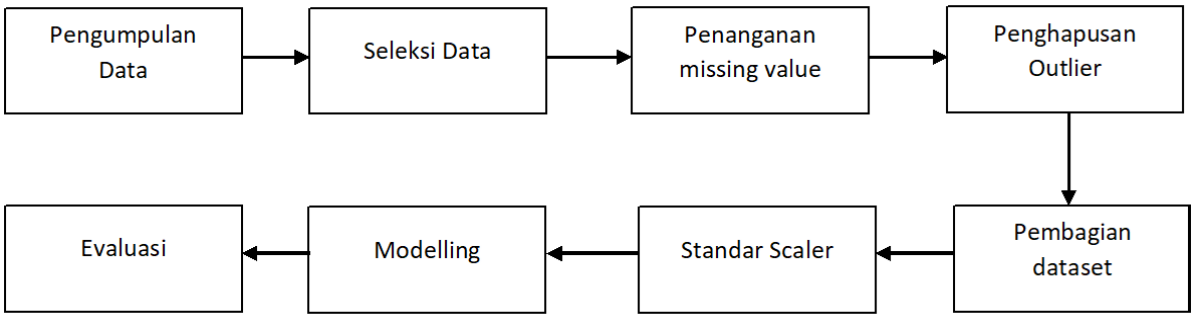
Metode *ensemble* seperti *voting classifier* memberikan keunggulan dalam menggabungkan prediksi dari beberapa model untuk meningkatkan stabilitas dan akurasi (Bagas et al., 2023; Saputra et al., 2024; Sari et al., 2024). *Voting classifier* bekerja dengan menggabungkan prediksi beberapa algoritma (Husaini et al., 2023; Kusyanti, 2019), sehingga hasil akhirnya ditentukan berdasarkan mayoritas suara atau rata-rata skor probabilitas. Dengan sifat data meteorologi yang kompleks, pendekatan ini diharapkan dapat memberikan prediksi yang lebih andal dibandingkan penggunaan model individu. Meskipun demikian, tinjauan

literatur menunjukkan bahwa penggunaan kombinasi *gradient boosting* dan *CatBoost* dalam pendekatan *voting classifier* belum banyak dieksplorasi, terutama dalam konteks pengelompokan intensitas curah hujan ke dalam kategori seperti tanpa hujan, hujan ringan, hujan sedang, hujan lebat, dan hujan ekstrem.

Penelitian ini bertujuan untuk mengisi celah tersebut dengan menggabungkan *gradient boosting* dan *CatBoost* dalam kerangka *voting classifier* untuk meningkatkan akurasi prediksi curah hujan. Selain itu, penelitian ini berfokus pada perbandingan kinerja model sebelum dan sesudah optimasi *hyperparameter* menggunakan metrik evaluasi seperti *accuracy*, *presisi*, *recall*, *f1-score*, dan *cohen's kappa*, sehingga memberikan analisis yang komprehensif mengenai efektivitas pendekatan yang diusulkan. Dengan demikian, penelitian ini tidak hanya memberikan kontribusi pada pengembangan model prediksi curah hujan, tetapi juga pada penerapan teknologi machine learning dalam mitigasi dampak perubahan iklim di Indonesia.

METODE

Penelitian ini merupakan studi eksperimental yang bertujuan untuk memberikan gambaran sistematis mengenai tahapan-tahapan yang dilakukan dalam proses klasifikasi curah hujan. Fokus penelitian utama dalam penelitian ini pada pengoptimalan model *ensemble* dengan *voting classifier* pada algoritma *gradient boosting* dan *CatBoost*. Alur pada penelitian ini berdasarkan gambar 1 dimulai dari pengumpulan data, seleksi data, penangannan missing value, sampai dilakukannya evaluasi pada tahap terakhir.



Gambar 1. Alur penelitian

Data penelitian ini berasal dari BMKG melalui *platform online* (https://dataonline.bmkg.go.id/) dengan total 1.461 data cuaca dari Stasiun Meteorologi Ahmad Yani, Semarang, untuk periode 2020–2023. Variabel yang digunakan mencakup suhu (Tn, Tx, Tavg), kelembapan relatif (RH_avg), curah hujan (RR), durasi sinar matahari (ss), serta kecepatan dan arah angin (ff_x, ff_avg, ddd_x). Curah hujan diukur dalam milimeter (mm), kecepatan angin dalam km/jam, dan arah angin dalam derajat. Intensitas harian curah hujan dikategorikan sebagai tanpa hujan (0 mm/hari), hujan ringan (0,5–20 mm/hari), hujan sedang (20–50 mm/hari), hujan lebat (50–150 mm/hari), dan hujan ekstrem (lebih dari 150 mm/hari). *Pre-processing* data melibatkan penanganan nilai yang hilang dengan metode imputasi rata-rata, di mana nilai yang tidak tersedia digantikan dengan rata-rata dari nilai-nilai yang ada pada fitur tersebut. Outlier dihapus menggunakan metode Rentang Interkuartil (IQR) untuk mengurangi kemungkinan bias akibat data yang tidak normal (Hazizah & Widiyaningtyas, 2024).

Tahapan selanjutnya adalah proses penyeimbangan data melalui metode *random oversampling*, bertujuan untuk menghindari bias model terhadap kelas mayoritas dan meningkatkan kemampuan model dalam mengenali pola pada kelas minoritas. Selanjutnya, pembagian data menggunakan teknik *stratified sampling* untuk memastikan distribusi kelas seimbang, teknik ini penting agar model tidak bias terhadap kelas mayoritas. Berikutnya

normalisasi fitur, dengan normalisasi model dapat memproses data dengan lebih efisien tanpa terpengaruh oleh perbedaan besar antara satu fitur dengan fitur lainnya.

Tahapan utama penelitian ini adalah pengembangan model menggunakan algoritma *gradient boosting* dan *CatBoost*. Model *baseline* untuk kedua algoritma dilatih tanpa optimasi hyperparameter. Untuk meningkatkan kinerja, dilakukan optimasi parameter seperti yang ditunjukkan pada tabel 1. Pada *gradient boosting*, parameter yang diuji mencakup n_estimators, learning_rate, max_depth, min_samples_split, *subsample*, dan min_samples_leaf. Sementara itu, parameter optimasi pada *CatBoost* meliputi iterations, learning_rate, depth, l2_leaf_reg, bagging_temperature, dan border_count, yang disesuaikan untuk memperoleh konfigurasi optimal dan meningkatkan performa model.

Tabel 1. Optimasi parameter

Model	Parameter	Value	Model	Parameter	Value
Gradient Boosting	n_estimators	[100, 200, 300]	*CatBoost*	iterations	[1000]
	learning_rate	[0.05, 0.1, 0.2]		learning_rate	[0.1]
	max_depth	[3, 5, 7]		depth	[7]
	min_samples_split	[2, 5, 10]		l2_leaf_reg	[7]
	subsample	[0.8, 1.0]		bagging_temperature	[2]
	Min_sample_leaf	[1,2]		Border_count	[128]

Gradient boosting dan *CatBoost* digabungkan menggunakan teknik *soft voting ensemble* untuk meningkatkan akurasi dan mengurangi variasi prediksi. Evaluasi model dilakukan dengan beberapa metrik: *accuracy* untuk mengukur persentase prediksi benar, *precision* untuk mengevaluasi keakuratan prediksi positif, *recall* untuk mengukur kemampuan model dalam mendeteksi semua kasus positif, *f1-score* untuk mengevaluasi keseimbangan antara precision dan recall, serta *cohen's kappa* untuk menilai kesepakatan prediksi model dengan kenyataan, terutama pada dataset kompleks atau tidak seimbang.

HASIL DAN PEMBAHASAN

Hasil

Proses penelitian dimulai dengan tahap *preprocessing* dengan penghapusan kolom "Date". Kolom ini dianggap kurang relevan karena data tanggal tidak secara langsung berkontribusi pada pola cuaca yang dapat digunakan untuk prediksi curah hujan. Dataset penelitian ini mengandung nilai yang hilang pada beberapa fitur numerik. Untuk mengatasi hal ini, nilai yang hilang diimputasi menggunakan rata-rata (*mean*) dari masing-masing kolom, guna menjaga konsistensi dan menghindari bias signifikan pada distribusi data.

Tabel 2. Distribusi data *outlier*

Keterangan	Jumlah Data
Jumlah baris sebelum menghapus *outlier*	1461
Jumlah baris setelah menghapus *outlier*	1363

Proses pembersihan data dilakukan dengan menghilangkan outlier untuk meningkatkan kualitas data yang akan digunakan dalam pemodelan. Berdasarkan tabel 2, jumlah baris data

awal sebelum penghapusan *outlier* adalah 1.461. Setelah proses identifikasi dan penghapusan *outlier*, jumlah baris data yang tersisa menjadi 1.363. Penghapusan ini bertujuan untuk mengurangi pengaruh data ekstrem yang dapat mengganggu kinerja model.

Tabel 3 menunjukkan distribusi kelas curah hujan sebelum dan sesudah proses penyeimbangan data. Untuk mengatasi data yang tidak seimbang menggunakan teknik *random oversampling*, yang bertujuan untuk meningkatkan jumlah data pada kelas minoritas hingga setara dengan kelas lainnya. Setelah proses penyeimbangan data, dataset dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian menggunakan metode *stratified sampling*, yang memastikan distribusi kelas tetap konsisten di kedua subset. Selanjutnya, fitur numerik dinormalisasi menggunakan *StandardScaler*. Proses standardisasi ini bertujuan untuk mengubah nilai fitur numerik menjadi skala standar dengan rata-rata 0 dan deviasi standar 1. Langkah ini penting untuk memastikan bahwa algoritma *machine learning* yang sensitif terhadap skala data, seperti *gradient boosting* dan *CatBoost*, dapat bekerja secara optimal tanpa bias yang disebabkan oleh perbedaan skala antar fitur.

Tabel 3. Distribusi kelas curah hujan sebelum dan sesudah *oversampling*

Kategori Curah Hujan	Sebelum *Oversampling*	Setelah *Oversampling*
Tanpa Hujan	416	454
Hujan Ringan	454	454
Hujan Sedang	107	454
Hujan Lebat	22	454
Hujan Ekstrem	364	454

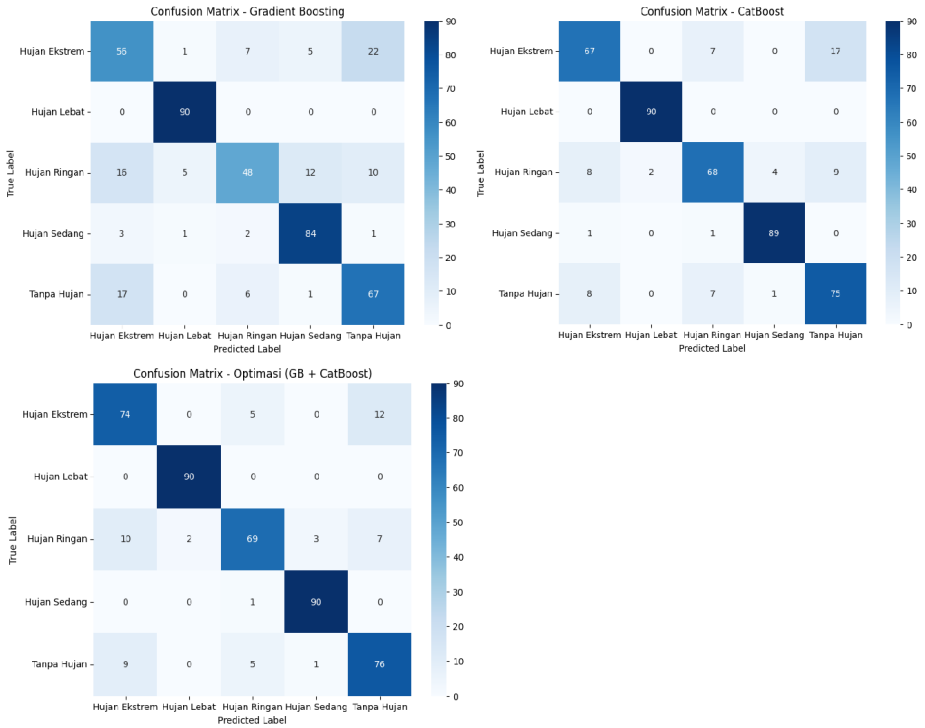
Pengembangan model pada penelitian ini dilakukan dengan menggunakan algoritma *gradient boosting* dan *CatBoost*, dengan fokus pada optimasi *hyperparameter* untuk meningkatkan performa model. Pada *gradient boosting*, optimasi *hyperparameter* dilakukan menggunakan *GridSearchCV*, sebuah teknik pencarian grid yang menguji kombinasi nilai *hyperparameter* secara menyeluruh berdasarkan metrik evaluasi tertentu. Proses optimasi *hyperparameter* pada *gradient boosting* mencakup beberapa parameter terbaik dengan rentang nilai sebagai berikut: n_estimators [200] untuk menentukan jumlah pohon keputusan, learning_rate [0.2] ntuk mengontrol kontribusi tiap pohon terhadap pembaruan model, Max_depth [7] untuk menangkap kompleksitas data tanpa *overfitting*, Min_samples_split [5] mengatur jumlah minimum sampel untuk membagi simpul, subsample [1.0], Min_sample_leaf [2] memastikan setiap daun pohon dapat memprediksi meskipun dengan dataset kecil.

Tabel 4. *Classification report*

Model	*Accuracy*	*Precision*	*Recall*	*F1-Score*	*Cohen's Kappa*
Gradient Boosting	75,99%	0,76	0,76	0,75	0,7599
CatBoost	85,68%	0,86	0,86	0,86	0,8210
Optimasi (*Gradient Boosting* + *CatBoost*)	87,89%	0,88	0,88	0,88	0,8486

Tabel 4 merupakan *classification report* dari hasil evaluasi perbandingan antara tiga model dalam prediksi curah hujan. Pada metrik *accuracy*, *gradient boosting* menunjukkan performa awal dengan *accuracy* model 75,99%, sementara *CatBoost* memberikan peningkatan hingga 85,68%. Setelah optimasi dan menggunakan pendekatan *ensemble* antara kedua model, *accuracy* model meningkat menjadi 87,89%. Selain itu, hasil terbaik pada metrik *precision*, *recall*, *f1-score*, dan *cohen's kappa* dicapai oleh model optimasi dan menggunakan pendekatan *ensemble* kedua model. Secara keseluruhan, optimasi *hyperparameter* dan metode *ensemble*

gradient boosting dengan *CatBoost* memberikan hasil yang lebih baik di semua metrik, menunjukkan bahwa kombinasi kedua model ini dengan optimasi mampu meningkatkan akurasi dan konsistensi prediksi.



Gambar 2. *Confusion matrix*

Gambar 2 merupakan *confusion matrix* dari performa model *Gradient Boosting* (GB), *CatBoost*, dan model optimasi (GB + *CatBoost*) dapat dibandingkan secara mendetail. Model GB memiliki beberapa kelemahan, terutama pada kategori "Hujan Ekstrem" dan "Tanpa Hujan". Untuk kategori "Hujan Ekstrem," model ini memiliki 56 prediksi *True Positives* (TP) namun menghasilkan 31 *False Positives* (FP) dan 17 *False Negatives* (FN), yang mengindikasikan banyaknya kesalahan dalam mendeteksi kelas ini. Demikian pula, untuk kategori "Tanpa Hujan," model menghasilkan 67 TP tetapi dengan 28 FP dan 24 FN, menunjukkan kesalahan signifikan. Model *CatBoos*t memperlihatkan peningkatan performa dengan jumlah FP dan FN yang lebih rendah dibandingkan GB.

Pada kategori "Hujan Ekstrem," *CatBoost* menghasilkan 67 TP dengan hanya 24 FP dan 15 FN, sementara untuk kategori "Hujan Ringan," model ini mencatat 68 TP dengan 15 FP dan 16 FN, menunjukkan pengurangan kesalahan yang cukup signifikan. Selain itu, kategori "Hujan Lebat" diprediksi dengan sangat baik, menghasilkan 90 TP tanpa FN pada semua model. Model optimasi (GB + *CatBoost*) memberikan hasil terbaik dengan pengurangan FP dan FN di hampir semua kategori. Pada kategori "Hujan Ekstrem," model ini menghasilkan 74 TP dengan 17 FP dan hanya 10 FN, sementara pada kategori "Hujan Ringan," model mencatat 69 TP dengan 12 FP dan 12 FN. Kategori "Tanpa Hujan" juga menunjukkan peningkatan, dengan 76 TP, 14 FP, dan hanya 8 FN. Dengan hasil ini, model optimasi (GB + *CatBoost*) memberikan performa yang paling konsisten dan akurat dibandingkan kedua model lainnya.

Pembahasan

Pada penelitian ini, kombinasi algoritma GB dan *CatBoost* dengan pendekatan *Voting Classifier* menunjukkan kinerja yang lebih unggul dibandingkan penerapan masing-masing model secara individu. Pemilihan algoritma GB dan *CatBoost* didasarkan pada keunggulan

keduanya dalam menangani data yang kompleks. GB efektif dalam mengurangi dampak *outlier* dan mampu menangani hubungan *non-linear* antara variabel input dan output. Sementara itu, *CatBoost* unggul dalam penanganan data kategorikal karena otomatis mengubah data menjadi format numerik, serta lebih stabil dalam menghindari *overfitting*, yang membuat *CatBoost* cocok untuk prediksi curah hujan.

Metode *voting classifier* menggabungkan kekuatan masing-masing model, sehingga dapat memanfaatkan keunggulan GB dalam menangani kesalahan prediksi sebelumnya dan kemampuan *CatBoost* dalam memberikan hasil yang lebih stabil. Setelah optimasi *hyperparameter* menggunakan teknik *GridSearchCV*, model *ensemble* ini mampu mencapai akurasi 87,89%, menunjukkan peningkatan dibandingkan dengan akurasi model baseline GB 75,99% dan *CatBoost* 85,68%. Peningkatan akurasi ini menunjukkan bahwa teknik *ensemble* dapat mengoptimalkan performa model dan memberikan prediksi curah hujan yang lebih akurat. Berdasarkan model terbaik yaitu kombinasi algoritma GB dan *CatBoost* dengan pendekatan *Voting Classifier*, hasil prediksi sesuai dengan data aktual mencakup berbagai kategori curah hujan aktual dan prediksi model, mencakup berbagai kategori seperti : tanpa hujan, hujan ringan, hujan lebat, dan hujan ekstrem. Dengan fitur input, Tn: 24.8, Tx: 32.3, Tavg: 28.1, RH_avg: 78.0, ss: 7.6, ff_x: 5.0, ddd_x : 350.0, dan ff_avg: 2.0, model memprediksi kategori curah hujan sebagai hujan ekstrem dengan prediksi model sesuai dengan data actual. Kolom actual menunjukkan data curah hujan actual dengan kategori seperti hujan ekstrem, hujan lebat, tanpa hujan, dan hujan ringan, sedangkan kolom predicted menunjukkan hasil prediksi dari model. Jika prediksi model sesuai dengan data aktual, maka mempresentasikan performa model yang sangat baik pada data uji.

Penelitian sebelumnya umumnya hanya menggunakan algoritma GB dan *CatBoost* secara terpisah tanpa menggabungkan keduanya dalam kerangka *ensemble* untuk meningkatkan akurasi model. Misalnya, beberapa studi seperti yang dilakukan oleh Usman & Sudibyo (2022) menggunakan *regresi logistik*, *random forest*, dan GB untuk memprediksi curah hujan berdasarkan sembilan variabel cuaca. Namun masih memiliki keterbatasan pada evaluasi model individu tanpa eksplorasi penggabungan algoritma untuk meningkatkan kinerja. Penelitian lain, seperti yang dilakukan oleh Istianto et al. (2024), hanya menggunakan algoritma GB dan *CatBoost* secara individu tanpa menggabungkan algoritma tersebut dalam kerangka *ensemble* untuk lebih meningkatkan akurasi model. Dengan pendekatan klasifikasi, penelitian ini memberikan kontribusi dalam menggabungkan kedua algoritma melalui teknik *soft voting*, yang terbukti meningkatkan stabilitas dan akurasi model. Selain itu, optimasi *hyperparameter* menggunakan *GridSearchCV* memastikan bahwa model beroperasi pada konfigurasi optimal, yang berkontribusi pada peningkatan akurasi.

Penelitian kami menunjukkan hasil baik dalam peningkatan akurasi prediksi curah hujan, melalui optimasi *hyperparameter* dan kombinasi algoritma GB dengan *CatBoost* dalam kerangka *voting classifier*. Namun, ada beberapa batasan, seperti penggunaan data hanya dari satu stasiun meteorologi. Penelitian mendatang disarankan menguji lokasi lain untuk mengukur generalisasi model, menggunakan kombinasi *ensemble* lebih kompleks seperti *stacking*, atau integrasi dengan *deep learning* seperti LSTM untuk menangani data temporal. *Feature selection* juga penting untuk memfokuskan model pada fitur relevan. Model ini berpotensi mendukung sistem peringatan dini, pengelolaan air, dan mitigasi bencana melalui prediksi cuaca yang lebih akurat.

SIMPULAN

Penelitian ini menunjukkan bahwa kombinasi GB dan *CatBoost* dengan pendekatan *voting classifier* berhasil meningkatkan akurasi prediksi curah hujan, dengan hasil prediksi yang lebih. *Accuracy* prediksi curah hujan dari baseline 75,99% (GB) dan 85,68% (*CatBoost*)

menjadi 87,89% setelah optimasi *hyperparameter* dan penerapan *ensemble*. Teknik *soft voting* dan optimasi *hyperparameter* berkontribusi dalam meningkatkan stabilitas dan *accuracy* klasifikasi curah hujan, hal ini bermanfaat untuk aplikasi seperti sistem peringatan dini cuaca dan mitigasi bencana. Meskipun penelitian ini terbatas pada data dari satu stasiun meteorologi, penelitian ini membuka peluang pengembangan lebih lanjut, seperti penggunaan metode *ensemble* lebih kompleks seperti *stacking* atau integrasi dengan model *deep learning* (LSTM) untuk menangani ketergantungan temporal.

REFERENSI

Afifah, D., Chusni, A., Nahar, A. N., Sirojuddin, M. A., Fatmawati, N., Islam, I. A., & Kudus, N. (2024). Persepsi Masyarakat Nelayan Dalam Menghadapi Perubahan Iklim Studi Desa Ujung Batu Kawasan Pesisir Utara Pulau Jawa (Ditinjau Aspek Sosial Ekonomi). *UTILITY: Jurnal Ilmiah Pendidikan Dan Ekonomi*, *8*(1), 42–58. https://doi.org/10.30599/utility.v8i1.3107

Ananda, I. K., Fanani, A. Z., Setiawan, D., & Wicaksono, D. F. (2024). Penerapan Random Oversampling dan Algoritma Boosting untuk Memprediksi Kualitas Buah Jeruk. *Edumatic: Jurnal Pendidikan Informatika*, *8*(1), 282-289. https://doi.org/10.29408/edumatic.v8i1.25836

Azhari, D. M., & Hidajat, M. S. (2024). Klasifikasi Stunting pada Balita menggunakan Algortima Gradient Bossting Clasifier. *Edumatic: Jurnal Pendidikan Informatika*, *8*(2), 507-515. https://doi.org/10.29408/edumatic.v8i2.27502

Azhari, R., Amanah, S., Fatchiya, A., & Kinseng, R. A. (2023). The Role of Agricultural Extension, Communication, and Farmer Organizations in Building Resilience of Smallholder Farmers. *Forum Penelitian Agro Ekonomi*, *41*(1), 45–63.

Bagas, M., Darmawan, A., Dewanta, F., & Astuti, S. (2023). Analisis Perbandingan Algoritma Decision Tree, Random Forest, dan Naïve Bayes untuk Prediksi Banjir di Desa Dayeuhkolot Comparative Analysis of Decision Tree, Random Forest, and Naïve Bayes Algorithm for Flood Prediction at Dayeuhkolot Village. *TELKA*, *9*(1), 52–61. https://doi.org/10.15575/telka.v9n1.52-61

Hastuti, N. T., & Budiman, F. (2024). Optimasi Klasifikasi Stunting Balita dengan Teknik Boosting pada Decision Tree. *Edumatic: Jurnal Pendidikan Informatika*, *8*(2), 655-664. https://doi.org/10.29408/edumatic.v8i2.27913

Hayu, C. S., Aprilia, C., Kamila Putri, U., Leana Putri, V., Alfi Hidayat, A., Ansori, N., & Negeri Semarang, U. (2024). Analisis Pola Debit Hujan terhadap Terjadinya Banjir di Daerah Aliran Kali Es Sawah Besar Pada 12 Februari 2024. *Jurnal Implementasi*, *4*(1), 65–78.

Hazizah, C. Y., & Widiyaningtyas, T. (2024). Analisis Metode Collaborative Filtering menggunakan KNN dan SVD++ untuk Rekomendasi Produk E-commerce Tokopedia. *Edumatic: Jurnal Pendidikan Informatika*, *8*(2), 595–604. https://doi.org/10.29408/edumatic.v8i2.27793

Husaini, A., Hoeronis, I., Lumana, H. H., & Puspareni, L. D. (2023). Early Detection of Stunting in Toddlers Based on Ensemble Machine Learning in Purbaratu Tasikmalaya. *Jurnal Sistem Dan Teknologi Informasi (JustIN)*, *11*(3), 487–495. https://doi.org/10.26418/justin.v11i3.66465

Irfannandhy, R., Handoko, L. B., & Ariyanto, N. (2024). Analisis Performa Model Random Forest dan CatBoost dengan Teknik SMOTE dalam Prediksi Risiko Diabetes. *Edumatic: Jurnal Pendidikan Informatika*, *8*(2), 714–723. https://doi.org/10.29408/edumatic.v8i2.27990

Istianto, A. F., Hadiana, A. I., & Umbara, F. R. (2024). Prediksi Curah Hujan Menggunakan Metode Categorical Boosting (CATBOOST). *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(4), 2930–2937. https://doi.org/10.36040/jati.v7i4.7304

Jasman, T. Z., Fadhlullah, M. A., Pratama, A. L., & Rismayani, R. (2022). Analisis Algoritma Gradient Boosting, Adaboost dan Catboost dalam Klasifikasi Kualitas Air. *Jurnal Teknik Informatika Dan Sistem Informasi*, *8*(2), 392–402. https://doi.org/10.28932/jutisi.v8i2.4906

Kusuma, A. C., Pratiwi, N. W. W., Humairah, N. A., & Yulistio, M. R. (2022). Analisis Dampak Kebijakan Populis Terhadap Keputusan Gubernur DKI Jakarta. *Jurnal Analisis Hukum*, *5*(1), 90–105. https://doi.org/10.38043/jah.v5i1.3491

Kusyanti, A. (2019). Metode Ensemble Classifier untuk Mendeteksi Jenis Attention Deficit Hyperactivity Disorder (SDHD) pada Anak Usia Dini. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIIK)*, *6*(3), 301–308. https://doi.org/10.25126/jtiik.201961313

Mabruroh, F., & Wiyanto, A. (2023). Analisis Fenomena Perubahan Iklim Terhadap Curah Hujan Ekstrim. *OPTIKA: Jurnal Pendidikan Fisika*, *7*(1), 94–100. https://doi.org/10.37478/optika.v7i1.2738

Pahlevi, O., Ayu, D., Wulandari, N., Rahayu, L. K., Leidiyana, H., & Handrianto, Y. (2024). Model Klasifikasi Risiko Stunting Pada Balita Menggunakan Algoritma CatBoost Classifier. *Bulletin of Computer Science Research*, *6*(4), 414–421.

Purify, A., Teknik Elektronika Pertahanan, P., Militer, A., Kusman, A., Widodo, S., & Silitonga, F. (2024). Perubahan Iklim Dan Risiko Keamanan Nasional: Kajian Mengenai Kesiapsiagaan Pertahanan Indonesia. *Jurnal Elektrosista*, *12*(1), 1–11.

Putri, F., & Arianto, D. B. (2024). Perbandingan Performa Random Forest Dan Gradient Boosting Dalam Prediksi Pada Dataset Customer Shopping Trends. *Kohesi: Jurnal Multidisiplin Saintek*, *5*(10), 1–9.

Runtulalo, Y. S., & Manongga, D. H. F. (2024). Clustering Tingkat Kemiripan Curah Hujan di Indonesia Berdasarkan Provinsi Menggunakan Metode Hierarchical Clustering dan GeoMap. *Progresif: Jurnal Ilmiah Komputer*, *20*(1), 325–336. https://doi.org/10.35889/progresif.v20i1.1583

Saputra, D. R. K., Via, Y. V., & Sihananto, A. N. (2024). Deteksi Anomali Menggunakan Ensemble Learning Dan Random Oversampling Pada Penipuan Transaksi Keuangan. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(3), 2779–2788. https://doi.org/10.23960/jitet.v12i3.4910

Sari, D. P., Halim, Z., & Waseso, B. (2024). Implementasi Machine Learning untuk Deteksi Intrusi pada Jaringan Komputer. *Jurnal Minfo Polgan*, *13*(2), 1389–1394. https://doi.org/10.33395/jmp.v13i2.14074

Sari, V. R., Firdausi, F., & Azhar, Y. (2020). Perbandingan Prediksi Kualitas Kopi Arabika dengan Menggunakan Algoritma SGD, Naive Bayes, dan Random Forest. *Edumatic: Jurnal Pendidikan Informatika*, *4*(2), 1-9. https://doi.org/10.29408/edumatic.v4i2.2202

Suwarman, R., Riawan, E., Simanjuntak, Y. S. M., & Irawan, D. E. (2022). Kajian Perubahan Iklim di Pesisir Jakarta Berdasarkan Data Curah Hujan dan Temperatur. *Buletin Oseanografi Marina*, *11*(1), 99–110. https://doi.org/10.14710/buloma.v11i1.42749

Usman, C. D., & Sudibyo, U. (2022). Klasifikasi Curah Hujan di Kota Semarang Menggunakan Machine Learning. *Pros. Sains Dan Teknol*, *1*(1), 1–5.