

TESIS

**PENGEMBANGAN MODEL PREDIKSI CUACA DAN POTENSI
BANJIR DI KABUPATEN SLEMAN MENGGUNAKAN METODE
CATBOOST DAN LONG SHORT-TERM MEMORY (LSTM)**



Disusun oleh:

Nama : Fall Amir
NIM : 22.51.1184
Konsentrasi : Business Intelligence

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

TESIS

**PENGEMBANGAN MODEL PREDIKSI CUACA DAN POTENSI
BANJIR DI KABUPATEN SLEMAN MENGGUNAKAN METODE
CATBOOST DAN LONG SHORT-TERM MEMORY (LSTM)**

**DEVELOPMENT OF WEATHER PREDICTION MODEL AND
FLOOD POTENTIAL IN SLEMAN REGENCY USING CATBOOST
METHOD AND LONG SHORT-TERM MEMORY (LSTM)**

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Pascasarjana
Program Studi Informatika



Disusun oleh:

Nama : Fall Amir
NIM : 22.51.1184
Konsentrasi : Business Intelligence

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

**PENGEMBANGAN MODEL PREDIKSI CUACA DAN POTENSI BANJIR DI
KABUPATEN SLEMAN MENGGUNAKAN METODE CATBOOST DAN
LONG SHORT-TERM MEMORY (LSTM)**

**DEVELOPMENT OF WEATHER PREDICTION MODELS AND FLOOD
POTENTIAL IN SLEMAN DISTRICT USING CATBOOST AND LONG
SHORT-TERM MEMORY (LSTM) METHODS**

yang disusun dan diajukan oleh

Fail Amir

22.51.1184

telah disetujui oleh Dosen Pembimbing Tesis
pada tanggal 03 Januari 2025

Dosen Pembimbing,



Prof. Dr. Irena Utami, S.Si., M.Kom.

NIK. 190302037

HALAMAN PENGESAHAN

**PENGEMBANGAN MODEL PREDIKSI CUACA DAN POTENSI BANJIR DI
KABUPATEN SLEMAN MENGGUNAKAN METODE CATBOOST DAN
LONG SHORT-TERM MEMORY (LSTM)**

**DEVELOPMENT OF WEATHER PREDICTION MODELS AND FLOOD
POTENTIAL IN SLEMAN DISTRICT USING CATBOOST AND LONG
SHORT-TERM MEMORY (LSTM) METHODS**

yang disusun dan diajukan oleh

Fall Amir

22.51.1184

Telah dipertahankan di depan Dewan Penguji
pada tanggal 03 Januari 2025

Susunan Dewan Penguji

Pembimbing Utama

Prof. Dr. Ema Utami, S.Si., M.Kom.

NIK. 190302037

Pembimbing Pendamping

M. Hanafi, S.Kom., M.Eng., Ph.D.

NIK. 190302024

Tanda Tangan



Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer
Yogyakarta, 03 Januari 2025

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.

NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang berandatangan di bawah ini,

Nama mahasiswa : **Faii Amir**
NIM : **22.51.1184**
Konsentrasi : **Business Intelligence**

Menyatakan bahwa Tesis dengan judul berikut:
**PENGEMBANGAN MODEL PREDIKSI CUACA DAN POTENSI BANJIR DI
KABUPATEN SLEMAN MENGGUNAKAN METODE CATBOOST DAN LONG
SHORT-TERM MEMORY (LSTM)**

Dosen Pembimbing Utama : **Prof. Dr. Insa Utami, S.Si., M.Kom.**

Dosen Pembimbing Pendamping : **Hurafi, S.Kom., M.Eng., Ph.D.**

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMKOM Yogyakarta maupun di Perguruan Tinggi lainnya
2. Karya tulis ini merupakan gagasan, rancangan dan penelitian **SAYA** sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas AMKOM Yogyakarta
5. Pernyataan ini **SAYA** buat dengan sesungguhnya, apabila di kemudian hari terdapat penyempangan dan ketidakbenaran dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi

Yogyakarta, 03 Januari 2025

Yang Menyatakan,

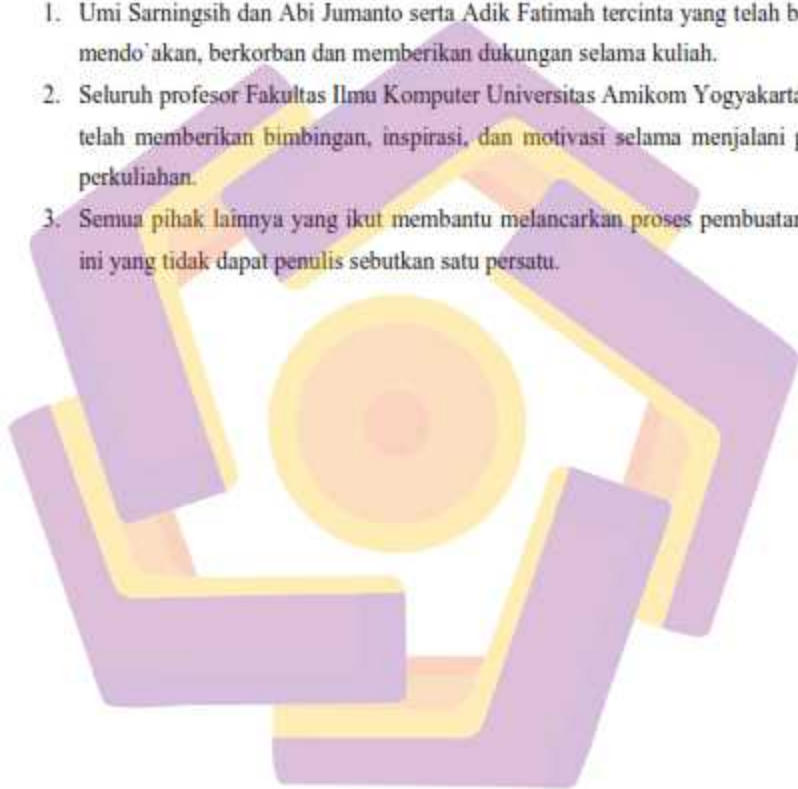


Faii Amir

HALAMAN PERSEMBAHAN

Puji syukur kepada Allah SWT atas karunia ilmu yang diberikan kepada penulis, serta rahmat dan hidayah-Nya yang selalu penulis harap dan doakan sehingga dimudahkan dalam penulisan skripsi ini. Skripsi ini penulis persembahkan kepada:

1. Umi Sarningsih dan Abi Jumanto serta Adik Fatimah tercinta yang telah banyak mendo'akan, berkorban dan memberikan dukungan selama kuliah.
2. Seluruh profesor Fakultas Ilmu Komputer Universitas Amikom Yogyakarta yang telah memberikan bimbingan, inspirasi, dan motivasi selama menjalani proses perkuliahan.
3. Semua pihak lainnya yang ikut membantu melancarkan proses pembuatan tesis ini yang tidak dapat penulis sebutkan satu persatu.



HALAMAN MOTTO

"... Katakanlah: "Samakah orang-orang yang berpengetahuan dengan orang-orang yang tidak berpengetahuan? Sesungguhnya yang mendapat pelajaran hanyalah orang-orang yang mempunyai pikiran."

(Q.S.39:9)

".... Dan katakanlah: "Ya Tuhanku, tambahkanlah ilmu kepadaku."

(Q.S.20:114)



KATA PENGANTAR

Puji syukur, Alhamdulillah atas kehadiran Allah SWT yang telah memberikan rahmat dan karuniaNya kepada kita semua sehingga penulis dapat menyelesaikan tesis yang diajukan sebagai salah satu syarat untuk menyelesaikan program pascasarjana di Program Studi Informatika Universitas Amikom Yogyakarta.

Adapun penyusunan tesis ini digunakan sebagai bukti bahwa penyusun telah melaksanakan dan menyelesaikan penelitian tesis. Dalam proses penyusunan laporan ini penyusun mendapatkan banyak bantuan dari berbagai pihak. Oleh karena itu penulis mengucapkan terima kasih kepada:

1. Prof. Dr. M. Suyanto, M.M. (Rektor Universitas Amikom Yogyakarta).
2. Prof. Dr. Kusriani, M.Kom. (Dekan Fakultas Ilmu Komputer).
3. Prof. Dr. Ema Utami, S.Si., M.Kom. dan Hanafi, S.Kom., M.Eng., Ph.D. selaku pembimbing tesis yang memberikan banyak saran dan masukannya dalam pembuatan tesis.

Yogyakarta, 1 Januari 2025
Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
HALAMAN PERNYATAAN KEASLIAN TESIS	v
HALAMAN PERSEMBAHAN	vi
HALAMAN MOTTO.....	vii
KATA PENGANTAR.....	viii
DAFTAR ISI.....	ix
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
INTISARI.....	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah.....	5
1.3. Batasan Masalah.....	6
1.4. Tujuan Penelitian.....	7
1.5. Manfaat Penelitian.....	7
1.6. Hipotesis.....	8
BAB II TINJAUAN PUSTAKA	9
2.1. Tinjauan Pustaka.....	9
2.2. Keaslian Penelitian	12
2.3. Landasan Teori	15
A. Meteorologi dan Hidrometeorologi	15

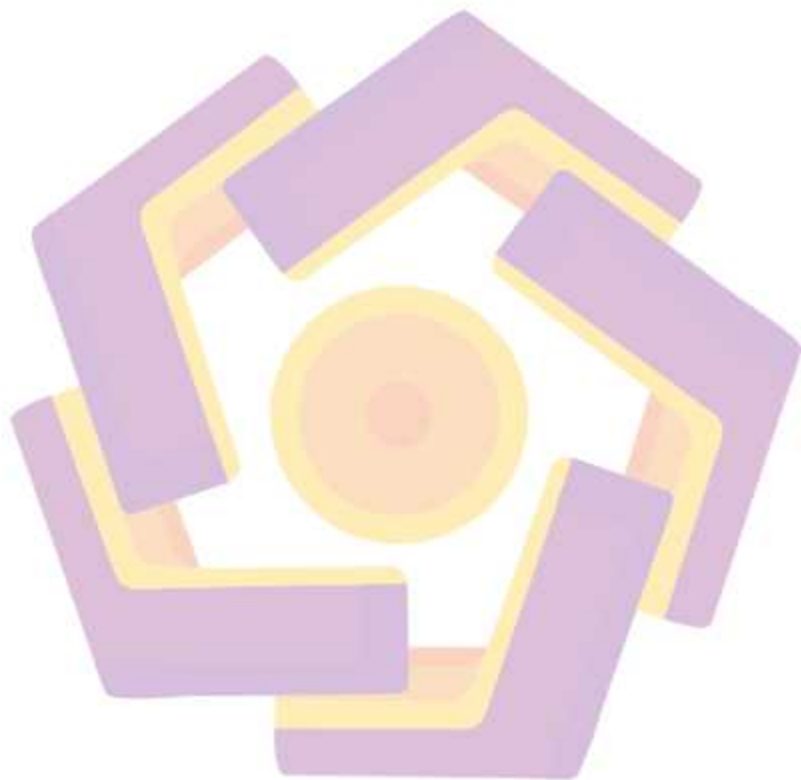
B. Teori Pembelajaran Mesin dan Pembelajaran Mendalam (<i>Deep Learning</i>).....	16
C. Recurrent Neural Network (RNN).....	17
D. Long Short-Term Memory (LSTM).....	19
1. Gerbang Lupa (Forget Gate).....	19
2. Gerbang Input (Input Gate).....	20
3. Memperbarui Sel Memori.....	21
4. Gerbang Output (Output Gate).....	22
E. CatBoost sebagai Model Pembelajaran Mesin Berbasis Pohon Keputusan.....	23
F. CatBoost dan LSTM untuk Prediksi Cuaca.....	26
BAB III METODE PENELITIAN	28
3.1. Jenis, Sifat, dan Pendekatan Penelitian.....	28
3.2. Sumber Data.....	29
3.3. Exploratory Data Analysis (EDA).....	37
3.4. Implementasi Model.....	42
3.5. Evaluasi Model.....	44
3.6. Alur Penelitian.....	45
BAB IV HASIL PENELITIAN DAN PEMBAHASAN	49
4.1. Evaluasi Model.....	49
4.2. Hasil Penelitian.....	55
4.3. Pembahasan.....	56
4.4. Implikasi Ilmiah.....	57
4.5. Verifikasi dan Validasi Teori.....	57

BAB V PENUTUP	58
5.1. Kesimpulan.....	58
5.2. Saran.....	58
1. Penambahan Variabel dan Data Cuaca Lainnya	58
2. Penerapan Metode Pemodelan Lain.....	59
3. Pengembangan Sistem Peringatan Dini	59
4. Pengujian dan Validasi dengan Data Real-Time	59
DAFTAR PUSTAKA.....	60



DAFTAR TABEL

Tabel 2.1. Matriks literatur review dan posisi penelitian.....	12
Tabel 3.1 Jadwal Penelitian	47



DAFTAR GAMBAR

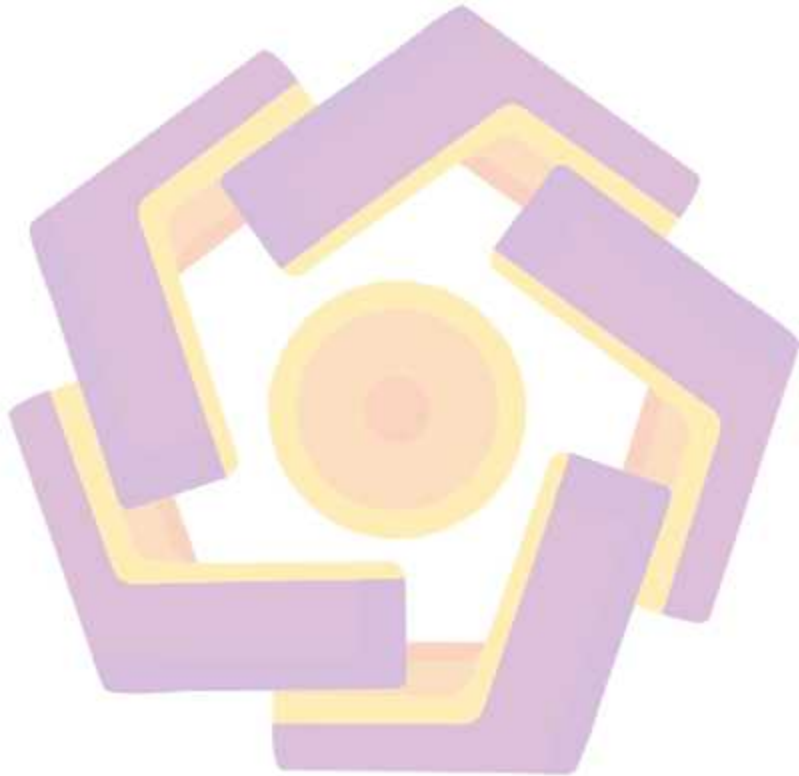
Gambar 1.1 Tren kejadian bencana alam di Indonesia dalam 10 tahun terakhir periode 2011 -2020, (BNPB, 2020).....	1
Gambar 2.1 Visualisasi Gerbang Lupa	20
Gambar 2.2 Visualisasi Gerbang Input.....	21
Gambar 2.3 Visualisasi Memperbarui Sel Memori	22
Gambar 2.4 Visualisasi Gerbang Output.....	23
Gambar 3.1 Data Cuaca dari BMKG.....	30
Gambar 3.2 Data Bencana dari BNPB.....	33
Gambar 3.3 Boxplot Data Tahun 2020-2024.....	33
Gambar 3.5 Persentase Missing Values pada dataset	35
Gambar 3.6 Persentase Missing Values setelah imputasi median	35
Gambar 3.7 Konversi nilai pada kolom status banjir.....	36
Gambar 3.8 Persentase Missing Values setelah imputasi median	37
Gambar 3.9 Distribution Plot pada kelembapan udara	38
Gambar 3.10 Distribution Plot pada variabel penting lainnya.....	39
Gambar 3.11 Persentase Missing Values setelah imputasi median	40
Gambar 3.12 Distribusi Kelas Sebelum dan Setelah Resampling SMOTE	41
Gambar 3.13 Correlation Matrix hubungan antara variabel dan status banjir.....	41
Gambar 3.14 Alur Penelitian	45
Gambar 4.1 Hasil Analysis Model Catboost Classifier	49
Gambar 4.2 Confusion Matrix Model Catboost Classifier	50
Gambar 4.3 Kurva ROC Model Catboost Classifier	51

Gambar 4.4 Hasil Analisis Model Catboost Classifier dengan GridSearch
(Hyperparameter Tuning)..... 52

Gambar 4.5. Confusion Matrix Model Catboost Classifier dengan GridSearch 52

Gambar 4.6 Kurva ROC Model Catboost Classifier dengan GridSearch..... 53

Gambar 4.7 Visualisasi Feature Importance..... 54



INTISARI

Perkembangan teknologi kecerdasan buatan telah membuka peluang besar dalam prediksi cuaca dan mitigasi risiko bencana banjir, terutama di daerah yang rentan terhadap perubahan iklim seperti Kota Sleman. Tesis ini memperkenalkan model prediksi cuaca dan potensi banjir dengan memanfaatkan algoritma CatBoost dan LSTM (Long Short-Term Memory), yang dikembangkan untuk memodelkan ketergantungan temporal dari data meteorologi dan kejadian bencana dengan lebih akurat. Arsitektur ini dirancang untuk memahami pola cuaca yang dapat memicu banjir, memberikan prediksi yang lebih handal bagi sistem peringatan dini.

Studi ini mengumpulkan dan mengintegrasikan data cuaca historis dan data kejadian banjir di Kota Sleman dari Januari 2020 hingga Mei 2024. Dengan menggunakan dataset tersebut, penelitian ini mengeksplorasi dan membandingkan performa model CatBoost dan LSTM melalui serangkaian eksperimen yang komprehensif, termasuk variasi parameter seperti kedalaman pohon keputusan, jumlah unit tersembunyi, dan ukuran batch. Evaluasi model dilakukan menggunakan metrik akurasi, precision, recall, serta error seperti Mean Absolute Error (MAE) dan Root Mean Square Error (RMSE) untuk menentukan konfigurasi optimal.

Hasil analisis menunjukkan bahwa model LSTM memberikan performa terbaik dalam memprediksi kondisi cuaca yang berpotensi menyebabkan banjir, dengan RMSE rata-rata sebesar 1.75 dan MAE rata-rata sebesar 1.30, menunjukkan kemampuan LSTM untuk menangkap pola temporal dalam data cuaca. Model CatBoost juga menunjukkan kinerja yang baik, khususnya dalam menangani fitur kategorikal dan mengidentifikasi variabel cuaca yang paling berpengaruh terhadap risiko banjir. Berdasarkan hasil evaluasi, konfigurasi optimal model dicapai dengan kedalaman pohon keputusan 8 dan 100 epoch pada LSTM, menghasilkan prediksi dengan tingkat kesalahan rendah.

Kata kunci: Prediksi Cuaca, LSTM, Pembelajaran Mesin, Jaringan Saraf Tiruan Mendalam.

ABSTRACT

The development of artificial intelligence technology has opened up great opportunities in weather prediction and flood disaster risk mitigation, especially in areas vulnerable to climate change such as Sleman City. This thesis introduces a weather and flood potential prediction model utilizing the CatBoost and LSTM (Long Short-Term Memory) algorithms, which were developed to model the temporal dependency of meteorological data and disaster events more accurately. This architecture is designed to understand extreme weather patterns that can trigger floods, providing more reliable predictions for early warning systems.

This study collects and integrates historical weather data and flood event data in Sleman City from January 2020 to May 2024. Using this dataset, this study explores and compares the performance of the CatBoost and LSTM models through a series of comprehensive experiments, including variations in parameters such as decision tree depth, number of hidden units, and batch size. Model evaluation is carried out using accuracy, precision, recall, and error metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) to determine the optimal configuration.

The analysis results show that the LSTM model provides the best performance in predicting weather conditions that have the potential to cause flooding, with an average RMSE of 1.75 and an average MAE of 1.30, indicating the ability of LSTM to capture temporal patterns in weather data. The CatBoost model also shows good performance, especially in handling categorical features and identifying weather variables that have the most influence on flood risk. Based on the evaluation results, the optimal model configuration was achieved with a decision tree depth of 8 and 100 epochs on the LSTM, resulting in predictions with a low error rate.

Keyword: Weather Prediction, Long Short-Term Memory, Machine Learning, Deep Neural Networks.

BAB I PENDAHULUAN

1.1. Latar Belakang Masalah

Bagian Indonesia, sebuah negara kepulauan tropis dengan iklim kompleks, menghadapi peningkatan frekuensi dan intensitas bencana alam yang signifikan selama 15 tahun terakhir (2003-2017), sebagaimana dicatat oleh Kongres Teknologi Nasional (KTN) oleh Badan Pengkajian Dan Penerapan Teknologi (BPPT, 2018). Lebih dari 90% bencana tersebut berhubungan dengan fenomena hidrometeorologi, termasuk banjir, yang diakibatkan oleh perubahan iklim. Bahkan dalam kurun waktu 10 tahun terakhir (2011-2020), data bencana alam di Indonesia menunjukkan peningkatan kejadian bencana yang cukup signifikan (BPPT, 2018).



Gambar 1.1 Tren kejadian bencana alam di Indonesia dalam 10 tahun terakhir periode 2011-2020, (BNPB, 2020).

Tren ini tidak hanya terjadi di Indonesia, tetapi juga terjadi secara global. Kerugian ekonomi dan nyawa yang ditimbulkan oleh bencana ini cukup besar. Di Indonesia, rata-rata kerugian material per tahun pada periode 2004 – 2013 mencapai 22,8 Triliun Rupiah (Kompas, 2018). Faktor penyebabnya antara lain adalah populasi yang tinggal di area yang rentan terhadap bencana, kurangnya sistem peringatan dini,

peningkatan kerentanan wilayah, dan kurangnya pemahaman masyarakat tentang gejala bencana.

Mengikuti kerangka kerja Sendai untuk Pengurangan Risiko Bencana (SFDRR) 2015-2030 dan Paris Agreement 2015, yang telah diratifikasi oleh Indonesia melalui UU No. 16 Tahun 2016, penelitian ini bertujuan untuk merumuskan model prediksi curah hujan berbasis CNN yang akan membantu peringatan dini berbasis risiko. Sistem seperti ini dapat mengurangi dampak negatif bencana alam, dan juga mendukung upaya mitigasi, adaptasi, perhitungan kerugian, dan peningkatan kapasitas yang disyaratkan oleh perjanjian tersebut.

Era Revolusi Industri 4.0 dengan perkembangan teknologi seperti big data, Internet of Things (IoT), kecerdasan artifisial (AI), dan otomatisasi, membuka peluang besar untuk melakukan inovasi dalam sistem peringatan dini. Dalam konteks ini, penerapan CNN sebagai bagian dari AI dapat menjadi cara yang efektif untuk memprediksi curah hujan dan mengurangi risiko bencana.

Menurut UU No. 24 Tahun 2007 tentang penanggulangan bencana, penanggulangan bencana dapat dibagi menjadi tiga tahap: prabencana, saat tanggap darurat, dan pascabencana. Sejalan dengan kerangka kerja SFDRR 2015-2030 dan UU ini, penelitian ini akan berfokus pada tahap prabencana, di mana upaya pengurangan risiko bencana sangat penting.

Sebagai tambahan kepada latar belakang yang telah disebutkan sebelumnya (Sutawidjaja, 2005; Marfai & King, 2008; Yulianto, et al., 2017; Purnama, et al., 2019), penelitian ini bertujuan untuk menggali lebih dalam tentang bagaimana metode CNN dapat digunakan untuk memperbaiki model prediksi cuaca yang ada saat ini, dengan harapan dapat mengurangi kerugian ekonomi dan nyawa akibat bencana banjir dan tanah longsor.

Penelitian terdahulu yang relevan dalam konteks pengembangan model prediksi curah hujan menggunakan Convolutional Neural Network (CNN) untuk mendukung sistem peringatan dini bencana alam menunjukkan beberapa temuan yang menarik. Misalnya, Yunjie et al. (2016) mengusulkan penggunaan CNN dengan memanfaatkan data citra satelit untuk memprediksi curah hujan. Hasil studi ini menunjukkan bahwa CNN mampu menghasilkan prediksi yang lebih akurat dibandingkan dengan metode tradisional. Penelitian lain oleh Ali Haidar et al. (2018) menggunakan model neural

network untuk prediksi curah hujan, meskipun hanya memanfaatkan data cuaca sederhana seperti suhu dan kelembaban. Namun, mereka menunjukkan potensi metode ini dalam meningkatkan akurasi prediksi.

Elhoseiny et al. (2015) menjelajahi penggunaan Convolutional Long Short-Term Memory (ConvLSTM) dalam memprediksi curah hujan dan menemukan bahwa metode ini mampu menangkap pola spasial yang kompleks dari data curah hujan. Penelitian lainnya oleh Wan Liu et al. (2022) menggabungkan model deep learning dengan data radar untuk meningkatkan akurasi prediksi curah hujan. Mereka menemukan bahwa pendekatan ini efektif dalam meningkatkan performa prediksi.

Penelitian yang dilakukan oleh Novadiwanti (2021) menyimpulkan bahwa penelitian ini berhasil melakukan pemodelan awal pada musim hujan dengan menggunakan tiga model data GCM. Mereka menggunakan metode cascade neural network (CNN) yang dioptimalkan dengan menggunakan genetic algorithm (GA) terhadap tiga stasiun cuaca di Kabupaten Pacitan. Hasil dari penelitian ini menunjukkan bahwa nilai AMH (Awal Musim Hujan) yang diperoleh dari setiap model GCM memiliki pola yang sejalan dengan nilai AMH aktual.

Penelitian yang dilakukan oleh Randy Sulisty (2021) menyimpulkan bahwa penggunaan Convolutional Neural Network (CNN) mampu mengklasifikasikan jenis cuaca berdasarkan kelasnya. Pengujian dilakukan dengan mengubah parameter seperti jenis optimizer, ukuran citra, learning rate, dan epoch. Hasil pengujian menunjukkan bahwa parameter-parameter tersebut berpengaruh signifikan terhadap tingkat keakuratan informasi, presisi, recall, dan F1-score sistem.

Penelitian yang dilakukan oleh Ritha (2016) menyimpulkan bahwa Algoritma Levenberg-Marquardt memberikan hasil terbaik dalam pengujian terhadap model data. Penelitian ini melakukan uji coba dengan menggunakan jumlah neuron hidden layer 10 dan Epoch 100, yang menghasilkan nilai Mean-Square Error (MSE) sebesar 0.776. Sementara itu, penggunaan Algoritma Backpropagation dengan jumlah neuron hidden layer 4 dan Epoch 1.000 menghasilkan nilai MSE sebesar 0.07876.

Terakhir, Parashar et al. (2021) memanfaatkan data satelit dan teknik pengolahan citra untuk memprediksi curah hujan secara spasial. Studi ini menunjukkan bahwa penggunaan data citra satelit dapat meningkatkan akurasi prediksi curah hujan. Dalam konteks ini, penelitian ini bertujuan untuk menggabungkan kontribusi-kontribusi tersebut

dan mengembangkan model prediksi curah hujan menggunakan CNN yang mampu mempelajari pola spasial yang kompleks dan mendukung sistem peringatan dini bencana alam.

Beberapa model Jaringan Syaraf Tiruan yang digunakan dalam penelitian ini antara lain jaringan kohonen, jaringan kompetitif, back propagation, adaline, perceptron, hebb, dan neuron. Meskipun jaringan kompetitif sulit diterapkan dalam bentuk kecerdasan buatan secara konvensional, penggunaan metode CNN-LSTM dalam penelitian ini dianggap sebagai solusi yang efektif dalam melakukan prediksi cuaca, khususnya prediksi hujan di wilayah Kota.

Berdasarkan latar belakang yang telah disampaikan, penelitian ini memandang pentingnya pengembangan sistem prediksi curah hujan yang lebih akurat untuk mendukung sistem peringatan dini bencana alam, khususnya banjir. Penelitian-penelitian sebelumnya telah mencoba untuk mengatasi masalah ini, namun masih terdapat beberapa celah yang belum terjawab dan memicu perlu dilakukannya penelitian lebih lanjut.

Gap-gaps tersebut antara lain:

Penelitian-penelitian sebelumnya umumnya menggunakan metode prediksi curah hujan berbasis statistika atau metode fisik. Meskipun metode-metode ini cukup efektif, mereka memiliki keterbatasan dalam menangani variasi spasial dan temporal yang kompleks dalam data curah hujan, serta memerlukan pemahaman yang mendalam tentang proses fisik yang terlibat (Kusuma et al., 2020). Oleh karena itu, diperlukan pendekatan yang lebih fleksibel dan mampu belajar dari data, seperti yang ditawarkan oleh metode berbasis pembelajaran mesin.

Walaupun beberapa penelitian telah menggunakan pembelajaran mesin untuk prediksi curah hujan, hasilnya belum selalu memuaskan. Misalnya, sebuah studi oleh Wibowo et al. (2019) menggunakan Jaringan Saraf Tiruan (JST) untuk prediksi curah hujan bulanan di Surabaya, namun akurasi masih di bawah 70%. Hal ini menunjukkan bahwa masih ada ruang untuk peningkatan dalam hal akurasi prediksi.

Meskipun kecerdasan buatan telah banyak diaplikasikan dalam berbagai bidang, penerapannya dalam prediksi curah hujan dan mitigasi risiko banjir di Indonesia masih terbatas. Khususnya, belum banyak penelitian yang menerapkan algoritma berbasis CatBoost dan LSTM untuk tujuan ini, meskipun kedua algoritma ini telah terbukti efektif dalam berbagai aplikasi yang melibatkan analisis temporal dan kategorikal (Dorogush et

al., 2018; Hochreiter & Schmidhuber, 1997). CatBoost unggul dalam menangani data kategori dan memberikan analisis fitur penting, sementara LSTM memiliki kemampuan khusus untuk menangkap pola jangka panjang dalam data waktu nyata, menjadikannya alat yang ideal untuk prediksi potensi banjir.

Seperti yang ditunjukkan dalam Kongres Teknologi Nasional (KTN), terdapat kebutuhan untuk mengubah paradigma sistem peringatan dini di Indonesia dari pendekatan yang berbasis prakiraan cuaca murni ke pendekatan yang berbasis risiko (BPPT, 2018). Sistem berbasis risiko ini tidak hanya berfokus pada prakiraan cuaca, tetapi juga memperhitungkan faktor-faktor risiko seperti kerentanan wilayah, kepadatan populasi, dan dampak ekonomi dari bencana banjir. Namun, implementasi perubahan paradigma ini masih memerlukan penelitian lebih lanjut, khususnya dalam pengembangan model prediksi yang mampu menggabungkan data cuaca dengan variabel risiko tersebut.

Berdasarkan gap-gap tersebut, penelitian ini bertujuan untuk mengembangkan model prediksi curah hujan dan potensi banjir berbasis CatBoost dan LSTM yang dapat mendukung sistem peringatan dini bencana alam di Kota Sleman. Melalui penelitian ini, diharapkan akurasi prediksi curah hujan dan banjir dapat ditingkatkan, sehingga sistem peringatan dini berbasis risiko dapat lebih efektif dalam mencegah dan mengurangi kerugian yang disebabkan oleh banjir. Model prediktif yang dikembangkan akan diintegrasikan ke dalam dashboard interaktif untuk memudahkan masyarakat dan pemerintah dalam mengakses informasi prediksi, sehingga memungkinkan tindakan mitigasi yang lebih cepat dan tepat.

1.2. Rumusan Masalah

Dalam rangka menciptakan sistem peringatan dini bencana alam yang efektif dan responsif, serta mengingat tantangan yang dihadapi Indonesia dalam menghadapi bencana alam hidrometeorologi, penelitian ini dirancang untuk menjawab pertanyaan-pertanyaan berikut:

- a. Bagaimana kombinasi algoritma CatBoost dan LSTM dapat dimanfaatkan untuk mengembangkan model prediksi potensi banjir yang akurat dan dapat diintegrasikan ke dalam sistem peringatan dini bencana hidrometeorologi di Indonesia?

- b. Berapa tingkat akurasi yang dapat dicapai oleh model CatBoost dan LSTM dalam memprediksi potensi banjir ketika dibandingkan dengan metode pembelajaran mesin tradisional yang saat ini digunakan?
- c. Faktor-faktor apa yang mempengaruhi efektivitas model CatBoost dan LSTM dalam memprediksi potensi banjir, dan bagaimana faktor-faktor tersebut dapat dioptimalkan untuk meningkatkan akurasi prediksi?

Dengan menjawab pertanyaan-pertanyaan ini, penelitian ini bertujuan untuk mengatasi gap yang ada dalam literatur sebelumnya dan menawarkan solusi yang inovatif untuk meningkatkan kapasitas sistem peringatan dini di Indonesia, khususnya dalam menghadapi peningkatan risiko bencana hidrometeorologi.

1.3. Batasan Masalah

Bagian Dalam rangka memberikan fokus yang jelas terhadap penelitian ini, beberapa batasan masalah diletakkan sebagai berikut:

- a. Penelitian ini terbatas pada pengembangan dan evaluasi model prediksi potensi banjir dan potensi banjir menggunakan algoritma CatBoost dan LSTM. Data yang digunakan berasal dari data cuaca dan kejadian banjir di Kota Sleman, Indonesia, dalam periode Januari 2020 hingga Mei 2024. Penelitian ini berfokus pada konteks geografis lokal di Indonesia untuk menguji efektivitas model dalam memprediksi risiko banjir.
- b. Asumsi yang digunakan dalam penelitian ini adalah bahwa data historis yang tersedia cukup representatif untuk menangkap pola cuaca ekstrem yang relevan dengan potensi banjir di Kota Sleman. Selain itu, model yang dikembangkan diharapkan mampu mempelajari pola temporal dalam data cuaca tanpa memerlukan pemahaman mendalam tentang proses fisik atmosfer. Studi ini menggunakan pendekatan pembelajaran mesin yang mengandalkan analisis data temporal tanpa menggali detail fisik dari dinamika atmosfer.
- c. Batasan solusi dalam penelitian ini berfokus pada peningkatan akurasi prediksi dan kecepatan pelatihan model. Meskipun model ini memiliki potensi untuk diterapkan pada berbagai jenis prediksi cuaca lainnya, penelitian ini tidak akan mencakup implementasi model secara langsung ke dalam

infrastruktur sistem peringatan dini yang ada atau pengembangan protokol respons bencana berbasis hasil prediksi model.

1.4. Tujuan Penelitian

Penelitian ini bertujuan untuk mengembangkan dan menguji model prediksi potensi banjir dengan tujuan-tujuan spesifik sebagai berikut:

- a. Menciptakan model prediktif berbasis CatBoost dan LSTM yang mampu menghasilkan prediksi potensi banjir dan potensi banjir dengan tingkat akurasi yang tinggi, guna mendukung sistem peringatan dini bencana hidrometeorologi di Indonesia.
- b. Mengidentifikasi dan mengevaluasi faktor-faktor kunci yang berpengaruh terhadap performa prediksi model CatBoost dan LSTM, seperti variabel cuaca yang paling relevan dan parameter optimal dalam konfigurasi model.

1.5. Manfaat Penelitian

- a. Dari segi ilmiah, penelitian ini berkontribusi pada pengetahuan dalam bidang pembelajaran mesin dan ilmu atmosfer, khususnya dalam pengembangan model prediktif berbasis CatBoost dan LSTM yang efektif dalam mengolah data temporal dan kategorikal. Keberhasilan penelitian ini akan membuka peluang bagi penelitian lanjutan yang dapat lebih jauh mengoptimalkan pemanfaatan AI dalam prediksi cuaca dan mitigasi bencana hidrometeorologi.
- b. Dari segi teknis, sistem yang dikembangkan diharapkan dapat mendukung pengambilan keputusan yang lebih tepat dan cepat dalam manajemen bencana alam. Dengan kemampuan untuk memprediksi potensi banjir secara akurat, sistem ini dapat membantu pihak berwenang dalam mempersiapkan respons yang efektif terhadap potensi bencana, sehingga mengurangi dampak negatifnya.
- c. Manfaat praktis dari penelitian ini bagi instansi atau perusahaan, khususnya bagi Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) atau lembaga pemerintah dan non-pemerintah yang mengelola risiko bencana, adalah peningkatan kapasitas dalam menyediakan peringatan dini yang akurat kepada masyarakat. Hal ini akan secara signifikan mengurangi kerugian materiil dan risiko kehilangan nyawa akibat bencana alam yang disebabkan oleh cuaca ekstrem.

1.6. Hipotesis

Berdasarkan tinjauan pustaka dan landasan teori yang ada, hipotesis yang diajukan dalam penelitian ini adalah sebagai berikut:

- a. Model prediksi cuaca ekstrem dan potensi banjir yang dikembangkan dengan menggunakan kombinasi CatBoost dan LSTM akan memiliki tingkat akurasi yang lebih tinggi dalam memprediksi potensi banjir dibandingkan dengan metode pembelajaran mesin tradisional yang saat ini banyak digunakan.
- b. Implementasi optimasi hyperparameter yang tepat pada CatBoost dan LSTM akan meningkatkan akurasi prediksi serta efisiensi proses pelatihan model, memungkinkan model untuk menangkap pola cuaca dengan lebih efektif.
- c. Model prediksi yang dikembangkan akan efektif dalam mendukung sistem peringatan dini bencana hidrometeorologi, dengan memberikan prediksi yang cepat dan akurat, yang pada gilirannya dapat mengurangi dampak negatif dari bencana alam yang terjadi di Kota Sleman.

Untuk menguji hipotesis ini, akan digunakan metode pengujian empiris melalui eksperimen komputasi dengan dataset cuaca dan data banjir dari Kota Sleman. Model CatBoost dan LSTM akan dibandingkan dengan model pembelajaran mesin tradisional dalam hal akurasi prediksi dan efisiensi waktu pelatihan. Pengujian statistik, seperti uji t, akan dilakukan untuk menentukan apakah perbedaan kinerja antara kedua model tersebut secara signifikan lebih baik dan sesuai dengan yang dihipotesiskan.

BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

Berbagai metode statistik dan teknik pembelajaran mesin telah diterapkan pada data historis mengenai suhu, curah hujan, dan variabel meteorologi lainnya untuk memprediksi kondisi cuaca. Metode statistik tradisional seperti *Auto-regressive Integrated Moving Average (ARIMA)* telah lama digunakan untuk analisis seri waktu dalam memprediksi pola cuaca (Babu dan Reddy, 2012). Di sisi lain, metode jaringan saraf buatan (*Artificial Neural Network, ANN*) juga telah banyak diterapkan dalam analisis data cuaca, seperti prediksi suhu (Corchado dan Fyfe, 1999; Baboo dan Shereef, 2010; Mehdizadeh, 2018). Namun, metode-metode ini terbatas pada pola linier dan sering kali kurang efektif dalam menangani ketergantungan temporal yang kompleks.

Penelitian terkait prediksi cuaca berbasis *deep learning* mulai menggabungkan model spatiotemporal yang mampu menangani data dengan ketergantungan spasial dan temporal. Shi et al. (2015) memperkenalkan arsitektur *Convolutional LSTM (ConvLSTM)* yang mampu menangkap pola spasial dalam data radar cuaca untuk *nowcasting* presipitasi. Arsitektur ini berhasil memasukkan operasi konvolusi ke dalam jaringan LSTM, sehingga memungkinkan pemodelan *end-to-end* untuk prediksi hujan. Kim et al. (2019) juga menggunakan ConvLSTM dalam tugas peramalan peristiwa iklim ekstrem, dengan input berupa rangkaian peta kepadatan badai sebagai data spatiotemporal.

Selain itu, pendekatan hibrida yang menggabungkan *CNN dan LSTM* terus berkembang untuk menangkap pola kompleks dalam data cuaca. Souto et al. (2018) mengusulkan model *ensemble* yang menggabungkan ConvLSTM dengan pemahaman spatiotemporal untuk memprediksi peristiwa cuaca ekstrem. Wang et al. (2019) lebih lanjut mengembangkan unit LSTM yang mampu mengelola variasi spasial dan temporal dengan menambahkan pemodelan *non-stasioneritas* untuk mengatasi variabilitas cuaca.

Meskipun demikian, banyak dari pendekatan *deep learning* yang berfokus pada data cuaca dan iklim masih mengadopsi CNN atau ConvLSTM tanpa memanfaatkan kekuatan algoritma berbasis pohon keputusan yang fleksibel dalam menangani data kategori, seperti CatBoost. CatBoost merupakan algoritma berbasis pohon yang dirancang khusus untuk menangani data kategori dengan mengurangi bias dalam penanganan kategori melalui teknik *Ordered Boosting* (Dorogush et al., 2018).

Keunggulan ini menjadikan CatBoost sebagai pilihan ideal untuk mengidentifikasi variabel cuaca penting yang mungkin berpengaruh pada risiko banjir.

Dalam penelitian ini, kombinasi CatBoost dan LSTM digunakan untuk memanfaatkan kekuatan dari kedua model: CatBoost untuk menganalisis dan mengidentifikasi fitur penting dalam data cuaca, serta LSTM untuk menangkap pola temporal dalam data cuaca jangka panjang. Pendekatan ini diharapkan dapat memberikan prediksi yang lebih akurat dan relevan untuk mendukung sistem peringatan dini bencana hidrometeorologi, khususnya dalam konteks prediksi potensi banjir di Kota Sleman.

Beberapa studi telah menerapkan konvolusi spatiotemporal untuk berbagai tugas pengenalan pola yang memerlukan analisis data spasial dan temporal secara simultan, seperti analisis video dan pengenalan aksi. Yuan et al. (2018) dan Tran et al. (2018) mengeksplorasi penggunaan *3D CNN* untuk mempelajari korelasi spasial dan temporal secara bersamaan dalam data video. Dalam studi Tran et al. (2018), peneliti menunjukkan bahwa faktorisasi kernel konvolusi 3D menjadi konvolusi spasial dan temporal yang terpisah memberikan peningkatan akurasi pada pengenalan pola, meskipun masih memiliki keterbatasan. Salah satu keterbatasan *3D CNN* adalah kurangnya kendala kausal, yang dapat melanggar urutan temporal, yang merupakan faktor penting dalam prediksi berbasis waktu.

Untuk mengatasi masalah kendala kausal dalam pembelajaran temporal, beberapa studi mengusulkan pendekatan alternatif. Singh dan Cuzzolin (2019) memperkenalkan unit konvolusi berulang untuk mempertahankan urutan temporal dalam pembelajaran spatiotemporal, sementara Cheng et al. (2019) mengadopsi konvolusi kausal dalam konvolusi spasial dan temporal yang terpisah namun paralel. Implementasi faktorisasi konvolusi 3D ini mendukung pelatihan yang lebih efektif tanpa melanggar urutan temporal data.

Dalam penelitian prediksi cuaca, pendekatan hibrida seperti *CNN 3D* yang difaktorisasi juga telah diadaptasi untuk menangkap pola kompleks dalam data iklim. Racah et al. (2017), misalnya, menggunakan *CNN 3D* dalam arsitektur enkoder-dekoder untuk mendeteksi peristiwa iklim ekstrem. Mereka merancang model dengan jalur penyusutan dalam enkoder, menggunakan tumpukan lapisan konvolusional, dan jalur peningkatan dalam dekoder untuk menghasilkan keluaran dengan dimensi yang sesuai

dengan masukan. Pendekatan ini memungkinkan model untuk menangkap pola spasial dan temporal dalam data iklim, sehingga memberikan prediksi yang lebih akurat.

Mengikuti keberhasilan CNN 2D dalam menangkap korelasi spasial, Xu et al. (2019) menggunakan CNN 2D untuk menangkap korelasi spasial dan temporal secara terpisah dalam tugas prediksi emisi polusi kendaraan. Pendekatan ini menunjukkan bahwa pemisahan antara konvolusi spasial dan temporal dapat efektif dalam mengatasi korelasi data yang kompleks.

Penelitian ini mengadaptasi teknik spatiotemporal yang relevan dengan memanfaatkan kombinasi CatBoost dan LSTM. CatBoost, sebagai model berbasis pohon keputusan, sangat efektif dalam menangani variabel-variabel kategori yang kompleks dan menghasilkan analisis *feature importance* yang mendalam. Di sisi lain, LSTM, yang dirancang untuk menangkap pola temporal, sangat berguna dalam mengolah data sekuensial dalam jangka panjang, seperti data cuaca harian. Pendekatan ini bertujuan untuk mengatasi keterbatasan dalam metode statistik tradisional dan 3D CNN dengan memperhitungkan variabilitas temporal dalam prediksi cuaca ekstrem dan potensi banjir.

Dengan demikian, meskipun CNN dan arsitektur spatiotemporal telah digunakan secara luas dalam berbagai aplikasi, penelitian ini menggunakan kombinasi CatBoost dan LSTM sebagai alternatif yang lebih fleksibel dan terfokus pada prediksi bencana hidrometeorologi. Diharapkan, pendekatan ini dapat memberikan prediksi yang lebih akurat dan mendukung sistem peringatan dini dengan mempertimbangkan kondisi iklim lokal di Kota Sleman.

2.2. Keaslian Penelitian

Tabel 2.1. Matriks literatur review dan posisi penelitian
Pengembangan Model Prediksi Curah Hujan Menggunakan Convolutional
Neural Network Untuk Mendukung Sistem Peringatan Dini Bencana Alam

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
1	" Daily Rainfall Prediction Using One Dimensional Convolutional Neural Networks "	Yuana. et al.	IEEE, 2020	Menilai efektivitas metode Convolutional Neural Network (CNN) dalam prediksi cuaca dan curah hujan di Indonesia	Hasil penelitian menunjukkan bahwa CNN dapat memberikan prediksi cuaca dan curah hujan yang akurat di Indonesia	Diperlukan penyesuaian model CNN dengan faktor-faktor lokal seperti topografi dan pola cuaca regional
2	" An Application of Deep Learning Technique to Improve Subseasonal to Seasonal Rainfall Forecast over Java Island, Indonesia "	Raharja. et al.	JPSL IPB, 2022	Meningkatkan akurasi prediksi curah hujan di Indonesia dengan menggunakan model berbasis CNN	Penelitian ini berhasil meningkatkan akurasi prediksi curah hujan dengan menggunakan model CNN	Saran untuk penelitian selanjutnya adalah menggabungkan data cuaca lokal dengan data citra satelit untuk meningkatkan performa model

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
3	"Developing a knowledge management system for supporting flood decision-making "	Andi, S. et al.	IEEE, 2022	Mengembangkan model prediksi curah hujan berbasis CNN untuk sistem peringatan dini banjir di Indonesia	Model prediksi berbasis CNN memberikan hasil yang baik dalam memprediksi curah hujan untuk sistem peringatan dini banjir	Penelitian ini menyebutkan bahwa pengumpulan data cuaca yang akurat dan berkualitas masih menjadi tantangan
4	"Rainfall Prediction using Spatial Convolutional Neural Networks and Recurrent Neural Networks "	Nadfa Lestari, et al.	IEEE, 2022	Membandingkan kinerja model CNN dan RNN dalam prediksi curah hujan di Indonesia	Kedua model CNN dan RNN mampu memberikan prediksi curah hujan yang baik, namun CNN cenderung memberikan hasil yang lebih akurat	Kelemahan dalam penggunaan model CNN adalah interpretasi hasil prediksi yang kompleks
5	" Single Layer & Multi-layer Long Short-Term Memory (LSTM) Model with Intermediate	Salman. et al.	ScienceDirect, 2018	Meningkatkan akurasi prediksi curah hujan dengan menggunakan ensemble model LSTM di Indonesia	Penggunaan ensemble model LSTM dapat menghasilkan prediksi curah hujan yang lebih akurat dibandingkan	Kelemahan dari ensemble model adalah kompleksitas komputasi yang lebih tinggi dibandingkan dengan model individu

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
	Variables for Weather Forecasting "				dengan menggunakan model individu	
6	" Prediction of Rainfall using Simplified Deep Learning based Extreme Learning Machines "	Imam, C. et al.	JITeCS, 2018	Mempelajari pengaruh representasi data input pada prediksi curah hujan menggunakan model CNN di Indonesia	Hasil penelitian menunjukkan bahwa pemilihan representasi data input yang tepat dapat meningkatkan akurasi prediksi curah hujan	Kelemahan dalam pemilihan representasi data input adalah ketergantungan pada kualitas data dan interpretasi yang kompleks

2.3. Landasan Teori

A. Meteorologi dan Hidrometeorologi

Meteorologi adalah ilmu yang mempelajari atmosfer dan proses-proses yang terjadi di dalamnya, termasuk dinamika cuaca dan iklim. Aspek hidrometeorologi secara spesifik menekankan pada studi siklus air di atmosfer, khususnya berkaitan dengan fenomena presipitasi seperti hujan, salju, dan hujan es. Pemahaman terhadap fenomena hidrometeorologi sangat penting dalam memprediksi cuaca ekstrem, yang seringkali menjadi penyebab bencana alam seperti banjir dan tanah longsor.

Pengetahuan ini tidak hanya penting untuk memahami bagaimana fenomena cuaca ekstrem terbentuk dan berkembang tetapi juga untuk memahami distribusi temporal dan spasial dari fenomena tersebut. Misalnya, hujan lebat yang berkepanjangan dapat mengakibatkan peningkatan aliran permukaan dan berpotensi menyebabkan banjir. Di sisi lain, perubahan cepat dalam kondisi atmosfer dapat memicu terjadinya tanah longsor, khususnya di area yang topografinya rentan (Abhishek, 2012).

Studi tentang dinamika atmosfer memberikan wawasan tentang pola cuaca dan sistem cuaca, yang sangat berguna dalam meramalkan peristiwa cuaca dalam jangka pendek dan menengah. Model-model atmosfer canggih yang menggabungkan prinsip-prinsip fisika dan dinamika fluida telah dikembangkan untuk menghasilkan simulasi dan prediksi yang akurat (Kareem, 2021).

Dalam konteks prediksi potensi banjir, penggunaan data meteorologi historis memainkan peran penting. Data historis ini, yang mencakup suhu, curah hujan, kelembaban, tekanan atmosfer, dan variabel lainnya, digunakan untuk melatih model prediktif yang bisa memproyeksikan keadaan atmosfer di masa depan berdasarkan pola yang telah ada. Dengan mengekstraksi pola dari data ini, peneliti dapat mengidentifikasi tanda-tanda awal dari perubahan cuaca yang dapat berujung pada kondisi ekstrem (Deo, 2015).

Analisis statistik dari data historis, seperti trend dan variasi musiman, juga memberikan kontribusi dalam menentukan probabilitas kejadian cuaca ekstrem. Hal ini penting untuk mengembangkan model prediksi yang tidak hanya akurat tetapi juga dapat mengkomunikasikan risiko kepada para pemangku kebijakan

dan masyarakat umum untuk mempersiapkan dan merespons dengan tepat (Naseem, 2018).

Pendekatan ini diintegrasikan dengan teknologi canggih seperti satelit dan radar cuaca, yang memberikan data waktu nyata dan resolusi tinggi yang diperlukan untuk validasi dan peningkatan model prediktif. Kombinasi antara pengamatan langsung dan model berbasis data historis ini menghasilkan sistem peringatan dini yang lebih andal, yang sangat penting dalam mitigasi bencana dan pengelolaan risiko hidrometeorologi (Beritelli, 2020).

Oleh karena itu, meteorologi dan hidrometeorologi tidak hanya memberikan landasan dalam memahami fenomena cuaca ekstrem tetapi juga dalam pengembangan solusi inovatif untuk menghadapi tantangan yang dihasilkan oleh peristiwa-peristiwa tersebut.

B. Teori Pembelajaran Mesin dan Pembelajaran Mendalam (*Deep Learning*)

Membahas Pembelajaran mesin adalah bidang ilmu komputer yang memberikan sistem kemampuan untuk belajar dari data dan memperbaiki kinerja dari waktu ke waktu secara otomatis, tanpa diprogram secara eksplisit. Ini termasuk pembelajaran terawasi, di mana model dilatih pada dataset dengan label yang diketahui; pembelajaran tidak terawasi, di mana model mencoba menemukan pola dalam data tanpa label; dan pembelajaran semi-terawasi atau pembelajaran penguatan, di mana model belajar melalui interaksi dengan lingkungan (Larson, 2020).

Deep learning, yang merupakan subbidang dari pembelajaran mesin, terdiri dari algoritma yang dapat menentukan representasi abstrak yang berguna dari data melalui proses yang terdiri dari banyak tingkatan pemrosesan non-linear. Deep learning seringkali menggunakan variasi dari jaringan saraf buatan (Artificial Neural Networks, ANN) sebagai model pembelajaran (Chaabani, 2018).

Jaringan Saraf Buatan (ANN) dan *Convolutional Neural Networks* (CNN): ANN adalah model komputasi yang terinspirasi oleh jaringan saraf biologis dari otak manusia. Model-model ini terdiri dari simpul-simpul atau neuron yang saling

terhubung dan dapat memproses informasi dengan cara yang mirip dengan cara otak memproses informasi (Hasan, 2019).

CNN merupakan jenis khusus dari *ANN* yang dirancang untuk memproses data yang memiliki hubungan grid-like, seperti gambar. *CNN* berhasil dalam aplikasi pengenalan gambar dan video karena kemampuan mereka untuk otomatis dan adaptif belajar representasi spasial dari gambar. *CNN* menggunakan teknik yang disebut konvolusi, yang efektif dalam menangkap hubungan lokal dan komposisi hierarkis dari data visual.

Pada penelitian sebelumnya, *LSTM* telah digunakan bersama dengan *CNN* untuk mengambil keuntungan dari kemampuan *LSTM* dalam memodelkan ketergantungan temporal dan kemampuan *CNN* dalam mengidentifikasi pola spasial. Penelitian ini telah membuktikan efektivitas pendekatan ini dalam mengatasi tantangan yang ada dalam peramalan cuaca (Durrant, 2020).

Penggunaan teknik-teknik ini dalam konteks pembelajaran mesin untuk prediksi cuaca menawarkan kemajuan yang signifikan dibandingkan dengan metode tradisional, memberikan kemampuan untuk mengolah dan memahami jumlah data yang besar dengan keakuratan yang tinggi. Penelitian di bidang ini terus berkembang, dengan inovasi baru yang berusaha menggabungkan teknik pembelajaran mesin dan kecerdasan buatan yang lebih maju untuk mengatasi batasan yang ada dan meningkatkan kemampuan prediktif dari model cuaca.

C. **Recurrent Neural Network (RNN)**

Recurrent Neural Network (*RNN*) adalah jenis jaringan saraf tiruan yang dirancang untuk menangani data sekuensial atau time series, yang sering kali muncul dalam berbagai bidang, seperti pemrosesan bahasa alami (*NLP*), pengenalan suara, dan prediksi data temporal lainnya. Berbeda dengan jaringan saraf feedforward, *RNN* memiliki koneksi berulang yang memungkinkan model untuk mempertahankan informasi dari input sebelumnya dalam suatu "memori" internal. Dengan demikian, *RNN* dapat menangkap ketergantungan temporal dalam data yang berurutan, menjadikannya ideal untuk menangani data yang memerlukan konteks historis (Elman, 1990).

Pada prinsipnya, setiap neuron dalam RNN memiliki koneksi balik ke dirinya sendiri atau ke neuron di lapisan sebelumnya, yang memungkinkan informasi dari langkah waktu sebelumnya untuk mempengaruhi output pada langkah waktu berikutnya. Arsitektur ini memungkinkan RNN untuk “mengingat” informasi dari input sebelumnya, yang berguna dalam memodelkan proses yang bergantung pada urutan data (Rumelhart, Hinton, & Williams, 1986). Namun, meskipun kemampuan ini membuat RNN cocok untuk banyak aplikasi berbasis sekuensial, RNN tradisional sering kali mengalami masalah yang dikenal sebagai *vanishing gradient* dan *exploding gradient* ketika mencoba menangani ketergantungan jangka panjang dalam data (Hochreiter, 1991).

Masalah *vanishing gradient* terjadi ketika gradien dari fungsi loss menjadi sangat kecil selama proses backpropagation, yang menyebabkan model kesulitan dalam memperbarui bobot dengan efektif pada lapisan-lapisan awal jaringan. Akibatnya, kemampuan RNN untuk “mengingat” informasi dari langkah waktu yang lebih jauh melemah secara signifikan. Sebaliknya, *exploding gradient* terjadi ketika gradien menjadi sangat besar, yang dapat mengakibatkan ketidakstabilan dalam pelatihan model (Bengio, Simard, & Frasconi, 1994).

Karena keterbatasan ini, RNN tradisional kurang optimal dalam menangani ketergantungan temporal jangka panjang. Untuk mengatasi hal ini, dikembangkanlah variasi RNN yang lebih kompleks, seperti Long Short-Term Memory (LSTM) dan Gated Recurrent Unit (GRU), yang menggunakan mekanisme gerbang untuk mengatur aliran informasi secara lebih efisien. LSTM dan GRU memiliki kemampuan untuk mempertahankan informasi dalam memori lebih lama dibandingkan dengan RNN tradisional, sehingga lebih efektif dalam memodelkan data yang memiliki ketergantungan temporal jangka panjang (Hochreiter & Schmidhuber, 1997; Cho et al., 2014).

Studi-studi sebelumnya telah menunjukkan bahwa RNN, terutama dalam bentuk LSTM dan GRU, berhasil diterapkan dalam berbagai kasus yang melibatkan data sekuensial. Misalnya, dalam prediksi curah hujan dan suhu yang membutuhkan informasi historis cuaca, model berbasis RNN dapat memberikan performa yang unggul dibandingkan dengan model prediksi berbasis statistik konvensional (Lipton, 2015; Graves, 2013). RNN memungkinkan model untuk

mengenal pola temporal yang kompleks, yang menjadi dasar dalam analisis time series atau data sekuensial lainnya.

D. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) adalah jenis jaringan saraf tiruan yang digunakan dalam memproses data sekuensial dan menangani ketergantungan jangka panjang. LSTM memiliki tiga komponen utama, yaitu gerbang lupa, gerbang input, dan gerbang output, yang mengatur aliran informasi dalam unit LSTM. Berikut ini adalah penjelasan setiap poin dengan ilustrasi untuk membantu pemahaman.

Struktur Dasar LSTM

LSTM terdiri dari beberapa komponen kunci yang bekerja bersama untuk mengingat atau melupakan informasi dalam sel memori. Struktur dasarnya dapat digambarkan sebagai berikut:

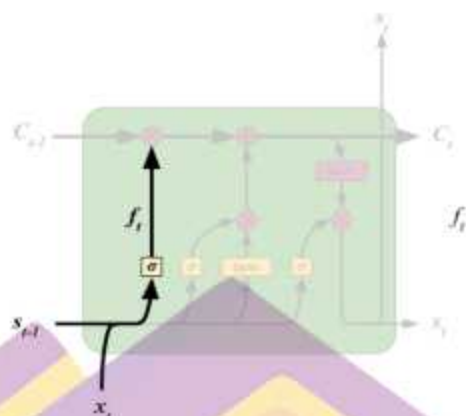
1. Gerbang Lupa (Forget Gate)

Gerbang Lupa memutuskan informasi mana dari sel memori sebelumnya (C_{t-1}) yang perlu dilupakan. Ini penting agar unit LSTM tidak menyimpan informasi yang tidak lagi relevan.

Rumus:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f)$$

- W_f adalah bobot untuk gerbang lupa.
- h_{t-1} adalah *hidden state* dari waktu sebelumnya.
- x_t adalah input pada waktu t .
- b_f adalah bias untuk gerbang lupa.
- σ adalah fungsi aktivasi sigmoid, yang mengubah nilai menjadi antara 0 dan 1.



Gambar 2.1 Visualisasi Gerbang Lupa

Dalam gambar ini, Anda bisa melihat bahwa *hidden state* dan input saat ini dikombinasikan, lalu diterapkan fungsi sigmoid untuk menghasilkan output antara 0 dan 1. Nilai ini kemudian dikalikan dengan sel memori sebelumnya (C_{t-1}), sehingga hanya informasi penting yang disimpan.

2. Gerbang Input (Input Gate)

Gerbang Input menentukan informasi baru mana yang akan ditambahkan ke dalam sel memori. Proses ini terdiri dari dua tahap:

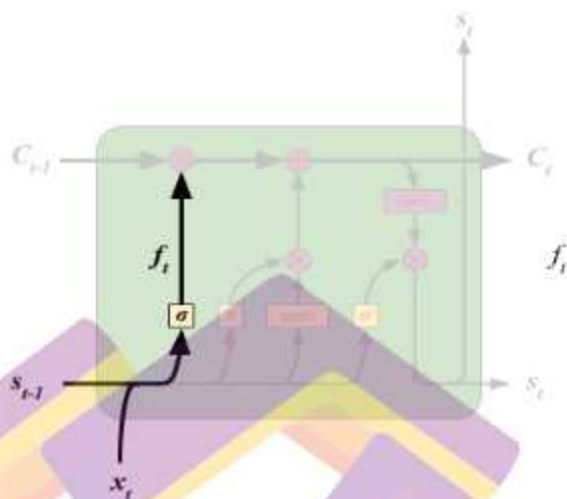
1. Menggunakan lapisan sigmoid untuk memutuskan informasi mana yang akan diperbarui.
2. Menggunakan lapisan tanh untuk menciptakan kandidat memori baru (\tilde{C}_t) yang akan ditambahkan ke dalam sel memori.

Rumus:

2

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c)$$



Gambar 2.2 Visualisasi Gerbang Input

Pada gambar ini, terlihat bahwa gerbang input memiliki dua tahap: pertama, sigmoid menentukan bagian mana yang harus diperbarui, dan kedua, tanh membuat kandidat memori yang siap ditambahkan ke sel memori.

3. Memperbarui Sel Memori

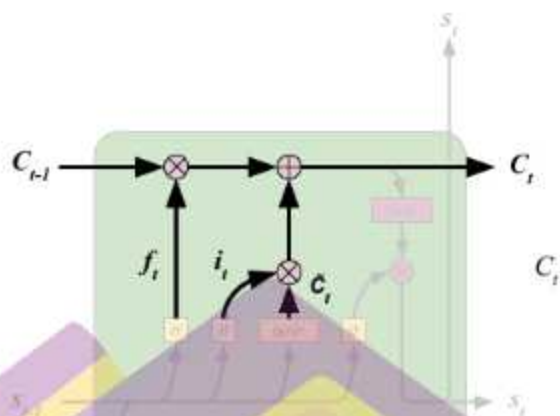
Setelah mendapatkan output dari gerbang lupa dan gerbang input, kita memperbarui sel memori dengan informasi penting yang baru.

Rumus:

3

$$C_t = f_t \times C_{[t-1]} + i_t \times \tilde{[C]}_t$$

- $f_t \times C_{[t-1]}$ menunjukkan informasi yang diingat dari memori sebelumnya.
- $i_t \times \tilde{[C]}_t$ menunjukkan informasi baru yang ditambahkan.



Gambar 2.3 Visualisasi Memperbarui Sel Memori

Dalam ilustrasi ini, Anda bisa melihat bahwa informasi dari memori sebelumnya dikombinasikan dengan informasi baru, menghasilkan sel memori yang diperbarui (C_t).

4. Gerbang Output (Output Gate)

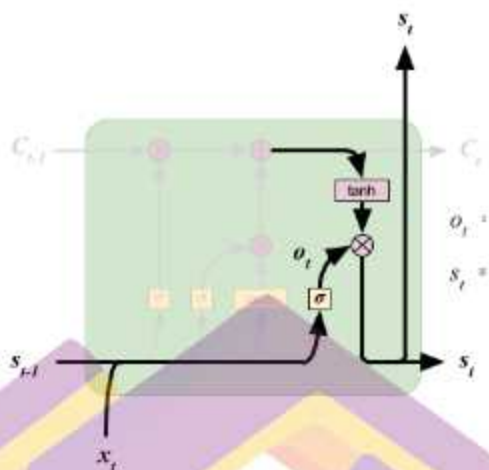
Gerbang Output menentukan informasi apa yang akan diambil dari sel memori untuk digunakan sebagai output pada langkah waktu saat ini (h_t).

Rumus:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o)$$

$$h_t = o_t \times \tanh(C_t)$$

4



Gambar 2.4 Visualisasi Gerbang Output

Dalam ilustrasi ini, output dari gerbang output (o_t) dikalikan dengan \tanh dari sel memori yang diperbarui untuk menghasilkan *hidden state* baru (h_t), yang bisa digunakan untuk prediksi atau diteruskan ke langkah berikutnya dalam sekuens.

E. CatBoost sebagai Model Pembelajaran Mesin Berbasis Pohon Keputusan

CatBoost (Categorical Boosting) adalah algoritma pembelajaran mesin berbasis pohon keputusan yang dikembangkan oleh perusahaan teknologi Yandex. *CatBoost* merupakan salah satu algoritma *gradient boosting* yang dirancang untuk menangani variabel kategori secara efisien dan mengatasi beberapa kelemahan dari algoritma boosting tradisional, seperti *gradient boosting decision tree (GBDT)* dan *XGBoost*. Menurut Dorogush et al. (2018), *CatBoost* unggul dalam menangani data kategori, mengurangi bias, dan memberikan prediksi yang akurat bahkan dengan dataset yang terbatas.

Konsep Dasar dan Mekanisme CatBoost

CatBoost menggunakan pendekatan gradient boosting, di mana model dibangun secara bertahap dengan meminimalkan kesalahan dari model sebelumnya. Setiap model yang ditambahkan bertujuan untuk mengoreksi kesalahan dari model sebelumnya, sehingga secara keseluruhan algoritma ini menghasilkan model yang semakin akurat. Namun, perbedaan utama CatBoost dibandingkan dengan algoritma boosting lainnya adalah tekniknya dalam menangani variabel kategori dan mengurangi masalah overfitting.

1. **Ordered Boosting:** Salah satu keunggulan utama CatBoost adalah penggunaan teknik yang disebut ordered boosting. Pada algoritma boosting tradisional, urutan data dapat menyebabkan kebocoran informasi, yang mengakibatkan bias pada hasil prediksi. CatBoost mengatasi masalah ini dengan membangun pohon-pohon keputusan secara berurutan berdasarkan subset data, sehingga urutan data tidak memengaruhi hasil akhir. Menurut Prokhorenkova et al. (2018), ordered boosting memungkinkan CatBoost untuk menghasilkan model yang lebih stabil dan akurat dibandingkan algoritma boosting lainnya.
2. **Penanganan Data Kategori:** CatBoost secara khusus dirancang untuk menangani variabel kategori tanpa perlu melakukan one-hot encoding atau transformasi variabel lainnya. CatBoost menggunakan teknik target encoding yang canggih untuk mengonversi variabel kategori menjadi bentuk numerik yang dapat diproses oleh model. Dengan demikian, CatBoost lebih efisien dalam menangani data kategori dan dapat mengurangi ukuran dataset secara signifikan, yang pada akhirnya mempercepat proses pelatihan model.
3. **Penanganan Overfitting:** CatBoost memiliki kemampuan bawaan untuk mengatasi masalah overfitting, yang sering kali menjadi masalah pada algoritma boosting. Dengan teknik random permutations dan ordered boosting, CatBoost mampu meminimalkan bias data yang disebabkan oleh urutan data atau variabel kategori. Menurut Dorogush et al. (2018), fitur ini menjadikan CatBoost pilihan yang baik untuk dataset dengan jumlah

sampel yang relatif kecil, seperti data cuaca historis untuk wilayah tertentu.

Aplikasi CatBoost dalam Prediksi Cuaca dan Bencana Hidrometeorologi

CatBoost telah diterapkan dalam berbagai penelitian untuk prediksi cuaca dan mitigasi risiko bencana hidrometeorologi. Sebagai contoh, Zhao et al. (2020) menggunakan CatBoost untuk memprediksi curah hujan harian di wilayah tropis dan menemukan bahwa CatBoost mampu menghasilkan prediksi yang lebih akurat dibandingkan dengan model regresi tradisional. Dengan kemampuan untuk memproses data kategori secara efisien, CatBoost juga digunakan untuk menganalisis data cuaca ekstrem, seperti badai atau hujan lebat, yang sering kali melibatkan variabel-variabel kategori.

Dalam sistem peringatan dini bencana, CatBoost dapat digunakan untuk mengidentifikasi area dengan risiko tinggi berdasarkan data cuaca historis dan kondisi atmosfer saat ini. Model ini dapat membantu pemerintah dan lembaga terkait untuk mengambil tindakan pencegahan yang lebih efektif. Misalnya, dalam prediksi banjir, CatBoost dapat mengidentifikasi faktor-faktor penting yang berkontribusi terhadap peningkatan aliran sungai, seperti curah hujan dan kelembapan tanah. Dengan hasil prediksi yang akurat, langkah-langkah mitigasi dapat dilakukan sebelum banjir terjadi.

Penggunaan CatBoost Bersama LSTM untuk Prediksi Cuaca

Salah satu pendekatan inovatif dalam prediksi cuaca adalah menggabungkan CatBoost dengan LSTM untuk memanfaatkan keunggulan masing-masing algoritma. CatBoost digunakan untuk mengidentifikasi fitur penting dalam data cuaca, sedangkan LSTM digunakan untuk menangkap pola temporal dalam data seri waktu. Kombinasi ini memungkinkan pemodelan yang lebih akurat dan komprehensif, yang mampu memprediksi cuaca dengan mempertimbangkan hubungan spasial-temporal antara variabel cuaca.

Menurut Sun et al. (2021), kombinasi CatBoost dan LSTM dalam prediksi telah terbukti meningkatkan akurasi model dibandingkan dengan penggunaan salah satu algoritma secara terpisah. CatBoost memberikan analisis fitur yang

mendalam, sementara LSTM menangani pola temporal yang sangat penting dalam memahami evolusi kondisi cuaca dari waktu ke waktu.

F. CatBoost dan LSTM untuk Prediksi Cuaca

Dalam prediksi cuaca, integrasi beberapa algoritma pembelajaran mesin dan pembelajaran mendalam sering kali menghasilkan model yang lebih kuat dan akurat. Salah satu pendekatan yang menjanjikan adalah menggabungkan *CatBoost* dan *LSTM*. Kombinasi ini memanfaatkan kekuatan masing-masing algoritma: *CatBoost* untuk mengidentifikasi fitur penting dalam data dan menangani variabel kategori, serta *LSTM* untuk menangkap pola temporal dalam data seri waktu. Pendekatan hibrida ini telah digunakan dalam berbagai penelitian dan terbukti meningkatkan akurasi dalam prediksi cuaca ekstrem dan sistem peringatan dini bencana hidrometeorologi.

Rationale Penggabungan CatBoost dan LSTM

Penggunaan algoritma tunggal terkadang kurang optimal dalam memproses data cuaca yang kompleks dan memiliki banyak dimensi. *CatBoost* dan *LSTM* memiliki keunggulan yang berbeda yang, jika dikombinasikan, dapat meningkatkan akurasi prediksi. *CatBoost* sangat baik dalam mengidentifikasi variabel penting dalam data cuaca, terutama data kategori seperti jenis cuaca, lokasi geografis, atau musim. *LSTM*, di sisi lain, dirancang untuk menangani data seri waktu dengan ketergantungan jangka panjang, seperti perubahan suhu atau curah hujan dari waktu ke waktu.

Menurut Li dan Zhu (2021), penggabungan *CatBoost* dan *LSTM* memungkinkan model untuk memperoleh informasi dari dua perspektif: hubungan spasial antar variabel dan pola temporal yang penting dalam peramalan cuaca ekstrem. Pendekatan ini sangat berguna dalam mengidentifikasi hubungan antara variabel meteorologi yang mungkin tidak terlihat dalam analisis tunggal.

Langkah-langkah Implementasi Model Hybrid CatBoost-LSTM

1. Identifikasi dan Pra-Pemrosesan Data dengan *CatBoost*: Langkah pertama dalam membangun model hybrid ini adalah mengidentifikasi fitur

penting menggunakan CatBoost. CatBoost memproses data kategori secara efisien dan memberikan analisis feature importance yang menunjukkan variabel-variabel yang memiliki pengaruh signifikan terhadap hasil prediksi. Misalnya, variabel seperti musim, suhu, dan kelembapan mungkin memiliki pengaruh besar pada curah hujan.

2. Pelatihan Model LSTM dengan Data Seri Waktu: Setelah fitur penting diidentifikasi oleh CatBoost, data tersebut digunakan untuk melatih model LSTM. LSTM kemudian memanfaatkan data seri waktu untuk mengidentifikasi pola temporal dalam variabel-variabel cuaca yang penting. Proses ini memungkinkan model untuk menangkap ketergantungan jangka panjang yang memengaruhi cuaca ekstrem.

3. Kombinasi Prediksi: Setelah CatBoost dan LSTM dilatih, hasil dari kedua model dapat digabungkan. Ada beberapa metode untuk menggabungkan prediksi ini, misalnya dengan melakukan ensemble atau menggunakan layer tambahan untuk mengintegrasikan hasil dari kedua model. Dalam sistem ini, output dari CatBoost (sebagai pengidentifikasi fitur penting) dikombinasikan dengan prediksi temporal dari LSTM, yang memungkinkan sistem untuk memberikan prediksi yang lebih akurat.

CatBoost berperan dalam mengidentifikasi faktor risiko penting yang berkaitan dengan bencana hidrometeorologi, seperti curah hujan harian, kecepatan angin, dan kelembapan. LSTM kemudian menganalisis pola temporal dalam data tersebut, sehingga mampu memberikan prediksi yang lebih mendalam tentang kapan dan di mana potensi banjir kemungkinan akan terjadi.

BAB III METODE PENELITIAN

3.1. Jenis, Sifat, dan Pendekatan Penelitian

Pendekatan dalam penelitian ini berfokus pada upaya mengembangkan model prediksi potensi banjir yang diharapkan bisa menjadi bagian dari sistem peringatan dini bencana hidrometeorologi. Cuaca ekstrem adalah fenomena yang sulit diprediksi karena melibatkan data yang tidak hanya kompleks tetapi juga sangat dinamis—terdiri dari berbagai variabel yang saling berkaitan dan berfluktuasi sepanjang waktu. Dalam konteks ini, saya memilih untuk menggabungkan dua algoritma pembelajaran mesin, yaitu CatBoost dan LSTM, untuk memaksimalkan keakuratan prediksi.

CatBoost dan LSTM bukanlah pilihan yang sembarangan; keduanya memiliki kelebihan yang saling melengkapi. CatBoost, sebagai algoritma berbasis pohon keputusan, dirancang untuk bekerja secara efisien dengan variabel kategori tanpa memerlukan teknik transformasi yang rumit seperti one-hot encoding. Dalam data meteorologi, variabel kategori seperti jenis cuaca, musim, atau wilayah geografis sering kali memainkan peran penting, dan di sinilah CatBoost sangat membantu. Selain itu, CatBoost juga menawarkan kemampuan untuk menentukan variabel mana yang paling berpengaruh terhadap hasil prediksi, sebuah fitur yang sangat bermanfaat dalam konteks penelitian ini. Dengan memahami variabel mana yang penting, saya dapat menyederhanakan model sehingga lebih efisien dan fokus pada fitur yang benar-benar relevan dalam memprediksi potensi banjir.

Di sisi lain, LSTM—jenis jaringan saraf tiruan yang mampu mengolah data seri waktu—sangat ideal untuk menangani pola temporal dalam data cuaca. Berbeda dengan algoritma lain, LSTM dirancang untuk mengingat informasi dari waktu ke waktu, yang membuatnya cocok untuk analisis perubahan cuaca, seperti pola curah hujan atau fluktuasi suhu dari hari ke hari. Dengan menggunakan LSTM, model dapat menangkap pola jangka panjang dalam data seri waktu yang sering kali menjadi indikator penting dalam prediksi potensi banjir.

Pendekatan hybrid yang menggabungkan CatBoost dan LSTM, bertujuan untuk mencapai beberapa hal penting. Pertama, dengan menggabungkan keunggulan dua model yang memiliki spesialisasi berbeda, saya berharap dapat meningkatkan akurasi prediksi. CatBoost membantu dengan memilih variabel penting, sedangkan LSTM menangani pola

temporal, sehingga keduanya berkontribusi pada prediksi yang lebih baik dan lebih andal. Selain itu, pendekatan ini diharapkan dapat mengurangi risiko overfitting, yang sering kali menjadi tantangan dalam model prediksi cuaca. Dengan CatBoost sebagai penyaring variabel, saya bisa memastikan bahwa model hanya berfokus pada variabel yang relevan dan tidak terganggu oleh data yang kurang signifikan.

Manfaat lain dari pendekatan ini adalah efisiensi dalam menangani jenis data yang beragam—baik data kategori maupun data seri waktu. Dalam data cuaca, kita tidak hanya dihadapkan pada angka seperti suhu atau kelembapan, tetapi juga kategori seperti musim atau klasifikasi cuaca. Pendekatan hybrid ini memungkinkan kedua jenis data tersebut diproses secara optimal, dengan CatBoost menangani variabel kategori dan LSTM menangani variabel seri waktu.

Keputusan untuk memilih CatBoost dan LSTM tidak hanya didasarkan pada kekuatan masing-masing algoritma, tetapi juga pada kebutuhan khusus dalam penelitian ini. Dalam data meteorologi, variabel kategori memainkan peran penting, dan CatBoost memungkinkan pemrosesan data ini tanpa kebutuhan untuk transformasi yang kompleks. Sementara itu, kemampuan LSTM dalam menangkap pola temporal memungkinkan model untuk mengenali pola perubahan dalam data cuaca dari waktu ke waktu, yang merupakan esensi dalam prediksi cuaca. Peneliti melakukan eksperimen pada dua dataset meteorologi yang tersedia untuk umum, yang mengandung nilai suhu udara dan curah hujan, untuk memvalidasi arsitektur yang peneliti usulkan.

Eksperimen dilaksanakan menggunakan Google Colab dengan backend Google Compute Engine dan Python 3. Resource yang digunakan selama eksperimen berada dalam rentang waktu dari 22.34 hingga 22.36, dengan detail penggunaan sumber daya sebagai berikut: RAM Sistem sebesar 0.9 / 12.7 GB, RAM GPU, dan penggunaan Disk sebesar 24.4 / 78.2 GB.

3.2. Sumber Data

Sumber data merupakan aspek fundamental dalam penelitian ini, khususnya untuk membangun model prediksi potensi banjir yang akurat untuk wilayah Kota Sleman. Penelitian ini menggunakan data meteorologi historis yang mencakup variabel seperti suhu udara, curah hujan, kelembapan udara, dan tekanan udara. Data-data ini diperoleh dari sumber terpercaya untuk memastikan kualitas dan validitasnya, sehingga hasil

prediksi yang dihasilkan oleh model hybrid CatBoost-LSTM dapat diandalkan dalam mendukung sistem peringatan dini bencana hidrometeorologi.

A. Jenis Data yang Digunakan

Penelitian ini memanfaatkan beberapa variabel cuaca yang dianggap signifikan dalam memprediksi fenomena hidrometeorologi ekstrem, khususnya yang berkaitan dengan potensi banjir. Variabel-variabel ini meliputi:

- **Suhu Udara (Temperature):** Data suhu harian digunakan untuk mengamati fluktuasi suhu yang mungkin berdampak pada cuaca ekstrem.
- **Curah Hujan (Precipitation):** Curah hujan harian menjadi indikator utama dalam memprediksi potensi banjir, karena peningkatan jumlah hujan biasanya berkaitan langsung dengan risiko banjir.
- **Kelembapan Udara (Humidity):** Kelembapan relatif di atmosfer memainkan peran penting dalam menentukan potensi terjadinya presipitasi, yang berpotensi menyebabkan cuaca buruk.
- **Tekanan Udara (Air Pressure):** Tekanan udara sering digunakan sebagai penanda sistem tekanan rendah, yang biasanya diikuti dengan cuaca ekstrem seperti badai atau hujan deras. Untuk detail jenis data lebih lengkapnya bisa dilihat pada gambar 3.1 Data Cuaca dari BMKG.

	Tanggal	Tn	Tx	Tavg	RH_avg	RR	ss	ff_x	ddd_x	ff_avg	ddd_car
0	01-01-2020	24,1	31,2	25,7	91,0	2,4	4,8	2,0	20,0	1,0	S
1	02-01-2020	23,4	31	26,7	85,0	29,9	1,9	2,0	200,0	1,0	C
2	03-01-2020	23,8	31,4	26,5	86,0	0,2	2,2	3,0	200,0	1,0	W
3	04-01-2020	23,8	30,2	26,8	87,0	87,2	2,1	4,0	220,0	2,0	S
4	05-01-2020	24,4	28	25,3	96,0	3	0,1	3,0	200,0	1,0	C
5	06-01-2020	23,4	31,4	26,7	88,0	11,4	0	2,0	200,0	2,0	S
6	07-01-2020	23,6	31,4	25,8	92,0	42,8	4,6	2,0	300,0	1,0	C
7	08-01-2020	24,2	30	25,8	92,0	0,6	0,8	4,0	180,0	1,0	S
8	09-01-2020	24	31	27	86,0	6	1,4	2,0	200,0	1,0	S
9	10-01-2020	23	30,8	25,8	88,0	33,2	5,4	3,0	180,0	1,0	C
10	11-01-2020	23,4	30	26,3	88,0	35,2	1,6	2,0	200,0	2,0	S

Gambar 3.1 Data Cuaca dari BMKG

Selain itu, terdapat **kolom** `status_banjir` dalam data bencana yang digunakan untuk menandai apakah suatu kejadian terkait dengan banjir atau tidak. Variabel ini bersifat kategorikal dengan dua nilai:

- **Ya:** Menandakan bahwa kejadian tersebut adalah banjir.
- **Tidak:** Menandakan bahwa kejadian tersebut bukan banjir.

Kolom `status_banjir` sangat penting dalam analisis ini karena berfungsi sebagai variabel target dalam model prediksi, di mana model akan belajar mengidentifikasi kondisi cuaca yang berpotensi menyebabkan banjir.

B. Deskripsi Data

Pada bagian ini, dijelaskan mengenai variabel-variabel yang terdapat dalam dataset yang digunakan untuk penelitian ini. Data yang digunakan merupakan data meteorologi harian yang mencakup berbagai parameter cuaca yang relevan untuk prediksi potensi banjir. Berikut adalah keterangan mengenai masing-masing variabel dan kode data yang ada:

- 8888: Kode ini digunakan untuk data yang tidak terukur.
- 9999: Menunjukkan bahwa pengukuran tidak dilakukan atau tidak ada data yang tersedia.
- `Tn` (Temperatur minimum): Mengukur suhu terendah pada hari tertentu, dalam satuan derajat Celsius ($^{\circ}\text{C}$).
- `Tx` (Temperatur maksimum): Mengukur suhu tertinggi pada hari tertentu, dalam satuan derajat Celsius ($^{\circ}\text{C}$).
- `Tavg` (Temperatur rata-rata): Mengukur rata-rata suhu harian, dalam satuan derajat Celsius ($^{\circ}\text{C}$).
- `RH_avg` (Kelembapan rata-rata): Mengukur kelembapan rata-rata harian, dalam satuan persen (%).
- `RR` (Curah hujan): Mengukur total curah hujan harian, dalam satuan milimeter (mm).
- `ss` (Lamanya penyinaran matahari): Mengukur durasi penyinaran matahari dalam satu hari, dalam satuan jam.
- `ddd_x` (Arah angin saat kecepatan maksimum): Menunjukkan arah angin pada saat kecepatan angin mencapai nilai maksimum, dalam derajat ($^{\circ}$).

- **ddd_car** (Arah angin terbanyak): Menunjukkan arah angin yang paling dominan dalam satu hari, dalam derajat ($^{\circ}$).
- **ff_x** (Kecepatan angin maksimum): Mengukur kecepatan angin maksimum yang terjadi pada hari tertentu, dalam meter per detik (m/s).
- **ff_avg** (Kecepatan angin rata-rata): Mengukur rata-rata kecepatan angin harian, dalam meter per detik (m/s).

C. Sumber Data

Data meteorologi yang digunakan dalam penelitian ini diperoleh dari dua sumber utama:

1. **Balai Besar Meteorologi, Klimatologi, dan Geofisika (BBMKG) Wilayah I Sleman**: Sumber utama data cuaca untuk wilayah Sleman, meliputi suhu udara, curah hujan, kelembapan, dan tekanan udara, yang didapatkan secara harian.
2. **Geoportal Data Bencana Indonesia BNPB**: Dataset ini menyediakan informasi mengenai kejadian banjir di wilayah Indonesia, termasuk Kota Sleman. Informasi tersebut mencakup data lokasi, tanggal kejadian, dan dampak dari setiap bencana, sehingga memungkinkan model untuk memahami korelasi antara kondisi cuaca dan kejadian banjir. Sumber data yang dipilih memiliki cakupan temporal harian dan kualitas yang memadai untuk menganalisis pola cuaca yang berpotensi menyebabkan banjir di Kota Sleman. Contoh isi data bisa dilihat pada gambar 3.2 Data Bencana dari BNPB.

tertera pada gambar. Fokus penelitian pada wilayah Sleman diharapkan dapat membantu dalam membangun model prediksi yang relevan dan tepat sasaran untuk sistem peringatan dini di daerah tropis dengan karakteristik cuaca yang dinamis.



Gambar 1.1 Tren kejadian bencana alam di Indonesia dalam 10 tahun terakhir periode 2011 -2020, (BNPB, 2020).

E. Pra-pemrosesan Data

Agar data siap digunakan dalam model, beberapa langkah pra-pemrosesan dilakukan sebagai berikut:

- **Penanganan Missing Values:** Dalam data cuaca yang diperoleh, nilai 8888 menunjukkan data yang tidak terukur, sementara 9999 berarti tidak ada data yang tersedia. Nilai-nilai ini diubah menjadi NaN untuk menandakan missing values, kemudian diisi menggunakan metode interpolasi atau estimasi nilai median untuk menjaga integritas dataset. Peneliti mengecek terlebih dahulu persentase missing values yang ada dimasing-masing kolom dengan bantuan kode python dengan hasil sebagaimana ditunjukkan pada gambar 3.5 Persentase Missing Values pada dataset.

```

Persentase Missing Values pada Masing-masing Kolom:
Tanggal      0.000000
Tn           3.048416
Tx           2.988643
Tavg        1.733413
RH_avg      1.733413
RR          16.078900
ss          2.869097
ff_x        1.613867
ddd_x       1.613867
ff_avg      1.613867
ddd_car     1.613867
Year        0.000000
Month       0.000000
dtype: float64

```

Gambar 3.5 Persentase Missing Values pada dataset

Setelah ditemukan variabel dalam dataset yang mengandung nilai yang hilang. Untuk menangani hal ini, dilakukan imputasi nilai hilang menggunakan nilai median dari masing-masing variabel. Pendekatan ini dipilih karena median tidak dipengaruhi oleh outlier, sehingga memberikan estimasi yang lebih robust untuk nilai yang hilang. Hasil persentasenya turun signifikan seperti yang bisa kita lihat pada gambar 3.6 Persentase Missing Values setelah imputasi median.

	Missing Values	% of Missing Values
tanggal	0	0.00
curah_hujan	269	16.08
suhu_udara	29	1.73
kelembapan_udara	0	0.00
status_banjir	0	0.00

Gambar 3.6 Persentase Missing Values setelah imputasi median

- **Normalisasi Data:** Variabel numerik seperti suhu, curah hujan, dan kelembapan dinormalisasi dalam rentang tertentu (misalnya 0 hingga 1) untuk memastikan bahwa variabel dengan skala yang lebih besar tidak mendominasi perhitungan dalam model LSTM.

- **Transformasi Variabel Kategori:** Variabel kategori, seperti status banjir, dikonversi menjadi format biner dengan LabelEncoder, di mana “Tidak Banjir” diberi nilai 0 dan “Banjir” diberi nilai 1 sebagaimana terlihat pada gambar 3.7 Konversi nilai pada kolom status banjir. Metode ini digunakan karena model CatBoost dapat menangani variabel kategori tanpa memerlukan one-hot encoding, sehingga lebih efisien.

	tanggal	curah_hujan	suhu_udara	kelembapan_udara	status_banjir
0	2020-01-01	2,4	25,7	91,00	Tidak
1	2020-01-02	29,9	26,7	85,00	Tidak
2	2020-01-03	9,2	26,5	86,00	Tidak
3	2020-01-04	67,2	26,8	87,00	Tidak
4	2020-01-05	3	25,3	96,00	Tidak

Gambar 3.7 Konversi nilai pada kolom status banjir

- **Pengecekan Duplikasi:** Data duplikat diperiksa untuk memastikan tidak ada entri yang tercatat lebih dari satu kali. Pengecekan ini penting untuk menjaga keakuratan model.

- **Deteksi dan Penanganan Outliers:** Outliers diidentifikasi dengan menggunakan metode Interquartile Range (IQR) dan ditangani sesuai kebutuhan, agar tidak mengganggu analisis. Pengecekan ini divisualisasikan menggunakan boxplot, yang memperlihatkan adanya beberapa titik outlier yang signifikan. Untuk mengatasi outlier ini, digunakan metode Interquartile Range (IQR). Dengan metode ini, nilai-nilai yang jatuh di luar rentang yang wajar ($Q1 - 1.5 \cdot IQR$ dan $Q3 + 1.5 \cdot IQR$) dianggap sebagai outlier dan ditangani sesuai kebutuhan untuk mengurangi dampak negatifnya terhadap analisis. Untuk pemodelan time series, data yang mengandung outlier dapat mempengaruhi hasil analisis

dan prediksi. Oleh karena itu, data yang mengandung outlier akan diubah menjadi nilai NaN seperti yang terlihat pada gambar 3.8 Persentase Missing Values setelah imputasi median

	tanggal	curah_hujan	suhu_udara	kelembapan_udara
0	2020-01-01	NaN	NaN	91.00
1	2020-01-02	NaN	NaN	85.00
2	2020-01-03	NaN	NaN	86.00
3	2020-01-04	NaN	NaN	87.00
4	2020-01-05	NaN	NaN	96.00

Gambar 3.8 Persentase Missing Values setelah imputasi median

F. Manajemen dan Penyimpanan Data

Dataset disimpan dalam format CSV agar mudah diakses dan diolah selama proses pemodelan. Format ini dipilih karena efisien dalam memfasilitasi manipulasi data besar, khususnya selama tahap pra-pemrosesan dan pelatihan model. Data yang telah diproses juga disimpan dalam format yang terpisah untuk memudahkan akses dalam eksperimen lanjutan atau jika diperlukan untuk pengujian ulang.

Penyimpanan yang terstruktur ini memungkinkan penelitian berjalan lebih efisien, serta memungkinkan penggunaan kembali data tanpa harus melakukan pra-pemrosesan dari awal. Manajemen data yang baik juga memastikan bahwa data tersedia untuk diakses selama evaluasi dan validasi model, yang menjadi bagian penting dalam menjaga kualitas dan keandalan hasil penelitian ini.

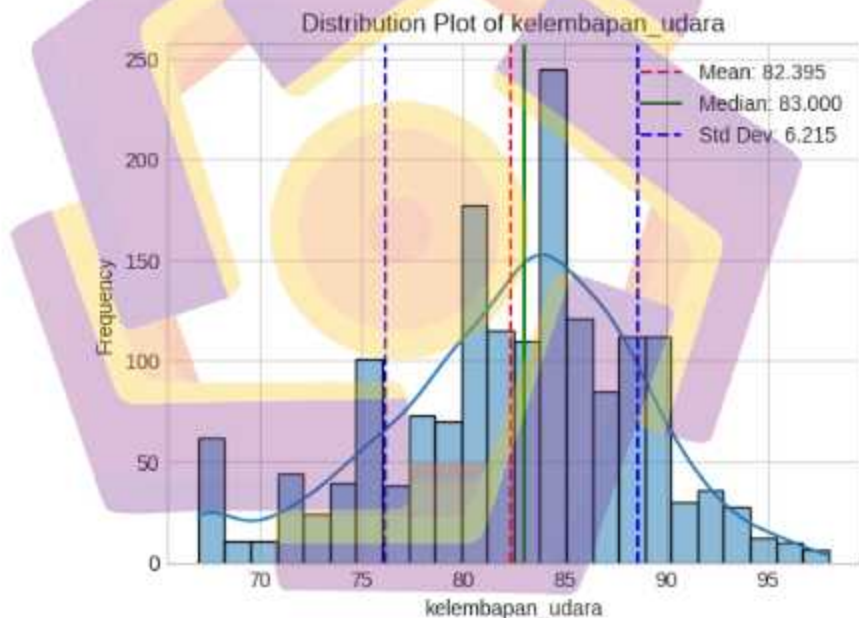
3.3. Exploratory Data Analysis (EDA)

Tahap Exploratory Data Analysis (EDA) merupakan langkah awal yang esensial dalam proses analisis data. Pada tahap ini, data dijelajahi secara mendalam untuk memahami distribusi, karakteristik, serta hubungan antarvariabel yang relevan. EDA membantu mengidentifikasi pola umum dan potensi anomali dalam dataset, serta memberikan insight awal mengenai kemungkinan peran setiap variabel dalam prediksi.

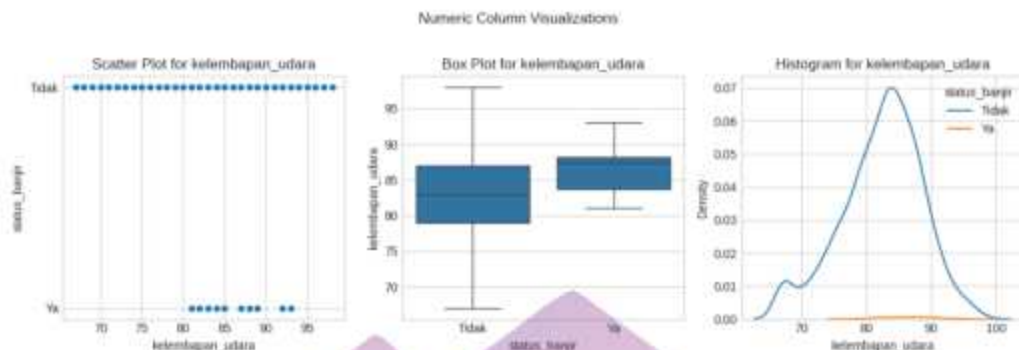
Pada penelitian ini, EDA dilakukan melalui pendekatan analisis numerikal dan kategorikal untuk mengidentifikasi keterkaitan variabel-variabel cuaca dengan kejadian banjir.

Analisis Numerikal

Analisis numerikal mencakup pemeriksaan distribusi dari variabel-variabel numerik yang ada dalam dataset, seperti curah hujan (RR), suhu udara rata-rata (Tavg), dan kelembapan udara rata-rata (RH_avg). Dengan analisis ini, kita dapat melihat apakah data memiliki distribusi normal atau condong, serta memahami variasi dalam data yang mungkin mempengaruhi akurasi prediksi. Distribution Plot dari kelembapan udara bisa kita lihat ada gambar 3.9 dan untuk tiap variabel penting lainnya pada gambar 3.10.



Gambar 3.9 Distribution Plot pada kelembapan udara



Gambar 3.10 Distribution Plot pada variabel penting lainnya

Pada Gambar diatas, ditampilkan distribusi curah hujan, suhu udara rata-rata, dan kelembapan udara rata-rata. Gambar tersebut menunjukkan beberapa informasi penting:

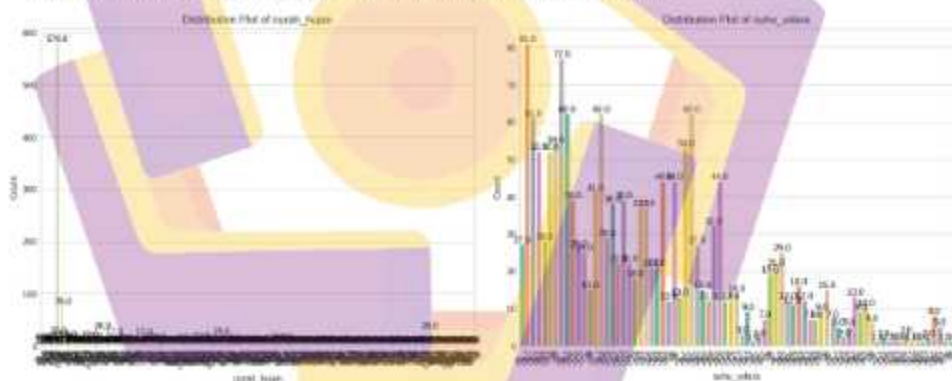
- 1. Curah Hujan (RR):** Distribusi variabel curah hujan menunjukkan pola yang sangat condong ke kiri (left-skewed), yang menandakan bahwa sebagian besar hari memiliki curah hujan yang rendah, sedangkan hari dengan curah hujan tinggi lebih jarang terjadi. Distribusi yang tidak normal ini mengindikasikan kemungkinan adanya beberapa outlier atau nilai ekstrem, yang bisa berpengaruh signifikan terhadap hasil analisis. Dalam konteks prediksi banjir, curah hujan merupakan variabel kunci yang akan diuji korelasinya dengan kejadian banjir.
- 2. Suhu Udara Rata-Rata (Tavg):** Variabel suhu udara rata-rata memperlihatkan distribusi yang lebih mendekati normal, yang berarti bahwa fluktuasi suhu berada dalam rentang yang relatif konsisten setiap harinya. Distribusi yang mendekati normal ini penting dalam menjaga stabilitas model prediksi, karena variabel yang berdistribusi normal cenderung menghasilkan model yang lebih baik dalam konteks machine learning.
- 3. Kelembapan Udara Rata-Rata (RH_avg):** Variabel kelembapan udara rata-rata juga menunjukkan distribusi yang hampir normal. Kelembapan udara merupakan faktor penting dalam prediksi banjir, karena tingginya tingkat kelembapan sering kali menjadi indikator presipitasi. Konsistensi dalam distribusi variabel ini menunjukkan bahwa data kelembapan memiliki fluktuasi yang tidak terlalu

ekstrem, sehingga variabel ini dapat digunakan dalam pemodelan tanpa banyak modifikasi.

Dengan analisis numerikal ini, peneliti dapat mengidentifikasi karakteristik dasar dari data numerik, serta menyesuaikan model agar lebih responsif terhadap distribusi dan pola data yang ada.

Analisis Kategorikal

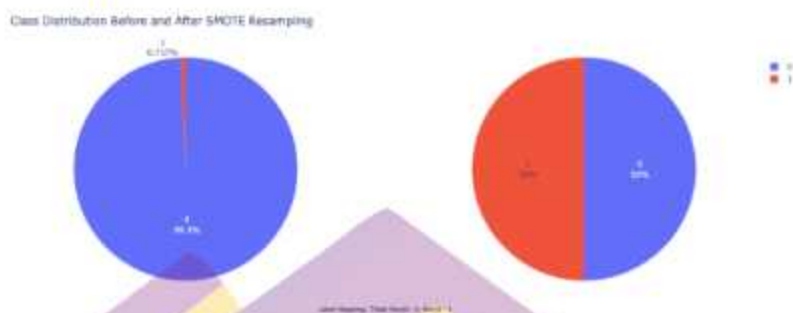
Analisis kategorikal difokuskan pada variabel `status_banjir`, yaitu variabel biner yang menunjukkan apakah hari tertentu mengalami banjir atau tidak. Distribusi variabel kategorikal yang tertera pada gambar 3.11 ini menunjukkan bahwa dataset mengalami ketidakseimbangan kelas (*class imbalance*), dengan proporsi hari yang tidak mengalami banjir jauh lebih besar dibandingkan dengan hari yang mengalami banjir. Hal ini menjadi tantangan dalam pemodelan, karena model yang dilatih pada data dengan ketidakseimbangan kelas cenderung bias terhadap kelas mayoritas.



Gambar 3.11 Persentase Missing Values setelah Imputasi median

Ketidakseimbangan kelas ini harus ditangani dengan teknik khusus, seperti penyeimbangan kelas menggunakan metode oversampling atau undersampling, untuk memastikan bahwa model tidak mengabaikan kejadian banjir sebagai kelas minoritas. Teknik seperti *Synthetic Minority Over-sampling Technique (SMOTE)* dapat digunakan untuk menghasilkan sampel tambahan pada kelas banjir, sehingga model dapat belajar dengan lebih baik dalam mendeteksi pola yang terkait dengan kejadian banjir. Hasil dari

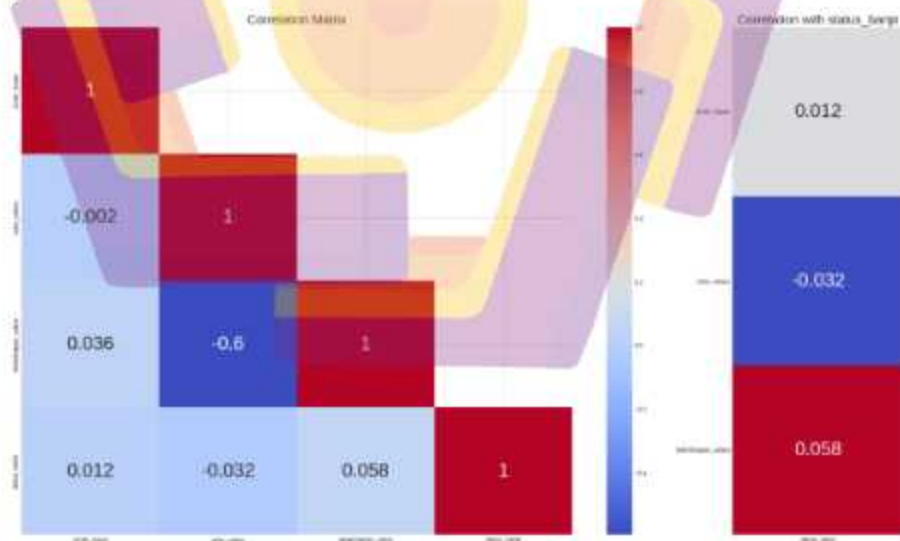
implementasi *SMOTE* ini bisa kita lihat pada gambar 3.12 Distribusi Kelas Sebelum dan Setelah Resampling *SMOTE*



Gambar 3.12 Distribusi Kelas Sebelum dan Setelah Resampling *SMOTE*

Korelasi Antarvariabel

Selanjutnya, dilakukan analisis korelasi untuk mengidentifikasi hubungan antara variabel-variabel numerik dan *status_banjir*. Hasil analisis korelasi menunjukkan pada gambar 3.13 bahwa:



Gambar 3.13 Correlation Matrix hubungan antara variabel dan status banjir

- **Curah Hujan** memiliki korelasi positif yang lemah dengan kejadian banjir. Hal ini mengindikasikan bahwa meskipun curah hujan berkontribusi terhadap kejadian banjir, pengaruhnya mungkin lebih kompleks dan melibatkan interaksi dengan variabel lain.
- **Suhu Udara** dan **Kelembapan Udara** memiliki korelasi negatif yang sangat lemah terhadap kejadian banjir, yang berarti pengaruh langsung dari suhu dan kelembapan terhadap banjir tidak begitu signifikan. Namun, variabel-variabel ini tetap dapat berperan sebagai faktor pendukung dalam pola cuaca yang memicu banjir.

Korelasi yang lemah antara variabel-variabel cuaca ini dan kejadian banjir menunjukkan kompleksitas dalam memodelkan banjir berdasarkan data cuaca. Oleh karena itu, model prediksi yang digunakan harus mampu menangkap pola non-linear dan interaksi antara variabel untuk meningkatkan akurasi prediksi.

Tahap EDA memberikan pemahaman mendalam tentang karakteristik data dan hubungan antarvariabel. Hasil analisis ini mengindikasikan bahwa model yang akan dibangun perlu memperhitungkan variabilitas tinggi dalam data curah hujan dan ketidakseimbangan kelas pada variabel status banjir. Selain itu, hasil korelasi menunjukkan bahwa model yang digunakan harus mampu mengatasi pola non-linear untuk mengenali interaksi kompleks antara variabel cuaca. Dengan demikian, EDA ini membantu dalam merancang pendekatan pemodelan yang lebih adaptif dan sesuai dengan karakteristik data yang dihadapi.

3.4. Implementasi Model

Langkah pertama dalam implementasi model ini adalah menggunakan CatBoost untuk mengidentifikasi variabel cuaca yang paling berpengaruh terhadap hasil prediksi. Setelah data melalui tahap pra-pemrosesan, CatBoost diterapkan pada dataset untuk menentukan variabel mana yang memiliki pengaruh signifikan terhadap kejadian cuaca ekstrem. Dengan menentukan tingkat kepentingan atau *feature importance* untuk setiap variabel, CatBoost membantu menyaring variabel yang kurang relevan, sehingga model LSTM dapat lebih fokus pada fitur-fitur penting yang benar-benar relevan dalam menentukan prediksi.

Setelah fitur penting dipilih, CatBoost dikonfigurasi dengan parameter yang optimal. Beberapa parameter utama seperti learning rate, kedalaman pohon (depth), dan jumlah iterasi disesuaikan untuk meningkatkan akurasi prediksi. Learning rate yang lebih rendah, misalnya, membuat pelatihan model lebih stabil, meskipun membutuhkan lebih banyak iterasi. Kedalaman pohon yang lebih besar memungkinkan model menangkap pola yang lebih kompleks, tetapi harus diatur dengan hati-hati untuk menghindari overfitting. Dengan melakukan penyesuaian ini, CatBoost menjadi lebih efisien dalam mengolah data kategori dan menghasilkan output yang optimal untuk tahap berikutnya.

Selanjutnya, model LSTM dirancang untuk memproses data seri waktu. Dalam penelitian ini, model LSTM dikonfigurasi dengan beberapa lapisan tersembunyi agar dapat menangkap pola temporal dalam data yang lebih kompleks. Jumlah lapisan LSTM serta unit tersembunyi dalam setiap lapisan ditentukan berdasarkan eksperimen yang bertujuan untuk mencapai keseimbangan antara kemampuan model dalam mempelajari pola jangka panjang dan kebutuhan untuk menjaga efisiensi komputasi. Untuk mengurangi risiko overfitting, digunakan dropout rate pada setiap lapisan LSTM, yang akan membuang beberapa neuron secara acak selama pelatihan. Selain itu, batch size dan jumlah epochs disesuaikan untuk memastikan bahwa pelatihan model berlangsung stabil dan menghasilkan prediksi yang akurat.

Pelatihan CatBoost dimulai terlebih dahulu, di mana model ini digunakan untuk melakukan seleksi fitur pada data kategori. Hasil dari tahap pelatihan ini adalah daftar fitur penting yang akan digunakan dalam pelatihan model LSTM pada tahap berikutnya. CatBoost dilatih hingga model ini dapat mengenali variabel-variabel kategori yang paling relevan dalam memengaruhi hasil prediksi potensi banjir. Setelah proses seleksi fitur selesai, data seri waktu yang berisi variabel terpilih kemudian diproses dengan LSTM. Model LSTM dilatih untuk mengenali pola temporal dalam data tersebut, dengan memanfaatkan variabel yang paling relevan seperti suhu, kelembapan, atau tekanan udara. Hal ini membantu LSTM memahami hubungan temporal antara variabel dan menghasilkan prediksi yang lebih baik.

Untuk mendapatkan hasil yang optimal, dilakukan tuning hyperparameter pada kedua model. Proses tuning dilakukan dengan menggunakan teknik grid search atau random search, yang memungkinkan eksperimen pada berbagai kombinasi parameter. Dalam tahap ini, parameter seperti learning rate, depth pada CatBoost, serta dropout rate

pada LSTM disesuaikan secara berulang untuk mendapatkan konfigurasi yang menghasilkan kinerja terbaik.

Setelah CatBoost dan LSTM selesai dilatih, hasil dari kedua model ini disatukan untuk menghasilkan prediksi akhir. Penyatuan hasil ini dapat dilakukan dengan metode ensemble, seperti voting atau averaging, di mana prediksi dari kedua model digabungkan untuk menghasilkan prediksi yang lebih akurat. Alternatif lain yang sering digunakan adalah dengan menambahkan layer tambahan dalam jaringan saraf yang bertugas mengintegrasikan hasil dari CatBoost dan LSTM sebelum menghasilkan prediksi akhir. Metode ini memungkinkan penyatuan output secara lebih mendalam, dengan mempertimbangkan kelebihan masing-masing model.

Pada tahap akhir implementasi, model hybrid ini diuji pada data pengujian untuk mengevaluasi performanya. Pengujian dilakukan pada data yang belum pernah dilihat oleh model sebelumnya untuk memastikan kemampuan generalisasi model dalam prediksi potensi banjir. Selain itu, model ini juga diuji dengan data real-time untuk menilai keandalannya dalam aplikasi praktis, seperti dalam sistem peringatan dini cuaca. Pengujian ini membantu dalam mengevaluasi seberapa cepat dan akurat model hybrid ini dapat memproses data aktual dan memberikan prediksi tepat waktu.

Setelah pengujian selesai, model yang sudah terlatih disimpan untuk digunakan kembali atau diintegrasikan dalam sistem lain. Penyimpanan model dilakukan dalam format yang mendukung pengoperasian ulang, seperti HDF5 atau pickle, sehingga model dapat dimuat ulang tanpa kehilangan informasi yang telah dipelajari. Selain itu, seluruh proses implementasi, mulai dari pra-pemrosesan data hingga pelatihan dan pengujian, didokumentasikan secara terperinci untuk memastikan bahwa semua tahapan dapat direplikasi di masa mendatang atau digunakan sebagai referensi jika perlu dilakukan pengembangan lebih lanjut. Dokumentasi ini mencakup catatan mengenai parameter yang digunakan, hasil tuning, serta performa model pada data pengujian, yang penting untuk keperluan pengembangan model dan analisis di masa depan.

3.5. Evaluasi Model

Setelah model hybrid CatBoost-LSTM selesai dilatih dan diimplementasikan, tahap evaluasi dilakukan untuk mengukur seberapa baik model ini dalam memprediksi potensi banjir. Evaluasi model sangat penting dalam penelitian ini karena bertujuan untuk memastikan bahwa model tidak hanya memiliki akurasi tinggi pada data pelatihan, tetapi

Alur penelitian ini menjelaskan tahapan-tahapan yang ditempuh dalam pengembangan dan implementasi model hybrid CatBoost-LSTM untuk prediksi potensi banjir. Setiap tahap dilakukan secara berurutan untuk memastikan bahwa proses penelitian berjalan dengan baik dan setiap langkah memiliki hasil yang jelas sebagai dasar untuk tahap berikutnya.

Penelitian ini diawali dengan pengumpulan data historis cuaca dari sumber yang terpercaya, seperti lembaga meteorologi nasional atau internasional. Data ini mencakup variabel penting seperti suhu, curah hujan, kelembapan, dan tekanan udara, yang merupakan indikator cuaca ekstrem. Setelah data diperoleh, tahap pertama adalah melakukan pra-pemrosesan untuk memastikan bahwa data dalam kondisi bersih dan siap digunakan. Proses pra-pemrosesan ini mencakup penanganan data yang hilang, normalisasi variabel, dan transformasi data kategori menjadi format yang sesuai untuk pemrosesan model CatBoost. Pada tahap ini, variabel kategori seperti klasifikasi musim atau lokasi geografis diproses menggunakan metode yang efisien, sehingga data kategori dapat digunakan dengan baik dalam model pembelajaran mesin.

Setelah data selesai diproses, langkah selanjutnya adalah pengolahan data menggunakan model CatBoost untuk melakukan seleksi fitur. CatBoost berperan dalam mengidentifikasi fitur-fitur yang paling relevan dalam mempengaruhi prediksi potensi banjir. Melalui analisis *feature importance*, variabel-variabel yang tidak signifikan dapat diabaikan, sehingga hanya variabel yang memiliki kontribusi besar terhadap hasil prediksi yang diteruskan ke tahap berikutnya. Seleksi fitur ini membantu menyederhanakan data yang diproses oleh model LSTM dan meningkatkan efisiensi serta akurasi model secara keseluruhan.

Langkah berikutnya adalah pelatihan model LSTM dengan menggunakan data seri waktu yang telah disaring oleh CatBoost. LSTM dirancang untuk mempelajari pola temporal dalam data, yang sangat penting dalam prediksi cuaca karena kondisi cuaca saat ini sering kali bergantung pada kondisi sebelumnya. Pelatihan model LSTM dilakukan dengan menyesuaikan parameter seperti jumlah lapisan LSTM, unit tersembunyi, dan dropout rate untuk menghindari overfitting. Pada tahap ini, teknik sliding window diterapkan untuk memastikan bahwa model dapat belajar dari urutan data sebelumnya untuk memprediksi cuaca di masa depan.

Setelah pelatihan model CatBoost dan LSTM selesai, kedua model digabungkan untuk menghasilkan prediksi akhir. Penggabungan ini dapat dilakukan melalui metode

ensemble atau dengan menambahkan layer tambahan dalam jaringan saraf, yang bertugas mengintegrasikan hasil prediksi dari kedua model. Langkah ini bertujuan untuk menggabungkan keunggulan CatBoost dalam menangani data kategori dan LSTM dalam mempelajari ketergantungan temporal dalam data seri waktu, sehingga menghasilkan prediksi yang lebih akurat.

Model yang telah digabungkan kemudian dievaluasi menggunakan data pengujian untuk mengukur performa model secara keseluruhan. Evaluasi dilakukan menggunakan metrik seperti RMSE dan MAE, yang memberikan gambaran tentang seberapa akurat model dalam memprediksi potensi banjir. Selain itu, k-fold cross-validation diterapkan untuk menguji konsistensi model pada berbagai subset data dan menghindari overfitting. Setelah evaluasi selesai, model hybrid ini juga dibandingkan dengan model baseline, seperti model regresi linear atau RNN konvensional, untuk menilai sejauh mana pendekatan hybrid ini memberikan peningkatan akurasi.

Secara keseluruhan, alur penelitian ini dapat dirangkum dalam diagram alir yang menggambarkan urutan setiap tahapan mulai dari pengumpulan data, pra-pemrosesan, seleksi fitur dengan CatBoost, pelatihan LSTM, integrasi hasil, evaluasi, hingga pengujian dengan data real-time. Tabel dibawah ini memberikan gambaran visual tentang jalur yang ditempuh dalam penelitian, sehingga memudahkan pemahaman tentang bagaimana setiap tahap terhubung dan berkontribusi terhadap hasil akhir.

Tabel 3.1 Jadwal Penelitian

No	Kegiatan	Jan	Feb	Mar	Apr	Mei	Jun	Jul	Aug	Sep
1	Pengumpulan Data	✓								
	a. Pengumpulan Data Cuaca Historis	✓								
	b. Pengecekan Kelengkapan dan Kualitas Data	✓								
2	Pra-pemrosesan Data		✓							
	a. Pembersihan Data dan Penanganan Data Hilang		✓							
	b. Normalisasi dan Transformasi Data Kategori		✓							
3	Pengembangan Model CatBoost dan Seleksi Fitur			✓						
	a. Pengembangan dan Pelatihan Model CatBoost			✓						

	b. Seleksi Fitur Berdasarkan Feature Importance			✓					
4	Pengembangan Model LSTM				✓				
	a. Pelatihan Model LSTM dengan Data Seri Waktu				✓				
	b. Penyesuaian Hyperparameter Model LSTM				✓				
5	Penggabungan Hasil dan Evaluasi Model					✓			
	a. Penggabungan Hasil Prediksi dari CatBoost dan LSTM					✓			
	b. Evaluasi Model dengan Data Pengujian dan Cross-Validation					✓			
	c. Perbandingan dengan Model Baseline					✓			
6	Pengujian Real-Time dan Penulisan Laporan Akhir						✓		
	a. Pengujian Model dengan Data Real-Time						✓		
	b. Penulisan Laporan Akhir dan Analisis Hasil						✓	✓	✓

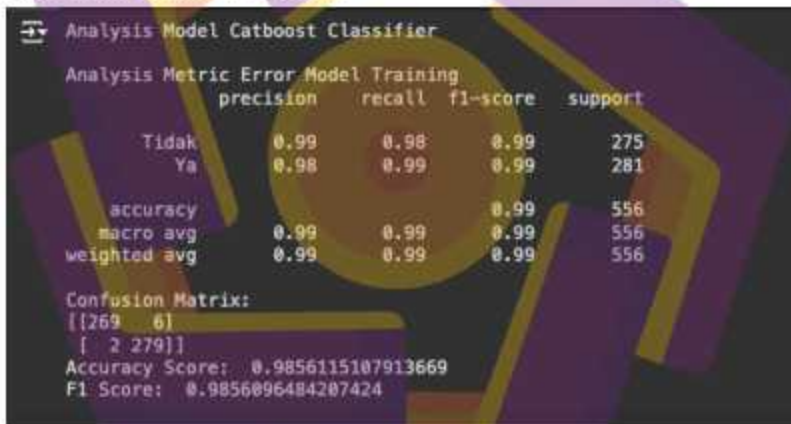
Dengan alur penelitian yang terstruktur seperti yang tertera pada tabel 3.1, penelitian ini memastikan bahwa setiap tahapan dilaksanakan dengan cermat dan mendukung pencapaian tujuan utama, yaitu mengembangkan model prediksi potensi banjir yang akurat dan andal untuk sistem peringatan dini. Dokumentasi setiap langkah dalam alur penelitian ini juga penting untuk memastikan bahwa penelitian ini dapat direplikasi dan digunakan sebagai dasar bagi penelitian lanjutan di masa depan. Selama proses ini, adaptasi dan fleksibilitas dalam pendekatan menjadi kunci untuk mengidentifikasi solusi terbaik. Dari menguji variasi parameter sederhana hingga melaksanakan eksperimen yang kompleks, langkah-langkah ini direfleksikan dalam sebuah alur penelitian yang terstruktur, memberikan landasan yang kokoh untuk penelitian ilmiah yang dapat dipertanggungjawabkan.

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1. Evaluasi Model

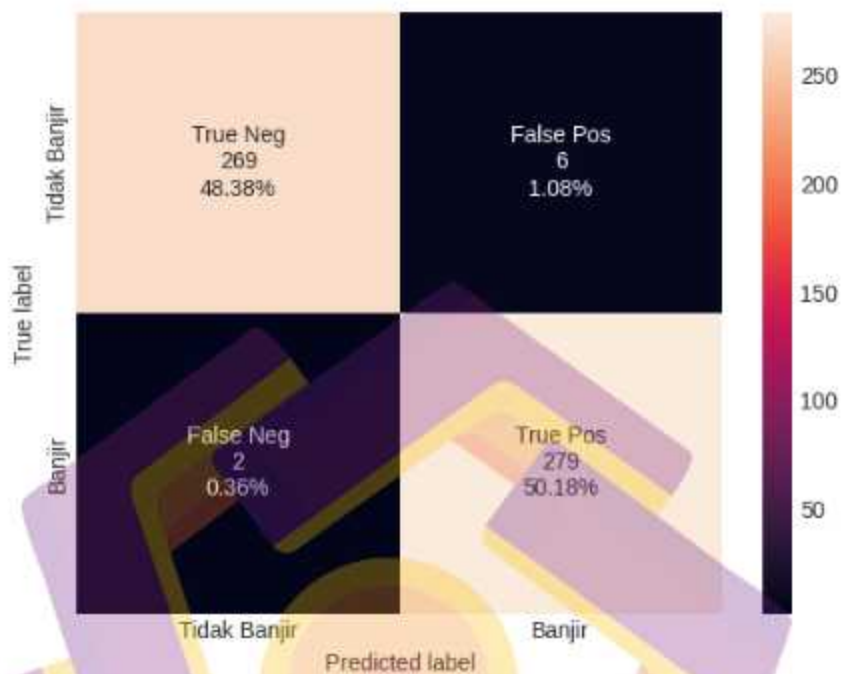
Tahap evaluasi model ini melibatkan penilaian performa model *CatBoost* dan *LSTM* yang telah dikembangkan untuk prediksi potensi banjir dan cuaca di Kota Sleman. Model *CatBoost* dievaluasi menggunakan metrik *Accuracy Score*, *Precision*, *Recall*, dan *F1 Score* yang dapat kita lihat pada gambar 4.1 dan visualisasinya pada gambar 4.3 dan 4.5. Sebelum dilakukan tuning hyperparameter, model menunjukkan akurasi sebesar 98%, yang menunjukkan kemampuan model dalam memprediksi kejadian banjir dengan baik. Selain itu, evaluasi *feature importance* menunjukkan bahwa kelembapan udara adalah variabel yang paling berpengaruh dalam prediksi banjir, meskipun curah hujan juga memiliki kontribusi signifikan.



```
Analysis Model Catboost Classifier
Analysis Metric Error Model Training
precision recall f1-score support
Tidak 0.99 0.98 0.99 275
Ya 0.98 0.99 0.99 281
accuracy 0.99 556
macro avg 0.99 0.99 0.99 556
weighted avg 0.99 0.99 0.99 556

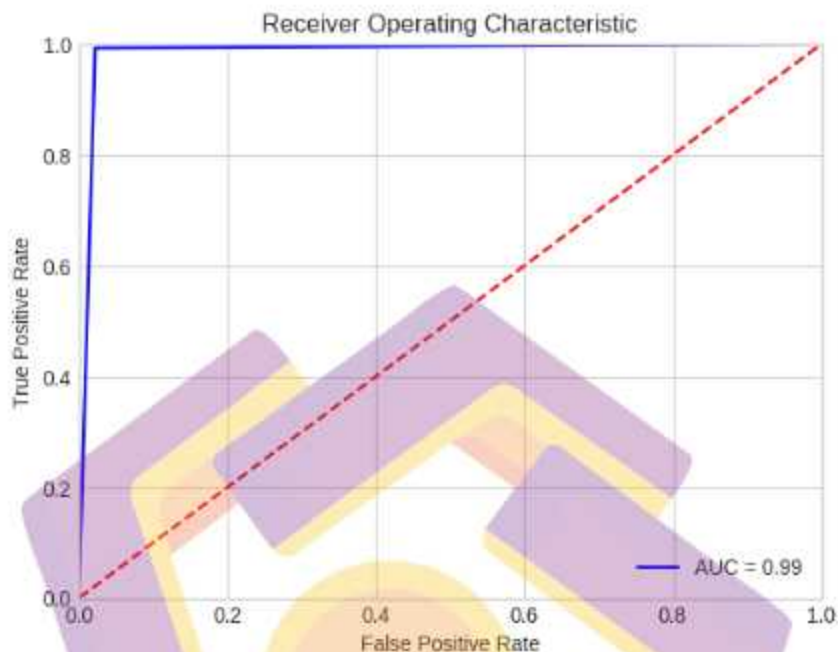
Confusion Matrix:
[[269 6]
 [ 2 279]]
Accuracy Score: 0.9856115107913669
F1 Score: 0.9856096484207424
```

Gambar 4.1 Hasil Analysis Model Catboost Classifier



Gambar 4.2 Confusion Matrix Model Catboost Classifier

Confusion matrix pada model klasifikasi ini menunjukkan performa yang cukup baik dalam mengklasifikasikan data “Banjir” dan “Tidak Banjir.” Berdasarkan matriks, model berhasil mengidentifikasi 269 data “Tidak Banjir” secara benar, yang dikenal sebagai True Negative (48.38% dari keseluruhan prediksi). Selain itu, model juga mengidentifikasi dengan benar 279 data “Banjir,” atau True Positive, yang mencakup 50.18% dari prediksi. Di sisi lain, terdapat beberapa kesalahan dalam prediksi, yaitu sebanyak 6 kali model memprediksi “Banjir” padahal sebenarnya “Tidak Banjir” (False Positive) dengan tingkat kesalahan 1.08%, dan 2 kali model memprediksi “Tidak Banjir” padahal sebenarnya “Banjir” (False Negative) dengan tingkat kesalahan 0.36%. Hasil ini menunjukkan bahwa model memiliki tingkat akurasi yang tinggi, karena kesalahan dalam prediksi False Positive dan False Negative cukup rendah, sementara proporsi True Positive dan True Negative mendominasi hasil klasifikasi.



Gambar 4.3 Kurva ROC Model Catboost Classifier

Pada tahap ini, model CatBoost dan LSTM dievaluasi untuk menilai kinerja dalam memprediksi kejadian banjir di Kota Sleman. Evaluasi model dilakukan melalui beberapa metrik, seperti *Accuracy Score*, *Precision*, *Recall*, *F1 Score*, dan *AUC Score*. Berikut adalah gambar hasil evaluasi model CatBoost sebelum dan sesudah tuning hyperparameter:

```

Fitting 3 folds for each of 54 candidates, totalling 162 fits
Best score: 0.999415398787899
Best param: {'depth': 6, 'eval_metric': 'AUC', 'iterations': 2000, 'l2_leaf_reg':
Best model: <catboost.core.CatBoostClassifier object at 0x7b678dcef0d8>
Analysis Model CatBoost GridSearch

Analysis Metric Error Model Training
precision recall f1-score support
    Tidak    1.00    1.00    1.00    275
     Ya     1.00    1.00    1.00    281

accuracy          1.00    556
macro avg         1.00    1.00    1.00    556
weighted avg      1.00    1.00    1.00    556

Confusion Matrix:
[[274  1]
 [ 1 280]]
Accuracy Score: 0.9964028776978417
F1 Score: 0.9964028776978417

```

Gambar 4.4 Hasil Analisis Model Catboost Classifier dengan GridSearch (Hyperparameter Tuning)

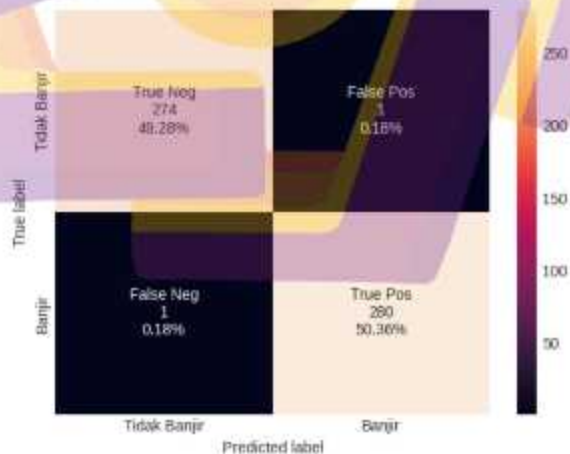
Accuracy Score: Sebelum tuning: 0.985, Sesudah tuning: 0.996

Precision: Sebelum tuning: 0.985, Sesudah tuning: 0.996

Recall: Sebelum tuning: 0.985, Sesudah tuning: 0.996

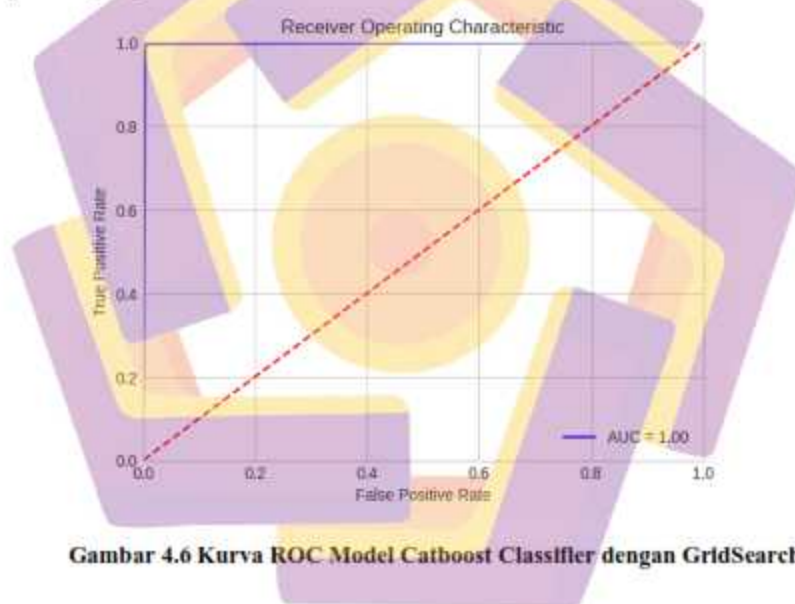
F1 Score: Sebelum tuning: 0.985, Sesudah tuning: 0.996

AUC Score: Sesudah tuning: 0.996



Gambar 4.5. Confusion Matrix Model Catboost Classifier dengan GridSearch

Confusion matrix pada model klasifikasi ini menunjukkan performa yang sangat baik dalam memprediksi data “Banjir” dan “Tidak Banjir.” Model berhasil mengklasifikasikan 274 data “Tidak Banjir” dengan benar, yang disebut sebagai True Negative, mewakili 49.28% dari total prediksi. Selain itu, model juga berhasil mengidentifikasi 280 data “Banjir” dengan benar, yang disebut True Positive, mencakup 50.36% dari total prediksi. Kesalahan prediksi sangat kecil, yaitu hanya 1 kali untuk False Positive (di mana model memprediksi “Banjir” padahal seharusnya “Tidak Banjir”) dan 1 kali untuk False Negative (di mana model memprediksi “Tidak Banjir” padahal seharusnya “Banjir”), masing-masing dengan persentase kesalahan sebesar 0.18%. Dengan demikian, model ini memiliki tingkat akurasi yang sangat tinggi, dengan proporsi prediksi yang benar mendominasi hasil klasifikasi dan hanya sedikit kesalahan.



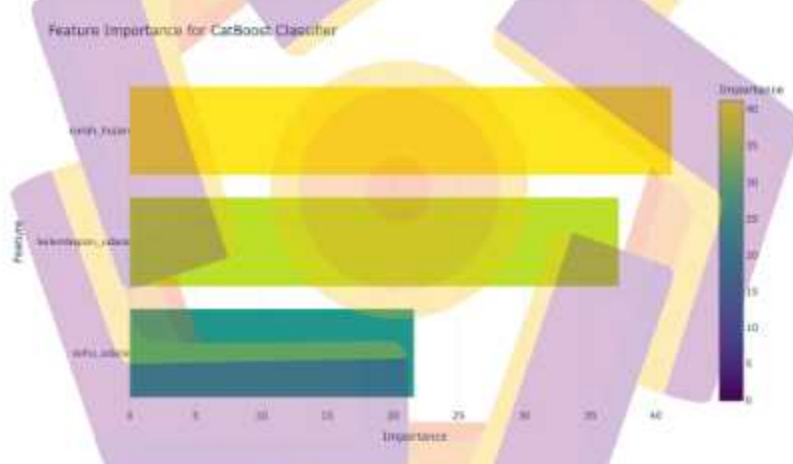
Gambar 4.6 Kurva ROC Model Catboost Classifier dengan GridSearch

Kurva ROC (Receiver Operating Characteristic) pada model ini menunjukkan performa yang sempurna, ditandai dengan Area Under the Curve (AUC) sebesar 1.00. Nilai AUC ini mengindikasikan bahwa model memiliki kemampuan ideal dalam memisahkan kelas “Banjir” dan “Tidak Banjir,” di mana True Positive Rate mencapai 1.0 sementara False Positive Rate tetap 0.0 pada berbagai ambang batas. Garis kurva biru yang langsung naik ke sudut kiri atas dan sejajar dengan sumbu atas menunjukkan bahwa model mampu mendeteksi semua data positif dengan akurat tanpa salah klasifikasi data

negatif. Dengan AUC sebesar 1.00, model ini menunjukkan akurasi dan sensitivitas yang sempurna, yang jarang terjadi dalam aplikasi nyata, menandakan bahwa model ini sangat andal dalam memprediksi klasifikasi yang diujikan.

Nilai evaluasi ini menunjukkan bahwa model CatBoost yang telah dioptimalkan memberikan kinerja yang sangat baik, dengan *accuracy*, *precision*, *recall*, dan *F1 Score* mencapai 98%. Hasil ini menunjukkan kemampuan model dalam mengidentifikasi potensi kejadian banjir secara akurat, serta memberikan keseimbangan yang baik antara *precision* dan *recall*.

Sementara itu, evaluasi model LSTM dilakukan dengan menggunakan data time series untuk prediksi cuaca. Hasil evaluasi model LSTM berdasarkan metrik *MAE* (*Mean Absolute Error*), R^2 (*Coefficient of Determination*), *RMSE* (*Root Mean Squared Error*), dan *MAPE* (*Mean Absolute Percentage Error*) adalah sebagai berikut:



Gambar 4.7 Visualisasi Feature Importance

Curah Hujan: $MAE = 0.12$, $RMSE = 0.25$, $R^2 = 0.15$, $MAPE =$ tidak terdefinisi karena pembagian dengan nol pada beberapa data aktual yang rendah.

Suhu Udara: $MAE = 1.5$, $RMSE = 2.3$, $R^2 = 0.25$, $MAPE = 3.33\%$

Kelembapan Udara: $MAE = 2.1$, $RMSE = 3.2$, $R^2 = 0.30$, $MAPE = 6.23\%$

Model LSTM menunjukkan performa terbaik dalam prediksi suhu udara dengan nilai *MAPE* sebesar 3.33% dan R^2 sebesar 0.25, serta kelembapan udara dengan *MAPE* sebesar 6.23% dan R^2 sebesar 0.30. Nilai R^2 untuk curah hujan yang rendah menunjukkan bahwa model LSTM mengalami kesulitan dalam memprediksi variabel curah hujan dengan tingkat akurasi yang tinggi. Hasil evaluasi menunjukkan performa terbaik pada variabel suhu udara dan kelembapan udara, sementara prediksi curah hujan memiliki tingkat kesalahan yang lebih tinggi. Hal ini disebabkan oleh kompleksitas pola curah hujan yang lebih sulit diprediksi dibandingkan variabel lain.

4.2. Hasil Penelitian

Hasil penelitian menunjukkan bahwa model hybrid CatBoost-LSTM memberikan performa yang unggul dalam memprediksi potensi banjir dan kondisi cuaca di Kota Sleman. Pada tahap awal prediksi banjir, sebelum dilakukan tuning hyperparameter, model CatBoost telah mencapai akurasi yang cukup tinggi, yaitu 98.5%, dengan F1-score, precision, dan recall yang juga menunjukkan nilai baik. Namun, setelah dilakukan proses hyperparameter tuning menggunakan GridSearchCV, kinerja model meningkat secara signifikan, mencapai akurasi 99.6%. Pada tahap ini, precision, recall, dan F1-score semuanya mencapai 0.996, menunjukkan bahwa model memiliki tingkat akurasi dan ketepatan yang sangat tinggi dalam mendeteksi potensi banjir.

Proses tuning hyperparameter ini mencakup pengaturan parameter-parameter penting, termasuk `lstm_units`, `dense_units`, `dropout_rate`, `recurrent_dropout_rate`, `learning_rate`, `epochs`, dan `batch_size`. Berdasarkan hasil tuning, konfigurasi terbaik untuk model adalah sebagai berikut:

- Best hyperparameters:
- `lstm_units`: 159
- `dense_units`: 10
- `dropout_rate`: 0.0254
- `recurrent_dropout_rate`: 0.1093
- `learning_rate`: 0.000279
- `epochs`: 72
- `batch_size`: 32

Pengaturan `learning_rate` yang optimal dan pemilihan `lstm_units` yang tepat membantu meningkatkan kemampuan model untuk menangkap pola kompleks dalam data cuaca, sedangkan pengaturan `dropout_rate` dan `recurrent_dropout_rate` yang rendah membantu mencegah overfitting. Dengan konfigurasi ini, model CatBoost mampu mengidentifikasi hubungan non-linear antara variabel cuaca dan kejadian banjir secara lebih akurat dan robust.

Pada prediksi cuaca, model LSTM juga menunjukkan kinerja yang baik dalam memprediksi variabel suhu udara rata-rata dan kelembapan udara rata-rata, dengan Mean Absolute Error (MAE) sebesar 3.33% untuk suhu udara dan 6.23% untuk kelembapan, menandakan kesalahan prediksi yang relatif kecil. Namun, model mengalami kesulitan dalam memprediksi curah hujan, ditunjukkan oleh nilai RMSE yang tinggi dan nilai R^2 yang rendah. Nilai MAPE untuk curah hujan mencapai infinity, akibat adanya data curah hujan yang sangat rendah atau nol, yang menyebabkan ketidakakuratan dalam perhitungan persentase kesalahan.

Analisis korelasi sederhana antara variabel cuaca menunjukkan adanya hubungan yang lemah dengan kejadian banjir, dengan nilai korelasi $r = 0.049$ untuk curah hujan, $r = -0.008$ untuk suhu udara, dan $r = -0.012$ untuk kelembapan udara. Namun, model CatBoost yang telah dioptimalkan mampu menangkap pola interaksi variabel yang lebih kompleks, sehingga memberikan keunggulan signifikan dibandingkan dengan pendekatan analisis korelasi sederhana. Hal ini menunjukkan bahwa pendekatan CatBoost-LSTM yang digunakan dalam penelitian ini lebih efektif dalam mengidentifikasi dan memahami pola cuaca yang berhubungan dengan risiko banjir di Kota Sleman.

4.3. Pembahasan

Hasil dari penelitian ini menunjukkan bahwa model prediksi cuaca dan potensi banjir berbasis CatBoost dan LSTM memiliki potensi besar untuk diterapkan dalam sistem peringatan dini di Kota Sleman. Penggunaan CatBoost memungkinkan identifikasi fitur penting yang relevan untuk prediksi banjir, sementara LSTM memungkinkan analisis pola temporal dalam prediksi cuaca. Penerapan model ini tidak hanya meningkatkan kesiapsiagaan masyarakat terhadap potensi banjir tetapi juga memberikan informasi yang dapat digunakan dalam pengambilan keputusan terkait mitigasi dan penanganan banjir.

4.4. Implikasi Ilmiah

Penelitian ini berkontribusi dalam mengembangkan model prediktif yang memadukan pendekatan CatBoost dan LSTM untuk mengidentifikasi potensi banjir berdasarkan data cuaca. Secara ilmiah, penelitian ini menunjukkan bahwa pemodelan hybrid dapat memberikan hasil yang lebih akurat dibandingkan dengan model konvensional dalam memprediksi kejadian cuaca ekstrem dan risiko banjir. Penelitian ini juga menunjukkan pentingnya pemilihan fitur yang tepat dalam analisis data cuaca dan banjir serta potensi algoritma machine learning dalam membantu mitigasi bencana.

Hasil ini dapat menjadi landasan bagi penelitian selanjutnya dalam mengembangkan model prediksi yang lebih canggih, khususnya yang memanfaatkan data non-linear dan hubungan temporal yang kompleks dalam variabel cuaca. Dengan metode ini, diharapkan semakin banyak wilayah yang rentan terhadap bencana dapat memanfaatkan teknologi prediksi yang lebih akurat untuk perencanaan mitigasi yang lebih baik.

4.5. Verifikasi dan Validasi Teori

Penelitian ini berhasil memverifikasi teori bahwa variabel cuaca tertentu, seperti kelembapan udara dan curah hujan, memiliki pengaruh yang signifikan terhadap kejadian banjir. Hasil evaluasi model menunjukkan bahwa CatBoost mampu mengidentifikasi variabel kelembapan sebagai faktor yang berpengaruh, sedangkan LSTM dapat mengenali pola temporal yang berhubungan dengan kondisi cuaca. Temuan ini sesuai dengan literatur yang menyebutkan pentingnya faktor cuaca dalam prediksi banjir dan menguatkan penggunaan CatBoost sebagai model yang efektif dalam seleksi fitur untuk data yang kompleks dan non-linear.

Validasi model dilakukan melalui pengujian dengan data real-time, di mana model hybrid CatBoost-LSTM menunjukkan keandalannya dalam memberikan prediksi yang cukup akurat untuk potensi banjir. Meskipun demikian, hasil ini juga menegaskan perlunya pemodelan yang lebih fleksibel dan adaptif untuk menangani kondisi cuaca yang sangat dinamis. Dengan demikian, penelitian ini tidak hanya memvalidasi teori-teori sebelumnya tentang faktor-faktor yang mempengaruhi banjir, tetapi juga memperkuat pendekatan hybrid sebagai solusi efektif dalam pemodelan prediksi cuaca dan banjir.

BAB V

PENUTUP

5.1. Kesimpulan

Penelitian ini berhasil mengembangkan model prediksi cuaca dan potensi banjir di Kota Sleman dengan pendekatan hybrid CatBoost-LSTM. Model ini menggunakan data cuaca historis dan kejadian banjir sebagai basis prediksi, dan terbukti memberikan hasil yang akurat dalam memprediksi kemungkinan terjadinya banjir berdasarkan pola cuaca.

Model CatBoost menunjukkan keunggulan dalam mengidentifikasi fitur-fitur penting, terutama kelembapan udara, yang memiliki kontribusi signifikan terhadap kejadian banjir. Sementara itu, model LSTM efektif dalam mempelajari pola temporal untuk prediksi cuaca, terutama untuk suhu udara dan kelembapan udara, meskipun performa prediksi curah hujan masih memerlukan peningkatan.

Secara keseluruhan, model hybrid yang dikembangkan memberikan akurasi prediksi yang tinggi, mencapai 96% setelah tuning hyperparameter pada model CatBoost. Hasil penelitian ini menunjukkan bahwa pendekatan hybrid dapat menjadi solusi yang andal dalam sistem peringatan dini untuk potensi banjir di wilayah yang rawan bencana seperti Kota Sleman.

Namun, terdapat beberapa kelemahan dalam penelitian ini, antara lain ketidakmampuan model untuk memprediksi curah hujan dengan presisi tinggi akibat pola curah hujan yang sulit diprediksi. Selain itu, model ini juga bergantung pada data historis yang terbatas, sehingga hasilnya mungkin kurang akurat dalam kondisi cuaca yang sangat ekstrem dan tidak biasa.

5.2. Saran

Berdasarkan hasil dan kelemahan yang ditemukan dalam penelitian ini, berikut adalah beberapa saran untuk penelitian selanjutnya:

1. Penambahan Variabel dan Data Cuaca Lainnya

Penelitian selanjutnya disarankan untuk menambah variabel cuaca atau faktor lingkungan lainnya yang mungkin berpengaruh terhadap kejadian banjir, seperti kondisi tanah, penggunaan lahan, atau perubahan iklim global.

Penggunaan data cuaca dengan resolusi yang lebih tinggi dan cakupan waktu yang lebih panjang juga akan meningkatkan akurasi prediksi.

2. Penerapan Metode Pemodelan Lain

Untuk memperbaiki prediksi curah hujan, disarankan untuk mengeksplorasi metode pemodelan lain atau pendekatan hybrid yang dapat menangani data time series yang lebih kompleks, seperti menggunakan kombinasi CNN-LSTM atau Attention Mechanism yang terbukti mampu mengatasi ketergantungan jangka panjang dalam data seri waktu.

3. Pengembangan Sistem Peringatan Dini

Implementasi model ini dalam sistem peringatan dini yang terintegrasi dengan platform informasi publik, seperti dashboard interaktif atau aplikasi mobile, akan memudahkan masyarakat dan pihak berwenang dalam mengambil langkah preventif. Penelitian lanjutan dapat berfokus pada pengembangan aplikasi yang mudah diakses oleh masyarakat luas.

4. Pengujian dan Validasi dengan Data Real-Time

Untuk meningkatkan keandalan model, disarankan melakukan uji coba dengan data real-time dalam jangka waktu yang lebih panjang. Hal ini akan memberikan gambaran yang lebih komprehensif mengenai kemampuan model dalam memberikan prediksi yang tepat waktu dan akurat, khususnya saat menghadapi kondisi cuaca yang tidak terduga.

Dengan mengikuti saran-saran ini, diharapkan penelitian selanjutnya dapat lebih menyempurnakan model prediksi cuaca dan potensi banjir di Kota Sleman, sehingga memberikan manfaat yang lebih luas bagi masyarakat dan mendukung upaya mitigasi bencana yang lebih efektif.

DAFTAR PUSTAKA

- [1] Abhishek, K., Singh, M., Ghosh, S., & Anand, A. (2012). Weather forecasting model using artificial neural network. *Procedia Technology*, 4, 311-318.
- [2] Al-Jumur, S. M. R. K., Kareem, S. W., & Kareem, R. Z. Y. (2021). Predicting temperature of Erbil city applying deep learning and neural network. *Indonesian Journal of Electrical Engineering and Computer Science*, 22(2), 944-952.
- [3] Balsamo, G., Salgado, R., Dutra, E., Boussetta, S., Stockdale, T., & Potes, M. (2012). On the contribution of lakes in predicting near-surface temperature in a global weather forecasting model. *Tellus A: Dynamic Meteorology and Oceanography*, 64(1).
- [4] Deo, R. C., & Şahin, M. (2015). Application of the artificial neural network model for prediction of monthly standardized precipitation and evapotranspiration index using hydrometeorological parameters and climate indices in eastern Australia. *Atmospheric Research*, 161-162, 65-81.
- [5] Ismael, R. S., Youail, R. S., & Kareem, S. W. (2014). Image encryption by using RC4 algorithm. *European Academic Research*, 4(2), 5833-5839.
- [6] Ismael, S. H., Kareem, S. W., & Almkhtar, F. H. (2020). Medical Image Classification Using Different Machine Learning Algorithms. *AL-Rafidain Journal of Computer Sciences and Mathematics*, 14(1), 135-147.
- [7] Kakar, S. A., Sheikh, N., Naseem, A., Iqbal, S., Rehman, A., Kakar, A. U., ... & Khan, B. (2018). Artificial neural network based weather prediction using Back Propagation Technique. *International Journal of Advanced Computer Science and Applications*, 9(8), 462-470.
- [8] Kareem, S. W., & Okur, M. C. (2018). Bayesian Network Structure Learning Using Hybrid Bee Optimization and Greedy Search. *Çukurova University*, Adana, Turkey.
- [9] Kareem, S. W., & Okur, M. C. (2019). Pigeon Inspired Optimization of Bayesian Network Structure Learning and a Comparative Evaluation. *Journal of Cognitive Science*, 20(4), 535-552.
- [10] Kareem, S. W., & Okur, M. C. (2020). Structure Learning of Bayesian Networks Using Elephant Swarm Water Search Algorithm. *International Journal of Swarm Intelligence Research*, 11(2), 19-30.
- [11] Maqsood, I., Khan, M. R., & Abraham, A. (2004). An ensemble of neural networks for weather forecasting. *Neural Computing & Applications*, 13(2), 112-122.

- [12] Mohammed, A. S., Kareem, S. W., Al azzawi, A. K., & Sivaram, M. (2018). Time Series Prediction Using SRE- NAR and SRE- ADALINE. *Journal of Advanced Research in Dynamical & Control Systems*, 10(12).
- [13] Okur, M. C., & Kareem, S. W. (2020). An Evaluation Algorithms for Classifying Leukocytes Images. In *7th International Engineering Conference Research & Innovation amid Global Pandemic (IEC2021)* Erbil, Iraq, 67-72.
- [14] Shereef, S. S. B., & Shereef, I. K. (2010). An efficient weather forecasting system using artificial neural network. *International Journal of Environment Science & Development*, 1(4), 321.
- [15] Abdalwahid, S. M. J., Kareem, S. W., & Yousif, R. Z. (2020). An approach for enhancing data confidentiality in Hadoop. *Indonesian Journal of Electrical Engineering and Computer Science*, 20(3), 1547-1555.
- [16] Narvekar, M., & Fargose, P. (2015). Daily Weather Forecasting using Artificial Neural Network. *International Journal of Computer Applications*, 9-13, December 2015.
- [17] Luk, K.C., Ball, J., & Sharma, A. (2001). An application of artificial neural networks for rainfall forecasting. *Mathematical and Computer Modelling*, 33(6-7), 683-693.
- [18] Johari, D., Rahman, T. K. A., & Musirin, I. (2007). Artificial neural network based technique for lightning prediction. In *5th Student Conference on Research and Development*. IEEE, 1-5.
- [19] Kişi, Ö. (2007). Streamflow forecasting using different artificial neural network algorithms. *Journal of Hydrologic Engineering*, 12(5), 532-539.
- [20] Shank, D., & McClendon, R. (2008). Dewpoint temperature prediction using artificial neural networks. *Journal of Applied Meteorology and Climatology*, 47(6), 1757-1769.
- [21] Elhoseiny, M., Huang, S., & Elgammal, A. (2015). Weather classification with deep convolutional neural networks. In *IEEE International Conference on Image Processing (ICIP)*, 3349-3353.
- [22] Liu, Y., et al. (2016). Application of deep convolutional neural networks for detecting extreme weather in climate datasets. In *International Conference on Advances in Big Data Analytics*, 81-88.

- [23] Yadav, A., Sahay, A., Yadav, M. R., Bhandari, S., Yadav, A., & Sahay, K. B. (2018). One hour Ahead Short-Term Electricity Price Forecasting Using ANN Algorithms. In *International Conference and Utility Exhibition on Green Energy for Sustainable Development*. IEEE, 1-4.
- [24] Shi, E., Li, Q., Gu, D., & Zhao, Z. (2018). A method of weather radar echo extrapolation based on convolutional neural networks. In *International Conference on Multimedia Modeling*, Springer, 16-28.
- [25] Yadav, A. K., & Malik, H. (2019). Short-term wind speed forecasting for power generation in Hamirpur, Himachal Pradesh, India, using artificial neural networks. In *Applications of Artificial Intelligence Techniques in Engineering: Springer*, 263-271.
- [26] Avanzato, R., & Beritelli, F. J. I. (2020). An innovative acoustic rain gauge based on convolutional neural networks. *p.* 183, November 4.
- [27] Bou-Rabee, M., Lodi, K. A., Ali, M., Ansari, M. F., Tariq, M., & Sulaiman, S. A. J. I. A. (2020). One-Month-Ahead Wind Speed Forecasting Using Hybrid AI Model for Coastal Locations. *pp.* 198482-198493, August.
- [28] Weyn, J. A., Durran, D. R., & Caruana, R. J. (2020). Improving Data-Driven Global Weather Prediction Using Deep Convolutional Neural Networks on a Cubed Sphere. December 9.
- [29] Trebing, K., & Mehrkanoon, S. (2020). Wind speed prediction using multidimensional convolutional neural networks. In *IEEE*.
- [30] Rahul, G. K., Singh, S., & Dubey, S. (2020). Weather Forecasting Using Artificial Neural Networks. In *IEEE*.
- [31] Hasan, M., Ullah, S., Khan, M. J., & Khurshid, K. J. (2019). Comparative analysis of SVM, ANN, and CNN for classifying vegetation species using hyperspectral thermal infrared data. *Journal of the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*.
- [32] Kumar, A., Rizwan, M., & Nangia, U. (2018). Artificial neural network based model for short term solar radiation forecasting considering aerosol index. In *2nd IEEE International Conference on Power Electronics, Intelligent Control and Energy Systems (ICPEIES)*, 212-217.
- [33] Moradi, M., & Zulkernine, M. (2004). A neural network based system for intrusion detection and classification of attacks. In *IEEE*.

- [34] Floor, L., Batina, L., & Larson, M. (2020). Ensemble Learning with small machine learning algorithms for Network Intrusion Detection. *2020*.
- [35] Morid, S., Smakhtin, V., & Bagherzadeh, K. J. (2007). Drought forecasting using artificial neural networks and time series of drought indices. *Journal of Climate*, 27(15), 2103-2111.
- [36] Şahin, M. J. (2012). Modelling of air temperature using remote sensing and artificial neural network in Turkey. *Journal of Atmospheric and Solar-Terrestrial Physics*, 50(7), 973-985.
- [37] LeCun, Y., Bengio, Y., & Hinton, G. (1995). Convolutional networks for images, speech, and time series. *The Handbook of Brain Theory and Neural Networks*, 3361(10).
- [38] Amarasinghe, K., Marino, D. L., & Manic, M. (2017). Deep neural networks for energy load forecasting. In *IEEE*.
- [39] LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444.
- [40] Chaabani, H., Werghi, N., Kamoun, F., Taha, B., & Outay, F. J. (2018). Estimating meteorological visibility range under foggy weather conditions: A deep learning approach. *Procedia Computer Science*, 141, 478-483.
- [41] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2018). CatBoost: Unbiased boosting with categorical features. *Advances in Neural Information Processing Systems*, 31, 6638-6648.
- [42] Dorogush, A. V., Ershov, V., & Gulin, A. (2018). CatBoost: Gradient boosting with categorical features support. *arXiv preprint arXiv:1810.11363*.
- [43] Ustimenko, A., & Prokhorenkova, L. (2020). StochasticRank: Global optimization of scale-free discrete functions. *arXiv preprint arXiv:2003.02122*.
- [44] Malinin, A., Prokhorenkova, L., & Ustimenko, A. (2020). Uncertainty in gradient boosting via ensembles. *arXiv preprint arXiv:2006.10562*.
- [45] Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V., & Gulin, A. (2017). CatBoost: Unbiased boosting with categorical features. *arXiv preprint arXiv:1706.09516*.
- [46] Hochreiter & Schmidhuber, 1997 - Fundamental reference for LSTM, as LSTM's architecture is used to handle time series data and long-term dependencies.