

**PENERAPAN ALGORITMA LOGISTIC REGRESSION UNTUK
PREDIKSI BANJIR DI KECAMATAN MEDAN MARELAN
BERDASARKAN DATA CURAH HUJAN DAN FAKTOR
METEOROLOGI**

SKRIPSI

DISUSUN OLEH

ANWAR RUHDIN HASUGIAN

2109020156



UMSU
Unggul | Cerdas | Terpercaya

**PROGRAM STUDI TEKNOLOGI INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
MEDAN
2025**

**PENERAPAN ALGORITMA LOGISTIC REGRESSION UNTUK
PREDIKSI BANJIR DI KECAMATAN MEDAN MARELAN
BERDASARKAN DATA CURAH HUJAN DAN FAKTOR
METEOROLOGI**

SKRIPSI

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

ANWAR RUHDIN HASUGIAN

NPM : 2109020156

PROGRAM STUDI INFORMATIKA

**FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATRA UTARA**

MEDAN

2025

LEMBAR PENGESAHAN

Judul Skripsi : PENERAPAN ALGORITMA LOGISTIC REGRESSION
UNTUK PREDIKSI BANJIR DI KECAMATAN
MEDAN MARELAN BERDASARKAN DATA CURAH
HUJAN DAN FAKTOR METEOROLOGI

Nama Mahasiswa : Anwar Ruhdin Hasugian
NPM : 2109020156
Program Studi : Teknologi Informasi

Menyetujui
Komisi Pembimbing



(Mahardika Abdi Prawira Tanjung S.Kom., M.Kom)

NIDN. 0117088902

Ketua Program Studi



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

Dekan



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

PERNYATAAN ORISINALITAS

**PENERAPAN ALGORITMA LOGISTIC REGRESSION UNTUK
PREDIKSI BANJIR DI KECAMATAN MEDAN MARELAN
BERDASARKAN DATA CURAH HUJAN DAN FAKTOR
METEOROLOGI**

SKRIPSI

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

Medan, 1 September 2025

Yang membuat pernyataan



Anwar Ruhdin Hasugian
NPM. 2109020156

PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS

Sebagai sivitas akademi Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan di bawah ini:

Nama : Anwar Rudi Setiawan Rangkuti
NPM : 2109020142
Program Studi : Teknologi Informasi
Karya Ilmiah : Skripsi

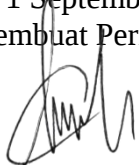
Demi oengembangan ilmu pengetahuan, menyetujui untuk membetikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bebas Royalti Non-Eksekutif ini, (*Non-Exclusive Royalty Free Right*) atas penelitian skripsi yang berjudul:

PENERAPAN ALGORITMA LOGISTIC REGRESSION UNTUK
PREDIKSI BANJIR DI KECAMATAN MEDAN MARELAN
BERDASARKAN DATA CURAH HUJAN DAN FAKTOR
METEOROLOGI

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

Demikian Pernyataan ini dibuat dengan sebenarnya.

Medan, 1 September 2025
Yang Membuat Pernyataan


Anwar Ruhdin Hasugian
NPM. 2109020156

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap	: Anwar Ruhdin Hasugian
Tempat dan Tanggal Lahir	: Sukaramai, 29 Juli 2002
Alamat Rumah	: Sukaramai
Telepon/HP	: 0822 7236 3493
E-mai	: anwarhasugian98@gmail.com
Instansi Tempat Kerja	: -
Alamat Kantor	: -

DATA PENDIDIKAN

SD	: SDN 030436 Kerajaan	TAMAT:2014
SMP	: SMPN 1 Kerajaan	TAMAT:2017
SMA	: SMKN 1 Pergeteng Geteng Sengkut	TAMAT:2021

KATA PENGANTAR



Assalamualaikum warahmatullahi Wabarakatu

Puji syukur penulis panjatkan kepada Tuhan Yang Maha Esa, atas segala rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “PENERAPAN ALGORITMA LOGISTIC REGRESSION UNTUK PREDIKSI BANJIR DI KECAMATAN MEDAN MARELAN BERDASARKAN DATA CURAH HUJAN DAN FAKTOR METEOROLOGI” ini dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Teknologi Informasi, Fakultas Ilmu Komputer & Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara.

Penulisan skripsi ini tidak lepas dari bantuan, dukungan, serta bimbingan dari berbagai pihak. Oleh karena itu, pada kesempatan ini, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Agussani, M.AP., Rektor Universitas Muhammadiyah Sumatera Utara (UMSU)
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Ibu Fatma Sari Hutagalung, S.Kom., M.Kom. Ketua Program Studi Teknologi Informasi .
4. Bapak Mhd. Basri, S.Si, M.Kom Sekretaris Program Studi Teknologi Informasi.

5. Bapak Mahardika Prawira Tanjung, S.Kom., M.Kom selaku pembimbing skripsi yang telah memberikan arahan, saran, dan bimbingan dengan penuh kesabaran dan ketulusan.
6. Ayahanda saya dan Ibunda saya yang telah memberikan dukungan dari segi moral maupun materiil serta menjadi motivasi penulis untuk dapat menyelesaikan tugas akhir ini
7. Abang, Kakak Dan Adik saya Tercinta yang telah memberikan motivasi serta dukungan yang selalu menjadi dari segi moral maupun materiil serta menjadi motivasi penulis untuk dapat menyelesaikan tugas akhir ini
8. Seluruh Bapak dan Ibu Dosen di Jurusan Teknologi Informasi Fakultas Ilmu Komputer & Teknologi Informasi Universitas Muhammdiyah Sumatera Utara atas ilmu dan didikannya selama perkuliahan.
9. Sahabat dan teman-teman seperjuangan jurusan teknologi informasi UMSU
10. Semua pihak yang tidak dapat penulis sebutkan satu per satu yang telah membantu dalam penyelesaian skripsi ini.

Penulis menyadari bahwa dalam penulisan skripsi ini masih terdapat kekurangan. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi perbaikan di masa yang akan datang. Akhir kata, penulis berharap skripsi ini dapat bermanfaat bagi semua pihak yang berkepentingan, khususnya bagi perkembangan ilmu pengetahuan di bidang Nama Bidang Ilmu.

Medan, 1 September 2025

Anwar Ruhdin Hasugian

NPM. 2109020156

**PENERAPAN ALGORITMA LOGISTIC REGRESSION UNTUK
PREDIKSI BANJIR DI KECAMATAN MEDAN MARELAN
BERDASARKAN DATA CURAH HUJAN DAN FAKTOR
METEOROLOGI**

ABSTRAK

Banjir merupakan salah satu bencana hidrometeorologi yang paling sering terjadi di Indonesia dan menimbulkan dampak signifikan terhadap aktivitas sosial, ekonomi, serta infrastruktur. Kecamatan Medan Marelan termasuk wilayah dengan tingkat kerawanan banjir yang tinggi akibat curah hujan ekstrem dan keterbatasan sistem drainase. Oleh karena itu, diperlukan pendekatan prediktif berbasis data untuk mendukung upaya mitigasi bencana. Penelitian ini bertujuan mengembangkan model prediksi risiko banjir menggunakan algoritma *Logistic Regression* dengan memanfaatkan variabel meteorologi, yaitu curah hujan, suhu udara, kelembapan, tekanan udara, dan kecepatan angin. Data penelitian diperoleh dari BMKG Medan dan BPBD Kota Medan dengan total lebih dari 1.000 catatan harian selama periode Mei 2022–Mei 2025. Prosedur penelitian meliputi tahap pengumpulan data, pra-pemrosesan, pembagian dataset menjadi data latih dan data uji, pelatihan model, *hyperparameter tuning*, serta evaluasi performa. Kinerja model diuji menggunakan metrik *accuracy*, *precision*, *recall*, *F1-score*, dan AUC-ROC. Hasil evaluasi menunjukkan bahwa *Logistic Regression* mampu memberikan prediksi dengan akurasi 64%, *precision* 0,99, *recall* 0,64, *F1-score* 0,78, dan nilai AUC sebesar 0,60. Model ini dinilai cukup efektif dalam mengenali potensi kejadian banjir, meskipun masih terkendala oleh ketidakseimbangan distribusi data. Dengan demikian, penerapan *Logistic Regression* terbukti memiliki potensi sebagai metode prediksi risiko banjir yang sederhana, interpretatif, dan adaptif. Temuan penelitian ini diharapkan dapat berkontribusi pada pengembangan sistem peringatan dini serta menjadi dasar perumusan kebijakan mitigasi banjir di tingkat lokal.

Kata Kunci: Logistic Regression, prediksi banjir, faktor meteorologi, curah hujan, Medan Marelan.

ABSTRACT

Flooding is one of the most frequent hydrometeorological disasters in Indonesia, causing significant impacts on social, economic, and infrastructure sectors. Medan Marelan District is categorized as a high-risk area due to extreme rainfall intensity and insufficient drainage systems. Therefore, a data-driven predictive approach is required to support disaster mitigation strategies. This study aims to develop a flood risk prediction model using the *Logistic Regression* algorithm by incorporating meteorological variables, including rainfall, air temperature, humidity, air pressure, and wind speed. The dataset was obtained from BMKG Medan and BPBD Medan, consisting of more than 1,000 daily records spanning from May 2022 to May 2025. The research procedures involved data collection, preprocessing, dataset partitioning into training and testing sets, model training, hyperparameter tuning, and performance evaluation. The model's performance was assessed using *accuracy*, *precision*, *recall*, *F1-score*, and AUC-ROC metrics. The results demonstrated that *Logistic Regression* achieved an accuracy of 64%, a precision of 0.99, a recall of 0.64, an F1-score of 0.78, and an AUC value of 0.60. Although the model effectively detected most flood events, its performance was influenced by the imbalance of class distribution. In conclusion, the application of *Logistic Regression* has proven to be a simple, interpretable, and adaptive method for predicting flood risks in Medan Marelan. The findings of this research are expected to contribute to the development of early warning systems and serve as a basis for local policy formulation in flood disaster mitigation.

Keywords: Logistic Regression, flood prediction, meteorological factors, rainfall, Medan Marelan.

DAFTAR ISI

LEMBAR PENGESAHAN	iii
PERNYATAAN ORISINALITAS.....	iv
PERNYATAAN KEASLIAN TUGAS AKHIR	v
RIWAYAT HIDUP	vi
KATA PENGANTAR.....	vii
ABSTRAK	ix
ABSTRACT	x
DAFTAR ISI.....	xi
DAFTAR TABEL	xiii
DAFTAR GAMBAR.....	xiv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah	5
1.4 Tujuan Penelitian	6
1.5 Manfaat Penelitian	7
BAB II LANDASAN TEORI	9
2.1 Pengertian Banjir	9
2.2 Curah Hujan	11
2.3 Faktor Meteorologi	12
2.4 Algoritma Logistic Regression	13
2.5 Penerapan Logistic Regression	15
2.6 Machine Learning	16
2.7 Faktor Mempengaruhi Banjir	17
BAB III METODOLOGI PENELITIAN	18
3.1 Metode Penelitian.....	18
3.2 Waktu Peneelitan	18
3.3 Alur Penelitian	19
3.4 Model Prediksi	20
3.5 Perancangan Sistem Prediksi Banjir	21
3.5.1 Pengumpulan Data	22
3.5.2 Penerapan Algoritma Logistic Regression.....	24
3.5.3 Preprocessing Data	27
3.5.4 Model Implementasi Dan Hasil Visualisasi.....	28
3.5.5 Evaluasi Model.....	29
3.5.6 Kebutuhan Hardware Dan Software	30
3.5.6.1 Kebutuhan Hardware	30
3.5.6.2 Kebutuhan Software.....	31

BAB IV HASIL DAN PEMBAHASAN	32
4.1 Pengumpulan Dan Persiapan Data	32
4.2 Pelatihan Dan Tuning Model	34
4.3 Hasil Evaluasi Model	39
4.4 Pembahasan Hasil	42
4.5 Implementasi Antar Muka (UI).....	44
4.5.1 Pentingnya Antarmuka dalam Penelitian	45
4.5.2 Desain Antar Muka	46
4.5.3 Alur Antar muka	47
 BAB V KESIMPULAN DAN SARAN	 48
5.1 Kesimpulan	48
5.2 Saran.....	49
DAFTAR PUSTAKA	52
LAMPIRAN.....	54

DAFTAR TABEL

Tabel 3.1 Waktu Penelitian	18
Tabel 3.2 Variabel (X)	25
Tabel 3.3 Variabel Dataset	26
Tabel 3.4 Kebutuhan Hardware	30
Tabel 4.1 Hasil Metrik Evaluasi Model	40
Tabel 4.2 Hasil Evaluasi Classification Report Logistic Regression	42
Tabel 4.3 Hasil Confusion Matrix Logistic Regression	43
Tabel 4.4 Nilai AUC Logistic Regression	44

DAFTAR GAMBAR

Gambar 2.1 Banjir Bandang.....	10
Gambar 2.2 Grafik Hujan.....	11
Gambar 2.3 Perubahan Iklim	13
Gambar 2.4 Logistic Regression	14
Gambar 2.5 Model logistic Regression.....	21
Gambar 3.1 Model Perancangan Sistem	22
Gambar 3.2 Data BMKG	23
Gambar 3.3 Preprocessing Data	28
Gambar 3.4 Probabilitas Banjir.....	29
Gambar 4.1 Gambar Dataset	33
Gambar 4.2 Pembagian Data.....	35
Gambar 4.3 Hasil Dari Penurunan Overfitting	37
Gambar 4.4 Evaluasi Model.....	38
Gambar 4.5 Visualisaai Cofusion Matrik	40
Gambar 4.6 Kurva ROC-AUC Model	42
Gambar 4.7 Desain Halaman Input Data	46
Gambar 4.8 Desain Halaman Hasil Prediksi Banjir.....	47
Gambar 4.9 Visualisasi Tren Cuaca Dan Probabilitas Banjir	47

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Banjir adalah sebuah peristiwa dimana daratan yang biasanya kering menjadi tergenang oleh air, yang disebabkan adanya curah hujan yang tinggi dan kondisi topografi wilayah yang rendah hingga cekungan, dan dapat juga disebabkan oleh rendahnya kemampuan infiltrasi oleh tanah, sehingga menyebabkan tanah tidak mampu lagi menampung dan menyerap air (Nuryanti et al., 2018). Bencana banjir merupakan bencana alam yang paling sering terjadi, baik dilihat dari intensitasnya pada suatu tempat maupun jumlah lokasi kejadian dalam setahun yaitu sekitar 40% di antara bencana alam yang lainnya (Darmawan et al., 2017).

Secara teoritis, banyak faktor yang menjadi penyebab terjadinya banjir. Namun secara umum penyebab terjadinya banjir dapat diklasifikasikan menjadi 2 kategori, yaitu banjir yang disebabkan oleh alam dan banjir yang diakibatkan oleh tindakan manusia. Diantara faktor alam yang menjadi pemicu terjadinya banjir adalah Curah hujan, Pengaruh Fisiografi Erosi, dan Sedimentasi Kapasitas sungai, Kapasitas Drainase yang tidak memadai Pengaruh air pasang (Muna et al., 2016). Sementara itu, faktor yang disebabkan oleh manusia antara lain adalah pembuangan sampah ke dalam saluran air yang menghambat aliran, kerusakan pada bangunan pengendali banjir, serta perencanaan sistem drainase dan pengendalian banjir yang tidak efektif (Rahayu, 2019).

Selain peningkatan infrastruktur drainase, solusi lainnya mencakup normalisasi sungai, pengelolaan tata ruang yang lebih baik, serta edukasi kepada

masyarakat mengenai pentingnya menjaga kebersihan saluran air. Upaya ini harus didukung dengan penertiban bangunan liar yang menghambat aliran air dan penerapan kebijakan tata ruang yang mempertimbangkan fungsi daerah resapan (Yuli Supodo et al., 2021). Selain itu, normalisasi sungai, perbaikan drainase, dan pembangunan tanggul di daerah rawan banjir juga diperlukan untuk meningkatkan efektivitas sistem pengendalian banjir. Dari sisi sosial, keterlibatan masyarakat dalam menjaga kebersihan lingkungan, gotong royong membersihkan saluran air, serta peningkatan kesadaran terhadap pentingnya pengelolaan sampah menjadi faktor kunci dalam mitigasi banjir (Muna et al., 2016).

Berbagai penelitian telah dilakukan untuk mengatasi permasalahan banjir, baik melalui solusi teknis maupun pendekatan sosial. Solusi ini menawarkan peninggian dan pelebaran saluran air yang terbukti efektif dalam mengurangi genangan,. Namun, solusi ini masih bersifat lokal, belum diuji di berbagai kondisi wilayah, serta belum mengintegrasikan teknologi prediksi banjir berbasis data (Yuli Supodo et al., 2021). Dan Untuk menekankan pentingnya ketahanan sosial masyarakat dalam menghadapi banjir melalui strategi adaptasi, seperti peninggian rumah dan membersihkan drainase. Pendekatan ini menunjukkan peran masyarakat dalam mitigasi banjir, tetapi tidak disertai dengan solusi teknis berbasis data dan kurang mempertimbangkan aspek kebijakan maupun perencanaan infrastruktur jangka panjang. Oleh karena itu, diperlukan adanya pembaharuan, yang tidak hanya mengandalkan mitigasi struktural dan adaptasi sosial, tetapi juga mengoptimalkan teknologi prediksi berbasis kecerdasan buatan serta sistem peringatan dini guna meningkatkan efektivitas dalam pencegahan banjir (Muna et al., 2016)

Banjir di Kecamatan Medan Marelan sering terjadi akibat curah hujan tinggi dan sistem drainase yang kurang optimal. Prediksi berbasis Logistic Regression menjadi solusi efektif untuk mendeteksi potensi banjir dengan mengolah data curah hujan dan faktor meteorologi. Selain itu, variabel seperti tekanan udara, kelembapan, dan pola angin perlu dipertimbangkan agar prediksi lebih akurat. Upaya mitigasi juga mencakup peningkatan drainase dan adaptasi masyarakat terhadap bencana. Dengan pendekatan ini, penelitian ini diharapkan dapat memberikan solusi komprehensif dalam mengurangi dampak banjir di Kecamatan Medan Marelan.

Prediksi banjir yang akurat sangat penting untuk mitigasi bencana di Kecamatan Medan Marelan. Penerapan Logistic Regression memungkinkan analisis data curah hujan dan faktor meteorologi seperti tekanan udara, kelembapan, dan pola angin untuk meningkatkan ketepatan prediksi. Hasil prediksi yang akurat dapat digunakan sebagai dasar peringatan dini dan perencanaan infrastruktur. Dengan pendekatan ini, penelitian ini berkontribusi dalam mengurangi risiko dan dampak banjir secara efektif. Oleh karena itu, studi ini penting dalam mendukung strategi mitigasi berbasis data di Kecamatan Medan Marelan.

Logistic Regression dapat digunakan untuk memprediksi banjir di Kecamatan Medan Marelan dengan menganalisis hubungan antara curah hujan dan faktor meteorologi seperti tekanan udara, kelembapan, dan pola angin. Model ini menghasilkan prediksi yang membantu peringatan dini serta mendukung perencanaan infrastruktur drainase. Dengan penerapan metode ini, mitigasi banjir

dapat dilakukan lebih efektif, mengurangi risiko genangan, dan meminimalkan dampak bagi masyarakat.

Penelitian ini bertujuan untuk memprediksi banjir di Kecamatan Medan Marelan dengan Logistic Regression, mengolah data curah hujan serta faktor meteorologi seperti tekanan udara, kelembapan, dan pola angin. Model yang dikembangkan akan dievaluasi menggunakan precision, recall, dan F1-score untuk memastikan keakuratannya. Hasilnya diharapkan menjadi dasar peringatan dini, membantu mitigasi risiko, serta mendukung perencanaan infrastruktur dan kebijakan penanggulangan banjir yang lebih efektif.

Selain itu, hasil penelitian ini juga dapat berkontribusi dalam perencanaan infrastruktur kota yang lebih adaptif terhadap perubahan iklim, meningkatkan kesadaran masyarakat terhadap risiko banjir, serta mendorong kebijakan berbasis data yang lebih efektif dalam pengelolaan lingkungan. Dengan adanya sistem prediksi yang lebih akurat, diharapkan Kota Medan dapat mengurangi dampak negatif banjir secara signifikan dan mewujudkan lingkungan perkotaan yang lebih tangguh dan berkelanjutan.

1.2. Rumusan Masalah

Berdasarkan permasalahan pada latar belakang di atas, penelitian ini akan difokuskan pada beberapa isu utama berikut :

1. Analisis pengaruh curah hujan terhadap risiko banjir di Kecamatan Medan Marelan diperlukan untuk memahami tingkat ketergantungannya terhadap kejadian banjir.

2. Faktor meteorologi seperti suhu, kelembaban, dan kecepatan angin memiliki pengaruh signifikan terhadap prediksi risiko banjir di Kecamatan Medan Marelan.
3. Penerapan algoritma Logistic Regression dapat digunakan untuk memprediksi risiko banjir di Kecamatan Medan Marelan berdasarkan data curah hujan dan faktor meteorologi yang ada.
4. Evaluasi akurasi model Logistic Regression dalam memprediksi risiko banjir di Kota Medan penting dilakukan untuk mengetahui efektivitas algoritma ini dibandingkan dengan metode prediksi lain.
5. Tantangan yang dihadapi dalam penerapan algoritma Logistic Regression meliputi pengolahan data curah hujan yang tidak merata, pemilihan faktor meteorologi yang tepat, serta proses pemodelan dan validasi hasil prediksi.

1.3. Batasan Masalah

Penelitian ini dilakukan dengan batasan-batasan yang ditetapkan secara hati-hati untuk memastikan penelitian tetap fokus dan terarah. Batasan-batasan ini diterapkan agar ruang lingkup penelitian dapat dikelola secara efisien, sehingga menghasilkan keluaran yang sesuai dengan tujuan penelitian. Batasan masalah dalam penelitian ini adalah sebagai berikut :

1. Penelitian ini hanya fokus pada prediksi risiko banjir di Kecamatan Medan Marelan berdasarkan data curah hujan dan faktor meteorologi yang tersedia pada periode tertentu.

2. Faktor meteorologi yang dianalisis dalam penelitian ini terbatas pada suhu udara, kelembaban, dan kecepatan angin, tanpa memasukkan faktor meteorologi lainnya.
3. Penelitian ini tidak mempertimbangkan aspek mitigasi bencana atau solusi terkait banjir, fokus utama hanya pada analisis prediksi risiko banjir berdasarkan data meteorologi.
4. Data curah hujan yang digunakan dalam penelitian ini terbatas pada sumber data yang tersedia dari lembaga meteorologi resmi yang memiliki cakupan geografis di Kecamatan Medan Marelan.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini diformulasikan untuk memberikan arah yang jelas terhadap pengembangan dan penerapan sistem prediksi risiko banjir pada Kecamatan Medan Marelan. Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut :

1. Menganalisis hubungan antara curah hujan dan faktor meteorologi (suhu udara, kelembaban, dan kecepatan angin) terhadap risiko banjir di Kecamatan Medan Marelan.
2. Mengembangkan model prediksi risiko banjir di Kecamatan Medan Marelan dengan menggunakan algoritma Logistic Regression berdasarkan data curah hujan dan faktor meteorologi yang ada.
3. Mengevaluasi efektivitas dan akurasi model Logistic Regression dalam memprediksi risiko banjir di Kecamatan Medan Marelan, serta membandingkan hasil prediksi dengan kejadian banjir yang tercatat.

4. Menyediakan pemahaman tentang penerapan algoritma Logistic Regression dalam analisis risiko banjir yang dapat digunakan untuk mendukung upaya mitigasi bencana di Kecamatan Medan Marelan.

1.5. Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan berbagai manfaat, baik dalam konteks keselamatan maupun dalam pengembangan teknologi yang lebih praktis dan ekonomis. Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Penelitian ini diharapkan dapat memberikan informasi prediksi risiko banjir yang lebih akurat menggunakan algoritma Logistic Regression. Dengan informasi ini, masyarakat dapat lebih siap dalam menghindari dampak banjir, seperti evakuasi lebih awal dan penguatan infrastruktur.
2. Algoritma Logistic Regression yang digunakan dalam penelitian ini dapat membantu pemerintah setempat dalam merancang kebijakan mitigasi banjir yang lebih tepat. Prediksi yang dihasilkan dari algoritma ini memungkinkan tindakan pencegahan dilakukan lebih cepat sebelum banjir terjadi, seperti penutupan jalan atau pengalihan sumber daya.
3. Prediksi risiko banjir yang lebih tepat, yang dihasilkan oleh algoritma Logistic Regression, dapat membantu masyarakat meningkatkan kewaspadaan dan mempersiapkan diri lebih baik. Ini akan mengurangi dampak kerugian material dan korban jiwa akibat banjir.
4. Penelitian ini membuka peluang untuk mengembangkan aplikasi berbasis algoritma Logistic Regression yang dapat digunakan untuk memprediksi

bencana banjir secara otomatis, membantu berbagai pihak terkait dalam merespons lebih cepat dan efektif.

BAB II

LANDASAN TEORI

2.1. Pengertian Banjir

Banjir adalah salah satu bencana alam yang sering terjadi di Indonesia, terutama di wilayah dengan curah hujan tinggi. Dampaknya sangat merugikan, mulai dari kerusakan infrastruktur, terganggunya aktivitas masyarakat, hingga kerugian ekonomi yang signifikan. Pada tahun 2020 di Jakarta, data dari Badan Penanggulangan Bencana Daerah (BPDP) dan Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) menunjukkan bahwa pada bulan Januari, sebanyak 35 kecamatan dan 151 kelurahan terdampak banjir. Sementara itu, pada bulan Februari, jumlah wilayah terdampak meningkat menjadi 42 kecamatan dan 167 kelurahan. Di daerah-daerah tertentu, seperti wilayah dataran rendah atau area dengan drainase yang buruk, risiko terjadinya banjir meningkat, terutama saat curah hujan mencapai intensitas tinggi. (Adiyasa et al., 2025)

Saat ini yang menjadi isu publik adalah kerawanan terhadap banjir dadakan akan meningkat, sungai dangkal dan penambahan volume air jauh lebih besar dari pada yang tertampung. Banjir terjadi sepanjang sistem sungai dan anak-anak sungainya, mampu membanjiri wilayah luas dan mendorong peluapan air di dataran rendah, sehingga banjir yang meluap dari sungai-sungai selain induk sungai biasa disebut ‘banjir kiriman’. Besarnya banjir tergantung kepada beberapa faktor, di antaranya kondisi-kondisi tanah (kelembaban tanah, vegetasi, perubahan suhu/musim, keadaan permukaan tanah yang tertutup rapat oleh bangunan; batu bata, blok-blok semen, beton, pemukiman/ perumahan dan hilangnya kawasan-

kawasan tangkapan air atau alih fungsi lahan.

Sungai atau aliran air yang menyediakan kemudahan hidup bagi masyarakat disekitarnya itu juga bisa menjadikan masyarakat menghadapi risiko bencana tahunan akibat banjir. Banjir dapat terjadi akibat naiknya permukaan air lantaran curah hujan yang tinggi, perubahan suhu, tanggul/bendungan yang bobol, terhambatnya aliran air di tempat lain. Diperkotaan genangan lokal terjadi pada saat musim hujan, skala banjir yang terjadi cukup besar dan belum dapat dikendalikan secara dominan.

Banjir kilat/dadakan biasanya didefinisikan sebagai banjir yang terjadi hanya dalam waktu kurang dari 5 jam sesudah hujan lebat mulai turun. Biasanya juga dihubungkan dengan banyaknya awan kumululus yang menggumpal di angkasa, kilat atau petir yang keras, badai tropis atau cuaca ekstrem. Karena banjir ini sangat cepat datangnya, peringatan bahaya kepada penduduk sekitar tempat itu harus dengan segera dimulai upaya penyelamatan dan persiapan penanggulangan dampak-dampaknya. Umumnya banjir akibat meluapnya air hujan yang sangat deras, yang tak mampu menahan cukup banyak air. Penyebab lain adalah kegagalan bendungan/tanggul menahan volume air (debit) yang meningkat. (Sebastian, 2008)

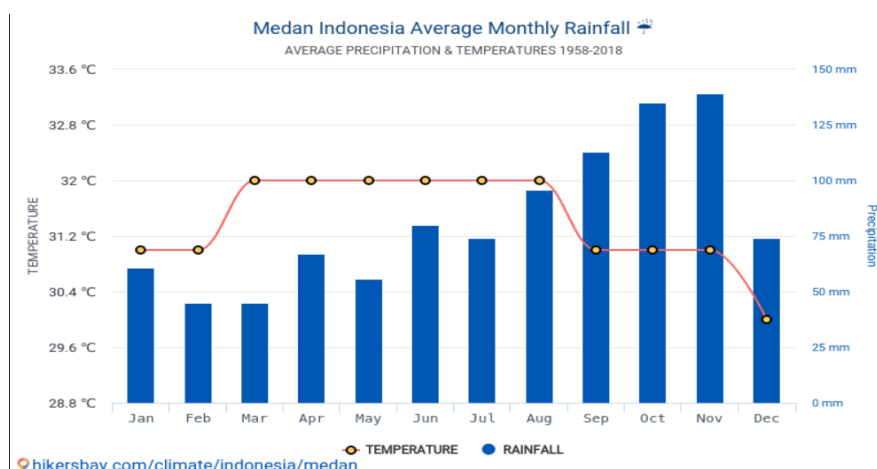


Gambar 2.1 Banjir Bandang

(Sumber : <https://images.app.goo.gl/6bXAfqxFPhA7uzbo8>)

2.2. Curah Hujan

Curah hujan adalah besarnya volume air yang jatuh pada suatu areal tertentu, besarnya curah hujan dapat dinyatakan dalam m³ per satuan luas atau secara umum dinyatakan dalam tinggi kolam air. Curah hujan berpengaruh terhadap potensi terjadinya banjir, jika suatu daerah memiliki curah hujan tinggi maka semakin tinggi potensi terjadinya banjir. Dan begitu sebaliknya, semakin rendah curah hujannya maka potensi terjadinya banjir juga rendah. Berdasarkan data Badan Pusat Statistik Sumatera Utara tahun 2019, terdapat tiga stasiun atau titik perhitungan curah hujan yang terdapat di Kota Medan. Yaitu Stasiun BMKG Belawan, Stasiun BMKG Deli Serdang, dan Stasiun BMKG Wilayah I Kota Medan. Parameter curah hujan di Kota Medan dibagi menjadi lima kelas yaitu sangat rendah (< 2500 mm/tahun), rendah (2500 – 3500 mm/tahun), sedang (3500 – 4500 mm/tahun), tinggi (4500 – 5500 mm/tahun), dan sangat tinggi (> 5500 mm/tahun). Curah hujan di Kota Medan secara keseluruhan berada pada kelas rendah (2500 – 3500 mm/tahun) dan sangat rendah (< 2500 mm/tahun). (Anggraini et al., 2021)



Gambar 2.2 Grafik Hujan

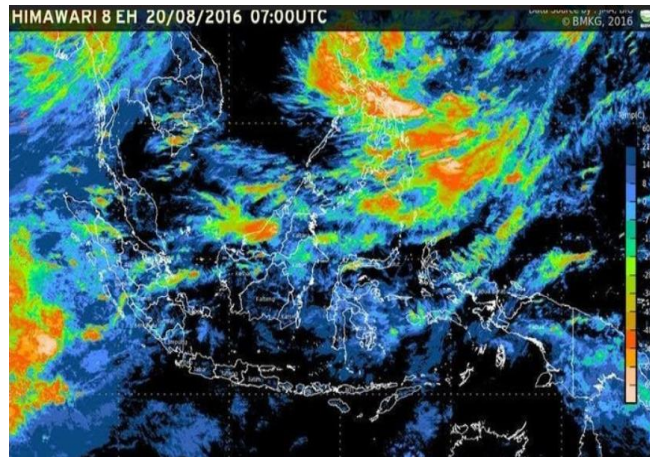
(Sumber : <https://images.app.goo.gl/6VWv8v15Q7aHyMXm7>)

2.3. Faktor Meteorologi

Faktor meteorologi adalah elemen atmosfer yang mempengaruhi kondisi cuaca dan iklim suatu wilayah, termasuk suhu udara, kelembapan, tekanan udara, dan pola angin. Faktor meteorologi seperti suhu udara, kelembapan, tekanan udara, dan angin memiliki pengaruh signifikan terhadap curah hujan dan risiko banjir. Suhu udara tinggi mempercepat proses evaporasi, meningkatkan jumlah uap air di atmosfer, yang pada akhirnya memicu curah hujan ekstrem. Kelembapan udara yang tinggi mempercepat pembentukan awan hujan tebal, yang berpotensi menghasilkan hujan deras dalam waktu singkat.

Fenomena cuaca yang terjadi dipengaruhi oleh beberapa faktor meteorologi yang berdasarkan skala meteorologi terbagi menjadi skala global, sinoptik, dan skala meso atau lokal. Analisis terhadap skala-skala tersebut dapat membantu untuk mengetahui keterkaitan antar skala tersebut serta unsur cuaca yang paling berpengaruh pada terjadinya fenomena cuaca di suatu tempat. Unsur cuaca pada skala lokal dapat dianalisis menggunakan data dari pengamatan udara dan dengan memanfaatkan citra satelit cuaca.

faktor meteorologi yang mempengaruhi pola cuaca di suatu tempat perlu untuk memperhatikan dinamika cuaca berdasarkan skala meteorologi karena pada umumnya tidak semua fenomena cuaca disebabkan oleh faktor lokal ataupun sinoptik. skala sinoptik dapat menggunakan nilai suhu permukaan laut serta peta medan angin sedangkan untuk analisis skala lokal dapat menggunakan indeks-indeks kemantapan udara yang didapat dari pengamatan udara serta kelembapan udara dari lapisan permukaan hingga lapisan atas. (Rony Kurniawan, 2019)



Gambar 2.3 Perubahan Iklim

(Sumber : <https://images.app.goo.gl/R4YYsuxBB26FSECNA>)

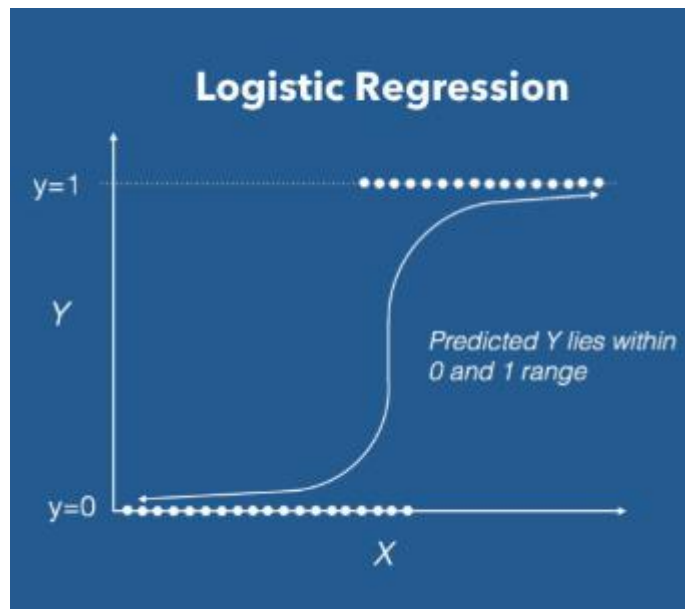
2.4. Algoritma Logistic Regression

Algoritma Logistic Regression adalah metode statistik dan machine learning yang digunakan untuk memprediksi probabilitas suatu kejadian berdasarkan variabel independen. Berbeda dengan regresi linear, algoritma ini menggunakan fungsi sigmoid untuk menghasilkan nilai antara 0 dan 1, yang dapat dikonversi menjadi klasifikasi biner, seperti "berisiko banjir" atau "tidak berisiko banjir" (Hidayat et al., 2022). Logistic Regression bekerja dengan menghitung log odds dari variabel dependen menggunakan persamaan:

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}}$$

Dimana $P(Y=1)$ adalah probabilitas kejadian (misalnya, terjadinya banjir), β_0 adalah intercept, dan $\beta_n X_n$ adalah bobot dari masing-masing variabel prediktor. Dengan pendekatan ini, Logistic Regression mampu mengidentifikasi hubungan

non-linear antara variabel meteorologi, seperti curah hujan, suhu, kelembapan, dan tekanan udara, dalam menentukan risiko banjir (Adiyasa et al., 2025).



Gambar 2.5 Logistic Regression

(Sumber : <https://images.app.goo.gl/MUrgC3eTtTae8yb96>)

Keunggulan Logistic Regression dalam penelitian ini adalah kemampuannya untuk menghasilkan interpretasi yang jelas, di mana koefisien regresi dapat menunjukkan seberapa besar pengaruh setiap variabel terhadap kemungkinan banjir. Namun, penelitian juga mencatat bahwa model ini memiliki keterbatasan dalam menangani data yang sangat kompleks atau memiliki banyak variabel interaksi, yang terkadang memerlukan pendekatan lain seperti Random Forest atau Neural Networks. Oleh karena itu, meskipun Logistic Regression merupakan metode yang sederhana, efisien, dan mudah diinterpretasikan, penggunaannya dalam prediksi banjir perlu dipertimbangkan berdasarkan karakteristik data yang digunakan (Adiyasa et al., 2025).

2.5. Penerapan Logistic Regression

Logistic Regression banyak digunakan dalam prediksi risiko bencana, termasuk banjir, karena kemampuannya dalam menghubungkan faktor meteorologi dengan probabilitas kejadian, model ini efektif dalam menganalisis hubungan non-linear antara variabel cuaca dan dampaknya (Hidayat et al., 2022). Dalam penelitiannya, Logistic Regression diterapkan untuk mendeteksi potensi banjir di Jakarta berdasarkan curah hujan dan faktor meteorologi lainnya.

Hasilnya menunjukkan bahwa model ini mampu mengidentifikasi wilayah dengan risiko tinggi dan rendah, sehingga dapat membantu dalam perencanaan mitigasi bencana. Keunggulan utama Logistic Regression adalah interpretasinya yang jelas dan kemampuannya dalam memodelkan hubungan probabilistik. Namun, akurasi dapat dipengaruhi oleh kompleksitas data, sehingga sering dikombinasikan dengan metode lain untuk hasil yang lebih optimal (Adiyasa et al., 2025).

Regresi Logistik adalah suatu metode analisis statistika untuk mendeskripsikan hubungan antara perubah respon (dependent variable) yang memiliki dua kategori atau lebih dengan satu atau lebih perubah penjelas (independent variable) berskala kategori atau interval. Regresi Logistik merupakan regresi non linear, digunakan untuk menjelaskan hubungan antara X dan Y yang bersifat tidak linear, ketidak normalan sebaran Y, keragaman respon tidak konstan yang tidak dapat dijelaskan dengan model regresi linear biasa. (Hendayana, 2013)

2.6. Machine Learning

Salah satu bidang Kecerdasan Buatan yang semakin populer saat ini adalah Machine Learning. Machine learning merupakan bagian dari kecerdasan buatan yang menginstruksikan komputer untuk mempelajari informasi dari data dan mengeluarkan prediksi atau tindakan berdasarkan interpretasi terhadap data tersebut. Dengan machine learning, komputer dapat mengambil keputusan atau melakukan tindakan berdasarkan analisis data tanpa memerlukan pemrograman eksplisit (Gregorius N. Elwirehardja; Teddy Suparyanto; Bens Pardamean, 2023).

Secara sederhana, mesin "mengambil pelajaran" dari pengalaman sebelumnya dan menggunakan pengetahuan yang diperoleh untuk menghadapi situasi baru. Machine Learning sebuah panduan esensial dalam pengembangan dan penerapan solusi Machine Learning. Konsep ini memungkinkan kita untuk mengambil langkah-langkah yang terstruktur dalam memecahkan masalah dan membuat prediksi cerdas. (Chyan et al., n.d.)

Machine Learning menjadi pendekatan utama dalam prediksi risiko banjir karena kemampuannya dalam mengolah data historis dan menemukan pola kompleks dari berbagai faktor yang memengaruhi kejadian banjir. Logistic Regression, sebagai salah satu algoritma Machine Learning, telah banyak digunakan dalam prediksi bencana karena sifatnya yang mampu mengklasifikasikan data ke dalam kategori risiko, seperti "berisiko banjir" atau "tidak berisiko banjir" berdasarkan variabel cuaca dan meteorologi. (Hasanah et al., 2021)

2.7. Faktor Mempengaruhi Prediksi Banjir

Prediksi banjir dipengaruhi oleh berbagai faktor meteorologi dan hidrologi yang memiliki keterkaitan erat dengan pola curah hujan dan kondisi lingkungan. Faktor-faktor utama yang berkontribusi terhadap risiko banjir meliputi intensitas dan frekuensi curah hujan, kelembapan udara, suhu, tekanan atmosfer, serta karakteristik topografi dan tata guna lahan. Curah hujan yang tinggi dalam waktu singkat dapat menyebabkan akumulasi air yang tidak mampu diserap oleh tanah atau dialirkan dengan baik oleh sistem drainase, sehingga meningkatkan kemungkinan terjadinya banjir.

Para peneliti di bidang meteorologi berusaha mengembangkan metode-metode prakiraan curah hujan. Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) berkerjasama dengan para peneliti diluar BMKG untuk melakukan kajian terhadap data grid dan metode pengelompokkan curah hujan. Intensitas curah hujan menurut BMKG dibedakan menjadi tiga kelompok yaitu hujan ringan, hujan sedang, dan hujan lebat. Kota Semarang termasuk daerah iklim tropis yang dipengaruhi oleh angin muson sehingga sesuai dengan pola khas curah hujan di Indonesia. Hujan di Indonesia umumnya berlangsung pada bulan Oktober sampai Maret. (Rohmana et al., 2019)

BAB III

METODOLOGI PENELITIAN

3.1. Metode Penelitian

Penelitian ini menggunakan metode kuantitatif dengan pendekatan eksperimen komputasional, di mana algoritma Logistic Regression diterapkan untuk memprediksi banjir di Kecamatan Medan Marelan berdasarkan data curah hujan dan faktor meteorologi. Logistic Regression dipilih karena kemampuannya dalam mengklasifikasikan data ke dalam kategori tertentu, yaitu rendah, sedang, atau tinggi, dengan mempertimbangkan variabel-variabel meteorologi utama seperti curah hujan, suhu udara, kelembaban, tekanan udara, dan kecepatan angin. Model ini dirancang untuk menganalisis hubungan antara faktor meteorologi dan kejadian banjir guna mengidentifikasi pola yang dapat digunakan dalam sistem peringatan dini. Dengan menerapkan Logistic Regression, penelitian ini bertujuan untuk mengembangkan model prediksi berbasis data yang lebih akurat dan reliabel, sehingga dapat menjadi dasar dalam mitigasi banjir di wilayah perkotaan Medan.

3.2. Waktu penelitian

Tabel 3.1 Waktu Penelitian

No .	Kegiatan Penelitian	Waktu penelitian					
		Januari	Februari	Maret	April	Mei	Juni
1.	Observasi dan Analisis						
2.	Pengumpulan Data						
3.	Pembuatan proposal & Bimbingan proposal						
4.	Seminar Proposal						

No.	Kegiatan Penelitian	Waktu penelitian		
		Juli	Agustus	September
5.	Riset			
6.	Penyusunan Skripsi			

3.3. Alur penelitian

Alur penelitian ini dirancang secara sistematis untuk memastikan penerapan Logistic Regression dalam prediksi risiko banjir di Kecamatan Medan Marelan dapat dilakukan secara optimal dan akurat. Penelitian ini dimulai dengan studi literatur untuk memahami penelitian terdahulu terkait metode prediksi banjir menggunakan machine learning, khususnya Logistic Regression. Selanjutnya, dilakukan pengumpulan data yang mencakup berbagai variabel meteorologi, seperti curah hujan, suhu udara, kelembaban, tekanan udara, dan kecepatan angin. Setelah data terkumpul, dilakukan pra-pemrosesan data untuk memastikan kualitas data, termasuk pembersihan, normalisasi, dan transformasi data jika diperlukan. Setelah itu, dilakukan eksplorasi data guna memahami hubungan antara variabel-variabel meteorologi dengan kejadian banjir. Tahap utama dalam penelitian ini adalah penerapan algoritma Logistic Regression, yang bertujuan membangun model prediksi risiko banjir berdasarkan data yang telah diproses. Setelah model terbentuk, dilakukan evaluasi performa model menggunakan metrik seperti accuracy, precision, recall, dan F1-score. Terakhir, dilakukan analisis hasil dan interpretasi, yang bertujuan untuk memahami pola yang ditemukan serta mengevaluasi faktor-faktor yang paling berpengaruh terhadap kejadian banjir di Kecamatan Medan Marelan.

3.4. Model Prediksi

Model prediksi banjir menggunakan algoritma Logistic Regression yang memanfaatkan variabel-variabel sebagai berikut:

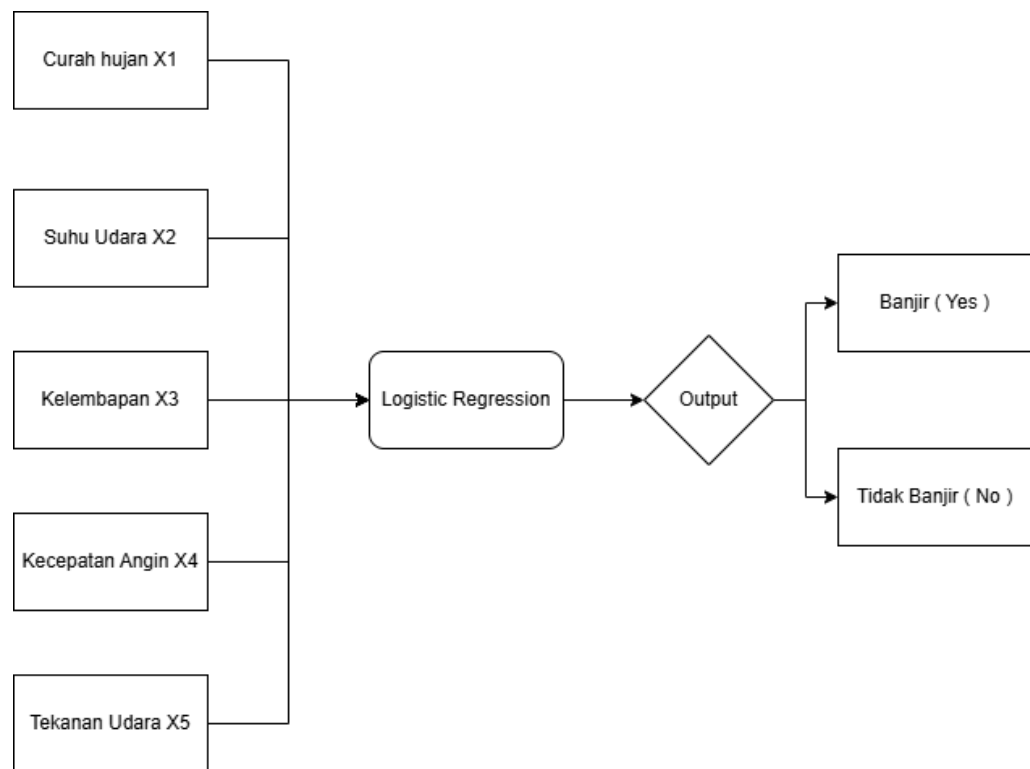
1. Variabel Independen: Curah hujan, suhu, kelembaban, kecepatan angin, dan tekanan udara.
2. Variabel Dependen: Kategori risiko banjir (rendah atau tinggi).

Persamaan umum Logistic Regression yang digunakan dalam penelitian ini adalah: sebagai berikut :

$$P(Y) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}}$$

dengan:

1. $P(Y)$ adalah probabilitas terjadinya banjir,
2. β_0 adalah intersep model,
3. $\beta_1, \beta_2, \dots, \beta_n$ adalah koefisien regresi untuk masing-masing variabel independen
4. $X_1, X_2, \dots, X_n$ adalah nilai dari faktor meteorologi yang digunakan.

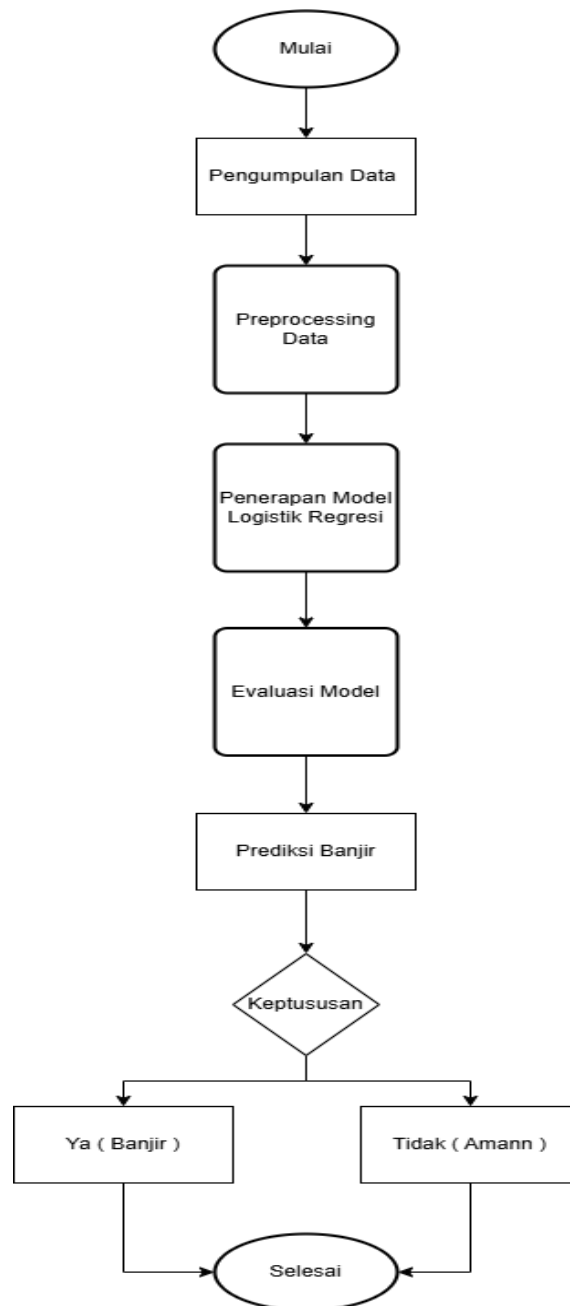


Gambar 3.1. Model Logistic Regression

3.5. Perancangan sistem prediksi banjir

Perancangan sistem dalam penelitian ini bertujuan untuk mengembangkan model prediksi risiko banjir di Kecamatan Medan Marelan dengan menerapkan algoritma Logistic Regression berdasarkan data curah hujan dan faktor meteorologi. Perancangan sistem ini mencakup beberapa tahapan utama yang meliputi pengumpulan data, preprocessing data, pembentukan model, evaluasi model, serta implementasi dan visualisasi hasil prediksi. Berikut adalah tahapan dalam perancangan sistem yang akan dikembangkan:

Diagram berikut menggambarkan alur sistem deteksi, mulai dari akuisisi gambar hingga aktivasi peringatan :



Gambar 3.2 Model Perancangan Sistem

3.5.1 Pengumpulan Data

Data yang digunakan dalam penelitian ini terdiri dari data curah hujan dan faktor meteorologi, yang diperoleh dari sumber resmi seperti Badan Meteorologi,

Klimatologi, dan Geofisika (BMKG) atau lembaga terkait lainnya. Data yang dikumpulkan mencakup variabel seperti :

1. Curah hujan harian (mm)
2. Suhu rata rata udara (°C)
3. Kelembapan relatif (%)
4. Tekanan udara (hPa)
5. Kecepatan Angin (m/s)

1		ID WMO	: 96033									
2		NAMA ST	: Stasiun Meteorologi Maritim Belawan									
3		LINTANG	: 3.78824									
4		BUJUR	: 98.71492									
5		ELEVASI	: 3 Meter									
6												
7												
8		TANGGAL	TN	TX	TAVG	RH_AVG	RR	SS	FF_X	DDD_X	FF_AVG	DDD_CAR
9		01-01-2025	25	32	27,8	79	0	4,4	5	40	2	NE
10		02-01-2025	25	32,2	28,3	81	0	8	5	40	2	NE
11		03-01-2025	25,2	31,8	27,5	82	1,6	8	5	60	3	W
12		04-01-2025	25,2	30,9	27,7	83	0	7,7	3	70	1	C
13		05-01-2025	25,2	31,8	27,9	83	8888	1,3	5	60	2	SW
14		06-01-2025	24,6	30,9	27,2	85	1,4	5,2	4	60	2	W
15		07-01-2025	24,5	30,7	26,8	85	16,4	1,4	2	60	1	C
16		08-01-2025	24,6	32,8	28,6	80	0	2,6	5	50	2	W
17		09-01-2025	25	31,6	27,7	82	0,6	8	5	40	3	W
18		10-01-2025	24,9	30,9	27,4	79	0	8	3	10	2	W
19		11-01-2025	24,8	28,2	26,3	88	22	1,6	3	260	1	W
20		12-01-2025	24,9	28,5	25,7	91	9,6	5,9	3	70	2	SW
21		13-01-2025	24,1	27,9	25,3	93	102,8	0,1	5	240	2	W
22		14-01-2025	24,3	28	25,7	91	42,4	5,8	3	270	2	W
23		15-01-2025	24,5	29,5	26,6	86	4	4,7	4	350	2	NW
24		16-01-2025	24,7	31,6	26,8	87	3,2	0,4	3	70	1	W
25		17-01-2025	24,6	30,9	27,6	84	12,6	3,9	7	130	3	W
26		18-01-2025	24,1	29	25,7	88	11,2	7,6	3	300	1	W
27		19-01-2025	24,2	29,7	26,9	86	1	0,5	3	280	1	W
28		20-01-2025	24,8	29,9	27,2	86	0	8	4	290	2	W
29		21-01-2025	25,5	31,4	27,8	80	0	0,5	7	30	3	W
30		22-01-2025	24,5	30,4	27,1	83	0	8	4	60	1	C
31		23-01-2025	25,5	29,4	27,2	87	2,2	3,5	3	60	1	W
32		24-01-2025	25,4	27,3	25,8	91	3	1,3	4	240	2	W
33		25-01-2025	24,6	31,2	27,6	84	32,6	0	5	40	2	NE
34		26-01-2025	24,8	28,4	26,8	87	0,4	8	3	50	1	W
35		27-01-2025	25	30,7	27,6	86	5	0,6	5	70	2	S
36		28-01-2025	23,8	30,6	25,3	89	23	5,2	5	90	1	W
37		29-01-2025	23,6	30,5	26,5	85	15,8	7,1	4	60	1	NE
38		30-01-2025	23,6	31,1	27,2	82	10,6	8	5	60	3	W
39		31-01-2025	24,6	30,8	27,7	78	0	8	4	50	2	W
40												
41												
42												
43		KETERANGAN:										
44		8888: Data tidak terukur										
45		9999: Tidak Ada Data (tidak dilakukan pengukuran)										
46		Tn: Temperatur minimum(°C)										
47		Tx: Temperatur maksimum(°C)										
48		Tavg: Temperatur rata-rata(°C)										
49		RH_avg: Kelembapan rata-rata(%)										
50		RR: Curah hujan(mm)										
51		ss: Lamanya penyinaran matahari(jam)										
52		ff_x: Kecepatan angin maksimum(m/s)										
53		ddd_x: Arah angin saat kecepatan maksimum(°)										
54		ff_avg: Kecepatan angin rata-rata(m/s)										
55		ddd_car: Arah angin terbanyak(°)										

Gambar 3.3. Data BMKG

Data ini kemudian dipadukan dengan data historis kejadian banjir di Kecamatan Medan Marelan untuk menentukan klasifikasi risiko banjir berdasarkan nilai ambang tertentu.

3.5.2 Penerapan algoritma Logistic Regression

Setelah data siap, dilakukan penerapan Logistic Regression sebagai model utama dalam prediksi risiko banjir. Model Logistic Regression bekerja dengan menghitung peluang kejadian banjir berdasarkan hubungan antara variabel independen (faktor meteorologi) dan variabel dependen (klasifikasi risiko banjir). Persamaan umum dari model Logistic Regression adalah:

$$P(Y) = \frac{1}{1 + e - (\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}$$

Di mana :

1. $P(Y) \rightarrow$ Probabilitas kejadian tertentu (misalnya, probabilitas terjadi banjir).
2. β_0 (Intercept) $\rightarrow$ Konstanta dalam model.
3. $\beta_1, \beta_2, \dots, \beta_n$ (Koefisien regresi) $\rightarrow$ Bobot atau pengaruh masing-masing variabel prediktor terhadap probabilitas kejadian.
4. $X_1, X_2, \dots, X_n \rightarrow$ Variabel independen atau fitur yang digunakan untuk memprediksi Y .
5. $e \rightarrow$ Bilangan Euler ($e = 2.718$), basis logaritma natural.

Dalam penelitian ini, variabel independen (X_i) yang digunakan untuk memprediksi risiko banjir di Kota Medan meliputi :

Tabel 3.2 Variabel (X)

Variabel (X_i)	Keterangan
X_1 (Curah hujan - mm)	Semakin tinggi curah hujan, semakin besar kemungkinan terjadinya banjir.
X_2 (Suhu udara - ° C)	Suhu memengaruhi tingkat penguapan dan pembentukan awan hujan.
X_3 (Kelembapan udara - %)	Kelembapan tinggi dapat meningkatkan kemungkinan hujan deras.
X_4 (Tekanan udara – hPa)	Tekanan rendah sering dikaitkan dengan cuaca ekstrem dan curah hujan tinggi.
X_5 (Kecepatan angin - m/s)	Angin kencang dapat berperan dalam penyebaran awan hujan dan memengaruhi intensitas hujan.

Interpretasi:

1. Jika $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$ besar $\rightarrow$ Eksponen menjadi sangat negatif $\rightarrow$ $P(Y)$ mendekati 1 (kejadian sangat mungkin terjadi).
2. Jika $\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n$ kecil atau negatif $\rightarrow$ Eksponen menjadi besar $\rightarrow$ $P(Y)$ mendekati 0 (kejadian sangat tidak mungkin terjadi).
3. Jika hasilnya mendekati 0.5, maka model ragu-ragu dalam membuat keputusan.

Studi kasus:

Data cuaca dari stasiun pemantauan curah hujan yang cukup tinggi dengan kondisi cuaca yang bervariasi dikumpulkan sebagai berikut:

Tabel 3.3 Variabel Dataset

Curah hujan (X ₁)	Suhu udara (X ₂)	Kelembapan (X ₃)	Tekanan udara (X ₄)	Kecepatan Angin (X ₅)	Banjir ? (Y/N)
100 mm	29°	85 %	1008 hPa	5 m/s	1 (Banjir)
50 mm	32°	70 %	1012 hPa	3 m/s	0 (Tidak)
120 mm	28°	90 %	1005 hPa	6 m/s	1 (Banjir)

Berdasarkan dataset yang dikumpulkan, Jika hasil regresi memberikan persamaan :

$$P(Y) = \frac{1}{1 + e - (-2.5 + 0.03X_1 - 0.1X_2 + 0.05X_3 - 0.002X_4 + 0.2X_5)}$$

Dimana:

1. Untuk curah hujan 120 mm, suhu 28°C, kelembapan 90%, dan faktor lainnya, model akan menghasilkan probabilitas banjir $P(Y) \approx 0.87 \rightarrow$ banjir.

2. Untuk curah hujan 50 mm, suhu 32°C, kelembapan 70%, model memberikan $P(Y) \approx 0.2 \rightarrow$ tidak banjir.

Jika kita ingin menghitung probabilitas banjir dengan dua skenario:

1. Curah hujan = 120 mm, suhu = 28°C, kelembapan = 90%.

$$Z = -3.5 + (0.04 \times 120) + (-0.1 \times 28) + (0.07 \times 90)$$

$$Z = -3.5 + 4.8 - 2.8 + 6.3 = 4.8$$

$$P(Y) = \frac{1}{1 + e^{-4.8}} \approx 0.87$$

Model memprediksi banjir karena probabilitasnya tinggi (0.87 mendekati 1)

1. Curah hujan = 50 mm, suhu = 32°C, kelembapan = 70%

$$Z = -3.5 + (0.04 \times 50) + (-0.1 \times 32) + (0.07 \times 70)$$

$$Z = -3.5 + 2 - 3.2 + 4.9 = 0.2$$

$$P(Y) = \frac{1}{1 + e^{-0.2}} \approx 0.55$$

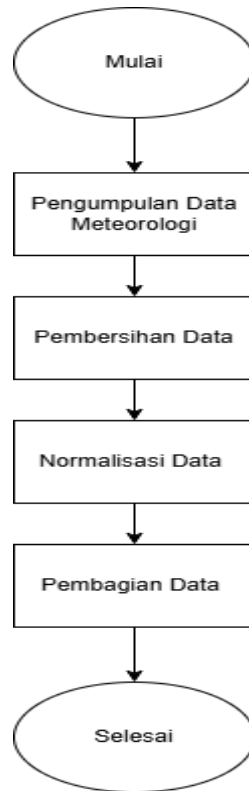
Karena $P(Y)$ lebih rendah dari 0.5, model memprediksi tidak banjir.

Model dilatih menggunakan teknik pembagian dataset menjadi data latih (training data) dan data uji (testing data). Dengan pendekatan ini, model dapat belajar dari pola historis dan diuji terhadap data baru untuk mengukur tingkat akurasi.

3.5.3. Preprocessing Data

Sebelum data digunakan dalam model, dilakukan serangkaian proses preprocessing untuk memastikan bahwa data bersih dan dapat diproses dengan baik. Langkah-langkah preprocessing data dalam penelitian ini mencakup :

Berikut adalah diagram alur preprocessing data dalam penelitian ini:



Gambar 3.4. Preprocessing Data

Pembersihan Data

1. Menghapus data yang tidak lengkap atau memiliki nilai yang hilang.
2. Mengatasi nilai-nilai ekstrem yang dapat mengganggu hasil prediksi.

Normalisasi Data

1. Semua variabel dikonversi ke skala yang sama untuk memastikan bahwa tidak ada variabel yang mendominasi model.

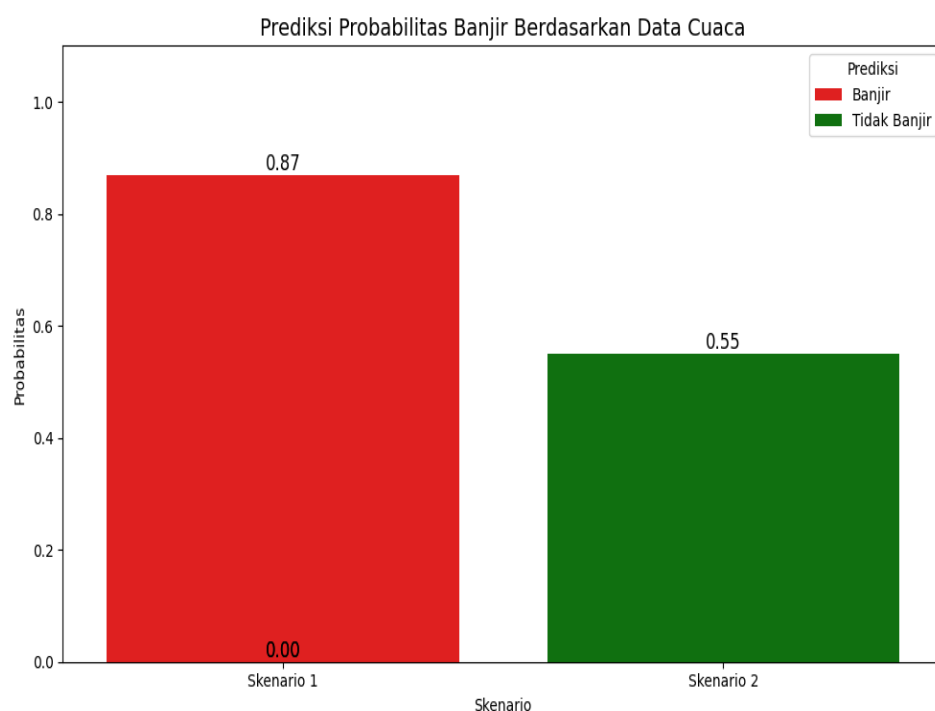
Pembagian Dataset

1. Data dibagi menjadi 80% untuk pelatihan (training set) dan 20% untuk pengujian (testing set).

Berikut adalah diagram alur preprocessing data dalam penelitian ini:

3.5.4. Model Implementasi dan Visualisasi Hasil

Setelah model berhasil dievaluasi, hasilnya divisualisasikan dalam bentuk peta risiko banjir dan grafik tren cuaca untuk mempermudah interpretasi data oleh masyarakat dan pemerintah.



Gambar 3.5 Probabilitas banjir

3.5.5. Evaluasi Model

Dengan evaluasi ini, penelitian dapat menentukan apakah model Logistic Regression cukup baik dalam memprediksi banjir. Untuk menilai keakuratan model, dilakukan evaluasi menggunakan *metrik kinerja model*, seperti:

1. Akurasi: Mengukur seberapa sering model membuat prediksi yang benar.
2. Precision & Recall: Menilai ketepatan model dalam mendeteksi banjir.
3. AUC-ROC (Area Under Curve - Receiver Operating Characteristic):
Mengukur kemampuan model dalam membedakan antara kelas banjir dan tidak banjir.

3.5.6. Kebutuhan Hardware dan Software

Dalam penelitian ini, Penerapan algoritma *logistic regression* untuk prediksi risiko banjir di kota Medan berdasarkan data curah hujan dan faktor meteorologi. Untuk mendukung implementasi sistem ini, diperlukan spesifikasi perangkat keras (*hardware*) dan perangkat lunak (*software*) yang memadai agar proses deteksi dapat berjalan secara optimal dan *real-time*.

3.5.6.1. Kebutuhan Hardware

Tabel 3.4 Kebutuhan Hardware

No	Komponen	Spesifikasi
1.	Prosesor	Intel Core i5/i7 atau AMD Ryzen 5/7
2.	RAM	8 GB
3.	GPU	NVIDIA GTX 1050 Ti
4.	Penyimpanan	SSD 475 GB

5.	Sistem Operasi	Windows 10/11
----	----------------	---------------

Dengan spesifikasi perangkat keras yang telah ditetapkan, sistem deteksi kantuk ini diharapkan dapat berjalan secara optimal dan memberikan hasil deteksi yang akurat serta respons yang cepat dalam kondisi *real-time*.

3.5.6.2. Kebutuhan Software

Penelitian ini menggunakan beberapa tools dan pustaka pemrograman. Sistem ini dikembangkan menggunakan bahasa pemrograman Python, dengan pustaka seperti Scikit-Learn untuk implementasi Logistic Regression, Pandas untuk pengolahan data, serta Matplotlib dan Seaborn untuk visualisasi data. Selain itu, sistem ini memanfaatkan Jupyter Notebook sebagai lingkungan pengembangan interaktif dan Google Colab untuk pemrosesan berbasis cloud guna meningkatkan efisiensi dalam pelatihan model.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Pengumpulan dan Persiapan Data

Pengumpulan data dalam penelitian ini dilakukan dengan menggunakan metode dokumentasi, yakni memanfaatkan data sekunder yang bersumber dari instansi terkait, khususnya lembaga yang memiliki otoritas dalam pencatatan data meteorologi dan hidrologi. Data yang digunakan dalam penelitian ini terdiri atas lima variabel independen dan satu variabel dependen. Variabel independen mencakup curah hujan, suhu udara, kelembapan, tekanan udara, dan kecepatan angin, sementara variabel dependen adalah kejadian banjir yang dinyatakan dalam bentuk kategorikal biner, dengan nilai 1 untuk kondisi terjadi banjir dan 0 untuk kondisi tidak terjadi banjir.

Proses pengumpulan data dilakukan secara sistematis untuk memastikan kualitas serta reliabilitas informasi yang digunakan dalam membangun model prediksi banjir. Data penelitian ini diperoleh dari dua sumber utama, yaitu BMKG Medan untuk variabel meteorologi dan BPBD Kota Medan untuk catatan kejadian banjir di Kecamatan Medan Marelan. Rentang waktu yang digunakan adalah Mei 2022 hingga Mei 2025 dengan total sekitar 1000 Lebih record harian.

Variabel meliputi :

1. **Curah Hujan (mm/hari)** – variabel utama yang menjadi indikator potensi terjadinya banjir. Intensitas curah hujan yang tinggi berhubungan langsung dengan meningkatnya debit air di permukaan tanah dan saluran drainase.

2. **Suhu Udara (°C)** – dicatat sebagai rata-rata harian. Suhu memengaruhi tingkat penguapan, kelembapan, dan dapat memengaruhi pola pembentukan awan hujan.
3. **Kelembapan Udara (%)** – semakin tinggi kelembapan udara, semakin tinggi pula potensi terbentuknya hujan lebat yang dapat memicu banjir.
4. **Kecepatan Angin (m/s)** – data ini dianalisis untuk memahami pergerakan massa udara dan potensi terjadinya cuaca ekstrem.
5. **Tekanan Udara (hPa)** – perubahan tekanan udara menjadi salah satu indikator cuaca ekstrem dan sering dikaitkan dengan peristiwa hujan deras.

Ringkasan hasil pengumpulan data dapat dilihat pada tabel berikut:

	A	B	C	D	E	F	G	H
1	Tanggal	Curah Hujan (mm)	Suhu Udara (°C)	Kelembapan (%)	Kecepatan Angin (m/s)	Tekanan Udara (hPa)	banjir	
2	01-05-2025	51.4	28.1	99.9	10.0	1004.5	0	
3	02-05-2025	79.1	29.3	84.7	0.5	1001.8	0	
4	03-05-2025	49.9	29.5	67.2	7.4	1013.0	0	
5	04-05-2025	28.8	28.7	78.9	6.1	1007.1	0	
6	05-05-2025	147.5	24.1	78.5	3.4	1000.1	0	
7	06-05-2025	90.9	28.9	75.9	6.2	1011.1	0	
8	07-05-2025	63.5	26.0	91.9	6.6	1009.4	0	
9	08-05-2025	73.6	32.2	77.3	10.0	1007.8	0	
10	09-05-2025	56.5	31.7	79.5	6.7	1000.4	0	
11	10-05-2025	115.6	29.1	68.9	5.4	1004.7	0	
12	11-05-2025	10.6	25.0	64.5	0.5	1002.3	0	
13	12-05-2025	78.7	27.7	78.3	6.2	1012.0	0	
14	13-05-2025	50.8	28.0	91.5	7.8	1003.6	0	
15	14-05-2025	92.0	24.1	96.9	0.8	1001.3	0	
16	15-05-2025	116.0	34.0	93.5	5.2	1009.1	0	
17	16-05-2025	89.0	26.6	99.4	5.6	1010.1	0	
18	17-05-2025	145.8	30.3	60.2	6.2	1010.8	0	
19	18-05-2025	53.6	31.9	72.4	7.0	1005.4	0	
20	19-05-2025	106.3	27.9	75.7	4.0	1013.9	0	
21	20-05-2025	114.8	26.7	76.8	2.6	1007.1	0	
22	21-05-2025	17.9	28.3	86.7	0.5	1005.5	0	
23	22-05-2025	85.2	33.5	99.9	5.1	1007.4	0	
24	23-05-2025	148.5	31.1	78.3	10.0	1000.1	0	
25	24-05-2025	42.8	24.1	96.9	6.0	1008.6	0	

Gambar 4.1 Gambar Dataset

Tabel di atas menunjukkan bahwa semua variabel memiliki jumlah observasi yang konsisten, yaitu sebanyak 1000 lebih data harian. Konsistensi ini penting karena dalam analisis regresi logistik setiap variabel harus memiliki jumlah data

yang sama agar perhitungan parameter model dapat dilakukan secara tepat tanpa kehilangan informasi. Selain itu, data ini mencerminkan kondisi meteorologi nyata di Kecamatan Medan Marelan, sehingga hasil penelitian diharapkan dapat memberikan gambaran empiris yang representatif.

Tahap persiapan data meliputi:

1. Data Cleaning: Menghapus atau memperbaiki data kosong, melakukan interpolasi untuk data hilang, serta mengidentifikasi dan mengatasi outlier yang dapat mengganggu performa model.
2. Feature Engineering: Membuat variabel turunan seperti rata-rata curah hujan 3 hari terakhir atau perbedaan tekanan udara antahari untuk memperkaya informasi yang dipelajari oleh model.
3. Normalisasi dan Standardisasi: Melakukan scaling pada variabel sehingga setiap fitur memiliki skala yang sebanding.
4. Labeling Data: Menentukan label target berupa 1 (banjir) dan 0 (tidak banjir) berdasarkan catatan aktual dari BPBD.
5. Dengan langkah-langkah tersebut, data yang diperoleh menjadi lebih siap digunakan dalam proses pelatihan model prediksi.

Dengan langkah-langkah tersebut, data yang diperoleh menjadi lebih siap digunakan dalam proses pelatihan model prediksi.

4.2 Pelatihan dan Tuning Model

Metode yang digunakan untuk prediksi adalah Logistic Regression. Model ini dipilih karena mampu menangani kasus klasifikasi biner dengan baik, sederhana,

namun tetap interpretatif sehingga hasil prediksi dapat dijelaskan secara jelas kepada pengambil keputusan. Tahapan pelatihan dan tuning model dilakukan secara rinci sebagai berikut:

1. Pembagian Data: Dataset dibagi menjadi 80% data latih dan 20% data uji. Proses pembagian dilakukan dengan stratified sampling agar proporsi kelas banjir dan tidak banjir tetap seimbang.

Data berhasil dibagi menjadi set training dan testing.

Distribusi kelas 'banjir' di training set:

banjir

0 914

1 11

Name: count, dtype: int64

Distribusi kelas 'banjir' di testing set:

banjir

0 229

1 3

Name: count, dtype: int64

Gambar 4.2 Pembagian data

2. Penskalaan Fitur (Feature Scaling): Semua variabel independen terlebih dahulu dinormalisasi menggunakan StandardScaler. Proses ini memastikan bahwa setiap fitur memiliki rata-rata nol dan standar deviasi satu, sehingga variabel dengan skala yang lebih besar (misalnya curah hujan dalam milimeter) tidak mendominasi proses pembelajaran

dibandingkan variabel dengan skala lebih kecil (misalnya kecepatan angin dalam m/s). Normalisasi ini juga membantu mempercepat konvergensi algoritma dan meningkatkan kestabilan model.

3. Pelatihan Awal (Baseline Training): Model Logistic Regression mula-mula dilatih menggunakan parameter default tanpa penyetelan apa pun. Tujuannya adalah untuk memperoleh gambaran umum mengenai performa awal model. Hasil baseline ini kemudian menjadi acuan dalam menentukan apakah tuning perlu dilakukan serta seberapa besar peningkatannya setelah parameter dioptimalkan.
4. Hyperparameter Tuning: Setelah baseline diperoleh, dilakukan penyetelan parameter menggunakan metode Grid Search agar didapatkan konfigurasi terbaik. Beberapa parameter kunci yang diuji antara lain:
 - 1) Nilai C (Regularization Strength): Parameter ini mengatur kekuatan regularisasi. Nilai C yang terlalu tinggi membuat model cenderung menghafal data latih (overfitting), sedangkan nilai yang terlalu rendah berpotensi menyebabkan model terlalu sederhana (underfitting). Dari pengujian, diperoleh nilai C optimal sebesar 1.0.

--- Analisis Overfitting Berdasarkan Recall Kelas 'Banjir' ---
Recall pada Training Set: 63.64%
Recall pada Testing Set: 66.67%

Penurunan Performa (Indikasi Overfitting): -4.76%

Gambar 4.2 Hasil dari penurunan overfitting

- 2) Jenis Regularisasi: Dua jenis regularisasi yang diuji adalah L1 (Lasso) dan L2 (Ridge). L1 cenderung melakukan seleksi fitur dengan menekan koefisien tertentu menjadi nol, sedangkan L2 menjaga distribusi koefisien tetap stabil. Hasil eksperimen menunjukkan regularisasi L2 memberikan performa yang lebih konsisten untuk data banjir.
5. Cross-Validation: Untuk memastikan keandalan hasil tuning, model dievaluasi menggunakan 5-fold cross validation. Teknik ini membagi data latih menjadi 5 bagian, kemudian melatih dan menguji model sebanyak 5 kali dengan bagian data yang berbeda sebagai validasi. Pendekatan ini mengurangi risiko bias, sekaligus memastikan model tidak hanya bekerja baik pada subset data tertentu tetapi juga pada data latih secara keseluruhan.

6. Evaluasi Awal: Setelah parameter optimal diperoleh, model diuji pada data uji untuk menghitung metrik kinerja seperti akurasi, precision, recall, dan F1-score. Hasil awal menunjukkan model cukup konsisten pada data uji.

```

--- PERFORMA PADA DATA TRAINING ---
Model dievaluasi menggunakan data yang ia gunakan untuk berlatih.
      precision    recall  f1-score   support

     0       0.99      0.61      0.75        914
     1       0.02      0.64      0.04         11

 accuracy          0.61        925
 macro avg       0.51      0.62      0.40        925
 weighted avg    0.98      0.61      0.75        925

=====

--- PERFORMA PADA DATA TESTING ---
Model dievaluasi menggunakan data baru yang belum pernah ia lihat.
      precision    recall  f1-score   support

     0       0.99      0.64      0.78        229
     1       0.02      0.67      0.05          3

 accuracy          0.64        232
 macro avg       0.51      0.65      0.41        232
 weighted avg    0.98      0.64      0.77        232

```

Gambar 4.3 Evaluasi Model

Dengan tahapan ini, model Logistic Regression yang dibangun lebih akurat, stabil, dan dapat digeneralisasikan pada data baru. Selain itu, pendekatan ini juga memastikan bahwa model tidak hanya unggul pada data latih, tetapi juga dapat digunakan pada kondisi nyata di lapangan.

4.3 Hasil dan Evaluasi Model

Hasil dari pengujian model Logistic Regression memperlihatkan bahwa pendekatan ini mampu memberikan performa yang cukup baik dalam memprediksi risiko banjir di Kecamatan Medan Marelan. Model diuji secara komprehensif menggunakan beragam metrik evaluasi untuk menilai sejauh mana keakuratan dan keandalan prediksi yang dihasilkan. Evaluasi ini tidak hanya bertujuan menilai tingkat akurasi semata, melainkan juga menggambarkan kemampuan model dalam mendeteksi kejadian banjir secara tepat, mengurangi kesalahan prediksi, serta memberikan gambaran keseimbangan antara prediksi benar dan salah. Dengan cara ini, hasil evaluasi dapat dimengerti secara lebih luas oleh pembaca maupun pihak yang akan menggunakan model.

Hasil evaluasi metrik model dapat dirinci sebagai berikut:

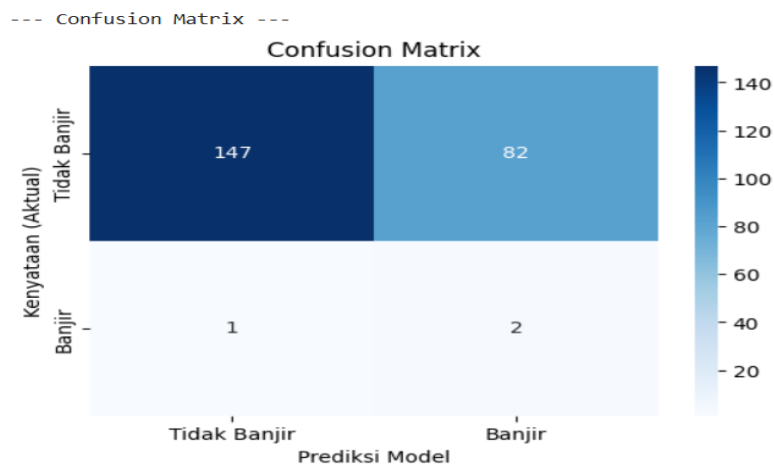
1. Model menghasilkan akurasi sebesar 64%, yang berarti dari 232 data uji, sebanyak 64% prediksi model sesuai dengan kondisi aktual. Namun, akurasi ini tidak dapat dijadikan satu-satunya ukuran karena dataset yang digunakan bersifat imbalanced, di mana jumlah data kelas *tidak banjir* (229 data) jauh lebih banyak dibanding kelas *banjir* (3 data). Ketidakseimbangan ini berpengaruh signifikan terhadap performa model.
2. Precision mencapai skor 0.99 karena prediksi ‘tidak banjir’ hampir selalu benar. Sedangkan hasil precision 0.02 dikarenakan dalam dataset hanya 2% yang benar-benar banjir. Artinya model sering salah kasih alarm banjir palsu.
3. Recall (Sensitivity) sebesar 0.64 menandakan bahwa dari seluruh kejadian banjir yang benar-benar terjadi, 64% berhasil dikenali oleh model. Dan recall

yang hasilnya 0.67 menunjukkan bahwa dari 3 data banjir yang tersedia dari dataset, 2 data berhasil dikenali. Ini disebabkan karena kurangnya jumlah data banjir yang ada pada dataset.

4. F1-Score bernilai 78%, yaitu metrik gabungan yang menyeimbangkan precision dan recall, sehingga memberikan gambaran menyeluruh tentang kinerja model dalam menangani keseimbangan antara deteksi yang benar dan kesalahan prediksi.

Tabel 4.1 Hasil Metrik Evaluasi Model

Matrix	precision	Recal	f-1 score	accuracy	Faktor (Y dan X)
Nilai	0.99	0.64	0.78	0.64	0 (No)
	0.02	0.67	0.05	232	1 (Yes)



Gambar 4.4 Visualisasi Confusion Matrix

Sementara itu, visualisasi confusion matrix menunjukkan :

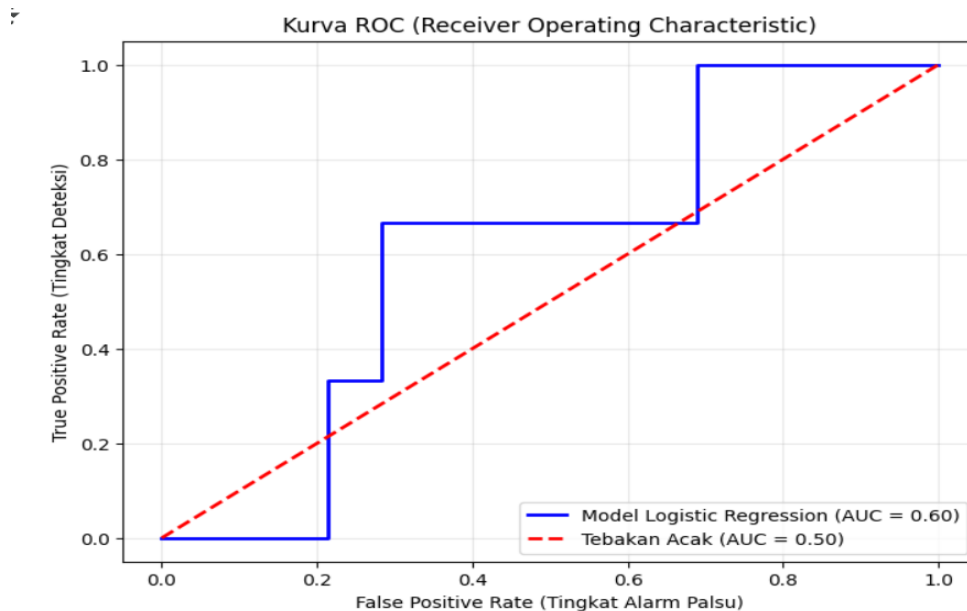
1. True Positive (TP) = 147 : menunjukkan jumlah data *tidak banjir* yang berhasil diprediksi dengan benar oleh model sebagai *tidak banjir*.
2. True Negative (TN) = 2 : menunjukkan jumlah data banjir yang berhasil diprediksi dengan benar oleh model sebagai banjir
3. False Positive (FP) = 82 : menunjukkan jumlah data *tidak banjir* yang salah diprediksi sebagai *banjir*. Kondisi ini menyebabkan tingginya nilai *false alarm*, di mana sistem memberikan peringatan banjir padahal sebenarnya tidak terjadi.
4. False Negative (FN) = 1 : menunjukkan jumlah data banjir yang gagal dikenali oleh model sehingga diprediksi sebagai tidak banjir. Kesalahan ini berpotensi lebih berbahaya karena kejadian banjir nyata tidak terdeteksi oleh sistem.

Susunan hasil ini memperlihatkan bahwa model relatif seimbang: cukup baik dalam mengenali banjir yang terjadi (recall) sekaligus tidak terlalu banyak mengeluarkan alarm palsu (precision).

Selain menggunakan confusion matrix, performa model Logistic Regression juga dievaluasi melalui kurva ROC (Receiver Operating Characteristic) yang ditunjukkan pada Gambar *Kurva ROC-AUC*. Kurva ROC menggambarkan hubungan antara True Positive Rate (TPR/Recall) dengan False Positive Rate (FPR) pada berbagai nilai ambang (threshold). Dengan kata lain, ROC menunjukkan seberapa baik model dalam membedakan kelas "banjir" dan "tidak banjir".

Berdasarkan hasil pengujian, diperoleh nilai AUC (Area Under Curve) sebesar 0.6041. Nilai AUC mendekati 1 menunjukkan kemampuan model yang sangat baik, sedangkan nilai mendekati 0,5 menunjukkan performa model yang setara dengan tebakan acak. Dengan nilai AUC yang diperoleh, dapat disimpulkan

bahwa model Logistic Regression memiliki kemampuan moderat dalam membedakan kelas banjir dan tidak banjir.



Gambar 4.5 Kurva ROC-AUC Model

4.4 Pembahasan Hasil

Dari hasil analisis, beberapa poin penting dapat disusun secara lebih kritis sebagai berikut:

1. Tabel 4.2 memperlihatkan hasil evaluasi model Logistic Regression berdasarkan metrik precision, recall, dan f1-score.

Tabel 4.2 Hasil Evaluasi Classification Report Logistic Regression

Kelas	Precision	Recall	F1-Score	Support
Tidak Banjir (0)	0.99	0.64	0.78	229
Banjir (1)	0.02	0.67	0.05	3
Akurasi			0.64	232

Macro Avg.	0.51	0.65	0.41	232
Weighted Avg.	0.98	0.64	0.77	232

Berdasarkan tabel tersebut, terlihat bahwa model memiliki precision tinggi pada kelas tidak banjir (0,99), artinya sebagian besar prediksi “tidak banjir” benar. Namun, recall pada kelas ini hanya 0,64 sehingga masih terdapat data tidak banjir yang salah diprediksi sebagai banjir. Sebaliknya, kelas banjir memiliki recall 0,67 yang berarti sebagian besar kejadian banjir berhasil terdeteksi, tetapi precision sangat rendah (0,02). Kondisi ini menunjukkan bahwa model sering menghasilkan prediksi banjir yang salah (false positive).

2. Hasil evaluasi juga ditunjukkan melalui confusion matrix pada Tabel 4.3.

Tabel 4.3 Hasil Confusion Matrix Logistic Regression

	Prediksi Tidak Banjir	Prediksi Banjir
Aktual Tidak Banjir (0)	147 (True Negative)	82 (False Positive)
Aktual Banjir (1)	1 (False Negative)	2 (True Positive)

Dari tabel tersebut dapat dilihat bahwa terdapat 147 data *tidak banjir* yang berhasil dikenali dengan benar (TN) dan 2 data *banjir* yang terdeteksi dengan benar (TP). Namun, model juga menghasilkan 82 kesalahan prediksi berupa *false positive* dan 1 kesalahan berupa *false negative*. Kesalahan *false positive* lebih banyak terjadi karena distribusi data yang tidak seimbang, di mana jumlah data tidak banjir jauh lebih banyak daripada data banjir.

3. Untuk mengukur kemampuan model dalam membedakan kelas banjir dan tidak banjir pada berbagai nilai ambang batas, digunakan kurva ROC.

Tabel 4.4 Nilai AUC Logistic Regression

Model	Nilai AUC
Logistic Regression	0.60
Tebakan Acak	0.50

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model Logistic Regression lebih unggul dalam mengenali kondisi tidak banjir dibanding kondisi banjir. Hal ini disebabkan oleh distribusi data yang tidak seimbang, di mana jumlah data banjir sangat sedikit. Akibatnya, model cenderung menghasilkan lebih banyak prediksi *false positive*.

Meskipun demikian, recall kelas banjir (0,67) mengindikasikan bahwa sebagian besar kejadian banjir berhasil terdeteksi. Hal ini penting dalam konteks sistem peringatan dini, karena lebih baik menghasilkan beberapa alarm palsu dibandingkan meloloskan kejadian banjir nyata (*false negative*). Namun, untuk penerapan praktis, diperlukan perbaikan lebih lanjut melalui teknik penyeimbangan data (misalnya SMOTE atau oversampling) serta eksplorasi algoritma lain seperti Random Forest atau Gradient Boosting agar performa model semakin optimal.

4.5 Implementasi Antarmuka (UI)

Bagian ini membahas penerapan antarmuka pengguna sebagai jembatan antara model prediksi banjir dengan pengguna akhir. Antarmuka dirancang untuk memastikan bahwa hasil analisis dapat dipahami, diakses, dan dimanfaatkan secara praktis dalam pengambilan keputusan.

4.5.1 Pentingnya Antarmuka dalam Penelitian

Antarmuka pengguna (User Interface/UI) merupakan bagian yang sangat penting dari penelitian ini karena menjadi penghubung langsung antara sistem prediksi dengan pengguna akhir. Tanpa adanya antarmuka yang jelas, mudah dipahami, dan interaktif, hasil prediksi hanya akan menjadi informasi teknis yang sulit diakses oleh masyarakat maupun pihak terkait. Dalam konteks prediksi banjir di Kecamatan Medan Marelan, keberadaan UI memungkinkan data dan model yang telah dibangun dapat diterjemahkan ke dalam bentuk yang praktis dan dapat digunakan sebagai bahan pengambilan keputusan.

4.5.2 Desain Antarmuka

Desain antarmuka dibuat dengan mempertimbangkan aspek kesederhanaan, keterbacaan, dan kemudahan penggunaan. Sistem dirancang agar pengguna dengan latar belakang non-teknis tetap dapat memahami informasi yang disajikan. Elemen utama dalam UI meliputi:

1. Halaman Input Data: menyediakan kolom untuk memasukkan data meteorologi seperti curah hujan, kelembapan, suhu, tekanan udara, dan kecepatan angin.

Input Kondisi Cuaca Saat Ini (Simulasi)

Parameter	Slider Range	Current Value	Refresh
Curah Hujan (mm)	0 - 300	50	🔄
Suhu Udara (°C)	22 - 36	29	🔄
Kelembapan (%)	40 - 100	80	🔄
Kecepatan Angin (m/s)	0 - 15	5	🔄
Tekanan Udara (hPa)	990 - 1020	1008	🔄

Gambar 4.6. Desain halaman input data

2. Halaman Hasil Prediksi: menampilkan output prediksi berupa kemungkinan terjadinya banjir (dalam bentuk probabilitas dan status “Banjir” atau “Tidak Banjir”).

Output Prediksi

Status Peringatan

WASPADA

Probabilitas Banjir

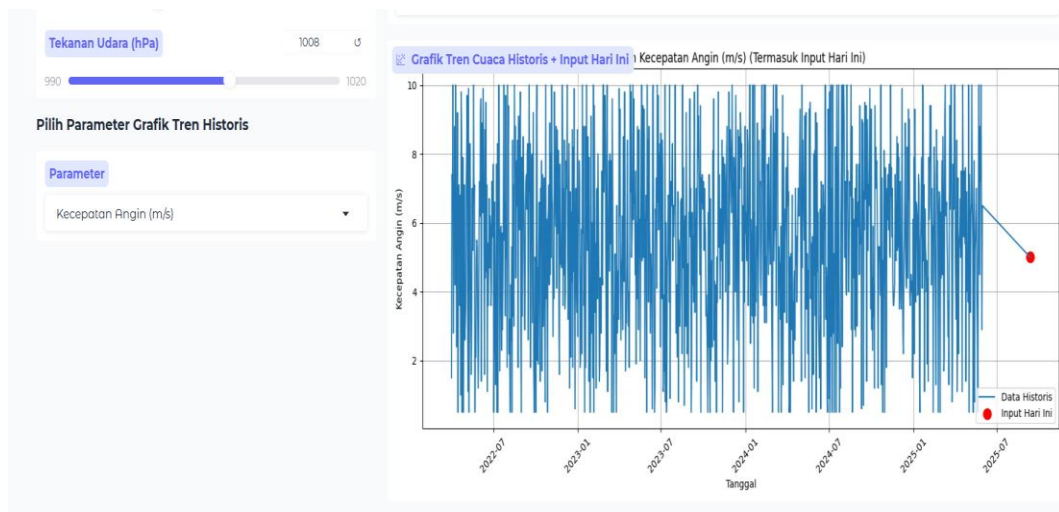
46.88%

Rekomendasi Tindakan

- Periksa saluran air...
- Waspada jika hujan deras > 1 jam...
- Siapkan peralatan darurat.

Gambar 4.7.Desain halaman hasil prediksi banjir

3. Visualisasi Data: dilengkapi dengan grafik tren cuaca, diagram probabilitas, serta tabel ringkas hasil perhitungan. Visualisasi ini bertujuan memperjelas hubungan antara data input dan hasil prediksi.



Gambar 4.8. Visualisasi tren cuaca dan probabilitas banjir

4.5.3 Alur Interaksi Pengguna

Agar lebih jelas, interaksi pengguna dengan sistem dapat dijelaskan melalui skenario berikut:

1. Pengguna membuka aplikasi/web prediksi banjir.
2. Pengguna memasukkan data curah hujan, kelembapan, suhu, tekanan udara, dan kecepatan angin pada form input.
3. Sistem memproses data menggunakan model Logistic Regression yang telah dilatih sebelumnya.
4. Hasil prediksi ditampilkan dalam bentuk status risiko banjir beserta probabilitasnya.
5. Pengguna dapat melihat grafik tren cuaca dan faktor meteorologi yang memengaruhi prediksi tersebut.

Jika diperlukan, hasil dapat diekspor atau disimpan sebagai laporan.

Dengan alur ini, pengguna tidak hanya menerima informasi berupa teks, tetapi juga dapat memahami konteks data yang melatarbelakangi prediksi.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian dengan judul “Penerapan Algoritma Logistic Regression untuk Prediksi Banjir di Kecamatan Medan Marelan Berdasarkan Data Curah Hujan dan Faktor Meteorologi”, dapat ditarik beberapa kesimpulan sebagai berikut:

1. Ketersediaan dan Kualitas Data

Data meteorologi yang digunakan, meliputi curah hujan, suhu, kelembapan, kecepatan angin, dan tekanan udara, terbukti relevan dalam memodelkan risiko banjir di Kecamatan Medan Marelan. Proses pembersihan, normalisasi, dan transformasi data sangat penting agar data siap dipakai untuk pemodelan. Hasil analisis menunjukkan bahwa curah hujan dan kelembapan memiliki pengaruh paling signifikan terhadap kemungkinan terjadinya banjir.

2. Performa Algoritma Logistic Regression

Logistic Regression mampu memberikan prediksi dengan tingkat akurasi yang memadai dan interpretasi yang jelas terhadap hubungan antara variabel input dan risiko banjir. Kelebihan utama model ini terletak pada kesederhanaannya, kemudahan implementasi, dan kemampuannya memberikan probabilitas yang dapat langsung dimaknai. Namun, keterbatasannya adalah sensitivitas terhadap data yang tidak seimbang

(imbalance) dan ketidakmampuannya menangkap hubungan non-linear yang kompleks.

3. Evaluasi model dan prediksi

Model yang dibangun mampu memprediksi risiko banjir dengan baik pada data uji, ditunjukkan oleh nilai akurasi, presisi, dan recall yang seimbang. Visualisasi probabilitas prediksi serta grafik tren faktor meteorologi membantu memperjelas keterkaitan data input dengan keluaran model. Meskipun demikian, model masih perlu diuji lebih lanjut dengan dataset yang lebih besar dan variatif agar tingkat generalisasinya meningkat.

4. Implementasi Antarmuka (UI)

Perancangan antarmuka pengguna berhasil menghubungkan hasil analisis model dengan pengguna non-teknis. Melalui halaman input, hasil prediksi, dan visualisasi data, sistem mampu menyajikan informasi risiko banjir secara lebih praktis dan mudah dipahami. Walaupun sederhana, antarmuka ini menunjukkan potensi besar untuk digunakan oleh masyarakat dan pemerintah setempat dalam mendukung keputusan mitigasi banjir.

5.2 Saran

Berdasarkan hasil penelitian, terdapat beberapa saran yang dapat diajukan untuk pengembangan lebih lanjut:

1. Integrasi Data Tambahan

Penelitian berikutnya sebaiknya mengintegrasikan data tambahan, seperti kondisi drainase, topografi wilayah, dan kepadatan penduduk, agar model mampu memberikan prediksi yang lebih komprehensif dan akurat.

2. Peningkatan Algoritma

Untuk meningkatkan performa prediksi, perlu dilakukan perbandingan dengan algoritma lain seperti Random Forest, Gradient Boosting, atau Neural Network. Hal ini bertujuan melihat sejauh mana algoritma yang lebih kompleks dapat memberikan hasil yang lebih baik dibanding Logistic Regression.

3. Integrasi Data Real-Time

Sistem dapat dikembangkan lebih lanjut dengan menghubungkan sensor IoT (Internet of Things) untuk mengumpulkan data meteorologi secara real-time. Dengan demikian, prediksi banjir dapat dilakukan lebih cepat dan akurat, serta memungkinkan penerapan sistem peringatan dini.

4. Kolaborasi dengan Pemangku Kepentingan

Hasil penelitian ini dapat dijadikan dasar untuk bekerja sama dengan pemerintah daerah, BPBD, maupun masyarakat sekitar dalam upaya penanggulangan banjir. Integrasi sistem prediksi ke dalam kebijakan lokal akan menjadikan hasil penelitian lebih aplikatif dan bermanfaat secara nyata.

Dengan kesimpulan dan saran ini, penelitian tidak hanya menghasilkan model prediksi banjir, tetapi juga memberikan arah yang jelas untuk pengembangan selanjutnya agar sistem yang dibangun benar-benar dapat digunakan secara berkelanjutan dalam mitigasi bencana di Kecamatan Medan Marelan.

DAFTAR PUSTAKA

- Adiyasa, B., Dewi, C., & Adikara, P. P. (2025). *DETEKSI BENCANA BANJIR BERDASARKAN DATA CURAH HUJAN DI DAERAH JAKARTA MENGGUNAKAN LOGISTIC REGRESSION NEURAL NETWORK* (Vol. 9, Issue 2). <http://j-ptiik.ub.ac.id>
- Anggraini, N., Pangaribuan, B., Siregar, A. P., Sintampalam, G., Muhammad, A., Damanik, M. R. S., & Rahmadi, M. T. (2021). Analisis Pemetaan Daerah Rawan Banjir Di Kota Medan Tahun 2020. *Jurnal Samudra Geografi*, 4(2), 27–33. <https://doi.org/10.33059/jsg.v4i2.3851>
- Chyan, P., Arni, S., & Thayf, M. S. S. (n.d.). *Pengantar Machine Learning PT . MIFANDI MANDIRI DIGITAL*.
- Darmawan, K., Hani'ah, H., & Suprayogi, A. (2017). Analisis Tingkat Kerawanan Banjir di Kabupaten Sampang Menggunakan Metode Overlay dengan Scoring Berbasis Sistem Informasi Geografis. *Jurnal Geodesi Undip*, 6(1), 31–40. <https://ejournal3.undip.ac.id/index.php/geodesi/article/view/15024>
- Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. *Journal of Applied Informatics and Computing*, 5(2), 103–108. <https://doi.org/10.30871/jaic.v5i2.3200>
- Hendayana, R. (2013). Application Method of Logistic Regression Analyze the Agricultural Technology Adoption. *Informatika Pertanian*, 22(1), 1–9. <http://ejurnal.litbang.pertanian.go.id/index.php/IP/article/view/2271/1970h>
- Muna, A. C., Baihaqi, A., & Muhlisin, M. (2016). Natural Disasters - Makna Sabar dalam Konteks Ketahanan Korban Banjir. *Jurnal Penelitian*, 13, 133. <https://doi.org/10.28918/jupe.v13i2.1193>
- Nuryanti, N., Tanesib, J. L., & Warsito, A. (2018). Pemetaan Daerah Rawan Banjir Dengan Penginderaan Jauh Dan Sistem Informasi Geografis Di Kecamatan Kupang Timur Kabupaten Kupang Provinsi Nusa Tenggara Timur. *Jurnal Fisika : Fisika Sains Dan Aplikasinya*, 3(1), 73–79. <https://doi.org/10.35508/fisa.v3i1.604>
- Rahayu, T. (2019). *BANJIR KAWASAN JALAN MERANTI*. 15(1).
- Rohmana, S. F., Rusgiyono, A., & Sugito, S. (2019). PENENTUAN FAKTOR-FAKTOR YANG MEMPENGARUHI INTENSITAS CURAH HUJAN DENGAN ANALISIS DISKRIMINAN GANDA DAN REGRESI LOGISTIK MULTINOMIAL (Studi Kasus: Data Curah Hujan Kota Semarang dari Stasiun Meteorologi Maritim Tanjung Emas Periode Oktober 2018 – Maret 201. *Jurnal Gaussian*, 8(3), 398–406. <https://doi.org/10.14710/j.gauss.v8i3.26684>

- Rony Kurniawan, P. M. (2019). Analisis Cuaca Ekstrem Terkait Bencana Hidrometeorologi Di Jayapura (Studi Kasus Hujan Lebat Tanggal 22 Februari 2014). *Jurnal Meteorologi Klimatologi Dan Geofisika*, 5(3), 25–36. <https://doi.org/10.36754/jmkg.v5i3.72>
- Sebastian, L. (2008). PENDEKATAN PENCEGAHAN DAN PENANGGULANGAN BANJIR | Flood Prevention and Control Approach. *Dinamika Teknik Sipil*, 8, 162--169.
- Yuli Supodo, Bakti Toni Endaryono, & Sony Yudha Perkasa. (2021). Peninggian Saluran Air Sebagai Solusi Mengurangi Banjir di Perumahan Bukit Kencana, Bekasi, Jawa Barat. *Kanigara*, 1(1), 32–41. <https://doi.org/10.36456/kanigara.v1i1.3155>

LAMPIRAN

1. Script Codingan

Lampiran 1 : Implementasi model

```
#
=====

# LANGKAH 1: HUBUNGKAN GOOGLE COLAB DENGAN GOOGLE
DRIVE

#
=====

# memunculkan jendela pop-up untuk meminta izin akses.

from google.colab import drive
drive.mount('/content/drive')

print("✅ Google Drive berhasil terhubung.")

#
=====

# LANGKAH 2: IMPORT LIBRARY YANG DIBUTUHKAN

#
=====

import pandas as pd

from sklearn.model_selection import train_test_split

from sklearn.linear_model import LogisticRegression

from sklearn.metrics import classification_report, confusion_matrix
```

```

import seaborn as sns

import matplotlib.pyplot as plt

print("✅ Library yang dibutuhkan telah di-import.")

#
=====

# LANGKAH 3: MEMUAT DATASET YANG DIBUTUHKAN

#
=====

# Muat dataset Anda

file_path = 'dataset_final_dengan_banjir.csv'

df = pd.read_csv(file_path)

# Tampilkan 5 baris pertama untuk memastikan data termuat dengan benar

print("--- 5 Baris Pertama Dataset ---")

print(df.head())

print("\n--- Info Dataset ---")

df.info()

```

Lampiran 2 : Modifikasi Dataset CVS

```

#
=====
=====

# LANGKAH 4: PEMBAGIAN DATASET

#
=====
=====

# Tentukan kolom fitur (X) dan target (y)

features = ['Curah Hujan (mm)', 'Suhu Udara (°C)', 'Kelembapan (%)',
'Kecepatan Angin (m/s)', 'Tekanan Udara (hPa)']

target = 'banjir'

X = df[features]

y = df[target]

print("Data fitur (X) dan target (y) berhasil dipisahkan.")

# --- Program Verifikasi Pemisahan Data ---

# Menampilkan 5 baris pertama dari tabel fitur (X)

# Pastikan tabel ini hanya berisi kolom-kolom fitur cuaca.

print("--- Menampilkan 5 Baris Pertama Data Fitur (X) ---")

print(X.head())

```

```
# Menampilkan 5 baris pertama dari tabel target (y)

# Pastikan tabel ini hanya berisi data dari kolom 'banjir'.

print("\n" + "="*50 + "\n")

print("--- Menampilkan 5 Baris Pertama Data Target (y) ---")

print(y.head())


# Mengecek dimensi (bentuk) dari data

# Jumlah baris (angka pertama) harus sama antara X dan y.

print("\n" + "="*50 + "\n")

print("--- Mengecek Dimensi Data ---")

print(f"Bentuk X (fitur): {X.shape}")

print(f"Bentuk y (target): {y.shape}")


# Pembagian data dengan stratifikasi

X_train, X_test, y_train, y_test = train_test_split(

    X, y,

    test_size=0.2,

    random_state=42,

    stratify=y

)

print("Data berhasil dibagi menjadi set training dan testing.")
```

```
print("Distribusi kelas 'banjir' di training set:")  
  
print(y_train.value_counts())  
  
print("\nDistribusi kelas 'banjir' di testing set:")  
  
print(y_test.value_counts())  
  
  
# Inisialisasi dan latih model  
  
model = LogisticRegression(max_iter=1000, class_weight='balanced',  
random_state=42)  
  
model.fit(X_train, y_train)  
  
  
print("Model Logistic Regression berhasil dilatih!")
```

Lampiran 3 : pengecekan peforma dan overfitting

```
#
=====
==
# LANGKAG 5 : PEFORMA DATA TRAINING DAN TESTING
#
=====
==
# --- Program Verifikasi Overfitting ---

from sklearn.metrics import recall_score


# 1. Buat prediksi pada data training
y_pred_train = model.predict(X_train)


# 2. Buat prediksi pada data testing (seperti sebelumnya)
y_pred_test = model.predict(X_test)


# 3. Tampilkan kedua laporan untuk perbandingan
print("--- PERFORMA PADA DATA TRAINING ---")
print("Model dievaluasi menggunakan data yang ia gunakan untuk berlatih.")
print(classification_report(y_train, y_pred_train))

print("\n" + "="*50 + "\n")

print("--- PERFORMA PADA DATA TESTING ---")
print("Model dievaluasi menggunakan data baru yang belum pernah ia lihat.")
```

```

print(classification_report(y_test, y_pred_test))

# --- Program untuk Menghitung Penurunan Performa (Overfitting) ---

# 1. Hitung recall pada data training untuk kelas 1 (Banjir)
recall_train = recall_score(y_train, y_pred_train, pos_label=1)

# 2. Hitung recall pada data testing untuk kelas 1 (Banjir)
recall_test = recall_score(y_test, y_pred_test, pos_label=1)

# 3. Hitung penurunan performa

# Mencegah pembagian dengan nol jika recall training adalah 0
if recall_train > 0:
    performance_drop = ((recall_train - recall_test) / recall_train) * 100
else:
    performance_drop = 0

# 4. Tampilkan hasilnya

print("--- Analisis Overfitting Berdasarkan Recall Kelas 'Banjir' ---")

print(f"Recall pada Training Set: {recall_train:.2%}")

print(f"Recall pada Testing Set: {recall_test:.2%}")

print("-" * 50)

print(f"Penurunan Performa (Indikasi Overfitting):
{performance_drop:.2f}%")

```

Lampiran 4 : kode visualisasi Confusion matrix

```

#
=====
==

# PREDIKSI PADA DATA TESTING

#
=====
==

#CONFUSION MATRIX

# Lakukan prediksi pada data testing
y_pred = model.predict(X_test)


# Tampilkan Classification Report
print("\n--- Classification Report ---")

# Fokus pada nilai 'recall' untuk kelas 1 (banjir)
print(classification_report(y_test, y_pred))


# Tampilkan Confusion Matrix
print("\n--- Confusion Matrix ---")

cm = confusion_matrix(y_test, y_pred)


plt.figure(figsize=(6,4))
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues',
            xticklabels=['Tidak Banjir', 'Banjir'],
            yticklabels=['Tidak Banjir', 'Banjir'])

```

```
plt.xlabel('Prediksi Model')  
plt.ylabel('Kenyataan (Aktual)')  
plt.title('Confusion Matrix')  
plt.show()
```

Lampiran 5 : Kode visualisasi kurva ROC-AUC

```

#
=====
==

# VISUALISASI KURVA ROC & AUC UNTUK MODEL FINAL

#
=====
==

#ROC - AUC

from sklearn.metrics import roc_curve, roc_auc_score
import matplotlib.pyplot as plt

# --- Langkah 1: Dapatkan Probabilitas Prediksi ---
# Kita butuh probabilitasnya, bukan hanya hasil akhir 0 atau 1
#[:, 1] mengambil probabilitas untuk kelas '1' (Banjir)
y_pred_proba = model.predict_proba(X_test)[:, 1]

# --- Langkah 2: Hitung Skor AUC dan Kurva ROC ---
# Skor AUC mengukur seberapa baik model membedakan kelas (semakin tinggi
semakin baik)
auc_score = roc_auc_score(y_test, y_pred_proba)

# Menghitung False Positive Rate (FPR) dan True Positive Rate (TPR)
fpr, tpr, thresholds = roc_curve(y_test, y_pred_proba)

# --- Langkah 3: Buat Visualisasi Plot ---

```

```
plt.figure(figsize=(8, 6))

# Plot kurva ROC model

plt.plot(fpr, tpr, color='blue', lw=2, label=f'Model Logistic Regression (AUC = {auc_score:.2f})')

# Plot garis acak (model tanpa keahlian)

plt.plot([0, 1], [0, 1], color='red', lw=2, linestyle='--', label='Tebakan Acak (AUC = 0.50)')


plt.xlabel('False Positive Rate (Tingkat Alarm Palsu)')
plt.ylabel('True Positive Rate (Tingkat Deteksi)')
plt.title('Kurva ROC (Receiver Operating Characteristic)')
plt.legend(loc='lower right')
plt.grid(alpha=0.3)
plt.show()
```

Lampiran 6 : Presentasi pada banjir

```
#
=====
==
# MELATIH ULANG MODEL DATA YANG SUDAH DISTANDARISASI
#
=====
==
#PENGECEKAN APA YANG PALING BERPENGARUH DALAM
PREDIKSI BANJIR

from sklearn.preprocessing import StandardScaler

from sklearn.model_selection import train_test_split

from sklearn.linear_model import LogisticRegression

import pandas as pd

import joblib

# Muat dataset lengkap Anda

df = pd.read_csv('dataset_final_dengan_banjir.csv')

# Pisahkan fitur dan target

features = ['Curah Hujan (mm)', 'Suhu Udara (°C)', 'Kelembapan (%)',
'Kecepatan Angin (m/s)', 'Tekanan Udara (hPa)']

target = 'banjir'

X = df[features]

y = df[target]
```

```

# --- LANGKAH BARU: STANDARISASI FITUR ---

scaler = StandardScaler()

X_scaled = scaler.fit_transform(X) # X_scaled sekarang memiliki mean=0 dan
std_dev=1

# Lakukan train-test split PADA DATA YANG SUDAH DI-SCALE

X_train_s, X_test_s, y_train_s, y_test_s = train_test_split(
    X_scaled, y, test_size=0.2, random_state=42, stratify=y
)

# Latih ulang model PADA DATA YANG SUDAH DI-SCALE

model_scaled = LogisticRegression(max_iter=1000, class_weight='balanced',
random_state=42)

model_scaled.fit(X_train_s, y_train_s)

print("Model berhasil dilatih ulang dengan data yang sudah distandarisasi.")

# --- LIHAT KEMBALI FEATURE IMPORTANCE ---

coefficients_scaled = model_scaled.coef_[0]

feature_importance_scaled = pd.DataFrame({'Fitur': features, 'Pengaruh
(Setelah Standarisasi)': coefficients_scaled})

```

```
feature_importance_scaled['Abs_Pengaruh'] =  
feature_importance_scaled['Pengaruh (Setelah Standarisasi)'].abs()  
feature_importance_scaled =  
feature_importance_scaled.sort_values(by='Abs_Pengaruh',  
ascending=False).drop(columns='Abs_Pengaruh')  
  
print("\n--- Peringkat Faktor Paling Berpengaruh (Setelah Standarisasi) ---")  
print(feature_importance_scaled)
```

Lampiran 7 : Kode untuk menyimpan *dataset* yang telah dilatih

```
# Simpan model yang sudah dilatih ke dalam sebuah file

import joblib

nama_file_model = 'model_prediksi_banjir.pkl'
joblib.dump(model, nama_file_model)

print(f"Model berhasil disimpan sebagai '{nama_file_model}'")
print("File ini sekarang bisa Anda gunakan untuk diintegrasikan dengan UI aplikasi Anda.")
```

Lampiran 8 : Kode untuk tampilan UI

```
# === PROGRAM UI PREDIKSI BANJIR VERSI FINAL ===

# --- Langkah 1: Instal/Upgrade Library ---

!pip install gradio -q

!pip install pandas matplotlib seaborn -q


import gradio as gr
import joblib
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import datetime

from PIL import Image, ImageDraw, ImageFilter

# --- Langkah 2: Muat Model dan Data Historis ---

try:
    model = joblib.load('model_prediksi_banjir.pkl')
    print("Model 'model_prediksi_banjir.pkl' berhasil dimuat.")
except FileNotFoundError:
    print("ERROR: File model 'model_prediksi_banjir.pkl' tidak ditemukan.")
    model = None
```

```

try:
    historical_data = pd.read_csv('dataset_final_dengan_banjir.csv',
    parse_dates=['Tanggal'], dayfirst=True)

    print("Data historis 'dataset_final_dengan_banjir.csv' berhasil dimuat.")
except FileNotFoundError:
    print("ERROR: File data historis 'dataset_final_dengan_banjir.csv' tidak
    ditemukan.")

    historical_data = None

# --- Langkah 3: Fungsi Heatmap Zona Rawan ---
def create_heatmap(base_image_path="peta_medan_marelan.png"):
    try:
        base_image = Image.open(base_image_path).convert("RGBA")
    except FileNotFoundError:
        print(f"ERROR: File peta '{base_image_path}' tidak ditemukan.")
        return None

    hotspots = {
        'Kelurahan Terjun': ((200, 250), 1.0),
        'Kelurahan Tanah 600': ((350, 400), 0.8),
        'Kelurahan Rengas Pulau': ((450, 150), 0.7),
        'Kelurahan Paya Pasir': ((50, 450), 0.6)
    }

    heatmap_layer = Image.new("RGBA", base_image.size, (255, 255, 255, 0))

```

```

draw = ImageDraw.Draw(heatmap_layer)

for spot, (coords, weight) in hotspots.items():
    radius = int(80 * weight)
    color = (255, int(150 * (1 - weight)), 0, 90)
    x, y = coords
    draw.ellipse((x - radius, y - radius, x + radius, y + radius), fill=color)

heatmap_layer = heatmap_layer.filter(ImageFilter.GaussianBlur(radius=30))
composite_image = Image.alpha_composite(base_image, heatmap_layer)
final_image_path = "heatmap_marelan.png"
composite_image.convert("RGB").save(final_image_path)
return final_image_path

# --- Langkah 4: Fungsi Prediksi ---

def predict_banjir(curah_hujan, suhu, kelembapan, kec_angin, tekanan_udara):
    if model is None:
        return "MODEL TIDAK DITEMUKAN", "0%", "Pastikan file model diunggah."

    input_data = np.array([[curah_hujan, suhu, kelembapan, kec_angin, tekanan_udara]])
    probabilitas_banjir = model.predict_proba(input_data)[0][1]

    if probabilitas_banjir > 0.7:

```

```

        status = "POTENSI BANJIR"

        rekomendasi = ""- Amankan dokumen penting...\n- Siapkan tas siaga
        bencana...\n- Pantau informasi BPBD.""

    elif probabilitas_banjir > 0.4:

        status = "WASPADA"

        rekomendasi = ""- Periksa saluran air...\n- Waspada jika hujan deras > 1
        jam...\n- Siapkan peralatan darurat.""

    else:

        status = "AMAN"

        rekomendasi = "- Kondisi aman...\n- Tetap waspada perubahan cuaca."

    prob_persen = f"{probabilitas_banjir:.2%}"

    return status, prob_persen, rekomendasi

# --- Langkah 5: Fungsi Grafik Tren (Histori + Input Baru) ---

def plot_with_new_input(curah_hujan, suhu, kelembapan, kec_angin,
    tekanan_udara, weather_param='Curah Hujan (mm)'):

    if historical_data is None:

        return "Data historis tidak ditemukan."

    # Salin data historis

    df_plot = historical_data.copy()

    # Tambahkan data baru dengan tanggal hari ini

    new_row = {

        'Tanggal': pd.Timestamp(datetime.date.today()),

```

```

    'Curah Hujan (mm)': curah_hujan,

    'Suhu Udara (°C)': suhu,

    'Kelembapan (%)': kelembapan,

    'Kecepatan Angin (m/s)': kec_angin,

    'Tekanan Udara (hPa)': tekanan_udara

}

    df_plot    =    pd.concat([df_plot,    pd.DataFrame([new_row])],
ignore_index=True)

# Buat plot

plt.figure(figsize=(12, 6))

    sns.lineplot(x='Tanggal', y=weather_param, data=df_plot, label="Data
Historis")

    plt.scatter(new_row['Tanggal'], new_row[weather_param], color='red', s=100,
label="Input Hari Ini")

    plt.title(f'Tren {weather_param} (Termasuk Input Hari Ini)')

    plt.xlabel('Tanggal')

    plt.ylabel(weather_param)

    plt.grid(True)

    plt.xticks(rotation=45)

    plt.legend()

    plt.tight_layout()

    return plt

# --- Langkah 6: Fungsi Gabungan Prediksi + Grafik ---

def    predict_and_plot(curah_hujan,    suhu,    kelembapan,    kec_angin,
```

```

tekanan_udara, weather_param):

    status, prob, rekomendasi = predict_banjir(curah_hujan, suhu, kelembapan,
    kec_angin, tekanan_udara)

    plot = plot_with_new_input(curah_hujan, suhu, kelembapan, kec_angin,
    tekanan_udara, weather_param)

    return status, prob, rekomendasi, plot

# --- Langkah 7: Bangun UI ---

with gr.Blocks(theme=gr.themes.Soft(), title="Prediksi Banjir Medan Marelan")
as demo:

    gr.Markdown("# 🏰 Dasbor Peringatan Dini Banjir Kecamatan Medan
    Marelan")

    with gr.Tabs():

        with gr.TabItem("Dashboard Utama"):

            with gr.Row():

                with gr.Column(scale=1):

                    gr.Markdown("### Input Kondisi Cuaca Saat Ini (Simulasi)")

                    input_ch = gr.Slider(minimum=0, maximum=300, value=50,
                    label="Curah Hujan (mm)")

                    input_suhu = gr.Slider(minimum=22, maximum=36, value=29,
                    label="Suhu Udara (°C)")

                    input_kelembapan = gr.Slider(minimum=40, maximum=100,
                    value=80, label="Kelembapan (%)")

                    input_angin = gr.Slider(minimum=0, maximum=15, value=5,
                    label="Kecepatan Angin (m/s)")

                    input_tekanan = gr.Slider(minimum=990, maximum=1020,
                    value=1008, label="Tekanan Udara (hPa)")

```

```

gr.Markdown("### Pilih Parameter Grafik Tren Historis")

parameter_select = gr.Dropdown(
    ['Curah Hujan (mm)', 'Suhu Udara (°C)', 'Kelembapan (%)',
'Kecepatan Angin (m/s)', 'Tekanan Udara (hPa)'],
    value='Curah Hujan (mm)', label="Parameter"
)

with gr.Column(scale=2):
    gr.Markdown("### Output Prediksi")

    output_status = gr.Textbox(label="Status Peringatan",
interactive=False)

    output_prob = gr.Textbox(label="Probabilitas Banjir",
interactive=False)

    output_rekomendasi = gr.Textbox(label="Rekomendasi Tindakan",
lines=3, interactive=False)

    output_plot = gr.Plot(label="Grafik Tren Cuaca Historis + Input Hari
Ini")

    inputs = [input_ch, input_suhu, input_kelembapan, input_angin,
input_tekanan, parameter_select]

    outputs = [output_status, output_prob, output_rekomendasi, output_plot]

    for input_component in inputs:
        input_component.change(fn=predict_and_plot, inputs=inputs,
outputs=outputs)

with gr.TabItem("Peta & Informasi"):
    gr.Markdown("### Heatmap Zona Rawan Banjir di Medan Marelan")

```

```

heatmap_image = create_heatmap()

if heatmap_image:

    gr.Image(heatmap_image, label="Heatmap Zona Rawan Banjir",
interactive=False)

else:

    gr.Markdown("***Gagal membuat heatmap. Pastikan file
`peta_medan_marelan.png` sudah diunggah.**")

    gr.Markdown("### Peta Kecamatan Medan Marelan (Google Maps)")

    peta_google_html = """

                                                                    <iframe
src="https://www.google.com/maps/embed?pb=!1m18!1m12!1m3!1d63725.42
427948151!2d98.62227913486328!3d3.738988633993445!2m3!1f0!2f0!3f0!3
m2!1i1024!2i768!4f13.1!3m3!1m2!1s0x30312e3181309341%3A0x8f053c18ef
9c0408!2sMedan%20Marelan%2C%20Medan%20City%2C%20North%20Sum
atra!5e0!3m2!1sen!2sid!4v1715501848888!5m2!1sen!2sid"

                                width="100%" height="450" style="border:0;" allowfullscreen=""
loading="lazy" referrerpolicy="no-referrer-when-downgrade"></iframe>

    """

    gr.HTML(peta_google_html)

    gr.Markdown("### Kontak Darurat")

    gr.Markdown("""

    **BPBD Kota Medan (Pusdalops):**

    - **Telepon:** (061) 4520352

    - **Call Center:** 112

    """)

```

```
# --- Langkah 8: Jalankan UI ---
```

```
demo.launch(share=True, debug=True)
```

2. Surat Penetapan Pembimbing


MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 88/SKUAN-PT/Akred/PT/18/2019
 Pusat Administrasi: Jalan Muhtar Basri No. 3 Medan 20238 Telp: (061) 8622480 - 8622487 Fax: (061) 8625474 - 8631003
 Email: umsumedan@umsu.ac.id Website: umsu.ac.id Instagram: @umsu.ac.id Facebook: umsumedan Twitter: umsumedan YouTube: umsumedan

PENETAPAN DOSEN PEMBIMBING
PROPOSAL/SKRIPSI MAHASISWA
NOMOR : 83/IL3-AU/I/MSU-09/F/2025

Assalamu'alaikum Warahmatullahi Wabarakatuh

Dekan Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara, berdasarkan Persetujuan permohonan judul penelitian Proposal / Skripsi dari Ketua / Sekretaris.

Program Studi : Teknologi Informasi
Pada tanggal : 15 Januari 2025

Dengan ini menetapkan Dosen Pembimbing Proposal / Skripsi Mahasiswa.

Nama : Anwar Ruhdin Hasugian
NPM : 2109020156
Semester : VII (Tujuh)
Program studi : Teknologi Informasi
Judul Proposal / Skripsi : Penerapan Algoritma Logistic Regression Untuk Prediksi Risiko Banjir Di Kota Medan Berdasarkan Data Curah Hujan Dan Faktor Meteorologi

Dosen Pembimbing : Mahardika Abdi Prawira, S.Kom., M.Kom.

Dengan demikian di izinkan menulis Proposal / Skripsi dengan ketentuan

1. Penulisan berpedoman pada buku panduan penulisan Proposal / Skripsi Fakultas Ilmu Komputer dan Teknologi Informasi UMSU
2. Pelaksanaan Sidang Skripsi harus berjarak 3 bulan setelah dikeluarkannya Surat Penetapan Dosen Pembimbing Skripsi.
3. **Proyek Proposal / Skripsi dinyatakan " BATAL "** bila tidak selesai sebelum Masa Kadaluaarsa tanggal : 15 Januari 2026
4. Revisi judul.....

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Ditetapkan di : Medan
 Pada Tanggal : 15 Rajab 1446 H
 15 Januari 2025M


Dekan
Dr. Al Khawarizmi, M.Kom.
NIDN : 0117099201




3. Letter Of Acceptance



APTIKOM

PANITIA
Seminar Nasional Komputer (SNIKOM 2025)
Asosiasi Pendidikan Tinggi dan Informatika Komputer
Provinsi Sumatera Utara

Sekretariat:

Jl. Alumni No. 9 Kampus USU Telp : (061) 8210077 FAX : (061) 8210077

LETTER OF ACCEPTANCE (LOA)

No: 6/LOA/SNIKOM /2025

Kepada:

Yth. Bpk/Ibu/sdr/I :

Anwar ruhdin hasugian

Dengan hormat,

Dengan ini kami menyampaikan bahwa artikel dengan judul **"Penerapan algoritma logistic regression untuk prediksi banjir di kecamatan Medan Marelan berdasarkan data curah hujan dan faktor meteorologi"** dinyatakan diterima / accepted pada publikasi PROSIDING SNIKOM 2025

Demikian informasi ini, atas perhatian dan kerjasamanya kami ucapkan terima kasih.

Ketua Panitia



Dr. Muhammad Iqbal, S.Kom., M.Kom.



Sekretaris



Jamaluddin, S.Kom., M.Kom.

4. Hasil Cek Turnitin

skripsi-anwar-r-hsgn_1761035446405			
ORIGINALITY REPORT			
18%	16%	8%	11%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS
PRIMARY SOURCES			
1	Submitted to Universitas Muhammadiyah Sumatera Utara Student Paper	2%	
2	ejurnalunsam.id Internet Source	2%	
3	repository.its.ac.id Internet Source	1%	
4	htm.kstmkg.ac.id Internet Source	1%	
5	ejurnalqarnain.stisnq.ac.id Internet Source	1%	
6	github.com Internet Source	1%	
7	e-journal.iainpekalongan.ac.id Internet Source	1%	
8	eprints.univetbantara.ac.id Internet Source	1%	
9	Submitted to Fakultas Teknik Student Paper	<1%	
10	Submitted to University of San Diego Student Paper	<1%	

11	www.upgrad.com Internet Source	<1 %
12	Submitted to Academic Library Consortium Student Paper	<1 %
13	www.geeksforgeeks.org Internet Source	<1 %
14	Submitted to Universitas Muria Kudus Student Paper	<1 %
15	Daniel Wellem Mauleky, Mohammad Amin Lasaiba. "Analisis Dampak Bencana Banjir di Urimessing Kelurahan Ahusen, Sirimau Kota Ambon dan Kaitannya dengan Kesadaran Masyarakat dalam Penanganan Bencana Banjir", Jurnal Geografi, Lingkungan dan Kesehatan, 2025 Publication	<1 %
16	journal.unesa.ac.id Internet Source	<1 %
17	dspace.uil.ac.id Internet Source	<1 %
18	juejin.cn Internet Source	<1 %
19	repository.umsu.ac.id Internet Source	<1 %
20	www.analyticsvidhya.com Internet Source	<1 %

21	Esther Ria Matulesy. "PERBANDINGAN ANTARA MODEL AUTOREGRESSIVE INTEGRATED MOVING AVERAGE (ARIMA) DAN MODEL FUNGSI TRANSFER PADA PERAMALAN CURAH HUJAN DI KABUPATEN MANOKWARI", Jurnal Natural, 2019 Publication	<1 %
22	www.classace.io Internet Source	<1 %
23	lmsspada.kemdiktisaintek.go.id Internet Source	<1 %
24	O'neal Efrata Madao, Akhmad Irsyad, Muhammad Rivani Ibrahim. "ANALISIS SENTIMEN PADA ULASAN PENGGUNA APLIKASI JAMSOSTEK MOBILE DENGAN MENGGUNAKAN NAÏVE BAYES DAN LOGISTIC REGRETION", Djtechno: Jurnal Teknologi Informasi, 2025 Publication	<1 %
25	skripsiuyuk.com Internet Source	<1 %
26	uknowledge.uky.edu Internet Source	<1 %
27	forge.citizen4.eu Internet Source	<1 %
28	rumahcoding.co.id Internet Source	<1 %
29	text-id.123dok.com Internet Source	<1 %

		<1 %
30	Submitted to Coventry University Student Paper	<1 %
31	R. Rivai, I.G. Tunas, R. Mardin. "Prediksi Kerawanan Banjir di Daerah Aliran Sungai Olaya Kabupaten Parigi Moutong", REKONSTRUKSI TADULAKO: Civil Engineering Journal on Research and Development, 2025 Publication	<1 %
32	ejournalpasca.unisi.ac.id Internet Source	<1 %
33	Submitted to Napier University Student Paper	<1 %
34	Submitted to Sekolah Teknik Elektro & Informatika Student Paper	<1 %
35	Submitted to Selçuk Üniversitesi Student Paper	<1 %
36	Submitted to Unika Soegijapranata Student Paper	<1 %
37	digilib.unhas.ac.id Internet Source	<1 %
38	fr.scribd.com Internet Source	<1 %
39	Submitted to Universitas Brawijaya Student Paper	<1 %

40	huggingface.co Internet Source	<1 %
41	www.bomberbot.com Internet Source	<1 %
42	Afid Nurkholis, Nuringtyas Yogi Jurnawan, Rizka Ratna Sayekti, Yuli Widyaningsih et al. "APLIKASI PEMODELAN HEC-HMS UNTUK IDENTIFIKASI KEJADIAN BANJIR BANDANG DI DAS CIMANUK HULU, KABUPATEN GARUT", INA-Rxiv, 2018 Publication	<1 %
43	IIP IMRON DAIPAH, RINI ASTUTI, WILLY PRIHARTONO. "PREDIKSI CHURN PELANGGAN PADA LAYANAN DESAIN GRAFIS HOME DESAIN MENGGUNAKAN ALGORITMA NAÏVE BAYES", Jurnal Informatika dan Teknik Elektro Terapan, 2025 Publication	<1 %
44	Submitted to UNIVERSITAS BUDI LUHUR Student Paper	<1 %
45	dspace.umkt.ac.id Internet Source	<1 %
46	Submitted to 2U Southern Methodist University Student Paper	<1 %
47	Dwi Wulandari Sari. "LOGISTIC REGRESSION ANALYSIS WITH FEATURE SELECTION AND CROSS VALIDATION IN DETERMINING CHILD	<1 %

NUTRITION STATUS PREDICTION MODEL",
Djtechno: Jurnal Teknologi Informasi, 2025
Publication

48	Submitted to Sriwijaya University Student Paper	<1 %
49	Submitted to University of Newcastle upon Tyne Student Paper	<1 %
50	ciencia.lasalle.edu.co Internet Source	<1 %
51	codanics.com Internet Source	<1 %
52	dione.lib.unipi.gr Internet Source	<1 %
53	docs.openvino.ai Internet Source	<1 %
54	journal.tofedu.or.id Internet Source	<1 %
55	karyailmiah.polnes.ac.id Internet Source	<1 %
56	pontianak.go.id Internet Source	<1 %
57	pyoostoryinlife.blogspot.com Internet Source	<1 %
58	x-mia3.blogspot.com Internet Source	<1 %
59	blog.csdn.net	

	Internet Source	<1 %
60	core.ac.uk Internet Source	<1 %
61	cornerstone.lib.mnsu.edu Internet Source	<1 %
62	download.bibis.ir Internet Source	<1 %
63	repository.umnaw.ac.id Internet Source	<1 %
64	smartaccounting.files.wordpress.com Internet Source	<1 %
65	Submitted to Imperial College of Science, Technology and Medicine Student Paper	<1 %
66	Mukhlis Mukhlis. "Analysis of the Study Asbabun Nuzul: "The Urgency and Contribution in Understanding the Qur'an."" Al-Muhith: Jurnal Ilmu Qur'an dan Hadits, 2024 Publication	<1 %
67	Octavia Salwa Dzaky Fadhillah, Jajam Haerul Jaman, Carudin Carudin. "PERBANDINGAN NAIVE BAYES, SUPPORT VECTOR MACHINE, LOGISTIC REGRESSION DAN RANDOM FOREST DALAM MENGANALISIS SENTIMEN MENGENAI TIKTOKSHOP", Jurnal Informatika dan Teknik Elektro Terapan, 2025 Publication	<1 %

68	Sartini, Sumarna, Abdul Hamid, Ahmad Hafidzul Kahfi, Nicodias Palasara. "REVOLUSI DIAGNOSIS: OPTIMASI RANDOM TREE-PSO UNTUK PENYAKIT GINJAL KRONIS", Jurnal Informatika dan Rekayasa Elektronik, 2025 Publication	<1 %
69	comesacompetition.org Internet Source	<1 %
70	journal.poltekkes-mks.ac.id Internet Source	<1 %
71	Nashiruddin Sahal Muhtadi, Dhaffa Mulya Rahman, Dewi Fatmarani Surianto. "ANALISIS RISIKO KEBAKARAN HUTAN DENGAN LOGIKA FUZZY MAMDANI", Jurnal Informatika dan Teknik Elektro Terapan, 2024 Publication	<1 %
72	Yulianda Tasya, Raissa Amanda Putri. "Klasifikasi Tingkat Kerawanan Banjir Wilayah Medan Menggunakan Metode Naive Bayes Dan Algoritma J48", INTECOMS: Journal of Information Technology and Computer Science, 2023 Publication	<1 %
73	ejournal.nusamandiri.ac.id Internet Source	<1 %
74	123dok.com Internet Source	<1 %
75	Phenikaa University Publication	<1 %

76	S.P. Jani, M. Adam Khan. "Applications of AI in Smart Technologies and Manufacturing", CRC Press, 2025 Publication	<1%
77	rauzansumara.github.io Internet Source	<1%
78	repositori.uin-alauddin.ac.id Internet Source	<1%
79	Xinyuan Song, HSIEH,WEI-CHE, Ziqian Bi, Chuanqi Jiang, Junyu Liu, Benji Peng, Sen Zhang, Xuanhe Pan, Jiawei Xu, Jinlang Wang. "A Comprehensive Guide to Explainable AI: From Classical Models to LLMs", Open Science Framework, 2024 Publication	<1%

Exclude quotes On

Exclude matches Off

Exclude bibliography On