

ANALISIS KORELASI DAN REGRESI LINIER BERGANDA UNTUK PREDIKSI CURAH HUJAN HARIAN DI SIDOARJO

Sofia Ramadhani Megantara, Rosyid Abdillah Ramdhan,
Reyana Dinda Maulan Feriza, Anggraini Puspita Sari*
Informatika, Universitas Pembangunan Nasional “Veteran” Jawa Timur
Jl. Rungkut Madya No.1, Gunung Anyar
24081010091@student.upnjatim.ac.id

ABSTRAK

Kabupaten Sidoarjo mengalami banjir berulang pada musim hujan akibat curah hujan tinggi dan kondisi geografis sebagai dataran rendah di bagian bawah Sub DAS Pucang. Belum ada analisis kuantitatif mendalam untuk mengidentifikasi faktor cuaca dominan yang berpengaruh signifikan terhadap kejadian banjir. Penelitian ini bertujuan menganalisis hubungan statistik antara curah hujan harian (X_1), kecepatan angin rata-rata (X_2), kelembaban udara (X_3), dan hujan lebat (X_4) terhadap kejadian banjir (Y) di Kabupaten Sidoarjo. Data BMKG Juanda (394 observasi) dianalisis menggunakan korelasi Pearson dan regresi linier berganda. Hasil menunjukkan curah hujan memiliki hubungan sangat kuat dengan $r = 0,821$ ($p < 0,001$), diikuti hujan lebat dengan $r = 0,978$ ($p < 0,001$), dan kelembaban udara dengan $r = 0,295$ ($p < 0,001$), sedangkan kecepatan angin tidak signifikan dengan $r = -0,094$ ($p = 0,078$). Model regresi $Y = 0,049 + 0,001X_1 + 0,000X_2 - 0,001X_3 + 0,933X_4$ menghasilkan $R^2 = 0,959$. Curah hujan dan hujan lebat berpengaruh signifikan secara parsial dan simultan ($p < 0,001$). Temuan mengkonfirmasi curah hujan sebagai faktor dominan pemicu banjir, sehingga monitoring curah hujan harian perlu diprioritaskan untuk sistem peringatan dini banjir Kabupaten Sidoarjo.

Kata kunci : curah hujan harian, kejadian banjir, analisis korelasi, regresi linier berganda, Sidoarjo

1. PENDAHULUAN

Indonesia adalah sebuah negara yang terdiri dari banyak pulau, yang berada di garis ekuator dengan ciri khas iklim tropis, yang ditandai oleh dua periode utama, yakni musim kemarau (April hingga Oktober) dan musim hujan (Oktober hingga April). Fluktuasi curah hujan yang signifikan di wilayah ini dipengaruhi oleh berbagai elemen cuaca, seperti laju angin, tingkat kelembaban, suhu udara, serta tekanan atmosfer [1]. Pengetahuan mengenai keterkaitan antara unsur-unsur cuaca tersebut dengan curah hujan sangat krusial untuk memperbaiki ketepatan ramalan hujan dan memperkuat mekanisme peringatan awal terhadap bencana yang berkaitan dengan cuaca dan udara [1].

Kabupaten Sidoarjo, yang berada di wilayah Provinsi Jawa Timur, termasuk salah satu daerah dengan tingkat curah hujan yang cukup besar dan berfluktuasi, terutama selama periode musim hujan [2]. Dari segi geografi, daerah ini terletak di bagian bawah Sub Daerah Aliran Sungai (DAS) Pucang, yang merupakan komponen dari jaringan DAS Brantas [3]. Pola curah hujan di lokasi tersebut terbentuk oleh interaksi berbagai unsur cuaca yang saling mempengaruhi, khususnya laju angin dan tingkat kelembaban udara. Metode korelasi dan regresi berganda dapat diterapkan guna menilai kekuatan hubungan dan besarnya efek dari laju angin rata-rata serta tingkat kelembaban udara pada curah hujan harian di Sidoarjo [4]. Kondisi sebagai dataran rendah dengan kapasitas drainase terbatas menyebabkan Kabupaten Sidoarjo rentan mengalami banjir berulang.

Permasalahan utama adalah tingginya frekuensi banjir yang menyebabkan kerugian ekonomi dan sosial signifikan. Tingkat intensitas hujan berperan sebagai

indikator utama kondisi cuaca yang memberikan kontribusi signifikan terhadap kejadian banjir, terutama ketika curah hujan harian ≥ 50 mm [5]. Kecepatan angin dan kelembaban udara memiliki peran penting dalam mekanisme pembentukan serta penyebaran hujan di suatu daerah [6]. Meskipun curah hujan diketahui sebagai pemicu utama, belum ada analisis kuantitatif mendalam untuk mengidentifikasi faktor cuaca dominan yang secara statistik berpengaruh signifikan terhadap kejadian banjir. Keterbatasan pemahaman ini menghambat pengembangan sistem peringatan dini yang akurat dan responsif.

Beberapa penelitian telah mengeksplorasi pola curah hujan melalui integrasi metode korelasi statistik dan model regresi linier berganda di wilayah Indonesia. Penelitian oleh Luthfiarta dkk. [1] Menggunakan pendekatan regresi linier berganda untuk meramalkan curah hujan. Penelitian tersebut memanfaatkan empat prediktor utama, yaitu suhu, kelembaban relatif, tekanan atmosfer, dan kecepatan angin. Data berasal dari Stasiun Meteorologi BMKG Ahmad Yani Semarang. Hasilnya menunjukkan koefisien determinasi R^2 mencapai 25,5%. Koefisien tersebut mengindikasikan kontribusi signifikan dari model regresi multivariat dalam menjelaskan variasi curah hujan. Normelia dkk. [7] menerapkan metode regresi linier berganda untuk estimasi curah hujan di Stasiun Geofisika Sleman. Penelitian ini menghasilkan koefisien determinasi R^2 sebesar 0,1136. Nilai tersebut berarti sekitar 11,36% dari variabilitas curah hujan yang dapat di akuntansi oleh variabel-variabel prediktor yang dipilih. Pangestu dkk. [8] melakukan evaluasi korelasi statistik antara data curah hujan

BMKG dan pengukuran satelit TRMM di Stasiun BMKG Sumatera Utara. Koefisien korelasi Pearson menunjukkan efektivitas dalam mendeteksi kesamaan distribusi temporal dan kekuatan hubungan antar variabel curah hujan. Nilai-nilai koefisien tersebut secara statistik bermakna.

Penelitian ini menggunakan data curah hujan, kelembaban udara, kecepatan angin, dan kejadian banjir yang diperoleh dari Stasiun Meteorologi Juanda yang dikelola oleh Badan Meteorologi, Klimatologi, dan Geofisika (BMKG) serta data kejadian banjir dari BPBD Kabupaten Sidoarjo, dengan periode pengamatan dari November 2024 hingga November 2025. Periode satu tahun hidrologis ini dipilih untuk mencakup siklus lengkap musim hujan, musim kemarau, dan awal musim hujan berikutnya, sehingga dapat merekam seluruh variasi intensitas curah hujan dan kejadian banjir yang terjadi di Kabupaten Sidoarjo. Penelitian ini bertujuan untuk melaksanakan pengujian normalitas data guna memastikan validitas penerapan metode statistik parametrik, kemudian menganalisis korelasi statistik antara curah hujan harian, kategori hujan lebat dengan intensitas ≥ 50 mm, kelembaban udara rata-rata, dan kecepatan angin rata-rata terhadap kejadian banjir di wilayah penelitian. Selanjutnya, penelitian ini akan membangun model regresi linier berganda untuk menentukan faktor-faktor cuaca yang berpengaruh secara parsial maupun simultan terhadap kejadian banjir, serta mengidentifikasi parameter cuaca dominan yang dapat digunakan sebagai indikator utama dalam pengembangan sistem peringatan dini banjir di Kabupaten Sidoarjo. Hasil penelitian diharapkan dapat menghasilkan informasi kuantitatif berupa nilai koefisien korelasi beserta tingkat signifikansinya, identifikasi variabel cuaca yang paling berpengaruh, dan model persamaan regresi yang dapat dimanfaatkan untuk memprediksi probabilitas terjadinya banjir berdasarkan kondisi cuaca harian.

Untuk mengatasi permasalahan tersebut, penelitian ini menggabungkan uji normalitas, analisis korelasi Pearson, dan regresi linier berganda. Uji normalitas memastikan data dapat dianalisis dengan statistik parametrik [9]. Analisis korelasi mengukur kekuatan hubungan antara curah hujan harian, kecepatan angin, kelembaban udara, dan hujan lebat terhadap banjir, sementara regresi berganda membangun model prediksi untuk mengidentifikasi variabel dominan [7]. Hasil analisis diharapkan dapat diintegrasikan ke dalam sistem peringatan dini banjir di Kabupaten Sidoarjo.

2. TINJAUAN PUSTAKA

2.1. Curah Hujan

Curah hujan menjadi salah satu parameter utama yang selalu diperhatikan dalam analisis hidrologi karena berperan besar dalam proses terjadinya banjir. Intensitas hujan harian yang tinggi dalam waktu singkat dapat meningkatkan limpasan permukaan,

terutama pada wilayah dengan kapasitas drainase yang terbatas dan tingkat infiltrasi tanah yang rendah [11]. Penelitian terdahulu yang dilakukan oleh Saputra dan Nugroho mengkaji hubungan antara unsur-unsur iklim terhadap curah hujan harian di wilayah Jawa Timur menggunakan analisis korelasi statistik dengan variabel suhu udara, kelembapan relatif, tekanan udara, dan kecepatan angin. Hasil penelitian tersebut menunjukkan bahwa kelembaban udara memiliki korelasi paling kuat terhadap curah hujan harian, diikuti oleh suhu udara dan tekanan udara, sementara kecepatan angin menunjukkan hubungan yang relatif lemah. Dengan ini mengindikasikan bahwa curah hujan harian dipengaruhi oleh lebih dari satu variabel meteorologi dan memiliki sifat fluktuatif, sehingga pendekatan multivariat diperlukan dalam pemodelannya. Oleh karena itu, penelitian Saputra dan Nugroho menjadi landasan dalam penelitian ini untuk menerapkan analisis korelasi sebagai tahap awal dan regresi linier berganda sebagai metode prediksi curah hujan harian di Kabupaten Sidoarjo.

2.2. Banjir

Banjir di Indonesia umumnya dipengaruhi oleh kombinasi faktor hidrometeorologi dan kondisi fisik wilayah, seperti curah hujan, kapasitas sistem drainase, tata guna lahan, serta karakteristik geomorfologi. Beberapa studi dalam lima tahun terakhir menunjukkan bahwa curah hujan dengan intensitas dan durasi tinggi merupakan faktor dominan pemicu banjir, terutama di wilayah perkotaan dan dataran rendah di Provinsi Jawa Timur. Curah hujan harian yang tinggi berperan sebagai pemicu utama terjadinya banjir musiman, dan kondisi ini semakin diperparah oleh keterbatasan kapasitas drainase serta perubahan tata guna lahan yang menurunkan kemampuan infiltrasi tanah. Selain itu, hasil pemetaan dataran rendah banjir menunjukkan bahwa wilayah dataran rendah memiliki tingkat kerentanan yang lebih tinggi akibat intensitas hujan yang besar dan akumulasi aliran permukaan yang signifikan [12]. Sejalan dengan hal tersebut, Saputra dan Nugroho menegaskan bahwa kejadian banjir di Jawa Timur cenderung meningkat seiring dengan peningkatan intensitas dan durasi curah hujan, khususnya pada wilayah dengan topografi datar dan sistem pengaliran air yang terbatas, sehingga curah hujan menjadi faktor kunci dalam analisis dan mitigasi risiko banjir.

2.3. Korelasi

Korelasi digunakan untuk menunjukkan tingkat dan arah hubungan antara curah hujan dan tinggi muka air sungai, di mana hubungan kedua variabel tersebut umumnya bersifat positif, sehingga peningkatan curah hujan akan diikuti oleh kenaikan debit dan tinggi muka air sungai akibat bertambahnya limpasan permukaan yang mengalir menuju sungai. Penelitian dalam lima tahun terakhir menunjukkan bahwa korelasi antara curah hujan dan tinggi muka air sungai cenderung kuat pada musim hujan, terutama ketika intensitas hujan

tinggi dan terjadi secara berturut-turut, yang menyebabkan kondisi tanah menjadi jenuh, kapasitas infiltrasi menurun, dan volume limpasan meningkat sehingga berdampak langsung pada kenaikan muka air sungai [13]. Selain itu, Pratama dan Wibowo (2023) melalui analisis korelasi Spearman menunjukkan bahwa curah hujan ekstrem memiliki hubungan yang kuat dengan kejadian banjir perkotaan, di mana respon muka air sungai menjadi lebih cepat dan signifikan pada wilayah dengan dominasi lahan terbangun. Kekuatan korelasi ini juga dipengaruhi oleh karakteristik daerah aliran sungai (DAS), seperti luas DAS, kemiringan lereng, serta kondisi tutupan lahan, di mana DAS dengan permukaan kedap air cenderung menunjukkan korelasi yang lebih tinggi antara curah hujan dan tinggi muka air sungai dibandingkan DAS dengan tutupan vegetasi alami [14]. Dengan demikian, hubungan korelasi antara curah hujan dan tinggi muka air sungai banyak dimanfaatkan dalam pengembangan program peringatan dini banjir, yang dimana hasil analisis korelasi akan digunakan untuk memprediksi potensi kenaikan muka air sungai secara lebih akurat. Melalui sistem peringatan dini yang dianalisis dari data curah hujan dan hujan lebat, upaya penanggulangan banjir dapat dilakukan secara lebih cepat dan efektif sehingga risiko kerugian akibat banjir bisa diminimalkan. Korelasi pearson yang umum ditemui memiliki persamaan:

$$r = \frac{Cov(X,Y)}{S_X \times S_Y} \quad (2.1)$$

Keterangan:

- $Cov(X, Y)$ = kovarians antar variabel X dan Y
- S_X dan S_Y = deviasi standar dari variabel X dan Y

2.4. Regresi Linier

Regresi linier merupakan metode statistik yang banyak digunakan dalam kajian hidrologi untuk mengetahui hubungan fungsional antara variabel bebas dan variabel terikat, seperti curah hujan sebagai variabel independen dan tinggi muka air sungai atau debit sungai sebagai variabel dependen. Metode ini digunakan untuk menggambarkan sejauh mana perubahan curah hujan dapat memengaruhi perubahan tinggi muka air sungai secara kuantitatif dan sering dimanfaatkan dalam studi bencana hidrologi untuk memprediksi kenaikan debit atau tinggi muka air sungai berdasarkan data curah hujan harian atau bulanan. Penelitian di Indonesia menunjukkan bahwa model regresi linier mampu memberikan gambaran awal mengenai potensi banjir. Sejalan dengan hal tersebut, Pratama dan Wibowo melalui analisis regresi linier menunjukkan bahwa curah hujan, khususnya hujan dengan intensitas tinggi, berpengaruh signifikan terhadap peningkatan tinggi muka air sungai dan kejadian banjir perkotaan. Selain itu, hasil regresi linier dapat digunakan untuk menilai kekuatan pengaruh curah hujan terhadap kejadian banjir melalui nilai koefisien regresi dan koefisien determinasi, di mana koefisien regresi yang bernilai positif menunjukkan bahwa peningkatan curah hujan diikuti

oleh kenaikan tinggi muka air sungai, sedangkan koefisien determinasi menggambarkan besarnya kontribusi curah hujan dalam menjelaskan variasi tinggi muka air sungai [14]. Dengan demikian, regresi linier menjadi salah satu metode yang efektif dan sederhana dalam analisis hidrologi, khususnya untuk mendukung perencanaan pengelolaan sumber daya air serta pengembangan sistem peringatan dini banjir berbasis data curah hujan [10]. Regresi linier memiliki persamaan:

$$Y = a + bX \quad (2.2)$$

Keterangan:

- Y = variabel dependen (terikat)
- X = variabel independen (bebas)
- a = konstanta (*intercept*)
- b = koefisien regresi (*slope*)

2.5. Regresi Berganda

Regresi linier berganda digunakan untuk menentukan besarnya pengaruh beberapa variabel bebas secara simultan terhadap satu variabel terikat, seperti kejadian banjir atau tinggi muka air sungai. Dalam kajian hidrologi, variabel bebas yang umum digunakan meliputi curah hujan, kapasitas sistem drainase, penggunaan lahan, serta karakteristik topografi wilayah. Metode ini memungkinkan identifikasi variabel yang paling dominan mempengaruhi terjadinya banjir berdasarkan nilai koefisien regresinya. Penelitian terdahulu yang dilakukan di Surakarta menunjukkan bahwa curah hujan memberikan kontribusi paling besar dalam model prediksi banjir berbasis regresi linier berganda dibandingkan variabel lainnya, yang mengindikasikan bahwa peningkatan intensitas dan frekuensi hujan berperan signifikan dalam meningkatkan potensi banjir di wilayah perkotaan [11]. Penggunaan data harian dalam pemodelan regresi berganda dinilai mampu menggambarkan respon sistem hidrologi secara lebih rinci terhadap perubahan curah hujan dan kondisi lingkungan, sehingga hasil analisis menjadi lebih representatif dan dapat mendukung perencanaan serta pengambilan keputusan dalam upaya mitigasi dan pengurangan risiko banjir secara lebih akurat. Regresi berganda memiliki persamaan:

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon \quad (2.3)$$

Keterangan:

- Y = variabel dependen (yang diprediksi)
- $X_1, X_2, \dots, X_n$ = variabel independen (prediktor)
- α = konstanta
- $\beta_1, \beta_2, \dots, \beta_n$ = koefisien regresi
- ε = variabel pengganggu

2.6. Uji Simultan

Uji simultan digunakan untuk mengetahui apakah seluruh variabel bebas secara bersama-sama memiliki pengaruh yang signifikan terhadap variabel terikat dalam suatu permasalahan regresi, seperti kejadian banjir, debit sungai, atau tinggi muka air sungai. Dalam analisis regresi linier berganda, uji simultan umumnya dilakukan menggunakan uji F

untuk menguji hipotesis nol bahwa seluruh koefisien regresi sama dengan nol secara bersamaan. Penolakan hipotesis nol menunjukkan bahwa variabel-variabel bebas yang digunakan dalam model memiliki pengaruh kolektif terhadap variabel terikat [11]. Uji simultan memiliki peran penting dalam kajian banjir karena fenomena banjir umumnya tidak dipengaruhi oleh satu faktor tunggal, melainkan oleh interaksi berbagai variabel hidrometeorologi dan fisik wilayah, seperti curah hujan, kecepatan angin rata-rata, hujan lebat, dan kelembapan udara rata-rata yang secara bersama-sama memengaruhi besarnya debit air sungai.

Hal ini sejalan dengan penelitian Sari dan Wibowo yang menggunakan regresi linier berganda untuk menganalisis faktor-faktor yang memengaruhi tinggi muka air sungai dan menunjukkan bahwa variabel curah hujan, luas daerah aliran sungai, dan kemiringan lereng secara simultan berpengaruh signifikan berdasarkan hasil uji F. Hasil uji simultan tersebut menegaskan bahwa model regresi yang signifikan secara keseluruhan layak digunakan sebagai alat analisis dan prediksi sebelum dilakukan interpretasi lebih lanjut melalui uji parsial. Beberapa penelitian hidrologi juga menunjukkan bahwa model regresi yang lolos uji simultan dapat dimanfaatkan sebagai dasar pemodelan risiko banjir dan perencanaan mitigasi bencana, termasuk pengembangan sistem peringatan dini banjir, perencanaan tata ruang, serta pengelolaan sumber daya air berbasis data hidrometeorologi [12]. Persamaan yang dimiliki uji simultan:

$$F = \frac{R^2 / k}{(1 - R^2) / (n - k - 1)} \quad (2.4)$$

Atau

$$F = \frac{SSR / k}{SSE / (n - k - 1)} \quad (2.5)$$

Keterangan:

- F = nilai F hitung
- R^2 = koefisien determinasi
- k = jumlah variabel independen
- n = jumlah sampel/observasi
- SSR = *Sum of Square Regression* (jumlah kuadrat regresi)
- SSE = *Sum of Square Error* (jumlah kuadrat error/residual)

2.7. Central Limit Theorem (CLT)

Central Limit Theorem (CLT) menyatakan bahwa distribusi rata-rata sampel akan mendekati distribusi normal ketika ukuran sampel cukup besar, meskipun data asal tidak berdistribusi normal. Prinsip ini menjadi tumpuan utama dalam statistik inferensial karena memungkinkan peneliti melakukan pengujian hipotesis dan pemodelan statistik parametrik tanpa harus sepenuhnya bergantung pada bentuk distribusi data awal. Menurut Montgomery, Peck, dan Vining, dalam konteks analisis regresi, normalitas residual atau rata-rata sampel dapat dipenuhi melalui prinsip CLT apabila jumlah data pengamatan mencukupi, sehingga metode inferensial seperti uji t dan uji F tetap

valid digunakan. Dalam penelitian curah hujan dan banjir, data curah hujan harian umumnya bersifat tidak normal, cenderung positively skewed, serta mengandung nilai ekstrem yang tinggi, sehingga sering menjadi kendala dalam penerapan metode statistik parametrik. Namun, dengan menggunakan CLT, data pengamatan harian atau bulanan yang dikumpulkan dalam jumlah besar dapat menghasilkan distribusi rata-rata yang mendekati normal [10]. Penerapan CLT ini menjadi penting untuk menganalisis hubungan antara curah hujan, debit sungai, dan tinggi muka air sungai, karena dengan jumlah data yang memadai, seperti data harian selama beberapa tahun, peneliti dapat menerapkan regresi linier maupun regresi linier berganda untuk mengkaji pengaruh curah hujan terhadap kejadian banjir. Dengan demikian, CLT mendukung validitas penggunaan uji statistik seperti uji t dan uji simultan (uji F) dalam analisis regresi, sehingga hasil signifikansi yang diperoleh dapat diinterpretasikan secara ilmiah dan menjadi dasar yang kuat dalam penelitian banjir [15]. Persamaan yang dimiliki Central Limit Theorem (CLT):

$$Z = \frac{\bar{X} - \mu}{\sigma / \sqrt{n}} \quad (2.6)$$

Atau dalam bentuk distribusi

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \quad (2.7)$$

Keterangan:

- Z = nilai standar normal
- $\bar{X}$ = rata-rata sampel
- μ = rata-rata populasi
- σ = standar deviasi populasi
- n = ukuran sampel
- $\sigma / \sqrt{n}$ = kesalahan baku

2.8. Uji Parsial

Uji parsial digunakan untuk menentukan pengaruh masing-masing variabel bebas secara individu terhadap variabel terikat dalam model regresi yang diterapkan. Uji ini umumnya dilakukan menggunakan uji t untuk menguji signifikansi koefisien regresi dari setiap variabel independen dengan asumsi bahwa variabel bebas lainnya dianggap konstan, sehingga dapat diketahui apakah suatu variabel memiliki pengaruh yang signifikan secara statistik terhadap variabel terikat [13]. Dalam analisis regresi hidrologi, uji parsial memiliki peranan penting karena kejadian banjir dipengaruhi oleh berbagai faktor yang tingkat pengaruhnya tidak selalu sama. Variabel seperti curah hujan, kecepatan angin rata-rata, dan kelembapan udara rata-rata dapat memberikan kontribusi yang berbeda-beda terhadap peningkatan debit atau tinggi muka air sungai.

Hasil uji parsial membantu mengidentifikasi variabel yang paling dominan dalam memengaruhi banjir, di mana variabel dengan nilai signifikansi lebih kecil dari tingkat kepercayaan yang ditetapkan (misalnya $\alpha = 0,05$) dinyatakan berpengaruh signifikan terhadap peningkatan risiko banjir atau kenaikan tinggi muka air sungai. Hal ini sejalan dengan

penelitian Sari dan Wibowo yang menunjukkan bahwa curah hujan memiliki pengaruh signifikan secara parsial terhadap tinggi muka air sungai dibandingkan variabel fisik lainnya. Temuan tersebut menegaskan bahwa curah hujan sering menjadi faktor utama dalam pemodelan banjir pada berbagai penelitian hidrologi [12]. Uji parsial memiliki persamaan:

$$t_i = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \quad (2.7)$$

Keterangan:

- t = nilai t hitung
- $\hat{\beta}_i$ = koefisien regresi variabel independen ke- i
- $SE(\hat{\beta}_i)$ = standar error koefisien
- 0 = nilai hipotesis ($H_0: \beta_i = 0$)

3. METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif dengan jenis penelitian eksplanatori, yang bertujuan untuk menganalisis hubungan antara variabel-variabel penelitian serta menjelaskan keterkaitan yang terjadi pada data curah hujan di Kabupaten Sidoarjo. Metode analisis yang digunakan meliputi analisis korelasi dan regresi linier berganda, yang diterapkan untuk mengetahui tingkat hubungan dan pengaruh variabel independen terhadap curah hujan sebagai variabel dependen. Penggunaan analisis korelasi dan regresi linier berganda memungkinkan evaluasi hubungan antarvariabel dilakukan secara sistematis dengan tingkat keandalan statistik yang tinggi.

Penelitian dilakukan di Kabupaten Sidoarjo pada rentang waktu November 2024 hingga November 2025. Data yang digunakan merupakan data sekunder yang diperoleh dari BMKG dan BPBD Sidoarjo. Data yang digunakan akan dikumpulkan dengan metode dokumentasi, yaitu mengumpulkan data curah hujan, kecepatan angin rata-rata, dan kelembaban udara. Data diolah dengan uji korelasi pearson untuk mengetahui hubungan antara curah hujan dan kejadian banjir, serta dilakukan regresi linier berganda untuk menganalisis regresi yang melibatkan lebih dari satu variabel bebas (independen) untuk memprediksi variabel terikat (dependen).

Uji pendahuluan yang dilakukan pada penelitian ini adalah uji normalitas. Uji normalitas data dilakukan menggunakan uji Kolmogorov–Smirnov untuk mengukur deviasi antara distribusi sampel dan distribusi normal. Curah hujan ≥ 50 mm menyebabkan banjir di Kabupaten Sidoarjo. Pada uji normalitas menggunakan *python* sebagai perangkat lunak yang akan membantu dalam menganalisis data.

4. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data curah hujan harian dan unsur cuaca yang diperoleh dari BMKG, dengan periode pengamatan dari 1 November 2024 sampai 30 November 2025. Jumlah data yang dianalisis sebanyak 394 observasi. Variabel yang digunakan dalam penelitian ini terdiri dari:

- RR: curah hujan harian (mm)
- FF_AVG: kecepatan angin rata-rata (m/s)

- RH_AVG: kelembaban udara rata-rata (%)
- HUJAN_LEBAT: hujan ≥ 50 mm
- Banjir: kejadian banjir (0 = tidak banjir, 1 = banjir)

Variabel RR, FF_AVG, RH_AVG, dan HUJAN_LEBAT digunakan sebagai variabel independen, sedangkan variabel Banjir digunakan sebagai variabel dependen. Uji normalitas dilakukan untuk mengetahui apakah data setiap variabel penelitian berdistribusi normal. Dalam penelitian ini, uji normalitas dilakukan menggunakan uji Kolmogorov–Smirnov (K–S). Kriteria pengujian yang digunakan adalah:

- Jika nilai p -value $> 0,05$, maka data berdistribusi normal.
- Jika nilai p -value $\leq 0,05$, maka data tidak berdistribusi normal.

Tabel 1. Hasil uji normalitas kolmogorov-smirnov

Variasi	Statistik K-S	p-value	Keterangan
RR	0.3105	0.0000	Tidak normal
FF_AVG	0.3579	0.0000	Tidak normal
RH_AVG	0.0660	0.0863	Normal

Uji Normalitas RR: Statistic=0.3105, p-value=0.0000
 Uji Normalitas FF_AVG: Statistic=0.3579, p-value=0.0000
 Uji Normalitas RH_AVG: Statistic=0.0660, p-value=0.0863

Gambar 1. Hasil uji normalitas kolmogorov-smirnov

Berdasarkan hasil uji Kolmogorov–Smirnov di atas dapat dijelaskan bahwa:

- Variabel curah hujan (RR) memiliki nilai p -value sebesar 0,0000, yang lebih kecil dari taraf signifikansi 0,05, sehingga data RR tidak berdistribusi normal.
- Variabel kecepatan angin rata-rata (FF_AVG) juga memiliki nilai p -value 0,0000 $< 0,05$, sehingga data FF_AVG tidak berdistribusi normal.
- Variabel kelembaban udara rata-rata (RH_AVG) memiliki nilai p -value 0,0863 $> 0,05$, sehingga data RH_AVG berdistribusi normal.

Dengan demikian, bahwa tidak semua variabel penelitian berdistribusi normal. Namun analisis statistik parametrik seperti korelasi Pearson dan regresi linear berganda tetap digunakan, karena jumlah sampel yang digunakan dalam penelitian ini cukup besar ($n = 394$). Berdasarkan Central Limit Theorem (CLT), pada ukuran sampel besar, distribusi data mendekati normal sehingga analisis parametrik masih dapat diterapkan.

Analisis korelasi pearson digunakan untuk mengetahui hubungan linear antara dua variabel, baik antar variabel independen maupun antara variabel independen dengan variabel dependen. Hubungan antarvariabel dinyatakan signifikan apabila nilai p -value lebih kecil dari tingkat signifikan yang ditetapkan. Kriteria pengujian korelasi pearson adalah:

- Jika p -value $< 0,05$, maka terdapat hubungan yang signifikan.

- b. Jika $p\text{-value} \geq 0,05$, maka tidak terdapat hubungan yang signifikan.

Tabel 2. Korelasi antar variabel independen

Pasangan Variabel	r	p-value	Keterangan
RR - FF_AVG	-0.167	0.0016	Signifikan
RR - RH_AVG	0.452	0.0000	Signifikan
FF_AVG - RH_AVG	-0.271	0.0000	Signifikan

Korelasi RR dengan FF_AVG:
 $r = -0.167$, $p\text{-value} = 0.0016$

Korelasi RR dengan RH_AVG:
 $r = 0.452$, $p\text{-value} = 0.0000$

Korelasi FF_AVG dengan RH_AVG:
 $r = -0.271$, $p\text{-value} = 0.0000$

Gambar 2. Korelasi antar variabel independen

Hasil curah hujan tersebut memiliki hubungan negatif dan signifikan dengan kecepatan angin. Curah hujan memiliki hubungan positif dan signifikan dengan kelembaban udara. Kecepatan angin memiliki hubungan negatif dan signifikan dengan kelembaban udara. Hubungan ini menunjukkan bahwa saat curah hujan meningkat, kelembaban udara cenderung meningkat, sedangkan kecepatan angin cenderung menurun.

Tabel 3. Korelasi variabel independen dengan kejadian banjir

Pasangan Variabel	r	p-value	Keterangan
RR - Banjir	0.821	0.0000	Signifikan
FF_AVG - Banjir	-0.094	0.0776	Tidak signifikan
RH_AVG - Banjir	0.295	0.0000	Signifikan
HUJAN_LEBAT - Banjir	0.978	0.0000	Signifikan

Korelasi RR dengan Kejadian Banjir:
 $r = 0.826$, $p\text{-value} = 0.0000$

Korelasi FF_AVG dengan Kejadian Banjir:
 $r = -0.096$, $p\text{-value} = 0.0712$

Korelasi RH_AVG dengan Kejadian Banjir:
 $r = 0.289$, $p\text{-value} = 0.0000$

Korelasi HUJAN_LEBAT dengan Kejadian Banjir:
 $r = 0.978$, $p\text{-value} = 0.0000$

Gambar 3. Korelasi variabel independen dengan kejadian banjir

Curah hujan dan hujan lebat memiliki hubungan sangat kuat dan signifikan dengan kejadian banjir. Kecepatan angin tidak memiliki hubungan yang signifikan dengan kejadian banjir. Kelembaban udara memiliki hubungan positif dan signifikan dengan kejadian banjir, meskipun kekuatannya sedang. Analisis regresi linear berganda digunakan untuk mengetahui pengaruh simultan dan parsial variabel RR, FF_AVG, RH_AVG, dan HUJAN_LEBAT

terhadap kejadian banjir. Model regresi yang diperoleh adalah:

$$Y = 0.049 + 0.001X_1 + 0.000X_2 - 0.001X_3 + 0.933X_4$$

Keterangan:

- X_1 = curah hujan (RR)
- X_2 = kecepatan angin rata-rata (FF_AVG)
- X_3 = kelembaban udara rata-rata (RH_AVG)
- X_4 = hujan lebat (HUJAN_LEBAT)

Tabel 4. Uji parsial (uji t)

Pasangan Variabel	r	p-value	Keterangan
RR	0.0010	0.000	Signifikan
FF_AVG	-0.0005	0.912	Tidak signifikan
RH_AVG	-0.0007	0.132	Tidak signifikan
HUJAN_LEBAT	0.9331	0.000	Signifikan

	coef	std err	t	P> t	[0.025	0.975]
const	0.0494	0.037	1.345	0.180	-0.023	0.122
RR	0.0010	0.000	4.046	0.000	0.001	0.002
FF_AVG	-0.0005	0.004	-0.111	0.912	-0.009	0.008
RH_AVG	-0.0007	0.000	-1.510	0.132	-0.002	0.000
HUJAN_LEBAT	0.9331	0.020	47.818	0.000	0.895	0.971

Gambar 4. Uji parsial (uji t)

Curah hujan dan hujan lebat berpengaruh positif dan signifikan terhadap kejadian banjir, sedangkan kecepatan angin dan kelembaban udara tidak berpengaruh signifikan secara parsial. Nilai probabilitas uji F sebesar $5,06 \times 10^{-242}$ ($< 0,05$), sehingga dapat disimpulkan bahwa seluruh variabel independen secara simultan berpengaruh signifikan terhadap kejadian banjir. Nilai $R^2 = 0,959$, yang berarti bahwa 95,9% variasi kejadian banjir dapat dijelaskan oleh curah hujan, kecepatan angin, kelembaban udara, dan hujan lebat, sedangkan sisanya 4,1% dijelaskan oleh faktor lain di luar model.

Tabel 5. Hasil akhir

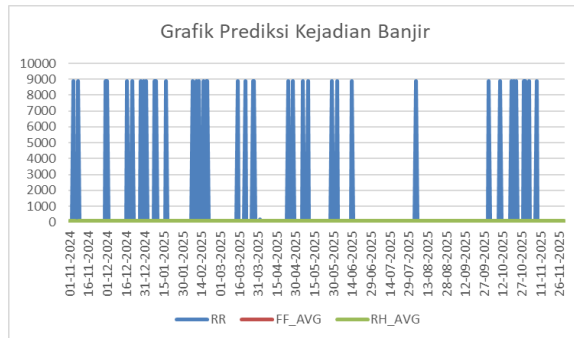
Keterangan	Nilai
R-squared	0.959
Adj. R-squared	0.959
F-statistic	2054.
Prob (F-statistic)	5.06e-242
Log-Likelihood	549.32
AIC	-1089.
BIC	-1069.

OLS Regression Results			
Dep. Variable:	BANJIR	R-squared:	0.959
Model:	OLS	Adj. R-squared:	0.959
Method:	Least Squares	F-statistic:	2054.
Date:	Mon, 05 Jan 2026	Prob (F-statistic):	5.06e-242
Time:	16:19:32	Log-Likelihood:	549.32
No. Observations:	356	AIC:	-1089.
Df Residuals:	351	BIC:	-1069.
Df Model:	4		
Covariance Type:	nonrobust		

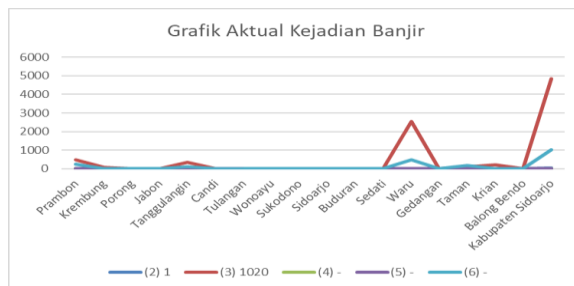
Gambar 5. Hasil akhir

Hasil penelitian menunjukkan bahwa kejadian banjir di Kabupaten Sidoarjo sangat dipengaruhi oleh intensitas curah hujan, khususnya hujan lebat dengan ambang ≥ 50 mm, sedangkan kelembaban udara dan

kecepatan angin tidak memberikan pengaruh yang berarti. Hasil ini sejalan dengan teori hidrometeorologi yang menyatakan bahwa intensitas dan hujan ekstrem merupakan pemicu utama terjadinya banjir, terutama pada daerah dengan kapasitas drainase yang terbatas. Temuan ini menguatkan bahwa pemantauan curah hujan harian menjadi indikator kunci dalam upaya mitigasi dan peringatan dini banjir di Kabupaten Sidoarjo.



Gambar 6. Grafik prediksi kejadian banjir



Gambar 7. Grafik aktual kejadian banjir

Pada grafik tersebut menampilkan variasi curah hujan harian selama periode pengamatan. Terlihat fluktuasi curah hujan dari nilai rendah hingga tinggi, dengan beberapa hari menunjukkan curah hujan ≥ 50 mm yang dikategorikan sebagai hujan lebat. Hari-hari dengan intensitas hujan tinggi ini menjadi indikator awal meningkatnya potensi kejadian banjir. Grafik tersebut menunjukkan hasil prediksi kejadian banjir berdasarkan ambang batas curah hujan sebesar 50 mm. Nilai 1 menunjukkan hari dengan potensi banjir ($RR \geq 50$ mm), sedangkan nilai 0 menunjukkan hari tanpa potensi banjir ($RR < 50$ mm). Pola grafik memperlihatkan bahwa kejadian banjir hanya muncul pada saat curah hujan melampaui ambang batas tersebut. Grafik sebar ini memperlihatkan hubungan antara curah hujan dan status prediksi banjir. Titik-titik banjir (nilai 1) terkonsentrasi pada curah hujan ≥ 50 mm, sementara pada curah hujan di bawah 50 mm didominasi oleh kondisi tidak banjir. Pola ini menunjukkan hubungan yang jelas antara intensitas curah hujan dan kemungkinan terjadinya banjir dalam model prediksi yang digunakan.

5. KESIMPULAN DAN SARAN

Berdasarkan pemeriksaan mendalam menggunakan analisis korelasi dan regresi linier

berganda pada data curah hujan harian di Kabupaten Sidoarjo dari bulan November 2024 hingga November 2025, dengan total 394 observasi, diperoleh temuan bahwa indikator hujan lebat (≥ 50 mm) merupakan faktor paling dominan yang mempengaruhi kejadian banjir dengan hubungan sangat kuat dan signifikan ($r = 0,978$, $p\text{-value} = 0,0000$), diikuti oleh curah hujan sebagai elemen utama dengan hubungan yang sangat erat dan bermakna secara statistik ($r = 0,821$, $p\text{-value} = 0,0000$), dan kelembaban udara yang menunjukkan kaitan positif signifikan meski kekuatannya sedang ($r = 0,295$, $p\text{-value} = 0,0000$), sedangkan kecepatan angin tidak menunjukkan hubungan yang berarti ($r = -0,094$, $p\text{-value} = 0,0776$). Model regresi linier berganda yang diperoleh adalah $Y = 0,049 + 0,001X_1 + 0,000X_2 - 0,001X_3 + 0,933X_4$, dengan koefisien determinasi R^2 mencapai 0,959, yang berarti sekitar 95,9% perubahan dalam kejadian banjir dapat dijelaskan oleh keempat variabel tersebut (curah hujan, kecepatan angin, kelembaban udara, dan indikator hujan lebat), sementara 4,1% sisanya dipengaruhi oleh faktor eksternal lainnya seperti kapasitas drainase dan topografi. Dari pengujian parsial, indikator hujan lebat memiliki dampak paling kuat dan signifikan terhadap banjir dengan koefisien 0,933 ($p\text{-value} = 0,0000$), diikuti curah hujan dengan koefisien 0,001 ($p\text{-value} = 0,0000$), sedangkan kecepatan angin ($p\text{-value} = 0,912$) dan kelembaban udara ($p\text{-value} = 0,132$) tidak memberikan pengaruh yang berarti secara parsial, namun secara keseluruhan, keempat variabel ini secara bersama-sama mempengaruhi banjir dengan signifikan berdasarkan uji F ($p\text{-value} = 5,06 \times 10^{-242}$). Untuk riset mendatang, sebaiknya ditambahkan variabel tambahan seperti durasi hujan, intensitas hujan maksimum per jam, curah hujan kumulatif, kapasitas saluran drainase, penggunaan lahan, kemiringan tanah, dan ciri-ciri geografis daerah untuk meningkatkan daya prediksi model, serta pertimbangkan penerapan teknik pembelajaran mesin seperti Random Forest atau Neural Network untuk memperbaiki ketepatan ramalan, perluas jangka waktu pengumpulan data agar bisa menangkap pola musiman dalam waktu lama, dan gabungkan model ini dengan sistem peringatan dini banjir serta teknologi sistem informasi geografis (SIG) untuk membantu upaya pencegahan bencana yang lebih efisien di Kabupaten Sidoarjo.

DAFTAR PUSTAKA

- [1] A. Luthfiarta, A. Febriyanto, H. Lestiawan and W. Wicaksono, "Analisa prakiraan Cuaca dengan Parameter Suhu, Kelembaban, Tekanan Udara, dan Kecepatan Angin Menggunakan Regresi Linear Berganda," *JOINS (Journal of Information System)*, vol. 5, pp. 10-17, 2020.
- [2] J.. A. Rohma and Yulfiah, "Pemetaan Kawasan Berpotensi Mengalami Kekeringan Berdasarkan Kondisi Ketersediaan Air Tanah Tersedia (KAT) di Kabupaten Sidoarjo, Jawa Timur," *Jurnal Ilmu Lingkungan*, vol. 22, 2024.

- [3] B. B. W. S. Brantas, "Langkah Melakukan Pelaporan," Kementerian Pekerjaan Umum dan Perumahan Rakyat, [Online]. Available: <https://sda.pu.go.id/balai/bbwsbrantas/>. [Accessed 29 Desember 2025]
- [4] J. Pebralia, "Analisis Curah Hujan Menggunakan Machine Learning Metode Regresi Linier Berganda Berbasis Python dan Jupyter Notebook," *JIFP(Jurnal Ilmu Fisika dan Pembelajarannya*, vol. 6, no. 2, pp. 22-30, 2022.
- [5] M. Ismail, "Intensitas Curah Hujan Tinggi, Jantung Kota Sidoarjo Dikepung Banjir," *Beritajatim.com*, 20 November 2025. [Online]. Available: <https://beritajatim.com/intensitas-curah-hujan-tinggi-jantung-kota-sidoarjo-dikepung-banjir>.
- [6] E. S. A. Ramadhany, N. Prianbikasatiarsa, P. Melani, G. G. Wisnawa and A. Realita, "ANALISIS KARAKTERISTIK GELOMBANG, ANGIN PERMUKAAN DAN CURAH HUJAN DI TIGA WILAYAH PEMBAGIAN ZONA HUJAN TAHUN 2021-2022," *Jurnal Inovasi Fisika Indonesia (IFI)*, vol. 14, 2025.
- [7] R. Normelia, T. D. Fortuna, E. P. Putri and E. Widodo, "Penerapan metode regresi linier berganda untuk memperkirakan curah hujan (studi kasus: Stasiun Geofisika Sleman)," *Jurnal Ilmiah Matematika (Konvergensi)*, vol. 9, no. 1, pp. 8-18, 2022
- [8] I. T. Pangestu, A. Zakaria and S. Tugiono, "Analisis Korelasi Data Curah Hujan BMKG dengan TRMM (Studi Kasus Stasiun BMKG di Sumatera Utara)," *Jurnal Rekayasa Sipil dan Desain*, vol. 8, no. 1, pp. 157-180, 2020.
- [9] R. SIANTURI, "UJI NORMALITAS SEBAGAI SYARAT PENGUJIAN HIPOTESIS," *JURNAL PEMBELAJARAN DAN MATEMA*
- [10] Ghasemi, A., & Zahediasl, S. (2020). Normality tests for statistical analysis: A guide for non-statisticians. *International Journal of Endocrinology and Metabolism*, 18(2), e102863.
- [11] Putri, R. N., Hidayat, R., & Prasetyo, Y. (2021). Pengaruh intensitas curah hujan harian terhadap limpasan permukaan dan kejadian banjir. *Jurnal Teknik Pengairan*, 12(1), 1–10
- [12] Utami, M. R., & Prabowo, A. (2024). Pemetaan daerah rawan banjir berdasarkan intensitas hujan dan topografi. *Jurnal Teknik Sipil dan Lingkungan*, 9(1), 55–65.
- [13] Saputra, A., & Nugroho, S. (2022). Analisis regresi berganda faktor-faktor penyebab banjir di wilayah perkotaan Surakarta. *Jurnal Sumber Daya Air*, 18(2), 101–110.
- [14] Rahman, F., Setiawan, B., & Lestari, I. (2023). Analisis hubungan curah hujan dan tinggi muka air sungai berbasis data deret waktu. *Jurnal Wilayah dan Lingkungan*, 11(3), 215–226.
- [15] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2021). *Introduction to linear regression analysis!*
- [16] A. Pratama and A. Wibowo, "Korelasi curah hujan ekstrem terhadap kejadian banjir perkotaan," *Jurnal Teknik Hidraulik*, vol. 14, no. 2, pp. 85–96, 2023
- [17] N. Sari and A. Wibowo, "Analisis faktor-faktor yang memengaruhi tinggi muka air sungai menggunakan regresi linier berganda," *Jurnal Teknik Sumber Daya Air*, vol. 14, no. 2, pp. 75–86, 2022.