

PERBANDINGAN AKURASI METODE *SUPPORT VECTOR MACHINE* DAN *K-NEAREST NEIGHBOUR* DALAM PREDIKSI CURAH HUJAN POTENSI BANJIR

Muhammad Fadawkas Oemarki¹, Muchammad Dimas Mufti Baskara², Iin Ernawati³

Program Sarjana Informatika / Fakultas Ilmu Komputer

Universitas Pembangunan Nasional Veteran Jakarta

Jl. RS. Fatmawati No. 1, Pondok Labu, Jakarta Selatan 12450

2210511116@mahasiswa.upnvj.ac.id¹, 2210511099@mahasiswa.upnvj.ac.id², iinernawati@upnvj.ac.id³

Abstrak. Peristiwa banjir telah menjadi suatu peristiwa yang kerap kali terjadi selama musim penghujan. Wilayah dataran rendah seperti DKI Jakarta tentunya menjadi daerah paling terdampak banjir. Terlebih lagi apabila tidak adanya mitigasi bencana, baik dari segi kesiapsiagaan maupun transparansi informasi. Seringkali hujan hanya menjadi akibat terjadinya banjir tanpa memerhatikan faktor-faktor cuaca lainnya. Padahal, informasi terkait keadaan cuaca beberapa saat sebelum adanya hujan dapat menjadi acuan besarnya curah hujan yang akan menyebabkan banjir. Maka dari itu, penelitian ini mengusulkan implementasi salah satu metode *data mining* guna memperoleh luaran berupa prediksi besaran curah hujan yang berpotensi menyebabkan banjir. Penelitian ini menggunakan metode CRISP-DM (*Cross-Industry Standard Process Model for Data Mining*). Metode prediksi curah hujan yang digunakan dalam penelitian ini adalah *Support Vector Machine* (SVM) dan *K-Nearest Neighbour* (K-NN) untuk mengklasifikasikan curah hujan berpotensi banjir. Penelitian ini akan memanfaatkan dataset cuaca historis yang mencakup variabel seperti suhu, kelembapan, arah angin, dan curah hujan, untuk mengembangkan model yang dapat memprediksi potensi banjir dengan akurasi tertinggi antara metode SVM dan K-NN. Melalui penemuan metode paling efektif dalam prediksi potensi banjir, penelitian ini diharapkan dapat meningkatkan kinerja dan ketepatan prediksi banjir yang dapat digunakan dalam mitigasi risiko bencana dan meminimalisir risiko yang terjadi. Hasil penelitian ini mengungkapkan bahwa metode *Support Vector Machine* memiliki akurasi rata-rata sebesar 90.19%, yang mana memiliki akurasi lebih baik dibandingkan metode *K-Nearest Neighbour* yang memiliki akurasi rata-rata sebesar 88.32%.

Kata Kunci: Banjir, Cuaca, K-NN, Prediksi, SVM

1 Pendahuluan

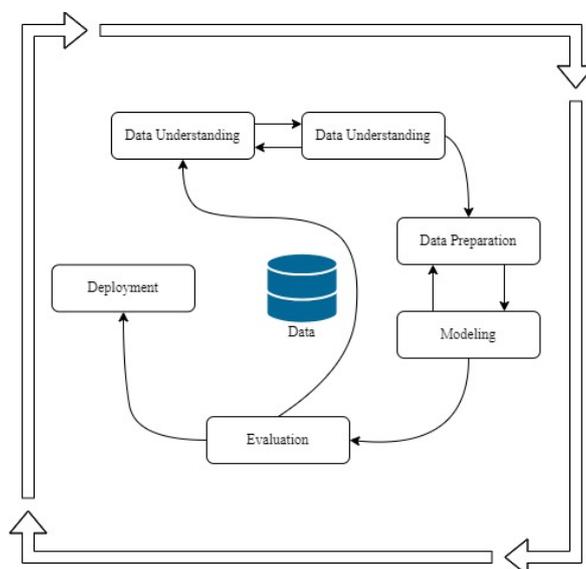
Banjir merupakan fenomena alam yang terjadi di saat ada aliran air yang menggenangi daratan yang kerap kali menjadi persoalan bagi sebagian masyarakat yang menetap baik di wilayah dataran rendah maupun bantaran sungai. Air akan menggenangi pemukiman warga sekitar ketika tinggi permukaan air sungai melebihi muka air normal. Hal ini terjadi akibat tingginya curah hujan dan kiriman air dari wilayah sekitar yang memiliki drainase buruk sehingga air meluap dari sungai yang akhirnya menimbulkan genangan air pada lahan rendah di tepi sungai. Banjir dapat terjadi karena dua kategori utama faktor: 1). Faktor alam seperti curah hujan yang berlebihan, erosi dan sedimentasi, karakteristik topografi dan geofisik sungai, serta kapasitas sungai dan drainase yang tidak memadai. Selain itu, penurunan permukaan tanah dan kerusakan pada bangunan pengendali banjir juga menjadi faktor penting. 2). Faktor manusia termasuk perubahan penggunaan lahan, pembuangan sampah yang tidak tepat, pembangunan di area yang rentan banjir, serta perencanaan sistem pengendalian banjir yang tidak efektif [1]. Peristiwa alam seperti banjir bukanlah fenomena baru di area perkotaan, termasuk di wilayah DKI Jakarta, yang merupakan bagian dari siklus tahunan [2].

Secara geografis, wilayah Provinsi DKI Jakarta memiliki karakteristik topografi dengan kemiringan lereng yang relatif landai dan berada di area dataran rendah [3]. Beberapa faktor utama yang mengakibatkan terjadinya banjir di DKI Jakarta antara lain adalah Pasang laut yang terjadi di Teluk Jakarta dan curah hujan lokal yang tinggi mengakibatkan air sungai meluap. Peristiwa banjir di Jakarta merupakan sejarah yang terulang tiap tahunnya. Sejarah mencatat banjir terparah di Jakarta terjadi pada tahun 2007, 2014, dan 2020. Banjir yang terjadi pada ketiga tahun tersebut sama-sama disebabkan oleh satu faktor terbesar, yakni curah hujan yang tinggi. Banjir Jakarta 2007 yang menenggelamkan setidaknya 60% wilayah Jakarta, pada saat itu disebabkan oleh tingginya curah hujan di angka 250 mm/hari [4]. Sementara pada tahun 2014, banjir yang menenggelamkan Bundaran HI ini disebabkan oleh tingginya curah hujan di angka 104 mm/hari. Pada hari pertama tahun 2020, DKI Jakarta diterpa hujan deras dengan curah hujan yang mencapai 377 mm/hari, yang merupakan rekor terbesar dalam 150 tahun terakhir sehingga menyebabkan banjir di berbagai wilayah [5]. Maka, dapat disimpulkan bahwa curah hujan ekstrim menjadi satu dari beragam faktor terbesar penyebab terjadinya banjir di wilayah DKI Jakarta.

Curah hujan merupakan ketinggian air hujan yang jatuh di area yang datar, tanpa menguap, meresap atau mengalir. Sebagai contoh, curah hujan 1 mm berarti bahwa dalam area satu meter persegi, air hujan menumpuk setinggi satu milimeter atau menumpuk sebanyak satu liter dalam periode waktu tertentu [6]. Besarnya curah hujan tentunya dipengaruhi dengan seberapa deras hujan yang terjadi di suatu wilayah. Sehingga, terdapat beberapa faktor yang mempengaruhi tingginya curah hujan. Penelitian [7] menyimpulkan bahwa suhu permukaan laut di Selat Makassar memiliki pengaruh signifikan terhadap intensitas curah hujan di Kota Balikpapan. Ketika suhu permukaan laut naik, curah hujan meningkat, dan sebaliknya di saat suhu permukaan laut turun, maka besaran curah hujan juga menurun. Penelitian [8] menjelaskan bahwa hasil dari model regresi yang dikembangkan, kelembaban yang berpengaruh positif terhadap curah hujan. Namun dijelaskan lebih lanjut oleh [9], bahwa suhu dan kelembaban udara berdampak hanya 16.1% terhadap prediksi curah hujan. Hal ini disebabkan ada beberapa aspek yang mempengaruhi selain temperatur dan kelembaban udara, di antaranya adalah tekanan udara dan kecepatan angin. Berdasarkan hasil uji hipotesis yang dilakukan oleh [10], ditemukan keterkaitan yang relevan antara tekanan, suhu, dan kelembaban udara serta kecepatan angin terhadap besaran curah hujan. Hasil analisis regresi menyatakan bahwa kelembaban udara dan kecepatan angin mempunyai nilai koefisien regresi positif terhadap curah hujan, sehingga curah hujan di kota Medan dapat diakibatkan oleh faktor tersebut.

Metode yang akan digunakan dalam prediksi curah hujan berpotensi banjir adalah metode *Support Vector Machine* karena faktor efektivitas yang berasal dari kemampuan metode dalam mengatasi *overfitting*, yang merupakan karakteristik data pada saat tahap pelatihan atau training memberikan hasil akurasi yang nyaris sempurna [11]. Sedangkan, *K-Nearest Neighbour* digunakan karena mudah diimplementasikan pada dataset yang kecil, tetapi dapat menghadapi keterbatasan dan kurang efisiensi saat berurusan dengan dataset yang besar dan kompleks [12]. Penelitian ini akan membandingkan akurasi antara metode SVM dan K-NN untuk melihat metode yang paling efektif digunakan dalam prediksi curah hujan berpotensi banjir.

2 Metodologi Penelitian



Gambar 2.1. Diagram alir penelitian

2.1 Business Understanding

Tahap awal penelitian dengan memahami masalah pada banjir yang kerap kali terjadi setelah hujan dengan intensitas tinggi. Pada tahapan ini diperlukan pemahaman tentang faktor penyebab curah hujan yang tinggi agar dapat diketahui data-data apa saja yang bisa digunakan dalam melakukan prediksi besaran curah hujan untuk mengetahui potensi terjadinya banjir. Dalam proses pencarian data, perlu dilakukan identifikasi stakeholder yang terlibat dan kebutuhan informasi yang dimiliki terkait data-data yang mempengaruhi besaran curah hujan. Pada penelitian ini, data didapatkan dari website resmi data online milik Badan Meteorologi Klimatologi dan Geofisika (BMKG).

2.2 Data Understanding

Tahapan selanjutnya merupakan pemahaman data yang mengacu kepada jenis-jenis data yang diperlukan dan dapat dilakukan penghitungan berbasis *Support Vector Machine* dan *K-Nearest Neighbour*. Pada tahapan ini diperlukan pemahaman tentang besaran curah hujan, jenis-jenis intensitas hujan berdasarkan curah hujan, dan pada level mana besaran curah hujan berpotensi menyebabkan terjadinya banjir. Proses memahami data dimulai dari eksplorasi data curah hujan dan variabel-variabel terkait lainnya yang berada di dalam penelitian ini. Dataset penelitian terdiri dari tujuh atribut yang meliputi *Tanggal*, *Tmin*, *Tmax*, *Tavg*, *RHavg*, *ddd_x*, dan *RR*.

2.3 Data Preparation

Pada tahapan kali ini, dilaksanakan perbaikan pada data sebelum dilakukan penerapan model. Berikut adalah beberapa proses yang dilakukan:

- a. Proses pembersihan data (cleansing), yaitu proses koreksi pada suatu data di dalam dataset. Dilakukan konversi data ke tipe data numerik agar dapat dilakukan operasi pada proses modeling. Proses ini juga melakukan penghapusan baris yang memiliki atribut bernilai missing value (NaN) pada data curah hujan (RR), sementara untuk atribut missing value selain data curah hujan, akan dilakukan penempatan data dengan nilai rata-rata data tersebut.
- b. Klasifikasi curah hujan, yakni dengan memberikan label pada masing-masing atribut curah hujan berdasarkan besaran nilai data tersebut. Berdasarkan kategori yang didapatkan melalui BMKG, penelitian ini mengkategorikan curah hujan yang berpotensi mengakibatkan banjir ke dalam tiga kategori, yakni “Tidak Berpotensi Banjir”, “Waspada”, dan “Berpotensi Banjir” yang masing-masing diberi label 0, 1, dan 2.

2.4 Modeling

Tahapan ini sudah memasuki pada analisis statistik dasar seperti mean, median, dan distribusi data untuk memahami karakteristik data. Selanjutnya, dilakukan proses pembagian data training dan testing yang akan diterapkan pada metode *Support Vector Machine* dan *K-Nearest Neighbour*. Pemilihan algoritma *Support Vector Machine* (SVM) untuk membangun model prediksi curah hujan dan potensi banjir meliputi implementasi algoritma berdasarkan dataset yang telah disesuaikan. Setelah mendapatkan nilai akurasi model, maka dilanjutkan dengan melakukan proses evaluasi terhadap hasil kinerja model dengan metode *K-Fold Cross Validation*

2.5 Evaluation

Hasil dari proses penerapan model selanjutnya adalah melakukan pengukuran kinerja model menggunakan *K-Fold Cross Validation*. Pada tahap ini perlu dilakukan analisis hasil evaluasi untuk memahami seberapa baik model dapat memprediksi potensi banjir berdasarkan curah hujan. Hasil evaluasi dapat menjabarkan rentang akurasi berdasarkan jumlah iterasi lipatan proses evaluasi. Evaluasi ini membantu dalam menentukan apakah model sudah memenuhi tujuan bisnis yang telah ditetapkan.

2.6 Deployment

Tahapan akhir ini merupakan proses yang tidak selamanya dilakukan, namun jika model ini digunakan untuk keperluan lain yang sesuai dengan tujuan bisnis, maka model ini akan diterapkan. Implementasi model yang telah dievaluasi dapat diterapkan dalam lingkungan praktis, seperti sistem peringatan dini atau sistem manajemen risiko banjir. Untuk memastikan keberlanjutan dari manfaat penggunaan model ini, diperlukan adanya diskusi tentang langkah-langkah yang diperlukan untuk memastikan model dapat digunakan secara efektif dan berkelanjutan. Bila ada fluktuasi atau ketidakakuratan dari model yang telah dikembangkan, maka dapat dilakukan pemeliharaan, perbaikan, atau pembaharuan model di masa depan sesuai dengan perubahan data atau kebutuhan bisnis.

3 Hasil dan Pembahasan

Data didapatkan melalui situs web Data Online BMKG pada <https://dataonline.bmkg.go.id/home>. Data cuaca merupakan data harian yang dimulai dari Januari 2018 hingga Februari 2024. Setelah dilakukan penggabungan keseluruhan masing-masing file data bulanan, maka diketahui data yang dipakai di dalam penelitian ini memiliki 2191 baris dan 7 kolom dengan masing-masing atribut berupa *Tanggal*, *Tmin*, *Tmax*, *Tavg*, *RHavg*, *ddd_x*, dan *RR*. Alasan penggunaan data sekunder adalah ketersediaan dan aksesibilitas data. Data sekunder biasanya sudah melalui proses validasi dan verifikasi oleh pihak yang mengumpulkannya, seperti BMKG, sehingga dapat dipercaya dan diandalkan.

Tabel 3.1. Keterangan dari atribut setiap kolom pada dataset cuaca harian Provinsi DKI Jakarta

No.	Kolom	Keterangan
1	<i>Tanggal</i>	Tanggal perekaman data cuaca harian Provinsi DKI Jakarta dari Januari 2018 hingga Februari 2024
2	<i>Tmin</i>	Temperatur minimum
3	<i>Tmax</i>	Temperatur maksimum
4	<i>Tavg</i>	Temperatur rata-rata
5	<i>RHavg</i>	Kelembapan rata-rata
6	<i>ddd_x</i>	Arah angin pada saat kecepatan maksimum
7	<i>RR</i>	Curah hujan

Tabel 3.2. Dataset cuaca harian Provinsi DKI Jakarta

	<i>Tanggal</i>	<i>Tmin</i>	<i>Tmax</i>	<i>Tavg</i>	<i>RHavg</i>	<i>ddd_x</i>	<i>RR</i>
0	01-09-2018	24	32.8	27.6	77	320	10.6
1	02-09-2018	25	32.2	27.5	80	310	8888
2	03-09-2018	24	31.8	27.8	79	300	17.6
3	04-09-2018	26	31.2	27.9	65	300	0
4	05-09-2018	26	31.8	28.5	65	120	0
...
2186	27-08-2020	25.8	34	29	64	340	nan
2187	28-08-2020	26.2	34.4	29.3	64	100	0
2188	29-08-2020	26.4	34.4	29.9	64	80	0
2189	30-08-2020	26.4	33.8	29.2	68	320	nan
2190	31-08-2020	26.8	33	29.3	68	300	0

3.1 Praproses Data

3.1.1 Import Library

Tahapan praproses data diawali dengan melakukan impor terhadap library yang dibutuhkan pada penelitian prediksi curah hujan potensi banjir, yakni *pandas* untuk memproses data, *numpy* untuk proses aritmatika numerik, dan *sklearn* untuk split data training dan testing, klasifikasi dengan algoritma *support vector machine* dan *k-nearest neighbour*, serta evaluasi kinerja menggunakan *k-fold cross validation*.

3.1.2 Encoding

Tahapan proses pada penelitian ini selanjutnya adalah *encoding*, yang bertujuan untuk melakukan konversi data ke dalam tipe data numerik agar dapat dilakukan pemodelan *machine learning* dengan memperhatikan apabila terdapat nilai non-numerik di dalamnya yang tidak dapat diubah sehingga dapat menyebabkan kesalahan saat konversi, maka diatur menggunakan “errors='coerce'” untuk semua atribut. Pada penelitian ini, seluruh atribut terlibat dalam proses encoding sehingga menghasilkan tipe data yang sama secara komprehensif.

3.1.3 Data Cleaning

Pra Proses selanjutnya adalah proses pembersihan data yang dilakukan dengan cara melakukan penghapusan baris data yang bernilai *missing value* pada kolom *RR*. Untuk nilai atribut lainnya yang juga masih bernilai *missing value*, maka nilai rata-rata atribut akan mengisi data tersebut. Setelah proses *data cleaning*, diketahui data yang valid terdapat sebanyak 1773 baris data.

3.2 Klasifikasi Data

Pada tahapan ini, dilaksanakan proses klasifikasi curah hujan yang memiliki potensi menyebabkan banjir. Penelitian ini mengklasifikasikan untuk curah hujan yang bernilai di bawah atau sama dengan 50 mm/hari, maka diklasifikasikan tidak berpotensi menyebabkan banjir. Namun jika curah hujan bernilai lebih dari 50 mm/hari dan kurang dari 100 mm/hari, maka akan diklasifikasikan sebagai peringatan/waspada banjir. Selebihnya, maka akan diklasifikasikan berpotensi banjir. Hasil klasifikasi tersebut didapati adalah sebagai berikut:

Tabel 3.3. Hasil klasifikasi besaran curah hujan yang berpotensi mengakibatkan banjir

	<i>potensi banjir</i>
0	1599
1	40
2	134

Dari hasil klasifikasi tersebut didapat sebanyak 1599 data curah hujan tidak berpotensi banjir, 40 data curah hujan waspada banjir, dan 134 data curah hujan berpotensi banjir.

3.3 Pemisahan Data Training dan Testing

Pemisahan data latih (*training*) dan data uji (*testing*) dengan menetapkan ukuran untuk data testing sebanyak 20%, sehingga data training berjumlah 80% dari data yang dipakai dalam penelitian. Hal ini dilakukan untuk melatih model pada jumlah data yang besar, sehingga diharapkan model yang dibangun memiliki akurasi yang tinggi.

3.4 Normalisasi Data

Pada tahapan ini, terjadi proses normalisasi data latih X_{train} dan data uji X_{test} memiliki rata-rata nol dan varians satu. Hal ini perlu dilakukan karena beberapa algoritma *machine learning*, termasuk SVC, dapat memberikan hasil yang lebih baik ketika fitur-fiturnya berada dalam rentang yang seragam. Dengan dilakukannya normalisasi data, maka dapat dipastikan bahwa skala semua fitur dalam data latih dan data uji seragam, sehingga pemodelan dari algoritma yang akan diterapkan dapat memberikan bobot yang setara dan hasil yang tidak bias terhadap fitur-fitur dengan skala yang lebih besar.

3.5 Pemodelan

3.5.1 Support Vector Machine

Implementasi algoritma SVM pada penelitian ini diawali dengan inisialisasi model *Support Vector Classifier* (SVC) menggunakan konfigurasi secara default dalam membangun model yang akan digunakan untuk klasifikasi. Model SVC kemudian dilatih menggunakan data latih yang telah dinormalisasi X_{train_scaled} dan label yang sesuai y_{train} . Pada proses ini, model mempelajari pola dari data latih. Setelah model dilatih, skor akurasi dari model tersebut dihitung menggunakan data uji yang telah dinormalisasi, yaitu X_{test_scaled} . Skor akurasi memberikan indikasi sejauh mana baik model dapat melakukan prediksi dengan baik untuk data baru. Hasil skor akurasi prediksi curah hujan yang berpotensi mengakibatkan bencana banjir di wilayah DKI Jakarta menggunakan algoritma *Support Vector Machine* pada penelitian ini adalah sebesar 90.14%.

3.5.2 K-Nearest Neighbour

Implementasi algoritma K-NN pada penelitian ini diawali dengan pembuatan objek model *KNeighborsClassifier* dari modul *neighbors* di *scikit-learn*. Penelitian ini menetapkan parameter `n_neighbors` ke 3, karena terdapat tiga klasifikasi curah hujan yang berpotensi menyebabkan banjir. Selanjutnya dilakukan pelatihan model K-NN menggunakan metode `fit()`. Data latih yang telah dinormalisasi (X_{train_scaled}) digunakan bersama dengan label kelas yang sesuai (y_{train}). Pada proses ini, model mempelajari pola dari data latih untuk membuat prediksi yang baik. Setelah model dilatih, dilakukan evaluasi kinerja model menggunakan data uji yang telah dinormalisasi (X_{test_scaled}) dan label kelas yang sesuai (y_{test}). Nilai akurasi model merupakan persentase dari prediksi yang benar dibandingkan dengan total jumlah sampel uji. Hasil skor akurasi prediksi besaran curah hujan yang memiliki potensi mengakibatkan bencana banjir di wilayah DKI Jakarta menggunakan algoritma *K-Nearest Neighbour* pada penelitian ini adalah sebesar 89.01%.

3.6 Evaluasi Model

Evaluasi hasil kerja model yang diterapkan pada penelitian ini menggunakan metode *K-Fold Cross Validation*. Tidak hanya untuk SVM, penggunaan algoritma *K-Fold Cross Validation* juga untuk mengukur kinerja algoritma K-NN secara lebih menyeluruh dengan melakukan pembagian dataset menjadi data latih (*training*) dan data uji (*testing*) [13]. Metode ini digunakan dalam memperkirakan kesalahan prediksi untuk evaluasi kinerja model [14]. *K-Fold Cross Validation* adalah metode yang umum digunakan dalam melakukan evaluasi kinerja model *machine*

learning untuk memvalidasi kinerja secara obyektif. Tujuan utamanya adalah untuk mengukur sejauh mana model akan bekerja dengan baik pada data baru. Dengan membagi data menjadi beberapa lipatan atau "fold" yang berbeda, metode ini memungkinkan untuk melatih model pada subset data dan menguji kinerjanya pada subset lain, sehingga setiap sampel data digunakan baik untuk pelatihan maupun pengujian. Dengan demikian, evaluasi model menjadi lebih obyektif karena model diuji pada data yang tidak digunakan selama pelatihan.

3.6.1 Support Vector Machine

Evaluasi algoritma SVM pada proses penelitian ini dibuat dengan membuat *k-fold cross validation* dengan 5 lipatan. *K-fold cross validation* melakukan pembagian data menjadi beberapa *k* subset yang sama besar, sedangkan *Stratified K-Fold* memastikan bahwa distribusi kelas dalam setiap lipatan sama dengan distribusi kelas dalam keseluruhan dataset. Urutan data pengujian algoritma ini juga diacak sebelum membaginya menjadi beberapa lipatan untuk pengujian. Selanjutnya, iterasi dilakukan untuk setiap lipatan yang dihasilkan oleh *Stratified K-Fold*. Pada setiap iterasi, model SVC dilatih pada data latih dari lipatan saat ini dan kemudian diuji pada data uji dari lipatan yang sama. Setelah semua lipatan dievaluasi, luaran berupa nilai akurasi dari masing-masing lipatan, nilai maksimum dan minimum, nilai rata-rata, serta besaran standar deviasi akurasi dari semua lipatan akan dicetak.

Tabel 3.4. Hasil akurasi model SVM berdasarkan evaluasi menggunakan *K-Fold Cross Validation*

Data	Accuracy
Accuracy Fold-1	90.14%
Accuracy Fold-2	90.14%
Accuracy Fold-3	90.14%
Accuracy Fold-4	90.40%
Accuracy Fold-5	90.11%

Dari hasil evaluasi ini, didapatkan bahwa akurasi model *Support Vector Machine* berada dalam angka yang cenderung stabil pada evaluasi dengan metode *K-Fold Cross Validation*. Dari akurasi masing-masing lipatan (*fold*) pada evaluasi model, didapatkan akurasi maksimum sebesar 90.40%, akurasi minimum sebesar 90.11%, dan akurasi rata-rata sebesar 90.19%. Hasil pengujian juga menyajikan nilai standar deviasi hasil pengujian model sebesar 0.0012, yang menunjukkan bahwa kinerja model sangat stabil dan konsisten di berbagai subset data. Nilai standar deviasi yang rendah menunjukkan bahwa variasi dalam kinerja model di antara lipatan sangat kecil, dan model memberikan hasil yang serupa pada setiap lipatan. selama pelatihan.

3.6.2 K-Nearest Neighbour

Evaluasi algoritma K-NN pada proses penelitian ini dilakukan dengan membuat *k-fold cross validation* dengan 5 lipatan. *Stratified K-Fold cross validation* diinisialisasi dengan menetapkan jumlah lipatan menjadi 5 dan melakukan pengacakan data sebelum pengujian. Selanjutnya, proses iterasi dilakukan untuk setiap lipatan yang dihasilkan oleh *Stratified K-Fold cross validation*. Pada setiap iterasi, model K-NN dilatih pada data latih dari lipatan saat ini dan kemudian diuji pada data uji dari lipatan yang sama. Setelah semua lipatan dievaluasi, luaran berupa nilai akurasi dari masing-masing lipatan, nilai maksimum dan minimum, nilai rata-rata, serta besaran standar deviasi akurasi dari semua lipatan akan dicetak.

Tabel 3.5. Hasil akurasi model K-NN berdasarkan evaluasi menggunakan *K-Fold Cross Validation*

Data	Accuracy
Accuracy Fold-1	88.45%
Accuracy Fold-2	89.01%
Accuracy Fold-3	87.32%
Accuracy Fold-4	88.98%
Accuracy Fold-5	87.85%

Dari hasil evaluasi ini, didapatkan bahwa akurasi model *K-Nearest Neighbour* berada dalam angka yang cenderung fluktuatif pada evaluasi hasil kerja model menggunakan metode *K-Fold Cross Validation*. Dari akurasi masing-masing lipatan (*fold*) pada evaluasi model, didapatkan akurasi maksimum sebesar 89.01%, akurasi minimum sebesar 87.32%, dan akurasi rata-rata sebesar 88.32%. Hasil pengujian juga menyajikan nilai standar deviasi hasil pengujian model sebesar 0.0073, yang menunjukkan bahwa kinerja model cukup stabil di berbagai subset data.

Nilai standar deviasi yang rendah menunjukkan bahwa variasi dalam kinerja model di antara lipatan sangat kecil, dan model memberikan hasil yang serupa pada setiap lipatan.

4 Kesimpulan

4.1 Kesimpulan

Berdasarkan hasil evaluasi dari tingkat keakuratan metode *Support Vector Machine* dan *K-Nearest Neighbour* untuk melakukan proses prediksi potensi banjir, didapatkan bahwa algoritma SVM memiliki tingkat akurasi yang lebih baik dalam melakukan prediksi dibandingkan K-NN dengan jumlah data latih dan data uji yang sama.

- a) Model SVM menunjukkan kinerja yang stabil dan konsisten pada evaluasi hasil kerja model dengan *K-Fold Cross Validation*. Akurasi model SVM memiliki rentang yang relatif sempit, dengan akurasi maksimum sebesar 90.40%, akurasi minimum sebesar 90.11%, dan akurasi rata-rata sebesar 90.19%. Standar deviasi hasil pengujian yang rendah, yaitu sebesar 0.0012, menunjukkan bahwa variasi dalam kinerja model di antara lipatan sangat kecil, dan model memberikan hasil yang serupa pada setiap lipatan.
- b) Model K-NN menunjukkan fluktuasi dalam kinerja pada evaluasi hasil kerja model dengan *K-Fold Cross Validation*. Akurasi model K-NN memiliki rentang yang sedikit lebih luas dibandingkan SVM, dengan akurasi maksimum sebesar 89.01%, akurasi minimum sebesar 87.32%, dan akurasi rata-rata sebesar 88.32%. Meskipun fluktuatif, standar deviasi hasil pengujian model K-NN juga relatif rendah, yaitu sebesar 0.0073, menunjukkan bahwa kinerja model cukup stabil di berbagai subset data.

Secara keseluruhan, meskipun kedua model menunjukkan kinerja yang baik dalam memprediksi potensi banjir, namun model SVM memiliki keunggulan dalam stabilitas dan konsistensi kinerja dibandingkan K-NN, yang dibuktikan oleh nilai standar deviasi yang lebih rendah dan rentang akurasi dari masing-masing lipatan pada evaluasi *K-Fold Cross Validation*.

4.2 Saran

Berdasarkan penelitian yang dilakukan, disarankan pada penelitian selanjutnya untuk melakukan *tuning* pada *hyperparameter* kedua model guna meningkatkan kinerja prediksi. Untuk model *Support Vector Machine* (SVM), optimalisasi *hyperparameter* seperti *kernel function*, parameter C untuk regularisasi, dan parameter lainnya dapat dilakukan untuk memperbaiki kemampuan model dalam memisahkan kelas potensi banjir. Sementara itu, pada model *K-Nearest Neighbour* (K-NN), penting untuk menyesuaikan jumlah tetangga dan metrik jarak yang digunakan untuk membangun model. Metode *cross validation* seperti *K-Fold Cross Validation* diperlukan untuk *tuning hyperparameter* untuk memastikan evaluasi kinerja model yang obyektif. Dengan demikian, melalui penyesuaian *hyperparameter* dan evaluasi yang cermat, diharapkan kinerja kedua model dapat ditingkatkan, meningkatkan kemampuan model dalam melakukan prediksi potensi banjir.

Referensi

- [1] Razikin, Pahrul, Rosalina Kumalawati, and Deasy Arisanty. 2017. "Strategi Penanggulangan Bencana Banjir Berdasarkan Persepsi Masyarakat Di Kecamatan Barabai Kabupaten Hulu Sungai Tengah." *JPG (Jurnal Pendidikan Geografi)* 4(1): 27–39.
- [2] Pendidikan, Jurnal, Lingkungan Dan, and Pembangunan Berkelanjutan. 2021. "Artikel Info ABSTRAK Received: 27." 22: 50–60.
- [3] Alawiyah, Ana Mardhiyana, and Harintaka Harintaka. 2021. "Identifikasi Genangan Banjir Di Wilayah DKI Jakarta Menggunakan Citra Satelit Sentinel-1." *JGISE: Journal of Geospatial Information Science and Engineering* 4(2): 95.
- [4] Nugroho, Sutopo Purwo. 2002. "Evaluasi Dan Analisis Curah Hujan Sebagai Faktor Penyebab Bencana Banjir Jakarta." *Jurnal Sains & Teknologi Modifikasi Cuaca* 3(2): 91–97.
- [5] Jannah, Miftahul, Joko Sujono, and Adam Pamudji Raharjdo. 2023. "Kajian Perubahan Iklim Di Dki Jakarta Berdasarkan Data Curah Hujan." *Teknisia* 28(1): 44–54.
- [6] Prakoso, Dipa. 2018. "Analisis Pengaruh Tekanan Udara, Kelembaban Udara Dan Suhu Udara Terhadap Tingkat Curah Hujan Di Kota Semarang." *Jurnal Universitas Negeri Semarang*: 1–77. <http://lib.unnes.ac.id/id/eprint/36742>.
- [7] Kusmardiyanti, Retni, Mustaid Yusuf, Djayus Djayus, and Rahmiati Rahmiati. 2022. "Studi Pengaruh Suhu Permukaan Laut Di Selat Makassar Terhadap Intensitas Curah Hujan Kota Balikpapan." *Geosains Kutai Basim* 5(2).
- [8] Al-Azki, Muhammad Wildan, Nurul Hitayuwana, Zulfa Aulia Khususna, and Edy Widodo. 2019. "Analisis Temperature Dan Kelembaban Terhadap Curah Hujan Di Kabupaten Sleman Provinsi Daerah Istimewa Yogyakarta." "Seminar Nasional Teknologi" *Creative and Innovative Education In The Industry 4.0: The Current Trends*: 77–85. http://www.academia.edu/download/60736032/Prosiding_Semnas_UNY20190929-115443-19fobt.pdf#page=97.

- [9] Septiani, Nadia. 2024. "Pengaruh Suhu, Kelembaban Udara Terhadap Prediksi Curah Hujan Dan Relevansi Pada Fenomena Hujan Es Di Bandar Lampung."
- [10] Pradipta, Nur, Pasukat Sembiring, and Pengarapen Bangun. 2013. "Analisis Pengaruh Curah Hujan Di Kota Medan." *Saintia Matematika* 1(5): 459–68.
- [11] Pushpita Anna Octaviani, Yuciana Wilandari, Dwi Ispiriyanti. 2014. "Penerapan Metode SVM Pada Data Akreditasi Sekolah Dasar Di Kabupaten Magelang." *Jurnal Gaussian* 3(8): 811–20.
- [12] Soleha, Romi et al. 2021. "IJCIT (Indonesian Journal on Computer and Information Technology) Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa." *IJCIT (Indonesian Journal on Computer and Information Technology)* 6(2): 118–27.
- [13] Hafid, Hardianti. 2023. "Penerapan K-Fold Cross Validation Untuk Menganalisis Kinerja Algoritma K-Nearest Neighbor Pada Data Kasus Covid-19 Di Indonesia." *Journal of Mathematics* 6(2): 161–68.
<http://www.ojs.unm.ac.id/jmathcos>.
- [14] Mardiana, Linda, Dadan Kusnandar, and Neva Satyahadewi. 2022. "Analisis Diskriminan Dengan K Fold Cross Validation Untuk Klasifikasi Kualitas Air Di Kota Pontianak." *Buletin Ilmiah Mat. Stat. dan Terapannya (Bimaster)* 11(1): 97–102.