



## Perbandingan Algoritma Klasifikasi Data Mining dalam Memprediksi Curah Hujan di Jawa Barat

Salsabila Rahadatul Aisy<sup>1</sup>, Maulana Kusuma Ramadhan<sup>2</sup>, Arizqa Shafa Salsabila<sup>3</sup>,  
Robert Kurniawan<sup>4</sup>

<sup>1,2,3</sup>Prodi Statistika, Politeknik Statistika STIS, Jakarta, Indonesia

<sup>4</sup>Prodi Komputasi Statistik, Politeknik Statistika STIS, Jakarta, Indonesia

<sup>2</sup>[212112181@stis.ac.id](mailto:212112181@stis.ac.id)

<sup>3</sup>[212111922@stis.ac.id](mailto:212111922@stis.ac.id)

<sup>4</sup>[robertk@stis.ac.id](mailto:robertk@stis.ac.id)

Corresponding author email: <sup>1</sup>[212112347@stis.ac.id](mailto:212112347@stis.ac.id)

**Abstract:** This study compares three classification algorithms: Support Vector Machine (SVM), Random Forest (RF), and Artificial Neural Network (ANN) to predict rainfall in West Java in 2022. The climate data used were obtained from the Google Earth Engine (GEE) API, including rainfall, wind speed, temperature, humidity, and sunlight duration. Performance evaluation was conducted using accuracy, precision, recall, F1-score, and ROC curve metrics. The results show that ANN with 10 hidden layers provided the best results with an accuracy of 90.28%. Wind speed was found to have the most significant influence on rainfall. This study recommends using different algorithms and data ratios to obtain more optimal classification results.

**Keywords:** Rainfall, Support Vector Machine, Random Forest, Artificial Neural Network, Performance Evaluation

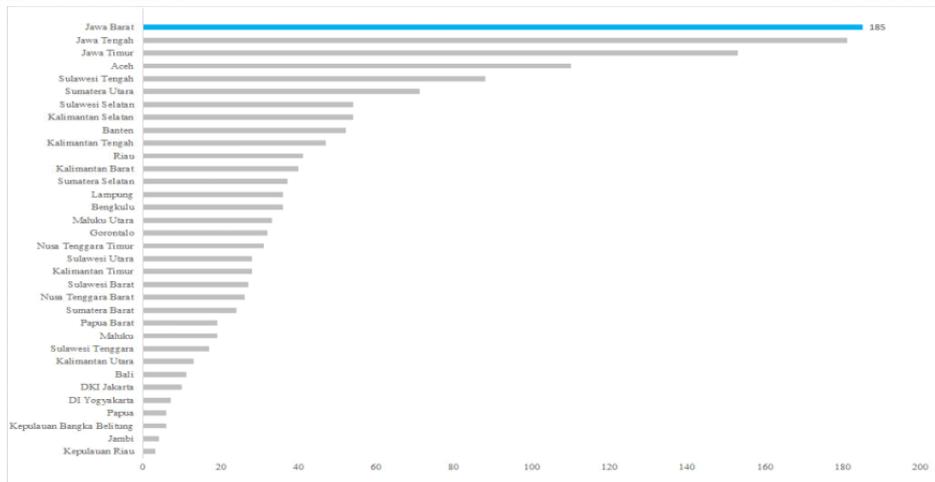
**Abstrak:** Penelitian ini membandingkan tiga algoritma klasifikasi, yaitu Support Vector Machine (SVM), Random Forest (RF), dan Artificial Neural Network (ANN) untuk memprediksi curah hujan di Jawa Barat pada tahun 2022. Data iklim yang digunakan diperoleh dari API Google Earth Engine (GEE), mencakup curah hujan, kecepatan angin, suhu, kelembapan, dan lama penyinaran matahari. Evaluasi kinerja dilakukan dengan menggunakan metrik akurasi, presisi, recall, F1-score, dan kurva ROC. Hasil penelitian menunjukkan bahwa ANN dengan 10 hidden layer memberikan hasil terbaik dengan akurasi 90,28%. Variabel kecepatan angin memiliki pengaruh terbesar terhadap curah hujan. Penelitian ini menyarankan penggunaan algoritma dan rasio data yang berbeda untuk mendapatkan hasil klasifikasi yang lebih optimal.

**Kata kunci:** Curah Hujan, Support Vector Machine, Random Forest, Artificial Neural Network, Evaluasi Kinerja

### I. PENDAHULUAN

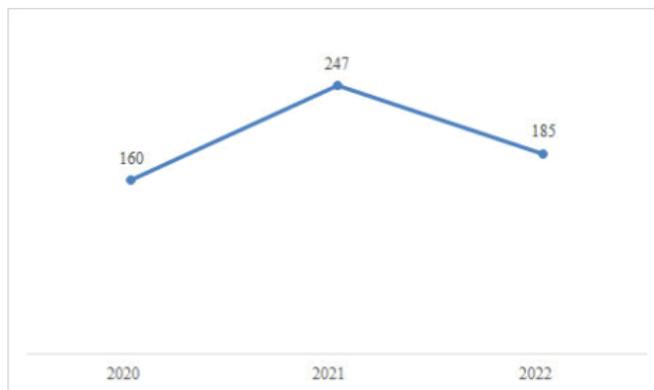
Sampai saat ini, Indonesia masih memiliki berbagai masalah lingkungan yang harus dihadapi. Sebagai negara tropis yang memiliki dua musim, yakni musim penghujan dan musim kemarau, Indonesia juga harus berhadapan dengan masalah lingkungan yang disebabkan oleh perubahan pola curah hujan yang juga disebabkan karena perubahan iklim [1]. Pola curah hujan yang tidak menentu dapat berdampak pada terjadinya banjir yang akan menghambat segala aspek kehidupan manusia.

Badan Pusat Statistik (BPS) mencatat bahwa pada tahun 2022, Provinsi Jawa Barat merupakan wilayah dengan jumlah kejadian banjir tertinggi di Indonesia (Gambar 1). Kondisi tersebut ternyata sudah dialami Jawa Barat selama beberapa tahun terakhir. Sejak tahun 2020 hingga tahun 2022, Jawa Barat terus menempati peringkat pertama sebagai wilayah dengan kejadian banjir terbanyak. Berdasarkan Gambar 2, dapat dilihat bahwa pada tahun 2020, jumlah kejadian banjir di Provinsi Jawa Barat tercatat sebanyak 160 kejadian, kemudian meningkat cukup signifikan pada tahun 2021 menjadi sebanyak 247 kejadian, hingga pada tahun 2022 menurun menjadi 185 kejadian. Meskipun sempat mengalami penurunan, apabila kondisi ini terus menerus terjadi, maka akan banyak kerugian yang dialami, baik dari segi moral maupun materil.



Gambar 1. Jumlah Kejadian Banjir Menurut Provinsi Tahun 2022

Dengan semakin canggihnya teknologi, telah banyak pihak yang memanfaatkan kecanggihannya tersebut sebagai salah satu cara untuk membantu penyelesaian masalah-masalah yang terjadi, salah satunya dengan memanfaatkan metode-metode dalam data mining. Data mining adalah suatu proses membedah demi mendapatkan sebuah pola atau fakta yang tersembunyi [2]. Terdapat beberapa metode yang dapat digunakan dalam data mining, seperti klasifikasi, pengelompokan (*clustering*), asosiasi, generalisasi, dan visualisasi data [3].



Gambar 2. Jumlah kejadian banjir Jawa Barat (2020 - 2022)

Metode klasifikasi dapat diterapkan untuk melakukan prediksi curah hujan yang dapat digunakan untuk memperoleh informasi dalam perencanaan pengelolaan risiko banjir. Dari metode-metode dalam klasifikasi ini, masalah lingkungan yang memiliki hubungan dengan cuaca, khususnya bencana banjir, dapat dipantau karena adanya informasi yang akurat [4]. Diperlukan prediksi curah hujan yang akurat karena aspek-aspek yang mempengaruhi curah hujan akan mengalami perubahan yang dinamis dalam atmosfer [5]. Menurut [29], penyebaran curah hujan berkaitan erat dengan sistem kecepatan angin dan tekanan udara. Selain itu, perubahan tutupan awan sebagai respon dari polusi udara memiliki pengaruh besar terhadap iklim [30]. Faktor lain yang mempengaruhi curah hujan adalah suhu, kelembapan, dan arah angin yang digunakan oleh [31] sebagai parameter dalam meneliti pengaruh tutupan. Oleh karena itu, mendapatkan metode yang sesuai dalam mengklasifikasikan curah hujan di suatu wilayah sangatlah penting.



Di Ghana, dengan menggunakan model klasifikasi berbeda, yakni Decision Tree, Multilayer Perceptron (MLP), Random Forest, Extreme Gradient Boosting (XGB), dan K-Nearest Neighbor (KNN) yang digunakan untuk memprediksi curah hujan di seluruh zona ekologi diperoleh hasil bahwa metode Random Forest, XGB, dan MLP adalah kandidat yang baik dalam prediksi curah hujan di berbagai zona [7]. Penelitian lain yang juga memanfaatkan metode klasifikasi untuk memprediksi curah hujan, yakni dengan metode Feed Forward Neural Network (FFNN) diperoleh hasil prediksi curah hujan dengan akurasi sebesar 93,55% [5]. Di samping itu, sebuah studi yang mencakup variabel prediktif baru dalam prediksi curah hujan, mendapatkan hasil bahwa pendekatan SVM adalah metode terbaik yang ditunjukkan melalui perbandingan dengan metode-metode lain, di antaranya adalah MLP, Extreme Learning Machine, serta algoritma lain, seperti Decision Tree dan K-Nearest Neighbor [8]. Kemudian, dalam [9] metode klasifikasi yang digunakan hanya terbatas dalam metode naive bayes saja dengan akurasi sebesar 92% sehingga tidak terdapat metode lain yang dapat menjadi pembanding dari metode yang digunakan tersebut.

Dari penelitian-penelitian tersebut, dapat diketahui bahwa penggunaan metode-metode dalam data mining memiliki kinerja yang baik dalam hal melakukan prediksi atau klasifikasi curah hujan. Oleh karena itu, penting untuk menyelidiki berbagai algoritma klasifikasi yang terbaik dalam pengklasifikasian curah hujan agar dapat dimanfaatkan dalam upaya pencegahan banjir di Jawa Barat. Dengan demikian, penelitian ini bertujuan untuk membandingkan metode klasifikasi dalam data mining, meliputi *Support Vector Machine*, *Random Forest*, dan *Artificial Neural Network* untuk mendapatkan metode terbaik dalam mengklasifikasikan curah hujan di Jawa Barat dengan nilai-nilai evaluasi antara lain, yakni akurasi, presisi, *recall*, *F1-Score* dan kurva AUC. Serta, untuk melihat variabel yang paling berpengaruh terhadap curah hujan melalui nilai importance variabel.

## II. METODE PENELITIAN

### 2.1 Support Vector Machine

Support Vector Machine (SVM) adalah sebuah model machine learning yang berpotensi untuk memberikan solusi dalam masalah regresi, prediksi, dan klasifikasi [10]. Pada SVM, persamaan yang digunakan untuk memisahkan klasifikasi dengan menggunakan hyperplane yang optimum seperti  $f(x) = w \cdot x + b$ , dengan  $w$  sebagai beban vektor dari  $x$  yang akan berfungsi untuk memperkirakan vektor  $Y$  dengan vektor pembobotnya yang memiliki  $n \times n$  dimensi dan  $b$  parameter bias. Untuk data yang tidak dipisahkan secara linear, dalam memperoleh hyperplane yang optimal dapat menggunakan fungsi kernel. Beberapa jenis kernel yang biasanya digunakan diantaranya ada linear, radial basis function (RBF), polynomial, dan sigmoid [11]. Dari beberapa jenis kernel tersebut nantinya akan ditentukan kernel mana yang menghasilkan akurasi yang paling baik. Dalam menganalisis kernel terbaik dipergunakan parameter cost dan gamma yang telah dioptimalkan [12].

### 2.2 Random Forest

Random Forest (RF) adalah teknik machine learning yang termasuk salah satu metode yang baik untuk melakukan klasifikasi dan regresi. Random forest merupakan versi yang lebih baik dari decision tree tergantung dari strategi *bagging* nya. Dalam *bagging*, banyak situasi kesalahan stokastik dilakukan dengan memilih set pelatihan secara independen dan sewenang-wenang dari set pelatihan; menjumlahkan antisipasi dari setiap model train independen dengan mengambil nilai rata-ratanya. Random forest dapat mengevaluasi pentingnya pengungkapan faktor-faktor terhadap estimasi yang sesuai dengan pedoman yang berlaku, “semakin relevan variabel ilustrasinya, semakin signifikan dampaknya terhadap prediksi” untuk menggunakan model RF dalam pemilihan faktor [13]. Salah satu



faktor yang dapat mempengaruhi hasil prediksi random forest adalah dengan menentukan jenis parameter yang digunakan, di antaranya adalah *mtry* parameter [14].

### 2.3 Artificial Neural Network

*Artificial Neural Networks* (ANN), atau Jaringan Saraf Tiruan (JST), adalah sistem pemrosesan informasi yang dirancang untuk meniru cara kerja otak manusia dalam menyelesaikan masalah melalui proses belajar dengan mengubah bobot sinapsisnya. ANN didasarkan pada kemampuan otak manusia dalam mengorganisasi neuron untuk melaksanakan tugas-tugas tertentu, khususnya pengenalan pola dengan efektivitas tinggi. ANN adalah generalisasi model matematis dari pemahaman manusia (human cognition) dan berfungsi untuk memodelkan hubungan kompleks antara input dan output guna menemukan pola-pola data. ANN bekerja sebagai sistem komputasi cerdas dalam bidang machine learning, dengan algoritma yang terdistribusi paralel, terdiri dari unit-unit sederhana, dan mampu menyimpan serta memanfaatkan pengetahuan yang diperoleh secara eksperimen untuk berbagai tujuan. Untuk mencari hasil prediksi yang baik dapat dilakukan dengan menentukan jumlah neuron, arsitektur ANN, dan hidden layer yang sesuai [15].

### 2.4 Dataset

Tabel 1. Deskripsi Data

Parameter	Satuan	Keterangan
Curah Hujan	Millimeter (mm)	Jumlah curah hujan yang tercatat dalam satu hari
Kecepatan Angin	Knot	Kecepatan angin rata-rata
Suhu	Derajat Celsius (°C)	Suhu rata-rata dalam derajat celsius
Kelembapan	Persen (%)	Kelembapan rata-rata
Lama Penyinaran Matahari	Jam	Lama penyinaran matahari dalam sehari

Sementara untuk klasifikasi curah hujan yang digunakan bersumber dari klasifikasi Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), meliputi (1) kategori berawan apabila curah hujan sebesar 0 mm/hari; (2) kategori hujan ringan 0,5 - 20 mm/ hari; (3) kategori hujan sedang 20 - 50 mm/hari; (4) kategori hujan lebat 50 - 100 mm/hari; (5) kategori hujan sangat lebat 100 - 150 mm/hari; dan (6) kategori hujan ekstrem apabila nilai curah hujan >150 mm/hari. Kendati demikian, pada kasus ini, curah hujan akan dibagi ke dalam dua kategori saja, yaitu hujan dan tidak hujan. Termasuk dalam kategori hujan apabila curah hujan > 0 mm/hari, sedangkan kategori tidak hujan apabila curah hujan bernilai 0 mm/hari. Pembagian kategori curah hujan dalam dua kategori ini dilakukan untuk menghindari ketimpangan antar kelas dan menyesuaikan dengan tujuan dalam penelitian.

### 2.5 Tahapan Analisis

Terdapat beberapa tahapan analisis yang dilakukan dalam penelitian seperti yang secara lebih detail dapat terlihat dalam Gambar 3.

#### 2.5.1 Preprocessing

Dalam tahap ini nantinya akan dilakukan pengolahan data dari yang awalnya data mentah menjadi data yang siap untuk diolah. Untuk penelitian ini hanya dilakukan preprocessing secara umum yang tahapannya dengan melakukan pengecekan terhadap ada atau tidaknya *missing value* dari data yang akan digunakan untuk penelitian. Setelah dipastikan datanya aman dari adanya *missing value*, selanjutnya akan dilakukan standarisasi data yang berskala numerik sehingga data numerik tersebut memiliki rentang pengukuran yang sama. Dengan demikian, misklasifikasi dapat diminimalkan.



### 2.5.2 Pembagian Data Training dan Testing

Pembagian data dilakukan untuk menentukan data *training* dan *testing*. Berdasarkan penelitian [6], pembagian data *training* dan *testing* menggunakan perbandingan 80:20 sehingga pada penelitian ini, digunakan pula perbandingan data training dan testing yang sama Adapun data *training* yang digunakan adalah 80% dari total koleksi data 364 curah hujan Provinsi Jawa Barat yang telah mengalami proses pengecekan *missing value*, 20% lainnya adalah data *testing*.

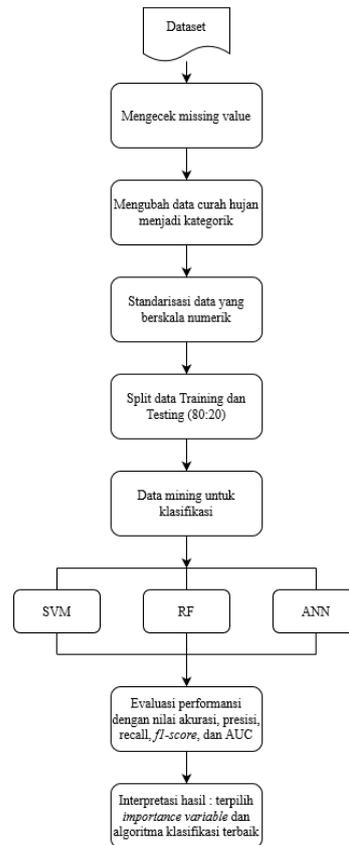
### 2.5.3 Metrik Evaluasi

Untuk mengetahui metode terbaik yang dapat digunakan dalam klasifikasi curah hujan di Jawa Barat pada tahun 2022, terdapat berbagai metrik evaluasi dengan tujuan mengevaluasi efisiensi algoritma-algoritma yang digunakan. Akan tetapi, pada penelitian ini, ukuran evaluasi yang diterapkan, meliputi nilai akurasi, presisi, recall, *f1-score*, *confusion matrix* yang menjadi dasar penghitungan nilai-nilai evaluasi yang telah disebutkan, serta kurva ROC (*Receiver Operating Characteristics*).

**Tabel 2.** Confusion matrix

		Reference	
		Positive (P)	Negative (N)
Prediction	Positive (P)	True Positive (TP)	False Negative (FN)
	Negative (N)	False Negative (FN)	True Negative (TN)

Kurva ROC adalah deskripsi dua dimensi kinerja klasifikasi yang dapat memungkinkan perbandingan klasifikasi dalam hal sensitivitas/spesifisitas [16]. Dalam kurva ROC ini, tersedia nilai skalar tunggal yang dapat menjadi pembanding dari dua atau lebih metode klasifikasi. Ukuran ini biasa disebut dengan luas area di bawah ROC/*Area Under ROC Curve* (AUC). Nilai AUC berada diantara 0,5 sampai 1 karena AUC merupakan luas satuan persegi [17].



Gambar 3. Diagram alir analisis data

## 2.6 Penelitian Terkait

Penelitian mengenai prediksi curah hujan pernah dilakukan oleh [7] dengan judul “Rainfall Prediction Using Machine Learning Algorithms for the Various Ecological Zones of Ghana”. Variabel yang digunakan adalah curah hujan, suhu (minimum dan maksimum), lama penyinaran matahari, dan kecepatan angin. Dalam penelitian ini, curah hujan dikategorikan menjadi hujan dan tidak hujan. Metode yang digunakan adalah Decision Tree, Random Forest, Multiplayer Perceptron, Extreme Gradient Boosting, dan K-Nearest Neighbor. Selain membandingkan metode, [7] juga membandingkan tiga jenis rasio perbandingan data *training* dan *testing*, yaitu 70:30, 80:20, dan 90:10. Secara umum, metode Random Forest, Extreme Gradient Boosting, dan Multiplayer Perceptron menghasilkan kinerja yang baik dalam semua kasus.

Berdasarkan penelitian [5] yang berjudul “Prediction of Rainfall using Artificial Neural Network”, prediksi curah hujan merupakan hal yang sangat krusial dan perlu segera dilakukan karena menyangkut iklim dan bencana banjir yang dapat memengaruhi kehidupan manusia dan perekonomian negara. Penelitian ini menggunakan empat variabel, yakni suhu, tekanan udara, tutupan awan, dan curah hujan. [5] menyimpulkan bahwa metode *Feed Forward Neural Network* (FFNN) mampu memprediksi curah hujan dengan akurasi 93,55 persen.

Penelitian [6] yang berjudul “Prediction of Rainfall Using Data Mining Techniques” membagi data *training* dan *testing* dengan perbandingan 80:20. Metode yang digunakan adalah Support Vector Regression, Random Forest, dan Decision Tree. Penelitian ini menyimpulkan bahwa di antara ketiga metode, Random Forest memberikan hasil terbaik dengan R-Squared 0,981 dan Adjusted R-Squared 0,980.



Penelitian [18] membandingkan metode Random Forest dan Support Vector Machine untuk memprediksi curah hujan. Penelitian ini menggunakan dua jenis model prediksi, yaitu Single-Mode Forecasting Model (SMFM) dan Multiple-Mode Forecasting Model (MMFM). Hasil penelitian menunjukkan bahwa secara keseluruhan, metode Support Vector Machine memiliki kinerja yang lebih baik dalam memprediksi curah hujan dibandingkan Random Forest.

Dalam [2] menggunakan empat algoritma klasifikasi berbeda untuk meramalkan curah hujan, yakni metode Klasifikasi dan Regression Tree, pendekatan Naive Bayes, K-Nearest Neighbor, dan 5-10-1 Pattern Recognition Neural Network. Hasil peramalan keempat metode tersebut selanjutnya diuji akurasi pada dataset uji. Adapun variabel bebas yang digunakan adalah suhu rata-rata, suhu titik embun, kelembapan, tekanan permukaan laut, dan kecepatan angin. Berdasarkan penelitian tersebut didapatkan hasil bahwa Naive Bayes dan Decision Tree lebih mudah dipahami dan digunakan sedangkan pada K-Nearest Neighbor didapatkan kesulitan pada penentuan nilai K. Sementara itu, Pattern Recognition Neural Network memberikan hasil yang lebih baik dibandingkan algoritma lain.

### III. HASIL DAN PEMBAHASAN

#### 3.1 Klasifikasi Curah Hujan dengan Support Vector Machine (SVM)

Pemodelan klasifikasi curah hujan dengan metode SVM dilakukan dengan mencari berbagai kombinasi kernel, gamma, dan juga cost demi mendapatkan akurasi klasifikasi yang terbaik. Pada mulanya berbagai kernel, yang meliputi kernel radial, kernel linear, kernel polynomial, dan kernel sigmoid dimodelkan lalu dilakukan pengujian pada data testingnya. Dari pemodelan tersebut, ternyata diperoleh hasil bahwa model SVM dengan kernel radial memberikan nilai akurasi tertinggi dibandingkan model-model lain dengan kernel yang berbeda, yakni sebesar 0,8194.

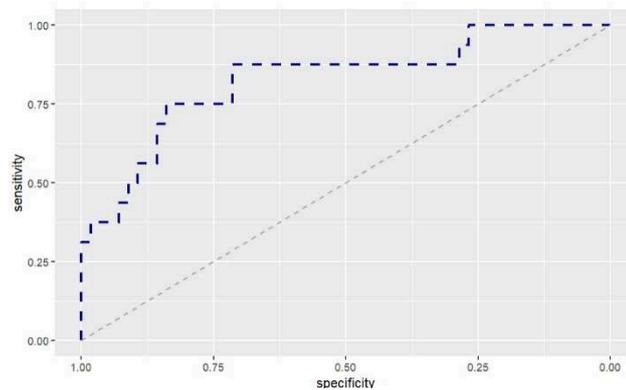
Selanjutnya dijalankan algoritma untuk mendapatkan nilai gamma dan cost terbaik dengan dua metode, yaitu *grid search cross validation* dan *random search cross validation*. Dari pencarian parameter terbaik tersebut, metode *grid search cross validation* memberikan nilai gamma terbaik sebesar 0,5 dengan nilai cost sebesar 1,2589. Sementara dengan metode *random search cross validation*, nilai gamma terbaik yang dihasilkan adalah sebesar 0,0120 dan nilai cost sebesar 173,4965.

**Tabel 4.** Confusion Matrix SVM

		<i>Reference</i>	
		Hujan	Tidak Hujan
<i>Prediction</i>	Hujan	54	10
	Tidak Hujan	2	6

Kemudian, setelah hasil dari masing-masing parameter terbaik ini dimodelkan dengan kernel radial, diketahui dari Tabel 4 bahwa parameter yang didapat dari metode *grid search cross validation* memberikan nilai akurasi yang lebih baik dari pemodelan dengan kombinasi lain, yaitu sebesar 0,8333. Hal ini berarti bahwa metode SVM dengan kernel radial, dan nilai gamma serta cost yang dihasilkan dari proses *grid search cross validation* tersebut dapat melakukan klasifikasi secara tepat sebesar 83,33%. Artinya, dari sebanyak 72 observasi yang merupakan data *testing*, metode SVM terpilih dapat mengklasifikasikan kelas hujan/tidak hujan secara tepat sebanyak 60 observasi. Di samping itu, juga digunakan nilai presisi yang menyatakan ukuran seberapa besar klasifikasi kelas hujan yang dihasilkan oleh model adalah benar-benar merupakan kelas hujan. Pada pemodelan ini, diperoleh nilai presisi

sebesar 0,8438 atau 84,38%. Lalu, apabila melihat nilai recall yang dimiliki oleh metode SVM ini, diketahui bahwa recallnya adalah sebesar 0,9643. Nilai tersebut menunjukkan kemampuan metode ini untuk mendeteksi dari semua hujan yang benar-benar hujan sebesar 96,43% terdeteksi oleh model. Berikutnya, *F1-Score* yang dihasilkan dalam pemodelan ini adalah sebesar 0,9. Nilai *F1-Score* yang mendekati 1 ini menunjukkan kemampuan metode SVM dalam menyeimbangkan presisi dan recall terbilang cukup baik.



**Gambar 4.** Kurva ROC metode SVM

Dari gambar 4, dihasilkan nilai AUC sebesar 0,8281 yang menandakan bahwa hasil klasifikasi curah hujan dengan metode SVM terpilih termasuk dalam kategori baik.

### 3.2 Klasifikasi Curah Hujan dengan Random Forest (RF)

Klasifikasi curah hujan dengan menggunakan metode RF juga akan mencari berbagai kombinasi sehingga didapatkan hasil akurasi yang terbaik. Akan tetapi, pada mulanya pemodelan dilakukan tanpa penyesuaian apapun atau dalam kondisi *default* pada parameternya. Dalam metode ini, parameter yang digunakan adalah *mtry*, yakni banyak pohon yang harus dibuat agar mendapat hasil yang lebih akurat [19]. *Mtry* yang dihasilkan dalam kondisi *default* adalah sebesar akar kuadrat dari banyaknya variabel prediktor yang digunakan. Oleh karena dalam penelitian ini variabel prediktor yang digunakan adalah sebanyak 4 variabel, maka banyaknya *mtry* dalam kondisi *default* adalah 2. Dari pemodelan kemudian dilakukan uji pada data *testing* yang ternyata memberikan nilai akurasi sebesar 0,8056 atau sebesar 80,56%.

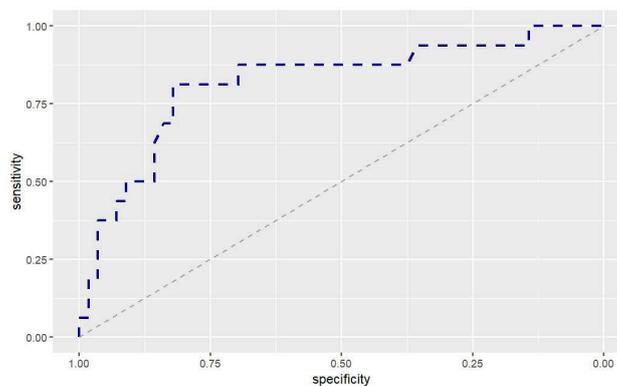
**Tabel 5.** Confusion Matrix RF

	<i>Reference</i>	
	Hujan	Tidak Hujan
<i>Prediction</i> Hujan	51	10
Tidak Hujan	5	6

Berikutnya demi mendapatkan hasil yang lebih baik lagi, dilakukan pencarian parameter terbaik dengan metode *grid search cross validation* dan *random search cross validation* yang keduanya memberikan hasil bahwa parameter terbaik dalam kasus ini adalah dengan *mtry* sebanyak 1. Dengan



ini, maka selanjutnya kembali dibentuk model untuk nantinya dilakukan uji pada data *testing*. Dari pengujian tersebut, dapat dilihat dari tabel V bahwa nilai akurasi yang diperoleh adalah sebesar 0,7917 artinya dari sejumlah data *testing* yang dimiliki, metode RF dengan penyesuaian parameter ini dapat mengklasifikasikan kelas hujan/tidak hujan secara tepat sebesar 79,17%. Kemudian, dari nilai presisi ditunjukkan bahwa diperoleh nilai sebesar 0,8360 yang berarti bahwa metode ini dapat mengklasifikasikan kelas hujan dengan benar sebesar 83,60%. Selanjutnya, jika melihat nilai recall yang dihasilkan, ternyata diperoleh nilai sebesar 0,9107. Nilai recall tersebut menandakan dari semua kelas hujan yang benar-benar hujan, 91,07% dapat terdeteksi oleh model. Untuk nilai *F1-Score* yang dihasilkan model RF ini, diperoleh nilai sebesar 0,8718. Nilai ini dapat dikatakan cukup mendekati 1 sehingga dapat mengimplikasikan bahwa kinerja model dalam menyeimbangkan presisi dan recall juga cukup baik.



**Gambar 5.** Kurva ROC metode RF

Dari gambar 5 dihasilkan nilai AUC sebesar 0,8192 yang menandakan bahwa hasil klasifikasi curah hujan dengan metode RF terpilih juga termasuk dalam kategori baik.

### 3.3 Klasifikasi Curah Hujan dengan Artificial Neural Network (ANN)

Pemodelan klasifikasi curah hujan dengan ANN ini dilakukan langsung menggunakan suatu fungsi yang bertujuan untuk mencari parameter terbaik yang dapat memaksimalkan hasil akurasi dari klasifikasi. Berbeda dengan metode-metode sebelumnya, pada ANN ini metode yang digunakan dalam pencarian parameter terbaik adalah dengan metode *grid search cross validation* saja. Parameter yang akan dicari, yaitu *hidden layer*. Dari hasil pencarian parameter terbaik, diperoleh nilai *size* atau *hidden layer* sebanyak 4. Kendati demikian, hasil akurasi dirasa belum optimal sehingga dilakukan berbagai kombinasi nilai parameter hingga mendapatkan *hidden layer* sebanyak 10.

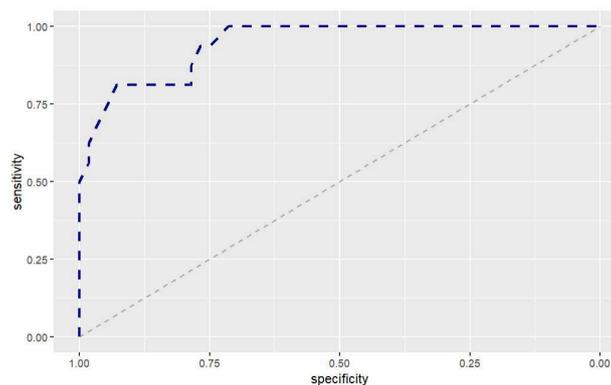
Dari parameter dengan 10 *hidden layer* tersebut, kemudian kembali dibentuk model dari data *training* untuk nantinya dapat dilakukan uji pada data *testing*. Setelah dilakukan pengujian, dapat dilihat berdasarkan tabel VI diperoleh nilai akurasi yang cukup tinggi dibandingkan dengan metode-metode sebelumnya, yakni sebesar 0,9028 yang berarti bahwa dari sebanyak 72 observasi data *testing*, sebesar 90,28% atau sebanyak 65 observasi dapat secara tepat mengklasifikasikan kelas hujan/tidak hujan.



**Tabel 6.** Confusion Matrix ANN

		<i>Reference</i>	
		Hujan	Tidak Hujan
<i>Prediction</i>	Hujan	52	3
	Tidak Hujan	4	13

Di samping itu, diperoleh pula nilai presisi sebesar 0,9455 yang berarti bahwa metode ANN dengan parameter terbaik ini dapat mengklasifikasikan kelas hujan dengan benar sebesar 94,55%. Lalu, didapatkan juga nilai recall sebesar 0,9286 yang menandakan bahwa kemampuan metode ANN untuk mendeteksi kelas hujan dari yang benar-benar hujan dari semua kelas hujan adalah sebesar 92,86%. Selanjutnya juga terdapat nilai *F1-Score* yang bernilai sebesar 0,9369. Nilai ini dapat dilihat akan mendekati 1 sehingga mengartikan bahwa kinerja model dalam menyeimbangkan presisi dan recall terbilang baik.



**Gambar 6.** Kurva ROC metode ANN

Dari gambar 6 dihasilkan nilai AUC sebesar 0,9459 yang menandakan bahwa hasil klasifikasi curah hujan dengan metode RF terpilih juga termasuk dalam kategori yang sangat baik.

### 3.4 Pemilihan Model Terbaik

Setelah dilakukan pemodelan dengan ketiga metode klasifikasi, ditampilkan komparasi yang menampilkan nilai-nilai evaluasi yang diperoleh berdasarkan hasil dari *confusion matrix* sebagaimana dalam Tabel 7. Untuk mengetahui model terbaik yang dapat digunakan dalam mengklasifikasikan curah hujan di Jawa Barat pada tahun 2022, maka digunakan beberapa nilai evaluasi, antara lain nilai akurasi, presisi, *recall*, *F1-Score*, dan AUC (*Area Under The ROC Curve*).

Dapat dilihat dari Tabel 7 bahwa secara umum, ketiga metode klasifikasi memiliki akurasi yang terbilang cukup baik untuk mengklasifikasikan kelas hujan/tidak hujan ke kelas yang benar. Nilai akurasi tersebut berkisar antara 0,79 sampai 0,90. Dengan akurasi minimum adalah 0,7917 (RF), sedangkan akurasi maksimum adalah sebesar 0,9028 (ANN). Karena tujuan dari penelitian adalah untuk mendapatkan metode terbaik dalam mengklasifikasikan curah hujan ke dalam kelas yang tepat dalam upaya mengontrol jumlah kejadian banjir di Jawa Barat, maka diperlukan metode dengan nilai recall yang tinggi. Nilai recall dari metode SVM mengungguli nilai dari recall dua metode lain, yakni bernilai



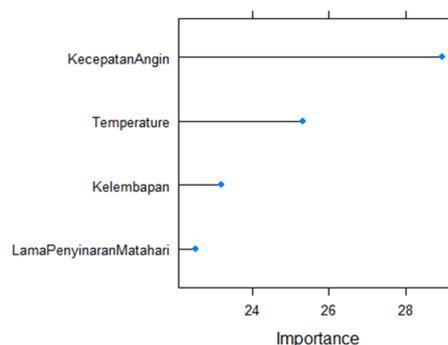
sebesar 96,43%. Sementara nilai recall dari metode RF dan ANN berturut-turut adalah sebesar 91,07% dan 92,86%.

Kendati demikian, ukuran-ukuran evaluasi lainnya, yakni nilai presisi, F1-Score, dan AUC dengan nilai tertinggi ditempati oleh metode klasifikasi ANN sehingga metode ini terpilih sebagai metode klasifikasi terbaik dalam mengklasifikasi curah hujan di Provinsi Jawa Barat. Hasil ini selaras dengan penelitian [5] yang menunjukkan bahwa FFNN, sebagai bentuk paling sederhana dari Neural Network, mampu menjadi algoritma prediktif dalam memprediksi curah hujan dengan akurasi 93,55 persen.

**Tabel 7.** Kinerja dari ketiga metode klasifikasi

Metode	Akurasi	Presisi	Recall	F1-Score	AUC
<i>Support Vector Machine</i>	0,8333	0,8438	0,9643	0,9000	0,8281
<i>Random Forest</i>	0,7917	0,8360	0,9107	0,8718	0,8192
<i>Artificial Neural Network</i>	0,9028	0,9455	0,9286	0,9369	0,9459

Pada penelitian [2] membandingkan empat jenis algoritma berbeda, yaitu Klasifikasi dan Regression Tree, pendekatan Naive Bayes, K-Nearest Neighbor, dan 5-10-1 Pattern Recognition Neural Network. Model terbaik juga didapatkan oleh metode *Pattern Recognition Neural Network* dengan 10 *hidden layer*. Akurasi yang dihasilkan model ini sebesar 82,1 persen. Berdasarkan Gambar 7, diketahui pula *importance variable* dari metode klasifikasi terbaik dalam penelitian ini, *Artificial Neural Network* (ANN), adalah variabel kecepatan angin. Dengan ini, mengimplikasikan bahwa variabel kecepatan angin memiliki pengaruh yang paling besar terhadap curah hujan dibandingkan variabel-variabel lainnya yang meliputi, temperatur, kelembapan, dan lama penyinaran matahari sehingga dapat diambil kebijakan yang berhubungan dengan pencegahan banjir yang disebabkan oleh terjadinya hujan melalui variabel kecepatan angin.



**Gambar 7.** Importance Variabel metode terbaik

#### IV. KESIMPULAN

Dari hasil klasifikasi menggunakan tiga metode yang berbeda tersebut, diperoleh hasil bahwa metode *Artificial Neural Network* (ANN) dengan 10 hidden layer memberikan hasil yang paling baik dengan akurasi sebesar 90,28%. Kemudian, diketahui juga dari metode klasifikasi terbaik (ANN), variabel prediktor yang memiliki pengaruh paling besar terhadap variabel respon, yaitu variabel kecepatan angin. Untuk mendapatkan hasil yang lebih optimal dalam pengklasifikasian curah hujan di



Jawa Barat, maka dapat dilakukan klasifikasi dengan algoritma yang berbeda, dengan rasio data *training* dan *testing* yang juga berbeda, serta dapat pula menambahkan jumlah variabel yang diteliti. Dengan demikian, diketahui metode terbaik dalam memprediksi curah hujan yang kemudian dapat digunakan sebagai antisipasi kejadian banjir di wilayah rawan seperti Jawa Barat.

## REFERENSI

1. I. G. W. Sena, J. W. Dillak, P. Leunupun, and A. J. Santoso, “Predicting Rainfall Intensity using Naïve Bayes and Information Gain Methods (Case Study: Sleman Regency),” *J. Phys. Conf. Ser.*, vol. 1577, no. 1, 2020, doi: 10.1088/1742-6596/1577/1/012011.
2. D. Gupta and U. Ghose, “A comparative study of classification algorithms for forecasting rainfall,” *2015 4th Int. Conf. Reliab. Infocom Technol. Optim. Trends Futur. Dir. ICRITO 2015*, pp. 1–6, 2015, doi: 10.1109/ICRITO.2015.7359273.
3. S. H. Liao, P. H. Chu, and P. Y. Hsiao, “Data mining techniques and applications - A decade review from 2000 to 2011,” *Expert Syst. Appl.*, vol. 39, no. 12, pp. 11303–11311, 2012, doi: 10.1016/j.eswa.2012.02.063.
4. Naufal Rasyid, Trevy Jonatya Novella, and Ahlijati Nuraminah, “Implementasi Metode K-Nearest Neighbour Dalam Memprediksi Curah Hujan Di Kota Bogor,” *Jural Ris. Rumpun Ilmu Tek.*, vol. 1, no. 1, pp. 98–106, 2022, doi: 10.55606/jurritek.v1i1.592.
5. A. Kala and S. G. Vaidyanathan, “Prediction of Rainfall Using Artificial Neural Network,” *Proc. Int. Conf. Inven. Res. Comput. Appl. ICIRCA 2018*, no. Icirca, pp. 339–342, 2018, doi: 10.1109/ICIRCA.2018.8597421.
6. S. R. Devi, “Prediction of Rainfall Using Data Mining Techniques School of Electronics Engineering,” *2018 Second Int. Conf. Inven. Commun. Comput. Technol.*, no. Iccict, pp. 1507–1512, 2018.
7. N. K. A. Appiah-Badu, Y. M. Missah, L. K. Amekudzi, N. Ussiph, T. Frimpong, and E. Ahene, “Rainfall Prediction Using Machine Learning Algorithms for the Various Ecological Zones of Ghana,” *IEEE Access*, vol. 10, pp. 5069–5082, 2022, doi: 10.1109/ACCESS.2021.3139312.
8. E. G. Ortiz-García, S. Salcedo-Sanz, and C. Casanova-Mateo, “Accurate precipitation prediction with support vector classifiers: A study including novel predictive variables and observational data,” *Atmos. Res.*, vol. 139, pp. 128–136, 2014, doi: 10.1016/j.atmosres.2014.01.012.
9. A. A. Rizqi and D. Kusumaningsih, “Klasifikasi Curah Hujan di Kota Bogor Provinsi Jawa Barat dengan Menggunakan Metode Naive Bayes,” *J. Semin. Nas. Mhs. Fak. Teknol. Inf.*, no. September, pp. 542–550, 2022.
10. O. S. Ojo and S. T. Ogunjo, “Machine learning models for prediction of rainfall over Nigeria,” *Sci. African*, vol. 16, p. e01246, 2022, doi: 10.1016/j.sciaf.2022.e01246.
11. L. Tapak, H. Mahjub, O. Hamidi, and J. Poorolajal, “Real-data comparison of data mining methods in prediction of diabetes in Iran,” *Healthc. Inform. Res.*, vol. 19, no. 3, pp. 177–185, 2013, doi: 10.4258/hir.2013.19.3.177.
12. H. Al Azies, D. Trishnanti, and E. Mustikawati P.H, “Comparison of Kernel Support Vector Machine (SVM) in Classification of Human Development Index (HDI),” *IPTEK J. Proc. Ser.*, vol. 0, no. 6, p. 53, 2019, doi: 10.12962/j23546026.y2019i6.6339.
13. P. C. S. Reddy, Y. Sucharitha, and G. S. Narayana, “Development of Rainfall Forecasting Model Using Machine Learning With Singular Spectrum Analysis,” *IJUM Eng. J.*, vol. 23, no. 1, pp. 172–186, 2022, doi: 10.31436/IJUMEng.V23I1.1822.
14. Z. Wang, Y. Wang, R. Zeng, R. S. Srinivasan, and S. Ahrentzen, “Random Forest based hourly building energy prediction,” *Energy Build.*, vol. 171, pp. 11–25, 2018, doi: 10.1016/j.enbuild.2018.04.008.
15. S. Chinchankar, S. Shinde, V. Gaikwad, A. Shaikh, M. Rondhe, and M. Naik, “ANN modelling of surface roughness of FDM parts considering the effect of hidden layers, neurons, and process parameters,” *Adv. Mater. Process. Technol.*, vol. 10, no. 1, pp. 22–32, 2024, doi: 10.1080/2374068X.2022.2091085.
16. J. Tuimala, A. Kallio, F. Red, and C. Blood, *Encyclopedia of Systems Biology*. 2013.
17. T. Fawcett, “ROC graphs: Notes and practical considerations for researchers,” *HP Lab.*, no. April, 2005, [Online]. Available: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:ROC+Graphs:+Notes+and+Practical+Considerations+for+Researchers#0>.
18. P. S. Yu, T. C. Yang, S. Y. Chen, C. M. Kuo, and H. W. Tseng, “Comparison of random forests and



- support vector machine for real-time radar-derived rainfall forecasting,” *J. Hydrol.*, vol. 552, pp. 92–104, 2017, doi: 10.1016/j.jhydrol.2017.06.020.
19. U. Shah, S. Garg, N. Sisodiya, N. Dube, and S. Sharma, “Rainfall prediction: Accuracy enhancement using machine learning and forecasting techniques,” *PDGC 2018 - 2018 5th Int. Conf. Parallel, Distrib. Grid Comput.*, pp. 776–782, 2018, doi: 10.1109/PDGC.2018.8745763.
  20. M. L. Felix and K. Jung, “Impacts of Spatial Interpolation Methods on Daily Streamflow Predictions with SWAT,” *Water (Switzerland)*, vol. 14, no. 20, 2022, doi: 10.3390/w14203340.
  21. A. G. El-Din and D. W. Smith, “A neural network model to predict the wastewater inflow incorporating rainfall events,” *Water Res.*, vol. 36, no. 5, pp. 1115–1126, 2002, doi: 10.1016/S0043-1354(01)00287-1.
  22. N. Samsiahsani, I. Shlash, M. Hassan, A. Hadi, and M. Aliff, “Enhancing Malaysia Rainfall Prediction Using Classification Techniques,” *J. Appl. Environ. Biol. Sci.*, vol. 7, no. April, pp. 20–29, 2017.
  23. A. Kusiak, X. Wei, A. P. Verma, and E. Roz, “Modeling and prediction of rainfall using radar reflectivity data: A data-mining approach,” *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 4, pp. 2337–2342, 2013, doi: 10.1109/TGRS.2012.2210429.
  24. S. Widaningsih, “Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm,” *J. Tekno Insentif*, vol. 13, no. 1, pp. 16–25, 2019, doi: 10.36787/jti.v13i1.78.
  25. A. Brotokuncoro, H. Fahmi, and W. Herry Utomo, “Implementasi Konsep Information Retrieval Dengan Metode Case Insentive Search Pada Mesin Pencari Dokumen Quality,” *Jhpp*, vol. 2, pp. 1–9, 2024.
  26. K. R. B and G. A., “Rainfall Prediction Using Data Mining Techniques - A Survey,” pp. 23–30, 2013, doi: 10.5121/csit.2013.3903.
  27. L. J. Anreaja, N. N. Harefa, J. G. P. Negara, V. N. H. Pribyantara, and A. B. Prasetyo, “Naive Bayes and Support Vector Machine Algorithm for Sentiment Analysis Opensea Mobile Application Users in Indonesia,” *JISA (Jurnal Inform. dan Sains)*, vol. 5, no. 1, pp. 62–68, 2022, doi: 10.31326/jisa.v5i1.1267.
  28. Badan Pusat Statistik, “Jumlah Kejadian Bencana Alam Menurut Provinsi, 2022,” 2023. <https://www.bps.go.id/id/statistics-table/3/TUZaMGVteFVjSEJ4T1RCMllyRjRTazVvVDJocVFUMDkjMw==/jumlah-kejadian-bencana-alam-menurut-provinsi--2022.html?year=2022>.
  29. A. A. Anandari, *Analisis Regresi Deret Fourier: Aplikasi Data Curah Hujan*. Sukabumi: Jejak Publisher, 2023.
  30. A. Sharma, *Environmental Chemistry*. Delhi: Khrisna Prakhasan Media, 2007.
  31. M. Bonell and L. A. Bruinjnzeel, *Forests, Water and People in the Humid Tropics: Past, Present, and Future Hydrological Research for Integrated Land and Water Management*. United Kingdom: Cambridge University Press, 2004.