

## PERBANDINGAN METODE *RANDOM FOREST* DAN *SUPPORT VECTOR MACHINE* PADA KLASIFIKASI CURAH HUJAN DI PROVINSI DAERAH ISTIMEWA YOGYAKARTA

Sugeng Fauzi Mahendra<sup>1\*</sup>, Yudi Setyawan<sup>2</sup>, Maria Titah Jatipaningrum<sup>3</sup>

<sup>1,2,3</sup>Jurusan Statistika, Fakultas Sains Terapan, Institut Sains & Teknologi AKPRIND Yogyakarta

Email : [sugeng.fmahendra@gmail.com](mailto:sugeng.fmahendra@gmail.com)

\*corresponding author

**Abstract.** *Rainfall is one of the climate components. According to the Meteorology, Climatology and Geophysics Agency, rainfall is the thickness of rainwater that collects in one millimeter area. There are several factors that affect rainfall including temperature, humidity, duration of irradiation and wind direction speed. Classification is a data mining technique that aims to accurately predict unknown categories for each data case. This study aims to classify rainfall in the Special Region of Yogyakarta using the Random Forest and Support Vector Machine methods and to compare the average accuracy with three times set.seed and three proportions of training data and data testing of each method. For the Random Forest method using set.seed (107), the proportion of training data and testing data is 70:30, the highest accuracy value is 71.8% with an estimated number of rainfall events for 79 days out of 110 days. For the Support Vector Machine method through set.seed (107), the proportion of training data and testing data is 60:40, the highest accuracy value is 93.83% with an estimated rainfall of 137 days out of 146 days. Based on the accuracy calculated through three set.seed times and three proportions, the support vector machine method has a higher accuracy value with an accuracy value of 90.25% than the random forest method with an accuracy value of 70.5%.*

**Keywords:** *Rainfall, Classification, Random Forest, Support Vector Machine*

**Abstrak.** Curah hujan merupakan salah satu dari komponen iklim. Menurut Badan Meteorologi, Klimatologi dan Geofisika, curah hujan merupakan ketebalan air hujan yang terkumpul dalam satu luasan milimeter. Terdapat beberapa faktor yang mempengaruhi curah hujan diantaranya suhu, kelembapan, lama penyinaran dan kecepatan arah angin. Klasifikasi adalah teknik data mining yang bertujuan untuk memprediksi kategori yang tidak diketahui untuk setiap kasus data secara akurat. Penelitian ini bertujuan untuk melakukan klasifikasi curah hujan di Daerah Istimewa Yogyakarta menggunakan metode *Random Forest* dan *Support Vector Machine* dan melakukan perbandingan rerata keakuratan dengan tiga kali set.seed dan tiga proporsi data training dan data testing dari masing-masing metode. Untuk metode *Random Forest* melalui set.seed (107), proporsi data training dan data testing 70:30 diperoleh nilai akurasi terbesar 71,8% dengan prakiraan angka kejadian curah hujan selama 79 hari dari 110 hari. Untuk metode *Support Vector Machine* melalui set.seed (107), proporsi data training dan data testing 60:40 diperoleh nilai akurasi terbesar 93,83% dengan prakiraan angka kejadian curah hujan 137 hari dari 146 hari. Berdasarkan keakuratan yang dihitung melalui tiga kali set.seed dan tiga proporsi, metode *support vector machine* memiliki nilai akurasi yang lebih tinggi dengan nilai akurasi 90,25% dari metode *random forest* dengan nilai akurasi 70,5%.

**Kata Kunci:** Curah Hujan, Klasifikasi, *Random Forest*, *Support Vector Machine*

### 1. Pendahuluan

Curah hujan merupakan salah satu komponen dalam iklim dan hal yang dapat berpengaruh pada cuaca. Menurut Badan Meteorologi, Klimatologi dan Geofisika atau disingkat BMKG menyebutkan bahwa curah hujan adalah ketebalan air hujan yang terkumpul dalam luasan 1 milimeter. Curah hujan 1 milimeter artinya tempat yang datar dengan luasan satu meter persegi dapat tertampung air setinggi satu milimeter atau

tertampung air sebanyak satu liter (Prawaka, 2016). Terdapat beberapa faktor yang mempengaruhi curah hujan diantaranya suhu, kelembapan, lama penyinaran, dan arah angin (BMKG). Badan Meteorologi, Klimatologi dan Geofisika sendiri adalah suatu Lembaga Pemerintah Nonkementerian Indonesia (LPNK) yang mempunyai tugas melaksanakan tugas pemerintah di bidang meteorologi, klimatologi, kualitas udara dan geofisika.

Klasifikasi adalah metode untuk memprediksi suatu kejadian atau keputusan yang akan datang berada di suatu titik. Dalam statistika ilmu data menyediakan banyak sekali algoritma klasifikasi seperti *logistic regression*, *support vector machine*, pengklasifikasi *Naive Bayes*, dan pohon keputusan atau *decision tree* salah satunya yaitu *random forest*.

Penelitian ini menerapkan metode *Random Forest* dan *Support Vector Machine* (SVM) yang diharapkan mampu mengetahui bagaimana klasifikasi pada curah hujan di Provinsi DIY dengan masing-masing metode dan bagaimana perbandingan keakuratan kedua metode yang digunakan dalam melakukan klasifikasi. Data sampel diperoleh dari Badan Meteorologi, Klimatologi dan Geofisika Provinsi Daerah Istimewa Yogyakarta Tahun 2021.

Adapun penelitian sebelumnya yang berkaitan adalah oleh (Nurkhaliza & Wijayanto, 2022) dengan judul “Perbandingan Algoritma Klasifikasi *Support Vector Machine* dan *Random Forest* pada Prediksi Status Indeks Mitigasi dan Kesiapan Bencana (IMKB) Satuan Kerja BPS di Indonesia Tahun 2020”. Kemudian dilakukan oleh (Pratama, Latipah, & Sari, 2022) yang berjudul “Optimasi Klasifikasi Curah Hujan Menggunakan *Support Vector Machine (SVM)* dan *Recursive Feature Elimination (RFE)*”. Selanjutnya penelitian dengan metode sejenis juga dilakukan oleh (Zeniarta, Salam, & Ma'ruf, 2022) yang melakukan penelitian dengan judul “Seleksi Fitur dan Perbandingan Algoritma Klasifikasi untuk Prediksi Kelulusan Mahasiswa”. Dan dilakukan oleh (Mauludiyah, 2020) yang berjudul “Klasifikasi Indeks Pembangunan Manusia Kabupaten/Kota Di Indonesia menggunakan Metode *Random Forest*”.

## 2. Metode Penelitian

### a. Sumber Data

Data yang digunakan dalam penelitian adalah data sekunder yang diperoleh dari Badan Meteorologi, Klimatologi dan Geofisika Provinsi Daerah Istimewa Yogyakarta, data semua variabel dapat diakses melalui website <https://dataonline.bmkg.go.id/home>.

### b. Variabel Penelitian

Adapun variabel yang digunakan dalam penelitian terdiri dari variabel dependen dan independen.

1. Variabel dependen ( $Y$ ) dalam penelitian ini yaitu Curah Hujan dengan kategori Sangat Ringan, Ringan, Sedang, Lebat dan Sangat Lebat.
2. Variabel independent ( $X$ ) yang digunakan sebanyak empat variabel yaitu Suhu, Kelembapan, Lama Penyinaran Matahari dan Kecepatan Arah Angin.

### c. Tahapan Analisis Data

Berikut merupakan tahapan analisis data dengan menggunakan metode *Random Forest* dan *Support Vector Machine*:

- 1) Mengumpulkan data sekunder.
- 2) Melakukan analisis statistika deskriptif.

- 3) Melakukan tiga kali pengacakan pada *data training* dan *data testing* dengan tiga proporsi 60:40, 70:30 dan 80:20.
- 4) Melakukan analisis klasifikasi *Random Forest*
  - a) Input data
  - b) Menentukan parameter *mtry* ( $m$ ) dan banyak pohon atau *ntree* ( $k$ )
  - c) Melakukan pengacakan atau *set.seed* sebanyak tiga kali
  - d) Pembuatan suatu *bootstrap sample* dari gugus *data training* dengan pengembalian melalui tiga kali pengacakan.
  - e) Membangun pohon klasifikasi menggunakan gugus data training baru yang terbentuk dari proses *bootstrap*.
  - f) Mengulangi langkah (d) dan (e) sebanyak  $k$  kali sehingga diperoleh  $k$  buah pohon acak. Setiap pohon klasifikasi akan menghasilkan satu keputusan sehingga didapatkan  $k$  buah keputusan. Penentuan klasifikasi didasarkan pada keputusan terbanyak (*majority vote*).
  - g) Menentukan nilai rerata akurasi tertinggi dan penentuan ketepatan klasifikasi *Random Forest* terbaik berdasarkan nilai akurasi tertinggi.
  - h) Menghitung akurasi tertinggi berdasarkan confusion matrix
- 5) Melakukan analisis klasifikasi *Support Vector Machine* pada kernel RBF
  - a) Input data
  - b) Menentukan parameter *cost* dan *gamma*, dengan melakukan beberapa percobaan sehingga nilai *cost* sebesar 0.1, 1, 10, 100 dan *gamma* sebesar 1
  - c) Melakukan pengacakan atau *set.seed* sebanyak tiga kali
  - d) Menentukan nilai rerata akurasi tertinggi dan penentuan ketepatan klasifikasi SVM terbaik berdasarkan nilai akurasi tertinggi.
  - e) Melakukan perhitungan dengan membangun model atau membentuk model menggunakan fungsi  $f(x)$  yaitu RBF dengan persamaan (14)
  - f) Penentuan hasil ketepatan klasifikasi SVM terbaik berdasarkan nilai akurasi tertinggi
  - g) Menghitung akurasi tertinggi berdasarkan confusion matrix
- 6) Perbandingan hasil terbaik dengan melihat nilai validitas rerata akurasi
- 7) Kesimpulan.

#### d. Metode Analisis Data

##### 1). Random Forest

Metode *random forest* adalah pengembangan dari metode *CART*, yaitu dengan menerapkan metode *bootstrap selection aggregating (bagging)* dan *random feature selection*. Berdasarkan penelitian yang dilakukan *Random Forest* memiliki beberapa kelebihan antara lain dapat menghasilkan eror yang lebih rendah, memberikan hasil yang bagus dalam klasifikasi, dapat mengatasi data training dalam jumlah sangat besar secara efisien dan metode yang efektif untuk mengestimasi *missing* data (Breiman, 2001). Metode ini bertujuan untuk membangun pohon keputusan yang terdiri dari *root node*, *internal node*, dan *leaf node* dengan mengambil atribut dan data secara acak sesuai ketentuan.

*Random Forest* menggunakan *information gain* untuk perhitungan dalam membangun pohon (Han, Kamber, & Pei, 2012). *Information gain* adalah perubahan *entropy* setelah dilakukan pembagian berdasarkan sebuah atribut ke dalam subset yang

lebih kecil. Secara matematis untuk menghitung nilai *information gain* menggunakan persamaan di bawah ini:

$$Gain(A) = Entropy(S) - Entropy_A(S)$$

Kemudian untuk menghitung nilai *Entropy* ( $S$ ) dan  $Entropy_A(S)$  dapat menggunakan persamaan sebagai berikut:

$$Entropy(S) = -\sum_{i=1}^n p_i \log_2 p_i \quad (1)$$

$$Entropy(S_i) = \sum_{i=1}^n \frac{A_i}{c} \quad (2)$$

$$Entropy_A(S) = \sum_{i=1}^n \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (3)$$

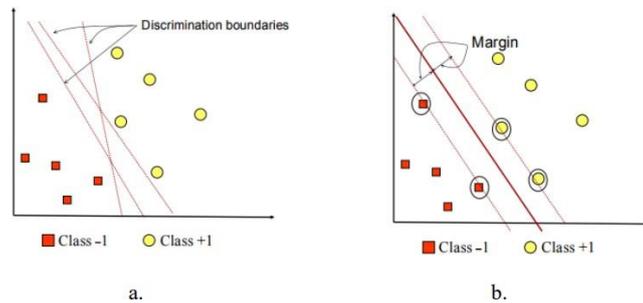
## 2) Support Vector Machine

*Support Vector Machine* adalah salah satu metode machine learning yang bekerja atas prinsip *Structural Risk Minimization* (SRM) yang bertujuan untuk menemukan hyperplane terbaik untuk memisahkan dua kelas data. *Support Vector Machine* bekerja dengan memaksimalkan margin yang merupakan jarak pemisah antara kedua kelas data tersebut (Pratiwi, 2017).

Pada dasarnya SVM memiliki prinsip *linier*, akan tetapi kini SVM telah berkembang sehingga dapat bekerja pada masalah *non-linier*. Cara kerja SVM pada masalah non-linier adalah dengan memasukkan konsep kernel pada ruang berdimensi tinggi. Pada ruang yang berdimensi tinggi ini, nantinya akan dicari pemisah atau yang sering disebut hyperplane.

### a). Linear Separable

Metode SVM dengan hyperplane yang berbentuk garis disebut dengan *Linear separable*. Terlihat pada Gambar 1



Gambar 1. Garis Linier Pemisah Dua Kelas

Pada Gambar 1 menunjukkan dua kelas dapat dipisahkan oleh sepasang bidang pembatas yang sejajar. Upaya dalam mencari lokasi *hyperplane* ini merupakan proses inti dari SVM. sehingga diperoleh persamaan klasifikasi linier SVM sebagai berikut:

$$f(x) = w \cdot x + b \quad (4)$$

Dan persamaan hyperplane yang membatasi kedua kelas sebagai berikut:

$$H_1 : w \cdot x_i + b \geq 1 \text{ untuk } y_i = +1 \quad (5)$$

$$H_2 : w \cdot x_i + b \leq -1 \text{ untuk } y_i = -1 \quad (6)$$

Penggabungan dari persamaan (5) dan (6) menghasilkan pertidaksamaan:

$$y_i(w \cdot x_i + b) \geq 1, \text{ untuk } \forall i = 1, 2, \dots, k \quad (7)$$

Margin antara dua kelas dapat dihitung dengan mencari jarak antara kedua hyperplane  $H_1$  atau  $H_2$ . Nilai margin terbesar dapat ditemukan dengan memaksimalkan nilai jarak antara  $H_1$  dan  $H_2$  dengan titik terdekatnya dihitung menggunakan rumus:

$$\text{Margin: } (H_1 - H_2) \cdot \frac{w}{\|w\|} = \left( \frac{(1-b)}{w} - \frac{(-b-1)}{w} \right) \cdot \frac{w}{\|w\|} = \frac{2}{\|w\|} \quad (8)$$

Oleh karena itu memaksimalkan  $\frac{2}{\|w\|}$  sama dengan meminimumkan  $\|w\|^2$  dan untuk menyederhanakan penyelesaian ditambahkan faktor  $\frac{1}{2}$  (Aisyah & Fithriasari, 2016). Dengan demikian, model persamaannya menjadi:

$$\min \frac{1}{2} \|w\|^2 \text{ atau } -\frac{1}{2} w^T w \quad (9)$$

dengan syarat  $y_i(w \cdot x_i + b) \geq 1, i = 1, 2, 3, \dots, k$  dan  $\|w\|$  merupakan vektor normal.

Dalam klasifikasi terkadang dapat dijumpai bidang pemisah yang tidak bisa diambil dengan linier sehingga diperlukan penyelesaian khusus untuk permasalahan ini. Untuk data-data yang tidak dapat dipisahkan secara linier tersebut ditambahkan variabel *slack*  $\xi_i \geq 0$  ke pertidaksamaan sehingga kendala dan fungsi tujuan menjadi:

$$y_i(w \cdot x_i + b) - 1 + \xi_i \geq 0, \text{ untuk } \forall i = 1, 2, \dots, k$$

$$\min \frac{1}{2} \|w\|^2 + C \sum_{i=1}^k \xi_i \quad (10)$$

dengan  $y_i(w \cdot x_i + b) - 1 + \xi_i \geq 0, \xi_i \geq 0, \text{ untuk } \forall i = 1, 2, \dots, k$

Hal ini dilakukan untuk mengurangi jumlah kesalahan klasifikasi (misclassification error) yang dinyatakan dengan adanya variabel *slack*  $\xi_i$ . Penggunaan variabel *slack* ini adalah untuk mengatasi kasus ketidaklayakan dengan cara memberi pinalti.  $C$  adalah parameter yang menentukan besar pinalti akibat kesalahan dalam klasifikasi data dan nilainya ditentukan oleh pengguna (Aisyah & Fithriasari, 2016). sehingga peran dari  $C$  adalah meminimalkan kesalahan pelatihan. Selanjutnya, bentuk persamaan akan menjadi:

$$\min L_p = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^k \xi_i - \sum_{i=1}^k a_i \{y_i(w \cdot x_i + b) - 1 + \xi_i\} - \sum_{i=1}^k a_i \xi_i \quad (11)$$

Dengan dilakukan turunan persamaan dual problem pada data linier untuk data non-linier, maka diperoleh persamaan lagrangian menjadi:

$$L_D = \sum_{i=1}^k a_i - \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k a_i a_j y_i y_j (x_i x_j^T) \quad (12)$$

Dengan kendala  $0 \leq a_i \leq C, \text{ untuk } i = 1, 2, \dots, k$

#### b). Non-Linear Separable

Pada Support Vector Machine terdapat sebuah fungsi kernel, yaitu fungsi yang digunakan untuk menyelesaikan masalah non linier. Kernel berfungsi memungkinkan untuk mengimplementasikan suatu model pada ruang dimensi lebih tinggi (ruang fitur). fungsi kernel yang akan digunakan dari fungsi non-linier  $\varphi$  (Nugroho, Satrio, & Witarto, 2003). Selanjutnya fungsi keputusan klasifikasi dari data  $x$  sign(f(x)):

$$f(x) = \text{sign} \left( \sum_{i=1}^k a_i y_i K(x_i, x_j^T) + b \right) \quad (13)$$

Berikut fungsi kernel yang digunakan pada penelitian ini adalah Kernel Radial Basis Function (RBF)

$$K(\vec{x}_i, \vec{x}_j) = \exp\left(-\frac{\|\vec{x}_i - \vec{x}_j\|^2}{2\sigma^2}\right) \text{ atau } \exp\left(-\gamma\|x_i - x_j\|^2\right) \quad (14)$$

### 3). Confusion Matrix

Confusion Matrix adalah pengukuran performa untuk masalah klasifikasi *machine learning* dan memiliki tabel dengan 4 kombinasi berbeda dari nilai prediksi dan nilai actual. Berikut merupakan hasil dari confusion matrix (Sasongko, 2016).

Tabel 1. Confusion Matrix

Aktual	Prediksi	
	Positive	Negative
Positive	TP	FN
Negative	FP	TN

Keterangan:

TP = Jumlah prediksi yang tepat bersifat positif (True Positive)

TN = Jumlah prediksi yang tepat bersifat negatif (True Negative)

FP = Jumlah prediksi yang tepat bersifat positif (False Positive)

FN = Jumlah prediksi yang tepat bersifat negatif (False Negative)

Beberapa persyaratan yang telah didefinisikan untuk matrik klasifikasi diantaranya Accuracy, Recall atau Sensitivity, Specificity, Tingkat False Positive, Tingkat False Negative dan Presisi.

## 3. Hasil dan Pembahasan

### 1. Analisis Deskriptif

Analisis deskriptif merupakan analisis yang digunakan untuk mengetahui secara keseluruhan gambaran curah hujan di Provinsi Daerah Istimewa Yogyakarta. Berdasarkan analisis deskriptif dapat diketahui bahwa curah hujan harian, suhu, kelembapan, lama penyinaran matahari dan kecepatan arah angin pada tahun 2021 terjadi fluktuasi yang artinya curah hujan setiap harinya tidak sama atau berbeda-beda.

Berdasarkan data penelitian dilakukan analisis statistika deskriptif dari curah hujan di Provinsi Daerah Istimewa Yogyakarta Tahun 2021. Pada Tabel 2 ditampilkan hasil dari analisis statistika deskriptif.

Tabel 2 Karakteristik Variabel

Keterangan	Rata-rata	Nilai Max	Nilai Min	Standar Deviasi
Suhu	26,362	28,3	23,9	0,783
Kelembapan	82,413	96	64	4,854
Lama Penyinaran	5,332	10,9	0	2,743
Kecepatan Angin	228,991	360	10	63,535
Curah Hujan	31,432	5925,3	0	346,59

Dapat dilihat dari Tabel 2 diatas yang menjelaskan statistik deskriptif 4 variabel independen yang digunakan untuk menggambarkan variabel dependen yaitu curah hujan di Provinsi Daerah Istimewa Yogyakarta.

## 2. Data Training dan Data Testing

Data kategori curah hujan di Provinsi Daerah Istimewa Yogyakarta yang berjumlah 365 hari data dibagi ke dalam *data training* dan *data testing*. Pembagian dilakukan dengan persentase yaitu 60%, 70% dan 80% pada *data training* dan 40%, 30% dan 20% pada *data testing*.

## 3. Hasil Klasifikasi Random Forest

Pembangunan pohon klasifikasi ini menerapkan *random feature selection* yaitu pemilihan  $m$  (*mtry*) variabel secara acak dari  $p$  (variabel input). demikian diperoleh percobaan untuk  $m$  yang akan digunakan adalah  $m_1 = 2$ ,  $m_2 = 1$  dan  $m_3 = 4$  dengan ntree sebanyak 5 yaitu 25, 50, 100, 500 dan 1000.

Berdasarkan perhitungan nilai rerata akurasi-akurasi tertinggi dengan tiga kali pengacakan. Rerata akurasi tertinggi terdapat pada proporsi 60:40 yaitu 0,705 atau 70,5%. Kemudian nilai akurasi tertinggi dihasilkan oleh model dengan  $m = 1$  dan  $k = 500$  yaitu sebesar 0,718. Sehingga pembentukan pohon dilakukan dengan proporsi 70:30 set.seed (107)  $m = 1$  dan  $k = 500$ . Berikut hasil klasifikasi menggunakan model yang terbaik

Tabel 3. Hasil Klasifikasi data testing set.seed (107)  $m = 1$  dan  $k = 500$

No	Suhu	Kelembapan	Lama Penyinaran	Kecepatan Arah Angin	Y	pred RF
1	26,7	84	5,4	280	2	0
2	27,3	82	5	190	1	0
3	26,8	80	2,5	230	1	1
4	25,3	90	6,4	200	1	0
5	26,2	89	2,5	220	1	1
6	26,9	80	5,3	250	3	0
...	...	...	...	...	...	...
...	...	...	...	...	...	...
...	...	...	...	...	...	...
105	26,4	83	10,1	200	1	0
106	27	84	2,9	250	0	0
107	25,9	91	6,2	270	1	1
108	26,4	87	2,6	260	2	1
109	26,7	84	1,2	180	0	0
110	27,4	77	6,6	270	0	0

Berikut pohon keputusan klasifikasi metode random forest menggunakan model  $m = 1$  dan  $k = 500$  pada proporsi data 70:30.



Berdasarkan perhitungan percobaan yang dilakukan diperoleh nilai-nilai akurasi setiap proporsi dan tiga kali pengulangan pada metode *Support Vector Machine*. Pembentukan model klasifikasi akan dilakukan dengan proporsi 60:40 set.seed (107) Cost = 100 dan Gamma = 1 yang memiliki jumlah support vector terendah.

Tabel 5 Nilai Alpha Kernel RBF C = 100

No	Alpha	Data ke..
1	5,02940363	2
2	100,0000000	3
3	37,12470190	4
....	....	....
51	-100,0000000	137
52	3,40027005	139
53	57,94243625	141
54	-99,84610827	143

Dengan memanfaatkan nilai alpha dan bias. Sehingga didapat persamaan model optimal hyperplane yang digunakan berdasarkan fungsi svm kernel RBF, sebagai berikut:

Tabel 6. Model SVM Kernel RBF

No	$f(x) = \text{sign} \sum_{i=1}^k a_i y_i \exp(-\gamma \ x_i - x_j\ ^2) + b$
1	0,999734823
2	-1,117736406
3	0,999902150
....	....
52	0,99992829
53	1,00026508
54	-1,0000851

Selanjutnya perbandingan hasil klasifikasi dan data actual pada proporsi *data testing* 60:40 set.seed (107) dengan nilai Gamma sebesar 1 dan nilai Cost sebesar 100.

Tabel 7. Hasil Klasifikasi data testing set.seed (107) Cost = 100 dan Gamma =1

No	Suhu	Kelembapan	Lama Penyinaran	Kecepatan Arah Angin	Y	pred_SVM
1	26,7	84	5,4	280	2	0
2	27,3	82	5	190	1	1
3	26,8	80	2,5	230	1	0
4	25,3	90	6,4	200	1	1
5	26,2	89	2,5	220	1	1
...	....	....	....	....	....	....
...	....	....	....	....	....	....
...	....	....	....	....	....	....
No	Suhu	Kelembapan	Lama Penyinaran	Kecepatan Arah Angin	Y	pred_SVM
142	26,4	87	2,6	260	2	2
143	26,7	84	1,2	180	0	0
144	27,2	82	4,1	210	3	3

145	27,4	77	6,6	270	0	0
146	25,9	86	7,5	210	2	2

Selanjutnya menampilkan hasil klasifikasi kategori *data testing* menggunakan metode *Support Vector Machine*.

Tabel 8. Klasifikasi Kategori Support Vector Machine

Data Testing	Prediksi				
	Sangat Ringan	Ringan	Sedang	Lebat	Sangat Lebat
Sangat Ringan	95	3	2	1	0
Ringan	1	28	1	0	0
Sedang	0	1	10	0	0
Lebat	0	0	0	3	0
Sangat Lebat	0	0	0	0	1

Berdasarkan tabel di atas, maka dapat dihitung nilai akurasi melalui hasil prediksi yang memiliki kinerja benar. Perhitungan sebagai berikut:

$$Akurasi = \frac{95+28+10+3+1}{146} = \frac{137}{146} = 0,9383$$

$$Akurasi = 0,9383 \approx 93,83\%$$

Nilai akurasi klasifikasi menggunakan *Support Vector Machine* adalah sebesar 0,9383 atau 93,83%. Nilai akurasi 0,90 – 1 dikelompokkan ke dalam klasifikasi kategori sangat baik.

#### 5. Klasifikasi Metode Terbaik

Berdasarkan perhitungan rerata nilai akurasi-akurasi tertinggi yang telah dilakukan pada kedua metode dengan tiga kali pengacakan dan tiga proporsi, perbandingan tingkat akurasi yang diperoleh sebagai berikut:

Tabel 9 Akurasi Klasifikasi Terbaik

Metode	Random Forest	Support Vector Machine
Tingkat Akurasi	70,5%	90,25%

Berdasarkan Tabel 9. Dapat dilihat bahwa metode yang memiliki nilai akurasi terbesar yaitu metode *Support Vector Machine* dengan nilai akurasi 0,9383 atau 93,83%. Maka metode *Support Vector Machine* menjadi metode dengan nilai akurasi klasifikasi terbaik.

## 4. Kesimpulan

Berdasarkan analisis dan pembahasan yang telah dilakukan pada penelitian ini, dapat ditarik beberapa kesimpulan sebagai berikut:

- Pada statistika deskriptif diperoleh bahwa untuk masing-masing variabel terjadi fluktuasi yang artinya setiap harinya curah hujan, Suhu, kelembapan udara, lama penyinaran matahari dan kecepatan angin tidak sama atau berbeda-beda. Kemudian pada tahun 2021 kondisi jumlah curah hujan berdasarkan kategori untuk data dengan kategori sangat ringan sebanyak 248 hari, data dengan kategori ringan

sebanyak 76 hari, data dengan kategori sedang sebanyak 32 hari, data dengan kategori lebat sebanyak 7 hari dan data dengan kategori sangat lebat sebanyak 2 hari.

- b) Prakiraan kondisi curah hujan menggunakan metode *Random Forest* melalui tiga kali pengacakan set.seed 107, 284 dan 775 dengan tiga proporsi 60:40, 70:30 dan 80:20 menghasilkan tingkat akurasi yang tertinggi pada proporsi 70:30 set.seed (107) yakni 71,8% dengan klasifikasi kategori menunjukkan prakiraan angka kejadian curah hujan selama 79 hari dari 110 hari.
- c) Prakiraan kondisi curah hujan menggunakan metode *Support Vector Machine* melalui tiga kali pengacakan set.seed 107, 284 dan 775 dengan tiga proporsi 60:40, 70:30 dan 80:20 menghasilkan tingkat akurasi yang tertinggi pada proporsi 60:40 set.seed (107) yakni 98,83% dengan klasifikasi menunjukkan prakiraan angka kejadian curah hujan selama 137 hari dari 146 hari.
- d) Berdasarkan rerata nilai akurasi-akurasi tertinggi yang dihasilkan dari kedua metode melalui tiga kali pengacakan diperoleh bahwa metode *Random Forest* mencapai rerata nilai akurasi sebesar 0,705 atau 70,5% dan *Support Vector Machine* mencapai rerata nilai akurasi sebesar 90,25 sehingga metode yang terbaik yaitu metode *Support Vector Machine*, hal ini dibuktikan dengan rerata nilai akurasi dicapai sebesar 90,25% yang dapat dikategorikan sangat baik.

#### Daftar Pustaka

- Breiman, L. (2001). *Random Forest*. *Machine Learning*, 5-32.
- Cortes, C., & Vapnik, V. (1995). *Support-Vector Networks Machine Learning*. USA.
- Mauludiyah, K. (2020). Klasifikasi Indeks Pembangunan Manusia Kabupaten/Kota di Indonesia Menggunakan Metode *Random Forest*. *Jurnal Ilmiah*, 1-8.
- Nurkhaliza, A. A., & Wijayanto, A. W. (2022). Perbandingan Algoritma Klasifikasi *Support Vector Machine* dan *Random Forest* pada Prediksi Status Indeks Mitigasi dan Kesigapan Bencana (IMKB) Satuan Kerja BPS di Indonesia Tahun 2020. *Jurnal Informatika Universitas Pamulang*, 54-59.
- Pratama, A. R., Latipah, S. A., & Sari, B. N. (2022). Optimasi Klasifikasi Curah Hujan Menggunakan *Support Vector Machine (SVM)* dan *Recursive Feature Elimination (RFE)*. *Jurnal Ilmiah Penelitian dan Pembelajaran Informatika*, 314-324.
- Setyawan, Y., Noeryanti, & Suryowati, K. (2018). *Statistika Dasar Dilengkapi dengan Software R*. Yogyakarta: Akprind Press.
- Sumanthi, K., Kannan, S., & Nagarajan, K. (2016). Data Mining: Analysis of student database using Classification Techniques. *International Journal of Computer Applications*, 22-27.
- Zeniarja, J., Salam, A., & Ma'ruf, F. A. (2022). Seleksi Fitur dan Perbandingan Algoritma Klasifikasi untuk Prediksi Kelulusan Mahasiswa. *Jurnal Rekayasa Elektrika*, Vol. 18, No.2, Hal 102-108