

# Perbandingan Algoritma SVM dan *Naïve Bayes* untuk Klasifikasi Kebakaran Hutan dan Lahan

Tiara Aurilia Viona, Putri Yuli Utami\*, Izhan Fakhruzi, Rifqi Anugrah  
Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Muhammadiyah  
Pontianak, Indonesia

\*E-mail koresponden: [putriyuli@unmuhpnk.ac.id](mailto:putriyuli@unmuhpnk.ac.id)

Diserahkan 23 Januari 2024; Direview 29 Oktober 2024; Dipublikasikan 30 November 2024

## Abstrak

*Kabupaten Ketapang merupakan satu diantara kabupaten di Provinsi Kalimantan Barat yang sering terjadi bencana kebakaran hutan dan lahan (karhutla). Tujuan dilakukannya penelitian ini untuk memprediksi terjadinya bencana karhutla di Kabupaten Ketapang agar penanganan karhutla dapat lebih optimal. Berdasarkan data dari Badan Penanggulangan Bencana Daerah (BPBD) Provinsi Kalimantan Barat total kejadian bencana karhutla di Kabupaten Ketapang mulai dari tahun 2015 sampai tahun 2022 sudah terjadi sebanyak 235 kali. Tujuan penelitian ini untuk memprediksi terjadinya kebakaran hutan dan lahan serta melakukan klasifikasi daerah yang masuk kategori rawan dan sangat rawan di Ketapang. Pada penelitian ini klasifikasi dilakukan dengan membandingkan metode SVM dan Naïve Bayes untuk mendapatkan akurasi yang optimal. Dataset yang digunakan adalah data hotspot dan data iklim Kabupaten Ketapang tahun 2021 – 2022 yang berjumlah 4.697 data. Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes mencapai akurasi 99,67%, sedangkan SVM memiliki akurasi 98,24%, mengindikasikan bahwa Naïve Bayes lebih unggul dalam klasifikasi daerah rawan karhutla. Penelitian ini dapat diperluas dengan mempertimbangkan faktor tambahan seperti vegetasi, topografi, dan pengaruh aktivitas manusia untuk meningkatkan keakuratan prediksi serta pengembangan sistem peringatan dini berbasis kecerdasan buatan.*

**Kata kunci:** CRISP-DM, Data Mining, Klasifikasi, Naïve Bayes, SVM

## Abstract

*Ketapang Regency is one of the regencies in West Kalimantan Province that is prone to forest and land fires (karhutla). The purpose of this study is to predict the occurrence of karhutla in Ketapang Regency so that karhutla management can be more optimal. Based on data from the Regional Disaster Management Agency (BPBD) of West Kalimantan Province, there have been a total of 235 karhutla disasters in Ketapang Regency from 2015 to 2022. The purpose of this study is to predict the occurrence of forest and land fires and to classify areas in Ketapang that are categorized as prone and highly prone. In this study, classification was carried out by comparing the SVM and Naïve Bayes methods to obtain optimal accuracy. The dataset used was hotspot and climate data for Ketapang Regency from 2021 to 2022, totaling 4,697 data points. The results showed that the Naïve Bayes*

algorithm achieved an accuracy of 99.67%. In comparison, SVM had an accuracy of 98.24%, indicating that Naïve Bayes was superior in classifying areas prone to forest and land fires. This research can be expanded by considering additional factors such as vegetation, topography, and the influence of human activities to improve prediction accuracy and develop an artificial intelligence-based early warning system.

**Keywords:** Classification, CRISP-DM, Data Mining, Naïve Bayes, SVM

## PENDAHULUAN

Kabupaten Ketapang terletak di Provinsi Kalimantan Barat dengan titik koordinat  $0^{\circ} 19' 26,51''$  Lintang Selatan sampai  $3^{\circ} 4' 16,59''$  Lintang Selatan dan  $109^{\circ} 47' 36,55''$  Bujur Timur sampai  $111^{\circ} 21' 37,36''$  Bujur Timur. Kabupaten Ketapang merupakan satu diantara kabupaten yang sering terjadi bencana kebakaran hutan dan lahan (karhutla). Berdasarkan data dari Badan Penanggulangan Bencana Daerah (BPBD) Provinsi Kalimantan Barat total kejadian bencana karhutla di Kabupaten Ketapang mulai dari tahun 2015 sampai tahun 2022 sudah terjadi sebanyak 235 kali. Bencana karhutla terjadi disebabkan oleh beberapa faktor misalnya faktor alam seperti musim kemarau panjang sehingga jumlah titik panas (*hotspot*) meningkat. Namun dapat juga disebabkan oleh faktor kegiatan manusia seperti pembakaran lahan secara sengaja dan sembarangan dalam membuang puntung rokok [1].

Dampak yang ditimbulkan akibat dari kebakaran hutan seperti penurunan jumlah keanekaragaman hayati, satwa liar yang kehilangan habitat hidup aslinya [2]. Dilihat dari segi kesehatan, kebakaran hutan dapat mengganggu kesehatan makhluk hidup yang disebabkan oleh asap hasil dari kebakaran tersebut. Asap tersebut juga dapat mengganggu penglihatan makhluk hidup karena dapat membuat jarak pandang penglihatan menurun [3].

Tindakan untuk meminimalkan dampak yang merugikan tersebut dibutuhkan suatu pencegahan dengan melakukan prediksi terjadinya kebakaran hutan dan lahan salah satu metode yang dapat dimanfaatkan adalah *data mining* yaitu metode klasifikasi. Klasifikasi merupakan metode yang dapat melakukan identifikasi objek menjadi sebuah kelas, kelompok atau kategori berlandaskan dari definisi, prosedur dan karakteristik yang sudah ditetapkan [4]. Klasifikasi dalam penelitian ini digunakan untuk mengidentifikasi daerah rawan dan sangat rawan, salah satu metode untuk melakukan klasifikasi diantaranya *Support Vector Machine* [5] dan *Naïve Bayes* [6]. Adapun atribut yang digunakan yaitu kecamatan, tanggal (WIB), tingkat kepercayaan, radius kemungkinan, curah hujan, kelembapan dan suhu.

*Support Vector Machine (SVM)* adalah metode mengenali pola dan melakukan prediksi dari masukkan atau pembelajaran yang diberikan, metode ini termasuk *supervised learning* karena membutuhkan pembelajaran sehingga *SVM* dapat menemukan pembatas antar kelas dengan memaksimalkan margin atau garis batas tersebut[7]. Kelebihan menggunakan metode *SVM* adalah proses komputasi yang cepat karena dalam menentukan jarak antar kelas digunakan *support vector*[8]. Penelitian yang telah memanfaatkan *SVM* menghasilkan tiga kategori *data hotspot* yaitu 2.214 titik termasuk resiko rendah, 15.412 titik termasuk resiko sedang dan 4.479 titik termasuk resiko tinggi [9]. Hasil dari penelitian ini adalah nilai kebenaran 91,475% evaluasi yang digunakan yaitu *Polynomial Kernel* dan nilai 93,667% pada *Confusion Matrix* sebagai validasi. Selanjutnya penelitian [10] menggunakan metode *SVM* dengan akurasi *SVM* yang terbaik yaitu 99.8% dan di daerah Kumpeh Ulu Muaro Jambi pada bulan agustus tahun 2019 diidentifikasi mengalami kebakaran daerah gambut

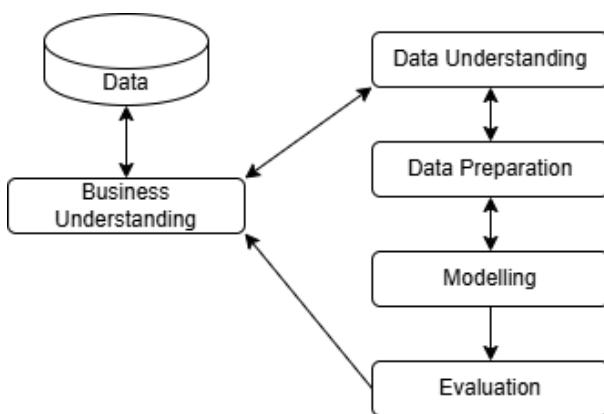
seluas 1.396,89 ha, seluas 1.089,54 ha teridentifikasi merupakan lahan terbuka dan 7.069,5 ha merupakan wilayah vegetasi. Model SVM dalam klasifikasi ketersediaan air tanah dengan nilai *F1-score* hijau 89.58%, setengah hijau 65.62%, dan kering 83.44% untuk *kernel linear*. Sementara itu untuk *kernel polynomial* diperoleh *F1 score* hijau 83.58%, setengah hijau 25.68%, dan kering 66.59% [11].

*Naïve bayes* merupakan algoritma yang dikembangkan oleh seorang ilmuwan *inggris* yaitu Thomas Bayes. Metode *Naïve Bayes* digunakan untuk memperkirakan peluang di masa yang akan datang berlandaskan peristiwa di masa lampau sehingga disebut dengan teorema *Bayes* yang kemudian dihubungkan dengan “*naïve*” yang memiliki arti situasi diantara atribut yang ada saling bebas [12]. Kelebihan menggunakan *Naïve Bayes* adalah metode tersebut menggunakan sedikit data training untuk menghitung parameter yang diperlukan dalam melakukan klasifikasi [13]. Penelitian yang telah memanfaatkan *Naïve Bayes* adalah [14] menerapkan metode *Naïve Bayes* sehingga diperoleh nilai akurasi tertinggi pada tahun 2019 yaitu sebesar 99.52% sehingga disimpulkan bahwa metode *Naïve Bayes* efektif digunakan untuk klasifikasi lokasi kebakaran hutan dan lahan. Penelitian lainnya yang dilakukan ialah membandingkan kedua metode antar *Naïve Bayes* dan SVM yang memiliki akurasi yang baik dalam menentukan klasifikasi [16].

Penelitian ini melakukan komparasi akurasi metode SVM dan *Naïve Bayes* untuk menemukan model klasifikasi yang paling optimal pada kasus karhutla serta melakukan klasifikasi daerah yang termasuk *cluster* (kelas) rawan dan sangat rawan di Ketapang. Hasil penelitian ini dapat digunakan untuk membantu pemerintah dalam mengambil kebijakan yang efektif dalam menanggulangi kasus karhutla.

## METODE PENELITIAN

Metode CRISP-DM merupakan metode yang digunakan dalam penelitian ini, CRISP-DM, atau *Cross Industry Standard Process for Data Mining*, adalah metodologi terstruktur yang digunakan untuk proyek penambangan data. Ini terdiri dari enam fase: pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penerapan [21]. Gambar 1 adalah prosedur penelitian yang digunakan dalam penelitian ini.



Gambar 1 Prosedur penelitian CRISP-DM

### ***Business Understanding***

Tahap ini dilakukan studi literatur mengenai kasus kebakaran hutan dan lahan di Kabupaten Ketapang, kemudian mempelajari model klasifikasi *SVM* dan *Naïve Bayes*, selanjutnya

mempelajari penggunaan *tools* RapidMiner serta mempelajari tentang *Confusion Matrix*.

### **Data Understanding**

Penelitian ini menggunakan data sekunder yang bersumber dari situs resmi BRIN <https://hotspot.brin.go.id/> dan data yang diperoleh adalah *dataset hotspot* Kabupaten Ketapang pada tahun 2021 dan 2022. Dataset tersebut berisi 4.697 entri yang mencakup parameter id, tanggal (WIB), waktu (WIB), *longitude*, *latitude*, tingkat kepercayaan, satelit, radius kemungkinan, kecamatan, kabupaten, provinsi dan tipe.

### **Data Preparation**

Pada tahap ini dilakukan eliminasi data (*Data Cleaning*), pemilihan data dan atribut (*Data Selection*) serta melakukan transformasi data (*Data Transformation*) yang digunakan untuk penelitian.

### **Data Selection**

*Dataset* bencana yang digunakan dari tahun 2021 – 2022 sebanyak 4.697 *records* akan diseleksi sehingga parameter yang akan digunakan adalah kecamatan, bujur, lintang, tanggal (WIB), tingkat kepercayaan dan radius kemungkinan. Kemudian *dataset* tersebut parameteranya dilengkapi dengan data yang diperoleh dari website BMKG [https://dataonline.bmkg.go.id/data\\_iklim](https://dataonline.bmkg.go.id/data_iklim), adapun parameter yang ditambah adalah curah hujan, kelembapan dan suhu. *Dataset* yang digunakan pada penelitian terlihat pada Tabel 2.

### **Data Cleaning**

Pada tahap ini dilakukan proses pembersihan data seperti *missing value*. Pada *dataset* Gambar 2 data sudah lengkap dan tidak ada *missing value* sehingga bisa dilanjutkan ke tahap selanjutnya.

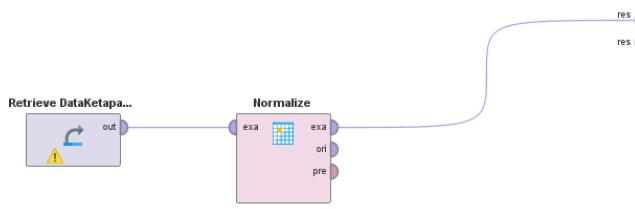
**Gambar 2** Proses *cleaning* saat *input* dataset di RapidMiner

### **Data Transformation**

Pada tahap ini dilakukan proses merubah bentuk data agar sesuai dalam proses *data mining*. Pada dataset ini dilakukan normalisasi data menggunakan *tools* RapidMiner dengan ilustrasi seperti pada Gambar 3. Langkah-langkah untuk melakukan normalisasi di RapidMiner yaitu memasukkan *dataset* yang akan diolah, kemudian memasukkan operator *Normalize*, setelah itu hubungkan *dataset* dengan operator *Normalize*, pada operator *Normalize* pilih metode normalisasi yang akan dilakukan, pada penelitian ini menggunakan metode Z-

*Transformation. Z-score normalization (Z-Transformation)* merupakan metode normalisasi berdasarkan *mean* (nilai rata-rata) dan *standard deviation* (deviasi standar) dari data. Metode ini sangat berguna jika tidak diketahui nilai aktual minimum dan maksimum dari data[15].

Setelah proses pengolahan data selesai dilakukan, diperoleh hasil sebagaimana ditunjukkan pada Tabel 3. Data yang dihasilkan tersebut telah memenuhi kriteria dan dinyatakan sesuai untuk digunakan dalam proses data mining.



**Gambar 3.** Proses Normalisasi data di RapidMiner

## Pemodelan

Pada tahap ini dilakukan pemodelan memanfaatkan *Machine Learning* untuk menentukan model *data mining*, *tools data mining* dan algoritma *data mining*. Penelitian ini menggunakan metode klasifikasi SVM dan *Naïve Bayes*. Sebelum proses pemodelan dengan SVM dan *Naïve Bayes*, jumlah *cluster* (kelas) yang akan dibentuk ditentukan terlebih dahulu. Penelitian ini menggunakan dua *cluster* yaitu *cluster* rawan dan *cluster* sangat rawan. *Cluster* tersebut dibagi berdasarkan tingkat kepercayaan atau *confidence hotspot* [16].

Jadi berdasarkan level dan perlakuan yang dilakukan dapat disimpulkan untuk *confidence*  $30\% \leq C \leq 80\%$  akan menjadi *cluster* rawan dan untuk *confidence*  $80\% \leq C \leq 100\%$  akan menjadi *cluster* sangat rawan.

## Proses Klasifikasi SVM

Langkah-langkah *SVM* untuk melakukan klasifikasi adalah [17]

- a. Input *dataset*
- b. Cari nilai kelasnya yakni kelas +1 (*negatif*) dan kelas -1 (*positif*). Nilai kelas didapatkan dengan Persamaan 1.

$$y(x) = w^T \Phi(x) + b \quad (1)$$

Pada persamaan 1, diketahui  $x$  adalah nilai *vector input*,  $w$  adalah nilai *weight*,  $\Phi(x)$  adalah fungsi basis dan  $b$  adalah nilai bias.

- c. Kemudian akan dicari dua data dari nilai kelas berbeda yang memiliki jarak terdekat (*support vector*).
- d. Selanjutnya akan dicari garis pemisah (*hyperplane*).

Persamaan 2 untuk mencari *hyperplane*.

$$Y_i (w x_i + b) \geq 1 \quad (2)$$

- e. Setelah itu cari jarak maksimum *hyperplane* dengan *support vector* (*margin*) untuk mendapatkan *hyperplane* terbaik dari dua data.

Persamaan 3 digunakan untuk mencari jarak maksimum.

$$\text{Min } \frac{1}{2} \|w\|^2 = \min \frac{1}{2} (w_{12} + w_{21}) \quad (3)$$

**Tabel 1** Dataset awal

id	Tanggal	Waktu	Latitude	Longitude	Tingkat Kepercayaan	Satelite	Radius Kemungkinan	Kecamatan	Kabupaten	Provinsi	Tipe
3567303	02/01/2022	12:24:42	-263.206196	1.102798996	8	Noaa20	1125	Kendawangan	Ketapang	Kalimantan Barat	cluster
3567317	02/01/2022	13:18:46	-265.371752	11.036894226	8	Snpp	1125	Kendawangan	Ketapang	Kalimantan Barat	cluster
3567369	04/01/2022	12:39:29	-266.9885755	11.027286148	8	snpp	1280	Kendawangan	Ketapang	Kalimantan Barat	cluster
....	....	....	.....	.....	....	....	.....	.....	.....	.....	....
3758366	31/12/2022	12:50:00	-0.5443345904350281	11.001763153076100	7	aqua	3000	Simpanghulu	Ketapang	Kalimantan Barat	cluster

**Tabel 2** Dataset yang digunakan dalam penelitian

Tingkat kepercayaan	Curah Hujan (mm)	Temperatur rata-rata (C)	Kelembapan rata-rata (%)	Radius Kemungkinan	Tanggal (WIB)	Latitude	Longitude	kecamatan
0,0	0,0	-0,4	0,2	0,2	2021-01-17 00:00:00	-2818864743333300,0	11031722259666600,0	Kendawangan
0,0	0,0	-0,4	0,2	-0,3	2021-01-17 00:00:00	-264375353,0	1111241684,0	Manisamata
0,0	0,0	-0,4	0,2	-0,3	2021-01-17 00:00:00	-272007442,0	11032154846,0	Kendawangan
0,0	0,0	-0,4	-0,6	3,2	2021-01-20 00:00:00	-0,9	110493896484375,0	Sungai Laur
0,0	0,0	-0,4	-0,6	0,0	2021-01-20 00:00:00	-0,8	11049296951,0	Sungai Laur

**Tabel 3** Hasil normalisasi dataset

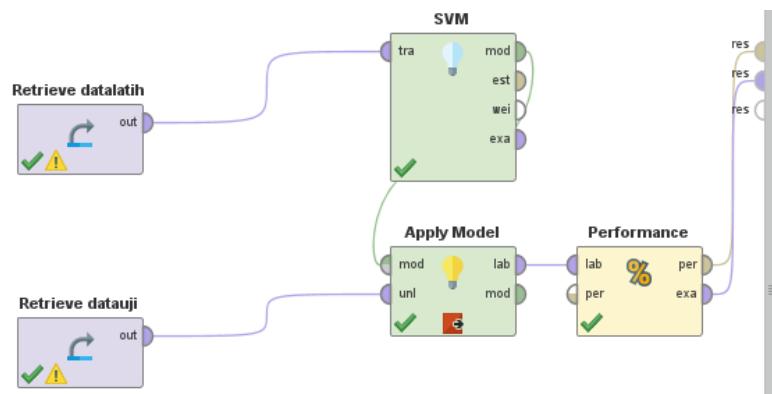
tingkat kepercayaan	Curah Hujan (mm)	temperatur rata-rata (C)	Kelembapan rata-rata (%)	Radius Kemungkinan	Tanggal (WIB)	Latitude	Longitude	kecamatan	Status
0,0	0,0	-0,4	0,2	0,2	2021-01-17 00:00:00	-2818864743333300,0	11031722259666600,0	Kendawangan	Rawan
0,0	0,0	-0,4	0,2	-0,3	2021-01-17 00:00:00	-264375353,0	1111241684,0	Manisamata	Rawan
0,0	0,0	-0,4	0,2	-0,3	2021-01-17 00:00:00	-272007442,0	11032154846,0	Kendawangan	Rawan
0,0	0,0	-0,4	-0,6	3,2	2021-01-20 00:00:00	-0,9	110493896484375,0	Sungai Laur	Rawan
0,0	0,0	-0,4	-0,6	0,0	2021-01-20 00:00:00	-0,8	11049296951,0	Sungai Laur	Rawan

**Tabel 4** Nilai Confidence

Tingkat Kepercayaan	Level	Perlakuan
$0\% \leq C \leq 30\%$	Rendah	Perlu Diawasi
$30\% \leq C \leq 80\%$	Sedang	Hati-hati
$80\% \leq C \leq 100\%$	Tinggi	Segera dilakukan Penaggulangan

### SVM di Rapid Miner

Gambar 4 merupakan proses klasifikasi SVM menggunakan *tools* RapidMiner, adapun langkah-langkah untuk melakukan klasifikasi adalah pertama *input data training* dan *data testing*, setelah itu *input* operator *Support Vector Machine* (SVM), selanjutnya *input* operator *Applymodel* untuk menyambungkan *data testing* dengan model SVM yang telah terbentuk, dan terakhir *input* operator *performance* untuk menguji model algoritma SVM.

**Gambar 4** Model SVM di Rapid Miner

### Proses Klasifikasi Naïve Bayes

Langkah-langkah dalam proses klasifikasi menggunakan metode *Naïve Bayes* yaitu [18], Pertama, dilakukan pengamatan terhadap data latih. Kemudian, hitung total dan probabilitas dari setiap kelas. Rumus untuk menghitung probabilitas kelas dapat menggunakan Persamaan 4.

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (4)$$

Pada Persamaan 4, diketahui X adalah data yang kelasnya belum ditemukan, H adalah hipotesis dari data yang kelas yang ditentukan,  $P(H|X)$  adalah probabilitas hipotesis H berlandaskan dari kondisi X (*posteriori probabilitas*),  $P(H)$  adalah probabilitas hipotesis H (*prior probabilitas*),  $P(X|H)$  adalah probabilitas X berlandaskan pada kondisi hipotesis H dan  $P(X)$  adalah probabilitas dari X.

Apabila data tersebut bersifat numerik maka langkah pertama mencari nilai *mean* dan standar deviasi dari setiap parameter, dimana rumus untuk menghitung *mean* ada pada Persamaan 5.

$$\mu = \frac{\text{Jumlah Data}}{\text{Banyak Data}} \quad (5)$$

Rumus untuk menghitung standar deviasi terdapat pada Persamaan 6.

$$\sigma = \sqrt{\frac{\sum_i^n (x_i - \mu)^2}{N-1}} \quad (6)$$

Pada Persamaan 6, diketahui  $\sigma$  adalah standar deviasi,  $x$  adalah nilai  $x$  ke-  $i$ ,  $n$  adalah total sampel dan  $\mu$  adalah rata-rata hitung (*mean*). Kedua, mencari nilai probabilistik dengan menghitung total keseluruhan data dalam kategori kemudian dibagi dengan total seluruh data yang ada.

Selanjutnya, hitung nilai probabilitas dari data uji yang merupakan data numerik. Untuk menghitung probabilitas, digunakan Persamaan 7 distribusi Gaussian:

$$P = (X_i = x_i | Y = y_j) = \frac{1}{\sqrt{2\pi\sigma_{ij}^2}} \times e^{-\frac{(x_i - \mu_{ij})^2}{2\sigma_{ij}^2}} \quad (7)$$

Pada Persamaan 7, diketahui  $P$  adalah probabilitas,  $X_i$  adalah atribut ke- $i$ ,  $X_j$  adalah Nilai atribut ke- $i$ ,  $C$  adalah kelas yang dicari,  $C_i$  adalah *subclass*  $Y$  yang dicari,  $\mu$  adalah nilai mean dari semua atribut dan  $\sigma$  adalah standar deviasi yang menyajikan varian dari semua atribut.

Kemudian, hitung probabilitas akhir dari masing-masing kelas. Nilai probabilitas akhir tiap kelas diperoleh dengan mengalikan seluruh nilai distribusi Gaussian dari atribut-atribut dalam kelas yang sama sesuai Persamaan 8:

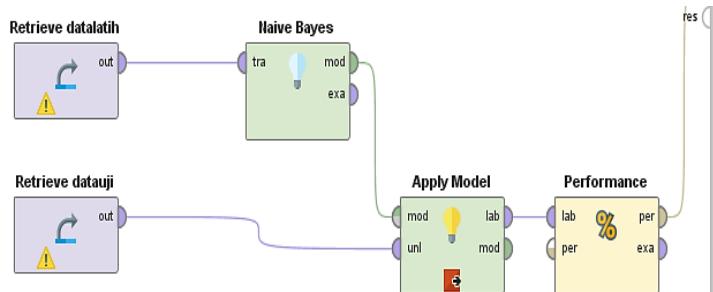
$$P(X|Kelas) = P(V1|Kelas) \times P(V2|Kelas) \times P(V3|Kelas) \times P(V4|Kelas) \times P(V5|Kelas) \times P(V6|Kelas) \times P(V7|Kelas) \quad (8)$$

Tahap terakhir yaitu menghitung probabilitas menggunakan perhitungan nilai probabilitas akhir kelas ke persamaan *Naïve Bayes Classifier*. Persamaan 9 merupakan rumus probabilitas akhir.

$$P(Kelas|X) = P(Kelas) \times P(X) \quad (9)$$

### ***Naïve Bayes* di RapidMiner**

Gambar 5 merupakan proses klasifikasi *Naïve Bayes* menggunakan *tools* RapidMiner, adapun langkah-langkah untuk melakukan klasifikasi adalah pertama *input data training* dan *data testing*, setelah itu *input* operator *Naïve Bayes*, selanjutnya *input* operator *Apply Model* untuk menyambungkan *data testing* dengan model *Naïve Bayes* yang telah terbentuk, dan terakhir *input* operator *performance* untuk menguji model algoritma *Naïve Bayes*.



**Gambar 5** Model algoritma *Naïve Bayes* di RapidMiner

### **Evaluasi**

Langkah ini dilakukan prosedur evaluasi hasil klasifikasi yang dilakukan oleh algoritma SVM dan *Naïve Bayes*. Proses ini dilakukan perhitungan untuk mencari nilai *accuracy* menggunakan tabel *Confusion Matrix*. *Confusion Matrix* adalah metode yang digunakan untuk mengukur suatu kinerja model [19]. *Confusion matrix* dapat disebut juga sebagai sebuah tabel yang berfungsi untuk menggambarkan performa dari suatu model [20].

*Accuracy* merupakan total seberapa sering model melakukan klasifikasi dengan benar. Persamaan 10 digunakan untuk menghitung *accuracy*.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (10)$$

*Precision*, saat model memprediksi positif, seberapa sering hasil prediksi model benar. Persamaan 11 dari *Confusion matrix* untuk menghitung *precision*.

$$Precision = \frac{TP}{TP+FP} \quad (11)$$

*Recall*, saat kelas aktual positif, seberapa sering model memprediksi positif.

Persamaan 12 dilakukan untuk menghitung *recall* yaitu:

$$Recall = \frac{TP}{FN+TP} \quad (12)$$

*True Positive (TP)* artinya ada berapa total data aktual yang bernilai positif serta model memprediksinya positif, *True Negative (TN)* artinya ada berapa total data aktual yang bernilai negatif serta model memprediksinya negatif, *False Negative (FN)* artinya ada berapa total data aktual yang bernilai positif akan tetapi model memprediksinya negatif dan *False Positive (FP)* artinya ada berapa total data yang aktual kelasnya bernilai negatif akan tetapi model memprediksinya positif.

## HASIL DAN PEMBAHASAN

Pada langkah ini dilakukan komparasi akurasi model algoritma SVM dan *Naïve Bayes* untuk menemukan algoritma yang paling optimal dalam klasifikasi karhutla.

accuracy: 98.69%			
	true Rawan	true Sangat Rawan	class precision
pred. Rawan	1732	21	98.80%
pred. Sangat Rawan	2	0	0.00%
class recall	99.88%	0.00%	

**Gambar 6** Hasil pengujian algoritma SVM

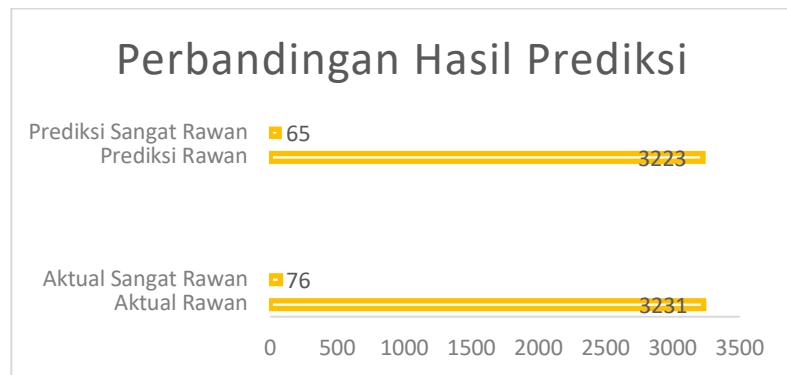
Berdasarkan Gambar 6 hasil nilai akurasi sebesar 98,69% dihasilkan nilai *precision* yang mencirikan daerah Rawan sebesar 98,80%, serta nilai *precision* yang mencirikan daerah Rawan sebesar 0,00%. Selain itu, nilai *recall* yang mencirikan daerah Rawan sebesar 99,88% dan nilai *recall* yang mencirikan daerah Sangat Rawan sebesar 0,00%. Sehingga dapat diketahui bahwa algoritma SVM gagal dalam mencirikan daerah Sangat Rawan pada data.

accuracy: 99.94%			
	true Rawan	true Sangat Rawan	class precision
pred. Rawan	1734	1	99.94%
pred. Sangat Rawan	0	20	100.00%
class recall	100.00%	95.24%	

**Gambar 7** Hasil pengujian algoritma *Naïve Bayes*

Berdasarkan Gambar 7 hasil pengujian algoritma *Naïve Bayes* diperoleh nilai akurasi 99,94% dimana nilai *precision* yang mencirikan daerah Rawan 99,94% dan nilai *precision*

yang mencirikan daerah Sangat Rawan adalah 100%. Nilai *recall* yang mencirikan daerah Rawan 100% dan Nilai *recall* yang mencirikan daerah Sangat Rawan 95,24%.



**Gambar 8** Perbandingan data aktual dan data prediksi *Naïve Bayes*

Gambar 8 menunjukkan bahwa model *Naïve Bayes* dapat memprediksi status rawan sebanyak 3.223 data dari data aktual rawan sebanyak 3.212 data sedangkan untuk status sangat rawan model memprediksi sebanyak 65 data dari data aktual rawan sebanyak 76 data. Hasil ini dinilai sangat baik karena hanya memiliki sedikit perbedaan dengan data aktual.

### **Evaluation**

Berdasarkan hasil penelitian yang dilakukan menggunakan 4.697 data, dapat diketahui bahwa nilai akurasi yang konvergen dari algoritma SVM sebesar 98,24%, sedangkan nilai akurasi yang konvergen algoritma *Naïve Bayes* sebesar 99,67%. Sehingga dapat disimpulkan bahwa algoritma *Naïve Bayes* dapat melakukan klasifikasi lebih akurat daripada algoritma SVM dalam kasus kebakaran hutan dan lahan.

## **KESIMPULAN**

Penelitian ini menerapkan teknik data mining untuk klasifikasi kebakaran hutan dan lahan di Kabupaten Ketapang menggunakan RapidMiner. Hasil analisis menunjukkan bahwa algoritma *Naïve Bayes* memiliki tingkat akurasi yang lebih tinggi dibandingkan dengan algoritma SVM. Setelah membandingkan hasil prediksi *Naïve Bayes* dengan data aktual kejadian karhutla, dapat disimpulkan bahwa *Naïve Bayes* merupakan algoritma dengan performa terbaik untuk klasifikasi kebakaran hutan dan lahan di wilayah ini. Oleh karena itu, model berbasis *Naïve Bayes* diharapkan dapat digunakan untuk mengidentifikasi area rawan dan sangat rawan karhutla, sehingga dapat mendukung pemerintah dalam merancang kebijakan mitigasi dan penanggulangan bencana kebakaran hutan dan lahan secara lebih efektif.

## **DAFTAR PUSTAKA**

1. A. Loren, M. Ruslan, F. H. Yusran, and F. Rianawati, “Analisis Faktor Penyebab Kebakaran Hutan,” *Enviro Scientiae*, vol. 11, pp. 1–999, 2015.
2. Endrawati, J. Purwanto, S. Nugroho, and R. Agung S, “IDENTIFIKASI AREAL BEKAS KEBAKARAN HUTAN DAN LAHAN MENGGUNAKAN ANALISIS SEMI OTOMATIS CITRA SATELIT LANDSAT,” 2017. [Online]. Available: <https://modis-catalog.lapan.go.id/monitoring>

3. Sutiyanti, Juanda, and S. Syam Saguni, “REPRESENTASI KERUSAKAN LINGKUNGAN DI INDONESIA DALAM PUISI MEDIA DARING INDONESIA (KAJIAN EKOKRITIK),” Makassar, 2019.
4. H. Muhamad, C. A. Prasojo, N. A. Sugianto, L. Surtiningsih, and I. Cholissodin, “OPTIMASI NAÏVE BAYES CLASSIFIER DENGAN MENGGUNAKAN PARTICLE SWARM OPTIMIZATION PADA DATA IRIS,” *Jurnal Teknologi Informasi da Ilmu Komputer (JTIK)*, vol. 4, no. 3, pp. 180–184, 2017.
5. J. Yandi, T. B. Kurniawan, E. S. Negara, and M. Akbar, “Prediksi Lokasi Titik Panas Kebaran Hutan Menggunakan Model Regresion SVM (Support Vector Machine) pada Data Kebakaran Hutan Daops Manggala Agni Oki Provinsi Sumatera Selatan Tahun 2019,” *InfoTekjar : Jurnal Nasional Informatika dan Teknologi Jaringan*, vol. 6, no. 1, pp. 10–15, 2021, doi: 10.30743/infotekjar.v6i1.4101.
6. R. Oktaria Mardiyanto, Fitriani, R. Joko Purnomo, Kusrini, and D. Maulina, “PEMETAAN LOKASI KEBAKARAN HUTAN DAN LAHAN DI NTB DENGAN MENGGUNAKAN ALGORITMA NAIVE BAYES,” *TEKNIMEDIA*, vol. 2, no. 2, pp. 69–75, 2021.
7. K. Supribadi, N. Khakhim, and T. H. Purwanto, “Analisis Metode Support Vector Machine (SVM) Untuk Klasifikasi Penggunaan Lahan Berbasis Penutup Lahan Pada Citra Alos Avnir-2,” *Majalah Geografi Indonesia*, vol. 28, no. 1, pp. 71–80, 2014.
8. R. R. Fiska, “Penerapan Teknik Data Mining dengan Metode Support Vector Machine (SVM) untuk Memprediksi Siswa yang Berpeluang Drop Out (Studi Kasus di SMKN 1 Sutera),” *SATIN-Sains dan Teknologi Informasi*, vol. 3, no. 1, pp. 15–23, 2017, [Online]. Available: <http://jurnal.stmik-amik-riau.ac.id>
9. A. Vatresia, R. Regen, F. P. Utama, and R. Febrianti, “Support Vector Machine Analysis for Potential Hotspot Over Papua Island Support Vector Machine untuk Potensi Hotspot pada Pulau Papua,” *Jurnal Teknologi Lingkungan*, vol. 24, no. 1, p. 81, 2023.
10. E. Saputra, U. Khaira, and Z. Abidin, “Implementasi Algoritme Support Vector Machines untuk Klasifikasi Area Terbakar di Lahan Gambut,” *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 10, no. 1, pp. 19–24, 2021.
11. Angellina, D. Erny Herwindiati, and J. Hendryli, “Performa Support Vector Machine Pada Klasifikasi Lahan dan Air Tanah,” *Jurnal Media Informatika Budidarma*, vol. 7, no. 1, 2023, doi: 10.30865/mib.v7i1.5279.
12. A. Ashari Muin, “Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi),” *Jurnal Ilmiah Ilmu Komputer*, vol. 2, no. 1, 2016, [Online]. Available: <http://ejournal.fikom-unasman.ac.id>
13. M. Farid Rifai, H. Jatnika, B. Valentino, and S. Tinggi Teknik PLN, “Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS),” *Jurnal Pengkajian dan Penerapan Teknik Informatika*, vol. 12, no. 2, pp. 131–144, 2019.
14. T. A. Pratiwi, M. Irsyad, and R. Kurniawan, “Klasifikasi Kebakaran Hutan dan Lahan Menggunakan Algoritma Naïve Bayes (Studi Kasus: Provinsi Riau),” *Jurnal Sistem dan Teknologi Informasi (Justin)*, vol. 9, no. 2, p. 101, Apr. 2021, doi: 10.26418/justin.v9i2.42823.

15. D. Azzahra Nasution, H. H. Khotimah, and N. Chamidah, “PERBANDINGAN NORMALISASI DATA UNTUK KLASIFIKASI WINE MENGGUNAKAN ALGORITMA K-NN,” *CESS (Journal of Computer Engineering System and Science)*, vol. 4, no. 1, pp. 2502–7131, 2019.
16. L. Giglio, W. Schroeder, J. V Hall, and C. O. Justice, “MODIS Collection 6 Active Fire Product User’s Guide Revision C,” 2020.
17. W. Winjani, M. Fatchan, and N. Nurhidayanti, “PREDIKSI KENAIKAN GAJI KARYAWAN DENGAN PENDEKATAN ALGORITMA SUPPORT VECTOR MACHINE PADA PT. DAEINDO INTERNATIONAL,” *JINTEKS*, vol. 4, no. 2, pp. 114–117, 2022.
18. M. Nizam Fadli, I. Sudahri Damanik, E. Irawan, S. Tunas Bangsa, and S. Utara, “Penerapan Metode Naive Bayes Dalam Menentukan Tingkat Kenyamanan Pada Rumah Sakit Terhadap Pasien,” *KLIK: Kajian Ilmiah Informatika dan Komputer* , vol. 2, no. 3, pp. 117–122, 2021, [Online]. Available: <https://djournals.com/klik>
19. Karsito and S. Susanti, “PENGAJUAN KREDIT RUMAH DENGAN ALGORITMA NAÏVE BAYES DI PERUMAHAN AZZURA RESIDENCIA,” *SIGMA (Jurnal Tekologi Pelita Bangsa)*, vol. 9, no. 3, pp. 43–48, 2019.
20. I. Widhi Saputro and B. Wulan Sari, “Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa Naïve Bayes Algorithm Performance Test for Student Study Prediction,” *Citec Journal*, vol. 6, no. 1, pp. 1–11, 2019.
21. Singgalen YA. Penerapan Metode CRISP-DM untuk Optimalisasi Strategi Pemasaran STP (Segmenting, Targeting, Positioning) Layanan Akomodasi Hotel, Homestay, dan Resort. *Jurnal media informatika Budidarma*. 2023.