

Klasifikasi Kebakaran Hutan dan Lahan Menggunakan Algoritma Naïve Bayes (Studi Kasus: Provinsi Riau)

Trya Ayu Pratiwi^{a1}, Muhammad Irsyad^{a2}, Rahmad Kurniawan^{a3}

^aUniversitas Islam Negeri Sultan Syarif Kasim Riau

Jalan. HR. Soebrantas Km. 15 Panam Pekanbaru, Riau

¹111651203413@students.uin-suska.ac.id

²irsyadtech@uin-suska.ac.id

³rahmadkurniawan@uin-suska.ac.id

Abstrak

Kebakaran hutan dan lahan yang terjadi di Indonesia telah menimbulkan banyak kerugian bagi masyarakat. Kebakaran hutan umumnya terjadi pada bulan Agustus dan September, bertepatan dengan musim kemarau pada sebagian besar wilayah Indonesia. Salah satu indikator terjadinya kebakaran hutan adalah titik panas. Penelitian ini menggunakan salah satu teknik data mining yaitu mengklasifikasi titik panas yang ada di Provinsi Riau. Penelitian ini menggunakan dataset kebakaran hutan Kabupaten Pelalawan pada tahun 2015 sampai 2019 dengan menggunakan algoritma Naïve Bayes. Titik panas yang akan dianalisis terdiri dari suhu, kelembaban, curah hujan, kecepatan angin, dan kelas. Akurasi tertinggi dari dataset kebakaran hutan dan lahan pada tahun 2019 adalah sebesar 96.95%. Metode klasifikasi menggunakan algoritma Naïve Bayes dapat digunakan untuk memprediksi kemunculan titik panas di masa yang akan datang sehingga dapat melakukan tindakan pencegahan sebelum terjadinya kebakaran hutan dan lahan.

Kata kunci: Kebakaran Hutan dan Lahan, Titik Panas, Naïve Bayes

Classification of Forest and Land Fire Using Naïve Bayes Algorithm (Case Study: Riau Province)

Abstract

Forest and land fires that occurred in Indonesia have caused many losses for the community. Forest fires generally occur in August and September, coinciding with the dry season in most parts of Indonesia. One indicator of the occurrence of forest fires is hotspots. This study uses one of the data mining techniques, namely classifying hotspots in Riau Province. This study used a dataset of forest fires in Pelalawan Regency from 2015 to 2019 using the Naïve Bayes algorithm. The hot spots to be analyzed consist of temperature, humidity, rainfall, wind speed, and class. The highest accuracy of the dataset of forest and land fires in 2019 is 96.95%. The classification method using the Naïve Bayes algorithm can be used to predict the emergence of hotspots in the future so that they can take preventive measures before forest and land fires occur.

Keywords: Forest and Land Fires, Hotspots, Naïve Bayes

I. PENDAHULUAN

Kebakaran hutan dan lahan yang terjadi di Indonesia telah menimbulkan banyak kerugian bagi masyarakat. Luas kebakaran hutan dan lahan di Indonesia telah meningkat secara drastis pada tahun 2015 dengan luas 2.611.411,44 Ha [3]. Luas kebakaran hutan dan lahan terburuk di Indonesia terjadi di Provinsi Riau, Jambi, Sumatera Selatan dan Kalimantan Tengah [3]. Provinsi Riau merupakan salah satu wilayah dengan kebakaran hutan dan lahan yang cukup tinggi dan terjadi setiap tahun. Hasil data luas kebakaran hutan dan lahan di Provinsi Riau dari pantauan Karhutla Monitoring System, area yang terbakar pada tahun 2015 yakni sekitar 183.808,59 Ha, pada tahun 2016 sekitar 85.219,51 Ha, pada 2017 sekitar 6.866,09 Ha, pada tahun 2018 sekitar 37.236,27

Ha hingga pada tahun 2019 sekitar 76.267,00 Ha [3]. Kebakaran hutan umumnya terjadi pada bulan Agustus dan September, bertepatan dengan musim kemarau pada sebagian besar wilayah Indonesia [15]. Curah hujan yang sangat rendah pada musim kemarau mempengaruhi terjadinya kebakaran hutan karena menyebabkan hutan dan lahan mudah terbakar [7].

Dikutip dari www.kompas.com (2019) berbagai upaya pemerintah dalam mencegah terjadinya kebakaran hutan di Provinsi Riau yaitu dengan melakukan water bombing, menyediakan masker gratis, meliburkan aktivitas sekolah, tetapi hal tersebut bukan solusi yang tepat untuk jangka panjang [9]. Tindakan yang tegas dan nyata seperti penegakan hukum, dan diperlukan sistem alarm kebakaran yang dapat berfungsi sebagai sistem peringatan dini [16].

Dalam pengembangan sistem peringatan dini, diperlukan hasil dari Data Mining yang akurat dan dapat digunakan untuk memberikan informasi yang efektif dalam pencegahan kebakaran hutan. Oleh karena itu, penelitian ini bertujuan untuk menggunakan Data Mining dalam mengklasifikasi titik panas yang ada di Provinsi Riau.

Data Mining adalah proses ekstraksi suatu data yang bersifat implisit dan tidak berguna menjadi informasi dari data yang jumlahnya besar [32]. Salah satu bagian dari Data Mining adalah metode klasifikasi [30]. Salah satu penerapan klasifikasi pada kebakaran hutan adalah K-Nearest Neighbor untuk klasifikasi indeks cuaca kebakaran berdasarkan data Automatic Weather Station [25]. Berdasarkan hasil penelitian Reza Noviansyah (2018) diperoleh bahwa algoritma K-Nearest Neighbor akurat digunakan untuk menghasilkan klasifikasi tingkat bahaya kebakaran hutan dengan atribut data berupa suhu, kelembaban, curah hujan dan kecepatan angin, yang menghasilkan akurasi persentase keberhasilan 80,16% dengan nilai K=5. Namun algoritma K-Nearest Neighbor ini mempunyai kelemahan seperti menentukan nilai K optimal, tidak menangani missing value secara implisit, sensitif terhadap data outlier dan nilai komputasi yang tinggi [5].

Penelitian lainnya dilakukan oleh Negara & Kurniawan (2019) menggunakan Decision Tree dan Bayesian Network. Algoritma ini memiliki kelebihan dalam menjelaskan hubungan setiap atribut data cuaca yang mempengaruhi tingkat hotspot menggunakan Decision Tree dan Bayesian Network memiliki nilai akurasi yang baik dan tingkat kesalahan yang rendah. Namun, algoritma Decision Tree memiliki kelemahan seperti terjadi tumpang tindih jika kelas dan kriteria yang digunakan sangat banyak, ini menyebabkan waktu keputusan sangat lama dan membutuhkan banyak memori [2]. Walaupun Bayesian Network memiliki kelebihan dan banyak digunakan oleh para peneliti, Bayesian Network juga memiliki kelemahan yaitu sulit untuk mencapai kesepakatan beberapa pakar dalam membangun struktur Bayesian Network, pakar juga akan memerlukan waktu yang lama hanya untuk membangun struktur Bayesian Network dan Pakar akan sulit menentukan conditional probability berdasarkan pendapat pakar [19].

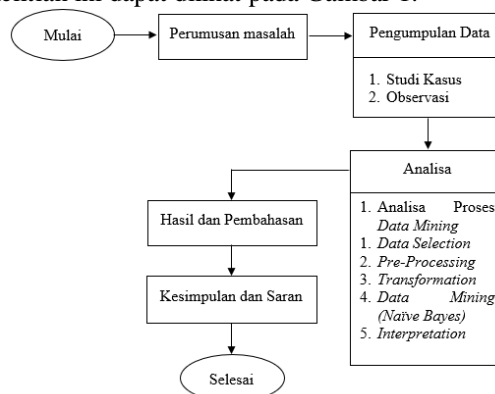
Penelitian yang dilakukan oleh Widiastuti (2012) yang membahas tentang perbandingan algoritma akan dilihat berdasarkan nilai correctly classified instance, incorrectly classified, kappa statistic, true positive, false positive, dan confusion matrix. Dalam penelitian ini, algoritma yang dibandingkan adalah algoritma SVM, Decision Tree, dan Naive Bayes. Dapat disimpulkan algoritma yang memiliki kinerja yang lebih unggul adalah Decision Tree.

Berdasarkan tinjauan pustaka, telah dilakukan penelitian terhadap analisis perbandingan algoritma SVM, Naive Bayes, dan Decision Tree dalam klasifikasi [31]. Decision Tree menjadi algoritma terbaik tetapi tidak maksimal dalam waktu komputasi (running time), untuk membangun sebuah model algoritma yang tercepat dari ketiga algoritma tersebut adalah Naive Bayes [20]. Maka penelitian ini menggunakan algoritma Naive Bayes untuk menganalisis metode klasifikasi untuk memprediksi

kemunculan titik panas di Provinsi Riau. Metode klasifikasi Naive Bayes dipilih karena memiliki pemodelan dan hasil akurasi yang baik dalam set pelatihan data dan algoritma Naive Bayes merupakan algoritma yang paling baik dalam hal waktu komputasi (waktu yang dibutuhkan untuk membangun sebuah model), dan sering bekerja dengan running time yang tercepat dengan tingkat akurasi yang tinggi [20]. Penelitian ini diharapkan dapat menghasilkan model yang akurat untuk memprediksi kemungkinan munculnya titik panas di masa yang akan datang sehingga pihak yang berwenang dapat melakukan tindakan pencegahan terjadinya kebakaran hutan dan lahan.

II. METODOLOGI PENELITIAN

Metodologi penelitian adalah tahapan atau proses yang disusun secara sistematis dan logis dalam melakukan suatu penelitian yang berguna untuk mencapai target yang diharapkan. Tahapan penelitian yang akan dilakukan pada penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Tahapan penelitian

A. Perumusan Masalah

Perumusan masalah adalah tahap utama dari metodologi penelitian. Dalam penelitian yang akan dilakukan rumusan masalahnya yaitu bagaimana menganalisis metode klasifikasi untuk memprediksi kemunculan titik panas menggunakan algoritma *Naive Bayes* sehingga diharapkan dapat menghasilkan model yang akurat.

B. Pengumpulan Data

Pengumpulan data dilakukan dengan dua tahap yaitu studi pustaka dan observasi. Penjelasan dari pengumpulan data adalah sebagai berikut:

1) *Observasi*: Pada tahapan observasi ini, dilakukan dengan mengumpulkan data kebakaran hutan. Pengumpulan data dapat dilakukan dengan data primer merupakan data Kabupaten Pelalawan yang diambil dari Badan Meteorologi Klimatologi Dan Geofisika Provinsi Riau dengan format .xlsx.

2) *Studi Literatur*: Studi pustaka atau biasa disebut dengan literature adalah tahapan awal dalam teknik pengumpulan data pada suatu penelitian. Metode ini akan dilakukan dengan cara mencari sumber-sumber referensi

terhadap berbagai buku, karya ilmiah maupun jurnal-jurnal yang dianggap penting dan mendukung.

C. Analisa

Analisa adalah tahapan yang dilakukan setelah pengumpulan data dari metodologi penelitian. Analisa merupakan metode khusus yang digunakan untuk menganalisis masalah. Pada tahapan analisa proses *Data Mining* peneliti akan menjelaskan tentang bagaimana tahapan-tahapan untuk mencari klasifikasi dari sejumlah data kebakaran hutan dengan menggunakan algoritma *Naïve Bayes* sebagai berikut:

1) *Data Selection*: Pada tahap ini dilakukan tahapan seleksi data. Pemilihan data dalam proses seleksi menggunakan 5 atribut yaitu suhu, curah hujan, kecepatan angin, kelembaban tanah dan kelas.

2) *Pre-processing*: Tahapan untuk pembersihan terhadap data-data yang terdapat data kosong (*missing value*), inkonsisten data, dan ketidak sesuaian data (*outlier*).

3) *Transformation*: Data yang telah melalui tahapan proses *pre-processing* ditransformasikan kemudian disimpan ke dalam bentuk yang bisa diterapkan pada *tools* yang akan digunakan nantinya.

Data yang berjenis numerikal seperti suhu, kelembaban, curah hujan, kecepatan angin, dan kelas harus dilakukan proses inisiasi data terlebih dahulu kedalam bentuk nominal yang dapat dilihat pada Tabel I hingga Tabel V.

TABEL I
DATA SUHU

Suhu	Inisialisasi data
<25°	Rendah
26°-29°	Sedang
>29°	Tinggi

TABEL II
DATA KELEMBABAN

Kelembaban	Inisialisasi data
<70-75	Kering
76-80	Sedang
80-85	Lembab
>85	Basah

TABEL III
DATA CURAH HUJUAN

Curah hujan	Inisialisasi data
0	Berawan
0.5-20 mm/hari	Ringan
20-50 mm/hari	Sedang
50-100 mm/hari	Lebat

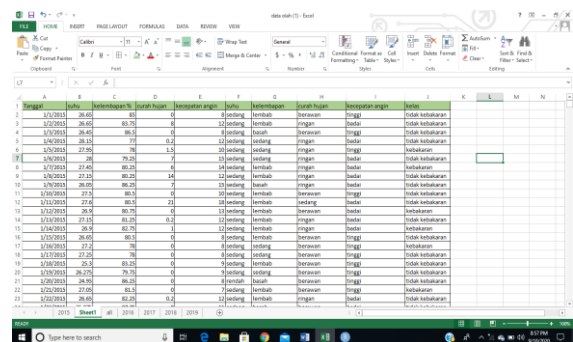
TABEL IV
DATA KECEPATAN ANGIN

Kecepatan angin (Knot)	Inisialisasi data
1-4	Tenang
4-7	Sedang
7-11	Tinggi
>11	Badai

TABEL V
DATA KELAS

Kelas	Inisialisasi data
0-30	Tidak kebakaran
31-80	Kebakaran

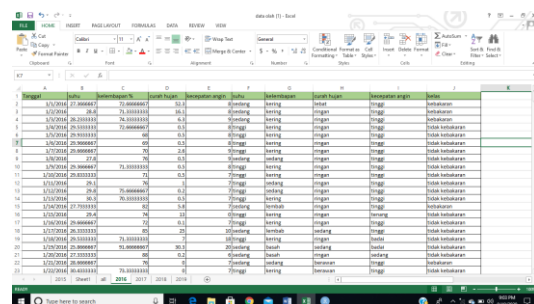
Contoh dataset 2015 yang telah dilakukan inisialisasi berdasarkan atribut dan kelas dapat dilihat Pada pada Gambar 2.



Gambar 2. Inisialisasi data berdasarkan atribut

4) *Data Mining*: Berdasarkan dari tahapan *Data Mining* untuk algoritma *Naïve Bayes*, adapun langkah-langkah dari algoritma *Naïve Bayes* adalah sebagai berikut:

1. Hitung $P(C_i)$ untuk setiap atribut, dalam kasus dataset pada penelitian ini yaitu atribut kelas yang terdiri dari 2 kelas yaitu tidak kebakaran dan kebakaran,
2. Kemudian hitung $P(X|C_i)$, $i=1,2$ untuk setiap kelas atau atribut. Setelah itu bandingkan, jika $P(X|C_1) > P(X|C_2)$ maka kesimpulannya C_1 adalah kelas titik kebakaran = "False". jika $P(X|C_1) < P(X|C_2)$ maka kesimpulannya C_1 adalah kelas kebakaran = "True".
3. Perhitungan manual *naïve bayes* dengan menggunakan dataset pada tahun 2015 dapat dilihat pada Gambar 3.



Gambar.3 Dataset 2016

a. Menghitung probabilitas tiap kelas (P(Ci))

$$C1 \text{ (Class kelas = "tidak kebakaran")} = 232/366 = 0.633879781$$

$$C2 \text{ (Class kelas= "kebakaran")} = 134/366 = 0.366120219$$

b. Menghitung probabilitas kemunculan setiap nilai untuk tiap atribut (P(X|Ci))

TABEL VI
ATRIBUT SUHU

Suhu	tidak kebakaran	kebakaran	P(tidak kebakaran)	P(kebakaran)
Rendah	6	0	0.025862069	0
Sedang	183	98	0.788793103	0.731343284
Tinggi	43	36	0.185344828	0.268656716
Total	232	134		

Pada model Naive Bayes terdapat nilai peluang 0 (lihat Tabel VI). Untuk menangani kasus ini dapat digunakan salah satu teknik smoothing yaitu Laplace estimation. Berikut implementasi Laplace estimation untuk atribut suhu:

$$P(\text{Suhu="rendah"}|\text{kelas="tidak kebakaran"}) = \frac{6+1}{232+3} = 0.029787234$$

$$P(\text{Suhu="rendah"}|\text{kelas="kebakaran"}) = \frac{0+1}{134+3} = 0.00729927$$

TABEL VII
ATRIBUT KELEMBABAN

kelembaban	tidak kebakaran	kebakaran	P(tidak kebakaran)	P(kebakaran)
Kering	42	33	0.181034483	0.246268657
Sedang	55	61	0.237068966	0.455223881
Lembab	83	27	0.357758621	0.201492537
Basah	52	13	0.224137931	0.097014925
Total	232	134		

TABEL VIII
ATRIBUT CURAH HUJAN

Curah hujan	tidak kebakaran	kebakaran	P(tidak kebakaran)	P(kebakaran)
Berawan	82	72	0.353448276	0.537313433
Ringan	116	54	0.5	0.402985075
Sedang	25	5	0.107758621	0.037313433
Lebat	9	3	0.038793103	0.02238806
Total	232	134		

TABEL IX
ATRIBUT KECEPATAN ANGIN

kecepatan angina	tidak kebakaran	kebakaran	P(tidak kebakaran)	P(kebakaran)
Tenang	35	12	0.150862069	0.089552239
Sedang	40	34	0.172413793	0.253731343
Tinggi	95	57	0.409482759	0.425373134
Badai	62	31	0.267241379	0.231343284
Total	232	134		

c. Kalikan semua probabilitas tiap atribut (lihat Tabel VII hingga IX) dengan probabilitas tiap kelas

$P(X|Ci)*P(Ci)$, Jika dilakukan observasi terhadap (suhu="tinggi"), (kelembaban="kering"), (curah hujan="berawan"), dan (kecepatan angin="tinggi") maka peluang *posterior* dapat diestimasi:

$$P(\text{suhu="tinggi"}|\text{kelas="tidak kebakaran"}) * P(\text{kelembaban="kering"}|\text{kelas="tidak kebakaran"}) * P(\text{curah hujan="berawan"}|\text{kelas="tidak kebakaran"}) * P(\text{kecepatan angin="badai"}|\text{kelas="tidak kebakaran"}) = 0.185344828 * 0.181034483 * 0.353448276 * 0.409482759 = 0.004856275$$

$$P(\text{suhu="tinggi"}|\text{kelas="kebakaran"}) * P(\text{kelembaban="kering"}|\text{kelas="kebakaran"}) * P(\text{curah hujan="berawan"}|\text{kelas="kebakaran"}) * P(\text{kecepatan angin="badai"}|\text{kelas="kebakaran"}) = 0.268656716 * 0.246268657 * 0.537313433 * 0.425373134 = 0.002505975$$

d. Bandingkan Hasil Per Kelas, kelas dengan nilai $P(X|Ci) * P(Ci)$ tertinggi adalah kelas yang terpilih $P(X|Ci) * P(Ci)$ tidak kebakaran > $P(X|Ci) * P(Ci)$ kebakaran

Dari hasil diatas, terlihat bahwa nilai probabilitas tertinggi ada pada klasifikasi (**P|kebakaran**) sehingga dapat disimpulkan bahwa status tersebut termasuk dalam klasifikasi "**Kebakaran**".

5) *Evaluation*: Proses pencarian pola atau informasi baru yang menarik dan bermanfaat pada suatu kumpulan data yang terpilih dengan menggunakan Tools R Studio algoritma Naive Bayes.

III. PEMBAHASAN

Semua paragraf mesti di-indent. Semua paragraf mesti rata kiri dan rata kanan.

Pada tahap pembahasan penulis menggunakan pengujian dengan *tools R Studio* untuk melakukan pengecekan *confusion matrix* terhadap klasifikasi yang telah dilakukan oleh algoritma *Naive Bayes*.

A. Pembagian Data

Proses pembagian data *training* dan data *testing* yang dilakukan di *R Studio* dengan menggunakan metode *K-fold cross validation* dengan nilai K=10. Proses pembagian data dilakukan secara otomatis dengan menggunakan salah satu metode yang terdapat pada *package klaR* yaitu *cv*. *Package* ini membagi data dengan menggunakan *K-fold cross validation* dengan nilai K yang dapat ditentukan oleh pengguna.

B. Pembuatan Model Klasifikasi

Dalam pembuatan model klasifikasi menggunakan beberapa *package* yaitu *e1071*, *caret* dan *gmodels*. Metode yang telah dibangun dapat memprediksi kelas dan dapat melihat nilai akurasi terhadap metode tersebut. Nilai akurasi terhadap metode klasifikasi dapat dilihat pada Gambar 4. Berikut adalah contoh pernyataan R yang digunakan untuk membangun metode klasifikasi kebakaran hutan dan lahan 2015 menggunakan algoritma *Naive Bayes* yang dapat dilihat pada Gambar 4.

```

1 library(readxl)
2 data=read.delim("clipboard")
3 view(data)
4 library(klaR)
5 library(caret)
6 x=data.frame(data)
7 y=data.frame(data)$kelas
8 model=train(x,y,'nb',trControl = trainControl(method="cv",number=10),
9           tuneGrid = data.frame(fL = 1,usekernel= FALSE, adjust = FALSE),
10          verbose = TRUE)
11 model
12 predict(model$finalModel,x)
13 predict(model$finalModel,x)$class
14 table(predict(model$finalModel,x)$class,y)
    
```

Gambar 4. Pembuatan model klasifikasi

Package yang digunakan untuk mengklasifikasikan titik panas adalah package *klaR* dan *caret*. Package *caret* digunakan untuk mengklasifikasikan data menggunakan algoritma *Naïve Bayes*, karena memiliki fitur yang baik sedangkan package *klaR* memiliki *Naïve Bayes classifier*. Baris ketujuh dari program merupakan pernyataan untuk melatih model. Baris program ini akan memproses model *Naïve Bayes* menggunakan *10-fold cross validation*. X merupakan *predictor* dan y merupakan label, sedangkan 'nb' menjelaskan bahwa model yang digunakan adalah *Naïve Bayes*. *TrainController* merupakan bagian bahwa proses pembagian data menggunakan *10-fold cross validation ('cv')*. Sintaks model berfungsi untuk mencetak hasil model klasifikasi yang berisi nilai akurasi dan kappa, dapat dilihat pada Gambar 5. Selain itu juga terdapat sintaks *predict* yang dapat menampilkan beberapa kelas yang dapat diprediksi berdasarkan peluang *posterior*.

```

Naïve Bayes
365 samples
 5 predictor
 2 classes: 'kebakaran', 'tidak kebakaran'

No pre-processing
Resampling: cross-validated (10 fold)
Summary of sample sizes: 328, 328, 328, 328, 330, ...
Resampling results:

Accuracy   Kappa
0.9945195  0.9882522

Tuning parameter 'fL' was held constant at a value of 1
    
```

Gambar 5. Hasil akurasi algoritma *Naïve Bayes*

C. Perhitungan Akurasi

Setelah mendapatkan metode klasifikasi pada algoritma *Naïve Bayes*, proses selanjutnya adalah melakukan perhitungan akurasi. Perhitungan akurasi dapat dilakukan dengan menggunakan *confusion matriks* yang dapat dilihat pada Tabel X.

TABEL X
CONFUSION MATRIX UNTUK KLASIFIKASI DUA KELAS

Actual	Predicted	
	Kebakaran	Tidak kebakaran
Kebakaran	TN	FP
Tidak kebakaran	FN	TP

Secara umum, evaluasi dapat dirumuskan sebagai berikut :

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

Keterangan :

- TN = nilai *true negatives*
- TP = nilai *true positives*
- FP = nilai *false positives*
- FN = nilai *false negatives*

Berdasarkan sintaks program baris ke empat belas, TN adalah jumlah kebakaran yang berhasil diprediksi sebagai kebakaran. TP adalah jumlah tidak kebakaran yang berhasil diprediksi tidak kebakaran. FP adalah jumlah kebakaran yang diprediksi tidak kebakaran. Sedangkan, FN menyatakan tidak kebakaran yang diprediksi kebakaran. Akurasi terbaik untuk wilayah kabupaten pelalawan adalah sebesar 100%. Untuk dataset pelalawan selama 4 tahun (lihat Tabel XI) terdapat 528 TP sedangkan 933 TN dan tidak ada data yang salah dalam proses klasifikasi (tidak ada FP dan FN).

TABEL XI
AKURASI MODEL PADA DATASET

Dataset	TP	TN	FP	FN	Akurasi
2015	137	228	0	0	100%
2016	134	232	0	0	100%
2017	108	257	0	0	100%
2018	149	216	0	0	100%

Akurasi model dan kappa statistik yang digunakan untuk menentukan dataset tertinggi. Berdasarkan Tabel XII dapat disimpulkan bahwa nilai akurasi dan kappa tertinggi adalah dataset tahun 2015 dengan nilai akurasi 99.45% dan kappa 98.82%.

TABEL XII
AKURASI DAN KAPPA UNTUK MODEL KLASIFIKASI PADA DATASET

Dataset	Akurasi	Kappa
2015	99.45%	98.82%
2016	98.64%	97.07%
2017	99.44%	98.61%
2018	99.17%	98.27%

D. Penerapan Model dengan Data Baru

Pada tahapan ini akan dilakukan validasi model terhadap dataset dari tahun 2015-2018. Kemudian diterapkan dengan model akurasi tertinggi pada dataset baru tahun 2019. Model dengan akurasi tertinggi terdapat pada tahun 2015. Oleh karena itu, data yang digunakan sebagai data latih merupakan dataset pada tahun 2015, sedangkan data yang digunakan sebagai data uji adalah dataset pada tahun 2019. Sintaks untuk menerapkan model pada data baru tahun 2019 dapat dilihat pada Gambar 6.

Package yang digunakan merupakan *e1071*, package ini membangun model menggunakan fungsi *Naïve Bayes* yang terdapat pada baris kesepuluh pada program dengan pembagian data dilakukan secara manual. Untuk membangun *matriks confusion* diperlukan package tambahan yaitu *gmodel*. Nilai akurasi dan kappa pada dataset baru 2019 adalah 99.52% dan 99% (lihat Gambar 7).


```

1 library(readxl)
2 data=read.delim("clipboard")
3 view(data)
4 library(e1071)
5 library(caret)
6 data_train=data[1:365, ]
7 data_test=data[365:1826, ]
8 prop.table(table(data_train$kelas))
9 prop.table(table(data_test$kelas))
10 data_classifier<-naiveBayes(data_train,c1<-data_train[1:365,5])
11 data_prediction<-predict(data_classifier,data_test)
12 library(gmodels)
13 crossTable(data_prediction,data_test$kelas,
14             prop.chisq = FALSE,prop.t = FALSE,
15             dnn = c("predicted","actual"))
16 hasil=confusionMatrix(table(data_prediction,data_test$kelas))
17 hasil
    
```

Gambar 6. Penerapan model dengan data baru

```

> hasil
Confusion Matrix and Statistics

data_prediction FALSE TRUE
FALSE 876 4
TRUE 3 579

Accuracy : 0.9952
95% CI : (0.9902, 0.9981)
No Information Rate : 0.6012
P-value [Acc > NIR] : <2e-16

Kappa : 0.99
    
```

Gambar 7. Nilai akurasi dan kappa pada dataset baru 2019

Pada Tabel XIII menunjukkan *matriks confusion* untuk model klasifikasi data baru terdapat 876 data kebakaran(F) yang diklasifikasikan sebagai data kebakaran (F) sedangkan terdapat 4 data tidak kebakaran(T) yang diklasifikasikan sebagai data kebakaran(F) terdapat 3 data tidak kebakaran(T) yang diklasifikasikan sebagai data kebakaran(F) dan terdapat 579 data tidak kebakaran(T) yang diklasifikasikan sebagai data tidak kebakaran(T). Dari hasil ini dapat disimpulkan bahwa kinerja *Naive Bayes* terhadap kebakaran hutan di Kabupaten Pelalawan sangat baik.

TABEL XIII
MARIKS CONFUSION UNTUK MODEL PADA DATASET

predictive/actual	False (kebakaran)	True (tidak kebakaran)
False (kebakaran)	876	4
True (tidak kebakaran)	3	579

IV. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan maka dapat diambil kesimpulan sebagai berikut:

- Penelitian ini berhasil mengklasifikasikan kemunculan titik panas kebakaran hutan di kabupaten pelalawan pada tahun 2015 hingga 2019 menggunakan algoritma *Naive Bayes*.
- Dataset yang digunakan untuk klasifikasi terdiri dari suhu, kelembaban, curah hujan, kecepatan angin, dan kelas yang menghasilkan nilai akurasi tertinggi adalah dataset tahun 2015 dengan nilai akurasi adalah **99.45%**. Sedangkan metode klasifikasi dengan akurasi tertinggi diterapkan pada

dataset baru yaitu pada tahun 2019 dengan nilai akurasinya adalah **99.52%**.

- Berdasarkan perhitungan data mining menggunakan algoritma *Naive Bayes*, dapat ditarik kesimpulan bahwa kelas klasifikasi “*True*” / kebakaran hutan lebih besar daripada kelas klasifikasi “*False*” / tidak kebakaran hutan.
- Metode klasifikasi menggunakan algoritma *Naive Bayes* dapat digunakan untuk memprediksi kemunculan titik panas sehingga dapat melakukan tindakan pencegahan sebelum terjadinya kebakaran hutan dan lahan pada masa yang akan datang.
- Penelitian selanjutnya diharapkan dapat membangun suatu *aplikasi* berbasis web untuk menampilkan hasil klasifikasi titik panas menggunakan algoritma *Naive Bayes*.

UCAPAN TERIMA KASIH

Terimakasih peneliti ucapkan kepada Universitas Islam Negeri Sultan Syarif Kasim dan juga kepada pembimbing dalam menyelesaikan penelitian ini, dan begitu juga peneliti sangat berterimakasih dan penghargaan atas diterimanya penelitian ini untuk diterbitkan pada publikasi JUSTIN : Jurnal Sistem dan Teknologi Informasi di Universitas Tanjungpura.

DAFTAR PUSTAKA

- [1] Andi Nugroho, A., Iwan, I., Iroh Nur Azizah, K., & Hakim Raswa, F. (2019). Peatland Forest Fire Prevention Using Wireless Sensor Network Based on Naive Bayes Classifier. *KnE Social Sciences*, 2019, 20–34. <https://doi.org/10.18502/kss.v3i23.5134>
- [2] Asfihan, A. (2019). *Decision Tree Adalah: Jenis, Manfaat, Kelebihan dan Kekurangannya*.
- [3] BNPB. (2019). Rekapitulasi Luas Kebakaran Hutan dan Lahan (Ha) Per Provinsi Di Indonesia Tahun 2011-2018. *SiPongi - Karhutla Monitoring Sistem*, (1), 26–27. <https://doi.org/10.1103/PhysRevB.65.020504>
- [4] Bustami. (2014). *PENERAPAN ALGORITMA NAIVE BAYES*. 8(1).
- [5] Cahya. (2018). *Implementasi Data Mining Algoritma k-Nearest Neighbors (k-NN) menggunakan PHP dan MySQL untuk memprediksi kelulusan mahasiswa tepat waktu*.
- [6] Dewi, S. (2016). Komparasi 5 Metode Algoritma Klasifikasi Data Mining Pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan. *None*, 13(1), 60–66.
- [7] Itsnaini, N., Sasmito, B., Sukmono, A., & Prasasti, I. (2017). Analisis Hubungan Curah Hujan Dan Parameter Sistem Peringkat Bahaya Kebakaran (Spbk) Dengan Kejadian Kebakaran Hutan Dan Lahan Untuk Menentukan Nilai Ambang Batas Kebakaran. *Jurnal Geodesi Undip*, 6(2), 62–70.
- [8] KLHK. (2019). *Kementerian Lingkungan Hidup dan Kehutanan*.
- [9] Kompas.com. (2019). 4 Upaya Pemerintah Padamkan Kebakaran Hutan. Retrieved from <https://www.kompas.com/tren/read/2019/09/17/213000265/4-upaya-pemerintah-padamkan-kebakaran-hutan?page=all#page3>. website: <https://www.kompas.com/tren/read/2019/09/17/213000265/4-upaya-pemerintah-padamkan-kebakaran-hutan?page=all#page3>.
- [10] LAPAN. (2019). Lembaga Penerbangan dan Antariksa Nasional. 2016. Informasi Titik Panas (Hotspot) Kebakaran Hutan/Lahan. Retrieved from <https://modis-catalog.lapan.go.id/monitoring/>
- [11] Laroussi, H. M. M. A. (2015). *Implementasi algoritma naive bayes sebagai proses seleksi penerima beasiswa libyan embassy berbasis web*.
- [12] Maulida, L. (2018). Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov . Dki Jakarta Dengan K-Means. *JISKa*, 2(3), 167–174.
- [13] Menhut. (2016). *Peraturan Menteri Kehutanan Republik Indonesia Nomor P.32 Tahun 2016 tentang Pengendalian*

- Kebakaran Hutan*. Jakarta (ID): Menteri Kehutanan Republik Indonesia.
- [14] Misfaul, M., Dana, M., Kurniawan, W., & Fitriyah, H. (2018). *Rancang Bangun Sistem Deteksi Titik Kebakaran Dengan Metode Naive Bayes Menggunakan Sensor Suhu dan Sensor Api Berbasis Arduino*. 2(9), 3384–3390.
- [15] Mulyana, E. (2019). Bencana Kabut Asap Akibat Kebakaran Hutan Dan Lahan Serta Pengaruhnya Terhadap Kualitas Udara Di Provinsi Riau Februari – Maret 2014. *Jurnal Sains Dan Teknologi Indonesia*, 16(3), 1–7. <https://doi.org/10.29122/jsti.v16i3.3417>
- [16] Negara, B. S., & Kurniawan, R. (2019). *Riau Forest Fire Prediction using Supervised Machine Learning*. (September 2019).
- [17] Novriansyah, D. (2015). *Konsep Data Mining VS Sistem Pendukung Keputusan*. Yogyakarta: Deepublish.
- [18] Nugroho, P. (2017). *Klasifikasi tingkat kerawanan kebakaran hutan dan lahan menggunakan algoritme naive bayes di pulau pisau, kalimantan tengah*.
- [19] Oktaviana, W. (2017). *“Optimasi basis pengetahuan menggunakan algoritma FP-Growth untuk membangun struktur Bayesian Network Penyakit mata*. 1–5.
- [20] Pattekari, S.A.; Parveen, A. (2012). Prediction system for heart disease using Naïve Bayes. *International Journal of Advanced Computer and Mathematical Sciences*, 3(3), 290–294.
- [21] Pei J, Kamber M, H. J. (2012). *Data Mining: Concepts and Techniques 3rd ed*. Massachusetts (US): Morgan Kaufmann Publisher.
- [22] *Peraturan Menteri Kehutanan Nomor: P. 12/Menhut-Ii/2009 tentang Pengendalian Kebakaran Hutan*. (2009). <https://doi.org/10.18860/ling.v5i1.609>
- [23] Pramesti, D. F., Furqon, M. T., & Dewi, C. (2017). *Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan / Lahan Berdasarkan Persebaran Titik Panas (Hotspot)*. 1(9), 723–732.
- [24] Putrada, A. M., Abdurrohman, M., & Putrada, A. G. (2018). Increasing Smoke Classifier Accuracy using Naïve Bayes Method on Internet of Things. *Kinetik: Game Technology, Information System, Computer Network, Computing, Electronics, and Control*, 4(1), 19–26. <https://doi.org/10.22219/kinetik.v4i1.704>
- [25] Reza Noviansyah, M., Rismawan, T., Marisa Midyanti, D., Sistem Komputer, J., & MIPA Universitas Tanjungpura Jl Hadari Nawawi, F. H. (2018). Penerapan Data Mining Menggunakan Metode K-Nearest Neighbor Untuk Klasifikasi Indeks Cuaca Kebakaran Berdasarkan Data Aws (Automatic Weather Station) (Studi Kasus: Kabupaten Kubu Raya). *Jurnal Coding, Sistem Komputer Untan*, 6(2), 48–56. Retrieved from <http://jurnal.untan.ac.id/index.php/jcskommipa/article/view/26672>
- [26] Rizka, Y. A. (2016). *Klasifikasi kemunculan titik panas pada lahan gambut di sumatera dan kalimantan menggunakan algoritme naive bayes*.
- [27] Roswintiarti, D. O. (2016). *Informasi Titik Panas (Hotspot) Kebakaran Hutan/Lahan. 1*.
- [28] Saleh, A. (2015). *Implementasi Metode Klasifikasi Naive Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga*. 2(3), 207–217.
- [29] Setiawan, A. (2018). Kebijakan Penanganan Kebakaran Hutan dan Lahan di Indonesia. *Political Ecology*, VI(May), 1–13.
- [30] Suntoro, J. (2019). *Data Mining: Algoritma dan Implementasi dengan Pemrograman PHP*. Jakarta: PT ELEX MEDIA KOMPUTINDO.
- [31] Widiastuti, D., Informasi, J. S., & Gunadarma, U. (2012). *ANALISA PERBANDINGAN ALGORITMA SVM, NAIVE BAYES, DAN DECISION TREE DALAM MENGLASIFIKASIKAN SERANGAN (ATTACKS)*. 1–8.
- [32] Witten, Ian H. Frank, E. H. (2011). Data Mining Practical Machine Learning Tools and Techniques. In *Encyclopedia of Ecology, Five-Volume Set* (third). <https://doi.org/10.1016/B978-008045405-4.00153-1>
- [33] Wu, X., Kumar, V., ... Dan, J. H. (2008). *Top 10 algorithms in data mining*. <https://doi.org/10.1007/s10115-007-0114-2>