

Terbit online pada laman : <http://teknosi.fti.unand.ac.id/>

Jurnal Nasional Teknologi dan Sistem Informasi

| ISSN (Print) 2460-3465 | ISSN (Online) 2476-8812 |



Studi Kasus

Implementasi Forward Selection dan Bagging untuk Prediksi Kebakaran Hutan Menggunakan Algoritma Naïve Bayes

Fitriyani

Universitas ARS, Jl. Sekolah Internasional 1-2, Bandung, 40282, Indonesia

INFORMASI ARTIKEL

Sejarah Artikel:

Diterima Redaksi: 12 November 2021

Revisi Akhir: 13 Maret 2022

Diterbitkan Online: 02 Mei 2022

KATA KUNCI

Prediksi Kebakaran Hutan,
Seleksi Fitur,
Forward Selection,
Bagging,
Naïve Bayes

KORESPONDENSI

E-mail: fitriyani@ars.ac.id

A B S T R A C T

Kebakaran hutan tidak hanya menimbulkan kerusakan ekonomi dan ekologi, akan tetapi juga mengancam kehidupan manusia dengan pencemaran udara karena asap yang ditimbulkan. Tingginya angka kejadian kebakaran hutan menentukan pentingnya prediksi dilakukan. Algerian Forest Fire merupakan dataset kebakaran hutan yang digunakan dalam penelitian ini, dimana dataset ini akan diolah dengan model yang diusulkan. Dataset ini memiliki fitur-fitur yang tidak relevan dan akan mempengaruhi terhadap kinerja dari model yang diusulkan, sehingga pemilihan fitur yang relevan menggunakan Forward Selection. Metode Bagging digunakan untuk menangani ketidakseimbangan kelas yang ada pada dataset ini dan algoritma Naïve Bayes sebagai algoritma machine learning yang diimplementasikan dalam penelitian ini. Hasil akurasi terbaik adalah sebesar 98,40% pada model Naive Bayes, Bagging dan Greedy Forward Selection dan 92,63% pada model Naïve Bayes dan Bagging.

1. PENDAHULUAN

Kebakaran hutan merupakan bencana yang menimbulkan kerusakan ekonomi dan ekologi serta mengancam kehidupan manusia, oleh karena itu memprediksi masalah lingkungan yang kritis sangat penting untuk mengurangi ancaman ini [1]. Dampak kebakaran hutan seperti mematikan pohon lebih dari 80%, tidak ada pohon yang mampu bertahan pasca kebakaran hutan di rawa gambut, krisis air, pencemaran udara yang dapat menyebabkan hipoksia (kekurangan oksigen akibat kabut asap), dampak ekonomi yang tak terhitung. Kebakaran hutan membutuhkan biaya yang sangat besar terukur dari hilangnya nyawa manusia, sumber daya, kerusakan lingkungan dan biaya untuk perbaikan daerah yang terkena bencana [2]. Kemungkinan penyebab terjadinya kebakaran hutan dapat disebabkan oleh manusia seperti pembakaran, merokok, berburu, pembakaran jerami dan lain-lain. Kebakaran hutan juga dapat disebabkan oleh alam

seperti sambaran petir. Kabel listrik, suhu, kelembapan, jenis pohon, jarak dari lahan pertanian, luas area yang terbakar, merupakan faktor yang dapat mempengaruhi terjadinya kebakaran hutan [3]. Tingginya angka kejadian dan daya rusak kebakaran menentukan pentingnya prediksi atau deteksi dini kebakaran hutan [4].

Dataset yang digunakan dalam penelitian ini adalah dataset Algerian Forest Fire, dimana dataset ini berisi dua data wilayah kebakaran yang berbeda yaitu wilayah Bejaia di Timur Laut Aljazair dan wilayah Sidi Bel-abbes di Barat Laut Aljazair. Dataset ini merupakan dataset publik dari UCI Repository. Penelitian menggunakan dataset publik dianjurkan karena sebanyak 64,79% penelitian di dunia menggunakan dataset publik dan sisanya 35,21% menggunakan dataset privat [5]. Penggunaan dataset publik dapat membuat penelitian berulang terbantahkan dan diverifikasi [6].

Klasifikasi menggunakan Naïve Bayes dapat menghasilkan akurasi yang tinggi dan cepat ketika diaplikasikan pada data yang besar serta mudah dalam penerapannya[7][8], selain cepat algoritma ini juga mempunyai keuntungan dapat menangani multiclass dan klasifikasi biner [9]. Klasifikasi Naïve Bayes berdasarkan asumsi bahwa prinsip hipotesis posteriori maksimum untuk mengidentifikasi yang paling mungkin diklasifikasikan [10]. Bagging (Bootstrap Aggregating) merupakan teknik yang dapat meningkatkan klasifikasi dengan kombinasi klasifikasi secara acak pada dataset training [11]. Seleksi fitur dapat mengurangi dimensional pada data untuk meningkatkan kinerja dari mesin pembelajaran [12] dan merupakan langkah yang penting dalam mesin pembelajaran (machine learning) [13]. Forward Selection adalah seleksi fitur yang dimulai dari set fitur kosong dan setiap langkah fitur yang paling efektif memenuhi metric evaluasi akan ditambahkan dalam fitur ini [14].

Penelitian yang menggunakan dataset yang sama adalah penelitian [1], dimana penelitian ini menggunakan algoritma yang berbeda, algoritma yang digunakan adalah Simple Decision Tree (J48), Random Forest (RF), Bagging dan Boosting (AdaboostM1). Hasil terbaik dalam penelitian ini adalah 84.21% pada metode Boosting. Penelitian [12] menggunakan dataset Thoracic Surgery dan algoritma yang digunakan adalah K-Nearest Neighbor (K-NN) dan Forward Selection, hasil terbaik dalam penelitian ini sebesar 85.74%. [10] menggunakan algoritma Naïve Bayes untuk mengklasifikasikan dataset Music Emotion dengan hasil kinerja terbaik sebesar 68%. Penelitian terkait selanjutnya adalah penelitian [14], algoritma yang digunakan adalah K-Nearest Neighbor (KNN), Forward, Genetic Algorithm (GA), Gini Index, Information Gain dan Correlation. Dataset yang diolah adalah Customer Insurance, hasil terbaik sebesar 95.58% pada algoritma GA dan KNN. Pada penelitian [8] algoritma yang digunakan adalah Naïve Bayes dan Bagging, sedangkan dataset yang digunakan adalah Thoracic Surgery. Hasil penelitian ini menunjukkan nilai terbaik sebesar 83.33%. Penelitian ini menggunakan algoritma seleksi fitur Forward Selection untuk memilih fitur yang paling relevan. Metode ensemble yang digunakan untuk menangani ketidakseimbangan kelas adalah Bagging dan algoritma klasifikasi dalam penelitian menggunakan Naïve Bayes. Tujuan penelitian ini adalah untuk memprediksi kebakaran hutan berdasarkan model yang digunakan.

2. METODE

2.1. Dataset

Algerian Forest Fire adalah dataset digunakan dalam penelitian ini, dimana dataset ini dapat diunduh pada situs web UCI Repository. Dataset ini terbagi menjadi 2 wilayah kebakaran yaitu wilayah Timur Laut dan Barat Laut Algeria. Dataset ini mempunyai 244 record dan 14 atribut, dimana masing-masing wilayah mempunyai 122 record. Data kebakaran hutan ini diambil dari Bulan Juni 2012 sampai bulan September 2012 dengan jumlah kelas 'Fire' sebanyak 138 dan jumlah kelas 'Not Fire' sebanyak 106 kelas.

Deskripsi atribut pada dataset Algerian Forest Fire dapat dilihat pada Tabel 1.

Tabel 1. Deskripsi Atribut Algerian Forest Fire

| ATRIBUT | DESKRIPSI |
|-------------|-------------------------------------|
| Day | Hari |
| Month | Bulan (Juni-September) |
| Year | Tahun (2012) |
| Temperature | Suhu (22°C-42°C) |
| RH | Relative Humadity (21%-90%) |
| Ws | Wind Speed dalam km/h (6-29) |
| Rain | Total Day dalam mm (0-16.8) |
| FFMC | Fine Fuel Moisture Code (28.6-92.5) |
| DMC | Duff Moisture Code (1.1-65.9) |
| DC | Drought Code (7-220.4) |
| ISI | Initial Spread Index (0-18.5) |
| BUI | Buildup Index (1.1-68) |
| FWI | Fire weather Index (0-31.1) |
| Classes | Fire, Not Fire |

2.2. Dataset

Pengklasifikasi *Naive Bayes* berpendapat bahwa nilai dari atribut pada kelas tertentu tidak bergantung (*independence*) pada nilai atribut lainnya, pendapat ini dapat disebut *class-conditional independence* sehingga perhitungannya dapat dibuat lebih sederhana dan disebut naif [7]. Naïve Bayes dapat disebut juga dengan Simple Bayes dan Independence Bayes. Algoritma ini dapat memprediksi probabilitas keanggotaan kelas, seperti probabilitas pada data yang diberikan label kelas tertentu [9]. Formula dari teorema bayes:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Keterangan:

$P(H|X)$ = Probabilitas akhir bersyarat (*conditional probability*) suatu hipotesis H terjadi jika diberikan bukti (*evidence*) X terjadi.

$P(X|H)$ = Probabilitas sebuah bukti X terjadi akan mempengaruhi hipotesis H .

$P(H)$ = Probabilitas awal (priori) hipotesis H terjadi tanpa memandang bukti apapun.

$P(X)$ = Probabilitas awal (priori) bukti X terjadi tanpa memandang hipotesis atau bukti yang lain.

Persamaan Naïve Bayes dalam penelitian ini menggunakan Distribusi Gaussian karena dataset Algerian Forest Fire merupakan dataset numerik. Pada Distribusi Gaussian dihitung mean μ dan standar deviasi σ pada semua atribut [11]. Berikut formula dari Distribusi Gaussian:

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2)$$

2.3. Bagging

Teknik *ensemble* merupakan teknik yang sukses untuk menangani *dataset* yang tidak seimbang meskipun tidak secara

khusus dirancang untuk masalah data yang tidak seimbang [13]. Bagging mengadopsi bootstrap sampling untuk mengaplikasikan metode random kedalam data training untuk menghasilkan base learners dengan mengurangi ketergantungan dan perbedaan. Algoritma *bagging* untuk klasifikasi [17]:

Input: Dataset $D = \{(x_i, y_i)\}_{i=1}^n$ (3)
Base learning algorithm \mathcal{L}
 The number of iterations T
 1. for $t=1$ to T do
 2. $h_t = \mathcal{L}(D, \mathcal{D}_{bs})$ /*
 \mathcal{D}_{bs} is the bootstrap distribution */
 3. end for
Output: $H(x) = \max_y \sum_{t=1}^n I(h_t(x) = y)$
 /* $I(x)=1$ if x is true, and 0 otherwise */

2.4. Seleksi Fitur

Pemilihan fitur merupakan cara yang efektif untuk memecahkan masalah dengan menghapus fitur yang tidak relevan sehingga dapat mengurangi waktu komputasi, meningkatkan akurasi dan memahami model pembelajaran dengan baik [18]. *Subset selection* merupakan metode *feature selection* (seleksi fitur), *subset selection* adalah menemukan *subset* terbaik (fitur), *subset* yang terbaik mempunyai jumlah dimensi yang paling berkontribusi pada akurasi [16]. Metode ini akan menghilangkan sisanya atau dimensi yang tidak penting menggunakan *error function* dan ini dapat menyelesaikan permasalahan pada regresi dan klasifikasi. Pemecahan masalah tersebut dapat diselesaikan dengan menggunakan metode *heuristic* untuk hasil yang masuk akal tapi bukan merupakan solusi yang optimal tapi masuk akal dalam waktu [19].

2.5. Forward Selection

Pada Forward Selection, dimulai dengan tidak ada fitur: $F = \emptyset$. Pada setiap langkah, untuk semua kemungkinan x_i , melihat model pada set pelatihan dan menghitung $E(F \cup x_i)$ pada set validasi. Kemudian pilih input x_j yang menyebabkan kesalahan paling kecil [19].

$$j = \operatorname{argmin} E(F \cup x_i) \tag{4}$$

and we
 add x_j to F if $E(F \cup x_j) < E(F)$

Prosedur Forward Selection [20]:

1. Forward Selection dimulai pada saat tidak ada variabel pada model.
2. Untuk variabel pertama yang masuk ke model, pilih prediktor yang paling berkorelasi tinggi dengan target. (tanpa kehilangan keumuman, menunjukkan variabel ini) Jika hasilnya model tidak signifikan, berhenti dan laporkan bahwa tidak ada variable mungkin F dan F . Pilih variabel dengan sekuensial terbesar F-statistik.

Untuk variabel yang dipilih pada langkah 2, uji untuk signifikansi sekuensial F-statistik. Jika model yang dihasilkan tidak signifikan, hentikan, dan laporkan arus model tanpa menambahkan variabel dari langkah 2. Jika tidak, tambahkan

<https://doi.org/10.25077/TEKNOSI.v8i1.2021.1-8>

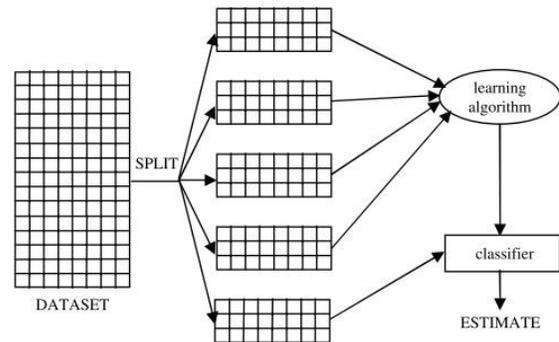
variabel dari langkah 2 ke dalam model dan kembali ke langkah 2.

2.6. K-Fold Cross Validation

Cross validation digunakan ketika tujuan dari machine learning adalah prediksi dan sangat penting untuk mengevaluasi kinerja dari machine learning [21]. Berikut persamaan dari k-fold cross validation [19]:

$$\begin{aligned} v_1 &= x_1 \mathcal{T}_1 = x_2 \cup x_3 \cup \dots \cup x_k \\ v_2 &= x_2 \mathcal{T}_2 = x_1 \cup x_3 \cup \dots \cup x_k \\ &\dots \\ v_k &= x_k \mathcal{T}_k = x_1 \cup x_2 \cup \dots \cup x_{k-1} \end{aligned} \tag{5}$$

Jika dataset terdiri dari jumlah N , dibagi menjadi k bagian yang sama, k biasanya berupa bilangan kecil seperti 5 atau 10. (Jika N tidak habis dibagi k , bagian akhir akan memiliki lebih sedikit jumlah daripada $k - 1$ lainnya). Serangkaian k dijalankan, masing-masing bagian k secara bergantian digunakan sebagai set pengujian dan bagian $k - 1$ lainnya digunakan sebagai set pelatihan [22]. Gambar 1 merupakan k-fold cross validation.



Gambar 1. K-Fold Cross Validation [22]

2.7. Confusion Matrix

Confusion matrix, confusion matrix merupakan alat untuk menganalisa seberapa baik kinerja dari pengklasifikasi dapat mengenali tupel dari kelas yang berbeda [7]. Confusion matrix merupakan matrik 2 dimensi yang menggambarkan perbandingan antara hasil prediksi dengan kenyataan [11]. Berikut adalah persamaan model confusion matrix:

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \tag{6}$$

$$\text{Sensitivity} = \text{Recall} = TP_{rate} = \frac{TP}{TP + FN}$$

$$\text{Specificity} = TN_{rate} = \frac{TN}{TN + FP}$$

$$FP_{rate} = \frac{FP}{FP + TN}$$

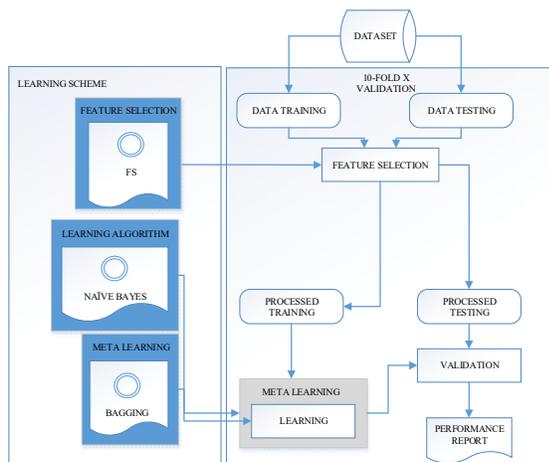
$$\text{Precision} = PPV = \frac{TP}{TP + FP}$$

Keterangan:

TP=True Positive
TN=True Negative
FP=False Positive
FN=False Negative

2.8. Kerangka Penelitian

Pada kerangka penelitian ini dataset akan dibagi menjadi 10 bagian menggunakan 10 fold x validation atau 10 fold cross validation yang terdiri dari dataset training dan testing. Tahapan selanjutnya dataset akan diolah menggunakan seleksi fitur Forward Selection, seleksi fitur ini digunakan untuk menemukan fitur atau atribut yang paling relevan, kemudian data akan diolah menggunakan metode ensemble Bagging untuk menangani ketidakseimbangan kelas karena jumlah kelas fire dan not fire tidak seimbang. Metode Bagging akan melakukan iterasi sebanyak yang ditentukan dengan algoritma machine learning Naïve Bayes. Setelah proses klasifikasi selesai maka akan keluar hasil dari kinerja pada model yang digunakan. Gambar 2 merupakan kerangka penelitian yang dilakukan.



Gambar 2. Kerangka Penelitian

2.9. Tahapan Penelitian

Tahapan penelitian yang dilakukan dalam penelitian ini dapat dilihat pada Gambar 3.



Gambar 3. Tahapan Penelitian

Pada Gambar 3, tahapan yang dilakukan dalam penelitian ini sebagai berikut:

1. Pengumpulan Data
Dataset Algerian Forest Fire dapat diunduh pada situs web UCI Repository di link berikut ini: <https://archive.ics.uci.edu/ml/machine-learning-databases/00547/>.
2. Pengolahan Data Awal
Pengolahan data awal yang dilakukan dalam penelitian adalah menghilangkan missing value, sehingga data tidak mengandung noise dan dapat diolah.
3. Model yang diusulkan
Model yang diusulkan dalam penelitian ini menggunakan algoritma klasifikasi Naïve Bayes dan seleksi fitur Forward Selection serta metode ensemble Bagging.
4. Eksperimen dan Pengujian Model
Eksperimen dan pengujian dilakukan dengan beberapa tahapan sebagai berikut:
 - a. Menyiapkan dataset untuk eksperimen.
 - b. Mendesain arsitektur Naïve Bayes.
 - c. Melakukan training dan testing terhadap model Naïve Bayes dan mencatat hasil kinerja dari model.
 - d. Mendesain arsitektur Naïve Bayes dan Bagging.
 - e. Melakukan training dan testing terhadap model Naïve Bayes dan Bagging, kemudian mencatat hasil kinerja dari model.
 - f. Mendesain arsitektur Naïve Bayes, Bagging dan Forward Selection.
 - g. Melakukan training dan testing terhadap model Naïve Bayes, Forward Selection, dan Bagging, kemudian mencatat hasil kinerja dari model.

Spesifikasi komputer yang digunakan dalam eksperimen ini dapat ditunjukkan pada Tabel 2.

Tabel 2. Spesifikasi Komputer

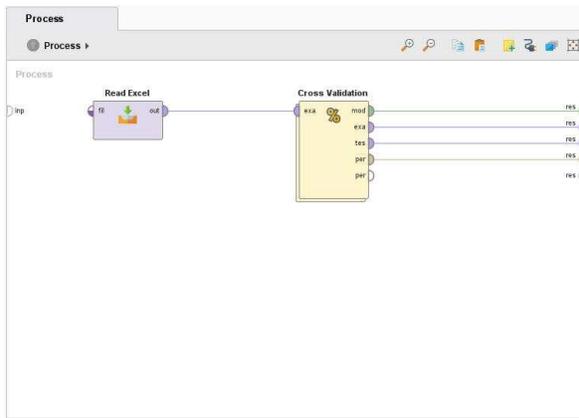
| | |
|----------------|-----------------------|
| Processor | AMD A9-9420 RADEON R5 |
| RAM | 4.00GB |
| Sistem Operasi | Windows 10 |
| Aplikasi | Rapidminer |

5. Evaluasi dan Validasi Hasil
Evaluasi dan validasi hasil yang digunakan dalam penelitian ini adalah nilai akurasi, nilai akurasi akan muncul setelah proses eksperimen selesai dilakukan. Nilai akurasi ini dapat dilihat seberapa baik atau bagus model yang digunakan.

3. HASIL

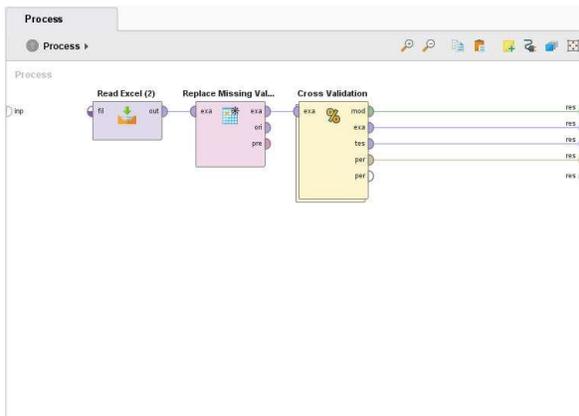
3.1. Eksperimen Naïve Bayes

Eksperimen Naïve Bayes diawali dengan pemanggilan dataset menggunakan operator Read Excel, kemudian dibagi dataset menjadi data training dan data testing menggunakan operator Cross Validation. Tampilan ini dapat dilihat pada Gambar 4 dan dataset yang digunakan adalah dataset Bejaia.



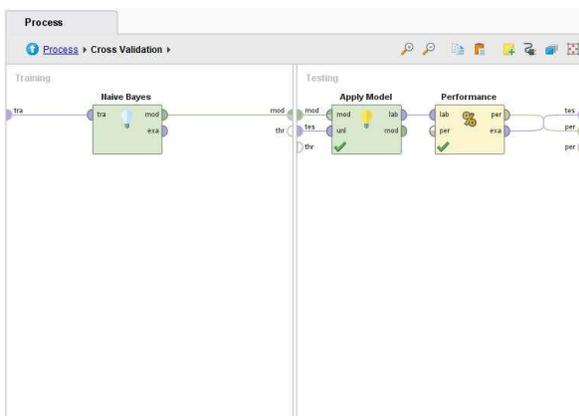
Gambar 4. Dataset dengan Cross Validation

Gambar 5 menunjukkan operator Replace Missing value, operator ini digunakan untuk menghilangkan missing value pada dataset Sidi, selanjutnya dataset diolah menggunakan operator Cross Validation.



Gambar 5. Dataset dengan Operator Missing Value

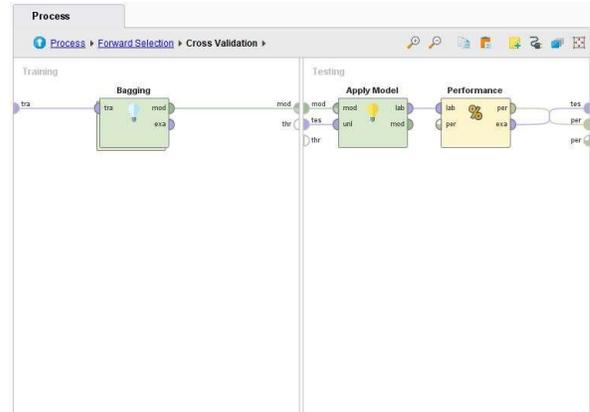
Gambar 6 merupakan proses algoritma Naïve Bayes mengolah dataset dan akan menghasilkan kinerja dengan menggunakan operator Performance.



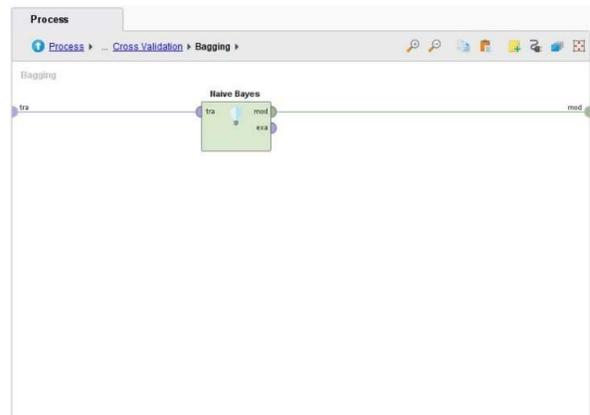
Gambar 6. Pengolahan Dataset dengan Naïve Bayes

3.2. Eksperimen Naïve Bayes dan Bagging

Pada Gambar 7 dapat dilihat ada perbedaan pada eksperimen sebelumnya, dimana sebelum operator Apply Model ada operator Bagging. Operator ini akan melakukan iterasi berdasarkan algoritma klasifikasi yang digunakan. Gambar 8 disisipkan operator Naïve Bayes sebagai algoritma klasifikasi.



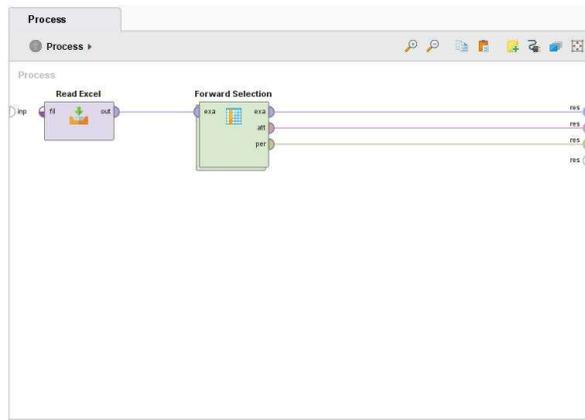
Gambar 7. Bagging pada Naïve Bayes



Gambar 8. Algoritma Naïve Bayes

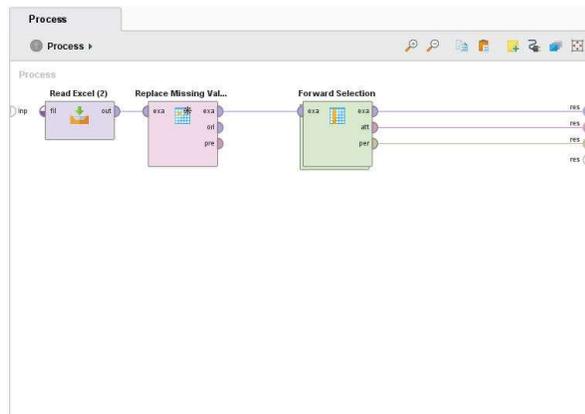
3.3. Eksperimen Naïve Bayes, Bagging dan Forward Selection

Gambar 9 merupakan tampilan proses eksperimen yang dilakukan pada aplikasi Rapidminer, dimana operator Read Excel digunakan untuk memanggil dataset Algerian Forest Fire untuk dataset Bejaia, sedangkan operator Forward Selection merupakan operator yang digunakan untuk menyeleksi fitur yang paling relevan pada dataset Algerian Forest Fire.

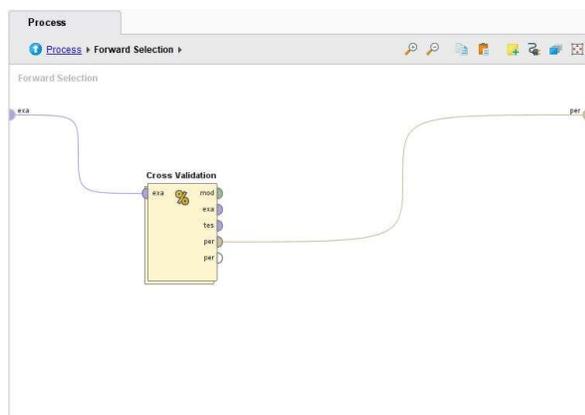


Gambar 9. Seleksi Fitur Dataset Bejaia

Dataset Sidi Bel-abbes pada Algerian Forest Fire mengandung missing value, maka ada perbedaan sedikit pada tampilan eksperimen di Rapidminer. Penambahan operator Replace Missing Value dibutuhkan untuk menghilangkan missing value pada dataset sehingga dataset dapat diolah. Gambar 10 merupakan proses seleksi fitur menggunakan dataset Sidi. Gambar 11 adalah tampilan cross validation, cross validation digunakan untuk mendapatkan hasil kinerja yang lebih akurat dengan membagi dataset menjadi 10 bagian.



Gambar 10. Seleksi Fitur Dataset Sidi Bel-abbes

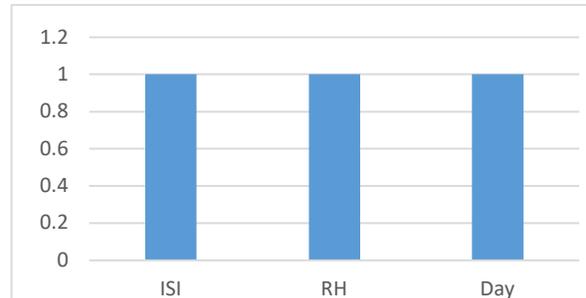


Gambar 11. Cross Validation

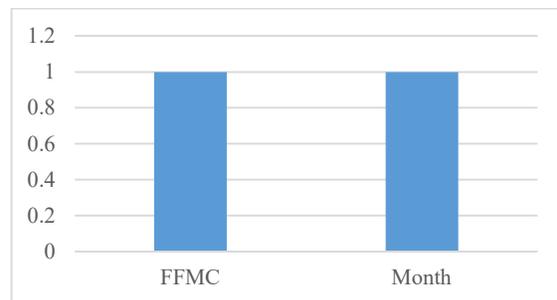
4. PEMBAHASAN

4.1. Hasil Seleksi Fitur Forward Selection

Seleksi fitur Forward Selection yang digunakan dalam penelitian ini adalah untuk memilih fitur atau atribut yang paling relevan, dimana seleksi fitur ini dapat meningkatkan kinerja dari model yang digunakan. Hasil seleksi fitur Forward Selection menggunakan dataset Bejaia adalah fitur ISI, RH dan Day yang dapat dilihat pada Gambar 12. Sedangkan hasil seleksi fitur Forward Selection pada dataset Sidi Bel-abbes adalah fitur FFMC dan Month yang dapat dilihat pada Gambar 13.



Gambar 12. Fitur Forward Selection Dataset Bejaia



Gambar 13. Fitur Forward Selection Dataset Sidi Bel-abbes

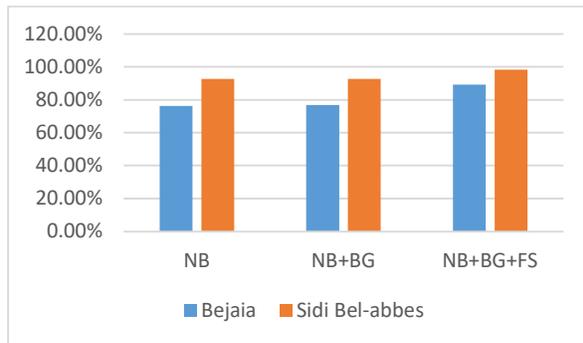
4.2. Hasil Naïve Bayes, Bagging dan Forward Selection

Berdasarkan eksperimen yang dilakukan dapat diketahui bahwa nilai akurasi terbaik ada pada dataset Sidi Bel-abbes untuk model Naïve Bayes (NB) sebesar 92.63%. Sedangkan model Naïve Bayes dan Bagging (NB+BG) sebesar 92.63% dan model Naïve Bayes, Bagging dan Forward Selection (NB+BG+FS) sebesar 98.40%. Sedangkan pada dataset Bejaia, nilai akurasi yang didapat pada model Naïve Bayes sebesar 76.15%, sedangkan model Naïve Bayes dan Bagging sebesar 76.92% dan model Naïve Bayes, Bagging dan Forward Selection sebesar 89.23%. Tabel 3 dapat dilihat perbandingan nilai akurasi pada setiap dataset.

Tabel 3. Perbandingan Nilai Akurasi

| Model | Bejaia | Sidi Bel-abbes |
|----------|--------|----------------|
| NB | 76.15% | 92.63% |
| NB+BG | 76.92% | 92.63% |
| NB+BG+FS | 89.23% | 98.40% |

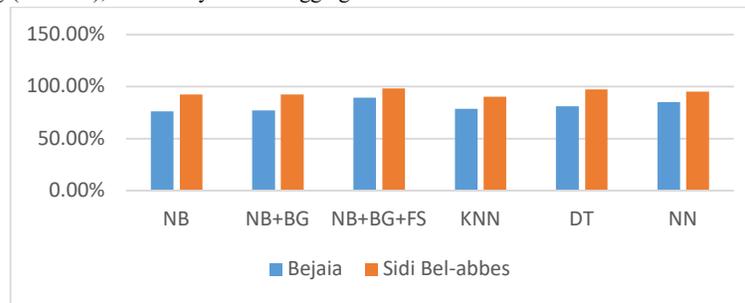
Gambar 14 dapat dilihat perbandingan nilai akurasi menggunakan grafik sehingga sangat jelas terlihat perbedaannya.



Gambar 14. Grafik Perbandingan Nilai Akurasi

4.3. Perbandingan Hasil Penelitian

Perbandingan hasil penelitian algoritma Naïve Bayes (NB), Naïve Bayes dan Bagging (NB+BG), Naïve Bayes dan Bagging



Gambar 15. Grafik Perbandingan Akurasi dengan Algoritma Lain

Perbandingan hasil penelitian ini dengan penelitian sebelumnya dapat dilihat pada Tabel 5, dimana penelitian [1] hasil terbaik ada pada algoritma Boosting sebesar 84.21%. Penelitian [12] hasil terbaik adalah 85.74% pada algoritma K-Nearest Neighbor dan Forward Selection (KNN+FS). Hasil akurasi 68% pada algoritma Naïve Bayes dalam penelitian [10]. Genetic Algorithm dan K-Nearest Neighbor (GA+KNN) mendapatkan nilai akurasi terbaik sebesar 95.58% dalam penelitian [14]. Pada penelitian [8] nilai akurasi terbaik ada pada algoritma Naïve Bayes dan Bagging (NB+BG) sebesar 83.33%. Gambar 16 dapat dilihat grafik perbandingan hasil penelitian ini dengan hasil penelitian lain.

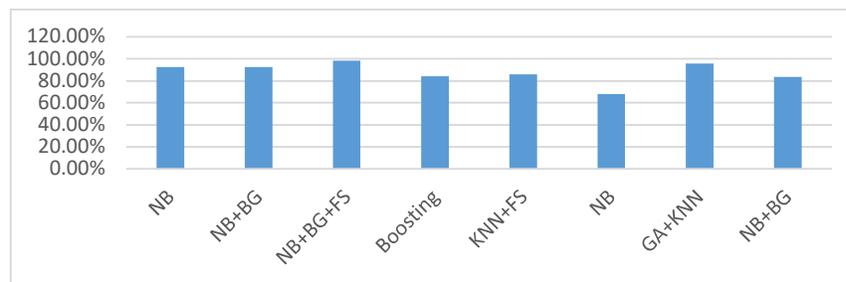
serta Forward Selection (NB+BG+FS), K-Nearest Neighbor (KNN), Decision Tree (DT) dan Neural Network (NN). Perbandingan hasil kinerja akurasi untuk dataset Bejaia dan dataset Sidi Bel-abbes dapat dilihat pada Tabel 4. Gambar 15 merupakan grafik perbandingan hasil akurasi dataset Bejaia dan dataset Sidi Bel-abbes dengan algoritma lain.

Tabel 4. Perbandingan Nilai Akurasi dengan Algoritma Lain

| Model | Bejaia | Sidi Bel-abbes |
|----------------|--------|----------------|
| NB | 76.15% | 92.63% |
| NB+BG | 76.92% | 92.63% |
| NB+BG+FS | 89.23% | 98.40% |
| KNN | 78.65% | 90.13% |
| Decision Tree | 81.09% | 97.56% |
| Neural Network | 85.13% | 95.13% |

Tabel 5. Perbandingan Hasil Penelitian dengan Penelitian Lain

| Model | Akurasi |
|----------|---------|
| NB | 92.63% |
| NB+BG | 92.63% |
| NB+BG+FS | 98.40% |
| Boosting | 84.21% |
| KNN+FS | 85.74% |
| NB | 68% |
| GA+KNN | 95.58% |
| NB+BG | 83.33% |



Gambar 15. Grafik Perbandingan Akurasi dengan Algoritma Lain

5. KESIMPULAN

Kebakaran hutan mengancam kehidupan manusia dan berdampak pada ekonomi dan ekologi kelangsungan hidup sehingga penting

untuk dapat memprediksi hal ini. Dataset Algerian Forest Fire memiliki fitur-fitur yang tidak relevan, hal ini dapat membuat kinerja dari model yang diusulkan menurun. Seleksi fitur Forward Selection dapat menangani permasalahan fitur yang tidak relevan dan meningkatkan kinerja dari model. Metode Bagging

digunakan untuk menangani ketidakseimbangan kelas dan algoritma Naïve Bayes merupakan algoritma klasifikasi yang sederhana dan mudah diimplementasikan. Hasil dari eksperimen yang dilakukan bahwa nilai akurasi terbaik yaitu 98.40% dari model Naïve Bayes, Bagging dan Forward Selection pada dataset Sidi Bel-abbes, sedangkan dataset Bejaia hanya sebesar 89.23%. Model Naïve Bayes dan Bagging pada dataset Sidi Bel-abbes sebesar 92.63%, sedangkan pada dataset Bejaia sebesar 76.92%. Pada dataset Sidi Bel-abbes ketika diimplementasikan model Naïve Bayes, hasil akurasi yang muncul adalah sama dengan model Naïve Bayes dan Bagging yaitu sebesar 92.63%. Sedangkan, pada dataset Bejaia untuk model Naïve Bayes adalah sebesar 76.15%. Hasil eksperimen dalam penelitian ini dapat disimpulkan bahwa seleksi fitur Forward Selection dan metode Bagging dapat meningkatkan kinerja dari algoritma yang digunakan sehingga untuk memprediksi kebakaran hutan, model ini menghasilkan nilai akurasi yang sangat baik. Penelitian selanjutnya dapat menggunakan seleksi fitur lain seperti Backward Selection, Greedy atau menggunakan metode ensemble lain seperti Adaboost untuk memprediksi kebakaran hutan atau lainnya.

UCAPAN TERIMA KASIH

Peneliti mengucapkan terimakasih kepada Faroudja ABID atas donasi dataset Algerian Forest Fire di UCI Repository sehingga dataset ini dapat digunakan dalam penelitian ini.

DAFTAR PUSTAKA

- [1] F. Abid and N. Izeboudjen, "Predicting Forest Fire in Algeria Using Data Mining Techniques: Case Study of the Decision Tree Algorithm," *Adv. Intell. Syst. Comput.*, vol. 1105 AISC, no. February, pp. 363–370, 2020.
- [2] A. Alonso-Betanzos *et al.*, "An intelligent system for forest fire risk prediction and fire fighting management in Galicia," *Expert Syst. Appl.*, vol. 25, no. 4, pp. 545–554, 2003.
- [3] V. Sevinc, O. Kucuk, and M. Goltas, "A Bayesian network model for prediction and analysis of possible forest fire causes," *For. Ecol. Manage.*, vol. 457, no. June, p. 117723, 2020.
- [4] H. Lin, X. Liu, X. Wang, and Y. Liu, "A fuzzy inference and big data analysis algorithm for the prediction of forest fire based on rechargeable wireless sensor networks," *Sustain. Comput. Informatics Syst.*, vol. 18, pp. 101–111, 2018.
- [5] R. S. Wahono, "A Systematic Literature Review of Software Defect Prediction : Research Trends , Datasets , Methods and Frameworks," *J. Softw. Eng.*, vol. 1, no. 1, 2015.
- [6] C. Catal and B. Diri, "A systematic review of software fault prediction studies," *Expert Syst. Appl.*, vol. 36, no. 4, pp. 7346–7354, 2009.
- [7] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*. 2012.
- [8] F. Fitriyani, "Metode Bagging Untuk Imbalance Class Pada Bedah Toraks Menggunakan Naive Bayes," *J. Kaji. Ilm.*, vol. 18, no. 3, p. 278, 2018.
- [9] F. Fitriyani, "Prediksi Diabetes Menggunakan Algoritma Naive Bayes dan Greedy Forward Selection," *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 2, pp. 61–69, 2021.
- [10] Y. An, S. Sun, and S. Wang, "Naive Bayes classifiers for music emotion classification based on lyrics," *Proc. - 16th IEEE/ACIS Int. Conf. Comput. Inf. Sci. ICIS 2017*, no. 1, pp. 635–638, 2017.
- [11] Fitriyani and R. S. Wahono, "Integrasi Bagging dan Greedy Forward Selection pada Prediksi Cacat Software dengan Menggunakan Naïve Bayes," *J. Softw. Eng.*, vol. 1, no. 2, pp. 101–108, 2015.
- [12] R. Sanjaya and F. Fitriyani, "Prediksi Bedah Toraks Menggunakan Seleksi Fitur Forward Selection dan K-Nearest Neighbor," *J. Edukasi dan Penelit. Inform.*, vol. 5, no. 3, p. 316, 2019.
- [13] I. H. Laradji, M. Alshayeb, and L. Ghouti, "Software defect prediction using ensemble learning on selected features," *Inf. Softw. Technol.*, vol. 58, pp. 388–402, 2015.
- [14] F. Abdi, K. Khalili-Damghani, and S. Abolmakarem, "Solving customer insurance coverage sales plan problem using a multi-stage data mining approach," *Kybernetes*, vol. 47, no. 1, pp. 2–19, 2018.
- [15] M. M. Saritas and A. Yasar, "Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification," *Int. J. Intell. Syst. Appl. Eng.*, vol. 7, pp. 88–91, 2019.
- [16] F. Fitriyani and T. Arifin, "Implementasi Greedy Forward Selection untuk Prediksi Metode Penyakit Kutil Menggunakan Decision Tree," *JST (Jurnal Sains dan Teknol.)*, vol. 9, no. 1, pp. 76–85, 2020.
- [17] X. Y. Liu and Z. H. Zhou, "Ensemble methods for class imbalance learning," *Imbalanced Learn. Found. Algorithms, Appl.*, pp. 61–82, 2013.
- [18] J. Cai, J. Luo, S. Wang, and S. Yang, "Feature selection in machine learning: A new perspective," *Neurocomputing*, vol. 300, pp. 70–79, 2018.
- [19] Alpaydm Ethem, *Introduction to Machine Learning Second Edition*, 2nd ed. London: MIT, 2010.
- [20] D. T. Larose, *Data Mining Methods and Models*. 2006.
- [21] D. Sumeet and D. Xian, *Data Mining and Machine Learning in Cybersecurity*. 2011.
- [22] M. Bramer, *Principles of Data Mining*, no. February. 2007.

BIODATA PENULIS



Fitriyani Memperoleh gelar S.T pada bidang Sistem Informasi di Universitas BSI Bandung dan gelar M.Kom pada bidang Ilmu Komputer di STMIK Nusa Mandiri. Minat penelitiannya adalah Data Mining, Machine Learning dan Software Engineering. Saat ini Dosen aktif di Universitas ARS.