

**IMPLEMENTASI ALGORITMA *NAÏVE BAYES* UNTUK
KLASIFIKASI KEBAKARAN HUTAN DI INDONESIA**

SKRIPSI

**OLEH:
NIKMATUN AWWALIYAH
NIM. 210601110053**



**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2024**

**IMPLEMENTASI ALGORITMA *NAÏVE BAYES* UNTUK
KLASIFIKASI KEBAKARAN HUTAN DI INDONESIA**

SKRIPSI

**Diajukan Kepada
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang
untuk Memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Matematika (S.Mat)**

**Oleh
Nikmatun Awwaliyah
NIM. 210601110053**

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM MALANG
2024**

IMPLEMENTASI ALGORITMA *NAÏVE BAYES* UNTUK KLASIFIKASI KEBAKARAN HUTAN DI INDONESIA

SKRIPSI

Oleh
Nikmatun Awwaliyah
NIM. 210601110053

Telah Disetujui Untuk Diuji

Malang, 19 November 2024

Dosen Pembimbing I


Hisyam Fahmi, M.Kom
NIP. 19890727 201903 1 018

Dosen Pembimbing II


Ari Kusumastuti, M.Pd., M.Si
NIP. 19770521 200501 2 004

Mengetahui,
Ketua Program Studi Matematika


Dr. Elly Susanti, M.Sc
NIP. 19741129 200012 2 005

IMPLEMENTASI ALGORITMA *NAÏVE BAYES* UNTUK KLASIFIKASI KEBAKARAN HUTAN DI INDONESIA

SKRIPSI

Oleh
Nikmatun Awwaliyah
NIM. 210601110053

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima sebagai Salah Satu Persyaratan
untuk Memperoleh Gelar Sarjana Matematika (S. Mat)

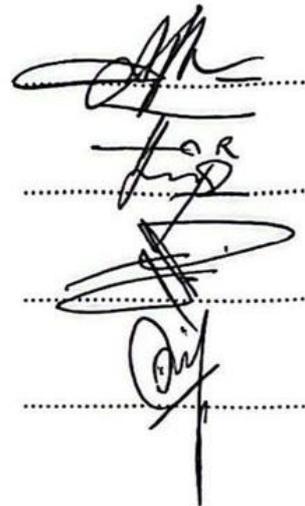
Tanggal 20 Desember 2024

Ketua Penguji : Abdul Aziz, M.Si.

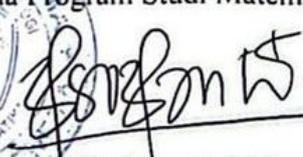
Anggota Penguji 1 : Dr. Fachrur Rozi, M.Si.

Anggota Penguji 2 : Hisyam Fahmi, M.Kom.

Anggota Penguji 3 : Ari Kusumastuti, M.Pd., M.Si.



Mengetahui,
Ketua Program Studi Matematika


Dr. Elly Susanti, M.Sc
NIP. 19741129 200012 2 005

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama : Nikmatun Awwaliyah
NIM : 210601110053
Program Studi : Matematika
Fakultas : Sains dan Teknologi
Judul Skripsi : Implementasi Algoritma *Naïve Bayes* untuk
Klasifikasi Kebakaran Hutan di Indonesia

Menyatakan dengan sebenarnya bahwa skripsi yang saya tulis ini benar-benar merupakan hasil karya sendiri, bukan merupakan pengambilan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan dan pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan skripsi ini hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 20 Desember 2024



Yang membuat pernyataan,


Nikmatun Awwaliyah

NIM. 210601110053

MOTO

"Dan bahwasanya seorang manusia tidak akan memperoleh selain apa yang telah diusahakannya."

(QS. An-Najm: 39)

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur kepada Allah SWT atas segala rahmat-Nya, karya ini penulis persembahkan kepada Ayah tercinta Amiruddin Zaka dan Ibunda tercinta

Aini Maslahah, yang telah berjuang tanpa lelah, memberikan cinta, doa, dan dukungan dalam setiap langkah perjalanan hidup penulis. Adik tersayang Matsna

Abidatun Bahiroh, yang memberikan keceriaan dan semangat di saat-saat sulit.

Serta teman-teman terdekat dan tersayang, yang selalu menjadi penyemangat dan tempat berbagi di setiap tantangan. Karya ini merupakan wujud dari cinta, doa,

dan dukungan yang telah diberikan kepada penulis.

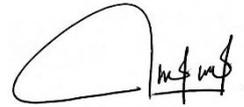
KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan proposal skripsi ini. Skripsi ini disusun sebagai bagian dari pemenuhan beban studi untuk mata kuliah Skripsi pada program studi Matematika di Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Penyusunan skripsi ini tentu tidak terlepas dari dukungan dan bantuan berbagai pihak, oleh karena itu penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan dan motivasi selama proses ini. Ucapan terima kasih kepada:

1. Prof. Dr. H. M. Zainuddin, M.A., selaku rektor Universitas Islam Negeri Maulana Malik Ibrahim.
2. Prof. Dr. Sri Harini, M.Si, selaku dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim.
3. Dr. Elly Susanti, S.Pd., M.Sc, selaku ketua Program Studi Matematika, Universitas Islam Negeri Maulana Malik Ibrahim.
4. Hisyam Fahmi, M.Kom, selaku dosen pembimbing I atas bimbingan, dukungan, dan arahan yang telah diberikan selama proses penyusunan skripsi ini. Semoga ilmu yang Bapak bagikan menjadi amal jariyah dan senantiasa bermanfaat.
5. Ari Kusumastuti, M.Si., M.Pd, selaku dosen pembimbing II atas kesabaran, arahan, serta masukan yang berharga dalam proses penyelesaian skripsi ini. Semoga segala bimbingan yang diberikan menjadi pahala dan berkah bagi Ibu.
6. Abdul Aziz, M.Si., selaku ketua penguji dalam ujian skripsi yang telah memberikan masukan dan saran yang bermanfaat bagi penulis
7. Dr. Fachrur Rozi, M.Si., selaku dosen penguji 1 dalam ujian skripsi yang telah memberikan masukan dan saran yang bermanfaat bagi penulis
8. Seluruh dosen Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim

9. Ayah Amiruddin Zaka, Ibu Aini Maslahah, Adik Matsna Abidatun Bahiroh, serta keluarga besar atas doa, dukungan, dan cinta kasih yang selalu menyertai. Segala bentuk dukungan menjadi kekuatan dalam menyelesaikan skripsi ini. Semoga segala kebaikan yang telah diberikan senantiasa mendapatkan balasan dari Allah SWT.
10. Seluruh mahasiswa angkatan 2021 dan teman teman terdekat penulis atas kebersamaan, dukungan, serta motivasi yang telah diberikan selama masa studi ini. Kehadiran dan semangat kalian menjadi bagian penting dalam perjalanan ini. Semoga kebersamaan ini terus terjalin dan membawa manfaat bagi kita semua.

Malang, 20 Desember 2024



Nikmatun Awwaliyah

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PENGAJUAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iv
PERNYATAAN KEASLIAN TULISAN	v
MOTO	vi
HALAMAN PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR SIMBOL	xiv
ABSTRAK	xvi
ABSTRACT	xvii
مستخلص البحث	xviii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	6
1.3 Tujuan Penelitian	6
1.4 Manfaat Penelitian	6
1.5 Batasan Masalah	7
1.6 Definisi Istilah	7
BAB II KAJIAN TEORI	9
2.1 Teori Pendukung.....	9
2.1.1 Data <i>mining</i>	9
2.1.2 Klasifikasi.....	12
2.1.3 Korelasi Pearson	14
2.1.4 Standardisasi Data	16
2.1.5 <i>Principal Component Analysis</i> (PCA).....	17
2.1.6 <i>Naïve Bayes</i>	18
2.1.7 <i>Gaussian Naïve Bayes</i>	21
2.1.8 <i>Confusion Matrix</i>	22
2.1.9 Kebakaran Hutan	23
2.2 Kajian Integrasi Topik dengan Al-Quran/Hadits.....	26
2.3 Kajian Topik Dengan Teori Pendukung	29
BAB III METODE PENELITIAN	31
3.1 Jenis Penelitian	31
3.2 Data dan Sumber Data	31
3.3 Teknik Analisis Data	33
BAB IV HASIL DAN PEMBAHASAN	38
4.1 <i>Exploratory Data Analysis</i>	38
4.1.1 Deskripsi Statistik.....	38
4.1.2 Uji Korelasi Pearson.....	42

4.2 Data <i>Preprocessing</i>	44
4.2.1 Standardisasi Data	44
4.2.2 <i>Principal Component Analysis (PCA)</i>	48
4.3 Implementasi Algoritma <i>Naïve Bayes</i>	52
4.4 Evaluasi Performa Model <i>Naïve Bayes</i> pada Data <i>Testing</i>	59
4.5 Kajian Keislaman dengan Hasil Penelitian	61
BAB V PENUTUP	63
5.1 Kesimpulan.....	63
5.2 Saran	64
DAFTAR PUSTAKA	65
LAMPIRAN	68
RIWAYAT HIDUP	78

DAFTAR TABEL

Tabel 2.1	<i>Confusion Matrix</i>	23
Tabel 3.1	Fitur- Fitur Data Kebakaran Hutan	32
Tabel 4.1	Deskripsi Statistik pada Data Kebakaran Hutan di Indonesia	40
Tabel 4.2	Nilai <i>P-value</i> Signifikansi Uji Korelasi Pearson	43
Tabel 4.3	Deskripsi Statistik pada Data <i>Training</i>	45
Tabel 4.4	Tampilan Sebelum dan Sesudah Standardisasi	46
Tabel 4.5	Presentase Variansi Komponen dan Variansi Kumulatif.....	49
Tabel 4.6	Nilai <i>P-value</i> Signifikansi Setelah PCA	51
Tabel 4.7	Contoh Data <i>Training</i>	53
Tabel 4.8	Hasil <i>5-fold Cross Validation</i>	58

DAFTAR GAMBAR

Gambar 3.1	Alur Penelitian.....	33
Gambar 4.1	Grafik Deskripsi Statistik.....	41
Gambar 4.2	<i>Heatmap Pearson Correlation Matrix</i>	42
Gambar 4.3	Grafik Mean Sebelum dan Sesudah Standardisasi.....	46
Gambar 4.4	Grafik Std. Deviasi Sebelum dan Sesudah Standardisasi	47
Gambar 4.5	Grafik Variansi Komponen	50
Gambar 4.6	Grafik Variansi Kumulatif	50
Gambar 4.7	Heatmap Korelasi Komponen PCA	51
Gambar 4.8	Jumlah Prediksi Tiap Kelas.....	57
Gambar 4.9	Probabilitas Prediksi Tiap Kelas Beberapa Data <i>Testing</i>	57
Gambar 4.10	<i>Confusion Matrix</i>	60

DAFTAR SIMBOL

μ	: Rata-rata data
σ	: Standar deviasi data
σ^2	: Varians
r_{xy}	: <i>Pearson correlation coefficient</i>

DAFTAR LAMPIRAN

Lampiran 1	Data Mentah Kebakaran Hutan di Indonesia.....	68
Lampiran 2	Data <i>Training</i>	69
Lampiran 3	Data <i>Testing</i>	70
Lampiran 4	Data <i>Training</i> Setelah Standardisasi	71
Lampiran 5	Data <i>Testing</i> Setelah Standardisasi.....	72
Lampiran 6	Data <i>Training</i> Setelah PCA.....	73
Lampiran 7	Data <i>Testing</i> Setelah PCA	74
Lampiran 8	Hasil Prediksi Probabilitas Posterior Data <i>Testing</i>	75
Lampiran 9	<i>Script Code Phyton</i>	76

ABSTRAK

Awwaliyah, Nikmatun. 2024. **Implementasi Algoritma *Naïve Bayes* untuk Klasifikasi Kebakaran Hutan di Indonesia**. Skripsi. Program Studi Matematika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Hisyam Fahmi, M.Kom. (II) Ari Kusumastuti, M.Pd., M.Si

Kata Kunci: *Naïve Bayes*, Klasifikasi, Kebakaran Hutan.

Kebakaran hutan di Indonesia menjadi permasalahan serius yang memerlukan penanganan segera. Berdasarkan data Pusdalops BNPB, sepanjang tahun 2023 tercatat 487 kejadian kebakaran hutan dan lahan di berbagai wilayah Indonesia. Peningkatan frekuensi kejadian ini menuntut adanya metode yang efektif untuk mengidentifikasi dan mengklasifikasikan kebakaran secara akurat, sehingga dapat mendukung upaya mitigasi yang lebih dini. Penelitian ini bertujuan mengevaluasi efektivitas algoritma *Naïve Bayes* dalam klasifikasi kebakaran hutan, dengan menggunakan data sekunder dari BNPB sebagai kelas kebakaran dan Open Meteo sebagai kelas bukan kebakaran. Metode penelitian yang digunakan adalah kuantitatif, dengan pendekatan analisis numerik terhadap faktor-faktor alam seperti suhu, cuaca, kecepatan angin, dan hembusan angin yang berpotensi memicu kebakaran hutan. Hasil penelitian menunjukkan bahwa model *Naïve Bayes* mampu mencapai tingkat akurasi sebesar 93,03% pada data *testing*, dengan precision sebesar 92,96% dan recall sebesar 96,17%. Tingginya nilai metrik performa ini membuktikan bahwa algoritma *Naïve Bayes* dapat menjadi alat yang andal dan akurat dalam mendeteksi kebakaran hutan. Dengan demikian, model ini memiliki potensi besar untuk diterapkan dalam mendukung upaya pencegahan dan mitigasi kebakaran hutan di Indonesia secara lebih efektif.

ABSTRACT

Awwaliyah, Nikmatun. 2024. **Implementation of the *Naïve Bayes* Algorithm for Forest Fire Classification in Indonesia**. Undergraduate Thesis. Mathematics Department, Faculty of Science and Technology, Universitas Islam Negeri Maulana Malik Ibrahim Malang. Advisors: (I) Hisyam Fahmi, M.Kom. (II) Ari Kusumastuti, M.Pd., M.Si.

Keywords: *Naïve Bayes*, Classification, Forest Fire.

Forest fires in Indonesia have become a serious issue requiring immediate attention. According to data from Pusdalops BNPB, 487 forest and land fire incidents were recorded across various regions of Indonesia throughout 2023. The increasing frequency of these incidents demands an effective method for accurately identifying and classifying fires, thereby supporting earlier mitigation efforts. This study aims to evaluate the effectiveness of the *Naïve Bayes* algorithm in classifying forest fires, using secondary data from BNPB for fire class and Open Meteo for non-fire class. The research method employed is quantitative, with a numerical analysis approach focusing on natural factors such as temperature, weather, wind speed, and wind gusts that could potentially trigger forest fires. The results of the study indicate that the *Naïve Bayes* model achieved an accuracy rate of 93.03% on testing data, with a precision of 92.96% and a recall of 96.17%. These high-performance metric values demonstrate that the *Naïve Bayes* algorithm can serve as a reliable and accurate tool for detecting forest fires. Thus, this model holds significant potential to be applied in supporting more effective forest fire prevention and mitigation efforts in Indonesia.

مستخلص البحث

اولية، نعمة. ٢٠٢٤. تنفيذ خوارزمية **Bayes** الساذجة لتصنيف حرائق الغابات في إندونيسيا. البحث الجامعي. قسمي الرياضيات ، كلية العلوم والتكنولوجيا ، جامعة مولانا مالك إبراهيم الإسلامية الحكومية مالانج. المشرف: (I) هسام فهمي، الماجستير. (II) أري كوسوماستوتي، الماجستير

الكلمات المفتاحية: الخلجان الساذجة ، التصنيف ، حرائق الغابات.

تعتبر حرائق الغابات في إندونيسيا مشكلة خطيرة تتطلب اتخاذ إجراءات فورية. استناداً إلى بيانات من نظام بوسدالوبس التابع لمكتب حماية الغابات في إندونيسيا، تم تسجيل ٤٨٧ حريق غابات وأراضٍ في أنحاء مختلفة من إندونيسيا خلال عام ٢٠٢٣. ويتطلب التواتر المتزايد لهذه الحوادث طريقة فعالة لتحديد الحرائق وتصنيفها بدقة، بحيث يمكن أن تدعم جهود التخفيف المبكرة. تهدف هذه الدراسة إلى تقييم فعالية كفاءة **BNPB** في تصنيف حرائق الغابات، باستخدام بيانات ثانوية من **Naive Bayes** خوارزمية كفاءة غير حرائق. طريقة البحث المستخدمة كمية، مع نهج التحليل العددي **Open Meteo** حرائق و للعوامل الطبيعية مثل درجة الحرارة والطقس وسرعة الرياح وهبوب الرياح التي لديها القدرة على إشعال الساذج تمكن من تحقيق معدل دقة بلغ **Naive Bayes** حرائق الغابات. أظهرت النتائج أن نموذج على بيانات الاختبار، مع دقة بلغت 92.96% واستدعاء بنسبة 96.17%. تُثبت القيمة 93.03% يمكن أن تكون أداة موثوقة ودقيقة في الكشف **Naive Bayes** العالية لمقياس الأداء هذا أن خوارزمية عن حرائق الغابات. وبالتالي، فإن هذا النموذج لديه إمكانات كبيرة لتطبيقه في دعم جهود الوقاية من حرائق الغابات والتخفيف من آثارها في إندونيسيا بشكل أكثر فعالية.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Dalam era digital yang terus berkembang, Data *mining* telah menjadi komponen penting untuk menggali informasi berharga dari data yang sangat besar. Data *mining* adalah proses mengungkap serta menganalisis keterkaitan dan tren dalam data yang tersimpan, dengan memanfaatkan teknik pengenalan pola, seperti metode statistik dan matematika (Kristianti & Hariyadi, 2022). Salah satu teknik utama dalam Data *mining* adalah klasifikasi, yang digunakan untuk mengelompokkan data ke dalam kategori tertentu berdasarkan pola-pola yang ditemukan. Teknik klasifikasi ini sangat membantu untuk membuat prediksi atau keputusan yang lebih baik, terutama ketika berhadapan dengan dataset yang kompleks dan beragam. Dalam konteks masalah lingkungan seperti kebakaran hutan, klasifikasi dapat digunakan untuk mendeteksi dan mengidentifikasi kejadian kebakaran secara lebih dini, sehingga langkah-langkah pencegahan dapat dilakukan lebih cepat dan efektif.

Seperti yang telah dijelaskan dalam Al-Qur'an surah Al-Waqi'ah (56:7-12) (Kementerian Agama RI, 2024):

وَكُنْتُمْ أَزْوَاجًا ثَلَاثَةً ۗ فَأَصْحَابُ الْمَيْمَنَةِ مَا أَصْحَابُ الْمَيْمَنَةِ ۘ وَأَصْحَابُ الْمَشْأَمِ مَا أَصْحَابُ

الْمَشْأَمِ ۙ وَالسَّابِقُونَ السَّابِقُونَ ۚ أُولَئِكَ الْمُقَرَّبُونَ ۚ ۙ فِي جَنَّاتِ النَّعِيمِ ۚ

Artinya :

“Dan kamu menjadi tiga golongan. Yaitu golongan kanan, alangkah mulianya golongan kanan itu. Dan golongan kiri, alangkah sengsaranya golongan kiri itu. Dan orang-orang yang paling dahulu (beriman). Merekalah yang paling

dahulu (masuk surga). Mereka itulah orang yang dekat (kepada Allah), Berada dalam surga kenikmatan.”

Ayat ini menggambarkan pembagian manusia pada Hari Kiamat menjadi tiga golongan, yaitu golongan kanan, golongan kiri, dan golongan yang lebih dahulu beriman. Golongan kanan merupakan mereka yang akan menerima catatan amal dengan tangan kanan, sebagai tanda keberuntungan dan kebahagiaan di akhirat. Sebaliknya, golongan kiri adalah mereka yang akan menerima catatan amal dengan tangan kiri, sebagai tanda kesengsaraan dan hukuman. Sementara itu, golongan yang paling dahulu beriman adalah orang-orang yang terdekat dengan Allah, yang memperoleh kenikmatan di surga.

Salah satu algoritma yang banyak digunakan dalam teknik klasifikasi adalah *Naïve Bayes*. Algoritma ini berbasis pada Teorema Bayes, Teorema Bayes digunakan untuk menghitung probabilitas suatu peristiwa berdasarkan informasi sebelumnya. Meskipun sederhana, *Naïve Bayes* memiliki sejumlah keunggulan, seperti efisiensi dalam hal waktu komputasi, karena sering kali memerlukan waktu paling cepat untuk membangun model, sambil tetap mempertahankan tingkat akurasi yang tinggi (Pratiwi et al., 2021). Keunggulan lain dari algoritma ini adalah kemampuannya untuk menangani data dengan berbagai jenis atribut, baik numerik maupun kategoris. Dalam konteks prediksi kebakaran hutan, *Naïve Bayes* dapat memungkinkan untuk memanfaatkan data historis kebakaran serta faktor-faktor lain, seperti cuaca, guna memprediksi risiko kebakaran dengan tingkat akurasi yang baik.

Kebakaran hutan di Indonesia merupakan permasalahan yang memerlukan penanganan segera. Berdasarkan data Pusdalops BNPB, sepanjang tahun 2023 tercatat 487 kejadian kebakaran hutan dan lahan (Karhutla) di berbagai wilayah

Indonesia, dengan penyebab yang bervariasi antara faktor alam dan aktivitas manusia. Meningkatnya frekuensi kebakaran hutan ini menuntut adanya metode yang efektif untuk mengidentifikasi dan mengklasifikasikan kejadian kebakaran secara akurat. Implementasi algoritma *Naïve Bayes* dalam klasifikasi kebakaran hutan diharapkan dapat memberikan solusi yang andal dalam memprediksi dan mendeteksi kejadian kebakaran, sehingga dapat mendukung upaya pencegahan dan penanggulangan lebih dini.

Dampak kebakaran hutan sangat beragam, mulai dari masalah lingkungan, ekonomi, sosial, dan kesehatan (Rudiyanto et al., 2022). Secara lingkungan, kebakaran hutan mengakibatkan kerusakan habitat alami yang penting bagi keberlangsungan kehidupan satwa liar dan tumbuhan endemik. Polusi udara yang dihasilkan oleh kebakaran hutan juga menimbulkan dampak buruk pada kualitas udara, meningkatkan risiko terhadap penyakit pernapasan, dan memengaruhi produktivitas pertanian. Sementara itu, masyarakat lokal yang secara langsung atau tidak langsung bergantung pada sumber daya alam yang terdapat di hutan juga menderita kerugian ekonomi dan sosial yang signifikan. Kebakaran hutan sering kali mengakibatkan kerugian properti, hilangnya mata pencaharian, dan gangguan terhadap kegiatan sehari-hari masyarakat, terutama yang tinggal di sekitar kawasan hutan. Oleh karena itu, penting untuk mengembangkan pendekatan yang efektif dalam mengklasifikasikan kejadian kebakaran hutan, sehingga langkah-langkah pencegahan yang tepat dapat diambil untuk meminimalkan dampak kebakaran hutan serta melindungi ekosistem dan masyarakat yang terdampak.

Identifikasi kejadian kebakaran hutan menjadi aspek penting dalam upaya pencegahan kebakaran hutan di Indonesia. Dengan adanya klasifikasi kebakaran

hutan, diharapkan dapat membantu mencegah terulangnya peristiwa kebakaran hutan dan lahan, serta memungkinkan langkah-langkah yang tepat untuk mengurangi risiko dan dampaknya (All Rosit et al., 2023). Di sinilah algoritma *Naïve Bayes* dapat memainkan peran penting. Dengan kemampuannya dalam memproses data dan memprediksi kejadian, serta dibantu dengan beberapa variabel lingkungan seperti suhu udara, kelembaban permukaan bumi, curah hujan, titik embun, kecepatan angin, dan hembusan angin, *Naïve Bayes* dapat digunakan untuk membangun model klasifikasi yang membantu memprediksi kebakaran hutan, sehingga tindakan pencegahan dapat dilakukan secara lebih efektif.

Peran algoritma *Naïve Bayes* dalam klasifikasi kebakaran hutan sangatlah penting karena dapat membantu mengembangkan sistem yang mampu memprediksi dan mencegah kejadian kebakaran dengan lebih baik. Algoritma *Naïve Bayes* berguna untuk menganalisis pola-pola yang berkaitan dengan kebakaran hutan, seperti kelembapan udara, suhu permukaan bumi, curah hujan, titik embun, kecepatan angin, dan hembusan angin. Dengan memanfaatkan data kebakaran hutan sebelumnya serta faktor lingkungan yang relevan, algoritma ini dapat membantu dalam mendeteksi dan mengidentifikasi kejadian kebakaran hutan di masa depan. Metode klasifikasi *Naïve Bayes* dipilih karena mampu menghasilkan pemodelan yang baik dan akurasi tinggi pada data pelatihan maupun pengujian. Selain itu, algoritma ini terkenal karena efisiensi waktu komputasinya, yakni waktu yang dibutuhkan untuk membangun model. *Naïve Bayes* sering kali memiliki waktu eksekusi tercepat sambil tetap mempertahankan tingkat akurasi yang tinggi (Oktaviana, 2017), menjadikannya pilihan ideal untuk pemantauan dan pencegahan kebakaran hutan secara *real-time*.

Penelitian sebelumnya telah dilakukan oleh Pratiwi, dkk pada tahun 2021 yang membahas tentang penggunaan Algoritma *Naïve Bayes* pada klasifikasi kebakaran hutan di Kabupaten Pelalawan. Atribut suhu, kelembaban, kecepatan angin, dan curah hujan digunakan dalam klasifikasi tersebut. Dengan menggunakan dataset 2017 tingkat akurasi tertinggi yang dihasilkan adalah 81,03%. Penelitian sebelumnya juga pernah dilakukan oleh Fitriyah & Fahmi pada tahun 2019 yang membahas penggunaan algoritma *Sequential Forward Floating Selection* (SFFS) dan *Sequential Backward Floating Selection* (SBFS) dalam kasus kebakaran hutan. Penelitian ini bertujuan untuk mengidentifikasi fitur-fitur lingkungan yang paling berpengaruh terhadap kebakaran hutan, seperti suhu permukaan bumi dan cakupan awan. Hasil menunjukkan bahwa suhu permukaan bumi adalah fitur paling signifikan dengan akurasi rata-rata 96,23% menggunakan SFFS. Metode SBFS memiliki akurasi lebih rendah, sekitar 82,41%. Penelitian lainnya tentang kebakaran hutan dilakukan oleh Mifaul et al., 2018 menerapkan algoritma *Naïve Bayes* dalam pengembangan sistem deteksi titik kebakaran. Metode tersebut menggunakan sensor api berbasis Arduino dan sensor suhu. Hasil penelitian menyimpulkan tiga hal. Pertama, saat diujikan di dalam ruangan tanpa penggunaan AC dengan 36 titik pengujian, sistem pendeteksi lokasi titik kebakaran memiliki akurasi sebesar 94%. Kedua, saat digunakan di ruangan yang menggunakan AC dengan jumlah pengujian yang sama, akurasi sistem pendeteksian kebakaran adalah 86%. Terakhir, sistem tersebut memerlukan waktu 0.48 detik untuk menentukan lokasi terjadinya kebakaran.

Tujuan utama dari penelitian ini adalah untuk mengimplementasikan algoritma *Naïve Bayes* dalam klasifikasi kebakaran hutan di Indonesia, dengan

harapan dapat meningkatkan pemahaman dan kemampuan dalam pencegahan kejadian kebakaran hutan. Melalui implementasi ini, penelitian ini bertujuan untuk menghasilkan model prediktif yang mampu mendeteksi dan mengidentifikasi kejadian kebakaran hutan, sehingga para pemangku kepentingan dapat mengambil tindakan yang tepat. Dengan demikian, penelitian ini diharapkan dapat memberikan dampak positif yang signifikan dalam upaya menjaga kelestarian lingkungan dan kesejahteraan masyarakat di Indonesia.

1.2 Rumusan Masalah

Berdasarkan uraian pada latar belakang di atas, rumusan masalah dalam penelitian ini adalah bagaimana evaluasi performa model klasifikasi yang dibangun menggunakan algoritma *Naïve Bayes* dalam mendeteksi dan mengidentifikasi kejadian kebakaran hutan di Indonesia?

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah yang ada di atas, maka tujuan penelitian yang hendak dicapai yakni untuk mengevaluasi keefektifan dan akurasi model klasifikasi yang dikembangkan dalam mendeteksi dan mengidentifikasi kejadian kebakaran hutan di Indonesia.

1.4 Manfaat Penelitian

Berdasarkan tujuan penelitian di atas maka setelah melakukan penelitian diharapkan dapat memberikan manfaat sebagai berikut:

1. Memperkaya pemahaman tentang penggunaan algoritma *Naïve Bayes* dalam konteks klasifikasi kejadian kebakaran hutan di Indonesia.
2. Memberikan informasi yang berharga kepada pembaca tentang faktor-faktor lingkungan yang signifikan yang dapat mempengaruhi terjadinya kebakaran hutan, sehingga dapat membantu masyarakat dan pihak terkait dalam mengambil keputusan yang tepat dalam upaya pencegahan kebakaran hutan.
3. Menyediakan tambahan informasi yang penting bagi instansi terkait pencegahan kebakaran hutan di Indonesia untuk meningkatkan efisiensi dan efektivitas langkah-langkah pencegahan dan penanganan bencana kebakaran hutan.

1.5 Batasan Masalah

Batasan masalah yang digunakan pada penelitian ini untuk menghindari pembahasan secara luas adalah data kebakaran hutan hanya berdasarkan faktor alam, yang meliputi kelembaban, suhu, curah hujan, titik embun, kecepatan angin, dan hembusan angin

1.6 Definisi Istilah

Terdapat beberapa istilah yang digunakan dalam penelitian ini, yakni sebagai berikut :

1. *Data mining* : Proses mengumpulkan, menganalisis, dan mengekstrak informasi atau pola penting dari kumpulan data yang besar

2. *Klasifikasi* : Proses pengelompokan atau pembagian objek, data, atau informasi ke dalam kategori atau kelompok berdasarkan karakteristik atau kriteria tertentu.
3. *Training* : Proses melatih model *Naïve Bayes* menggunakan data berlabel (misalnya, data kebakaran hutan) agar model dapat mengenali pola dan memprediksi kelas dari data baru berdasarkan pola tersebut.
4. *Testing* : Proses menguji model *Naïve Bayes* yang sudah dilatih menggunakan data baru yang belum pernah dilihat oleh model, untuk mengukur seberapa baik model tersebut memprediksi kelas secara akurat.

BAB II

KAJIAN TEORI

2.1 Teori Pendukung

2.1.1 Data *mining*

Menurut Efori Buulolo dalam bukunya yang berjudul *Data mining Untuk Perguruan Tinggi*, *Data mining* adalah proses yang terkait dengan pengumpulan data serta analisis data historis guna mengungkap pengetahuan, informasi, pola, keteraturan, atau hubungan dalam kumpulan data yang besar (Buulolo, 2020). *Data mining* adalah proses penemuan pola dan pengetahuan yang berharga dari kumpulan data yang besar atau kompleks. Tujuan utamanya adalah untuk mengekstraksi informasi yang tersembunyi, yang dapat membantu dalam pengambilan keputusan yang lebih baik, prediksi tren masa depan, dan mengidentifikasi hubungan yang relevan antar fitur. Teknik *Data mining* memanfaatkan pendekatan statistik, matematika, dan kecerdasan buatan guna menganalisis serta memahami pola-pola yang terdapat dalam data. Ini sering digunakan di berbagai bidang, termasuk bisnis, ilmu pengetahuan, kesehatan, keamanan, dan lainnya, untuk mendukung pengambilan keputusan yang lebih baik dan pemahaman yang lebih dalam tentang data. *Data mining* mengklasifikasikan data menjadi tiga jenis, yaitu data berdasarkan rentang waktu, data kategorikal, dan data numerik. (Pratiwi et al., 2021).

Dalam buku tentang *Data mining*, Algoritma, dan Implementasinya dengan menggunakan bahasa pemrograman PHP, konsep *Data mining* dibahas secara

mendalam dinyatakan memiliki lima peran berdasarkan fungsi-fungsinya, yang secara umum dapat diuraikan sebagai berikut :

1. Klasifikasi

Klasifikasi merupakan proses pengelompokan data ke dalam kelas-kelas tertentu berdasarkan pola variabel yang ditemukan, dengan tujuan memprediksi kelas dari data yang belum memiliki label (Wahyuni Kalumbang, 2021). Dalam proses klasifikasi, ada variabel yang menjadi fokus yang memiliki kategori-kategori tertentu. Sebagai contoh, dalam mengelompokkan pendapatan, bisa dibagi menjadi tiga kategori: tinggi, menengah, dan rendah. Demikian pula, dalam mengevaluasi pengajuan hipotek, dapat menentukan apakah itu dianggap sebagai kredit yang baik atau buruk.

2. Estimasi

Estimasi merupakan proses untuk menghitung atau menganalisis nilai yang belum diketahui dengan pasti (Pratiwi et al., 2021). Estimasi serupa dengan klasifikasi, tetapi variabel yang menjadi fokus dalam estimasi biasanya memiliki sifat numerik daripada kategorikal. Model dikonstruksi dengan menggunakan data lengkap yang mencakup nilai variabel yang akan diprediksi. Selanjutnya, dalam tahap evaluasi, nilai yang diperkirakan dari variabel target dibuat berdasarkan prediksi dari variabel tersebut. Sebagai contoh, dapat dilakukan estimasi tekanan darah sistolik pada pasien di rumah sakit dengan mempertimbangkan faktor-faktor seperti usia, jenis kelamin, berat badan, dan tingkat natrium dalam darah.

3. Prediksi

Prediksi adalah proses untuk meramalkan nilai suatu variabel berdasarkan nilai variabel lainnya (Pratiwi et al., 2021). Prediksi memiliki kemiripan dengan klasifikasi dan estimasi, namun perbedaannya terletak pada fokus prediksi terhadap nilai yang akan terjadi di masa depan. Sebagai contoh, dalam konteks bisnis, prediksi mungkin dilakukan terkait harga beras dalam tiga bulan mendatang.

4. *Clustering*

Clustering adalah proses untuk mengelompokkan data berdasarkan kesamaan karakteristik, di mana jarak antara elemen-elemen dalam kluster diminimalkan sedangkan jarak antara kluster diperbesar (Pratiwi et al., 2021).

5. Asosiasi

Asosiasi adalah proses untuk menemukan pola atau hubungan antara berbagai kombinasi item dari atribut tertentu (Pratiwi et al., 2021).

Data *mining* memiliki 5 tahapan (Pratiwi et al., 2021):

1. Seleksi Data

Seleksi data merupakan tahap awal dalam proses Data *mining* yang melibatkan pemilihan himpunan data, sampel data, dan pembentukan himpunan data target sebelum melakukan eksplorasi informasi dalam Data *mining*.

2. *Pre-processing*

Pre-processing adalah proses pembersihan data untuk mengatasi masalah seperti data yang mengganggu, duplikat, inkonsisten, dan kesalahan ketik yang dapat mengganggu proses *Data mining*.

3. *Transformasi*

Transformasi merujuk pada proses mengubah struktur atau format data menjadi bentuk yang sesuai untuk analisis dalam proses mining.

4. *Data mining*

Data mining adalah konsep yang digunakan untuk mengeksplorasi pengetahuan dari data besar dengan menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin dalam kumpulan data yang besar.

5. *Evaluasi*

Evaluasi melibatkan pemahaman dan analisis pola atau informasi yang ditemukan dalam proses *Data mining*, sehingga dapat dimengerti dan digunakan dengan mudah oleh pengguna.

2.1.2 **Klasifikasi**

Klasifikasi adalah proses dalam *Data mining* yang bertujuan untuk mengelompokkan atau mengkategorikan data ke dalam kelas atau kategori tertentu berdasarkan atribut atau karakteristik tertentu (Wahyuni Kalumbang, 2021). Tujuan utama dari klasifikasi adalah untuk membuat model yang dapat memprediksi kelas atau label dari data baru berdasarkan pola yang ditemukan dari data yang telah ada. Ini merupakan salah satu teknik penting dalam analisis data

untuk membuat keputusan atau prediksi berdasarkan data yang telah dikumpulkan dan diproses. Berikut merupakan beberapa metode klasifikasi yang terdapat dalam *Data mining* (Pratiwi et al., 2021):

1. *Neural Network*

Jaringan Syaraf Tiruan (*Neural Network*) adalah suatu model komputasi yang terinspirasi oleh cara kerja otak manusia, terdiri dari jaringan neuron buatan yang bekerja sama untuk memproses informasi. Model ini digunakan untuk mengenali pola dan memecahkan masalah kompleks dengan menyesuaikan bobot dan bias berdasarkan data yang dilatihkan, sehingga mampu melakukan prediksi atau klasifikasi pada data baru.

2. *Decision Tree*

Decision Tree, di sisi lain, adalah metode kuat dalam klasifikasi dan prediksi yang mendapat banyak perhatian.

3. *Naïve Bayes*

Algoritma *Naïve Bayes*, merupakan algoritma klasifikasi yang mengaplikasikan teorema Bayesian dalam statistika untuk memprediksi probabilitas suatu kelas.

4. *K-Nearest Neighbor*

K-Nearest Neighbor (KNN) adalah metode klasifikasi yang memanfaatkan data pembelajaran untuk menentukan kelas objek berdasarkan jaraknya yang terdekat dengan objek tersebut. Kepresisian algoritma KNN sangat dipengaruhi oleh keberadaan fitur-fitur yang tidak relevan atau jika bobot fitur tersebut tidak seimbang dengan relevansinya terhadap klasifikasi.

5. *Logic Regression*

Logistic regression, pada sisi lain, adalah teknik dalam analisis regresi yang digunakan ketika variabel responnya bersifat dikotomis.

2.1.3 Korelasi Pearson

Analisis korelasi merupakan metode statistik yang bertujuan untuk mengukur kekuatan hubungan antara dua variabel, dengan asumsi bahwa variabel lain yang berpotensi memengaruhi hubungan tersebut dikendalikan atau dianggap tetap sebagai variabel kontrol. Karena data yang dianalisis merupakan data numerik, maka teknik korelasi yang sesuai adalah *Pearson correlation*, yang menghitung derajat hubungan linier antara kedua variabel (Romadloni & Pardede, 2019).

Nilai yang diperoleh dari analisis Pearson berkisar antara -1 hingga +1. Nilai -1 menunjukkan adanya korelasi negatif sempurna, di mana ketika satu fitur meningkat, fitur lainnya akan menurun. Sebaliknya, nilai +1 menandakan korelasi positif sempurna, di mana peningkatan satu fitur diikuti oleh peningkatan fitur lainnya. Nilai 0 menunjukkan bahwa tidak ada hubungan linier yang signifikan antara kedua fitur tersebut. Selain nilai korelasi, analisis Pearson juga menghasilkan nilai *p-value*, yang digunakan untuk menentukan signifikansi hubungan antar fitur. Dalam uji korelasi Pearson, hipotesis yang diuji adalah sebagai berikut:

1. Hipotesis Nol (H_0) : Tidak terdapat hubungan linier yang signifikan antara dua variabel (koefisien korelasi = 0).

2. Hipotesis Alternatif (H_1) : Terdapat hubungan linier yang signifikan antara dua variabel (koefisien korelasi $\neq 0$).

P -value digunakan untuk menentukan apakah H_0 ditolak atau diterima. Jika p -value lebih kecil atau sama dengan tingkat signifikansi yang ditetapkan ($\alpha = 0,05$), maka H_0 ditolak, yang berarti terdapat hubungan linier yang signifikan secara statistik antara kedua variabel. Sebaliknya, jika p -value lebih besar dari tingkat signifikansi yang ditetapkan ($\alpha = 0,05$), maka H_0 diterima, yang menunjukkan bahwa hubungan linier antara variabel tidak signifikan (Alim & Arianto, 2023).

Berikut merupakan persamaan *Correlation Pearson* (Rozy et al., 2018):

$$r_{xy} = \frac{n(\sum xy) - (\sum x)(\sum y)}{(\sqrt{[n \sum x^2 - (\sum x)^2]}) (\sqrt{[n \sum y^2 - (\sum y)^2]})} \quad (2.1)$$

di mana,

r_{xy} : *Pearson's Product Moment Correlation Coefficient*

n : Banyaknya pasangan data x dan y

$\sum x$: Total jumlah dari fitur x

$\sum y$: Total jumlah dari fitur y

$\sum x^2$: Total jumlah dari kuadrat fitur x

$\sum y^2$: Total jumlah dari kuadrat fitur y

Metode Korelasi Pearson ini memiliki keunggulan dalam hal kecepatan dan efisiensi, terutama ketika diterapkan pada dataset dengan jumlah fitur yang sangat besar (Romadloni & Pardede, 2019). Dengan mempertimbangkan nilai p -

value dan hipotesis yang diuji, metode ini memberikan hasil yang lebih terukur dan memungkinkan pengambilan keputusan berbasis statistik yang lebih akurat.

2.1.4 Standardisasi Data

Standardisasi data merupakan salah satu teknik dari normalisasi data (Setiono, 2022). Standardisasi data adalah proses transformasi data untuk memastikan bahwa semua fitur dalam dataset memiliki skala yang seimbang ketika menggunakan algoritma yang sensitif terhadap perbedaan skala (Goejantoro & Desi Yuniarti, 2019). Dengan standardisasi, data diubah sedemikian rupa sehingga setiap fitur memiliki rata-rata nol (0) dan deviasi standar satu (1). Standardisasi biasanya diterapkan pada data yang mengikuti distribusi normal dan memiliki variabel numerik dengan skala yang berbeda-beda, seperti suhu dan kelembaban pada data kebakaran hutan. Proses ini dilakukan dengan menghitung rata-rata dan deviasi standar dari setiap fitur sebagai berikut (Goejantoro & Desi Yuniarti, 2019):

1. Rata Rata

$$\mu = \frac{1}{N} \sum_{i=1}^N x_i \quad (2.2)$$

di mana,

x_i : Nilai data ke- i pada fitur X

N : Jumlah data

μ : Rata rata dari fitur X

2. Varians

$$\sigma^2 = \frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2 \quad (2.3)$$

3. Standardisasi Data

$$\hat{x}_i = \frac{x_i - \mu}{\sigma} \quad (2.4)$$

di mana,

σ : Standar deviasi, $\sqrt{\sigma^2}$

\hat{x}_i : Nilai standardisasi dari data ke- i

2.1.5 *Principal Component Analysis (PCA)*

Principal Component Analysis (PCA) adalah teknik reduksi dimensi yang digunakan untuk mengatasi korelasi tinggi antar variabel dalam data. PCA digunakan untuk menyederhanakan data dengan mereduksi variabel yang memiliki korelasi tinggi, sehingga menghasilkan data dengan variabel yang lebih sedikit namun tetap mewakili informasi penting (Astoni & Haris, 2022). Teknik ini sangat berguna dalam kasus ketika terdapat korelasi tinggi antara fitur-fitur, yang dapat melanggar asumsi independensi pada algoritma *Naïve Bayes*.

PCA dapat diterapkan untuk memproses data dengan mengekstraksi pola yang terdapat dalam data. Teknik ini mampu mengenali pola-pola tertentu sehingga dapat membedakan atau mengidentifikasi kesamaan dengan data lainnya (Nugraheny, 2015). Selain itu, dengan menggunakan PCA, jumlah variabel yang digunakan dalam model dapat dikurangi dan menyisakan beberapa komponen utama tanpa kehilangan terlalu banyak informasi penting. Penentuan jumlah komponen utama (*principal components/PCs*) yang akan diambil dapat dilakukan dengan tiga metode berikut (Enzellina & Suhaedi, 2022):

1. Mengamati Scree Plot : Grafik ini membantu mengidentifikasi titik "elbow," yaitu jumlah komponen utama sebelum penurunan kontribusi variansi menjadi lebih lambat.
2. Proporsi Kumulatif Variansi : Mengambil jumlah komponen utama yang menjelaskan setidaknya 80% keragaman total data.
3. Nilai Eigen (Eigenvalues) : Memilih komponen utama dengan nilai eigen ≥ 1 , karena nilai ini menunjukkan kontribusi variansi yang signifikan dari setiap komponen.

Ketiga pendekatan ini memberikan panduan untuk menentukan jumlah komponen yang optimal sesuai kebutuhan analisis data.

PCA tidak hanya mengurangi kompleksitas data, tetapi juga memastikan bahwa variabel yang digunakan dalam model *Naïve Bayes* lebih mendekati asumsi independensi antar fitur. Hal ini membantu meningkatkan performa model *Naïve Bayes* dalam klasifikasi kebakaran hutan, di mana variabel seperti suhu, kelembaban, dan kecepatan angin seringkali berkorelasi tinggi

2.1.6 *Naïve Bayes*

Menurut buku *Data mining Untuk Perguruan Tinggi* yang ditulis oleh Efori Buulolo, Algoritma *Naïve Bayes classifier* merupakan salah satu teknik dalam *Data mining* yang memecahkan masalah dengan menggunakan prinsip probabilitas (Buulolo, 2020). *Naïve Bayes* adalah sebuah algoritma klasifikasi dalam *Data mining* yang didasarkan pada teorema *Bayes*. Algoritma ini digunakan untuk memprediksi probabilitas keanggotaan suatu data ke dalam kelas atau kategori tertentu berdasarkan fitur-fitur yang dimilikinya. Salah satu asumsi

utama dari algoritma *Naïve Bayes* adalah bahwa setiap fitur dalam data adalah independen satu sama lain, meskipun dalam kenyataannya tidak selalu demikian. Meskipun demikian, algoritma *Naïve Bayes* sering digunakan karena sederhana, efisien, dan sering kali memberikan hasil yang baik dalam klasifikasi data, terutama dalam kasus di mana data memiliki banyak fitur.

Rumus yang digunakan untuk mencari nilai peluang dalam algoritma *Naïve Bayesian classifier* adalah sebagai berikut (Bulolo, 2020):

$$P(H|X) = \frac{P(X|H) \cdot P(H)}{P(X)} \quad (2.5)$$

di mana,

$P(H)$: Probabilitas prior dari kelas H

$P(X)$: Probabilitas total dari fitur X

$P(X|H)$: Probabilitas *likelihood* dari fitur X diberikan kelas H

$P(H|X)$: Probabilitas posterior dari kelas H diberikan fitur X

Karena $P(X)$ adalah konstanta untuk semua kelas yang dibandingkan, maka hanya perlu memaksimalkan pembilangnya:

$$P(H | X) \propto P(X | H) \cdot P(H) \quad (2.6)$$

dengan asumsi independensi kondisional (asumsi *Naive*), likelihood $P(X | H)$ dapat diuraikan menjadi:

$$P(X | H) = \prod_{i=1}^n P(x_i | H) \quad (2.7)$$

di mana x_i adalah fitur ke- i dalam vektor fitur X dan n menunjukkan jumlah total fitur dalam vektor fitur X . Jadi rumus akhirnya adalah :

$$P(H | X) \propto P(H) \cdot \prod_{i=1}^n P(x_i | H) \quad (2. 8)$$

untuk melakukan klasifikasi, dipilih kelas H yang memaksimalkan $P(H|X)$:

$$\hat{H} = \operatorname{argmax}_H (P(H) \cdot \prod_{i=1}^n P(x_i | H)) \quad (2. 9)$$

di mana,

\hat{H} : Kelas yang diprediksi

argmax_H : Argumen dari nilai maksimum pada H

Penelitian terdahulu pernah dilakukan oleh (Nugroho & Khotimah, 2017) yang mengulas mengenai klasifikasi tingkat kerawanan kebakaran hutan dan lahan dengan memanfaatkan algoritma *Naïve Bayes* di wilayah Pulang Pisau, Kalimantan Tengah. Penelitian tersebut sukses dalam mengkategorikan tingkat kerawanan kebakaran hutan dan lahan di wilayah tersebut dengan menggunakan algoritma *Naïve Bayes*. Hasil dari model klasifikasi ini mencapai akurasi rata-rata sebesar 71.02% pada data *training*.

Putrada et al., (2018) melakukan penelitian yang membahas tentang klasifikasi asap dengan metode *Naïve Bayes* pada "Internet of Things" yang menunjukkan bahwa proses klasifikasi menggunakan metode *Naïve Bayes* memiliki tingkat akurasi antara 88% hingga 91%, yang menandakan bahwa hasil tersebut dapat dianggap memadai untuk keperluan klasifikasi. Penelitian lain juga dilakukan oleh Mifaul et al. (2018) menerapkan algoritma *Naïve Bayes* dalam merancang sistem deteksi titik kebakaran. Metode ini menggunakan sensor suhu dan sensor api yang berbasis Arduino, yang menghasilkan tiga temuan utama.

Pertama, ketika sistem pendeteksi lokasi titik kebakaran diuji dalam ruangan tanpa AC dengan total 36 titik pengujian, keakuratannya mencapai 94%. Temuan kedua menunjukkan bahwa saat sistem digunakan di ruangan yang menggunakan AC, dari 36 titik pengujian, tingkat keakuratannya adalah 86%. Temuan terakhir menunjukkan bahwa sistem memerlukan waktu 0.48 detik untuk membuat keputusan tentang lokasi terjadinya kebakaran.

2.1.7 Gaussian Naïve Bayes

Gaussian Naïve Bayes adalah salah satu model klasifikasi yang merupakan varian dari *Naïve Bayes*. Model ini digunakan untuk data kontinu, di mana setiap fitur pada data diasumsikan mengikuti distribusi Gaussian atau fungsi kepadatan probabilitas normal (*Normal Probability Density Function*) (Rozy et al., 2018). *Gaussian Naïve Bayes* (GNB) merupakan varian dari algoritma *Naïve Bayes* yang menggunakan distribusi Gaussian untuk menangani data kontinu. Algoritma ini berasumsi bahwa fitur-fitur dalam data berdistribusi normal, sehingga memungkinkan penggunaan parameter seperti rata-rata (μ) dan standar deviasi (σ) untuk menghitung probabilitas kondisi tiap fitur dalam suatu kelas. GNB sangat populer dalam klasifikasi berbasis statistik karena sederhana, cepat, dan efektif dalam menangani dataset besar, khususnya jika asumsi independensi antar fitur bisa diterima (Akil & Chaidir, 2024).

Teorema Bayes yang menjadi dasar GNB berguna dalam menghitung probabilitas bersyarat, yaitu peluang suatu kejadian berdasarkan informasi yang ada. Dalam GNB, setiap fitur dikondisikan untuk setiap kelas dengan distribusi Gaussian, yang memungkinkan pengklasifikasian data baru berdasarkan kalkulasi

probabilitas tertinggi. Rumus yang digunakan untuk menghitung probabilitas pada distribusi Gaussian adalah (A'yunnisa et al., 2022):

$$P(X | H) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(X - \mu)^2}{2\sigma^2}\right) \quad (2. 10)$$

di mana,

μ : Rata rata

σ^2 : Variansi untuk fitur tertentu

Secara empiris, GNB menunjukkan performa yang baik dalam berbagai aplikasi, seperti klasifikasi gambar tulisan tangan, diagnosa kesehatan, dan klasifikasi teks. Misalnya, dalam penelitian tentang klasifikasi penyakit jantung, GNB menunjukkan akurasi yang tinggi, membuatnya efektif dalam kasus diagnosis medis yang kompleks dengan data numerik (Akil & Chaidir, 2024)

2.1.8 *Confusion Matrix*

Confusion matrix merupakan alat evaluasi yang sering digunakan dalam klasifikasi untuk melihat kinerja model secara menyeluruh. Matrix ini menggambarkan jumlah prediksi yang benar dan salah, serta bagaimana data diklasifikasikan ke dalam kelas-kelas yang berbeda (Normawati & Prayogi, 2021). Dengan menggunakan *confusion matrix*, maka hal tersebut dapat membantu dalam menghitung metrik evaluasi seperti akurasi, presisi, recall, dan F1-score, yang membantu memberikan gambaran mendetail mengenai efektivitas model *Naïve Bayes* dalam melakukan klasifikasi pada data penelitian ini. Berikut merupakan contoh *Confusion matrix* untuk klasifikasi.

Tabel 2.1 *Confusion Matrix*

		<i>Prediction</i>	
		1	0
<i>Actual</i>	1	TP	FN
	0	FP	TN

di mana,

TP (*True Positive*) : Model memprediksi terjadi (1), dan benar benar terjadi (1).

FP (*False Positive*) : Model memprediksi terjadi (1), tapi sebenarnya tidak terjadi (0).

TN (*True Negative*) : Model memprediksi tidak terjadi (0), dan memang tidak terjadi (0).

FN (*False Negative*) : Model memprediksi tidak terjadi (0), tetapi sebenarnya terjadi (1).

Berikut merupakan rumus *Confusion matrix* untuk menghitung *accuracy*, *precision*, dan, *recall*:

$$accuracy = \frac{TP + TN}{Total} \quad (2.11)$$

$$precision = \frac{TP}{TP + FP} \quad (2.12)$$

$$recall = \frac{TP}{TP + FN} \quad (2.13)$$

2.1.9 Kebakaran Hutan

Kebakaran hutan dan lahan adalah kejadian kebakaran yang bisa disebabkan oleh tindakan manusia maupun faktor alam, yang terjadi dengan api

menjalar dan membakar area hutan dan lahan yang terkena dampaknya (Pratiwi et al., 2021). Kebakaran hutan terjadi ketika bahan bakar seperti tumbuhan dan material organik terpapar panas yang cukup untuk memicu oksidasi. Proses ini menghasilkan panas, gas, dan partikel yang menciptakan pola perilaku khas. Kebakaran bisa berkisar dari kecil hingga besar, dipengaruhi oleh cuaca, vegetasi, topografi, dan intervensi manusia. Secara ekologi, kebakaran adalah bagian dari dinamika alami hutan, tetapi kebakaran yang berlebihan bisa merusak keanekaragaman hayati dan fungsi ekosistem.

Beberapa elemen yang dapat memicu kebakaran hutan dan lahan meliputi suhu, tingkat kelembaban, curah hujan, dan kecepatan angin (Reza Noviansyah et al., 2018). Cuaca kering dan panas dapat menciptakan kondisi yang sangat rentan terhadap kebakaran, di mana vegetasi menjadi sangat mudah terbakar. Petir juga sering menjadi penyebab kebakaran hutan, terutama di daerah yang rentan terhadap aktivitas petir. Di samping itu, aktivitas manusia seperti pembakaran sampah, pembukaan lahan dengan menggunakan api, atau kelalaian dalam memadamkan api dapat memicu kebakaran yang merusak. Selain itu, praktek-praktek yang tidak berkelanjutan dalam pengelolaan hutan, seperti penebangan liar dan pembakaran hutan untuk pembersihan lahan, juga dapat meningkatkan risiko kebakaran.

Di Indonesia, kebakaran hutan dipengaruhi oleh faktor geografis, iklim, dan aktivitas manusia. Kebakaran lebih sering terjadi selama musim kemarau dengan kondisi cuaca kering dan angin kencang. Daerah dengan vegetasi mudah terbakar, seperti hutan gambut di Kalimantan dan Sumatera, menjadi fokus utama. Praktik pertanian lahan bakar dan pembukaan lahan tak terkontrol juga

meningkatkan risiko kebakaran. Pola kebakaran ini memengaruhi keanekaragaman hayati, kesehatan ekosistem, dan memiliki dampak sosial-ekonomi signifikan bagi masyarakat sekitar.

Dampak merugikan secara langsung dari kebakaran hutan ini mencakup kerusakan infrastruktur, kerugian aset pertanian, perkebunan, dan kehutanan, serta potensi kerugian nyawa manusia (Maharani et al., 2020). Di samping itu, ancaman terhadap dampak ekologi, ekonomi, dan kesehatan dari bencana kabut asap yang timbul juga menjadi risiko akibat kebakaran hutan. Secara ekologis, kebakaran hutan dapat menyebabkan kerusakan habitat yang luas dan kehilangan biodiversitas, karena menyebabkan kematian atau pengusiran berbagai jenis flora dan fauna serta mengubah struktur ekosistem hutan. Dampak sosial-ekonomi meliputi kerugian ekonomi akibat kerusakan infrastruktur dan sumber daya alam, hilangnya mata pencaharian bagi masyarakat yang bergantung pada hutan, serta dampak negatif terhadap kesehatan manusia akibat pencemaran udara dan hilangnya sumber air bersih.

Upaya pengendalian dan pencegahan kebakaran hutan adalah bagian penting dari pengelolaan hutan berkelanjutan di Indonesia. Langkah-langkah yang dilakukan meliputi peningkatan patroli, penegakan hukum, edukasi masyarakat, dan pemanfaatan teknologi seperti penginderaan jauh dan SIG. Kerja sama lintas sektor dan partisipasi aktif dari semua pemangku kepentingan diperlukan untuk melindungi keanekaragaman hayati, lingkungan, dan kesejahteraan masyarakat.

2.2 Kajian Integrasi Topik dengan Al-Quran/Hadits

Pada sub bab 1.1 dijelaskan mengenai penggolongan manusia berdasarkan amal perbuatan mereka, di mana manusia akan terbagi menjadi tiga golongan, yakni golongan kanan, golongan kiri, dan golongan yang lebih dahulu beriman. Golongan kanan yakni mereka yang beriman dan beramal saleh, adalah orang-orang yang dalam kehidupan di dunia telah menjalankan perintah Allah dan menjauhi larangannya. Mereka akan menerima catatan amal mereka dengan tangan kanan, sebagai tanda keberuntungan dan keselamatan di akhirat. Golongan ini akan memperoleh berbagai kenikmatan dan kedamaian di surga sebagai balasan atas keteguhan iman dan kesabaran mereka selama hidup di dunia. Sebagaimana yang dijelaskan dalam Surah Al-Insan (76:11-12) (Kementrian Agama RI, 2024):

فَوَقَاهُمُ اللَّهُ شَرَّ ذَلِكَ الْيَوْمِ وَلَقَّاهُمْ نَضْرَةً وَسُرُورًا ۝ ۱۱ وَجَزَّاهُمْ بِمَا صَبَرُوا جَنَّةً وَحَرِيرًا ۝ ۱۲

Artinya :

“Maka Tuhan memelihara mereka dari kesusahan hari itu, dan memberikan kepada mereka cahaya yang berseri-seri dan kegembiraan. Dan Dia memberi balasan kepada mereka karena kesabaran mereka (dengan) surga dan (pakaian) sutera.”

Asbabun nuzul (sebab turunnya) Surah Al-Insan (76:11-12) berkaitan dengan kisah pengorbanan dan kesabaran keluarga Ali bin Abi Thalib dan Fatimah. Dikisahkan bahwa mereka bernazar untuk berpuasa tiga hari jika Allah menyembuhkan kedua anak mereka, Hasan dan Husain, yang sedang sakit. Setelah sembuh, mereka berpuasa, tetapi pada setiap berbuka, mereka memberikan makanan yang seharusnya untuk diri mereka kepada seorang miskin, anak yatim, dan tawanan yang datang meminta pertolongan, sehingga mereka sendiri hanya berbuka dengan air. Karena kesabaran dan keikhlasan keluarga Ali dalam memenuhi nazar mereka dan memberikan makanan meskipun dalam keadaan sulit,

Allah menurunkan ayat ini sebagai penghargaan atas kesabaran mereka, menjelaskan bahwa orang-orang yang sabar akan dijauhkan dari kesulitan pada Hari Kiamat dan diberikan kenikmatan di surga.

Dalam Tafsir Al-Munir, Surah Al-Insan ayat 11-12 juga menjelaskan balasan istimewa bagi orang-orang yang beriman dan bersabar. Tafsir ini menyebutkan bahwa Allah akan menyelamatkan mereka dari berbagai kesulitan dan bencana pada Hari Kiamat, suatu hari yang penuh dengan ketakutan besar. Mereka yang bersabar akan dihiasi dengan perlindungan dari azab dan mendapat kesejahteraan serta kedamaian. Ayat 12 menambahkan bahwa Allah memberi mereka kebahagiaan sejati dan wajah yang berseri-seri, sebuah tanda kemuliaan yang sangat kontras dengan penderitaan yang pernah mereka alami di dunia. Sebagai penghargaan atas kesabaran mereka, Allah menganugerahkan mereka pakaian dari sutera halus dan tempat di surga yang penuh kenikmatan. Tafsir ini menekankan bahwa semua kenikmatan dan kemuliaan tersebut adalah balasan atas pengorbanan mereka di dunia, yang menunjukkan betapa besarnya pahala bagi orang yang tabah dan teguh imannya. Pesan utama dalam tafsir ini adalah bahwa kesabaran dan ketabahan di dunia tidak akan sia-sia, tetapi akan membawa pada keberkahan dan kemuliaan yang abadi di akhirat (Az-Zuhaili, 2013).

Dalam tafsir Al-Jalalayn, Surah Al-Insan (76:11-12) dijelaskan bahwa ayat-ayat ini menggambarkan imbalan yang dijanjikan Allah kepada orang-orang yang sabar. Mereka yang bertahan dalam ujian dan kesulitan di dunia akan dilindungi dari kesengsaraan pada hari kiamat, serta akan mendapatkan kegembiraan dan cahaya yang memancar dari wajah mereka sebagai tanda kebahagiaan. Surga yang mereka peroleh akan dipenuhi dengan berbagai kenikmatan, termasuk pakaian

sutera yang lembut dan dipan yang nyaman. Tafsir ini menekankan bahwa kenikmatan yang diperoleh di akhirat adalah hasil dari kesabaran dan keteguhan hati mereka dalam menjalani kehidupan di dunia, menunjukkan bahwa Allah menghargai setiap pengorbanan yang dilakukan oleh hamba-Nya (Al Mahalli & As Suyuti, 2003).

Surah Al-Insan (76:11-12) menjelaskan tentang balasan yang dijanjikan Allah kepada orang-orang yang sabar dan beriman. Dalam ayat ini, Allah menjanjikan perlindungan kepada mereka dari kesulitan pada Hari Kiamat dan memberikan kepada mereka cahaya kebahagiaan serta kegembiraan. Mereka akan mendapatkan surga sebagai imbalan atas kesabaran yang telah mereka tunjukkan selama hidup di dunia, di mana mereka akan mengenakan pakaian dari sutera yang lembut dan menikmati berbagai kenikmatan yang disediakan untuk mereka. Pada ayat-ayat ini ditekankan tentang pentingnya kesabaran dalam menghadapi ujian dan menunjukkan bahwa Allah akan memberikan balasan yang mulia bagi hamba-Nya yang tetap teguh dalam iman.

Kesimpulan dari ayat ini menegaskan bahwa kesabaran dan keteguhan iman dalam menghadapi tantangan hidup akan mendapatkan ganjaran yang luar biasa di akhirat. Allah tidak hanya melindungi mereka dari kesusahan pada Hari Kiamat, tetapi juga memberikan kenikmatan yang abadi di surga, sebagai imbalan atas pengorbanan dan keikhlasan mereka. Oleh karena itu, ayat ini mengingatkan tentang pentingnya mempertahankan keimanan dan kesabaran dalam menjalani kehidupan, karena setiap usaha yang dilakukan dalam jalan-Nya akan dibalas dengan kemuliaan dan kebahagiaan yang tiada tara.

2.3 Kajian Topik Dengan Teori Pendukung

Kebakaran hutan adalah salah satu bencana alam yang sering terjadi di Indonesia, terutama pada musim kemarau. Kebakaran ini tidak hanya menyebabkan kerugian ekologis dan ekonomi yang besar, tetapi juga menimbulkan dampak kesehatan dan sosial yang signifikan. Oleh karena itu, identifikasi dan klasifikasi kejadian kebakaran hutan sangat penting untuk pencegahan dan mitigasi.

Data *mining* adalah proses penemuan pola dalam data besar yang melibatkan metode dari berbagai bidang seperti statistik, *machine learning*, dan database. Dalam konteks ini, algoritma *Naïve Bayes* digunakan sebagai metode klasifikasi dalam Data *mining* untuk memprediksi kejadian kebakaran hutan. Algoritma *Naïve Bayes* adalah salah satu algoritma klasifikasi yang paling sederhana namun sangat efektif, berdasarkan Teorema Bayes dengan asumsi independensi antar fitur.

Untuk mengklasifikasikan dan memprediksi kejadian kebakaran hutan, berbagai fitur lingkungan dan cuaca digunakan, seperti:

1. Curah Hujan: Pengurangan curah hujan dapat meningkatkan risiko kebakaran.
2. Suhu Udara: Suhu yang lebih tinggi berhubungan dengan peningkatan risiko kebakaran.
3. Kelembaban Tanah dan Udara: Kelembaban rendah meningkatkan risiko kebakaran.
4. Kecepatan Angin: Angin kencang dapat menyebarkan api lebih cepat.
5. Hembusan Angin: Hembusan angin yang tiba-tiba dan kuat bisa memperparah penyebaran api dan mempercepat kebakaran.

6. Titik Embun: Titik embun yang lebih rendah menunjukkan udara yang lebih kering, yang dapat meningkatkan risiko kebakaran.
7. *Longitude*: Koordinat bujur lokasi yang mempengaruhi faktor risiko kebakaran berdasarkan posisi geografis.
8. *Latitude*: Koordinat lintang lokasi yang turut memengaruhi kerentanan area terhadap kebakaran.

Pembahasan dalam skripsi ini akan berfokus pada:

1. Analisis hasil implementasi algoritma *Naïve Bayes* .
2. Implikasi hasil penelitian terhadap kebijakan dan strategi pencegahan kebakaran hutan di Indonesia.

Dengan landasan teori yang kuat ini, penelitian diharapkan dapat memberikan kontribusi yang signifikan dalam upaya mitigasi kebakaran hutan melalui pemanfaatan teknologi data *mining* dan algoritma *Naïve Bayes*.

BAB III

METODE PENELITIAN

3.1 Jenis Penelitian

Penelitian ini menggunakan metode kuantitatif untuk mengklasifikasikan kejadian kebakaran hutan di Indonesia dengan mengimplementasikan algoritma *Naïve Bayes*. Metode kuantitatif dipilih karena fokus penelitian ini adalah pada pengambilan dan analisis data numerik yang berkaitan dengan faktor-faktor pemicu kebakaran hutan. Pendekatan ini memungkinkan peneliti untuk mengidentifikasi pola dan hubungan antar variabel secara objektif dan terukur, serta menghasilkan model prediksi yang akurat untuk kejadian kebakaran hutan. Dengan demikian, penelitian ini berkontribusi dalam menyediakan alat bantu yang dapat digunakan untuk pencegahan dan penanggulangan kebakaran hutan di Indonesia secara lebih efektif dan efisien.

3.2 Data dan Sumber Data

Dalam penelitian ini, data yang digunakan merupakan data sekunder yang diperoleh dari Badan Nasional Penanggulangan Bencana (BNPB). Data ini berisi informasi terkait kejadian kebakaran hutan di Indonesia, mencakup berbagai variabel yang relevan untuk analisis dan klasifikasi kejadian kebakaran hutan. Data BNPB ini secara khusus mencakup titik api yang menunjukkan lokasi terjadinya kebakaran, yang akan dijadikan sebagai label kebakaran hutan. Selain itu, dataset ini juga mencakup beberapa variabel penting, antara lain: garis lintang, garis bujur, suhu (°C), curah hujan (MM), kelembaban, titik embun (°C), kecepatan angin

(Kmph), dan hembusan angin (Kmph). Selanjutnya, data titik api dari BNPB akan digabungkan dengan data cuaca lainnya yang berada di luar titik api tersebut yang didapat dari Open Meteo, kemudian data cuaca dari Open Meteo akan dijadikan sebagai label bukan kebakaran hutan. Data di luar titik api dari Open Meteo juga diambil dalam rentang *latitude* -11 hingga 6 dan *longitude* 95 hingga 141 untuk memastikan bahwa data tersebut benar-benar mencakup wilayah daratan Indonesia. Data kebakaran Indonesia dapat diakses pada Lampiran 1.

Tabel 3.1 Fitur- Fitur Data Kebakaran Hutan

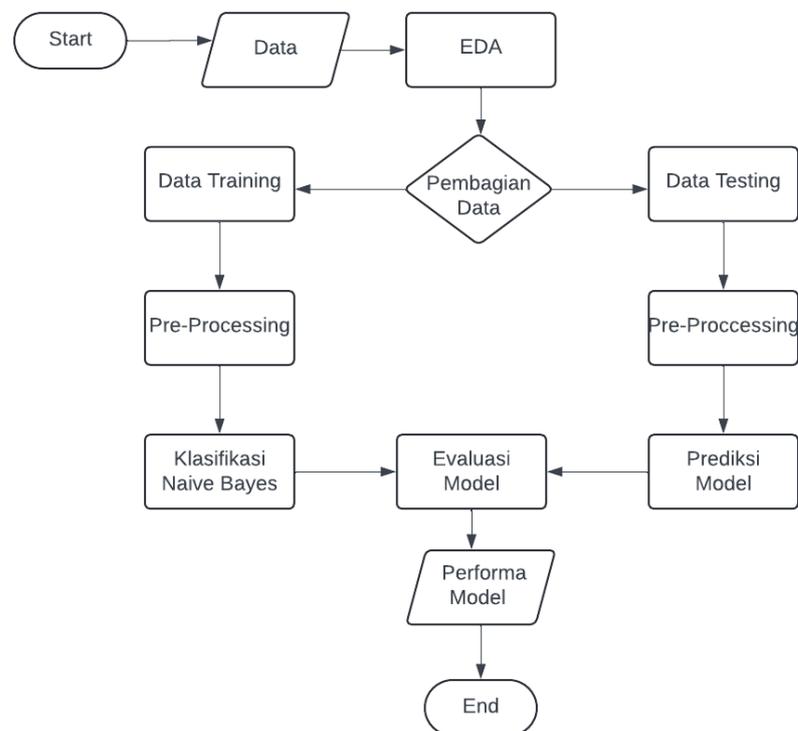
Variabel Fitur	Nama Fitur	Deskripsi Fitur
X1	Lat	Koordinat lintang lokasi kebakaran hutan
X2	Lon	Koordinat bujur lokasi kebakaran hutan
X3	tempC	Suhu udara di lokasi kebakaran hutan dalam derajat Celsius.
X4	humidity	Persentase kelembapan relatif udara di lokasi kebakaran hutan.
X5	DewPointC	Titik embun udara di lokasi kebakaran hutan dalam derajat Celsius.
X6	precipMM	Jumlah curah hujan dalam milimeter di lokasi kebakaran hutan.
X7	windspeedKmph	Kecepatan angin di lokasi kebakaran hutan dalam kilometer per jam.
X8	WindGustKmph	Kecepatan angin maksimum (gust) di lokasi kebakaran hutan dalam kilometer per jam.

Data yang merupakan gabungan dari informasi titik api dari BNPB dan data cuaca dari Open Meteo ini memiliki validitas dan reliabilitas tinggi, oleh karenanya dapat dijadikan sebagai landasan yang kokoh untuk analisis penelitian. Penggunaan

data sekunder gabungan ini memungkinkan penelitian dilakukan dengan lebih efisien dan efektif, mengingat keterbatasan waktu dan sumber daya. Selain itu, data yang sudah terverifikasi dari sumber-sumber tersebut mendukung peningkatan akurasi dan kredibilitas hasil penelitian.

3.3 Teknik Analisis Data

Berikut alur dari penelitian implementasi algoritma *Naïve Bayes* untuk klasifikasi kejadian kebakaran hutan di Indonesia :



Gambar 3.1 Alur Penelitian

1. Pengambilan Data

Data sekunder diambil dari Badan Nasional Penanggulangan Bencana (BNPB) yang merupakan titik api terjadinya kebakaran hutan, dan kemudian

digabungkan dengan data cuaca lain yang berada diluar titik api terjadinya kebakaran hutan yang diambil dari website Open Meteo.

2. *Exploratory Data Analysis* (EDA)

Pada tahap *Exploratory Data Analysis* (EDA), dilakukan analisis deskriptif terhadap statistik data untuk memahami distribusi, pusat, dan sebaran nilai dari setiap fitur. Selain itu, korelasi antar fitur juga diuji menggunakan metode korelasi Pearson untuk mengidentifikasi hubungan linier antar fitur, yang dapat membantu dalam menentukan adanya kesamaan atau tumpang tindih informasi di antara variabel-variabel dalam data.

3. Pembagian Data

Pada proses ini data dibagi menjadi dua subset: data *training* dan data *testing*. Pembagian ini biasanya dilakukan dengan rasio tertentu (misalnya 85% untuk *training* dan 15% untuk *testing*) untuk memastikan bahwa model dapat belajar dengan baik dari data *training* dan diuji pada data *testing*.

4. Data *training*

Data *training* digunakan untuk melatih model klasifikasi agar dapat mengenali pola dalam data. Model akan memanfaatkan data *training* untuk mempelajari hubungan antara fitur-fitur dalam data dan label atau kelas yang ingin diprediksi. Ini adalah dasar agar model dapat membuat prediksi yang akurat pada data yang baru.

5. Data *testing*

Data *testing* digunakan untuk mengevaluasi kinerja model setelah pelatihan. Data *testing* tidak digunakan dalam proses pelatihan, sehingga dapat memberikan gambaran seberapa baik model dapat memprediksi data baru

yang belum pernah dilihat sebelumnya. Evaluasi pada data *testing* menunjukkan keakuratan model dalam situasi nyata.

6. *Pre-processing*

Pada tahap ini, data *training* dan data *testing* masing-masing melalui proses *preprocessing* yang mencakup dua langkah utama: standardisasi data dan reduksi dimensi menggunakan *Principal Component Analysis* (PCA). Proses standardisasi dilakukan untuk menstandarisasi nilai fitur dengan mengubah data agar memiliki rata-rata nol dan standar deviasi satu. Langkah ini penting agar model tidak bias terhadap fitur yang memiliki skala berbeda, sehingga dapat meningkatkan akurasi prediksi. Selain itu, dilakukan reduksi dimensi dengan teknik *Principal Component Analysis* (PCA) untuk mengatasi masalah korelasi tinggi antar fitur. Dengan PCA, fitur-fitur yang berkorelasi tinggi dapat digabungkan menjadi beberapa komponen utama, sehingga memperkecil dimensi data tanpa kehilangan informasi penting. Untuk menggabungkan kedua langkah ini, standardisasi dan PCA dalam satu alur, penelitian ini menggunakan *pipeline*. *Pipeline* adalah alat yang memungkinkan setiap tahap *preprocessing* hingga pemodelan dilakukan secara terstruktur dan otomatis dalam satu kesatuan. Dengan *pipeline*, setiap langkah standardisasi, PCA, dan model *Naïve Bayes* diterapkan secara berurutan pada data. Ini memastikan bahwa setiap tahap pemrosesan diterapkan secara konsisten baik pada data *training* maupun *testing*.

7. Klasifikasi *Naïve Bayes*

Pada tahap ini, model Gaussian *Naïve Bayes* digunakan untuk melakukan klasifikasi. Model ini menghitung probabilitas dari setiap kelas berdasarkan

data *training*, dan menentukan kelas yang paling mungkin untuk setiap contoh dalam data *testing*. Metode ini sederhana namun efektif, terutama untuk data yang memiliki distribusi probabilitas. Selain itu, untuk memastikan model lebih robust dan tidak *overfitting*, dilakukan juga proses *cross-validation*. Teknik *cross-validation* yang digunakan adalah *5-fold cross-validation*. Dalam metode ini, data dibagi menjadi 5 bagian (*folds*) yang sama besar. Pada setiap iterasi, satu *fold* digunakan sebagai data *testing*, sementara 4 *fold* lainnya digunakan sebagai data *training*. Proses ini diulang sebanyak 5 kali, sehingga setiap *fold* digunakan sekali sebagai data *testing*. Penggunaan 5 *fold* dipilih karena memberikan keseimbangan yang baik antara variansi dan bias, efisien dalam waktu komputasi, merupakan praktik standar dalam penelitian, dan cocok untuk dataset ukuran sedang, sehingga memberikan hasil evaluasi yang stabil dan konsisten tanpa meningkatkan beban komputasi yang berlebihan.

8. Prediksi Model

Setelah pelatihan, model melakukan prediksi pada data *testing*. Pada tahap ini, model memprediksi kelas atau kategori untuk setiap data dalam subset data *testing*, berdasarkan pola yang telah dipelajari selama tahap *training*. Hasil prediksi ini kemudian digunakan untuk evaluasi.

9. Evaluation

Evaluasi model dilakukan menggunakan *Confusion Matrix*, yaitu tabel yang menunjukkan jumlah prediksi benar dan salah untuk setiap kelas. Dari *Confusion Matrix*, bisa di hitung metrik seperti akurasi, presisi, dan recall,

yang memberikan pemahaman mendalam tentang kinerja model dalam mengklasifikasikan data secara benar dan menangani kesalahan prediksi.

BAB IV

HASIL DAN PEMBAHASAN

4.1 *Exploratory Data Analysis*

Exploratory Data Analysis (EDA) adalah proses analisis statistik yang bertujuan untuk mengidentifikasi pola, hubungan, serta karakteristik utama dalam dataset secara deskriptif dan visual. EDA membantu mengungkap informasi penting dalam data kebakaran hutan, seperti distribusi variabel yang mungkin berperan dalam kejadian kebakaran, misalnya faktor cuaca, kelembapan, atau suhu.

4.1.1 Deskripsi Statistik

Salah satu metode utama dalam EDA adalah mendeskripsikan data secara statistik. Melalui deskripsi statistik ini, dapat dipahami variasi dan kecenderungan dalam data yang berpotensi memengaruhi persebaran kebakaran hutan, sehingga dapat digunakan sebagai dasar untuk analisis lebih lanjut atau pemodelan prediktif. Berikut merupakan langkah perhitungan untuk mengetahui nilai rata rata dan standard deviasi menggunakan salah satu fitur dalam dataset Lampiran 1 yakni, *humadity*.

1. Rata Rata, dihitung menggunakan Persamaan (2.2).

$$\begin{aligned}\mu &= \frac{x_1 + x_2 + x_3 + \dots + x_{2769}}{2769} \\ &= \frac{92 + 90 + 87 + \dots + 76,5}{2769} \\ &= 78,576\end{aligned}$$

Pada perhitungan ini, rumus diterapkan untuk mendapatkan nilai rata-rata fitur *humidity*. Dari seluruh data sebanyak 2769 sampel, diperoleh nilai rata-rata sebesar 78,576. Hal ini menunjukkan bahwa tingkat *humidity* rata-rata di wilayah penelitian cenderung tinggi, mendekati 79%. Nilai ini dapat menjadi indikator kondisi lingkungan yang memiliki *humidity* cukup tinggi, yang dapat memengaruhi risiko kebakaran hutan, mengingat tingkat *humidity* berperan dalam menjaga kestabilan suhu dan menghambat penyebaran api.

2. Standar Deviasi, dihitung menggunakan Persaman (2.3).

$$\begin{aligned}\sigma &= \sqrt{\frac{(x_1 - \mu)^2 + (x_2 - \mu)^2 + (x_3 - \mu)^2 + \dots + (x_{2769} - \mu)^2}{2769}} \\ &= \sqrt{\frac{(92 - 78,576)^2 + (90 - 78,576)^2 + (87 - 78,576)^2 + \dots + (76,5 - 78,576)^2}{2769}} \\ &= 11,021\end{aligned}$$

dengan,

μ : Rata rata dari fitur *humadity*

σ : Standar deviasi

σ^2 : Variansi

Pada perhitungan ini, rumus diterapkan untuk fitur *humidity* guna menentukan seberapa besar variasi nilai data *humidity* dari rata-ratanya (μ). Nilai standar deviasi untuk fitur kelembapan (*humidity*) sebesar 11,021 menunjukkan adanya tingkat penyebaran atau variasi data kelembapan di sekitar nilai rata-rata sebesar 78,576. Standar deviasi ini mengindikasikan

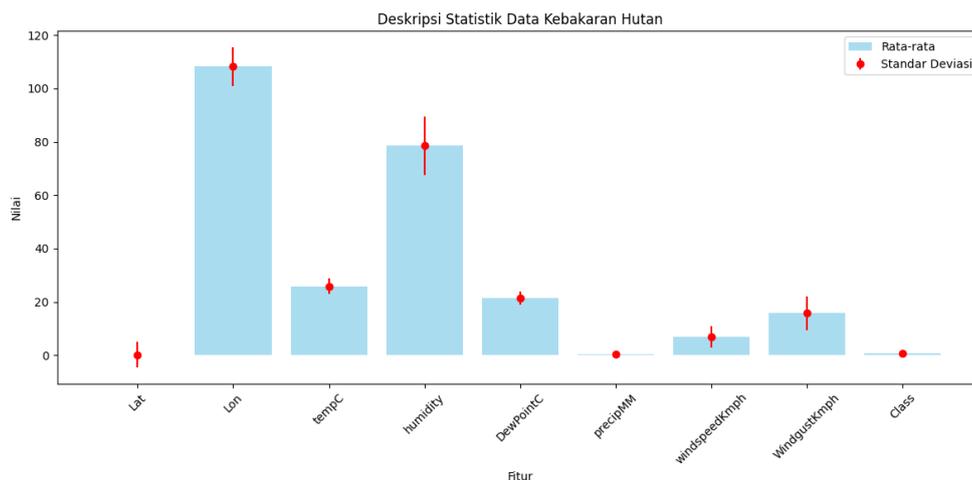
bahwa sebagian besar data kelembapan berkisar dalam rentang $\pm 11,021$ dari nilai rata-rata, yaitu antara sekitar 67,555 hingga 89,597. Perhitungan ini akan dilakukan pada semua fitur dalam dataset sehingga setiap fitur memiliki nilai standar deviasi yang menunjukkan tingkat penyebaran atau variabilitasnya masing-masing.

Berikut merupakan deskripsi statistik pada data kebakaran yang mencakup jumlah data, max, min, rata rata, dan standar deviasi:

Tabel 4.1 Deskripsi Statistik pada Data Kebakaran Hutan di Indonesia

Fitur	Nilai Tertinggi	Nilai Terendah	Rata-Rata	Std. Deviasi	Jumlah Data
Lat	5,988	-10,769	0,148	4,782	2769
Lon	130,368	95,331	108,207	7,336	2769
tempC	31,500	12,200	25,883	2,943	2769
humidity	100	40	78,576	11,021	2769
DewPointC	26,300	8,900	21,474	2,496	2769
precipMM	12,500	0	0,222	0,751	2769
windspeedKmph	30,800	0	6,751	3,988	2769
WindgustKmph	51	1,500	15,701	6,463	2769
Class	1	0	0,648	0,477	2769

Deskripsi statistik data kebakaran hutan di Indonesia menunjukkan bahwa rata-rata latitude sebesar 0,148 menunjukkan kebakaran cenderung dekat garis ekuator, sedikit di selatan, sementara rata-rata longitude sebesar 108,207 mengindikasikan persebaran ke arah barat Indonesia. Dengan standar deviasi latitude 4,782 dan longitude 7,336, titik kebakaran tersebar luas di seluruh wilayah Indonesia, tidak terkonsentrasi di satu lokasi atau bagian tertentu. Variasi besar pada kedua fitur ini menunjukkan kebakaran terjadi di berbagai wilayah Indonesia.



Gambar 4.1 Grafik Deskripsi Statistik

Berdasarkan Gambar 4.1, terdapat perbedaan signifikan pada rata-rata dan standar deviasi antar fitur. Fitur longitude memiliki rata-rata tinggi (108,207) karena cakupan geografis Indonesia, sementara latitude lebih rendah (0,148), mengindikasikan kedekatannya dengan garis ekuator. Fitur suhu (tempC) menunjukkan rata-rata 25,883°C dengan variasi moderat (2,943), mencerminkan kondisi tropis Indonesia.

Kelembaban (humidity) memiliki rata-rata tinggi (78,576) dan variasi besar (11,021), menunjukkan perbedaan kelembaban di berbagai daerah kebakaran. Curah hujan (precipMM) sangat rendah (0,222 mm) karena kebakaran hutan umumnya terjadi di daerah dengan curah hujan minim.

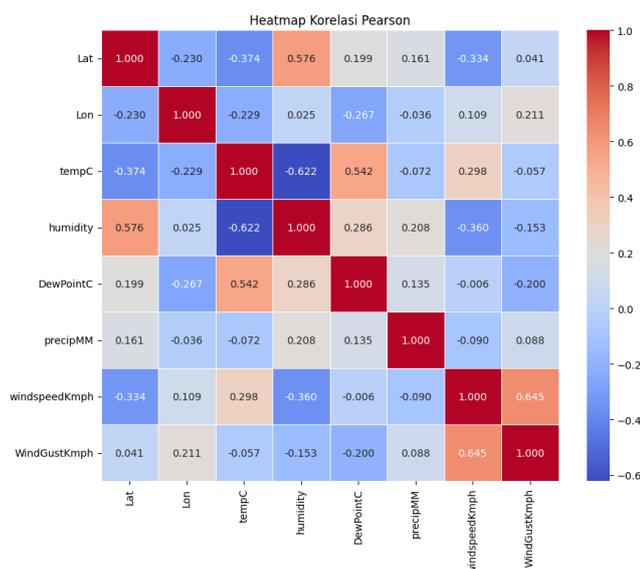
Kecepatan angin (windspeedKmph) rata-rata 6,751 km/jam, sedangkan hembusan angin (wind gust) jauh lebih tinggi (15,701 km/jam), dengan variasi besar, menunjukkan bahwa hembusan angin dapat mempercepat penyebaran kebakaran. Perbedaan ini mencerminkan kondisi lingkungan yang bervariasi, dipengaruhi faktor geografis, cuaca, dan waktu.

4.1.2 Uji Korelasi Pearson

Dalam proses *Exploratory Data Analysis* (EDA), uji korelasi Pearson digunakan untuk memahami hubungan linier antar variabel dalam dataset. Uji ini penting karena membantu mengidentifikasi pola atau hubungan yang mungkin signifikan antara variabel-variabel, yang nantinya dapat digunakan untuk menentukan fitur yang paling relevan dalam model klasifikasi. Pada tahap ini, korelasi Pearson dihitung untuk mengukur kekuatan dan arah hubungan antar fitur.

Selain itu, untuk menilai apakah hubungan tersebut signifikan secara statistik, digunakan nilai *p-value*. *P-value* digunakan untuk mengukur probabilitas bahwa hubungan antar fitur terjadi secara kebetulan atau tidak.

Pada Gambar 4.2 ditampilkan hasil nilai korelasi Pearson antara fitur-fitur dalam dataset. Sementara itu, Tabel 4.2 menyajikan nilai *p-value* dari uji korelasi Pearson untuk menguji signifikansi hubungan antar fitur. Keduanya digunakan untuk memahami tingkat hubungan linier dan signifikansinya sehingga fitur-fitur yang relevan dapat diidentifikasi untuk analisis lebih lanjut.



Gambar 4.2 Heatmap Pearson Correlation Matrix

Tabel 4.2 Nilai *P-value* Signifikansi Uji Korelasi Pearson

P-Value	Lat	Lon	tempC	humidity	DewPointC	precipMM	windspeedKmph	WindGustKmph
Lat	0	1,410E-34	7,861E-93	7,163E-245	2,689E-26	1,181E-17	2,477E-73	2,974E-02
Lon	1,410E-34	0	2,001E-34	1,720E-01	2,036E-46	5,443E-02	7,480E-09	2,774E-29
tempC	7,861E-93	2,001E-34	0	2,500E-296	1,296E-211	1,224E-04	5,907E-58	2,472E-03
humidity	7,163E-245	1,720E-01	2,500E-296	0	1,633E-53	1,620E-28	7,105E-86	3,897E-16
DewPointC	2,689E-26	2,036E-46	1,296E-211	1,633E-53	0	8,674E-13	7,188E-01	1,681E-26
precipMM	1,181E-17	5,443E-02	1,224E-04	1,620E-28	8,674E-13	0	2,040E-06	2,927E-06
windspeedKmph	2,477E-73	7,480E-09	5,907E-58	7,105E-86	7,188E-01	2,040E-06	0	0
WindGustKmph	2,974E-02	2,774E-29	2,472E-03	3,897E-16	1,681E-26	2,927E-06	0	0

Berdasarkan Gambar 4.2 dan Tabel 4.2, ditemukan bahwa beberapa fitur memiliki hubungan yang signifikan. Misalnya, fitur *Latitude* (Lat) dan Humidity memiliki hubungan linier positif yang signifikan dengan nilai korelasi $r = 0,576$ dan $p\text{-value} = 7,163 \times 10^{-245}$ ($p < 0,05$). Hal ini menunjukkan bahwa semakin tinggi *latitude*, tingkat kelembapan (humidity) cenderung meningkat. Hubungan linier positif yang signifikan juga ditemukan antara WindSpeedKmph dan WindGustKmph dengan $r = 0,645$ dan $p\text{-value} = 0$ ($p < 0,05$), yang mengindikasikan bahwa kecepatan angin berkorelasi erat dengan hembusan angin maksimum. Sebaliknya, hubungan linier negatif yang signifikan ditemukan antara Temperature (tempC) dan Humidity dengan $r = -0,622$ dan $2,500 \times 10^{-296}$ ($p < 0,05$), menunjukkan bahwa semakin tinggi suhu, kelembapan cenderung menurun. Hubungan linier negatif signifikan yang lain terlihat pada *Latitude* (Lat) dan Temperature (tempC), di mana $r = -0,374$ dan $p\text{-value} = 7,861 \times 10^{-93}$ ($p < 0,05$), menandakan bahwa semakin tinggi *latitude*, suhu cenderung lebih rendah.

Namun, tidak semua fitur memiliki hubungan linier yang signifikan. Sebagai contoh, hubungan antara *Longitude* (Lon) dan *Precipitation* (precipMM) memiliki nilai korelasi $r = -0,036$ dan $p\text{-value} = 5,443 \times 10^{-2}$ ($p > 0,05$), sehingga hubungan tersebut tidak signifikan secara statistik. Ini menunjukkan bahwa tidak ada hubungan linier yang berarti antara kedua fitur tersebut.

Sebagai kesimpulan, uji korelasi Pearson menunjukkan bahwa terdapat pasangan fitur yang berhubungan linier secara signifikan dalam dataset penelitian ini. Hubungan korelasi ini mengindikasikan keterkaitan yang kuat di antara beberapa fitur, yang dapat memengaruhi analisis lebih lanjut. Hasil uji juga menunjukkan bahwa hanya hubungan dengan nilai signifikansi ($p < 0,05$) yang dapat dianggap nyata secara statistik. Pemahaman terhadap korelasi dan signifikansi ini membantu dalam memilih fitur yang relevan serta mengidentifikasi interaksi antar variabel yang berpotensi memengaruhi performa model klasifikasi.

4.2 Data Preprocessing

4.2.1 Standardisasi Data

Sebelum melakukan standardisasi, data terlebih dahulu dibagi menjadi data *training* dan data *testing* dengan rasio 85:15. Rasio ini dipilih karena menunjukkan akurasi yang tinggi dalam pengujian model. Dari pembagian ini, diperoleh data *training* sebanyak 2.353 data dan data *testing* sebanyak 416 data. Pembagian data menjadi *training* dan *testing* sebelum dilakukannya proses standardisasi ini dilakukan untuk memastikan tidak adanya kebocoran informasi dari data *testing* ke dalam proses *training*. Dengan demikian, data *testing* benar-benar hanya digunakan untuk mengukur performa model secara objektif.

Setelah pembagian data, dilakukan proses standardisasi sebagai bagian dari preprocessing. Standardisasi bertujuan untuk memastikan bahwa setiap fitur memiliki skala yang sama, sehingga dapat menghilangkan perbedaan skala antar fitur dan meningkatkan efektivitas algoritma *machine learning*, terutama yang sensitif terhadap perbedaan skala. Proses standardisasi dilakukan hanya pada data

training, sehingga nilai rata-rata dan standar deviasi dihitung berdasarkan data tersebut. Data *testing* kemudian hanya ditransformasi menggunakan informasi dari data *training*. Dengan cara ini, data *testing* tidak menggunakan informasi statistik yang dihitung dari data *testing* itu sendiri, sehingga hasil pengujian mencerminkan kemampuan model yang realistis. Deskripsi Statistik pada data *training* ditunjukkan pada Tabel 4.3.

Tabel 4.3 Deskripsi Statistik pada Data *Training*

Fitur	Nilai Tertinggi	Nilai Terendah	Rata-Rata	Std. Deviasi	Jumlah Data
X1	5,988	-10,769	0,148	4,769	2353
X2	130,369	95,331	108,207	7,369	2353
X3	31,500	12,200	25,900	2,947	2353
X4	100	40	78,585	11,020	2353
X5	26,300	8,900	21,489	2,486	2353
X6	12,500	0	0,228	0,768	2353
X7	29	0	6,796	3,977	2353
X8	51	3,200	15,732	6,402	2353

Standardisasi dilakukan dengan cara mengubah setiap nilai pada fitur menjadi nilai yang memiliki rata-rata mendekati 0 dan standar deviasi mendekati

1. Rumus yang digunakan merujuk pada Persamaan (2.4), sehingga:

$$\hat{x}_1 = \frac{97 - 78,584}{11,020} = \frac{18,416}{11,020} = 1,671$$

$$\hat{x}_2 = \frac{79,500 - 78,584}{11,020} = \frac{0,918}{11,020} = 0,083$$

$$\hat{x}_3 = \frac{83 - 78,584}{11,020} = \frac{4,416}{11,020} = 0,401$$

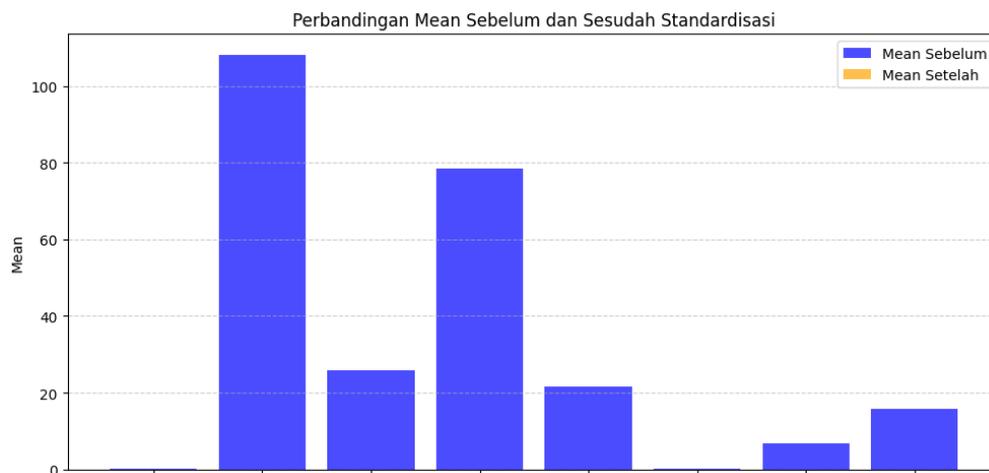
$$\vdots = \vdots$$

$$\hat{x}_{2353} = \frac{98 - 78,584}{11,020} = -\frac{19,416}{11,020} = 1,762$$

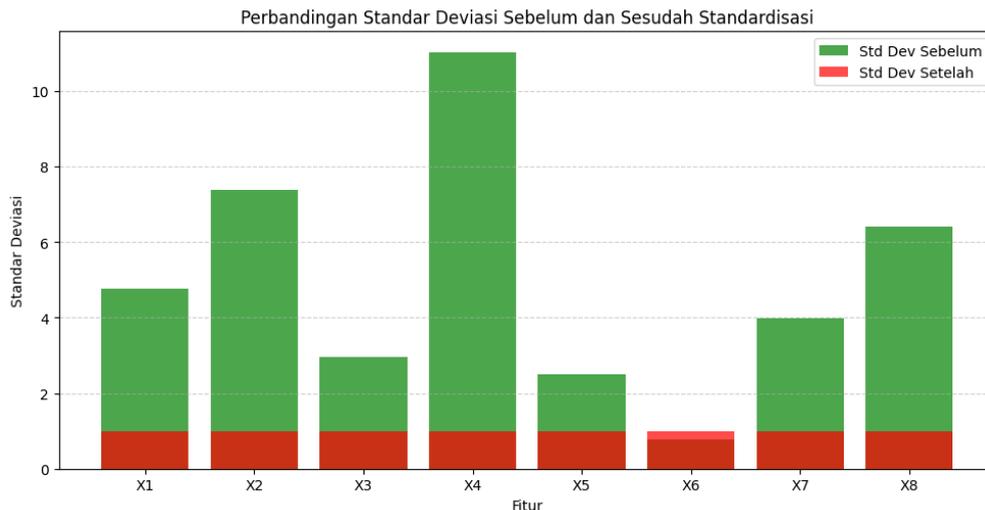
Dalam perhitungan di atas, dipilih fitur X4 (*humidity*) sebagai ilustrasi untuk menunjukkan proses standardisasi data. Nilai asli pada fitur X4 (*humidity*) dikonversi menggunakan rumus standardisasi sehingga memiliki rata-rata mendekati 0 dan standar deviasi mendekati 1. Perhitungan rata-rata (μ) dan standar deviasi (σ) dilakukan berdasarkan data *training* fitur X4. Nilai-nilai pada fitur tersebut kemudian distandardisasi menggunakan informasi ini. Proses ini berlaku untuk semua fitur dalam dataset, di mana setiap fitur akan distandardisasi menggunakan rata-rata dan standar deviasi dari data *training* masing-masing. Tabel 4.4 menunjukkan perbandingan nilai rata-rata dan standar deviasi sebelum dan setelah standardisasi.

Tabel 4.4 Tampilan Sebelum dan Sesudah Standardisasi

Fitur	Mean Sebelum	Std Dev Sebelum	Mean Setelah	Std Dev Setelah
X1	0,148	4,769	2,053E-10	1,000213
X2	108,207	7,369	4,016E-09	1,000213
X3	25,899	2,946	2,810E-10	1,000213
X4	78,584	11,020	-1,865E-10	1,000213
X5	21,489	2,486	2,742E-09	1,000213
X6	0,228	0,768	-1,238E-10	1,000213
X7	6,796	3,977	2,719E-09	1,000213
X8	15,731	6,401	1,354E-09	1,000213



Gambar 4.3 Grafik Mean Sebelum dan Sesudah Standardisasi



Gambar 4.4 Grafik Std. Deviasi Sebelum dan Sesudah Standardisasi

Gambar 4.3 menunjukkan bahwa sebelum standardisasi, nilai rata-rata (mean) untuk setiap fitur sangat bervariasi, mulai dari angka kecil seperti 0,148 pada fitur X1 hingga angka besar seperti 108,207 pada fitur X2. Namun, setelah proses standardisasi, rata-rata seluruh fitur menjadi mendekati nol, yang merupakan karakteristik utama data yang telah dinormalisasi dengan teknik standardisasi.

Gambar 4.4 memperlihatkan hal serupa. Sebelum standardisasi, standar deviasi setiap fitur berbeda secara signifikan, misalnya fitur X4 memiliki standar deviasi sebesar 11,020, sedangkan fitur X6 hanya 0,768. Setelah standardisasi, seluruh fitur memiliki standar deviasi yang seragam, yaitu mendekati nilai 1, yang menandakan bahwa penyebaran data telah distandardisasi dalam skala yang sama.

Sehingga dapat disimpulkan bahwa standardisasi berhasil mengubah skala data menjadi seragam, serta menjadikan rata-rata mendekati 0 dan standar deviasi mendekati 1, yang akan mempermudah dalam penerapan analisis lebih lanjut seperti PCA. Dengan data yang sudah distandardisasi, diharapkan model *machine*

learning yang digunakan dapat memberikan hasil yang lebih akurat. Hasil data setelah melalui proses standardisasi terletak pada Lampiran 4 dan Lampiran 5.

4.2.2 *Principal Component Analysis (PCA)*

Principal Component Analysis (PCA) adalah metode statistik yang digunakan untuk mereduksi dimensi data dengan mentransformasikan variabel-variabel awal menjadi sejumlah komponen utama. Setiap komponen utama merupakan kombinasi linear dari variabel asli dan dirancang untuk menangkap sebanyak mungkin variansi dalam data. Dalam penelitian ini, PCA diterapkan untuk menyederhanakan data tanpa kehilangan informasi yang signifikan. Selain mempercepat proses komputasi dan mengurangi risiko multikolinearitas, PCA juga digunakan sebagai solusi untuk menangani fitur-fitur yang memiliki korelasi tinggi.

Dalam penelitian ini, PCA diterapkan pada data yang telah distandardisasi untuk menghindari bias akibat perbedaan skala antar variabel. Variansi kumulatif sebesar 95% dipilih sebagai batas untuk menentukan jumlah komponen utama yang dipertahankan, karena nilai ini dianggap cukup untuk mempertahankan informasi yang signifikan dari data asli. Dengan mempertahankan 95% variansi, analisis dapat mengurangi jumlah dimensi tanpa kehilangan detail penting, sehingga transformasi data ke dalam ruang komponen utama tetap mewakili karakteristik data secara optimal.

Tabel 4.5 menunjukkan persentase variansi yang dijelaskan oleh setiap komponen utama serta variansi kumulatifnya. Berdasarkan tabel tersebut, terlihat bahwa komponen pertama mampu menjelaskan sebesar 30,616% dari variansi

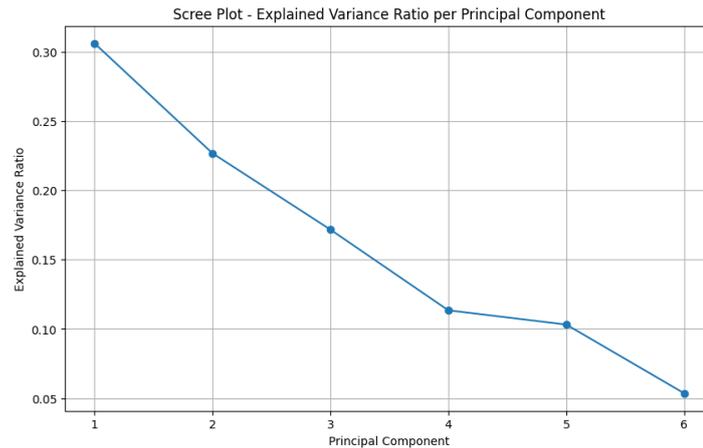
data, sedangkan komponen kedua menambahkan 22,678%, sehingga total variansi kumulatif menjadi 53,294%. Hingga komponen keenam, variansi kumulatif mencapai 97,255%, yang cukup untuk mempertahankan informasi yang signifikan dari data asli.

Tabel 4.5 Presentase Variansi Komponen dan Variansi Kumulatif

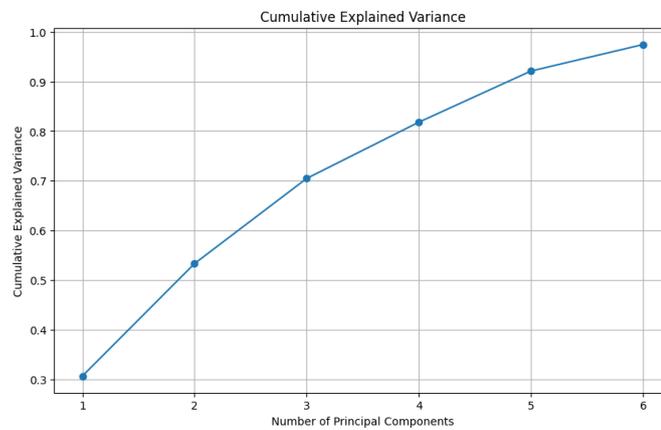
Komponen Utama	Variansi Komponen (%)	Variansi Kumulatif (%)
Komponen 1	30,616	30,616
Komponen 2	22,678	53,294
Komponen 3	17,167	70,461
Komponen 4	11,134	81,595
Komponen 5	10,311	91,906
Komponen 6	5,349	97,255

Gambar 4.5 menunjukkan grafik variansi yang dijelaskan oleh masing-masing komponen utama secara individu. Grafik ini menggambarkan kontribusi setiap komponen terhadap total variansi data. Berdasarkan grafik ini, terlihat bahwa komponen pertama memberikan kontribusi variansi terbesar dibandingkan komponen lainnya, sementara kontribusi variansi menurun secara bertahap pada komponen berikutnya. Informasi ini penting untuk memahami peran masing-masing komponen dalam representasi data yang lebih sederhana.

Gambar 4.6 menunjukkan grafik variansi kumulatif yang dijelaskan oleh setiap komponen utama. Dari grafik ini, terlihat bahwa variansi kumulatif mencapai 95% pada komponen keenam. Hal ini menunjukkan bahwa 6 komponen utama sudah cukup untuk mewakili data dengan baik, tanpa kehilangan informasi yang signifikan.



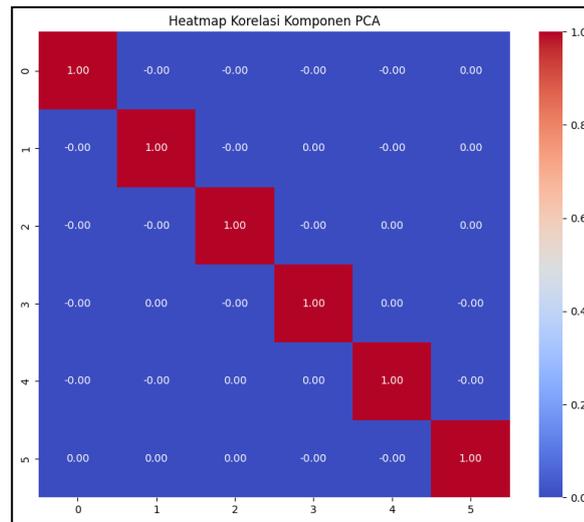
Gambar 4.5 Grafik Variansi Komponen



Gambar 4.6 Grafik Variansi Kumulatif

Berdasarkan Gambar 4.6, diketahui bahwa 6 komponen utama sudah cukup untuk menjelaskan 95% variansi dari data asli. Hal ini berarti dimensi data dapat direduksi dari jumlah fitur awal menjadi 6 komponen tanpa kehilangan informasi yang signifikan.

Kemudian, untuk mengetahui apakah PCA berhasil mengurangi masalah korelasi yang signifikan antar-fitur, dilakukan pengujian kembali korelasi Pearson terhadap komponen-komponen utama yang dihasilkan. Pengujian ini bertujuan untuk memastikan bahwa fitur-fitur baru dalam ruang PCA memiliki hubungan yang lebih independen, sesuai dengan asumsi yang dibutuhkan oleh algoritma *Naïve Bayes*. Hasil nilai korelasi ini disajikan pada Gambar 4.5.



Gambar 4.7 Heatmap Korelasi Komponen PCA

Tabel 4.6 Nilai *P-value* Signifikansi Setelah PCA

P-Value Fitur Setelah PCA	PC1	PC2	PC3	PC4	PC5	PC6
PC1	0	1	1	1	1	1
PC2	1	0	1	1	1	1
PC3	1	1	0	1	1	1
PC4	1	1	1	0	1	1
PC5	1	1	1	1	0	1
PC6	1	1	1	1	1	0

Pada Gambar 4.7 dan Tabel 4.6, terlihat bahwa penerapan PCA berhasil mengurangi masalah korelasi yang signifikan antar-fitur, sehingga menghasilkan komponen-komponen utama yang lebih independen. Batas 95% juga terbukti sudah cukup untuk menjadi solusi bagi fitur yang memiliki korelasi tinggi tanpa menghilangkan banyak informasi penting. Dengan demikian, reduksi dimensi ini tidak hanya meningkatkan efisiensi komputasi, tetapi juga memastikan data yang lebih sesuai untuk analisis dengan algoritma *Naïve Bayes* yang memerlukan independensi antar-fitur. Hasil data setelah melalui proses PCA terletak pada Lampiran 6 dan Lampiran 7.

4.3 Implementasi Algoritma *Naïve Bayes*

Pengklasifikasian dengan algoritma *Naïve Bayes* pada penelitian ini terdiri dari dua proses utama, yaitu proses pelatihan (*training*) dan pengujian (*testing*). Pada proses pelatihan, algoritma *Naïve Bayes* menggunakan data sampel yang terdiri atas variabel-variabel independen, seperti suhu permukaan, kecepatan angin, curah hujan, titik embun, kelembaban, hembusan angin, *longitude*, *latitude* serta kelas target yang ditentukan berdasarkan ada atau tidaknya kebakaran hutan. Kelas target pada data ini terbagi menjadi dua kategori, yaitu terjadi kebakaran hutan (kelas 1) dan tidak terjadi kebakaran hutan (kelas 0).

Pada proses pengujian, algoritma menghitung probabilitas untuk setiap kelas berdasarkan fitur-fitur pada data *testing*, dengan memanfaatkan rumus-rumus probabilistik sesuai dengan asumsi independensi pada setiap variabel. Proses klasifikasi menggunakan *Naïve Bayes* ini menghasilkan probabilitas posterior untuk masing-masing kelas. Data *testing* kemudian diklasifikasikan berdasarkan kelas dengan probabilitas tertinggi. Hasil klasifikasi dievaluasi menggunakan metrik akurasi untuk mengukur tingkat ketepatan model dalam memprediksi data *testing*. Tabel 4.7 di bawah ini menyajikan 4 data dari data *training* dataset kebakaran hutan Indonesia yang telah melalui proses standardisasi dan PCA yang akan digunakan sebagai contoh data *training* untuk implementasi algoritma *Naïve Bayes*. Setelah melalui proses reduksi dimensi menggunakan PCA, nama fitur awal yang berjumlah 8 berubah menjadi fitur baru dengan nama PC1, PC2, PC3, dan seterusnya, karena hanya 6 komponen utama yang dipertahankan. Oleh karena itu, fitur-fitur yang digunakan dalam model ini disederhanakan menjadi PC1 hingga PC6, mewakili informasi utama dari fitur asli.

Tabel 4.7 Contoh Data *Training*

PC1	PC2	PC3	PC4	PC5	PC6	Class
2,476	1,195	0,051	-0,171	0,241	-0,523	0
0,945	1,664	0,691	-0,455	0,126	0,653	0
-2,149	-0,775	-0,22	0,547	0,354	-0,618	1
-0,263	0,012	-1,263	1,288	0,565	-0,865	1

Kemudian sebuah data dipilih secara acak dari data *training* lainnya (di luar data yang tercantum dalam Tabel 4.7) dan diasumsikan sebagai data *testing* untuk diprediksi kelasnya menggunakan Persamaan (2.5). Data yang dipilih memiliki nilai PC1 -2,710, PC2 dengan nilai -1,991, PC3 dengan nilai -0,307, PC4 dengan nilai -0,092, PC5 dengan nilai -0,671, PC6 dengan nilai -0,748.

1. Probabilitas Prior $P(H)$

$$P(H = 0) = \frac{\text{Jumlah kelas 0}}{\text{Total data}} = \frac{2}{4} = 0,5$$

$$P(H = 1) = \frac{\text{Jumlah kelas 1}}{\text{Total data}} = \frac{2}{4} = 0,5$$

2. Rata Rata dan Varians untuk Masing Masing Kelas

Kelas 0 (bukan kebakaran hutan) :

$$\text{PC1} : \mu = \frac{2,475 + 0,945}{2} = 1,711$$

$$\sigma^2 = \frac{(2,475 - 1,711)^2 + (0,945 - 1,711)^2}{2} = 0,586$$

$$\text{PC2} : \mu = \frac{1,195 + 1,664}{2} = 1,430$$

$$\sigma^2 = \frac{(1,195 - 1,430)^2 + (1,664 - 1,430)^2}{2} = 0,055$$

$$\text{PC3} : \mu = \frac{0,051 + 0,691}{2} = 0,371$$

$$\sigma^2 = \frac{(0,051 - 0,371)^2 + (0,691 - 0,371)^2}{2} = 0,102$$

$$\begin{aligned} \text{PC4} & : \mu = \frac{-0,171 - 0,455}{2} = -0,313 \\ \sigma^2 & = \frac{(-0,171 - (-0,313))^2 + (-0,455 - (-0,313))^2}{2} = 0,020 \end{aligned}$$

$$\begin{aligned} \text{PC5} & : \mu = \frac{0,241 + 0,126}{2} = 0,184 \\ \sigma^2 & = \frac{(0,241 - 0,184)^2 + (0,126 - 0,184)^2}{2} = 0,003 \end{aligned}$$

$$\begin{aligned} \text{PC6} & : \mu = \frac{-0,523 + 0,653}{2} = 0,065 \\ \sigma^2 & = \frac{(-0,523 - 0,065)^2 + (0,653 - 0,065)^2}{2} = 0,346 \end{aligned}$$

Kelas 1 (kebakaran hutan) :

$$\begin{aligned} \text{PC1} & : \mu = \frac{-2,149 - 0,263}{2} = -1,206 \\ \sigma^2 & = \frac{(-2,149 - (-1,206))^2 + (-0,263 - (-1,206))^2}{2} = 0,889 \end{aligned}$$

$$\begin{aligned} \text{PC2} & : \mu = \frac{-0,775 + 0,012}{2} = -0,382 \\ \sigma^2 & = \frac{(-0,775 - (-0,382))^2 + (0,012 - (-0,382))^2}{2} = 0,155 \end{aligned}$$

$$\begin{aligned} \text{PC3} & : \mu = \frac{-0,220 - 1,263}{2} = -0,742 \\ \sigma^2 & = \frac{(-0,220 - (-0,742))^2 + (-1,263 - (-0,742))^2}{2} = 0,272 \end{aligned}$$

$$\begin{aligned} \text{PC4} & : \mu = \frac{0,547 + 1,288}{2} = 0,918 \\ \sigma^2 & = \frac{(0,547 - 0,918)^2 + (1,288 - 0,918)^2}{2} = 0,137 \end{aligned}$$

$$\begin{aligned} \text{PC5} & : \mu = \frac{0,354 + 0,565}{2} = 0,460 \\ \sigma^2 & = \frac{(0,354 - 0,460)^2 + (0,565 - 0,460)^2}{2} = 0,011 \end{aligned}$$

$$\text{PC6} : \mu = \frac{-0,618 - 0,865}{2} = -0,742$$

$$\sigma^2 = \frac{(-0,618 - (-0,742))^2 + (-0,865 - (-0,742))^2}{2} = 0,015$$

3. Probabilitas Likelihood $P(X|H)$

Kelas 0 (bukan kebakaran hutan) :

$$PC1 : P(-2,710|H) = \frac{1}{\sqrt{2\pi(0,586)}} \exp\left(-\frac{(-2,710 - 1,711)^2}{2(0,586)}\right) = 2,991 \times 10^{-8}$$

$$PC2 : P(-1,991|H) = \frac{1}{\sqrt{2\pi(0,055)}} \exp\left(-\frac{(-1,991 - 1,430)^2}{2(0,055)}\right) = 1,072 \times 10^{-46}$$

$$PC3 : P(-0,307|H) = \frac{1}{\sqrt{2\pi(0,102)}} \exp\left(-\frac{(-0,307 - 0,371)^2}{2(0,102)}\right) = 0,132$$

$$PC4 : P(-0,092|H) = \frac{1}{\sqrt{2\pi(0,020)}} \exp\left(-\frac{(-0,092 - (-0,313))^2}{2(0,020)}\right) = 0,837$$

$$PC5 : P(-0,671|H) = \frac{1}{\sqrt{2\pi(0,003)}} \exp\left(-\frac{(-0,671 - 0,184)^2}{2(0,003)}\right) = 7,678 \times 10^{-48}$$

$$PC6 : P(-0,748|H) = \frac{1}{\sqrt{2\pi(0,346)}} \exp\left(-\frac{(-0,748 - 0,065)^2}{2(0,346)}\right) = 0,261$$

Kelas 1 (kebakaran hutan) :

$$PC1 : P(-2,710|H) = \frac{1}{\sqrt{2\pi(0,889)}} \exp\left(-\frac{(-2,710 - (-1,206))^2}{2(0,889)}\right) = 0,119$$

$$PC2 : P(-1,991|H) = \frac{1}{\sqrt{2\pi(0,155)}} \exp\left(-\frac{(-1,991 - (-0,382))^2}{2(0,155)}\right) = 2,361 \times 10^{-4}$$

$$PC3 : P(-0,307|H) = \frac{1}{\sqrt{2\pi(0,272)}} \exp\left(-\frac{(-0,307 - (-0,742))^2}{2(0,272)}\right) = 0,541$$

$$PC4 : P(-0,092|H) = \frac{1}{\sqrt{2\pi(0,137)}} \exp\left(-\frac{(-0,092 - 0,918)^2}{2(0,137)}\right) = 0,026$$

$$PC5 : P(-0,671|H) = \frac{1}{\sqrt{2\pi(0,011)}} \exp\left(-\frac{(-0,671 - 0,460)^2}{2(0,011)}\right) = 4,403 \times 10^{-25}$$

$$PC6 : P(-0,748|H) = \frac{1}{\sqrt{2\pi(0,015)}} \exp\left(-\frac{(-0,748 - (-0,742))^2}{2(0,015)}\right) = 3,226$$

$$P(X|H = 0) = (2,991 \times 10^{-8}) \times (1,072 \times 10^{-46}) \times 0,132 \times 0,837 \\ \times (7,678 \times 10^{-48}) \times 0,261 = 7,100 \times 10^{-103}$$

$$P(X|H = 1) = 0,119 \times (2,361 \times 10^{-4}) \times 0,541 \times 0,026 \\ \times (4,403 \times 10^{-25}) \times 3,226 = 5,656 \times 10^{-31}$$

4. Probabilitas Posterior $P(H|X)$

$$P(H = 0|X) = \frac{P(X|H = 0) \cdot P(H = 0)}{P(X)}$$

$$\begin{aligned} P(H = 0|X) &\propto P(X | H = 0) \cdot P(H = 0) \\ &\propto 7,100 \times 10^{-103} \cdot 0,5 \\ &= 3,550 \times 10^{-103} \end{aligned}$$

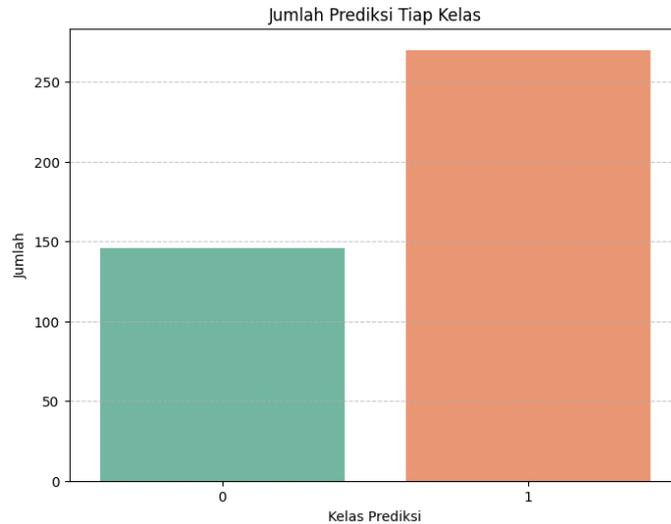
$$P(H = 1|X) = \frac{P(X|H = 1) \cdot P(H = 1)}{P(X)}$$

$$\begin{aligned} P(H = 1|X) &\propto P(X | H = 1) \cdot P(H = 1) \\ &\propto 5,656 \times 10^{-31} \cdot 0,5 \\ &= 2,828 \times 10^{-31} \end{aligned}$$

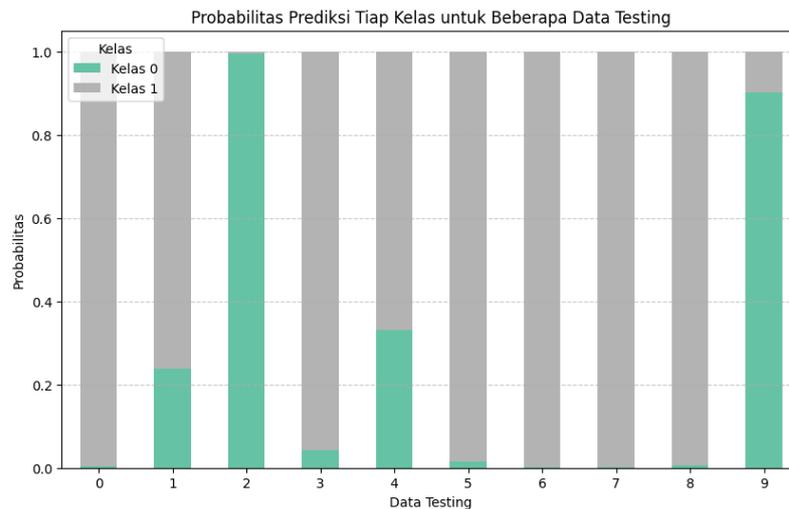
$$P(H = 1|X) = 2,828 \times 10^{-31} > P(H = 0|X) = 3,550 \times 10^{-103}$$

Karena nilai probabilitas posterior untuk kelas 1 lebih tinggi daripada kelas 0, maka data *testing* tersebut diprediksi termasuk dalam kelas 1.

Berikut merupakan visualisasi dari hasil implementasi algoritma *Naive Bayes* pada data *testing*. Gambar 4.8 menunjukkan jumlah prediksi yang dihasilkan oleh model untuk masing-masing kelas, sedangkan Gambar 4.9 memvisualisasikan probabilitas prediksi tiap kelas untuk beberapa sampel data *testing* secara detail. Visualisasi ini bertujuan untuk memberikan gambaran bagaimana model membedakan antara kelas 0 dan kelas 1 berdasarkan probabilitas yang dihitung.



Gambar 4.8 Jumlah Prediksi Tiap Kelas



Gambar 4.9 Probabilitas Prediksi Tiap Kelas Beberapa Data *Testing*

Gambar 4.8 memperlihatkan jumlah sampel yang diprediksi oleh model sebagai kelas 0 dan kelas 1. Dari visualisasi ini, terlihat bahwa kelas 1 memiliki jumlah prediksi yang lebih banyak dibandingkan kelas 0. Hal ini menunjukkan distribusi hasil klasifikasi model pada data *testing*.

Gambar 4.9 menunjukkan probabilitas prediksi untuk beberapa data *testing* yang dihasilkan oleh model *Naive Bayes*. Setiap bar menunjukkan probabilitas sebuah sampel termasuk ke dalam kelas 0 dan kelas 1. Plot ini menunjukkan

bagaimana keyakinan model dalam memprediksi kelas tertentu untuk masing-masing sampel, dengan probabilitas mendekati 1 menandakan keyakinan yang tinggi.

Setelah menyelesaikan perhitungan manual untuk mendapatkan probabilitas posterior dari model *Naïve Bayes*, dilakukan *cross-validation* pada data *training* sebagai langkah tambahan untuk memastikan bahwa model ini tidak mengalami *overfitting*. Dalam hal ini, *cross-validation* berfungsi untuk menguji apakah model memiliki performa yang stabil di berbagai subset data yang berbeda dan untuk melihat apakah terdapat penurunan akurasi yang signifikan saat data berubah. Dengan metode ini, data *training* dipecah menjadi lima subset atau “*folds*,” dan model dilatih secara berulang di tiap subset, dengan satu subset sebagai data validasi sementara empat lainnya digunakan untuk melatih model.

Penggunaan *5-fold cross-validation* dipilih karena menawarkan keseimbangan antara akurasi pengujian dan efisiensi komputasi, yaitu menghasilkan hasil yang cukup stabil tanpa memerlukan terlalu banyak pengulangan yang dapat membebani proses komputasi. Proses ini diulang sebanyak lima kali untuk memastikan bahwa setiap subset berperan sebagai data validasi satu kali. Hasil dari *5-fold cross-validation* dapat dilihat pada Tabel 4.8.

Tabel 4.8 Hasil *5-fold Cross Validation*

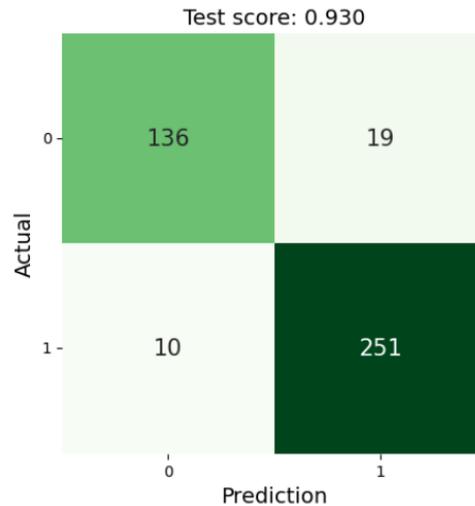
Fold	Cross-Validation Score
1	0,927
2	0,929
3	0,915
4	0,927
5	0,938
Rata-rata	0,927

Rata-rata skor *cross-validation* sebesar 0,927 atau 92,7% menunjukkan bahwa model cenderung stabil di berbagai subset data *training*, yang berarti tidak ada indikasi kuat terjadinya *overfitting* pada data ini. Dengan demikian, hasil *cross-validation* ini membantu memastikan bahwa model memiliki daya prediksi yang baik saat nantinya diuji pada data *testing* yang sesungguhnya, tanpa menunjukkan ketergantungan yang berlebihan pada satu subset data tertentu.

4.4 Evaluasi Performa Model *Naïve Bayes* pada Data *Testing*

Evaluasi performa model *Naïve Bayes* pada data *testing* dilakukan untuk mengetahui seberapa baik model dalam mengklasifikasikan kejadian kebakaran hutan dan bukan kebakaran hutan berdasarkan variabel-variabel yang telah ditetapkan. Dengan menggunakan confusion matrix, tingkat akurasi model dalam mendeteksi kedua kategori ini dapat diidentifikasi. Evaluasi pada data *testing* menjadi penting agar performa model lebih andal saat menghadapi data baru yang tidak digunakan dalam pelatihan, sehingga hasil klasifikasi dapat diaplikasikan dengan lebih tepat di lapangan.

Confusion matrix adalah matriks yang menunjukkan perbandingan antara hasil prediksi model dan data aktual. Matriks ini terdiri dari empat komponen utama: *True Positive* (TP) untuk prediksi benar pada kasus kebakaran hutan, *False Positive* (FP) untuk kesalahan prediksi yang menganggap terjadi kebakaran hutan padahal tidak, *True Negative* (TN) untuk prediksi benar pada bukan kebakaran hutan, dan *False Negative* (FN) yang menunjukkan kesalahan prediksi saat kebakaran hutan tidak terdeteksi. *Confusion matrix* memberikan informasi mengenai akurasi, presisi, dan sensitivitas model dalam klasifikasi kebakaran hutan.



Gambar 4.10 *Confusion Matrix*

Berdasarkan hasil *Confusion matrix* pada Gambar 4.10, model menunjukkan performa yang cukup baik dengan skor akurasi sebesar 93%. Pada kelas "bukan kebakaran hutan," model berhasil mengidentifikasi 136 sampel dengan benar (TN), namun terdapat 19 sampel yang salah diklasifikasikan sebagai "kebakaran hutan" (FP). Di sisi lain, untuk kelas "kebakaran hutan," model mampu mengidentifikasi 251 sampel dengan benar (TP) dan terdapat 10 sampel yang salah diklasifikasikan sebagai "bukan kebakaran hutan" (FN). Selanjutnya, nilai akurasi, recall, dan precision dapat dihitung berdasarkan matriks ini untuk menilai performa lebih lanjut.

$$accuracy = \frac{251 + 136}{136 + 19 + 251 + 10} \times 100\% = 93,030\%$$

$$precision = \frac{251}{251 + 19} \times 100\% = 92,960\%$$

$$recall = \frac{251}{251 + 10} \times 100\% = 96,170\%$$

Berdasarkan perhitungan metrik evaluasi terhadap data *testing*, diperoleh hasil sebagai berikut. Nilai akurasi sebesar 93,03% menunjukkan bahwa model

mampu melakukan prediksi dengan benar pada 93,03% dari seluruh data *testing*. Ini berarti bahwa secara keseluruhan, model memiliki kemampuan yang sangat baik dalam mengklasifikasikan data sebagai "kebakaran hutan" atau "bukan kebakaran hutan."

Precision sebesar 92,96% menunjukkan bahwa dari semua prediksi "kebakaran hutan" yang dibuat oleh model, 92,96% di antaranya benar-benar merupakan kejadian kebakaran hutan. Tingginya nilai precision ini menunjukkan bahwa model memiliki tingkat kesalahan yang rendah dalam memberikan peringatan positif terhadap kebakaran hutan.

Recall sebesar 96,17% menunjukkan bahwa dari semua kejadian kebakaran hutan yang sebenarnya, model berhasil mengidentifikasi 96,17% di antaranya dengan benar. Recall yang tinggi ini mengindikasikan bahwa model memiliki sensitivitas yang sangat baik dalam mendeteksi kejadian kebakaran hutan, yang penting dalam konteks mitigasi kebakaran hutan agar risiko dapat ditekan.

4.5 Kajian Keislaman dengan Hasil Penelitian

Kebakaran hutan adalah bencana alam yang dapat membawa dampak luas bagi kehidupan manusia dan lingkungan. Dalam Islam, setiap perbuatan baik yang bermanfaat untuk menjaga alam dan mencegah kerusakan dipandang sebagai bentuk ibadah dan amal shaleh yang akan membawa manusia menuju golongan kanan, yaitu golongan orang-orang yang diridhai oleh Allah SWT. Hasil penelitian ini, yang mengaplikasikan metode *Naïve Bayes* dalam mendeteksi kebakaran hutan, memiliki akurasi yang cukup tinggi dan berpotensi mendukung upaya pencegahan bencana secara dini.

Penelitian ini tidak hanya memberikan kontribusi dalam ilmu pengetahuan tetapi juga mengingatkan kita akan tanggung jawab manusia sebagai khalifah di bumi untuk memelihara alam. Ilmu pengetahuan modern, termasuk teknologi kecerdasan buatan, dapat menjadi sarana yang bermanfaat untuk menjaga ciptaan Allah SWT. Harapannya, penelitian ini dapat berperan dalam mengurangi kerusakan lingkungan dan membawa manusia menjadi bagian dari golongan yang berbuat baik, yang termasuk ke dalam golongan kanan yang dijanjikan balasan baik oleh Allah SWT.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan rumusan masalah dan hasil penelitian yang diperoleh, dapat disimpulkan bahwa model klasifikasi yang dibangun menggunakan algoritma *Naïve Bayes* menunjukkan performa evaluasi yang baik dalam mendeteksi dan mengidentifikasi kejadian kebakaran hutan di Indonesia berdasarkan data *testing*. Model ini mencapai tingkat akurasi sebesar 93,03%, yang menunjukkan kemampuan model untuk mengklasifikasikan data dengan tingkat ketepatan tinggi.

Analisis *Confusion matrix* pada data *testing* menunjukkan bahwa, untuk kelas "bukan kebakaran hutan," model berhasil mengidentifikasi 136 sampel dengan benar sebagai "bukan kebakaran hutan" (*True Negative*), meskipun terdapat 19 sampel yang salah diklasifikasikan sebagai "kebakaran hutan" (*False Positive*). Pada kelas "kebakaran hutan," model berhasil mengidentifikasi 251 sampel dengan benar (*True Positive*), sedangkan 10 sampel salah diklasifikasikan sebagai "bukan kebakaran hutan" (*False Negative*).

Secara keseluruhan, algoritma *Naïve Bayes* mampu memberikan performa klasifikasi yang efektif, dengan nilai *precision* sebesar 92,96% dan *recall* sebesar 96,17%, menunjukkan bahwa model memiliki ketepatan dan sensitivitas yang tinggi terhadap kejadian kebakaran hutan dalam data *testing*.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, beberapa rekomendasi dapat diberikan untuk penelitian selanjutnya guna memperluas cakupan analisis dan memperdalam pemahaman terkait prediksi kebakaran hutan di Indonesia. Saran ini diharapkan dapat memberikan arahan dalam pengembangan model klasifikasi yang lebih komprehensif dan akurat.

1. Pada penelitian ini, algoritma *Naïve Bayes* digunakan sebagai model klasifikasi yang terbukti cukup akurat dalam mendeteksi kejadian kebakaran hutan. Untuk penelitian selanjutnya, dapat ditambahkan penggunaan variabel lain untuk meningkatkan keakuratan hasil.
2. Dataset yang digunakan dalam penelitian ini cukup mendukung proses klasifikasi. Namun, agar dapat menangkap pola yang lebih luas dan kontekstual, penelitian mendatang bisa mempertimbangkan untuk menggunakan data dari periode yang lebih panjang.
3. Penelitian ini dapat menjadi dasar bagi instansi terkait untuk mengembangkan sistem pemantauan dini kebakaran hutan berbasis machine learning. Sistem ini dapat diintegrasikan dengan pusat kendali instansi untuk memberikan peringatan dini dan membantu pengambilan keputusan secara lebih cepat.
4. Penelitian ini dapat diimplementasikan dalam sistem peringatan publik, sehingga masyarakat sekitar kawasan hutan mendapatkan informasi lebih dini mengenai potensi kebakaran, yang dapat membantu dalam upaya pencegahan dan mitigasi.

DAFTAR PUSTAKA

- A'yunnisa, N., Salim, Y., & Azis, H. (2022). Analisis performa metode Gaussian *Naïve Bayes* untuk klasifikasi citra tulisan tangan karakter arab. *Indonesian Journal of Data and Science (IJODAS)*, 3(3), 115–121.
- Akil, I., & Chaidir, I. (2024). Classification of Heart Disease Diagnoses Using Gaussian *Naïve Bayes*. *Komputasi: Jurnal Ilmiah Ilmu Komputer Dan Matematika*, 21(2), 31–36.
<https://doi.org/10.33751/komputasi.v21i2.10114>
- Al Mahalli, J., & As Suyuti, J. (2003). *Tafsir Jalalain*.
- Alim, M. K., & Arianto, D. B. (2023). Analisis Korelasi Antara Faktor Ekonomi dan Distribusi Penduduk di Jawa Timur Tahun 2020 Menggunakan Metode Korelasi Pearson. *Jurnal Multidisiplin Saintek*, 01(04).
- All Rosit, H., Mardotillah, A., Aura Delazenitha, R., Mutiarani, S., & Vianney Christina Sulle, T. (2023). Identifikasi dan Mitigasi Kebakaran Hutan dan Lahan Melalui Zonasi Wilayah Rawan Kebarakan dengan Teknologi Geospasial. *Jurnal Widya Bhumi*, 3(1), 13–30.
- Astoni, K., & Haris, M. (2022). Analisis Penerapan *Principal Component Analysis* (PCA) pada Deteksi Kecurangan Kartu Kredit Menggunakan Random Forest. *Jurnal Elektro Telekomunikasi Terapan*, 1152–1161.
<https://doi.org/10.25124/jett.v9i1.5019>
- Az-Zuhaili, W. (2013). *Tafsir Al-Munir*.
- Buulolo, E. (2020). *Data Mining Untuk Perguruan Tinggi* (D. Novidiantoko, H. A. Susanto, & A. Y. Wati, Eds.; Pertama). CV Budi Utama.
- Enzellina, G., & Suhaedi, D. (2022). Penggunaan Metode *Principal Component Analysis* dalam Menentukan Faktor Dominan. *Jurnal Riset Matematika*, 2(2), 101–110.
- Fitrianah, D., & Fahmi, H. (2019). The Identification of determinant Parameter in Forest Fire Based on Feature Selection Algorithms. *Sinergi*, 23(3), 184.
<https://doi.org/10.22441/sinergi.2019.3.002>
- Goejantoro, R., & Desi Yuniarti. (2019). Comparison of K-Rata ratas And K-Medoids Grouping Data on Potential Forest / Land Fires Based on Hotspots Distribution (Case study : Hotspots Data in Indonesia on April 28, 2018). *Jurnal EKSPONENSIAL*, 10(2).

- Kementrian Agama RI. (2024). *Qur'an Kemenag*. Lajnah Pentashihan Mushaf Al-Qur'an.
- Kristianti, D., & Hariyadi, Ma. (2022). Support Vector Machine (SVM) dan Algoritma *Naïve Bayes* (NB) Untuk Mengklasifikasi Keterlambatan Pembayaran Sumbangan Pendidikan di Madrasah Ibtidaiyah. *Jurnal Pendidikan Tambusai*, 6(2), 13468–13478.
- Maharani, C., Asrial, A. A., Purba, B. A., & Miftahurrahmah. (2020). *Edukasi Dampak Kesehatan dan Upaya Perlindungan Diri dari Bencana Kabut Asap*. 3(1), 22–26.
- Mifaul, 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*. 9(2), 3384–3390.
- Normawati, D., & Prayogi, S. A. (2021). Implementasi *Naïve Bayes Classifier* Dan *Confusion matrix* Pada Analisis Sentimen Berbasis Teks Pada Twitter. *Jurnal Sains Komputer & Informatika*, 5(2).
- Nugraheny, D. (2015). Metode Nilai Jarak Guna Kesamaan atau Kemiripan Ciri suatu Citra (Kasus Deteksi Awan Cumulonimbus Menggunakan Principal Component Analysis). *Jurnal Angkasa*, VII.
- Nugroho, P., & Khotimah, H. (2017). Klasifikasi Tingkat Kerawanan Kebakaran Hutan dan Lahan Menggunakan Algoritme *Naïve Bayes* di Pulang Pisau, Kalimantan Tengah.
- Oktaviana, W. (2017). Optimasi basis pengetahuan menggunakan algoritma FP-Growth untuk membangun struktur Bayesian Network Penyakit mata. 1–5.
- Pratiwi, T. A., Irsyad, M., Kurniawan, R., Agustian, S., & Negara, B. S. (2021). Klasifikasi Kebakaran Hutan dan Lahan Menggunakan Algoritma *Naïve Bayes* di Kabupaten Palalawan. *Journal of Computer Engineering System and Science*, 6(1), 139–148.
- Putrada, A. M., Abdurohman, 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*, 19–26. <https://doi.org/10.22219/kinetik.v4i1.704>

- 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 Index Cuaca Kebakaran Berdasarkan Data AWS (Automatic Weather Station) (Studi Kasus : Kabupaten Kubu Raya). *Jurnal Coding, Sistem Komputer Untan*, 06(2), 48–56.
- Romadloni, N. T., & Pardede, H. F. (2019). Seleksi Fitur Berbasis Pearson Correlation Untuk Optimasi Opinion Mining Review Pelanggan. *Jurnal Resti*, 1(3), 505–510.
- Rozy, F., Rangkuti, S., Fauzi, M. A., Sari, Y. A., Dewi, E., & Sari, L. (2018). *Analisis Sentimen Opini Film Menggunakan Metode Naïve Bayes dengan Ensemble Feature dan Seleksi Fitur Pearson Correlation Coefficient* (Vol. 2, Issue 12). <http://j-ptiik.ub.ac.id>
- Rudiyana, A., Erik Dzulkifli, A., & Munazar, K. (2022). Forest Fires Classification by Using The K-Nearest Neighbor Method: Case Studi of West Kalimantan Province Forest. *Jurnal Teknologi Informasi Dan Multimedia*, 3(4), 195–202. <https://sipongi.menlhk.go.id/>
- Setiono, M. H. (2022). Komparasi Algoritma Decision Tree, Random Forest, SVM DAN K-NN dalam Klasifikasi Kepuasan Penumpang Maskapai Penerbangan. *INTI Nusa Mandiri*, 17(1), 32–39. <https://doi.org/10.33480/inti.v17i1.3420>
- Wahyuni Kalumbang, S. (2021). Perbandingan Regresi Logistik, Klasifikasi *Naive Bayes*, dan Random Forest (Comparison the Logistic Regression, *Naive Bayes* Classification, and Random Forest). *Jurnal Matematika Thales (JMT)*, 03(02).

LAMPIRAN

Lampiran 1 Data Mentah Kebakaran Hutan di Indonesia

<https://bit.ly/DatasetKebakaranIndonesia>

Contoh data :

No	X1	X2	X3	X4	X5	X6	X7	X8	Class
1.	4,182	101,240	25,7	92	24,2	0	1,6	12,2	0
2.	4,182	101,240	26,9	90	25,2	0	6,5	14,4	0
3.	4,182	101,240	27,7	87	25,4	0,1	6,4	17,6	0
4.	4,182	101,240	28,7	82	25,4	0,3	7,2	23	0
5.	4,182	101,240	28,6	85	25,7	0,3	3,3	20,9	0
6.	4,182	101,240	28,9	82	25,6	0,6	2,8	16,9	0
7.	4,182	101,240	26,5	90	24,7	5,1	3,6	18	0
8.	4,182	101,240	26,4	90	24,6	0,6	1,3	14	0
9.	4,182	101,240	26,3	91	24,6	0,3	2,3	10,8	0
10.	4,182	101,240	25,7	92	24,2	0	1,6	12,2	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2760.	0,319	101,018	27	78,5	23	0	4,5	11	1
2761.	1,869	98,592	27,5	80	23,5	0,25	5,5	10	1
2762.	-2,649	111,648	27	83,5	23,5	0,1	5,5	13	1
2763.	2,610	98,777	23,5	76,5	18,5	0	7	11	1
2764.	1,869	98,592	25,5	77,5	21	0,2	7,5	10,5	1
2765.	1,551	99,178	23,5	79,5	19	0,05	4,5	7	1
2766.	1,636	99,839	27,5	73	21,5	0	2	3,5	1
2767.	1,156	99,800	27	80	22,5	0,4	0,5	1,5	1
2768.	-1,824	106,080	27,5	81	23,5	0,1	7	10,5	1
2769.	-3,207	104,632	27	82	23	0	3,5	6	1

Lampiran 2 Data Training

<https://bit.ly/DataTrainingKebakaran>

Contoh data :

No	X1	X2	X3	X4	X5	X6	X7	X8
1.	4,182	101,167	24,5	97	24	0,4	2,3	9,4
2.	-7,720	114,005	25	79,5	22	0,25	5,5	9,5
3.	4,182	101,093	28,1	83	25	0,2	2,4	17,6
4.	-7,813	114,300	27,5	72,5	20	0	11,5	19
5.	-7,694	111,986	25	63,5	17	0	14	22
6.	-8,334	115,087	25,5	77	20,5	0	8	14,5
7.	-2,289	115,584	26	73	21	0	12,5	23,5
8.	-7,726	114,001	26	76,5	21,5	0,1	8	13
9.	-0,469	100,582	25,5	64,5	18	0	2,5	9,5
10.	-2,864	113,926	27,5	67,5	22	0	7,5	13,5
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2344.	-2,613	102,880	26,5	79,5	22,5	0,8	3	10
2345.	-6,285	106,361	29,5	64,5	22,5	0	9	10,5
2346.	5,518	116,492	19,7	90	17,9	0	3,6	16,9
2347.	-7,708	111,957	24,5	60	15,5	0	11,5	18,5
2348.	5,588	116,514	16,6	89	14,8	0,1	2,9	27
2349.	-3,889	114,994	28,5	72	21,5	0	5,5	13
2340.	-6,431	106,588	29	67,5	21,5	0,15	7,5	13,5
2341.	-2,454	104,970	27,5	78	23,5	0,05	7,5	14,5
2352.	-3,476	114,782	25,5	77,5	20,5	0	6	16
2353.	-7,845	111,989	25	67,5	17,5	0	16,5	30,5

Lampiran 3 *Data Testing*

<https://bit.ly/DataTestingKebakaran>

Contoh data :

No	X1	X2	X3	X4	X5	X6	X7	X8
1.	-7,734	112,601	26	66,5	18	0	7	18
2.	-1,923	114,230	27	82,5	22,5	0,75	2,5	7,5
3.	5,448	116,470	18,8	84	16,1	0	6,4	19,8
4.	2,565	100,601	28	77	23,5	0	12,3	16,9
5.	4,774	96,992	28	74	23	0	6	8
6.	-7,427	108,741	25,5	80	21,5	0	13,5	19,5
7.	-7,784	112,929	30	65,5	21,5	0	7,5	13
8.	-5,521	105,764	28,5	71,5	22	0	10	16
9.	-8,332	114,221	25,5	69,5	18,5	0	8	13,5
10.	5,237	116,330	23	99	22,9	0,5	3,8	7,6
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
407.	4,464	115,503	18,2	94	17,3	0	4	10,4
408.	-7,725	113,995	23,5	80,5	20	0,05	9	14,5
409.	5,788	95,637	27,5	80	25	1,3	19	40
400.	4,323	101,332	21,6	97	21,1	0,4	3,8	13
401.	4,323	101,332	22,3	94	21,2	0	9,2	28,8
412.	-7,064	107,482	26	45	12,5	0	9,5	14
413.	-3,103	102,495	27,5	66	20	0,45	4	11,5
414.	1,792	99,722	22,7	98	22,4	0	1,5	7,2
415.	-7,023	110,139	27	75,5	23	0,9	7,5	16
416.	-3,947	114,907	26,5	72,5	20	0	7	19

Lampiran 4 Data *Training* Setelah Standardisasi

<https://bit.ly/StandardisasiTraining>

Contoh data :

No	X1	X2	X3	X4	X5	X6	X7	X8
1.	0,846	-0,956	-0,475	1,671	1,010	0,224	-1,131	-0,989
2.	-1,650	0,787	-0,305	0,083	0,205	0,028	-0,326	-0,974
3.	0,846	-0,965	0,747	0,401	1,412	-0,037	-1,106	0,292
4.	-1,670	0,827	0,543	-0,552	-0,599	-0,297	1,183	0,511
5.	-1,645	0,513	-0,305	-1,369	-1,806	-0,297	1,811	0,979
6.	-1,779	0,934	-0,136	-0,144	-0,398	-0,297	0,303	-0,192
7.	-0,511	1,001	0,034	-0,507	-0,197	-0,297	1,434	1,214
8.	-1,652	0,786	0,034	-0,189	0,004	-0,167	0,303	-0,427
9.	-0,130	-1,035	-0,136	-1,278	-1,404	-0,297	-1,080	-0,974
10.	-0,632	0,776	0,543	-1,006	0,205	-0,297	0,177	-0,349
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2344.	-0,579	-0,723	0,204	0,083	0,407	0,744	-0,955	-0,896
2345.	-1,349	-0,251	1,222	-1,278	0,407	-0,297	0,554	-0,817
2346.	1,126	1,125	-2,104	1,036	-1,444	-0,297	-0,804	0,184
2347.	-1,648	0,509	-0,475	-1,687	-2,410	-0,297	1,183	0,433
2348.	1,141	1,128	-3,156	0,945	-2,691	-0,167	-0,980	1,761
2349.	-0,847	0,921	0,883	-0,598	0,004	-0,297	-0,326	-0,427
2340.	-1,380	-0,220	1,052	-1,006	0,004	-0,102	0,177	-0,349
2341.	-0,547	-0,439	0,543	-0,053	0,809	-0,232	0,177	-0,192
2352.	-0,760	0,892	-0,136	-0,098	-0,398	-0,297	-0,200	0,042
2353.	-1,677	0,513	-0,305	-1,006	-1,605	-0,297	2,440	2,307

Lampiran 5 Data *Testing* Setelah Standardisasi

<https://bit.ly/StandardisasiTesting>

Contoh data :

No	X1	X2	X3	X4	X5	X6	X7	X8
1.	-1,653	0,596	0,034	-1,097	-1,404	-0,297	0,051	0,354
2.	-0,434	0,818	0,373	0,355	0,407	0,679	-1,080	-1,286
3.	1,111	1,122	-2,410	0,492	-2,168	-0,297	-0,100	0,636
4.	0,507	-1,032	0,713	-0,144	0,809	-0,297	1,384	0,184
5.	0,970	-1,522	0,713	-0,416	0,608	-0,297	-0,200	-1,208
6.	-1,589	0,073	-0,136	0,128	0,004	-0,297	1,686	0,589
7.	-1,664	0,641	1,392	-1,188	0,004	-0,297	0,177	-0,427
8.	-1,189	-0,332	0,883	-0,643	0,205	-0,297	0,806	0,042
9.	-1,779	0,816	-0,136	-0,825	-1,203	-0,297	0,303	-0,349
10.	1,067	1,103	-0,984	1,853	0,567	0,354	-0,754	-1,271
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
407.	0,905	0,990	-2,613	1,399	-1,685	-0,297	-0,703	-0,833
408.	-1,651	0,786	-0,814	0,174	-0,599	-0,232	0,554	-0,192
409.	1,183	-1,706	0,543	0,128	1,412	1,395	3,069	3,792
400.	0,876	-0,933	-1,459	1,671	-0,157	0,224	-0,754	-0,427
401.	0,876	-0,933	-1,222	1,399	-0,116	-0,297	0,604	2,042
412.	-1,513	-0,098	0,034	-3,048	-3,617	-0,297	0,680	-0,271
413.	-0,682	-0,775	0,543	-1,142	-0,599	0,289	-0,703	-0,661
414.	0,346	-1,152	-1,086	1,762	0,366	-0,297	-1,332	-1,333
415.	-1,504	0,263	0,373	-0,280	0,608	0,875	0,177	0,042
416.	-0,859	0,909	0,204	-0,552	-0,599	-0,297	0,051	0,511

Lampiran 6 Data *Training* Setelah PCA

https://bit.ly/PCA_Training

Contoh data :

No	PC1	PC2	PC3	PC4	PC5	PC6	Class
1.	2,476	1,195	0,051	-0,171	0,241	-0,523	0
2.	-0,263	0,012	-1,263	1,288	0,565	-0,865	1
3.	0,945	1,664	0,691	-0,455	0,126	0,653	0
4.	-2,149	-0,775	-0,220	0,547	0,354	-0,618	1
5.	-2,710	-1,991	-0,307	-0,092	-0,671	-0,748	1
6.	-1,118	-0,698	-0,970	0,874	0,527	-0,814	1
7.	-1,645	-1,202	0,784	0,031	0,615	-0,044	1
8.	-1,024	-0,211	-0,844	0,955	0,562	-0,783	1
9.	-0,047	0,218	-1,891	-0,608	-1,654	0,400	1
10.	-1,172	0,234	-0,609	0,531	0,432	0,403	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
2344.	0,586	1,161	-0,588	0,753	-0,676	-0,384	1
2345.	-1,876	1,322	-0,780	0,388	-0,046	-0,318	1
2346.	1,992	-2,491	-0,597	-0,352	0,255	0,382	0
2347.	-2,446	-2,097	-1,202	-0,009	-1,134	-0,478	1
2348.	1,993	-4,356	-0,273	-0,818	-0,595	0,515	0
2349.	-0,969	0,304	-0,926	0,801	0,529	0,458	1
2340.	-1,616	0,839	-0,719	0,458	-0,309	-0,243	1
2341.	-0,471	1,000	0,060	-0,008	0,297	-0,426	1
2352.	-0,443	-0,696	-0,702	0,498	0,430	0,002	1
2353.	-3,121	-2,516	0,873	-0,382	-0,430	-0,952	1

Lampiran 7 Data *Testing* Setelah PCA

https://bit.ly/PCA_Testing

Contoh data :

No	PC1	PC2	PC3	PC4	PC5	PC6	Class
1.	-1,721	-1,139	-1,236	0,435	-0,527	-0,136	1
2.	0,727	0,705	-0,918	1,578	0,369	0,189	1
3.	1,319	-3,307	-0,425	-0,687	-0,221	0,340	0
4.	-0,700	1,051	1,198	-1,001	0,103	-0,367	0
5.	0,460	1,955	-0,237	-0,975	-0,523	0,253	1
6.	-1,588	-0,601	0,443	0,063	0,455	-1,594	1
7.	-2,083	0,669	-1,025	0,925	0,263	0,091	1
8.	-1,650	0,641	-0,086	0,044	0,007	-0,583	1
9.	-1,509	-0,994	-1,478	0,709	-0,148	-0,549	1
10.	2,574	-0,215	-0,172	0,801	1,265	-0,062	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
407.	2,490	-2,483	-1,258	-0,241	0,229	-0,347	0
408.	-0,713	-1,165	-0,843	0,666	0,414	-1,252	1
409.	-1,395	-0,037	5,267	-1,272	-0,800	-0,483	1
400.	2,482	-0,285	0,043	-0,597	-0,326	-0,814	0
401.	0,922	-1,290	1,840	-1,602	-0,047	-0,711	0
412.	-2,978	-1,792	-2,396	-0,328	-2,532	0,275	1
413.	-0,606	0,743	-1,137	0,223	-1,315	0,184	1
414.	2,587	0,734	-0,869	-0,505	0,097	-1,070	0
415.	-0,911	0,413	0,121	1,403	-0,082	-0,649	1
416.	-1,123	-0,825	-0,490	0,414	0,220	0,253	1

Lampiran 8 Hasil Prediksi Probabilitas Posterior Data *Testing*

https://bit.ly/Posterior_HasilPrediksi

Contoh data :

No	PC1	PC2	PC3	PC4	PC5	PC6	Class	Prob (0)	Prob (1)
1.	-1,721	-1,139	-1,236	0,435	-0,527	-0,136	1	0,005	0,995
2.	0,727	0,705	-0,918	1,578	0,369	0,189	1	0,237	0,763
3.	1,319	-3,307	-0,425	-0,687	-0,221	0,340	0	0,994	0,006
4.	-0,700	1,051	1,198	-1,001	0,103	-0,367	0	0,042	0,958
5.	0,460	1,955	-0,237	-0,975	-0,523	0,253	1	0,330	0,670
6.	-1,588	-0,601	0,443	0,063	0,455	-1,594	1	0,015	0,985
7.	-2,083	0,669	-1,025	0,925	0,263	0,091	1	0,001	0,999
8.	-1,650	0,641	-0,086	0,044	0,007	-0,583	1	0,002	0,998
9.	-1,509	-0,994	-1,478	0,709	-0,148	-0,549	1	0,005	0,995
10.	2,574	-0,215	-0,172	0,801	1,265	-0,062	0	0,902	0,098
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
407.	2,490	-2,483	-1,258	-0,241	0,229	-0,347	0	0,986	0,014
408.	-0,713	-1,165	-0,843	0,666	0,414	-1,252	1	0,049	0,951
409.	-1,395	-0,037	5,267	-1,272	-0,800	-0,483	1	0,234	0,766
400.	2,482	-0,285	0,043	-0,597	-0,326	-0,814	0	0,874	0,126
401.	0,922	-1,290	1,840	-1,602	-0,047	-0,711	0	0,873	0,127
412.	-2,978	-1,792	-2,396	-0,328	-2,532	0,275	1	0,008	0,992
413.	-0,606	0,743	-1,137	0,223	-1,315	0,184	1	0,023	0,977
414.	2,587	0,734	-0,869	-0,505	0,097	-1,070	0	0,845	0,155
415.	-0,911	0,413	0,121	1,403	-0,082	-0,649	1	0,016	0,984
416.	-1,123	-0,825	-0,490	0,414	0,220	0,253	1	0,014	0,986

Lampiran 9 Script Code Phyton

```

!pip install jcopml
import numpy as np
import pandas as pd

from sklearn.model_selection import train_test_split,
cross_val_score
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
from sklearn.metrics import accuracy_score
from sklearn.pipeline import Pipeline

import matplotlib.pyplot as plt
import seaborn as sns

from jcopml.pipeline import num_pipe, cat_pipe
df = pd.read_excel("Dataset.xlsx")
df
correlation_matrix = df.corr(method='pearson')
print(correlation_matrix)
x = df.iloc[:, :-1].values
y = df.iloc[:, -1].values
x_train, x_test, y_train, y_test = train_test_split(x, y,
test_size=0.15, random_state=42)

classification = Pipeline([
    ('scaler', StandardScaler()),          # Langkah
#standarisasi
    ('pca', PCA(n_components=0.95)),      # Langkah PCA
    ('algo', GaussianNB())               # Model Naive Bayes
])
scores = cross_val_score(classification, x_train, y_train,
cv=5)
# Menampilkan hasil cross-validation
print("Cross-validation scores:", scores)
print("Average cross-validation score:", scores.rata_rata())
y_pred = classification.predict(x_test)
y_pred

classification.predict_proba(x_test)

from jcopml.plot import plot_confusion_matrix

```

```
plot_confusion_matrix(x_train, y_train, x_test, y_test,
classification)

akurasi = classification_report(y_test, y_pred)
print(akurasi)

x_test_df = pd.DataFrame(x_test_pca)

# Menggabungkan x_test dengan y_test dan y_pred
ydata = pd.DataFrame(x_test_df) # Mulai dengan x_test
ydata['y_test'] = y_test        # Menambahkan kolom y_test
ydata['y_pred'] = y_pred        # Menambahkan kolom y_pred

ydata
```

RIWAYAT HIDUP



Nikmatun Awwaliyah, lahir di Jember pada 18 September 2003. Penulis merupakan putri pertama dari dua bersaudara dari pasangan Bapak Amiruddin Zaka dan Ibu Aini Maslahah. Pendidikan formal penulis dimulai di TK Dewi Masyithoh, yang diselesaikan pada tahun 2009. Selanjutnya, penulis melanjutkan pendidikan di MI Dewi Masyithoh dan lulus pada tahun 2015.

Pendidikan tingkat menengah pertama ditempuh di MTS "Unggulan" Nuris Jember dan lulus pada tahun 2018. Penulis kemudian menyelesaikan pendidikan menengah atas di MA "Unggulan" Nuris Jember pada tahun 2021. Pada tahun yang sama, penulis melanjutkan pendidikan tinggi di Universitas Islam Negeri Maulana Malik Ibrahim Malang, dengan memilih Program Studi Matematika di Fakultas Sains dan Teknologi. Selama menempuh pendidikan di universitas, penulis aktif berpartisipasi dalam berbagai kegiatan akademik dan non-akademik. Penulis pernah menjadi anggota kepanitiaan berbagai acara kampus, yang memberikan pengalaman dalam mengelola dan menyukseskan kegiatan mahasiswa. Selain itu, penulis juga pernah menjabat sebagai asisten praktikum, yang memperkuat kemampuan penulis dalam bidang pengajaran dan pendampingan akademik. Penulis juga aktif mengikuti berbagai seminar dan pelatihan, baik di dalam maupun di luar kampus, untuk meningkatkan wawasan dan keterampilan di bidang akademik dan profesional.



KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933

BUKTI KONSULTASI SKRIPSI

Nama : Nikmatun Awwaliyah
NIM : 210601110053
Fakultas / Jurusan : Sains dan Teknologi / Matematika
Judul Skripsi : Implementasi Algoritma *Naïve Bayes* untuk Klasifikasi Kebakaran Hutan di Indonesia
Pembimbing I : Hisyam Fahmi, M.Kom.
Pembimbing II : Ari Kusumastuti, M.Pd., M.Si.

No	Tanggal	Hal	Tanda Tangan
1.	20 Juni 2024	Konsultasi Topik dan Data	1.
2.	30 September 2024	Konsultasi Bab I, II, dan III	2.
3.	04 Oktober 2024	Konsultasi Bab I, II, dan III	3.
4.	04 Oktober 2024	ACC Bab I, II, dan III	4.
5.	07 Oktober 2024	Konsultasi Kajian Agama Bab I dan II	5.
6.	08 Oktober 2024	ACC Kajian Agama Bab I dan II	6.
7.	22 Oktober 2024	ACC Seminar Proposal	7.
8.	01 November 2024	Konsultasi Revisi Seminar Proposal	8.
9.	04 November 2024	Konsultasi Revisi Seminar Proposal	9.
10.	05 November 2024	Konsultasi Bab IV	10.
11.	06 November 2024	Konsultasi Bab IV	11.
12.	07 November 2024	Konsultasi Kajian Agama Bab IV	12.
13.	07 November 2024	Konsultasi Bab V	13.
14.	08 November 2024	ACC Kajian Agama Bab IV	14.
15.	08 November 2024	Konsultasi Bab IV dan V	15.



**KEMENTERIAN AGAMA RI
UNIVERSITAS ISLAM NEGERI
MAULANA MALIK IBRAHIM MALANG
FAKULTAS SAINS DAN TEKNOLOGI
Jl. Gajayana No.50 Dinoyo Malang Telp. / Fax. (0341)558933**

16.	08 November 2024	ACC Bab IV dan V	16.
17.	08 November 2024	ACC Seminar Hasil	17.
18.	11 November 2024	Konsultasi Revisi Seminar Hasil	18.
19.	13 November 2024	Konsultasi Revisi Seminar Hasil	19.
20.	19 November 2024	ACC Sidang Skripsi	20.
21.	20 Desember 2024	ACC Keseluruhan	21.

Malang, 20 Desember 2024

Mengetahui,
Ketua Program Studi Matematika



Dr. Elly Susanti, M.Sc.
NIP. 19741129 200012 2 005