

Deteksi Pola Kejadian Bencana Menggunakan Algoritma Naïve Bayes di Kabupaten Boyolali

Miranto Sri Agus Prasetyo ^{a,1,*}, Ari Wahyono ^{a,2}, Muhammad Abdul Aziz ^{a,3}

^a Universitas Boyolali, Jalan Pandanarang No. 405, Boyolali 57314, Indonesia

¹ lagusprasetyo628@gmail.com *; ² namaku.ariwahyono@gmail.com; ³ dotacome@gmail.com

* Korespondensi penulis

Submission 24/8/2023, Revision 30/5/2024, Accepted 30/5/2024

Abstract

Boyolali Regency, which is geographically located between Mount Merbabu and the still active Mount Merapi, does have high potential as a disaster-prone area. According to data from the Boyolali Regional Disaster Management Agency, during the period from 2022 to 2023, 936 disaster events were recorded in the region. These disasters have a devastating impact on ecosystems and the environment, causing significant material losses, causing psychological disorders, and even threatening human safety. In this context, detecting patterns of disaster events is an important step in disaster mitigation and management efforts. This research was conducted with the aim of understanding patterns of disaster events based on data collected by BPBD Boyolali. The method used is the Naive Bayes algorithm, which has proven to be a powerful tool in data analysis. The results of this research show a high level of accuracy of 92.52%, precision of 92.68%, and recall of 91.64% using the Naive Bayes algorithm. This indicates that the algorithm is effective in recognizing and classifying patterns of disaster events based on existing data. Thus, the results of this research can provide a valuable contribution to the development of an early disaster detection system, as well as assist the authorities in designing more effective mitigation strategies to protect the community and environment in Boyolali Regency.

Keywords: Naïve Bayes, Disaster, Data mining, Boyolali, Detection

Abstrak

Kabupaten Boyolali, yang secara geografis terletak di antara Gunung Merbabu dan Gunung Merapi yang masih aktif, memang memiliki potensi yang tinggi sebagai daerah rawan bencana. Menurut data dari Badan Penanggulangan Bencana Daerah Boyolali, selama periode dari tahun 2022 hingga 2023, tercatat telah terjadi 936 kejadian bencana di wilayah tersebut. Bencana-bencana ini menyebabkan dampak yang merusak pada ekosistem dan lingkungan, menyebabkan kerugian materiil yang signifikan, menghadirkan gangguan psikologis, dan bahkan mengancam keselamatan jiwa manusia. Dalam konteks ini, deteksi pola kejadian bencana menjadi langkah penting dalam upaya mitigasi dan penanggulangan bencana. Penelitian ini dilakukan dengan tujuan untuk memahami pola kejadian bencana berdasarkan data yang dikumpulkan oleh BPBD Boyolali. Metode yang digunakan adalah menggunakan algoritma *Naive Bayes*, yang telah terbukti menjadi alat yang kuat dalam analisis data. Hasil dari penelitian ini menunjukkan tingkat akurasi yang tinggi sebesar 92,52%, presisi sebesar 92,68%, dan *recall* sebesar 91,64% menggunakan algoritma *Naive Bayes*. Hal ini menandakan bahwa algoritma tersebut efektif dalam mengenali dan mengklasifikasikan pola kejadian bencana berdasarkan data yang ada. Dengan demikian, hasil penelitian ini dapat memberikan kontribusi yang berharga dalam pengembangan sistem deteksi dini bencana, serta membantu pihak berwenang dalam merancang strategi mitigasi yang lebih efektif untuk melindungi masyarakat dan lingkungan di Kabupaten Boyolali.

Kata kunci: *Naive Bayes*, Bencana, Data mining, Boyolali, Deteksi

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license.



1. Pendahuluan

Kabupaten Boyolali, secara geografis terletak di antara Gunung Merbabu dan Gunung Merapi, memiliki potensi yang tinggi untuk menjadi daerah rawan bencana. Data kejadian bencana alam yang dihimpun dari Badan Penanggulangan Bencana Daerah (BPBD) Kabupaten Boyolali selama tiga tahun terakhir, dari 2020 hingga 2022, mencatat adanya 936 kejadian bencana alam [1]. Dalam rangka mencegah dan menangani bencana, BPBD Boyolali telah melakukan berbagai kegiatan, seperti apel rutin siaga bencana, pemasangan atribut rambu rawan bencana, sosialisasi, dan edukasi mengenai bencana kepada masyarakat, serta pemetaan daerah rawan bencana. Bencana alam tidak hanya menyebabkan kerusakan ekosistem lingkungan dan kerugian material, tetapi juga memberikan gangguan psikologis bagi korban yang terdampak, bahkan dapat menyebabkan kematian. Oleh karena itu, deteksi terhadap pola kejadian bencana menjadi langkah penting untuk meminimalisir dampak bencana alam di Kabupaten Boyolali. Dengan pemahaman yang lebih baik tentang pola dan tren kejadian bencana, langkah-langkah mitigasi dan respons yang lebih efektif dapat diambil untuk melindungi masyarakat dan lingkungan di wilayah tersebut.

Penelitian mengenai bencana alam telah menarik perhatian para peneliti yang menggunakan berbagai metode, termasuk pendekatan berbasis algoritma *Machine Learning*. Beberapa penelitian terkait mencakup penggunaan Artificial Neural Network (ANN) [2], Algoritma *K-Means* [3] [4], dan algoritma *Fuzzy C-Means* untuk pengelompokan daerah bencana alam [5]. Selain itu, ada juga penelitian yang memanfaatkan *Internet of Things* (IoT) untuk monitoring cuaca dan banjir [6], serta kombinasi antara IoT dengan Algoritma C4.5 untuk deteksi bencana banjir [7]. Di sisi lain, ada juga penelitian yang menggunakan algoritma *YOLO* untuk deteksi korban bencana alam [8], serta pemanfaatan metode deep learning untuk mendeteksi dan menghitung jumlah korban bencana [9].

Keanekaragaman metode yang digunakan ini menunjukkan bahwa bencana alam tetap menjadi fokus utama bagi peneliti, khususnya dalam konteks penerapan algoritma data mining dan teknologi terkini. Dengan menggabungkan kecerdasan buatan, analisis data, dan teknologi sensor, para peneliti berharap dapat meningkatkan kemampuan dalam mendeteksi, memodelkan, dan merespons bencana alam dengan lebih efektif. Hal ini mencerminkan pentingnya upaya kolaboratif di antara berbagai bidang pengetahuan untuk menghadapi tantangan yang kompleks dan mendesak yang terkait dengan bencana alam.

Algoritma *Naïve Bayes* telah menjadi salah satu algoritma data mining yang populer dan sering digunakan dalam berbagai penelitian. Dalam konteks bencana alam, algoritma *Naïve Bayes* telah digunakan untuk klasifikasi [10] dan prediksi [11] bencana banjir. Selain itu, algoritma ini juga berhasil diterapkan dalam beberapa fokus penelitian lain, seperti prediksi kelulusan peserta sertifikasi *Microsoft Office* [12], analisis penggunaan akses internet [13], dan klasifikasi karya ilmiah [14]. Berbagai penelitian yang telah dilakukan menunjukkan bahwa *Naïve Bayes* memiliki kinerja yang baik dalam proses pencarian pola, deteksi, dan klasifikasi.

Dalam konteks penelitian ini, tujuan utamanya adalah untuk mengetahui pola kejadian bencana alam di wilayah Kabupaten Boyolali. Dengan menggunakan algoritma *Naïve Bayes*, penelitian ini bertujuan untuk menganalisis data historis tentang kejadian bencana alam untuk mengidentifikasi pola-pola yang mungkin ada. Dengan pemahaman yang lebih baik tentang pola kejadian bencana, diharapkan langkah-langkah mitigasi dan respons yang lebih efektif dapat diambil untuk melindungi masyarakat dan lingkungan di wilayah tersebut.

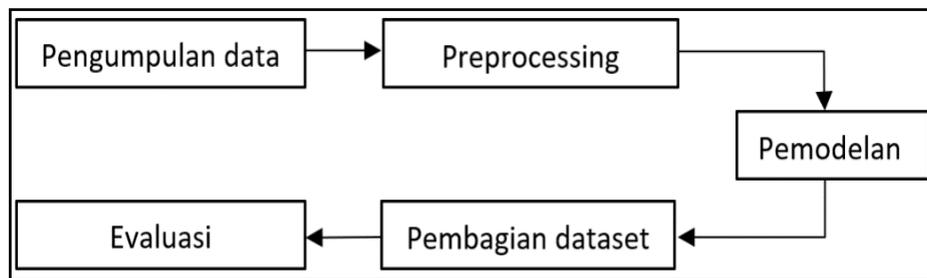
2. Metode Penelitian

Metodologi penelitian ini terdiri dari beberapa fase yang saling terkait, dimulai dari pengumpulan data yang melibatkan serangkaian langkah untuk mengumpulkan data yang relevan dan representatif tentang kejadian bencana di wilayah Kabupaten Boyolali. Setelah proses pengumpulan data selesai, langkah selanjutnya adalah tahap preprocessing di mana data tersebut menjalani serangkaian tindakan seperti pembersihan, transformasi, dan normalisasi untuk mempersiapkannya agar siap untuk analisis lebih lanjut.

Setelah data siap, proses berlanjut ke tahap pemodelan data di mana berbagai algoritma seperti *Naïve Bayes*, *decision tree*, dan *random tree* diterapkan untuk mengembangkan model prediktif berdasarkan data yang telah diproses sebelumnya. Pembagian dataset menjadi data latih dan data uji merupakan tahapan penting berikutnya, yang bertujuan untuk memastikan bahwa model yang dikembangkan dapat diuji secara independen untuk memastikan keandalannya.

Terakhir, proses evaluasi dilakukan untuk mengukur kinerja model yang telah dibuat dengan membandingkan hasil prediksi dengan nilai sebenarnya dari data uji. Parameter evaluasi seperti akurasi, precision, dan recall digunakan untuk mengevaluasi kinerja model dari berbagai algoritma yang telah diterapkan. Keseluruhan metodologi ini membentuk kerangka kerja yang komprehensif untuk mengidentifikasi, menganalisis, dan menginterpretasi pola kejadian bencana di wilayah Kabupaten Boyolali.

Harapannya, hasil penelitian ini akan memberikan wawasan yang berharga bagi pihak terkait dalam upaya mitigasi bencana di masa depan. Untuk alur metodologi penelitian ini bisa dilihat pada Gambar 1.



Gambar 1. Diagram Alir Penelitian

2.1. Pengumpulan data

Penelitian ini menggunakan data dari Badan Penanggulangan Bencana Daerah Boyolali untuk meneliti kejadian bencana alam dalam rentang waktu dari tahun 2020 hingga 2022. Total data yang terkumpul mencapai 936, mencerminkan berbagai insiden bencana yang terjadi selama periode tersebut. Dalam analisis yang mendalam terhadap dataset ini, teridentifikasi 487 kejadian angin puting beliung dan 449 kejadian tanah longsor, yang merupakan dua jenis bencana yang sering mewabah di wilayah Boyolali.

Pada Gambar 2, yang disajikan dalam penelitian ini, menguraikan 10 baris pertama dari *dataset*, memberikan gambaran awal tentang struktur dan isi data yang akan dihadapi dalam analisis. *Dataset* ini dianggap sebagai sumber informasi penting untuk memahami dinamika bencana alam di wilayah tersebut. Melalui analisis dan pemodelan yang cermat, penelitian ini bertujuan untuk mengungkap pola dan tren yang terkait dengan bencana alam, membantu pihak berwenang dan pemangku kepentingan dalam membangun strategi mitigasi dan penanggulangan yang lebih efektif.

Diharapkan, pemahaman yang lebih mendalam terhadap data ini akan memberikan kontribusi yang signifikan dalam upaya mitigasi bencana di masa mendatang. Dengan merumuskan strategi yang didukung oleh bukti-bukti data yang kuat, diharapkan dapat mengurangi dampak bencana tidak hanya pada tingkat masyarakat, tetapi juga pada ekosistem lokal. Oleh karena itu, hasil penelitian ini diharapkan dapat menjadi landasan yang kokoh dalam pembangunan kebijakan dan praktik-praktik yang responsif terhadap tantangan bencana di wilayah Boyolali dan mungkin juga daerah lainnya.

NO	HARI	TANGGAL	BULAN	TAHUN	DATARAN	JENIS TANAH	DESA	KECAMATAN	JENIS BENCANA
1	Senin	27	Desember	2021	Tinggi	Pasir	Sidomulyo	AMPEL	Angin Puting Beliung
2	Minggu	3	Januari	2021	Rendah	Kapur	WATU GEDE	Kemusu	Angin Puting Beliung
3	Sabtu	9	Januari	2021		Geluh	Jenengan	Sawit	Angin Puting Beliung
4	Jumat	8	Januari	2021	Rendah		Cepokosawit	Sawit	Angin Puting Beliung
5	Jumat	1	Februari	2021	Rendah	Lempung	Sempu	Andong	Angin Puting Beliung
6	Rebo	13	Januari	2021	Rendah	Geluh	Guwokajen	Sawit	Tanah Longsor
7	Sabtu	16	Januari	2021	Tinggi	Pasir	Sangup	TAMANSARI	Tanah Longsor
8	Selasa	19	Januari	2021	Rendah	Geluh	Gedangan	Teras	Tanah Longsor
9	Kamis	21	Januari	2021	Tinggi	Pasir	Tlogolele	Selo	Tanah Longsor
10	Jumat	22	Januari	2021	Tinggi	Pasir	tarubatang	Selo	Tanah Longsor

Gambar 2. Dataset Awal

2.2. Preprocessing

Dataset perlu dilakukan pembersihan karena ditemukan beberapa baris kosong. Untuk menyelesaikan ini, dilakukan penyesuaian dengan merujuk pada baris lain yang memiliki kesamaan. Selain itu, dilakukan pembenahan pada kolom "DESA dan KECAMATAN" dengan menerapkan aturan "*Capitalize Each Word*" untuk memastikan konsistensi penulisan. Selain itu, typo yang ditemukan dalam kolom "HARI" diperbaiki.

Transformasi juga dilakukan pada kolom "TANGGAL", di mana tanggal ditransformasi menjadi "PEKAN" yang disingkat sebagai "P". Penulisan dilakukan berdasarkan kriteria tanggal, di mana tanggal 1 hingga 7 masuk ke dalam "P1", tanggal 8 hingga 15 masuk ke dalam "P2", tanggal 16 hingga 21 masuk ke dalam "P3", tanggal 22 hingga 28 masuk ke dalam "P4", dan terakhir "P5" untuk tanggal 29 hingga 31.

Hasil transformasi dapat dilihat pada Gambar 3 yang disajikan di bawah ini. Dengan melakukan pembersihan dan transformasi ini, diharapkan *dataset* menjadi lebih rapi dan siap untuk analisis lebih lanjut.

NO	HARI	PEKAN	BULAN	TAHUN	DATARAN	JENIS TANAH	DESA	KECAMATAN	JENIS BENCANA
1	Senin	P4	Desember	2021	Tinggi	Pasir	Sidomulyo	Ampel	Angin Puting Beliuang
2	Minggu	P1	Januari	2021	Rendah	Kapur	Watugede	Kemusu	Angin Puting Beliuang
3	Sabtu	P2	Januari	2021	Rendah	Geluh	Jenengan	Sawit	Angin Puting Beliuang
4	Jumat	P2	Januari	2021	Rendah	Geluh	Cepokosawit	Sawit	Angin Puting Beliuang
5	Jumat	P1	Februari	2021	Rendah	Lempung	Sempu	Andong	Angin Puting Beliuang
6	Rabu	P2	Januari	2021	Rendah	Geluh	Guwokajen	Sawit	Tanah Longsor
7	Sabtu	P3	Januari	2021	Tinggi	Pasir	Sangup	Tamansari	Tanah Longsor
8	Selasa	P3	Januari	2021	Rendah	Geluh	Gedangan	Teras	Tanah Longsor
9	Kamis	P3	Januari	2021	Tinggi	Pasir	Tlogolele	Selo	Tanah Longsor
10	Jumat	P4	Januari	2021	Tinggi	Pasir	Tarubatang	Selo	Tanah Longsor

Gambar 3. Hasil Preprocessing

2.3. Pemodelan

Tahap ini akan dilakukan proses pemodelan dataset menggunakan algoritma *Naïve Bayes*. Algoritma ini menggunakan konsep probabilitas himpunan pada suatu kelas tertentu. Proses pemodelan algoritma diterapkan menggunakan aplikasi rapidminer. Penggunaan rumus *Naïve bayes* dapat dilihat pada persamaan (1) yang dimana X merupakan data tuple hasil dari pengujian suatu set data yang telah ditentukan masuk ke dalam kelas tertentu yang telah dibuat kategorinya. H adalah suatu hipotesis yang akan menentukan X masuk ke dalam kelas C. $P(H|X)$ adalah peluang atau probabilitas X, yang merupakan data tuple atau bukti yang diperoleh pada saat observasi, masuk ke dalam kelas H; dengan kata lain, mencari probabilitas X dimiliki oleh kelas H. $P(H | X)$ merupakan probabilitas *posterior*, di mana H dikondisikan pada X. $P(H)$ adalah probabilitas *prior*, atau probabilitas sebelumnya. $P(X|H)$ adalah probabilitas posterior di mana X dikondisikan pada H. H merupakan probabilitas sebelumnya dari X.

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

2.4. Pembagian *dataset* dan evaluasi

Penelitian ini membagi jumlah *dataset* yang digunakan menjadi dua bagian, dengan 80% dari data tersebut dialokasikan sebagai data latih untuk melatih model, sementara 20% sisanya digunakan sebagai data uji untuk menguji *performa* model yang telah dilatih. Dalam rangka mengukur kinerja sistem yang telah dikembangkan, penelitian ini menerapkan metode evaluasi menggunakan *confusion matrix*.

Confusion matrix adalah sebuah metode yang bertujuan untuk mengukur sejauh mana sistem mampu memprediksi data dengan akurat, dengan memperhatikan seberapa baik sistem tersebut dalam memprediksi dengan benar terhadap kelas-kelas yang telah ditentukan sebelumnya. Melalui *confusion matrix*, kinerja sistem dapat dipahami dengan melihat perbandingan antara nilai prediksi dan nilai aktual dalam klasifikasi multikelas. Pendekatan ini memberikan gambaran yang lebih komprehensif tentang keakuratan dan keandalan sistem dalam memprediksi data, yang merupakan langkah krusial dalam mengevaluasi efektivitas model yang dikembangkan dalam mengatasi masalah klasifikasi, atau bisa dilihat seperti Tabel 4.

Tabel 4. *Confusion Search*

		Prediksi Kelas	
		Positif	Negatif
Kelas Aktual	True	True Positive (TP)	False Negative (FN)
	False	False Positive (FP)	True Negative (TN)

True positive dan *true negative* merupakan klasifikasi yang benar pada setiap label, sementara *false negative* dan *false positive* hasil dari klasifikasi yang salah. Pengukuran evaluasi menggunakan *precision*,

recall, dan *accuracy*. *Precision* adalah perbandingan antara TP dengan banyak data yang diprediksi positif. Persamaan *precision* menggunakan formula :

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

Recall merupakan rasio antara data *True Positive* (TP) dengan jumlah seluruh data aktual yang memang benar-benar bernilai positif. Persamaan untuk menghitung *recall* dapat menggunakan formula berikut:

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Akurasi adalah perbandingan antara data *True Positif* dan *True Negatif* dengan semua jumlah *True Positif*, *True Negatif*, *False Positif*, dan *False Negatif*. Persamaan akurasi menggunakan formula:

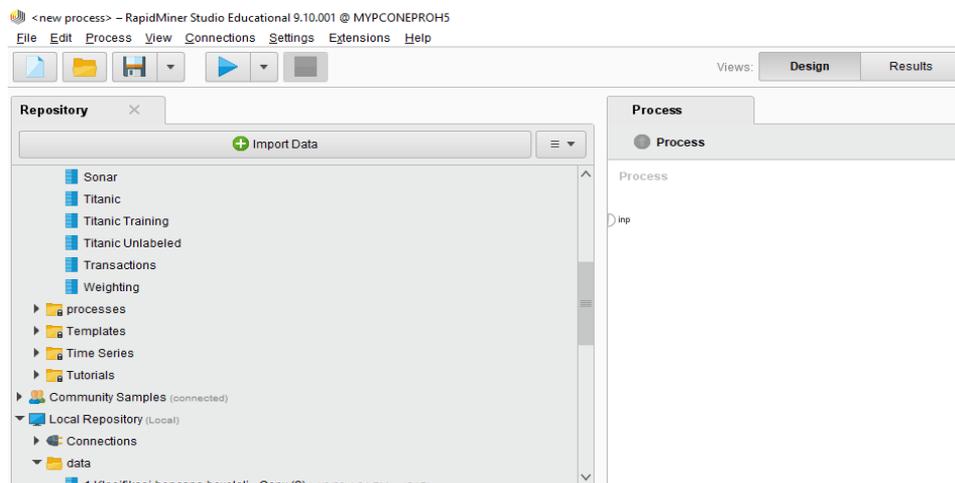
$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

3. Hasil dan Pembahasan

Pada tahapan ini membahas mengenai hasil dari penelitian yang telah dilakukan. Pada bagian ini digunakan aplikasi *rapidminer* untuk membuat model dari *dataset* menggunakan algoritma *Naïve Bayes*. *Rapidminer* merupakan suatu aplikasi bantuan yang digunakan untuk proses data mining. Beberapa penelitian menggunakan *rapidminer* sebagai alat bantuan diantaranya pengelompokkan data bencana alam [15], Analisis sebaran titik rawan bencana menggunakan *K-Means* [16], Penggunaan Algoritma C45 untuk evaluasi data sales [17]. Penggunaan *rapidminer* akan menghasilkan perhitungan dengan algoritma *Naïve bayes* untuk mengetahui pola dari kejadian bencana yang ada di Boyolali. Pola yang sudah teridentifikasi dapat digunakan sebagai acuan untuk memprediksi pola kejadian bencana data baru yang ada di wilayah Boyolali sehingga dapat digunakan sebagai alternatif langkah antisipasi terjadi bencana untuk meminimalisir dampak kerugian dari bencana di wilayah Kabupaten Boyolali.

3.1. Penggunaan *Rapidminer* dengan algoritma *Naive Bayes*

Dataset yang telah melalui proses *preprocessing* kemudian dimuat ke dalam aplikasi *RapidMiner*, sebagaimana yang terlihat pada gambar 2. Tahap *preprocessing* ini penting untuk mempersiapkan data sebelum dilakukan analisis lebih lanjut, termasuk membersihkan data dari nilai yang hilang atau tidak valid, melakukan normalisasi atau transformasi jika diperlukan, dan menyesuaikan format data agar sesuai dengan kebutuhan algoritma yang akan digunakan. Melalui langkah-langkah ini, data menjadi lebih siap untuk dieksplorasi dan dievaluasi menggunakan alat analisis seperti yang disediakan oleh *RapidMiner*. Gambar 2 menunjukkan tahapan ini secara visual, memperlihatkan bagaimana dataset yang telah diproses diimpor dan siap untuk digunakan dalam proses analisis data lebih lanjut menggunakan alat yang disediakan oleh *RapidMiner*.



Gambar 4. Proses Import Dataset

Setelah data dimuat, langkah selanjutnya adalah melakukan pengaturan terkait dengan kolom dan menentukan kolom "JENIS BENCANA" sebagai label dengan memilih opsi "Change Role" Label. Proses ini krusial dalam menetapkan kolom yang akan digunakan sebagai variabel independen dan variabel dependen dalam analisis data.

Setelah dataset berhasil diimpor dan pengaturan label selesai dilakukan, langkah berikutnya adalah menghubungkan operator yang akan digunakan dalam pemrosesan *RapidMiner*. Proses ini melibatkan penentuan urutan operasi yang diperlukan untuk analisis data, seperti pengelompokan, transformasi, atau pemodelan. Dengan menggunakan alat-alat yang tersedia dalam lingkungan *RapidMiner*, dataset dapat disiapkan untuk dilakukan langkah-langkah analisis lebih lanjut.

Hasil dari proses ini dapat dilihat pada Gambar 6, yang memberikan visualisasi yang jelas tentang struktur dan karakteristik dataset setelah melalui tahap pengaturan dan pemrosesan awal. Dengan dataset yang sudah diatur dan diproses dengan baik, dataset siap untuk dilakukan analisis lebih lanjut menggunakan berbagai alat yang tersedia dalam lingkungan *RapidMiner*. Gambar 6 memberikan gambaran yang komprehensif tentang data yang telah disiapkan dan siap untuk menjalani tahap analisis selanjutnya.

	TAHUN <i>integer</i>	DATARAN <i>polynomial</i>	JENIS TANAH <i>polynomial</i>	DESA <i>polynomial</i>	KECAMATAN <i>polynomial</i>	JENIS BEN... <i>binomial label</i>	J <i>polynomial</i>	K <i>polynomial</i>
1	2022	Rendah	Lempung	Gondanglegi	Klego	Angin Puting Beli...	?	?
2	2022	Rendah	Kapur	Juwangi	Juwangi	Angin Puting Beli...	?	?
3	2022	Tinggi	Pasir	Kiringan	Boyolali	Angin Puting Beli...	?	?
4	2022	Tinggi	Pasir	Mojosongo	Mojosongo	Angin Puting Beli...	?	?
5	2022	Rendah	Kapur	Watugede	Kemusu	Angin Puting Beli...	?	?
6	2022	Rendah	Kapur	Sarimulyo	Kemusu	Angin Puting Beli...	?	?
7	2022	Rendah	Kapur	Sarimulyo	Kemusu	Angin Puting Beli...	?	?
8	2022	Rendah	Kapur	Sarimulyo	Kemusu	Angin Puting Beli...	?	?
9	2022	Rendah	Kapur	Sarimulyo	Kemusu	Angin Puting Beli...	?	?
10	2022	Rendah	Kapur	Sarimulyo	Kemusu	Angin Puting Beli...	?	?
11	2022	Rendah	Kapur	Sarimulyo	Kemusu	Angin Puting Beli...	?	?
12	2022	Tinggi	Pasir	Klakah	Selo	Angin Puting Beli...	?	?
13	2022	Rendah	Lempung	Sendangrejo	Klego	Angin Puting Beli...	?	?
14	2022	Rendah	Lempung	Sangge	Klego	Angin Puting Beli...	?	?
15	2022	Tinggi	Pasir	Siswodipuran	Boyolali	Angin Puting Beli...	?	?
16	2022	Tinggi	Pasir	Pulisen	Boyolali	Angin Puting Beli...	?	?
17	2022	Tinggi	Pasir	Mudal	Boyolali	Angin Puting Beli...	?	?
18	2022	Tinggi	Pasir	Mudal	Boyolali	Angin Puting Beli...	?	?

Gambar 5. Proses Import Dataset

Setelah dataset berhasil diimpor, langkah berikutnya adalah memastikan bahwa data siap untuk digunakan dalam analisis lebih lanjut. Salah satu langkah krusial dalam proses ini adalah membagi dataset menjadi dua bagian: data latih dan data uji. Ini dilakukan untuk memungkinkan proses pelatihan model dengan menggunakan data latih, sementara data uji digunakan untuk menguji kinerja model yang telah dilatih. Dalam konteks ini, operator split data menjadi instrumen yang penting.

Dalam proses pembagian dataset, teknik sampling random seringkali digunakan untuk memastikan bahwa kedua jenis data memiliki representasi yang adil dan tidak bias. Dalam kasus ini, proporsi yang umumnya digunakan adalah 80% data latih dan 20% data uji. Proporsi ini telah terbukti efektif dalam banyak kasus untuk memastikan bahwa model yang dihasilkan memiliki kinerja yang baik pada data yang belum pernah dilihat sebelumnya.

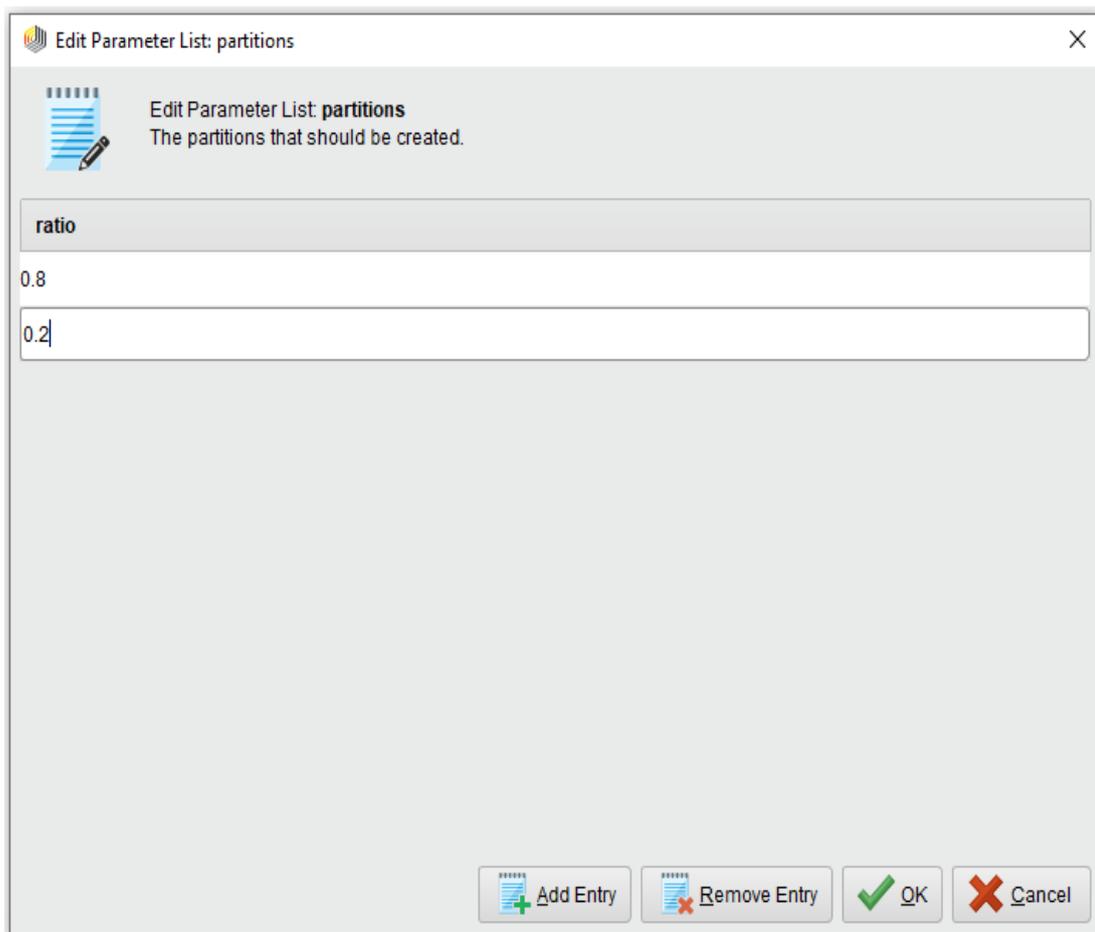
Penggunaan operator *split* data dalam *RapidMiner* mempermudah proses pembagian dataset dengan mengotomatiskan proses ini sesuai dengan proporsi yang diinginkan. Dengan demikian, pengguna dapat dengan mudah mengatur pembagian dataset tanpa harus melakukan langkah-langkah yang rumit secara manual.

Setelah pembagian dataset selesai dilakukan, langkah selanjutnya adalah memilih algoritma yang akan diterapkan pada dataset yang telah dipisahkan. Dalam konteks ini, algoritma *Naïve Bayes* dipilih karena kemampuannya yang terbukti dalam mengatasi masalah klasifikasi. Keunggulan utama dari algoritma ini

adalah kesederhanaannya yang membuatnya mudah diimplementasikan dan efisien dalam penggunaan sumber daya komputasi.

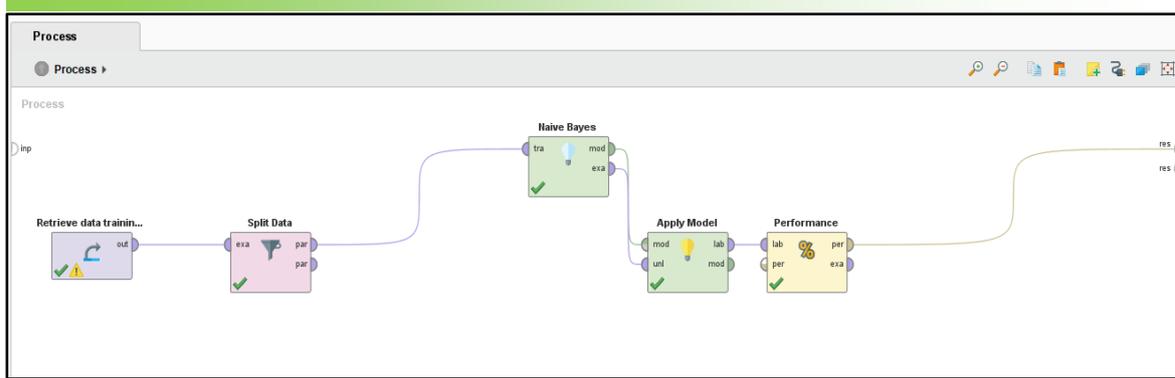
Dengan memilih algoritma *Naïve Bayes*, peneliti dapat melangkah ke tahap berikutnya dalam analisis data. Hasil dari proses ini akan memberikan pemahaman yang lebih dalam tentang pola dan tren dalam dataset yang diproses. Dengan demikian, langkah ini merupakan langkah penting dalam pengembangan model dan evaluasi performanya.

Hasil dari proses ini akan tercermin dalam visualisasi yang disajikan dalam Gambar 7. Gambar ini memberikan gambaran yang jelas tentang tahapan-tahapan yang telah dilalui dataset, mulai dari pembagian menjadi data latih dan data uji hingga pemilihan algoritma *Naïve Bayes* untuk analisis lebih lanjut. Dengan demikian, Gambar 4 memberikan pandangan yang lengkap tentang proses pemrosesan dan analisis data dalam lingkungan *RapidMiner*, menyoroti pentingnya langkah-langkah yang diambil dalam pengembangan model yang efektif.



Gambar 6. Proses Pemodelan Menggunakan *Rapidminer*

Setelah semua komponen terkoneksi dengan baik, langkah selanjutnya adalah memilih tombol "*Apply Model*" dalam aplikasi *RapidMiner* seperti pada Gambar 8. Tindakan ini akan memicu proses perhitungan menggunakan model yang telah dibuat dengan algoritma *Naïve Bayes* berdasarkan data latih yang telah diproses sebelumnya. Hasil dari proses ini akan menunjukkan bagaimana model tersebut berkinerja saat diterapkan pada data uji yang telah dipisahkan sebelumnya. Evaluasi kinerja model akan memberikan pemahaman yang lebih baik tentang seberapa baik model dapat memprediksi jenis bencana berdasarkan fitur-fitur yang ada dalam dataset. Hasil perhitungan ini kemudian dapat digunakan untuk membuat kesimpulan dan mengambil tindakan lebih lanjut, seperti melakukan penyesuaian pada model atau mengambil langkah-langkah mitigasi berdasarkan hasil prediksi. Dengan memilih tombol "*Apply Model*", *RapidMiner* akan menyajikan hasil perhitungan dengan jelas, yang memungkinkan pengguna untuk mengevaluasi dan menginterpretasi hasil dengan tepat.



Gambar 7. Proses Pemodelan Menggunakan Rapidminer

3.2. Evaluasi Hasil *Confusion matrix*

Evaluasi dilakukan untuk memahami tingkat akurasi dari model algoritma yang telah diterapkan pada dataset bencana di wilayah Kabupaten Boyolali. Hasil evaluasi menunjukkan bahwa performa dari algoritma *Naive Bayes* mencapai tingkat akurasi sebesar 92,52%. Ini berarti bahwa model *Naive Bayes* mampu dengan tepat memprediksi jenis bencana berdasarkan fitur-fitur yang ada dalam dataset dengan tingkat keakuratan yang tinggi. Selain itu, nilai precision dari model tersebut mencapai 92,68%, yang mengindikasikan seberapa baik model dalam memprediksi positif secara tepat. Sedangkan recall, atau yang sering disebut sebagai sensitivitas, mencapai 91,64%, yang menunjukkan seberapa baik model dalam menemukan semua instance positif. Hasil evaluasi ini memberikan pemahaman yang lebih baik tentang kinerja model *Naive Bayes* dalam konteks pengklasifikasian bencana di wilayah Kabupaten Boyolali, menunjukkan bahwa model tersebut dapat menjadi alat yang efektif dalam upaya mitigasi bencana dan pengambilan Keputusan, atau bisa dilihat pada Gambar 9 berikut.

accuracy: 92.52%			
	true Angin Puting Beliung	true Tanah Longsor	class precision
pred. Angin Puting Beliung	364	30	92.39%
pred. Tanah Longsor	26	329	92.68%
class recall	93.33%	91.64%	

Gambar 8. Proses Tampilan Hasil Akurasi

Pada tahap ini, juga dilakukan perbandingan antara algoritma *Naive Bayes* dengan dua algoritma lainnya, yaitu algoritma decision tree dan random tree. Hasil perbandingan pada Gambar 10 menunjukkan bahwa algoritma decision tree memperoleh nilai akurasi sebesar 89,32%, dengan nilai precision 84,29% dan recall 80,22%. Sementara itu, algoritma *random tree* mencapai tingkat akurasi sebesar 91,86%, dengan nilai precision 90,25% dan recall 91,23%. Dari hasil ini, dapat disimpulkan bahwa algoritma *random tree* memiliki kinerja yang sedikit lebih baik dibandingkan algoritma *decision tree*, namun kedua algoritma tersebut memiliki tingkat akurasi yang cukup tinggi. Meskipun demikian, algoritma *Naive Bayes* masih tetap menjadi pilihan yang baik dengan tingkat akurasi yang kompetitif sebesar 92,52% serta nilai precision dan recall yang memuaskan. Evaluasi ini memberikan pemahaman yang lebih luas tentang kinerja berbagai algoritma dalam konteks pengklasifikasian bencana di wilayah Kabupaten Boyolali, yang dapat menjadi landasan untuk pengambilan keputusan yang lebih baik dalam upaya mitigasi bencana.

No	Algoritma	Training (%)	Testing (%)	Accuracy (%)	Precision (%)	Recall (%)
1	Naive Bayes	80	20	92,52	92,68	91,64
2	Decision Tree	80	20	89,32	84,29	80,22
3	Random Tree	80	20	91,86	90,25	91,23

Gambar 9. Hasil Evaluasi perbandingan

4. Kesimpulan

Berdasarkan hasil pembahasan, algoritma Naïve Bayes dapat diterapkan untuk mendeteksi pola kejadian bencana menggunakan dataset kejadian bencana dari BPBD Boyolali. Model yang telah dibuat diharapkan menjadi satu opsi alternatif bagi BPBD dalam melakukan mitigasi bencana berdasarkan pola yang terjadi. Hasil penelitian ini menunjukkan bahwa berdasarkan nilai akurasi yang didapatkan, algoritma Naïve Bayes lebih baik dibanding algoritma decision tree dan random tree dalam penelitian ini. Selanjutnya, dari hasil penelitian tersebut, ada beberapa saran bagi penelitian yang akan datang. Pertama, dapat ditambahkan jumlah jenis bencana sebagai label pada penelitian yang akan datang, menyesuaikan dengan lokasi penelitian. Kedua, perlu diuji coba dengan dataset yang lebih besar agar dimungkinkan menghasilkan nilai akurasi yang lebih tinggi. Dan ketiga, pada penelitian yang akan datang, dapat juga ditambahkan curah hujan sesuai wilayah penelitian sehingga mungkin akan menghasilkan hasil yang lebih kompleks dan tentunya bernilai informasi yang sangat tinggi.

5. Daftar Pustaka

- [1] B. Boyolali, "Badan Penanggulangan Bencana Daerah Boyolali," 3 Oktober 2022. [Online]. Available: <https://bpbd.boyolali.go.id/406-rekap-data-kejadian-bencana/688-rekap-data-kejadian-bencana-2022-sampai-juni-6127xlsx>.
- [2] A. Wanto, S. Defit dan A. P. Windarto, "Algoritma Fungsi Pelatihan pada Machine Learning berbasis ANN untuk Peramalan Fenomena Bencana," *JURNAL RESTI(Rekayas a Sistem dan T eknol ogi Informasi)*, vol. 5, no. 2, pp. 254-264, 2021.
- [3] A. Maulana, R. D. D. dan N. D. N. , "Implementasi Algoritma K-Means Clustering dalam Pengelompokan Data Kerusakan Rumah Akibat Bencana Alam di Kabupaten Cirebon," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, 2024.
- [4] M. M. Effendi, "Analysis Prediksi Wilayah Rawan Banjir dengan Algoritma K-Means," *JOURNAL OF INFORMATION SYSTEM RESEARCH (JOSH)*, vol. 5, no. 2, pp. 697-703, 2024.
- [5] T. Yulianto dan F. Faisol, "Clustering Daerah Bencana Alam di Indonesia Menggunakan Metode Fuzzy C-Means," *UJMC (Unisda Journal of Mathematics and Computer science)*, vol. 9, no. 2, pp. 29-39, 31 Desember 2023.
- [6] M. B. Ulum dan F. Badri , "Sistem Monitoring Cuaca dan Peringatan," *JITET (Jurnal Informatika dan Teknik Elektro Terapan)*, vol. 11, no. 3, pp. 319-328, 2023.
- [7] Y. wahyudi, "Sistem Iot Untuk Deteksi," dalam *Seminar Nasional Teknologi dan Rekayasa*, 2020.
- [8] M. Sarosa dan Nailul Muna, "Implementasi Algoritma You Only Look Once (YOLO) untuk deteksi korban bencana alam," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, vol. 8, no. 4, 2021.
- [9] M. A. Islam, M. Z. S. Hadi dan R. Widyatatra, "Sistem Cerdas Pendeteksi Dan Penghitung Jumlah Korban Bencana Alam Menggunakan Algoritma Deep Learning," *JURNAL INOVTEK POLBENG*, vol. 8, no. 1, 2023.
- [10] S. Triyanto, A. Sunyoto dan M. R. Arief, "Analisis Klasifikasi Bencana Banjir Berdasarkan Curah Hujan Menggunakan Algoritma Naïve Bayes," *JOISIE (Journal of Information And Informatics Engineering)*, vol. 5, no. 2, 2021.
- [11] D. Fitriyah, W. Gunawan dan A. P. Sari, "Studi Komparasi Algoritma Klasifikasi C5.0, SVM dan Naive Bayes dengan Studi Kasus Prediksi Banjir," *Jurnal Teknologi Informasi Techno.Com*, vol. 21, no. 1, 2022.
- [12] M. F. Rifai, H. Jatnika dan B. Valentino, "Penerapan Algoritma Naïve Bayes Pada Sistem Prediksi Tingkat Kelulusan Peserta Sertifikasi Microsoft Office Specialist (MOS)," *PETIR (Jurnal Pengkajian Dan Penerapan Teknik Informatika)*, vol. 12, no. 2, 2019.
- [13] H. Susana, "Penerapan Model Klasifikasi Metode Naive Bayes Terhadap Penggunaan Akses Internet," *JURNAL RISET SISTEM INFORMASI DAN TEKNOLOGI INFORMASI (JURISISTEKNI)*, vol. 4, no. 1, 2022.
- [14] N. Nurdin, M. Suhendri, Y. Afrilia dan R. Rizal, "Klasifikasi Karya Ilmiah (Tugas Akhir) Mahasiswa Menggunakan Metode Naive Bayes Classifier (NBC)," *Jurnal Sistem Informasi*, vol. 10, no. 2, 2021.
- [15] Murdiaty, Angela dan C. Sylvia, "Pengelompokan Data Bencana Alam Berdasarkan Wilayah, Waktu, Jumlah Korban dan Kerusakan Fasilitas Dengan Algoritma K-Means," *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 4, no. 3, pp. 744-752, 2020.

-
- [16] T. I. Hermanto dan Y. Muhyidin, "Analisis Sebaran Titik Rawan Bencana dengan K-Means Clustering dalam Penanganan Bencana," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 5, no. 1, pp. 406-416, 2021.
- [17] D. M. Rajagukguk dan M. I. Panjaitan, "Evaluation Of Bolu Menara Sales Data With The C.45 Algorithm Using The Rapid Miner Application," *Journal Of Data Science*, vol. 1, no. 2, 2023.