

Klasifikasi Data Gempa Bumi di Pulau Sumatera Menggunakan Algoritma Naïve Bayes

Tobias Duha¹, Mitranikasih Laia², Amirudin Khorul Huda³, Agung Jasuma⁴

^{1,2}Program Studi Sistem Informasi, Universitas Nias Raya

^{3,4}Program Studi Ilmu Komputer, Universitas Nias Raya
Email: ¹bungtd@uniraya.ac.id, ²mitranikasih@uniraya.ac.id,

³amirudin@uniraya.ac.id, ⁴agungjasuma@uniraya.ac.id

ABSTRAK – Indonesia sebagai salah satu negara yang berada di wilayah lingkaran api pasifik (Ring of Fire), tempat pertemuan tiga lempengan tektonik dunia. Hal ini menyebabkan Indonesia sangat rentan terhadap bencana alam seperti gempa bumi, letusan gunung berapi dan tsunami. Fenomena alam seperti ini sangatlah sering terjadi, salah satunya di Pulau Sumatera. Penelitian ini bertujuan untuk mengklasifikasikan data gempa bumi di Kepulauan Sumatera berdasarkan hiposentrum dengan menggunakan Algoritma Naïve Bayes. Penelitian menggunakan dataset gempa bumi khususnya di Kepulauan Sumatera yang dibagi menjadi data latih dan data uji. Hasil dari penelitian menunjukkan bahwa klasifikasi dapat dilakukan dengan menggunakan Algoritma Naïve Bayes berdasarkan tiga kategori mulai dari gempa dangkal, gempa sedang, maupun gempa dalam.

Kata Kunci: Klasifikasi Gempa, Hiposentrum, Naïve Bayes, Confusion Matrix

ABSTRACT – Indonesia is one of the countries located in the Pacific Ring of Fire, where three tectonic plates meet. This makes Indonesia very vulnerable to natural disasters such as earthquakes, volcanic eruptions, and tsunamis. These natural phenomena occur very frequently, as evidenced by events such as those that have occurred on the island of Sumatra. This study aims to classify earthquake data in the Sumatra Islands based on hypocenter using the Naive Bayes Algorithm. The study uses earthquake datasets specifically from the Sumatra Islands, which are divided into training and testing data. The results of the study indicate that classification can be performed using the Naive Bayes Algorithm based on three categories, ranging from shallow earthquakes, moderate earthquakes, to deep earthquakes.

Keywords: Earthquake Classification, Naïve Bayes, Hypocenter, Confusion Matrix

PENDAHULUAN

Hiposentrum atau titik pusat gempa adalah sumber gempa bumi yang terletak pada kedalaman tertentu di dalam bumi. Kekuatan suatu gempa dipengaruhi oleh posisi hiposentrumnya yang dapat berada pada kedalaman berbeda. Gempa bumi dapat dikategorikan menjadi tiga jenis berdasarkan letak hiposentrumnya, yakni gempa bumi dalam, gempa bumi menengah, dan gempa bumi dangkal. Gempa bumi dalam terjadi ketika hiposentrum berada pada kedalaman >300 km di bawah permukaan bumi. Gempa bumi menengah terjadi ketika hiposentrum berada pada kedalaman antara 70 km hingga 300 km di bawah permukaan bumi. Sementara itu, gempa bumi dangkal terjadi ketika hiposentrum berada pada kedalaman <70 km di bawah permukaan bumi. Karena Indonesia terletak di wilayah kepulauan, maka Indonesia memiliki potensi gempa bumi yang tinggi. Data pada Badan Meteorologi, Klimatologi dan Geofisika (BMKG) mencatat terjadinya gempa

bumi di Indonesia secara berkala dalam selang waktu beberapa bulan saja. Bahkan potensi gempa bumi yang lokasi hiposentrumnya cukup dalam hampir terjadi setiap hari. Klasifikasi kekuatan gempa berdasarkan kedalaman hiposentrum diperlukan untuk mengetahui potensi kekuatan gempa yang tersebar di wilayah Indonesia. Hasil dari proses klasifikasi seringkali mengalami masalah yaitu ketidaktepatan dalam klasifikasi. Algoritme Naïve Bayes merupakan algoritme probabilitas sederhana untuk mengklasifikasikan data berdasarkan probabilitas data dengan menghitung frekuensi dan kombinasi nilai pada dataset yang digunakan. Di dalam algoritme Naïve Bayes setiap data diasumsikan sebagai variabel bebas yang dapat mempertimbangkan nilai dari variabel klasifikasi. Dalam penelitian ini penulis akan melakukan klasifikasi data menggunakan algoritma Naïve Bayes yang diterapkan untuk mengklasifikasikan dataset yang gempa bumi yang telah di processing berdasarkan hiposentrumnya. Dalam penelitian ini

objek penelitian berupa data gempa bumi di Pulau Sumatera yang akan digunakan sebagai data training dan data testing. Berbagai penelitian yang terkait dengan klasifikasi telah banyak dilakukan oleh para peneliti sebelumnya. Klasifikasi Data Mining untuk Prediksi Potensi Nasabah dalam Membuat Deposito Berjangka [1], Analisa Prediksi Efek Kerusakan Gempa dari Magnitudo (Skala Richter) dengan Metode Algoritma ID3 Menggunakan Aplikasi Data Mining Orange [2], Penerapan Data Mining dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandar Lampung dengan Algoritma Naïve Bayes [3], Penerapan Algoritma Decision Tree C4.5 untuk Prediksi Kelayakan Calon Pendonor Darah dengan Klasifikasi Data Mining [4], Penerapan Datamining pada Data Gempa Bumi Terhadap Potensi Tsunami di Indonesia [5].

METODE

Dalam penelitian ini, data histori gempa bumi yang digunakan selama 1 tahun terakhir di Pulau Sumatera dan sekitarnya dari katalog gempa BMKG (Badan Meteorologi, Klimatologi, dan Geofisika) daerah Sumatera dengan koordinat - 8.23° - 8.76° Lintang Selatan dan 91.78° - 108.27° Bujur Timur, mulai 1 Januari 2022 sampai 31 Januari 2023 dengan magnitudo 1 sd 10 Mw dengan kedalaman 1 sd 1000 km. Data yang disediakan oleh BMKG hanya dapat di download per 20 hari, oleh sebab itu data di download per 20 hari selama 1 tahun kemudian disatukan dalam 1 file histori gempa bumi. Variabel penelitian terdiri dari kedalaman hiposentrum, magnitudo, dan lokasi gempa, dengan data primer yang digunakan dalam proses klasifikasi berupa data kedalaman hiposentrum yang akan dibagi menjadi tiga klasifikasi: gempa dalam, gempa menengah, dan gempa dangkal.

2.1. Algoritma Naïve Bayes

Metode klasifikasi Naive Bayes didasarkan Teorema Bayes yang menggunakan probabilitas dan statistik. Teorema ini dikemukakan oleh ilmuwan Inggris bernama Thomas Bayes [6]. Algoritma Naive Bayes merupakan algoritma pembelajaran mesin yang digunakan untuk klasifikasi data. Algoritma ini berdasarkan teorema Bayes, yang menyatakan bahwa probabilitas suatu hipotesis atau kelas dapat dihitung berdasarkan probabilitas data dan probabilitas kondisional [7]. Algoritma Naive Bayes memanfaatkan probabilitas kondisional untuk menghitung probabilitas kelas dari sebuah data. Algoritma Naive Bayes dinamakan "naive" karena

mengasumsikan bahwa setiap fitur atau atribut data independen satu sama lain. Hal ini seringkali tidak realistis di dunia nyata, namun asumsi tersebut mempermudah perhitungan probabilitas kondisional dan membuat algoritma Naive Bayes cepat dan mudah diimplementasikan. Penerapan algoritma Naive Bayes dapat ditemukan dalam berbagai aplikasi, seperti klasifikasi email spam, klasifikasi dokumen teks, dan klasifikasi citra. Algoritma ini dapat memberikan hasil yang baik dalam klasifikasi data dengan jumlah atribut yang besar.

2.2. Gempa Bumi

Gempa bumi adalah kejadian ketika lapisan batuan di kulit bumi bergerak atau bergeser secara tiba-tiba, yang disebabkan oleh aktivitas pergerakan lempeng tektonik atau gunung berapi [8]. Gempa bumi yang terjadi akibat pergerakan lempeng tektonik disebut sebagai gempa bumi tektonik, sedangkan yang disebabkan oleh aktivitas gunung berapi disebut sebagai gempa bumi vulkanik. Saat lapisan batuan bergerak tiba-tiba, terjadi pelepasan energi yang menyebabkan terjadinya gelombang gempa bumi atau seismik [9]. Gempa bumi adalah getaran atau guncangan pada permukaan bumi yang disebabkan oleh pelepasan energi yang terjadi di dalam bumi. Penyebab terjadinya gempa bumi dapat bervariasi, termasuk aktivitas lempeng tektonik, letusan gunung berapi, pelepasan gas alam atau air tanah di dalam bumi, dan aktivitas manusia seperti pengeboran sumur minyak atau gas. Gempa bumi dapat menyebabkan kerusakan yang serius pada bangunan dan infrastruktur, serta menyebabkan korban jiwa dan luka-luka. Tingkat kerusakan yang disebabkan oleh gempa bumi dipengaruhi oleh kekuatan gempa, kedalaman episentrum, dan jarak dari lokasi gempa ke permukaan bumi.

2.3. Tsunami

Tsunami merupakan gelombang besar yang dihasilkan oleh peristiwa seperti gempa bumi, letusan gunung berapi, atau longsor laut yang menyebabkan pergerakan besar pada permukaan laut [10]. Gelombang tsunami dapat merambat dengan kecepatan yang sangat tinggi dan memiliki potensi untuk mengakibatkan kerusakan dan korban jiwa yang besar di wilayah pesisir. Tsunami sering kali disebut sebagai "gelombang laut raksasa" karena dapat mencapai ketinggian puluhan meter dan menyapu bersih apapun yang berada di jalurnya. Untuk mengurangi risiko tsunami, perlu dilakukan

upaya mitigasi bencana, termasuk pemetaan risiko, pembangunan infrastruktur yang tahan gempa dan tsunami, serta edukasi masyarakat mengenai tindakan evakuasi saat terjadi bencana.

2.4. Tahapan Penelitian

Terdapat beberapa tahapan penelitian yang dilakukan, antara lain sebagai berikut:

1. Pengumpulan Data

Tahap awal dari penelitian, yang bertujuan untuk mengumpulkan data yang relevan dan valid untuk dianalisis [11]. Dalam penelitian ini, data primer yang digunakan adalah data gempa bumi khususnya di Pulau Sumatera yang diperoleh dari Badan Meteorologi, Klimatologi dan Geofisik (BMKG).

2. Pre-Processing Data

Tahap ini dilakukan untuk mengolah data yang sudah ada dengan menggunakan normalisasi data. Pre-processing data bertujuan untuk mengolah data sebelum dilakukan analisis dan klasifikasi [12]. Hal ini dilakukan untuk mengatasi masalah yang mungkin terdapat pada data seperti atribut yang datanya hilang.

3. Visualisasi Data

Setelah melakukan pre-processing data, atau data telah di normalisasi, selanjutnya dilakukan penerapan pada algoritma Naive Bayes untuk menentukan dan melihat proses klasifikasi nya.

4. Evaluasi Algoritma Klasifikasi

Tahap ini meliputi pengujian dan analisis hasil dari algoritma yang digunakan. Hasil yang diperoleh dari algoritma akan dibandingkan dengan data yang sebenarnya untuk menentukan seberapa baik algoritma tersebut dalam mengklasifikasikan data. Metode klasifikasi Naive Bayes yang telah dioptimalkan pada data hiposentrum gempa dievaluasi dengan menggunakan Algoritma Confusion Matrix [13]. Klasifikasi dibagi menjadi tiga kelas,

sehingga dapat dilakukan pengukuran akurasi, presisi, dan recall dengan cara menghitung rata-rata dari nilai akurasi, presisi, dan recall pada setiap kelas [14]. Selain itu, dapat juga dilakukan analisis lebih lanjut dengan mengevaluasi Confusion Matrix dari setiap kelas untuk mengetahui tingkat kesalahan dalam klasifikasi.

$$Akurasi = \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i} \times 100\% \quad (1)$$

Dengan mengetahui nilai dari TP_i , TN_i , FN_i dan FP_i , maka dapat dilakukan perhitungan metrik precision, recall, dan akurasi dari algoritme klasifikasi yang digunakan [1]. Dimana:

- True Positive (TP_i) adalah jumlah data yang diklasifikasikan sebagai positif oleh sistem dan benar sesuai dengan kelas ke-i.
- True Negative (TN_i) adalah jumlah data yang diklasifikasikan sebagai negatif oleh sistem dan benar sesuai dengan kelas ke-i.
- False Negative (FN_i) adalah jumlah data yang seharusnya diklasifikasikan sebagai negatif namun diklasifikasikan salah oleh sistem untuk kelas ke-i.
- False Positive (FP_i) adalah jumlah data yang seharusnya diklasifikasikan sebagai positif namun diklasifikasikan salah oleh sistem untuk kelas ke-i.

2.5. Analisis Data Penelitian

Dalam penelitian, data primer yang digunakan masih berupa data mentah yang belum pre-processing atau belum diolah. Terdapat sekitar 2460 data yang akan digunakan dalam penelitian. Data mentah yang sudah ada perlu kembali untuk diolah atau di pre-processing. Proses dalam pengolahan data dilakukan dengan menggunakan normalisasi data [15].

Tabel 1. Data Gempa di Pulau Sumatera

No	Date	Time	Lat	Lon	Depth	Mag	Region
1	28/01/2023	24:31,5	2.99 S	102.96 E	207	4,1	Southern Sumatra, Indonesia
2	28/01/2023	12:32,0	6.69 S	106.13 E	130	3,3	Java, Indonesia
3	27/01/2023	34:13,8	7.44 N	93.02 E	116	4,7	Nicobar Islands, India Region
4	27/01/2023	12:05,1	1.68 N	99.41 E	160	4,4	Northern Sumatra, Indonesia
5	24/01/2023	09:10,2	5.42 S	106.66 E	281	3,4	Java, Indonesia
6	23/01/2023	07:26,8	2.62 S	102.43 E	184	3,3	Southern Sumatra, Indonesia
7	22/01/2023	28:01,5	2.04 N	98.94 E	118	4,1	Northern Sumatra, Indonesia
8	21/01/2023	55:54,7	2.13 S	102.17 E	188	4,4	Southern Sumatra, Indonesia

9	20/01/2023	07:23,5	0.90 N	100.07 E	175	3,6	Northern Sumatra, Indonesia
10	18/01/2023	05:30,5	4.41 S	104.76 E	140	2,9	Southern Sumatra, Indonesia
11	15/01/2023	37:59,7	1.11 N	99.92 E	170	3,3	Northern Sumatra, Indonesia
12	14/01/2023	32:18,1	6.51 S	106.53 E	148	4,7	Java, Indonesia
13	13/01/2023	09:06,9	0.66 N	99.60 E	105	2,9	Northern Sumatra, Indonesia
14	13/01/2023	04:00,6	4.31 N	96.77 E	130	3,3	Northern Sumatra, Indonesia
15	13/01/2023	19:22,6	4.27 N	96.81 E	128	3,9	Northern Sumatra, Indonesia
16	12/01/2023	01:01,4	2.76 N	98.93 E	152	2,8	Northern Sumatra, Indonesia
17	12/01/2023	44:53,6	2.33 S	101.97 E	157	4	Southern Sumatra, Indonesia
18	10/01/2023	22:23,3	2.14 S	101.70 E	121	2,5	Southern Sumatra, Indonesia
19	30/12/2022	19:45,4	0.86 N	99.69 E	131	4,2	Northern Sumatra, Indonesia
20	29/12/2022	56:04,4	3.41 N	98.19 E	107	1,9	Northern Sumatra, Indonesia

HASIL DAN PEMBAHASAN

3.1. Preprocessing Data

Preprocessing data menggunakan normalize features bertujuan untuk melakukan normalisasi nilai dari fitur-fitur dalam dataset agar memiliki skala yang sama. Hal ini dilakukan untuk menghindari bias pada model yang disebabkan oleh perbedaan skala antar fitur. Hasil dari preprocessing data menggunakan normalize features yang dimana setiap fitur dalam dataset akan memiliki nilai yang terdistribusi dalam rentang yang sama.

3.2. Hasil Klasifikasi Gempa Bumi

Hasil klasifikasi gempa bumi diklasifikasikan berdasarkan kedalamannya, yaitu gempa dangkal dengan hiposentrum kurang dari 70 km, gempa sedang dengan hiposentrum antara 70 km hingga 300 km, dan gempa dalam dengan hiposentrum lebih dari 300 km. Akurasi hasil klasifikasi dihitung dengan mempertimbangkan ketepatan dalam mengklasifikasikan gempa-gempa tersebut, yang dimana dibuat dalam tiga kategori mulai dari gempa dangkal, sedang, maupun dalam. Confusion matrix digunakan sebagai Algoritme dalam menghitung akurasi.

Tabel 2. Hasil Confusion Matrix menggunakan Naïve Bayes

		Predicted			
		Dangkal	Sedang	Dalam	Σ
Actual	Dangkal	0	0	3	3
	Sedang	0	1429	88	1517
	Dalam	0	19	183	202
	Σ	0	1448	274	1722

Dengan nilai Test and Score sebagai berikut:

AUC = 0.962
CA = 0.936
F1 = 0.939
Precision = 0.948
Recall = 0.936

KESIMPULAN DAN SARAN

Berdasarkan penelitian yang dilakukan, dapat disimpulkan bahwa algoritma Naïve Bayes dapat digunakan untuk melakukan klasifikasi data gempa bumi di Pulau Sumatera dengan cukup baik. Dalam penelitian ini, data gempa bumi di Pulau Sumatera diklasifikasikan menjadi tiga kategori, yaitu gempa dangkal, sedang, dan dalam. Algoritma Naïve Bayes memiliki asumsi dasar yang cukup sederhana, yaitu setiap fitur dianggap saling independen. Oleh karena itu, perlu dilakukan pengecekan lebih lanjut terhadap pengaruh korelasi antar fitur terhadap akurasi klasifikasi. Selain itu, dapat juga dilakukan penelitian lebih lanjut dengan menggunakan algoritma lainnya untuk membandingkan performa klasifikasi yang dihasilkan.

UCAPAN TERIMA KASIH

Terima kasih kepada semua pihak yang meluangkan waktu dalam memberikan arahan, bimbingan, dan masukan yang sangat berharga. Terima kasih juga kepada pihak-pihak yang telah memberikan dukungan, serta turut dalam membantu mendukung proses penelitian baik dari segi moril maupun material.

REFERENSI

- [1] N. Nuraeni, "Klasifikasi Data Mining untuk Prediksi Potensi Nasabah dalam Membuat Deposito Berjangka Data Mining Classification for Predicting Customer Potential in Making Term Deposits," *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, vol. 3, no. 01, pp. 65–75, 2021.
- [2] L. Irawan and L. Hermawan Hasibuan, "ANALISA PREDIKSI EFEK KERUSAKAN GEMPA DARI MAGNITUDO (SKALA RICHTER) DENGAN METODE ALGORITMA ID3 MENGGUNAKAN APLIKASI DATA MINING ORANGE," *Jurnal Teknologi Informasi*, vol. 14, no. 2, pp. 189–201, 2020, doi: 10.47111/JTI.
- [3] R. I. Borman and M. Wati, "Penerapan Data Mining Dalam Klasifikasi Data Anggota Kopdit Sejahtera Bandarlampung Dengan Algoritma Naïve Bayes," *JURNAL ILMIAH FAKULTAS ILMU KOMPUTER*, vol. 9, no. 1, pp. 25–34, 2020.
- [4] Y. Irawan, "Penerapan Algoritma Decision Tree C4.5 Untuk Prediksi Kelayakan Calon Pendonor Darah Dengan Klasifikasi Data Mining," vol. 2, no. 4, pp. 181–189, 2021.
- [5] D. P. Utomo and B. Purba, "Penerapan Datamining pada Data Gempa Bumi Terhadap Potensi Tsunami di Indonesia," *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, pp. 846–853, 2019.
- [6] N. Paramitha, E. Junianto, and S. Susanti, "Penerapan Teorema Bayes Untuk Diagnosis Penyakit Pada Ibu Hamil Berbasis Android," *JURNAL INFORMATIKA*, vol. 6, no. 1, pp. 53–61, 2019, [Online]. Available: <http://ejournal.bsi.ac.id/ejurnal/index.php/ji/article/view/4693>
- [7] I. Riswanto and R. H. Laluma, "KLASIFIKASI KELAYAKAN PINJAMAN PADA KOPERASI KARYAWAN MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER BERBASIS WEB," *Infotronik : Jurnal Teknologi Informasi dan Elektronika*, vol. 5, no. 1, pp. 11–16, Jun. 2020, doi: 10.32897/infotronik.2020.5.1.2.
- [8] Z. Bahri and M. Mungkin, "Penggunaan SCR Sebagai Alarm Peringatan Dini Pada Saat Terjadi Gempa Bumi," *Journal of Electrical Technology*, vol. 4, no. 3, pp. 101–105, 2019.
- [9] H. S. Kuyuk and O. Susumu, "Real-time classification of earthquake using deep learning," in *Procedia Computer Science*, Elsevier B.V., 2018, pp. 298–305. doi: 10.1016/j.procs.2018.10.316.
- [10] P. Artawan and K. Purnamawan, "RANCANGAN DETEKTOR GEMPA BERPOTENSI TSUNAMI BERBASIS WIRELESS SENSOR NETWORK DENGAN SISTEM MAGNETIC ALTITUDE (Kajian Teori)."
- [11] F. Nurona Cahya, N. Hardi, D. Riana, and S. Hadianti, "Klasifikasi Penyakit Mata Menggunakan Convolutional Neural Network (CNN)," *SISTEMASI: Jurnal Sistem Informasi*, vol. 10, no. 3, pp. 618–626, 2021, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [12] I. Widhi Saputro and B. Wulan Sari, "Uji Performa Algoritma Naïve Bayes untuk Prediksi Masa Studi Mahasiswa Naïve Bayes Algorithm Performance Test for Student Study Prediction," *Citec Journal*, vol. 6, no. 1, 2019.
- [13] M. Fluorida Fibrianda and A. Bhawiyuga, "Analisis Perbandingan Akurasi Deteksi Serangan Pada Jaringan Komputer Dengan Metode Naïve Bayes Dan Support Vector Machine (SVM)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 2, no. 9, pp. 3112–3123, 2018, [Online]. Available: <http://j-ptiik.ub.ac.id>
- [14] D. Azzahra Nasution, H. H. Khotimah, and N. Chamidah, "PERBANDINGAN NORMALISASI DATA UNTUK KLASIFIKASI WINE MENGGUNAKAN ALGORITMA K-NN," 2019.