

OPTIMASI PENENTUAN LOKASI BENCANA ALAM DENGAN REGRESI LINIER SEDERHANA DAN BERGANDA

Dwiki Jatikusumo^{1*}, Rahmat Rian Hidayat²

^{1,2} Prodi Teknik Informatika/Universitas Mercu Buana; Jl. Meruya Selatan No.01, DKI Jakarta, 11650; 021 5840816

Received: 26 Agustus 2024
Accepted: 5 Oktober 2024
Published: 12 Oktober 2024

Keywords:

disaster; linear regression;
multiple linear regression;
multiplatform.

Correspondent Email:

dwiki.jatikusumo@mercubua.ac.id

Abstrak. Pada tahun akhir-akhir ini, beberapa bencana terjadi di Indonesia khususnya daerah sekitar Jabodetabek, seperti banjir, cuaca ekstrem, tanah longsor dan kebakaran. Dari beberapa bencana yang terjadi tersebut, dapat dibuatkan sebuah sistem pendeteksi lokasi kejadian bencana alam dari berbagai data sumber *website* berita nasional. Kemudian dijadikan data proses yang dikombinasikan dengan algoritma regresi linier sederhana dan regresi linier berganda. Dengan adanya kombinasi algoritma tersebut, diharapkan mendapat akurasi yang tinggi dalam menentukan lokasi kejadian bencana alam berbasis *multiplatform*.

Abstract. In recent years, several disasters have occurred in Indonesia, especially in the area around Jabodetabek, such as floods, extreme weather, landslides, and fires. From some of these disasters that occur, a system can be made to detect the location of natural disasters from various national news website data sources. Then it is used as process data combined with simple linear regression and multiple linear regression algorithms. With the combination of these algorithms, it is expected to achieve high accuracy in determining the location of multiplatform-based natural disasters.

1. PENDAHULUAN

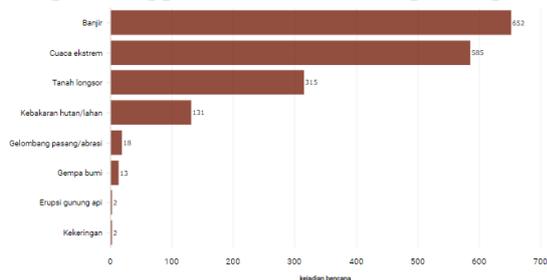
Pada tahun 2023, fenomena bencana alam banyak terjadi khusus-nya di Indonesia. Berikut ada data-data yang dirangkum oleh website katadata.co.id [1]. Badan Nasional Penanggulangan Bencana (BNPB) mencatat, terdapat 1.718 peristiwa bencana alam di Indonesia selama periode 1 Januari-8 Juni 2023. Banjir masih mendominasi jumlah bencana alam di Tanah Air dengan total 652 kejadian. Jumlah ini setara 37,95% dari total kejadian bencana alam nasional pada periode tersebut.

Berikutnya, ada 585 peristiwa cuaca ekstrem yang dilaporkan terjadi pada periode sama. Kemudian ada 315 kejadian tanah longsor, 131

kebakaran hutan dan lahan (karhutla), 18 gelombang pasang/abrasi, 13 gempa bumi, serta erupsi gunung api dan kekeringan masing-masing 2 kejadian. Berdasarkan wilayahnya, Jawa Barat adalah provinsi yang paling banyak mengalami bencana alam selama periode 1 Januari-8 Juni 2023, yaitu 324 kejadian. Diikuti Jawa Tengah dan Jawa Timur masing-masing 316 kejadian dan 88 kejadian. Menurut data BNPB, seluruh kejadian bencana itu membuat 2,85 juta orang menderita dan mengungsi, 5.487 orang luka-luka, 154 orang meninggal dunia, dan 8 orang hilang.

Menurut data BNPB, seluruh kejadian bencana itu membuat 2,85 juta orang menderita

dan mengungsi, 5.487 orang luka-luka, 154 orang meninggal dunia, dan 8 orang hilang.



Gambar 1. Data Kejadian Bencana Alam

Menurut data BNPB, seluruh kejadian bencana itu membuat 2,85 juta orang menderita dan mengungsi, 5.487 orang luka-luka, 154 orang meninggal dunia, dan 8 orang hilang.

Dari grafik tersebut, fokusnya terhadap banjir, cuaca ekstrem, tanah longsor dan kebakaran. Dikarenakan, empat teratas daftar bencana alam di Indonesia.

2. TINJAUAN PUSTAKA

Regresi Linear sederhana merupakan salah satu metode yang ada pada data mining yang dimana Regresi Linear sendiri merupakan suatu metode statistik yang dapat berfungsi untuk menguji sejauh mana hubungan atau korelasi antar sebab akibat dan antar variabel faktor penyebab (X) terhadap variabel [2]. Regresi linier merupakan satu cara prediksi yang menggunakan garis lurus untuk menggambarkan hubungan diantara dua variabel atau lebih. Variabel adalah besaran yang berubah-ubah nilainya. Selanjutnya variabel tersebut terbagi atas dua jenis yaitu variabel pemberi pengaruh dan variabel terpengaruh. Variabel pemberi pengaruh dapat dianalogikan sebab, sementara variabel terpengaruh merupakan akibat, Selanjutnya peramalan ini didasarkan pada asumsi bahwa pola pertumbuhan data historis yang bersifat linier, walaupun sebenarnya tidak 100% linier [3]. Regresi mengukur seberapa besar suatu variabel dapat mempengaruhi variabel lain sehingga nilai suatu variabel berdasarkan variabel lain dapat diprediksi. Prediksi adalah dugaan atau perkiraan tentang terjadinya suatu peristiwa atau kejadian di masa depan yang dapat bersifat kualitatif (tidak berupa angka) atau kuantitatif (berupa angka) [4].

Regresi linier berganda dipakai agar mencari korelasi antara variabel dependen terhadap

lebih dari satu variabel independen. Yang dimana, pada persamaan regresi linier berganda ini digunakan untuk mendapatkan penjelasan karakteristik hubungan atau korelasi yang terdapat pada data, meskipun ada beberapa variabel yang terabaikan [5]. Pengembangan dari model regresi linier sederhana adalah model regresi linier berganda. Untuk mengetahui hubungan antara variabel dependen dengan paling sedikit dua variabel bebas, regresi linier berganda merupakan salah satu metode [6]. Analisis regresi telah lama dikembangkan untuk mempelajari pola dan mengukur hubungan statistik antara dua atau lebih peubah (variabel) [7]. Regresi linier berganda merupakan model persamaan yang menjelaskan hubungan satu variabel tak bebas/ response (Y) dengan dua atau lebih variabel bebas/predictor (X_1, X_2, \dots, X_n). memprediksi nilai variabel tak bebas/ response (Y) apabila nilai-nilai variabel bebasnya/ predictor (X_1, X_2, \dots, X_n) diketahui. Disamping itu juga untuk dapat mengetahui bagaimanakah arah hubungan variabel tak bebas dengan variabel-variabel bebasnya [3]. Analisis regresi linear berganda diperlukan untuk mengetahui koefisien-koefisien regresi serta signifikasnsi sehingga dapat digunakan dalam menjawab hipotesis yang ada. Multiple linear regression (MLR) yaitu model regresi linear berganda atau lebih yang dimana cara kerjanya melakukan semacam prediksi atau peramalan dalam berbentuk data. Data yang dihasilkan dari pemodelan MLR tersebut adalah data pengaruhnya dari beberapa variabel independen terhadap variabel dependen sehingga terjadi tingkat akurasi dari kedua jenis variabel tersebut yang biasa disebut korelasi [8].

3. METODE PENELITIAN

Tahapan pengembangan yang digunakan penulis dalam riset ini yaitu menggunakan model SDLC (*System Development Life Cycle*) dengan menggunakan model/metode *agile development* pendekatan *Dynamic System Development Method* disingkat DSDM [9]. Model ini dapat dijelaskan berikut [10]:

1. Tahap studi kelayakan yang merupakan langkah awal dalam metode DSDM, pada langkah ini proyek akan dipelajari untuk memutuskan apakah DSDM layak digunakan dalam pengembangan atau tidak. Selain itu,

analisis risiko dan penetapan persyaratan teknis dilakukan pada tahap ini. Laporan kelayakan dan garis besar rencana pengembangan merupakan hasil dari langkah ini.

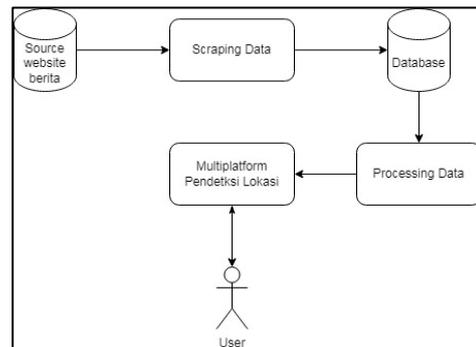
2. Fase studi bisnis, pada fase ini dilakukan pertemuan antara pakar bisnis dari sisi klien dan tim pengembangan, pertemuan ini dilakukan untuk menentukan kebutuhan pengguna dan membuat daftar fungsi yang diperlukan dalam sistem beserta prioritasnya. Teknologi yang akan digunakan ditentukan dalam langkah ini, juga deskripsi proses tingkat tinggi, Diagram *Entity Relationship*, mode objek, arsitektur sistem, arsitektur sistem, dan rencana pembuatan prototipe secara garis besar adalah hasil dari langkah ini.

3. Iterasi model fungsional, pada tahap ini analisis, *coding* dan *prototyping* dilakukan secara iteratif dan inkremental, *porotype* dianalisis untuk menyempurnakan model analisis. Tujuan utama dari langkah ini adalah untuk menentukan apa yang akan dikembangkan dalam perangkat lunak, kapan akan dikembangkan dan cara pengembangannya. Juga, fungsi yang diprioritaskan, pembuatan prototipe fungsional, persyaratan nonfungsional dan analisis risiko.

4. Fase desain dan pembangunan, fase ini juga merupakan fase iteratif, pada fase ini persyaratan yang diidentifikasi dari fase sebelumnya diimplementasikan dan diberi kode, kemudian akan dipublikasikan dan diuji oleh pengguna. Umpan balik pengguna digunakan untuk menyempurnakan sistem secara berulang. Jadi, perangkat lunak yang diuji dengan setidaknya persyaratan minimal adalah hasil dari langkah ini.

5. Tahap Implementasi, pada tahap ini perangkat lunak berpindah dari pengembangan ke produksi dan akan ditangani oleh pengguna, diberikan sesi pelatihan. Juga, panduan pengguna dikembangkan dan ditangani. Jika perangkat lunak memenuhi semua kebutuhan pengguna, maka tidak diperlukan proses pengembangan lainnya. Jika tidak, seluruh proses akan diulangi lagi.

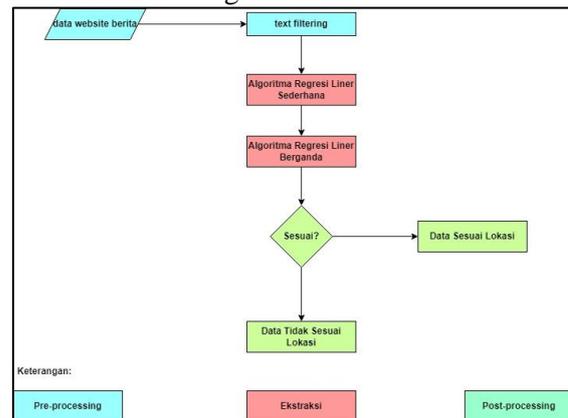
3.1. Desain Sistem



Gambar 2. Rancangan Usecase

Pada gambar di atas yang dibutuhkan pada pendeteksi lokasi kejadian bencana alam adalah data dari *website berita*, kita ambil dari internet kemudian disimpan di dalam database dan akan diperlihatkan peta dari lokasi kejadian bencana alam berbasis *multiplatform*.

Dalam pengesktrasian data dari data mentah menjadi data yang diolah dengan menentukan lokasi adalah sebagai berikut:



Gambar 2. Rancangan Usecase

1. *Text Filtering*, proses pemfilteran meningkatkan efisiensi algoritme jaro winkler dengan secara rasional mengurangi ukuran file teks yang diimpor. Ini menghilangkan kata-kata berulang yang tidak mengubah arti kalimat dan tidak memiliki nilai apa pun (misalnya, kata depan dan kata berhenti). Selain itu, menghapus semua *hyperlink*, gambar, rekaman audio, dan video.
2. Kemudian diproses menggunakan algoritma regresi linier sederhana dan regresi linier berganda, sampai data sesuai dengan lokasi.

4. HASIL DAN PEMBAHASAN

Dari desain sistem yang sudah disiapkan sebelumnya, berikut proses *preprocessing* yang dilakukan. Dalam pengesktrasian data dari data mentah menjadi data yang diolah dengan menentukan lokasi adalah sebagai berikut:

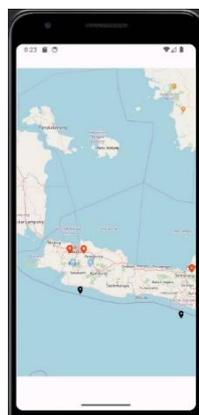
1. *Text filtering*
2. Tidak menggunakan algoritma regresi linier sederhana dan berganda
3. Menggunakan algoritma regresi linier sederhana dan berganda

Setelah *text filtering*, selanjutnya menggunakan algoritma regresi linier sederhana dan berganda yang dilakukan untuk mencari kata-kata lokasi yang sesuai dengan hasil pencarian data dari website berita.

Berikut contoh hasil dari pemetaan dengan menggunakan Openstreetmap API di browser dan aplikasi mobile Android menggunakan emulator. Jadi dari sini mendapatkan letak posisi kejadian bencana alam gempa bumi, banjir, kebakaran hutan, dan tanah longsor dari data website berita yang ada dari bulan Januari sampai Maret 2024. Warna merah merupakan icon untuk banjir, warna biru untuk tanah longsor, warna oranye untuk kebakaran hutan, dan warna hitam untuk gempa bumi.



Gambar 3. Hasil data proses tampilan web



Gambar 4. Hasil data proses tampilan *mobile* Android

Data diproses menggunakan *preprocessing* seperti tahap pada metodologi riset ini, data yang diolah sebanyak 6283 data data *website* berita sebelumnya sudah disimpan di dalam database untuk keperluan riset ini, dari bulan Januari sampai Maret 2024 hasilnya adalah 5381 yang terdapat kemungkinan lokasi kejadian berupa kata lokasi yang dicari setelah melakukan *text filtering*. Kemudian, dilakukan perhitungan akurasi terdapat yang kejadian dan tidak kejadian menggunakan tabel 2, 3, dan 4.

Selanjutnya adalah perbandingan yang merupakan kejadian atau tidak dengan ujicoba tanpa dan menggunakan algoritma regresi linier sederhana dan berganda. Di sini membedakan dari kata dan kalimat berikut “gempa bumi”, “banjir”, “kebakaran hutan”, “tanah longsor”. Dari beberapa referensi untuk mengetahui akurasinya [2], [5], [11], [12], [13], [14], [15], [16], [17].

4.1. Tidak menggunakan algoritma regresi linier sederhana dan berganda

Tabel 1 Tanpa Algoritma Regresi Linier Sederhana Dan Berganda

No	Kata dan Kalimat	Kejadian	Tidak Kejadian	Total	Persentase Akurasi
1	“gempa bumi”	1023	392	1415	72,29%
2	“kebakaran hutan”	983	328	1311	74,98%
3	“banjir”	974	377	1351	72,09%
4	“tanah longsor”	923	381	1304	70,78%

4.2. Menggunakan algoritma regresi linier sederhana dan berganda

Tabel 2 Menggunakan Algoritma Regresi Linier Sederhana Dan Berganda

No	Kata dan Kalimat	Kejadian	Tidak Kejadian	Total	Persentase Akurasi
1	“gempa bumi”	1023	392	1415	72,29%
2	“kebakaran hutan”	983	328	1311	74,98%
3	“banjir”	974	377	1351	72,09%
4	“tanah longsor”	923	381	1304	70,78%

4.3. Perbandingan

Berikut dari hasil perbandingan dari dua percobaan yang dilakukan.

Tabel 3 Hasil Rata-Rata

No	Kata dan Kalimat	Tanpa algoritma regresi linier sederhana dan berganda	Menggunakan algoritma regresi linier sederhana dan berganda
1	“gempa bumi”	72,29%	83,6%
2	“kebakaran hutan”	74,98%	89,39%
3	“banjir”	72,09%	93,19%
4	“tanah longsor”	70,78%	87,96%

5. KESIMPULAN

- a. Kesimpulan dari penelitian yang telah dilakukan, dari pengolahan data kejadian bencana alam didapat lokasi kejadiannya yang tidak menggunakan algoritma regresi linier sederhana dan berganda terdapat besaran akurasi 72,54%. Kemudian untuk data yang diproses menggunakan algoritma algoritma regresi linier sederhana dan berganda rata-rata 88,53%. Dalam perhitungan yang sudah dilakukan bisa dikatakan setidaknya lebih besar menggunakan algoritma regresi linier sederhana dan berganda dibandingkan tidak menggunakan algoritma dan teknik tersebut.
- b. Kedepannya, pengembangan penelitian selanjutnya dengan kombinasi dari algoritma regresi linier sederhana dan berganda tersebut, memungkinkan untuk membuat algoritma terbaru khususnya untuk algoritma pendeteksi lokasi dengan sumber website berita.

UCAPAN TERIMA KASIH

Ucapan terima kasih ditujukan kepada Lembaga Penelitian dan Pengabdian Masyarakat Universitas Mercu Buana dengan nomor kontrak 02-5/635/B-SPK/III/2024, sebagai pemberian dana untuk penelitian ini dilakukan.

DAFTAR PUSTAKA

- [1] katadata.co.id, “Ada 1,7 Ribu Bencana Alam di Indonesia hingga Awal Juni 2023, Mayoritas Banjir,” katadata.co.id, 2023. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2023/06/08/ada-17-ribu-bencana-alam-di-indonesia-hingga-awal-juni-2023-mayoritas-banjir>
- [2] S. Lestari, “Analisis Algoritma Regresi Linear Sederhana dalam Memprediksi Tingkat Penjualan Album KPOP,” *INSOLOGI: Jurnal Sains dan Teknologi*, vol. 2, no. 1, pp. 199–209, 2023, doi: 10.55123/insologi.v2i1.1692.
- [3] W. A. L. Sinaga, S. Sumarno, and I. P. Sari, “The Application of Multiple Linear Regression Method for Population Estimation Gunung Malela District,” *JOMLAI: Journal of Machine Learning and Artificial Intelligence*, vol. 1, no. 1, pp. 55–64, Mar. 2022, doi: 10.55123/jomlai.v1i1.143.
- [4] H. Hasanah, A. Farida, and P. P. Yoga, “Implementation of Simple Linear Regression for Predicting of Students’ Academic Performance in Mathematics,” *Jurnal Pendidikan Matematika (Kudus)*, vol. 5, no. 1, p. 38, Jun. 2022, doi: 10.21043/jpmk.v5i1.14430.
- [5] G. J. Simanjuntak, T. J. Marpaung, and Y. B. Siringoringo, “Penerapan Metode Regresi Linier Berganda Terhadap Pengaruh Jumlah Karyawan dan Jumlah Alat / Mesin,” *Jurnal Penelitian dan Pengabdian Masyarakat (COMSERVA)*, vol. 2, no. 3, pp. 246–266, 2022.
- [6] A. Dharmawan, T. Purwani, Y. Prihati, and C. P. Putri, “Machine Learning Implementation for Profit Estimation,” 2023. [Online]. Available: <http://ijsoc.goacademica.com>
- [7] S. Sulistyono and W. Sulistiyowati, “Peramalan Produksi dengan Metode Regresi Linier Berganda,” *PROZIMA (Productivity, Optimization and Manufacturing System Engineering)*, vol. 1, no. 2, pp. 82–89, Dec. 2017, doi: 10.21070/prozima.v1i2.1350.
- [8] S. Eta Agustin, “The Influence of Sugar Consumption and International Sugar Prices toward the Volume of Sugar Import

- in Indonesia,” 2021. [Online]. Available: <https://jurnal.polines.ac.id/index.php/jobs>
- [9] C. Putra and Y. Nora, “Application of Dynamic Systems Development Method in WEB-Based Promotion Media,” vol. 1, no. 3, pp. 196–204, 2020.
- [10] S. Al-saqqa, S. Sawalha, and H. Abdelnabi, “Agile Software Development : Methodologies and Trends,” *International Journal of Interactive Mobile Technologies (ijim)*, vol. 14, no. 11, pp. 246–270, 2020, doi: <https://doi.org/10.3991/ijim.v14i11.13269>.
- [11] M. Sitanggang and E. Simamora, “Increasing Accuracy of Classification in C4.5 Algorithm by Applying Principle Component Analysis for Diabetes Diagnosis,” vol. 6, no. 2, 2022, doi: [10.25217/numerical.v6i2](https://doi.org/10.25217/numerical.v6i2).
- [12] H. Husdi and H. Dalai, “Penerapan Metode Regresi Linear Untuk Prediksi Jumlah Bahan Baku Produksi Selai Bilfagi,” *Jurnal Informatika*, vol. 10, no. 2, pp. 129–135, Oct. 2023, doi: [10.31294/inf.v10i2.14129](https://doi.org/10.31294/inf.v10i2.14129).
- [13] W. Windasari, Wasiman, and A. Alfiyanto, “PERBANDINGAN AKURASI PREDIKSI METODE REGRESI LINEAR OLS DENGAN JARINGAN SYARAF TIRUAN: STUDI KASUS DATASET STARTUP,” *Journal of Data Science Theory and Application*, vol. 2, no. 1, pp. 17–22, 2023.
- [14] Z. Muttaqin and E. Srihartini, “PENERAPAN METODE REGRESI LINIER SEDERHANA UNTUK PREDIKSI PERSEDIAAN OBAT JENIS TABLET,” *Sistem Informasi*, vol. 9, no. 1, pp. 12–16.
- [15] R. D. Shaputra and S. Hidayat, “Implementasi regresi linear untuk prediksi penjualan pada aplikasi point of sales restoran,” *Automata*, vol. 2, p. 1, 2021, [Online]. Available: <https://103.220.113.195/AUTOMATA/article/view/17355>
- [16] A. Shabrina, P. Prasmono, and A. Ahdika, “Analisis Regresi Berganda pada Faktor-Faktor yang Mempengaruhi Kinerja Fisik Preservasi Jalan dan Jembatan Di Provinsi Sumatera Selatan,” vol. 1, no. 1, pp. 47–56, 2023.
- [17] M. F. Ruziqiana, L. Hidayah, and M. A. Rasyidi, “DETECTION OF CYBERBULLYING USING SVM, NAIVE BAYES, AND RANDOM FOREST ALGORITHM,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, pp. 2830–7062, 2024, doi: [10.23960/jitet.v12i3.5283](https://doi.org/10.23960/jitet.v12i3.5283).