



Penerapan *Data Mining* untuk Prediksi Jumlah Penduduk Miskin di Provinsi Sumatera Selatan Menggunakan Metode Regresi Linier Berganda

Muhammad Roihan Jannatun Adhen¹, Dafid²

¹Universitas Multi Data Palembang, Sumatera Selatan, Indonesia, mroihanadhen@gmail.com

²Universitas Multi Data Palembang, Sumatera Selatan, Indonesia, dafid@mdp.ac.id

Corresponding Author: mroihanadhen@gmail.com¹

Abstract: Poverty is a serious multidimensional problem in Indonesia that requires effective policies based on accurate data. This study aims to predict the number of poor people in South Sumatra Province using the Multiple Linear Regression method. The methodology employed is Knowledge Discovery in Databases (KDD), encompassing stages of data cleaning, integration, selection, transformation, data mining, and knowledge presentation. Data was sourced from the Central Bureau of Statistics (BPS) of South Sumatra Province for the 2014-2024 period, with variables including the poverty line, per capita expenditure, HDI, average years of schooling, access to clean water, unemployment, and labor force participation. The results indicate that the multiple linear regression model performs well, with evaluation values of RMSE at 15.759, MAE at 12.293, and MAPE at 23.17%. These prediction results are integrated into a web-based system to facilitate data visualization for users. It can be concluded that the multiple linear regression algorithm can be used to predict the number of poor people with high accuracy to assist in planning poverty alleviation programs.

Keyword: Poverty, Data Mining, Multiple Linear Regression, KDD, South Sumatra

Abstrak: Kemiskinan merupakan masalah multidimensional serius di Indonesia yang memerlukan kebijakan efektif berdasarkan data akurat. Penelitian ini bertujuan untuk memprediksi jumlah penduduk miskin di Provinsi Sumatera Selatan menggunakan metode Regresi Linier Berganda. Metodologi yang digunakan adalah *Knowledge Discovery in Databases* (KDD) yang mencakup tahapan *data cleaning*, *integration*, *selection*, *transformation*, *data mining*, hingga *knowledge presentation*. Data bersumber dari Badan Pusat Statistik (BPS) Provinsi Sumatera Selatan periode 2014-2024 dengan variabel meliputi garis kemiskinan, pengeluaran perkapita, IPM, rata-rata lama sekolah, akses air minum, pengangguran, dan partisipasi angkatan kerja. Hasil penelitian menunjukkan bahwa model regresi linier berganda memiliki kinerja yang baik dengan nilai evaluasi RMSE sebesar 15,759, MAE sebesar 12,293, dan MAPE sebesar 23,17%. Hasil prediksi ini diintegrasikan ke dalam sistem berbasis website untuk memudahkan visualisasi data bagi pengguna. Dapat disimpulkan bahwa algoritma regresi linier berganda dapat digunakan untuk memprediksi

jumlah penduduk miskin dengan akurasi tinggi guna membantu perencanaan program pengentasan kemiskinan.

Kata Kunci: Kemiskinan, Data Mining, Regresi Linier Berganda, KDD, Sumatera Selatan

PENDAHULUAN

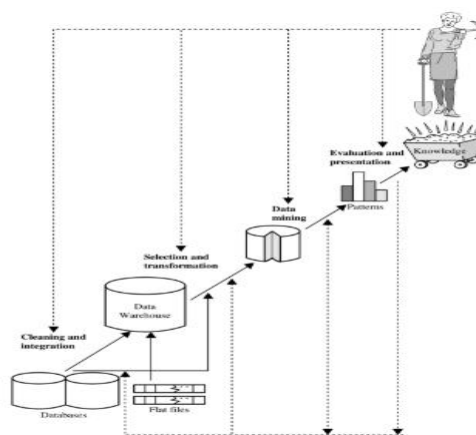
Kemiskinan adalah masalah serius dalam pembangunan Indonesia, meskipun pemerintah telah mengenalkan program-program untuk mengatasinya. Kemiskinan bukanlah hasil kehendak individu, tetapi akibat dari keadaan yang tidak dapat dihindari dengan sumber daya yang dimiliki. Ini juga merupakan masalah multidimensional yang terkait aspek ekonomi, sosial, politik, dan kesempatan merata dalam masyarakat. Dampaknya dirasakan dalam Pendidikan, Kesehatan, Kriminalitas, akses Pendidikan, Pengangguran, dan konflik sosial, pemahaman ini penting dalam merumuskan kebijakan efektif untuk mengatasi kemiskinan dan mendorong pemerataan pembangunan berkelanjutan (Maharani, Ningrum, Fatmawati, & Fadilla, 2024).

Dengan adanya data prediksi jumlah penduduk miskin di setiap kabupaten atau kota di provinsi Sumatera Selatan yang telah dikumpulkan dan diproses untuk menghasilkan pengetahuan baru. *Data mining* adalah proses untuk menggali informasi bermanfaat dari sekumpulan data besar melalui tahapan mengumpulkan, membersihkan, memproses, menganalisis dan interpretasi data (Aggrawal, 2015). Oleh karenanya, penerapan data mining sebagai alat untuk memprediksi jumlah penduduk miskin di masa depan berperan penting dalam menganalisis berbagai faktor.

Penelitian ini bertujuan untuk Menggunakan metode regresi linier berganda untuk memprediksi jumlah penduduk miskin di provinsi Sumatera Selatan. Metode regresi adalah teknik fundamental yang digunakan untuk membangun model guna memperkirakan suatu nilai berdasarkan informasi yang ada (Sinaga & Sipayung, 2024) (Han, Kamber, & Pei, 2022). Perhitungan nilai RMSE (Kesalahan Rata Rata Kuadrat), MAE (Kesalahan Rata Rata Absolute), dan MARE (Kesalahan Rata Rata Relatif Absolut) adalah hasil dari proses prediksi. Tujuan memuat pertanyaan artikel yang harus di jelaskan pada pembahasan dan di jawab pada kesimpulan.

METODE

Penelitian ini Menggunakan Metode *Knowledge Discovery in Databases* (KDD) dan Algoritma Regresi Linier Berganda. Proses ini menyediakan alur kerja sistematis dan iterative yang sangat sesuai dengan analisis prediktif. Proses KDD dapat dilihat pada gambar 1 berikut:



sumber : (Han, Kamber, & Pei, 2022)

Gambar 1. Tahapan Proses KDD

a) *Data Cleaning & Integration*

Pada tahap ini merupakan pondasi awal data dari badan pusat statistic (BPS) dengan bertujuan untuk melakukan proses pembersihan data untuk mematikan setiap data yang digunakan bersih dari inkonsistensi dan kesalahan input atau nilai yang hilang (*Missing Value*) dan proses penggabungan berbagai dataset menjadi sebuah dataset.

b) *Data Selection & Transformation*

Pada tahap ini berfokus pada pemilihan *variable* yang digunakan sebagai bahan penelitian dan data akan dilakukan *transformasi* ke dalam format yang sesuai untuk dianalisis lebih lanjut.

c) *Data Mining*

Tahap Penerapan Algoritma *Data Mining* yang digunakan yaitu Algoritma Regresi Linier Berganda. Bertujuan untuk mengetahui hubungan signifikan antara berbagai faktor sosial – ekonomi dan prediksi dari angka jumlah penduduk miskin.

d) *Pattern Evaluation*

Tahap ini untuk memastikan bahwa model prediksi yang digunakan tidak hanya akurat secara statistik, tetapi juga relevan. Metrik yang digunakan dalam evaluasi adalah RSME, MAE dan MAPE

e) *Knowlegde Presentation*

Tahap ini merupakan tahap akhir sekaligus penghubung antara analis data dan pelanggan, pada tahap ini dilakukan pengembangan aplikasi untuk menyajikan data secara visual yang interaktif

HASIL DAN PEMBAHASAN

Data Cleaning & Integration

Pada tahap ini dataset yang digunakan dalam penelitian ini bersumber dari Badan Pusat Statistik (BPS) Provinsi Sumatera Selatan pada periode tahun 2014 – 2024. Data yang dibutuhkan adalah Jumlah Penduduk Miskin, Garis Kemiskinan, Pengeluaran Perkapita, Indeks Pembangunan Manusia (IPM), Rata Rata Lama sekolah. Akses Air Minum Bersih, Tingkat Pengangguran Terbuka, Tingkat Partisipasi Angkatan Kerja. Selanjutnya dilakukan pemeriksaan data dengan pemeriksaan *Missing Value*, *Duplicate*, *inkonsistensi data*.

Data Selection & Transformation

Pada tahap ini dilakukan untuk menentukan *variable* yang relevan digunakan dalam proses pemodelan regresi linier berganda. Dari hasil integrasi yang dilakukan dataset terdiri dari delapan *variable* yang semuanya berhubungan dengan kemiskinan berdasarkan literatur dan publikasi resmi BPS. Seperti yang dapat dilihat pada table 1.

Tabel 1. Rincian Variable Data

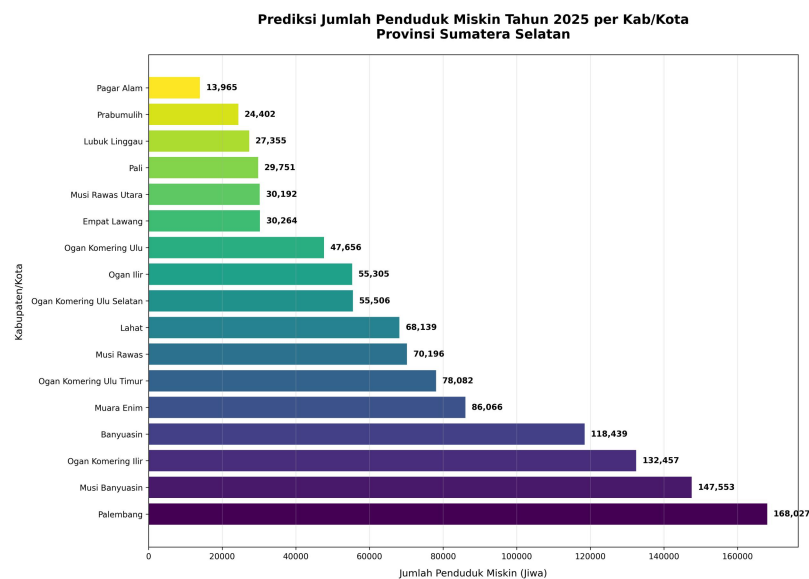
No	Keterangan	Satuan	Tipe Data	Variable
1	Jumlah Penduduk Miskin	Jiwa	Numerik	Y
2	Garis Kemiskinan	Rp/Kapita/Bulan	Numerik	X1
3	Pengeluaran Perkapita	Rp/Kapita/tahun	Numerik	X2
4	Indeks Pembangunan Manusia(IPM)	Indeks (0-100)	Numerik	X3
5	Rata rata Lama sekolah	Tahun	Numerik	X4
6	Akses Air Minum Layak	Presentase(%)	Numerik	X6
7	Tingkat Pengangguran Terbuka	Presentase	Numerik	X7
8	Tingkat Partisipasi Angkatan Kerja	Presentase	Numerik	X8

Sumber : Hasil Peneliti

Kemudian dilakukan transformasi data agar dapat diolah secara optimal pada algoritma regresi linier berganda pada tahap ini berfokus pada pembersihan nilai nol, konversi nilai nol menjadi median imputasi dan transformasi logaritmik.

Data Mining

Pada tahap ini dilakukan pemodelan dengan menggunakan google collab untuk memprediksi jumlah penduduk miskin dengan algoritma regresi linier berganda dalam penelitian ini menggunakan operator split data untuk pelatihan menggunakan 80% dan pelatihan menggunakan 20% dengan menggunakan *function LinearReggresion()* pada google collab maka didapatkan hasil estimasi jumlah penduduk miskin pada provinsi sumatera selatan seperti yang dapat dilihat pada gambar 2 berikut;



Sumber : Hasil Analisis Peneliti Menggunakan *Goggle Collab*
Gambar 2. Hasil Prediksi Jumlah Penduduk Miskin Tahun 2025 Provinsi Sumatera Selatan

Sehingga didapatkan hasil persamaan model regresi linier berganda sebagai berikut ;

$$\log(Y + 1) = -17.5968 - 0.000001 * X_2 - 0.000001 * X_1 + 0.606781 * X_3 - 0.436150 * X_4 + 0.003032 * X_5 + 0.005884 * X_6 + 0.003754 * X_7$$

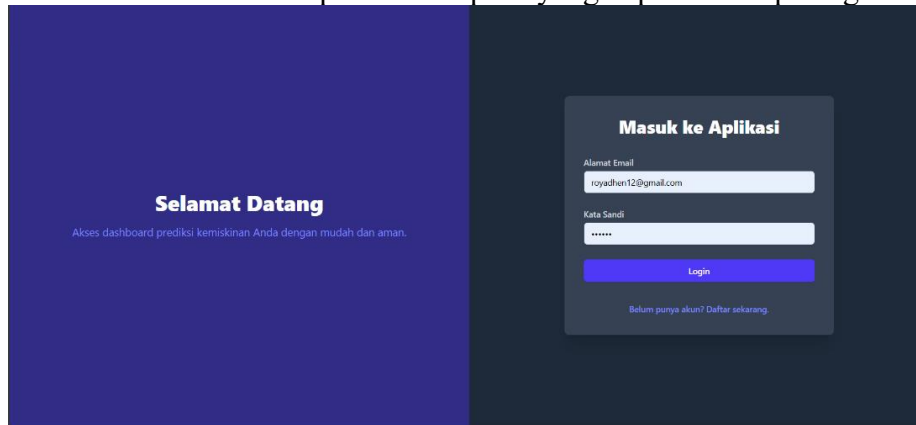
Pattern Evaluation

Hasil Evaluasi pada tahapan regresi linier berganda menunjukkan bahwa model prediksi jumlah penduduk miskin di provinsi sumatera selatan ataupun kabupaten kota yang ada di provinsi sumatera selatan dapat diukur melalui metrik evaluasi. Dalam hal ini, RMSE (*Root Mean Squared Error*) sebesar 15,759 sedangkan MAE (*Mean Absolute Error*) sebesar 12,293 dan nilai MAPE (*Mean Absolute Percentage Error*) sebesar 23.17%. Hasil ini menunjukkan bahwa model regresi linier berganda mampu memberikan prediksi yang sangat baik, terutama nilai RMSE, MAE, MAPE mampu mendekati 0 atau nilai actual. Dapat disimpulkan bahwa model regresi linier berganda memiliki kinerja yang baik dalam prediksi jumlah penduduk miskin di provinsi sumatera selatan.

Knowlegde Presentation

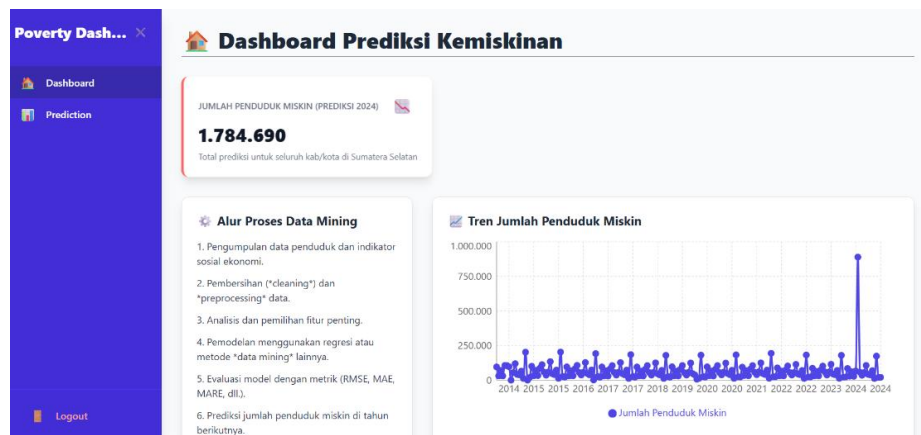
Pada tahap ini akan dilakukan pengembangan sistem untuk menampilkan hasil prediksi agar mudah dibaca oleh pelanggan, sistem berbentuk website dan terdiri dari halaman login dimana pengguna perlu menginputkan email dan password agar dapat

mengakses halaman dashboard, pada halaman dashboard pengguna dapat melihat jumlah penduduk dan alur proses data mining serta tren jumlah penduduk miskin, pada halaman prediksi pengguna dapat memilih tahun prediksi dan menginputkan file yang akan di prediksi, dan pada halaman hasil prediksi akan menampilkan seluruh data prediksi yang terdiri dari tren dan metrik evaluasi serta hasil prediksi. Seperti yang dapat dilihat pada gambar 3 – 6;



Sumber : Hasil Peneliti

Gambar 3. Halaman login

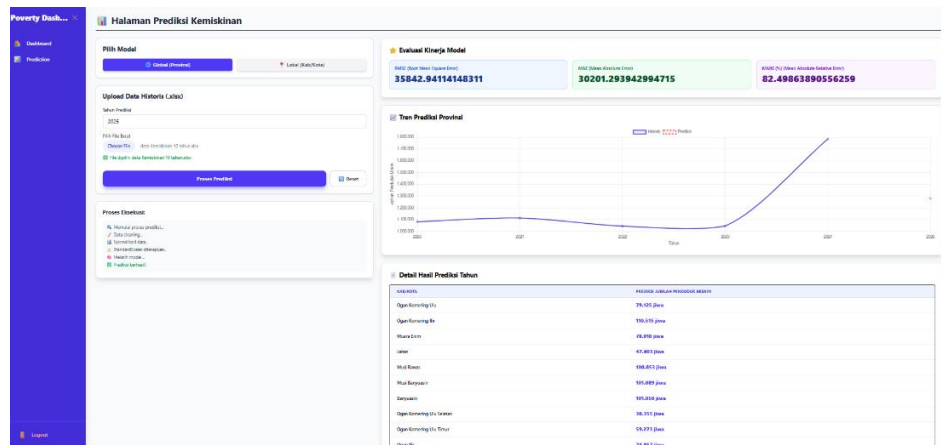


Sumber : Hasil Peneliti

Gambar 4. Halaman Dashboard

Sumber : Hasil Peneliti

Gambar 5. Halaman Prediksi



Sumber : Hasil Peneliti
Gambar 6. Halaman Hasil Prediksi

KESIMPULAN

Penelitian ini berhasil mengembangkan model prediksi jumlah penduduk miskin di Provinsi Sumatera Selatan tahun 2025 menggunakan algoritma regresi linier berganda dengan proses data mining lengkap mulai dari data cleaning, integration, selection, transformation, hingga knowledge presentation. Hasil Prediksi yang divalidasi dengan Metrik Evaluasi RMSE 15,759, MAE 12,293 dan MAPE 23,17% menandakan kinerja model dengan selisih dari nilai aktual relative kecil dan dapat disimpulkan bahwa model algoritma regresi linier berganda dapat digunakan untuk memprediksi jumlah penduduk miskin di provinsi sumatera selatan dengan akurasi tinggi.

Berdasarkan hasil penelitian ini, penulis menyadari adanya beberapa kekurangan sehingga penulis menyarankan untuk hal hal berikut; penelitian ini dapat dikembangkan lagi dengan model algoritma random forest , exponential smoothing atau algoritma lain yang digunakan untuk model prediksi untuk mengurangi RMSE/MAE lebih lanjut, terutama pada data time series kemiskinan yang fluktuatif dengan data periode lebih lama dan penambahan data dengan provinsi lain di Indonesia serta meningkatkan pengembangan website dengan terintegrasi dengan API BPS Real-time dan antarmuka lebih aplikatif bagi instansi pemerintah daerah dalam perencanaan program pengentasan kemiskinan.

REFERENSI

- Aggrawal, C. C. (2015). *Data Mining*. New York: Springer International Publishing. doi:10.1007/978-3-319-14142-8
- Han, J., Kamber, M., & Pei, J. (2022). *Data Mining Concepts and Techniques Third Edition*. Waltham Massachusetts: Morgan Kaufmann Publisher.
- Maharani, C., Ningrum, D. A., Fatmawati, A. E., & Fadilla, A. (2024). Dampak Kemiskinan Terhadap Kualitas Pendidikan Anak di Indonesia: Rekomendasi Kebijakan yang Efektif. *Journal of Macroeconomics and Social Development*, 1-10.
- Sinaga, L. M., & Sipayung, S. P. (2024). Analisis dan Prediksi Persentase Angka Kemiskinan di Indonesia Menggunakan Metode Regresi Berganda. *KAKIFIKOM (Kumpulan Artikel Karya Ilmiah Fakultas Ilmu Komputer)*, 116-125.