

MODEL DATA MINING DALAM MENGANALISIS FAKTOR-FAKTOR YANG MEMPENGARUHI KEMISKINAN DI SUMATERA SELATAN

Hardianto, Yesi Novaria Kunang, Edi Surya Negara, Tata Sutabri

Teknik Informatika-Pasca Sarjana, Universitas Binadarma Palembang

Jl. Jenderal Ahmad Yani No.3, 9/10 Ulu, Kecamatan Seberang Ulu I, Kota Palembang, Sumatera Selatan

rhdpaalembang@gmail.com

ABSTRAK

Pengentasan kemiskinan menjadi isu strategis dalam Rencana Pembangunan Nasional Indonesia, dengan target pengurangan kemiskinan ekstrim menjadi nol persen pada tahun 2024. Penelitian ini berfokus pada analisis kemiskinan di Provinsi Sumatera Selatan, yang meskipun memiliki potensi ekonomi tinggi, masih menghadapi tantangan besar dengan tingkat kemiskinan yang mencapai 11,78 persen pada tahun 2023. Melalui pendekatan multidimensi dan metodologi data mining, penelitian ini bertujuan untuk mengidentifikasi faktor-faktor yang mempengaruhi kemiskinan serta memberikan alternatif visualisasi data untuk mempermudah interpretasi hasil analisis. Metode yang digunakan dalam penelitian ini adalah analisis data mining dengan algoritma *machine learning*, termasuk regresi linier, *decision tree*, dan *random forest*. Data yang dianalisis berasal dari Survei Sosial Ekonomi Nasional (Susenas) Maret 2023. Hasil penelitian menunjukkan bahwa variabel aksesibilitas, pengeluaran konsumsi makanan, dan luas lantai per kapita merupakan faktor dominan yang mempengaruhi tingkat kemiskinan di wilayah tersebut. Hal ini memberikan wawasan penting bagi pengambil kebijakan dalam merancang intervensi yang lebih efektif untuk pengentasan kemiskinan. Pemanfaatan visualisasi data dalam penelitian ini tidak hanya mengidentifikasi faktor-faktor yang signifikan, tetapi juga mendukung pemahaman yang lebih baik tentang situasi kemiskinan di Sumatera Selatan.

Kata kunci : Data Mining, Kemiskinan, Machine Learning, Visualisasi Data

1. PENDAHULUAN

Pengentasan kemiskinan merupakan salah satu prioritas utama dalam Rencana Pembangunan Nasional Indonesia, baik jangka menengah maupun jangka panjang. Pemerintah Indonesia terus menggulirkan berbagai program perlindungan sosial untuk mengurangi angka kemiskinan, dengan target ambisius yang ditetapkan oleh Presiden Joko Widodo melalui Instruksi Presiden Nomor 4 Tahun 2022 untuk mengeliminasi kemiskinan ekstrim hingga 0% pada tahun 2024 [1].

Program-program pengentasan kemiskinan yang telah diimplementasikan mencakup berbagai aspek, mulai dari bantuan konsumsi hingga jaminan kesehatan dan kredit usaha. Beberapa program unggulan meliputi Bantuan Pangan Non Tunai, Program Keluarga Harapan (PKH), Jaminan Kesehatan Nasional (JKN), kartu sembako, subsidi energi, bantuan pertanian, Kredit Usaha Rakyat (KUR), serta alokasi Dana Desa. Alokasi anggaran untuk bantuan sosial ini menunjukkan tren peningkatan dalam lima tahun terakhir, dengan pengecualian pada tahun 2021 akibat penurunan penerimaan negara sebagai dampak pandemi COVID-19 [2].

Seiring dengan peningkatan nilai bantuan, persentase penduduk miskin di Indonesia mengalami penurunan gradual. Data Badan Pusat Statistik (2023) menunjukkan bahwa tingkat kemiskinan nasional turun dari 9,82% pada semester I tahun 2018 menjadi 9,36% pada periode yang sama di tahun 2023. Meskipun demikian, laju penurunan ini masih dianggap kurang signifikan mengingat target

pemerintah untuk mencapai tingkat kemiskinan 7% pada tahun 2024 [3].

Provinsi Sumatera Selatan menghadapi tantangan yang lebih besar dalam upaya pengentasan kemiskinan. Berbeda dengan tren nasional, angka kemiskinan di provinsi ini masih berada pada level dua digit. Badan Pusat Statistik Provinsi Sumatera Selatan (2023) melaporkan bahwa tingkat kemiskinan di provinsi ini hanya mengalami sedikit penurunan dari 12,78% pada Maret 2018 menjadi 11,78% pada tahun 2023 [4].

Lebih lanjut, Sumatera Selatan tercatat sebagai provinsi dengan tingkat kemiskinan ekstrim tertinggi kedua di Pulau Sumatera. Fenomena ini menimbulkan paradoks mengingat posisi Sumatera Selatan sebagai salah satu dari sepuluh daerah terkaya di Indonesia [2].

Kompleksitas fenomena kemiskinan di Sumatera Selatan menggarisbawahi pentingnya pendekatan multidimensi dalam menganalisis faktor-faktor penentu kemiskinan. Studi terdahulu oleh Ngunyi (2015) menegaskan bahwa kombinasi faktor sosial ekonomi berperan signifikan dalam menentukan status kemiskinan individu [5].

Namun, analisis konvensional yang umumnya didasarkan pada teori kemiskinan atau teori ekonomi yang sudah mapan mungkin tidak cukup untuk menangkap kompleksitas fenomena ini.

Perkembangan teknologi informasi membuka peluang baru untuk melakukan analisis kemiskinan melalui pendekatan data mining. Data mining, yang melibatkan proses ekstraksi informasi dan penemuan pola dari data empiris, menawarkan perspektif baru dalam menganalisis indikator sosial ekonomi,

termasuk kemiskinan. Metode ini memungkinkan eksplorasi sumber data yang lebih luas, melampaui batasan data resmi yang selama ini menjadi tumpuan utama [6].

Beberapa studi terdahulu telah mendemonstrasikan efektivitas pendekatan data mining dalam analisis kemiskinan. Gunawan mengaplikasikan metode *Cross-Industry Standard Process for Data Mining* (CRISP-DM) untuk mengelompokkan kabupaten/kota di Jawa Timur berdasarkan indikator sosial ekonomi, menunjukkan kesesuaian metode ini dalam menganalisis data ekonomi makro [7].

Kaunang menggunakan algoritma J-48 untuk menghitung tingkat kemiskinan periode 1999-2018, mencapai akurasi hingga 88% [8].

Sementara itu, Sari menerapkan algoritma K-Means untuk menganalisis faktor-faktor penentu kemiskinan di Provinsi Banten, mengidentifikasi rata-rata lama sekolah dan pengeluaran per kapita sebagai variabel kunci [9].

Pada skala nasional, Sudibyo menggunakan analisis K-Means untuk mengestimasi dan mengidentifikasi determinan kemiskinan di Indonesia [10].

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk mengaplikasikan pendekatan data mining dalam menganalisis faktor-faktor yang mempengaruhi kemiskinan di Provinsi Sumatera Selatan. Selain mengidentifikasi determinan kemiskinan secara multidimensi, studi ini juga berupaya mengestimasi tingkat kemiskinan pada level provinsi menggunakan teknik data mining. Penggunaan metode ini diharapkan dapat memberikan perspektif baru dan wawasan yang lebih intuitif bagi pemangku kebijakan dalam upaya pengentasan kemiskinan di Sumatera Selatan.

2. TINJAUAN PUSTAKA

Pengentasan kemiskinan merupakan tema utama dalam agenda pembangunan global, termasuk di Indonesia. Dalam konteks *Sustainable Development Goals* (SDGs), tujuan untuk mengakhiri kemiskinan menjadi prioritas utama. Untuk mencapai tujuan ini, dibutuhkan metode pengukuran kemiskinan yang akurat dan cepat agar kebijakan yang diambil tepat sasaran. Pengukuran kemiskinan yang akurat diperlukan untuk mengidentifikasi populasi miskin, memantau kebijakan, dan mengevaluasi efektivitas program pemerintah [11].

Kemiskinan dapat dipahami melalui pendekatan ekonomi, sosial, dan multidisipliner, dengan mempertimbangkan penyebabnya seperti faktor alamiah, kultural, dan struktural. Konsep pengukuran kemiskinan meliputi kemiskinan relatif, absolut, dan multidimensi. Di Indonesia, Badan Pusat Statistik (BPS) menggunakan pendekatan kebutuhan dasar yang berfokus pada pengeluaran individu sebagai indikator utama dalam mengukur kemiskinan. Terdapat beberapa indikator lain yang digunakan

untuk mengukur kemiskinan oleh Bappenas [12], antara lain: keterbatasan pangan, keterbatasan akses kesehatan, keterbatasan akses pendidikan, keterbatasan akses pekerjaan, keterbatasan akses terhadap layanan perumahan dan sanitasi, keterbatasan terhadap air bersih, keterbatasan akses terhadap tanah, keterbatasan akses terhadap sumber daya alam, tidak adanya jaminan rasa aman, keterbatasan akses partisipasi, dan besarnya beban rumah tangga

Penelitian ini bertujuan untuk menganalisis faktor-faktor yang mempengaruhi kemiskinan di Provinsi Sumatera Selatan dengan pendekatan multidimensi melalui penggunaan model data mining dengan visualisasi data. Metodologi yang digunakan dalam penelitian ini mengikuti langkah-langkah yang terstruktur, dimulai dengan identifikasi masalah yang merujuk pada perumusan masalah sebelumnya. Pendekatan yang diterapkan adalah visualisasi data melalui teknik data mining, yang akan membantu dalam memahami kondisi sosial-ekonomi yang berkaitan dengan pendidikan, kesehatan, pekerjaan, perumahan, dan pola pengeluaran rumah tangga di wilayah tersebut. Kerangka berpikir penelitian ini disusun dalam diagram alir yang jelas, mencakup langkah-langkah dari identifikasi masalah hingga kesimpulan dan rekomendasi.

Sumber data utama yang digunakan dalam penelitian adalah data dari Survei Sosial Ekonomi Nasional (Susenas) yang dilakukan oleh Badan Pusat Statistik (BPS). Susenas merupakan survei yang mengumpulkan informasi mengenai kependudukan, kesehatan, pendidikan, keluarga berencana, perumahan, serta konsumsi dan pengeluaran. Penelitian ini memanfaatkan data Susenas Maret 2023, yang melibatkan 11.070 rumah tangga di 17 Kabupaten/Kota di Provinsi Sumatera Selatan. Data yang dikumpulkan terdiri dari data kualitatif dan kuantitatif, yang memungkinkan analisis yang lebih komprehensif mengenai faktor-faktor yang mempengaruhi kemiskinan.

Pengolahan data dilakukan menggunakan perangkat lunak seperti Python, Visual Studio Code, dan Microsoft Excel. Data yang digunakan adalah data mentah dari Susenas yang akan diolah mengikuti langkah-langkah dalam proses *Knowledge Discovery in Databases* (KDD). Proses ini mencakup pemilihan data, pembersihan, normalisasi, dan transformasi data agar sesuai dengan algoritma yang akan digunakan. Algoritma yang dipilih dalam penelitian ini termasuk regresi linier, *decision tree*, dan *random forest*, yang akan digunakan untuk memprediksi tingkat kemiskinan berdasarkan variabel-variabel independen.

Setelah model dibangun, evaluasi dilakukan untuk memahami kontribusi masing-masing variabel terhadap kemiskinan. Hasil analisis akan divisualisasikan menggunakan alat seperti Matplotlib dan Seaborn, yang membantu dalam interpretasi data dan pemahaman distribusi serta hubungan antar variabel. Melalui tahapan ini, penelitian diharapkan dapat memberikan wawasan yang mendalam

mengenai faktor-faktor yang memengaruhi kemiskinan di Provinsi Sumatera Selatan, serta rekomendasi kebijakan yang dapat diterapkan untuk mengatasi masalah kemiskinan secara lebih efektif.

Dalam konteks penelitian terdahulu, Di Yang melakukan analisis mengenai perencanaan jangka panjang untuk pengentasan kemiskinan dengan judul *“The contribution of data-driven poverty alleviation funds in achieving mid-21st-Century multidimensional poverty alleviation planning”* [13].

Penelitian ini menggunakan metode machine learning dan analisis skenario untuk menunjukkan bagaimana simulasi Monte Carlo dapat membantu mengidentifikasi kebutuhan strategi pengentasan kemiskinan yang terintegrasi di Cina. Hasil dari penelitian ini menggarisbawahi pentingnya kebijakan pembangunan multidimensi untuk mencapai *Sustainable Development Goals* (SDGs) secara global, serta menunjukkan bahwa penggunaan indikator pengembangan multidimensi dapat mengurangi disparitas regional.

Selanjutnya, penelitian oleh Aprilliana Nur Aurelliaa dan Putriaji Hendikawati mengidentifikasi dua faktor utama yang mempengaruhi kemiskinan di Jawa Timur, yaitu kesejahteraan dan status pekerjaan [14].

Selain itu, studi oleh Stevencua dan Johan Setiawan (2017) mengembangkan visualisasi data untuk menggambarkan distribusi kemiskinan dan hubungannya dengan tingkat pendidikan di Indonesia [15].

Menggunakan data dari BPS periode 2013 hingga 2015, penelitian ini menghasilkan tiga dashboard interaktif yang menampilkan peta kemiskinan, peta pendidikan, dan peringkat daerah berdasarkan tingkat kemiskinan. Meskipun tingkat pendidikan berpengaruh terhadap kemiskinan, faktor lain seperti jumlah penduduk, infrastruktur, dan akses terhadap barang dan jasa juga perlu diperhatikan dalam analisis penyebab kemiskinan di suatu daerah.

2.1. Data Selection

Data utama yang digunakan dalam penelitian ini adalah data Survei Sosial Ekonomi Nasional (Susenas) Provinsi Sumatera Selatan periode Maret 2023. Selain data utama terdapat beberapa data penunjang lainnya yang digunakan dalam penelitian ini antara lain data Indeks Kesulitan Geografis yang dihasilkan dari Potensi Desa (Podes) Provinsi Sumatera Selatan tahun 2021 dan data inflasi Kota Palembang tahun 2023 yang dipublikasikan oleh BPS Provinsi Sumatera Selatan.

Output yang dihasilkan pada tahapan ini adalah berupa tabel data rumah tangga yang terdiri atas atribut-atribut antara lain: Status Perkawinan, Jenis Kelamin, Umur, Ijazah/STTB tertinggi, KRT Bekerja atau tidak, Kegiatan Terbanyak dalam seminggu, Lapangan Usaha Pekerjaan Utama, Status dalam Pekerjaan Utama, Jumlah Keluarga dalam 1 bangunan, Status kepemilikan bangunan, Luas Lantai bangunan tempat tinggal, Atap terluas, Dinding terluas, Bahan

Lantai Terluas, Fasilitas tempat buang air besar, sumber air minum utama, sumber air utama untuk mandi/cuci, sumber utama penerangan, daya terpasang, jenis bahan bakar utama untuk memasak, jumlah ART, luas lantai per kapita, Jumlah ART yang bekerja, Pengeluaran konsumsi Makanan, pengeluaran konsumsi Non Makanan, Nilai pengeluaran per kapita, Status ekonomi, indeks ke fasilitas pendidikan dan kesehatan, indeks kondisi infrastruktur, indeks aksesibilitas/transportasi, dan indeks kesulitan geografis (IKG).

2.2. Preprocessing

Beberapa variabel dalam Data Susenas Maret 2023 dilakukan preprocessing antara lain pengkodean pada variabel status bekerja dan sumber utama penerangan diberi pengkodean ‘1’ apabila bekerja dan penerangan dengan listrik PLN dan sebaliknya kode ‘0’. Selain itu juga dilakukan pemeriksaan untuk memastikan tidak adanya missing value dan duplikat yang menunjukkan tidak adanya *missing value*.

2.3. Transformation

Pada tahapan ini dilakukan pengelompokan data untuk semua variabel penelitian menjadi numerik dan kategorik. Adapun pengelompokan variabel tersebut adalah pada tabel berikut:

Tabel 1. Pengelompokan Data

Tipe Data	Pengelompokan
numerik	Status pernikahan, jenis kelamin, kualifikasi pendidikan tertinggi, apakah kepala rumah tangga bekerja, kegiatan utama yang dilakukan dalam seminggu, sektor pekerjaan utama, status pekerjaan dalam pekerjaan utama, status kepemilikan bangunan, luas lantai tempat tinggal, bahan atap terluas, bahan dinding terluas, bahan lantai yang terluas, ketersediaan fasilitas toilet, sumber utama air minum, sumber utama air untuk mandi/cuci, sumber penerangan utama, kapasitas listrik terpasang, jenis bahan bakar utama untuk memasak, status ekonomi.
kategori	Usia, jumlah anggota keluarga dalam satu bangunan, jumlah anggota rumah tangga, luas lantai per kapita, jumlah anggota rumah tangga yang bekerja, pengeluaran konsumsi makanan, pengeluaran konsumsi non-makanan, nilai pengeluaran per kapita, indeks untuk fasilitas pendidikan dan kesehatan, indeks kondisi infrastruktur, indeks aksesibilitas/transportasi, dan Indeks Kesulitan Geografis (IKG).

Proses selanjutnya adalah penerapan scaling pada data karena variabel-variabel dalam data Susenas Maret 2023 yang digunakan dalam penelitian ini memiliki keragaman skala yang cukup besar. Proses Scaling menyeragamkan skala ini, sehingga semua variabel memiliki bobot yang setara dalam analisis. Metode scaling yang digunakan antara lain:

- a. *No Scaler' (passthrough)*: Ini berarti tidak ada scaling yang diterapkan. Data tetap dalam bentuk aslinya.
- b. *StandardScaler*: Menstandarisasi nilai variabel dengan mengurangi mean dan membagi dengan standar deviasi. Menghasilkan distribusi dengan mean 0 dan variansi 1.
- c. *MinMaxScaler*: Menskalakan nilai variabel ke range tertentu, biasanya [0, 1]. Berguna ketika Anda ingin mempertahankan zero entries dalam data sparse.
- d. *MaxAbsScaler*: Menskalakan setiap nilai variabel dengan nilai absolutnya yang maksimum. Berguna untuk data yang sudah terpusat di sekitar nol atau data sparse.
- e. *RobustScaler*: Menggunakan statistik yang robust terhadap outlier. Mengurangi median dan menskalakan menggunakan interquartile range.
- f. *PowerTransformer*: Menerapkan transformasi power untuk membuat data lebih menyerupai distribusi normal Gaussian.
- g. *Normalizer*: Menormalkan sampel secara individual ke norm unit (panjang vektor 1).

2.4. Data Modelling

Memilih algoritma data mining yang sesuai dengan tujuan penelitian dan jenis data. Beberapa algoritma yang akan digunakan dalam penelitian ini antara lain:

- a. *Regresi Linier*: Untuk memprediksi tingkat kemiskinan berdasarkan variabel-variabel independen.
- b. *Decision Tree*: Untuk memahami hubungan non-linear antara variabel-variabel dan mengidentifikasi aturan yang terkait dengan kemiskinan.
- c. *Random Forest*: Untuk meningkatkan akurasi prediksi dan mengatasi *overfitting*.

Untuk setiap algoritma, tersebut digunakan percobaan 7 variasi model dengan kombinasi teknik scaling (normalisasi) dan encoding (*One Hot Encoder*) yang berbeda pada fitur-fitur data sehingga didapatkan 21 model prediksi kemiskinan dengan variabel yang telah ditentukan. Tujuannya adalah untuk menemukan konfigurasi terbaik yang dapat meningkatkan performa model

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan analisis data mining untuk mengidentifikasi faktor-faktor yang memengaruhi kemiskinan multidimensi di Provinsi Sumatera Selatan. Data utama yang digunakan bersumber dari Survei Sosial Ekonomi Nasional (Susenas) periode Maret 2023, dengan sampel sebanyak 11.070 rumah tangga yang tersebar di 17 Kabupaten/Kota.

Analisis data dilakukan menggunakan bahasa pemrograman Python dengan mengikuti tahapan

Knowledge Discovery in Databases (KDD). Tahapan analisis meliputi:

- a. *Data Selection*: Mengkombinasikan data KOR (keterangan pokok anggota rumah tangga) dan KP (konsumsi pengeluaran) Susenas, mencakup variabel demografis kepala rumah tangga, karakteristik rumah tangga, kondisi perumahan, akses layanan publik, dan pola konsumsi.
- b. *Preprocessing*: Melakukan pembersihan data, normalisasi, dan transformasi data untuk memastikan kualitas analisis. Proses ini mencakup penanganan missing values dan standarisasi format data.
- c. *Feature Selection*: Memilih variabel yang paling relevan berdasarkan analisis korelasi untuk menghindari *multicollinearity*.
- d. *Data Modelling*: Menerapkan tiga algoritma utama:
 - Regresi Linier untuk prediksi tingkat kemiskinan
 - Decision Tree untuk mengidentifikasi pola non-linear
 - Random Forest untuk meningkatkan akurasi prediksi
- 5) *Evaluasi dan Interpretasi*: Menggunakan visualisasi data untuk menginterpretasikan hasil analisis dan memahami kontribusi setiap variabel terhadap kemiskinan.

Ruang lingkup penelitian mencakup aspek sosial-ekonomi meliputi pendidikan, kesehatan, pekerjaan, perumahan, dan pola pengeluaran rumah tangga. Analisis dilakukan menggunakan perangkat lunak Python dengan library pendukung seperti pandas, Matplotlib, dan Seaborn untuk pengolahan dan visualisasi data.

4. HASIL DAN PEMBAHASAN

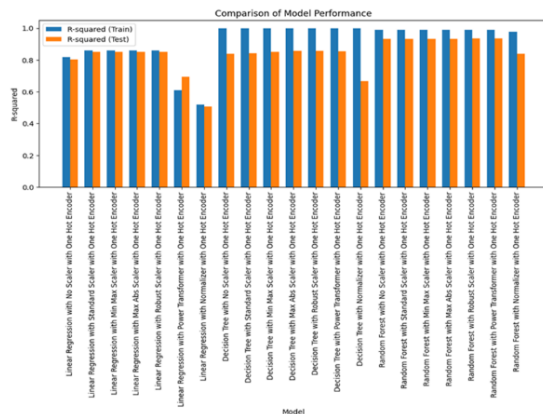
Hasil penerapan beberapa algoritma dalam machine learning tersebut menghasilkan performa dari masing-masing model. Adapun rincian performa masing-masing model adalah sebagai berikut:

- a. *Regresi Linier*: Pada penerapan algoritma regresi linier menghasilkan performa terbaik dengan menggunakan Standard Scaler untuk normalisasi fitur dan One Hot Encoder untuk meng-enkode fitur kategorikal. Pada model ini mampu menjelaskan sekitar 84,36% variasi dalam data target (variabel kemiskinan).
- b. *Decision tree*: Pada penerapan *decision tree*, Konfigurasi terbaik untuk model *Decision Tree* ditemukan dengan menggunakan *Power Transformer* untuk normalisasi dan *One Hot Encoder* untuk encoding. Model ini mencapai *R-squared* sempurna (1,000000) pada data pelatihan. Nilai *Adjusted R-squared* yang lebih rendah (0,696398)

mengindikasikan adanya potensi overfitting yang perlu di evaluasi lebih lanjut.

c. *Random Forest*:

Pada penerapan algoritma *random forest* ditemukan konfigurasi terbaik adalah dengan *Power Transformer* dan *One Hot Encoder* menunjukkan performa terbaik. Nilai *R-squared* 0,986825 dan *Adjusted R-squared* 0,939446 diperoleh dari model ini. Hasil tersebut menunjukkan bahwa model Random Forest dapat menjelaskan sekitar 93,94% variasi dalam data.



Gambar 1. Perbandingan performance dari Model

Hasil perbandingan model-model yang telah di-training menunjukkan bahwa model Random Forest dengan normalisasi menggunakan Power Transformer dan encoding fitur kategorikal menggunakan One Hot Encoder memiliki performa terbaik. Model ini mampu menjelaskan sekitar 93,94% variasi dalam variabel target (tingkat kemiskinan) berdasarkan fitur-fitur yang digunakan.

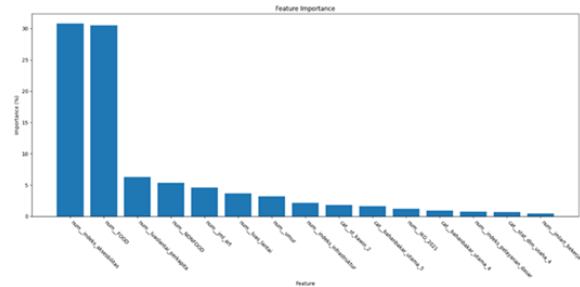
Indikasi yang diperoleh menunjukkan bahwa model Random Forest dapat menangkap dengan baik pola-pola kompleks yang terdapat dalam data. Metode ensemble seperti Random Forest dapat mengatasi masalah overfitting yang terjadi pada model Decision Tree, sehingga mampu memberikan generalisasi yang lebih baik terhadap data baru.

Analisis lebih mendalam dapat dilakukan untuk melihat masing-masing variabel pada model tersebut, analisis tersebut antara lain:

a. *Feature Importance*:

Karakteristik Faktor-Faktor yang Mempengaruhi Kemiskinan Rumah Tangga di Provinsi Sumatera Selatan berdasarkan hasil analisis feature importance dapat memberikan temuan komprehensif yang dapat digunakan sebagai dasar pengembangan kebijakan pengentasan kemiskinan. Variabel Indeks Aksesibilitas merupakan karakteristik yang memberikan kontribusi paling besar bagi model dalam memberikan prediksi yaitu sebesar 30,83%. Pentingnya variabel ini dapat dikaitkan dengan data Indeks Pembangunan Manusia (IPM) Sumsel yang pada tahun 2022 mencapai 71,05, naik dari tahun sebelumnya sebesar 70,72 (BPS Sumsel,

2023). Peningkatan IPM ini mencerminkan peningkatan akses terhadap layanan dasar seperti pendidikan dan layanan kesehatan.



Gambar 2. *feature importance* dalam model

Tabel 2. Kontribusi Variabel dalam analisis feature importance

No.	Feature	Importance (%)
1	num_indeks_aksesibilitas	30.82511
2	num_FOOD	30.529
3	num_luaslantai_perkapita	6.27969
4	num_NONFOOD	5.32868
5	num_jmlart	4.57825
6	num_luaslantai	3.66873
7	num_umur	3.1935
8	num_indeks_infrastruktur	2.13685
9	cat_st_kawin	1.77477
10	cat_bahanbakar_utama	1.61641
11	num_IKG_2021	1.19043
12	cat_bahanbakar_utama	0.89736
13	num_indeks_pelayanan_dasar	0.73406
14	cat_stat_dlm_usaha	0.66445
15	num_jmlart_bekerja	0.44592

Pengeluaran konsumsi makanan (FOOD) merupakan faktor terpenting kedua yang mempunyai pengaruh signifikan terhadap model yaitu sebesar 30,53%. Hal ini erat kaitannya dengan data garis kemiskinan Sumsel yang tercatat sebesar Rp562.837 per orang per bulan pada September 2023. Angka tersebut mencerminkan batas minimal pengeluaran untuk memenuhi kebutuhan dasar, termasuk pangan. Tingginya proporsi pengeluaran pangan pada analisis feature importance menunjukkan masih banyak rumah tangga yang kesulitan memenuhi kebutuhan pangannya. Proporsi pengeluaran pangan yang tinggi sering dijadikan indikator kesejahteraan rumah tangga, sebagaimana diungkapkan dalam Hukum Engel. Penelitian Anwar dalam studinya menggarisbawahi peran sentral ketahanan pangan dalam menentukan status ekonomi rumah tangga di Indonesia [15].

Variabel luas lantai per kapita merupakan indikator ketiga terpenting yang memberikan kontribusi terbesar terhadap prediksi model yaitu sebesar 6,28%. Variabel ini berkaitan erat dengan konsep kepadatan hunian yang dikemukakan oleh UN-Habitat tahun 2015 [16], yang menetapkan standar minimal luas lantai per orang untuk hunian yang layak sebesar 9 meter persegi per orang. Wijayanto dalam penelitiannya menegaskan signifikansi luas hunian

sebagai determinan kemiskinan di wilayah perkotaan Indonesia [17].

Data-data tersebut perlu diteliti lebih lanjut untuk mendapatkan gambaran yang lebih akurat mengenai situasi perumahan di Sumatera Selatan. Variabel selanjutnya seperti pengeluaran konsumsi non-pangan (5,33%), jumlah anggota rumah tangga (4,58%), dan total luas lantai (3,67%) juga memberikan kontribusi signifikan terhadap akuratnya prediksi model.

b. *Partial Dependence Plots (PDP)*

Analisis lebih dalam bagaimana masing-masing variabel berkontribusi terhadap prediksi kemiskinan, kita akan menginterpretasikannya dengan serangkaian grafik Partial Dependence Plot (PDP). Adapun masing-masing interpretasi dari masing-masing gambar diagram PDP tersebut adalah sebagai berikut:

1. Indeks Aksesibilitas (indeks_aksesibilitas):

Grafik menunjukkan penurunan tajam probabilitas kemiskinan saat indeks aksesibilitas meningkat. Ini mengindikasikan bahwa peningkatan akses ke layanan dasar pendidikan dan kesehatan sangat berpengaruh dalam mengurangi kemiskinan. Di Sumatera Selatan, peningkatan akses ini bisa menjadi kunci pengentasan kemiskinan.

2. Pengeluaran konsumsi makanan (FOOD):

Grafik menunjukkan penurunan probabilitas kemiskinan seiring peningkatan pengeluaran makanan. Namun, penurunan melambat setelah titik tertentu, menunjukkan ada batas di mana peningkatan pengeluaran makanan tidak lagi signifikan mengurangi kemiskinan. Ini mungkin mencerminkan kebutuhan untuk fokus pada aspek lain selain makanan setelah kebutuhan dasar terpenuhi.

3. Luas lantai perkapita (luaslantai_perkapita):

Terlihat penurunan probabilitas kemiskinan saat luas lantai per kapita meningkat. Ini menunjukkan bahwa kondisi tempat tinggal yang lebih baik berkorelasi dengan pengurangan kemiskinan. Program perumahan yang memadai di Sumatera Selatan bisa menjadi strategi efektif.

4. Pengeluaran konsumsi non-makanan (NONFOOD):

Grafik menunjukkan penurunan tajam probabilitas kemiskinan saat pengeluaran non-makanan meningkat. Ini mengindikasikan pentingnya kemampuan rumah tangga untuk memenuhi kebutuhan di luar makanan dalam mengurangi kemiskinan.

5. Jumlah anggota rumah tangga (jml_art):

Terlihat peningkatan probabilitas kemiskinan seiring bertambahnya jumlah anggota rumah tangga. Ini menunjukkan bahwa keluarga besar di Sumatera Selatan cenderung lebih rentan terhadap kemiskinan. Program pengendalian populasi dan dukungan keluarga besar mungkin diperlukan.

6. Luas lantai (luas_lantai):

Grafik menunjukkan penurunan probabilitas kemiskinan saat luas lantai meningkat, meskipun efeknya tidak selinier seperti luas lantai per kapita. Ini menegaskan pentingnya kualitas tempat tinggal dalam konteks kemiskinan.

7. Usia kepala rumah tangga (umur):

Terlihat fluktuasi kecil dalam probabilitas kemiskinan berdasarkan usia, dengan sedikit peningkatan di usia tengah. Ini mungkin mencerminkan tantangan ekonomi yang dihadapi oleh kepala rumah tangga pada usia produktif di Sumatera Selatan.

8. Indeks Infrastruktur (indeks_infrastruktur):

Grafik menunjukkan penurunan tajam probabilitas kemiskinan saat indeks infrastruktur meningkat. Ini menekankan pentingnya pembangunan infrastruktur dalam pengentasan kemiskinan di Sumatera Selatan.

9. Status perkawinan kepala rumah tangga (st_kawin):

Terlihat variasi dalam probabilitas kemiskinan berdasarkan status perkawinan. Ini mungkin mencerminkan perbedaan tantangan ekonomi yang dihadapi oleh rumah tangga dengan status perkawinan berbeda.

10. Bahan bakar utama memasak (bahanbakar_utama):

Grafik menunjukkan variasi dalam probabilitas kemiskinan berdasarkan jenis bahan bakar utama. Ini bisa mencerminkan akses ke energi modern dan kaitannya dengan tingkat kemiskinan di Sumatera Selatan.

11. Indeks Kesulitan Geografis (IKG_2021):

Terlihat peningkatan probabilitas kemiskinan seiring meningkatnya IKG. Ini menunjukkan bahwa daerah-daerah dengan kesulitan geografis di Sumatera Selatan cenderung lebih rentan terhadap kemiskinan.

12. Indeks Pelayanan Dasar (indeks_pelayanan_dasar):

Grafik menunjukkan penurunan probabilitas kemiskinan saat indeks pelayanan dasar meningkat. Ini menekankan pentingnya akses ke layanan pendidikan dan kesehatan dalam mengurangi kemiskinan.

13. Status bekerja kepala rumah tangga (stat_dlm_usaha):

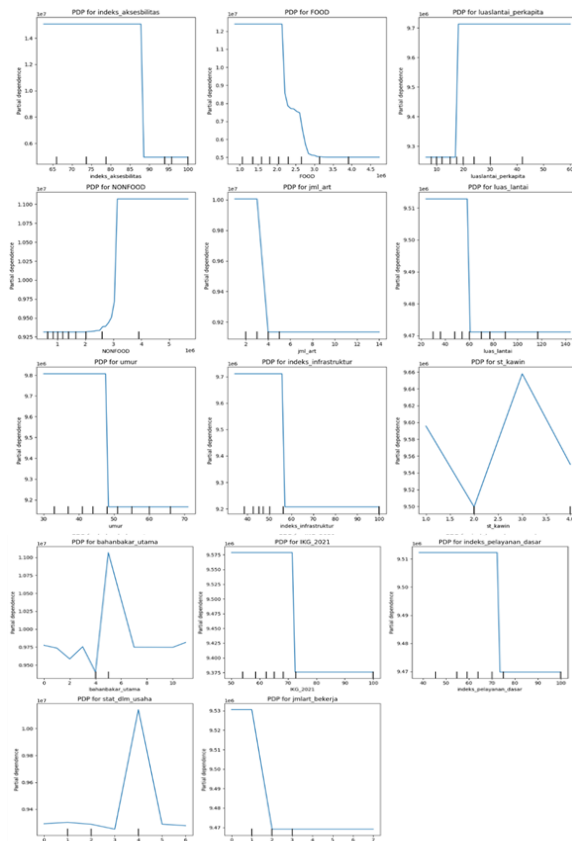
Terlihat perbedaan probabilitas kemiskinan berdasarkan status pekerjaan. Ini mencerminkan pentingnya kesempatan kerja dalam pengentasan kemiskinan di Sumatera Selatan.

14. Jumlah anggota rumah tangga yang bekerja (jmlart_bekerja):

Grafik menunjukkan penurunan probabilitas kemiskinan seiring bertambahnya jumlah anggota rumah tangga yang bekerja. Ini menekankan pentingnya kesempatan kerja bagi

anggota keluarga dalam mengurangi kemiskinan.

Analisis komprehensif ini memberikan wawasan penting tentang sifat multifaset kemiskinan di Sumatera Selatan dan menawarkan alternatif untuk intervensi yang terarah serta perumusan kebijakan.



Gambar 3. Analisis Partial Dependence Plots dalam model

5. KESIMPULAN

Berdasarkan hasil penelitian diatas dapat diperoleh beberapa kesimpulan sebagai berikut: Model Random Forest dengan Power Transformer dan One Hot Encoder memberikan performa terbaik dalam memprediksi kemiskinan di Sumatera Selatan, menjelaskan 93,94% variasi tingkat kemiskinan. Faktor signifikan yang mempengaruhi kemiskinan di Sumatera Selatan: Aksesibilitas ke layanan dasar kesehatan dan pendidikan (30,83%), Pengeluaran konsumsi makanan (30,53%), Luas lantai bangunan tempat tinggal per kapita (6,28%), Temuan ini dapat menjadi dasar kebijakan pengentasan kemiskinan, mengingat tingkat kemiskinan masih 11,82%. Penerapan teknik data mining dalam analisis kemiskinan di Provinsi Sumatera Selatan melalui visualisasi data dengan menggunakan heatmap, diagram feature importance dan partial dependence plots dapat mengidentifikasi pola tersembunyi dan korelasi antar variabel yang tidak terlihat menggunakan metode analisis tradisional.

DAFTAR PUSTAKA

- [1] K. PMK, "Indonesia Berhasil Turunkan Kemiskinan Ekstrem," 2024. [Online]. Available: <https://www.kemenkopmk.go.id/indonesia-berhasil-turunkan-kemiskinan-ekstrem>. [Accessed 20 4 2024]
- [2] K. K. RI, "Laporan Pemerintah tentang Pelaksanaan Anggaran Pendapatan dan Belanja Negara Semester Pertama 2023," Kementerian Keuangan RI, Jakarta, 2023.
- [3] B. P. Statistik, "Profil Kemiskinan di Indonesia Maret 2023," Badan Pusat Statistik (BPS), Jakarta, 2023
- [4] B. R. Statistik, "Profil Kemiskinan di Sumatera Selatan Maret 2023," BPS Provinsi Sumatera Selatan, Palembang, 2023
- [5] P. N. M. R. O. O. V. G. M. Anthony Ngunyi, "Multidimensional Analysis of the Determinant Poverty Indicators in the Lake Victoria Basin (Kenya)," *IOSR Journal Of Mathematics (IOSR-JM)*, pp. 73-87, 2015
- [6] J. A. Talingdan, "Data Mining Using Clustering Algorithm as Tool for Poverty Analysis," *ICSCA '19: Proceedings of the 2019 8th International Conference on Software and Computer Applications*, pp. 56-59, 2019
- [7] Gunawan, "The Opportunity of data mining for macroeconomic data analysis: A case analysis of East Java Province," *Jurnal Ekonomi dan Pembangunan n Vol. 29 No. 2*, pp. 183-198, 2021
- [8] F. J. Kaunang, "Penerapan Algoritma J48 Decision Tree Untuk Analisis Tingkat Kemiskinan di Indonesia," *Cogito Smart Journal | VOL. 4 | NO.2*, pp. 348-357, 2018
- [9] A. S. D. A. L. T. I. J. Yunita Ratna Sari, "PENERAPAN ALGORITMA K-MEANS UNTUK CLUSTERING DATA KEMISKINAN PROVINSI BANTEN MENGGUNAKAN RAPIDMINER," *CESS (Journal of Computer Engineering System and Science) Vol. 5 No. 2 Juli 2020*, pp. 192-198, 2020
- [10] A. I. K. S. S. S. Nugroho Arif Sudibyo, "PENERAPAN DATA MINING PADA JUMLAH PENDUDUK MISKIN DI INDONESIA," *Lebesgue : Jurnal Ilmiah Pendidikan Matematika, Matematika dan Statistika*, pp. 199-207, 2020
- [11] W. Bank, "The World Bank Annual Report 2019 : Ending Poverty, Investing in Opportunity," The World Bank, Washington, 2019
- [12] Harniati, Program-program Sektor Pertanian yang Berorientasi Penanggulangan Kemiskinan, Bogor: Pusat Analisis Sosial Ekonomi dan Kebijakan Pertanian Departemen Pertanian, 2010
- [13] D. Y. & W. L. & J. Y. & B. X. & X. Z. & H. W. & F. Pian, "The contribution of data-driven poverty alleviation funds in achieving mid-21st-Century multidimensional poverty alleviation

- planning," *HUMANITIES AND SOCIAL SCIENCES COMMUNICATIONS*, 2022
- [14] d. P. H. Aprilliana Nur Aurelliaa, "Faktor Dominan yang Mempengaruhi Kemiskinan Kabupaten/Kota di Provinsi Jawa Timur Tahun 2022," *PRISMA, Prosiding Seminar Nasional Matematika* 7, pp. 712-723, 2024
- [15] S. d. J. Setiawan, "Data Visualization of Poverty Level at Provinces in Indonesia from The year 2013-2015," *IJNMT, Vol. V, No. 1*, 2018