JTIF | Jurnal INOVATIF WIRA WACANA Vol. 04 No. 02 September 2025, hlm 181-188 **pISSN** (2961-7316)
eISSN (2962-5998)

PREDIKSI TINGKAT KEMISKINAN MENGGUNAKAN ALGORITMA *K-NEAREST NEIGHBORS* (KNN) BERBASIS *DATA MINING* (Studi Kasus: Kabupaten Sumba Timur)

(*PREDICTION OF POVERTY LEVEL USING K-NEAREST NEIGHBORS (KNN) ALGORITHM BASED ON DATA MINING (Case Study: East Sumba Regency)*

Melan Astriani Mburu Hamu[1], Arini Aha Pekuwali[2], Leonard Marten Doni Ratu[3]

[1,2,3]Program Studi Teknik Informatika, Universitas Kristen Wira Wacana Sumba

E-mail: [1]melanastriani2@gmail.com [2]arini.pekuwali@unkriswina.ac.id [3]leonard.ratu@unkriswina.ac.id.

KEYWORDS:
Poverty Rate Prediction, K-Nearest Neighbors Algorithm, Data Mining, East Sumba Regency

ABSTRACT

Poverty remains a significant social and economic challenge in East Sumba Regency, where more than a quarter of the population lives below the poverty line. The issue is not only the high poverty rate but also the suboptimal use of socio-economic data as a foundation for designing well-targeted policies. To support more effective decision-making, a data-driven analytical approach is needed—one that can process information predictively. This study aims to predict poverty levels using a data mining approach with the K-Nearest Neighbors (KNN) algorithm. The KNN algorithm was chosen because it can make predictions based on data similarity without requiring complex modeling, making it well-suited for the diverse nature of socio-economic data. The research process begins with data collection from relevant institutions, particularly those related to poverty indicators such as poverty rate, per capita expenditure, Human Development Index (HDI), access to sanitation, access to clean water, and labor force participation rate. Data pre-processing steps such as cleaning and normalization are carried out, followed by splitting the data into training and testing sets. The prediction model is built using RapidMiner software to facilitate process visualization and performance evaluation of the algorithm. This study is expected to serve as a reference to support local governments in making data-based decisions to identify areas with a high potential for poverty.

KATA KUNCI:
Prediksi Tingkat Kemiskinan, Algoritma K-Nearest Neighbors, Data Mining, Kabupaten Sumba Timur.

ABSTRAK

Kemiskinan merupakan tantangan sosial dan ekonomi yang masih signifikan di Kabupaten Sumba Timur, dengan tingkat kemiskinan yang tergolong tinggi dan lebih dari seperempat penduduknya hidup di bawah garis kemiskinan. Permasalahan yang dihadapi tidak hanya terkait tingginya angka kemiskinan, tetapi juga kurang optimalnya pemanfaatan data sosial-ekonomi sebagai dasar dalam merancang kebijakan yang tepat sasaran. Dalam upaya merekomendasikan keputusan yang lebih efektif, dibutuhkan pendekatan analitik berbasis data yang mampu mengolah informasi secara prediktif. Penelitian ini bertujuan untuk memprediksi tingkat kemiskinan masyarakat menggunakan pendekatan data mining dengan algoritma K-Nearest Neighbors (KNN). Algoritma KNN dipilih karena

mampu melakukan prediksi berdasarkan kemiripan data tanpa memerlukan model yang kompleks, sehingga cocok digunakan untuk data sosial-ekonomi yang variatif. Proses penelitian dimulai dengan pengumpulan data dari instansi terkait, khususnya yang berhubungan dengan indikator kemiskinan seperti persentase penduduk miskin, pengeluaran per kapita, indeks pembangunan manusia, akses sanitasi, akses air minum, dan tingkat partisipasi angkatan kerja. Selanjutnya dilakukan tahapan pre-processing seperti pembersihan dan normalisasi data, kemudian data dibagi menjadi data latih dan data uji. Model prediksi dibangun menggunakan software RapidMiner untuk mempermudah visualisasi proses serta evaluasi performa algoritma. Penelitian ini diharapkan dapat menjadi acuan yang mendukung pemerintah daerah dalam mengambil keputusan berbasis data untuk mengidentifikasi wilayah-wilayah yang berpotensi tinggi mengalami kemiskinan.

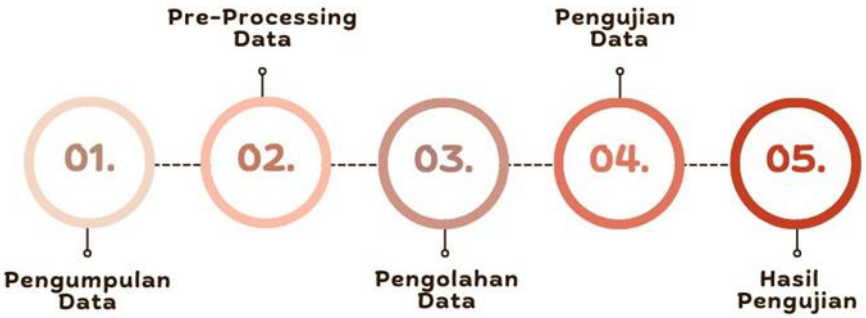
PENDAHULUAN

Kemiskinan merupakan persoalan kompleks yang mencakup berbagai dimensi kehidupan seperti pendidikan, kesehatan, dan kesejahteraan ekonomi [1]. Di Kabupaten Sumba Timur, tingkat kemiskinan masih menjadi isu utama dengan persentase penduduk miskin mencapai 27,04% pada tahun 2024, meskipun menunjukkan tren penurunan sejak 2022 [2]. Setiap tahun, Badan Pusat Statistik (BPS) menyediakan data sosial-ekonomi seperti pengeluaran per kapita, Indeks Pembangunan Manusia (IPM), akses sanitasi, akses air minum, dan tingkat partisipasi angkatan kerja (TPAK) [3]. Namun, data ini belum dimanfaatkan secara optimal dalam proses analisis prediktif [4].

Dalam upaya meningkatkan ketepatan pengambilan keputusan, metode *data mining* dapat digunakan untuk menggali pola dan tren tersembunyi dalam data tersebut [5]. Salah satu algoritma yang efektif *adalah K-Nearest Neighbors* (KNN), yang bekerja berdasarkan kedekatan antar data dan tidak memerlukan model matematis kompleks [6]. Algoritma ini cocok diterapkan pada data sosial-ekonomi yang bersifat numerik dan variatif [7]. Penelitian ini bertujuan membangun model prediksi tingkat kemiskinan menggunakan algoritma KNN dengan studi kasus di Kabupaten Sumba Timur. Diharapkan hasil dari model ini dapat digunakan sebagai alat bantu dalam proses identifikasi wilayah prioritas, sehingga kebijakan yang diterapkan dapat lebih tepat sasaran, efisien, dan berbasis data yang valid [8].

METODE PENELITIAN

Penelitian ini dilakukan untuk membangun model prediksi tingkat kemiskinan di Kabupaten Sumba Timur dengan menggunakan algoritma *K-Nearest Neighbors* (KNN) berbasis pendekatan *data mining*. Data yang digunakan dalam penelitian ini diperoleh dari Badan Pusat Statistik (BPS) Kabupaten Sumba Timur. Data yang digunakan meliputi indikator-indikator sosial ekonomi seperti persentase penduduk miskin, pengeluaran per kapita, Indeks Pembangunan Manusia (IPM), akses terhadap sanitasi layak, akses air minum layak, dan tingkat partisipasi angkatan kerja (TPAK) dari tahun 2020 hingga 2024, serta digunakan juga data prediksi untuk tahun 2025. Berikut alur penelitian yang digunakan:



Gambar 1. Alur Penelitian

Tahap pertama dalam penelitian ini adalah pengumpulan data yang berkaitan dengan tingkat kemiskinan di Kabupaten Sumba Timur. Data ini akan menjadi dasar dalam proses analisis untuk memprediksi tingkat kemiskinan menggunakan algoritma *K-Nearest Neighbors* (KNN).

Setelah data terkumpul, dilakukan tahap *pre-processing* untuk memastikan kualitas data sebelum dianalisis lebih lanjut. Proses ini mencakup pembersihan data dari nilai yang hilang (*missing value*) atau tidak valid, penghapusan data duplikat, serta normalisasi untuk menyeragamkan skala antar variabel numerik agar hasil perhitungan jarak pada algoritma KNN lebih akurat dan tidak bias terhadap atribut dengan nilai skala yang lebih besar.

Pada tahap pengolahan data, data yang telah diproses selanjutnya akan diolah dengan metode pemilihan fitur untuk menentukan faktor-faktor yang paling berpengaruh dalam memprediksi tingkat kemiskinan. Setelah itu, data dibagi menjadi dua bagian, yaitu data latih (*training data*) yang digunakan untuk membangun model prediksi dan data uji (*testing data*) yang digunakan untuk mengukur kinerja model.

Tahap pengujian dilakukan dengan menerapkan algoritma *K-Nearest Neighbors* (KNN) untuk memprediksi nilai persentase penduduk miskin berdasarkan kemiripan atribut sosial-ekonomi dengan data historis. Tahapan pengujian dimulai dengan menentukan jumlah K, yaitu jumlah tetangga terdekat yang akan digunakan dalam proses prediksi. Setelah itu, dilakukan perhitungan jarak antara data uji dengan seluruh data latih dalam dataset menggunakan rumus *Euclidean Distance* sebagai berikut:

$$d\,(x, y) = \sqrt{\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (1)$$

Tahap akhir dari penelitian ini adalah melakukan analisis terhadap hasil pengujian model untuk mengevaluasi kemampuan algoritma *K-Nearest Neighbors* (KNN) dalam memprediksi tingkat kemiskinan secara numerik. Evaluasi dilakukan dengan menggunakan metrik regresi seperti MAE, RMSE, dan *R-squared* guna mengukur tingkat kesalahan prediksi dan keakuratan model. Hasil evaluasi ini menjadi dasar untuk menilai apakah model yang dibangun sudah cukup andal dalam memperkirakan persentase kemiskinan berdasarkan data sosial-ekonomi. Jika model menunjukkan tingkat kesalahan yang rendah dan mampu menjelaskan variasi data secara signifikan, maka model tersebut dapat dimanfaatkan sebagai alat bantu dalam pengambilan keputusan, khususnya untuk mendukung perencanaan dan penyusunan strategi peningkatan kesejahteraan masyarakat di Kabupaten Sumba Timur secara lebih tepat sasaran dan berbasis data.

HASIL DAN PEMBAHASAN

Penelitian ini menghasilkan model prediksi tingkat kemiskinan berbasis algoritma *K-Nearest Neighbors* (KNN) dengan pendekatan *data mining*, menggunakan data sosial-ekonomi masyarakat Kabupaten Sumba Timur dari tahun 2020 hingga 2024. Enam variabel utama digunakan sebagai fitur dalam pemodelan, yaitu: persentase penduduk miskin, pengeluaran per kapita (ribu rupiah), indeks pembangunan manusia (IPM), akses terhadap sanitasi layak, akses air minum layak, dan tingkat partisipasi angkatan kerja (TPAK). Data diperoleh dari publikasi resmi Badan Pusat Statistik dan telah melalui tahapan pembersihan serta normalisasi. Berikut adalah format data yang dikumpulkan untuk masing-masing kecamatan:

Tabel 1. Data Latih (2020-2024)

Tahun	Kec/Kota	Penduduk Miskin (%)	Pengeluaran Per-Kapita (Ribu)	IPM	Akses Sanitasi (%)	Akses Air Minum (%)	TPAK
2020	Haharu	1.32	421	2.87	2.39	3.33	3.44
2020	Kahaungu Eti	1.29	433	3.09	2.35	3.08	3.15
2020	Kambata Mapambuhang	1.35	438	3	2.38	3.2	3.33
2020	Kambera	1.35	436	2.94	2.54	3.22	3.25
2020	Kanatang	1.33	418	2.91	2.45	3.1	3.17
2020	Karera	1.33	445	3.08	2.39	3.2	3.22
2020	Katala Hamu Lingu	1.28	427	3.02	2.47	3.18	3.26
2020	Kota Waingapu	1.3	414	3.07	2.39	3.29	3.32
…	…	…	…	…	…	…	…
2024	Wulla Waijelu	0.6	239	3.77	3.58	3.48	3.65

Setelah proses pengumpulan data selesai, dilakukan tahapan *pre-processing* untuk membersihkan data dari nilai tidak valid atau hilang pada kolom-kolom variabel sosial-ekonomi serta penghapusan data duplikat yang dapat menyebabkan distorsi hasil prediksi. Selain itu, dilakukan perhitungan rata-rata 5 tahun terakhir (2020–2024) yang digunakan sebagai nilai estimasi awal (*baseline*) untuk tahun 2025 sebagai pembanding hasil prediksi dari model. Berikut hasil perhitungan rata-rata (2020-2024) untuk data tahun 2025:

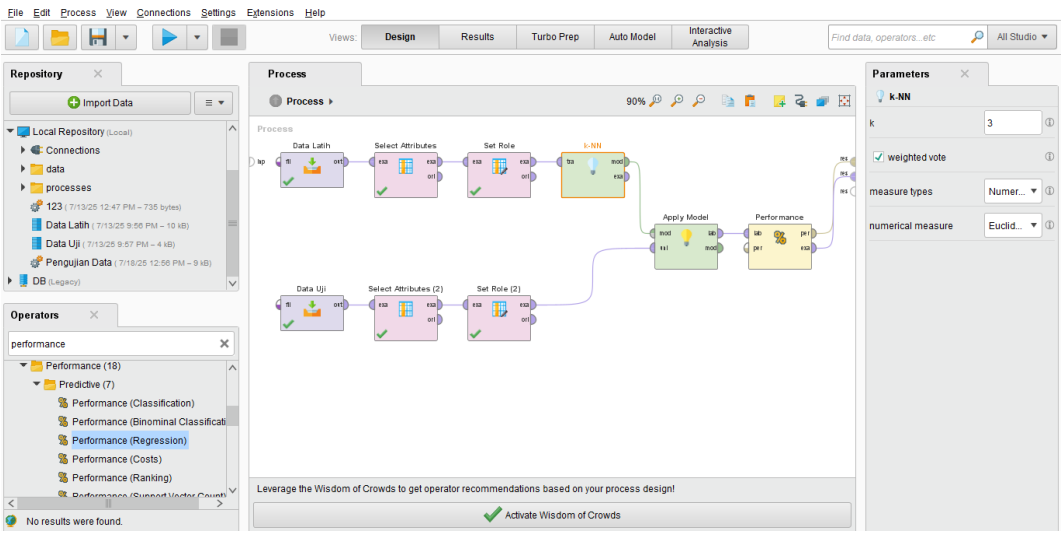
Tabel 2. Data Uji (2025)

Tahun	Kec/Kota	Penduduk Miskin (%)	Pengeluaran Per-Kapita (Ribu)	IPM	Akses Sanitasi (%)	Akses Air Minum (%)	TPAK
2025	Haharu	1.62	489	3.88	3.48	3.57	3.89
2025	Kahaungu Eti	1.13	576	2.85	2.63	3.24	3.28
2025	Kambata Mapambuhang	1.04	498	3.04	2.78	3.08	3.76
2025	Kambera	1.75	366	2.76	3.69	3.35	3.22
2025	Kanatang	1.21	628	3.21	2.66	3.1	3.72
2025	Karera	0.66	686	3.46	2.93	3.02	3.29

2025	Katala Hamu Lingu	1.85	698	2.61	3.34	3.65	3.7
2025	Kota Waingapu	0.59	648	3.6	3.52	3.81	3.97
2025	Lewa	0.88	430	3.09	3.3	3.18	3.61
2025	Lewa Tidahu	0.6	239	3.77	3.58	3.48	3.65
2025	Mahu	0.59	648	3.6	3.52	3.81	3.97
2025	Matawai La Pawu	1.85	698	2.61	3.34	3.65	3.7
2025	Ngadu Ngala	0.6	239	3.77	3.58	3.48	3.65
2025	Nggaha Ori Angu	1.04	498	3.04	2.78	3.08	3.76
2025	Paberiwai	0.59	648	3.6	3.52	3.81	3.97
2025	Pahunga Lodu	1.85	698	2.61	3.34	3.65	3.7
2025	Pandawai	1.62	489	3.88	3.48	3.57	3.89
2025	Pinu Pahar	0.6	239	3.77	3.58	3.48	3.65
2025	Rindi	1.62	489	3.88	3.48	3.57	3.89
2025	Tabundung	0.6	239	3.77	3.58	3.48	3.65
2025	Umalulu	1.85	698	2.61	3.34	3.65	3.7
2025	Wulla Waijelu	0.6	239	3.77	3.58	3.48	3.65

Tahapan selanjutnya yaitu pengolahan data untuk mempersiapkan data sebelum dibangun model prediksi tingkat kemiskinan menggunakan algoritma *K-Nearest Neighbors* (KNN). Pada tahapan ini dilakukan penentuan variabel yang digunakan serta membagi data latih (2020-2024) dan data uji (2025).

Setelah data latih dan data uji siap, maka tahap selanjutnya yaitu penerapan algoritma KNN untuk melakukan proses prediksi menggunakan *software RapidMiner*. Berdasarkan uji coba diperoleh bahwa nilai K = 3 merupakan nilai optimal yang digunakan dalam penelitian ini karena menghasilkan nilai RMSE dan MAE terkecil, sehingga model prediksi memiliki akurasi yang baik. Berikut desain proses pengujian data pada *RapidMiner*:



Gambar 2. Desain proses *RapidMiner*

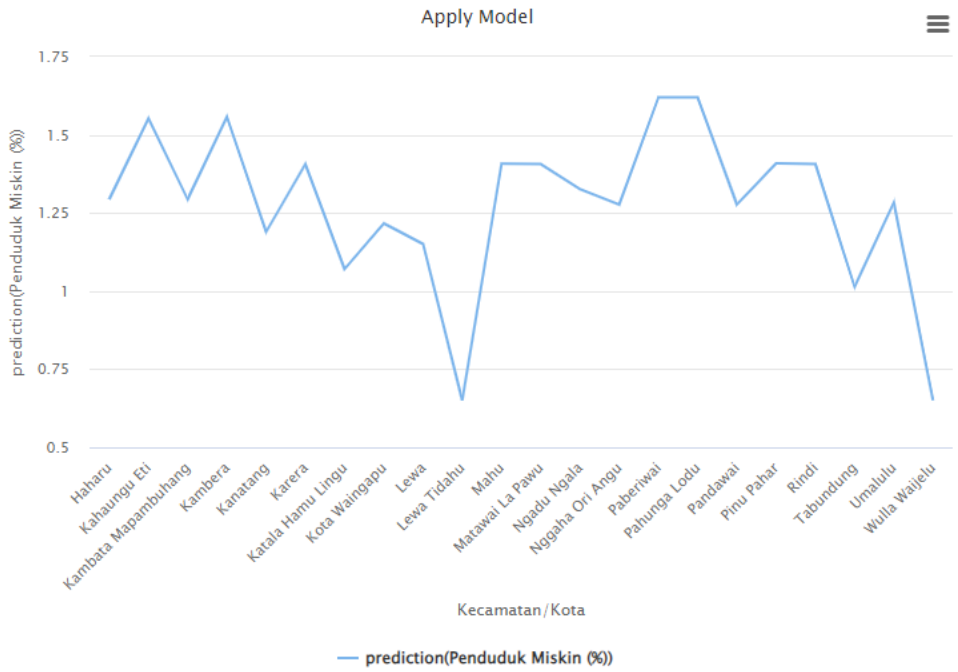
Desain proses diatas dimulai dengan *Read Excel* untuk mengimpor data latih tahun 2020-2024, dilanjutkan *Select Attributes* memilih variabel input dan target, serta *Set Role* untuk mengatur target sebagai label. Operator K-NN digunakan membangun model dengan K=3 dan metode *Euclidean Distance*. Selanjutnya, *Read Excel* juga digunakan untuk data uji tahun 2025, lalu *Select Attributes* (2) memilih variabel yang sama dan *Set Role* (2) mengatur target data uji sebagai label. Berikut hasil output prediksi tahun 2025 berdasarkan pengujian data pada *RapidMiner*:

Row No.	Kecamatan/...	Tahun	prediction(P...	Pengeluara...	IPM	Akses Sanit...	Akses Air Mi...	TPAK
1	Haharu	2025	1.294	438	3.310	2.870	3.530	3.660
2	Kahaungu Eti	2025	1.553	457	3.080	3.370	3.220	3.300
3	Kambata Ma...	2025	1.294	438	3.040	2.600	3.270	3.510
4	Kambera	2025	1.557	454	2.890	3.150	3.370	3.290
5	Kanatang	2025	1.190	405	3.070	2.830	3.240	3.480
6	Karera	2025	1.406	461	3.080	2.960	3.330	3.270
7	Katala Hamu ...	2025	1.070	537	2.840	3	3.290	3.370
8	Kota Waingapu	2025	1.216	529	3.120	3	3.590	3.600
9	Lewa	2025	1.151	426	3	2.910	3.360	3.400
10	Lewa Tidahu	2025	0.650	386	3.360	2.940	3.360	3.480
11	Mahu	2025	1.407	468	3.080	2.960	3.500	3.600
12	Matawai La P...	2025	1.406	463	2.950	2.780	3.320	3.420
13	Ngadu Ngala	2025	1.327	449	3.150	3.040	3.430	3.520
14	Nggaha Ori A...	2025	1.277	441	3	2.860	3.250	3.480
15	Paberiwai	2025	1.620	480	3.190	2.920	3.470	3.460
16	Pahunga Lodu	2025	1.620	474	3.070	2.960	3.360	3.490
17	Pandawai	2025	1.277	445	3.300	2.900	3.400	3.590
18	Pinu Pahar	2025	1.409	458	3.060	3.050	3.240	3.380
19	Rindi	2025	1.406	460	3.100	3.090	3.510	3.480
20	Tabundung	2025	1.014	391	3	2.960	3.330	3.470
21	Umalulu	2025	1.283	444	3.110	2.940	3.370	3.480
22	Wulla Waijelu	2025	0.650	378	3.220	2.800	3.390	3.550

ExampleSet (22 examples,3 special attributes,5 regular attributes)

Gambar 3. Hasil output prediksi 2025

Nilai prediksi yang dihasilkan model berkisar antara 0.650% hingga 1.620%, menunjukkan bahwa algoritma KNN mampu memberikan estimasi tingkat kemiskinan di setiap kecamatan berdasarkan variabel sosial-ekonomi input yang digunakan. Data prediksi ini akan digunakan sebagai dasar dalam analisis hasil dan penarikan kesimpulan pada tahap selanjutnya. Berikut gambar visualisasi hasil prediksi untuk tahun 2025:



Gambar 4. Visualisasi hasil prediksi 2025

Grafik di atas menunjukkan prediksi persentase penduduk miskin tahun 2025 pada setiap kecamatan di Kabupaten Sumba Timur dengan sumbu X berupa nama kecamatan dan sumbu Y nilai prediksi persentase penduduk miskin (%). Grafik menggunakan plot garis untuk menampilkan tren distribusi prediksi antar kecamatan, dengan nilai prediksi berkisar antara 0.65% hingga 1.62%. Puncak garis menunjukkan kecamatan dengan prediksi tingkat kemiskinan relatif lebih tinggi, sedangkan bagian lembah menunjukkan kecamatan dengan prediksi tingkat kemiskinan lebih rendah. Setelah dilakukan proses prediksi menggunakan algoritma *K-Nearest Neighbors* (KNN), evaluasi performa model dilakukan menggunakan operator *Performance* (*Regression*) pada *RapidMiner* untuk mengetahui akurasi model prediksi yang dihasilkan. Hasil evaluasi ditampilkan pada gambar berikut:

PerformanceVector

```
PerformanceVector:
root_mean_squared_error: 0.257 +/- 0.000
absolute_error: 0.207 +/- 0.152
squared_correlation: 0.090
```

Gambar 5. Evaluasi performa model

Gambar di atas menampilkan hasil evaluasi model KNN berupa metrik performa regresi dengan nilai *Root Mean Squared Error* (RMSE) sebesar 0.257 yang menunjukkan rata-rata kesalahan prediksi model hanya meleset sekitar 0.257% dari nilai aktual, serta nilai *Absolute Error* (MAE) sebesar 0.207 yang menandakan kesalahan prediksi model sangat kecil dan dapat diterima. Nilai *Squared Correlation* sebesar 0.090 menunjukkan korelasi antara variabel prediksi dan nilai aktual tergolong rendah, namun model tetap memiliki

akurasi yang cukup baik untuk data yang digunakan. Secara keseluruhan, nilai RMSE dan MAE yang rendah menunjukkan bahwa model KNN memiliki tingkat akurasi yang baik dalam memprediksi tingkat kemiskinan di Kabupaten Sumba Timur.

Pengujian dilakukan untuk mengevaluasi performa model prediksi tingkat kemiskinan menggunakan algoritma K-Nearest Neighbors (KNN) pada RapidMiner. Model ini menghasilkan prediksi tingkat kemiskinan tahun 2025 di setiap kecamatan dengan nilai berkisar antara 0.650% hingga 1.620%, dimana kecamatan dengan prediksi tertinggi yaitu Paberiwai (1.620%), Kambera (1.557%), dan Kahaungu Eti (1.553%), sedangkan yang terendah adalah Lewa Tidahu (0.650%), Wulla Waijelu (0.650%), dan Tabundung (1.014%). Hasil ini menunjukkan adanya variasi tingkat kemiskinan antar kecamatan yang dapat menjadi dasar bagi pemerintah dalam menetapkan prioritas intervensi program pengurangan kemiskinan secara tepat sasaran.

KESIMPULAN

Hasil pengujian model menunjukkan nilai evaluasi performa yang sangat baik, dengan nilai Root Mean Squared Error (RMSE) sebesar 0.257%, Mean Absolute Error (MAE) sebesar 0.207% serta Squared Correlation sebesar 0.090%. Nilai ini menunjukkan bahwa rata-rata kesalahan prediksi model sangat kecil, sehingga model memiliki tingkat akurasi tinggi dan dapat diandalkan. Berdasarkan hasil prediksi, kecamatan dengan tingkat kemiskinan prediksi tertinggi adalah Paberiwai (1.620%), sedangkan kecamatan dengan tingkat kemiskinan prediksi terendah adalah Lewa Tidahu (0.650%). Model yang dihasilkan memberikan gambaran wilayah mana yang memerlukan perhatian lebih besar dalam upaya peningkatan kesejahteraan masyarakat. Keberhasilan penerapan KNN pada penelitian ini juga menunjukkan bahwa pemanfaatan algoritma data mining dalam perencanaan pembangunan daerah dapat menjadi salah satu strategi inovatif untuk mendukung pengambilan keputusan berbasis data.

DAFTAR PUSTAKA

[1] A. de Haan, *Social Policy and Human Development: A Rights-Based Approach*. Geneva: UNRISD, 2015.

[2] Badan Pusat Statistik, *Kemiskinan dan Ketimpangan Pendapatan di Indonesia: Laporan Tahunan 2023*. Jakarta: BPS, 2023.

[3] Badan Pusat Statistik Kabupaten Sumba Timur, *Profil Sosial Ekonomi Kabupaten Sumba Timur 2020*. Waingapu: BPS Kabupaten Sumba Timur, 2020.

[4] N. A. Sudibyo, A. Iswardani, K. Sari, and S. Suprihatiningsih, "Penerapan data mining pada jumlah penduduk miskin di Indonesia," *Jurnal Informatika Sosial*, vol. 4, no. 3, pp. 87–102, 2020.

[5] M. Lantz, "Machine learning with R: expert techniques for predictive modeling," *Packt Publishing*, 3rd ed., 2019.

[6] M. Faisal, W. Utami, and S. Parmica, "Implementasi algoritma K-Nearest Neighbor (KNN) dalam memprediksi indeks kemiskinan," *Jurnal Data Science Indonesia*, vol. 5, no. 2, pp. 45–58, 2023.

[7] R. Sharda, D. Delen, and E. Turban, *Analytics, Data Science, & Artificial Intelligence: Systems for Decision Support*, 11th ed. London: Pearson, 2020, pp. 56–75.

[8] T. Rashid, *Data Mining Algorithms: Explained Using R*. Boca Raton: CRC Press, 2016, pp. 33–47.

[9] I. H. Witten, E. Frank, and M. A. Hall, *Data Mining: Practical Machine Learning Tools and Techniques*, 4th ed. Burlington: Morgan Kaufmann, 2016.

[10] World Bank, *Poverty and Shared Prosperity 2022: Correcting Course*. Washington, DC: World Bank Group, 2022.