INFORMATIKA
Jurnal Informatika, Manajemen dan Komputer, Vol. 16 No. 2, Desember 2024
eISSN : 2580-3042
pISSN : 1979-0694

TINJAUAN SISTEMATIS METODE LINEAR REGRESSION, K-NEAREST NEIGHBOR DAN RANDOM FOREST UNTUK PREDIKSI TINGKAT KEMISKINAN

Borianto[1], Billy Hendrik[2]
[1,2]Universitas Putra Indonesia YPTK Padang
Jl. Raya Lubuk Begalung Padang, Sumatera Barat, 25221
e-mail: borianto96@gmail.com

ABSTRAK

Tinjauan sistematis terhadap metode prediksi tingkat kemiskinan ini bertujuan untuk memberikan gambaran menyeluruh mengenai pendekatan, metode, dan hasil penelitian terdahulu di bidang ini. Pendekatan yang digunakan adalah tinjauan literatur dengan menganalisis berbagai algoritma machine learning, seperti Linear Regression, K-Nearest Neighbor (KNN), dan Random Forest, yang digunakan untuk memprediksi tingkat kemiskinan berdasarkan faktor sosial, ekonomi, dan demografi. Studi ini membahas elemen penting seperti preprocessing data, teknik data mining, dan tingkat akurasi model prediksi, dengan fokus pada data yang bersifat linear maupun data dengan kompleksitas tinggi atau non-linear antar variabel. Hasil kajian menunjukkan bahwa pemilihan algoritma bergantung pada karakteristik dataset, dan tinjauan ini diharapkan dapat menjadi landasan untuk pengembangan model prediksi yang lebih akurat dalam mendukung kebijakan pengentasan kemiskinan.

Kata kunci: *Prediksi tingkat kemiskinan, Machine learning, Regresi linear, K-Nearest Neighbor, Random Forest*

ABSTRACT

A systematic review of poverty rate prediction methods aims to provide a comprehensive overview of approaches, methods, and findings from previous studies in this field. The approach used is a literature review that analyzes various machine learning algorithms, such as Linear Regression, K-Nearest Neighbor (KNN), and Random Forest, which are employed to predict poverty rates based on social, economic, and demographic factors. This study discusses critical elements such as data preprocessing, data mining techniques, and the accuracy level of prediction models, focusing on both linear data and data with high complexity or non-linear relationships between variables. The findings reveal that the choice of algorithm depends on the dataset characteristics, and this review is expected to serve as a foundation for developing more accurate prediction models to support poverty alleviation policies.

Keywords: *Poverty level prediction, Machine learning, Linear regression, K-Nearest Neighbor, Random Forest.*

1. PENDAHULUAN

Kemiskinan merupakan permasalahan global yang kompleks dan multidimensional, memengaruhi aspek sosial, ekonomi, dan kesejahteraan masyarakat secara luas. Upaya pengentasan kemiskinan sering kali membutuhkan pemahaman mendalam mengenai faktor-faktor penyebabnya serta identifikasi kelompok masyarakat yang rentan terhadap kemiskinan. Dengan kemajuan teknologi informasi dan peningkatan ketersediaan data sosial-ekonomi, pendekatan berbasis data menjadi semakin relevan untuk membantu pengambilan keputusan yang lebih efektif.

Kemiskinan menjadi salah satu parameter yang dipakai untuk menilai tingkat kesejahteraan suatu bangsa. Apabila jumlah individu yang mengalami kesulitan ekonomi semakin meningkat, semakin rendah tingkat kesejahteraan negara tersebut (Sazwati et al., 2024).

Kemiskinan juga merupakan sebuah fenomena yang dihadapi hamper semua negara di dunia mengalami kemiskinan, terutama negara

INFORMATIKA
Jurnal Informatika, Manajemen dan Komputer, Vol. 16 No. 2, Desember 2024
eISSN : 2580-3042
pISSN : 1979-0694

berkembang yang selalu menjadi perhatian oleh pemerintah di berbagai Negara mana pun, termasuk Indonesia. Masalah kemiskinan ini bukanlah merupakan hal yang baru, kemiskinan muncul karena ketidakmampuan manusia sebagai masyarakat untuk memperoleh sumber daya yang cukup dalam mencukupi kebutuhan dasar hidupnya seperti pakaian, makanan dan tempat tinggal. Kemiskinan bisa menghambat kesejahteraan dan peradaban masyarakat, salah satunya dipengaruhi oleh rendahnya tingkat pendapatan yang diperoleh masyarakat untuk menjamin keberlangsungan hidupnya (Irawati & Pakereng, 2023).

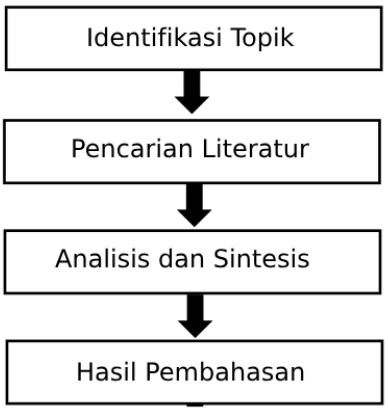
Peneliti sebelumnya telah menguji metode *Linear Regression* untuk menghitung kesalahan Rata-Rata Kuadrat, Kesalahan Rata-Rata Absolut, dan Kesalahan Rata-Rata Relatif Absolut (RMSE, MAE, dan MARE) (Laelatul Azizah et al., 2024).

Peneliti sebelumnya juga menguji dengan menerapkan metode K-Nearest Neighbor (KNN) untuk memprediksi indeks kemiskinan dengan tepat, hal tersebut akan menjadikan proses pemberian bantuan terhadap rakyat miskin akan lebih efektif dan efisien ditahun-tahun berikutnya. Selain itu penelitian ini menggunakan pendekata kuantitatif untuk mengukur nilai rata-rata indeks kemiskinan dalam kurun waktu tertentu. Data mining memiliki banyak algoritma dalam proses memprediksi suatu kasus atau masalah tertentu salah satunya adalah K-Nearest Neighbor (KNN) (Faisal et al., 2023).

Peneliti sebelumnya juga telah menguji dengan metode *Random Forest menggunakan Random Forest* yang bertujuan untuk memprediksi tingkat kesejahteraan masyarakat nelayan, dan juga untuk mengetahui performance algoritma random forest dalam memprediksi sebuah dataset. Metode ini digunakan untuk menganalisis dan memprediksi data dengan hasil akhir sebuah keputusan sejahtera dan tidak sejahtera untuk masyarakat nelayan (Sandi et al., 2023).

2. METODE PENELITIAN

Penelitian dengan metode *literature review* bertujuan untuk menggali, menganalisis, dan menyintesis berbagai karya ilmiah yang relevan untuk mendapatkan pemahaman mendalam tentang topik tertentu.



Gambar 1: Bagan Metode Penelitian

2.1. Mengidentifikasi Topik

Penelitian ini fakus pada prediksi tingkat kemiskinan menggunakan berbagai pendekatan dan algoritma. Tujuannya adalah mengevaluasi metode yang telah digunakan, memahami hasil yang dicapai, dan mengidentifikasi tantangan dalam penelitian terdahulu.

2.2. Pencarian Literatur

Proses ini dilakukan dengan menggunakan basis data akademik seperti Google Scholar, IEEE Xplore, atau database lainnya yang kredibel. Kata kunci seperti poverty prediction, machine learning, linear regression, KNN, dan Random Forest digunakan untuk menemukan artikel yang relevan. Artikel yang dipilih harus memenuhi kriteria inklusi, seperti relevansi, keandalan, dan publikasi terkini, sementara artikel yang tidak sesuai dengan fokus penelitian dikeluarkan.

2.3. Analisis dan Sintesis

Setelah literatur terkumpul, tahap analisis dan sintesis dilakukan untuk memahami metode, dataset, parameter, hasil, dan metrik evaluasi yang digunakan dalam masing-masing penelitian. Artikel-artikel tersebut dikelompokkan berdasarkan tema, pendekatan, atau hasil penelitian. Evaluasi kualitas literatur dilakukan untuk menilai keandalan dan validitas metode serta konsistensi hasilnya.

2.3. Hasil

Hasil analisis ini kemudian disusun dalam bentuk laporan atau artikel yang sistematis. Laporan ini biasanya terdiri dari pendahuluan, metodologi, hasil kajian, diskusi, dan kesimpulan. Temuan utama dirangkum untuk memberikan wawasan tentang kelebihan dan kekurangan metode yang dikaji, serta rekomendasi untuk penelitian di masa depan.

Dengan metode ini, penelitian literature review dapat memberikan kontribusi penting dalam

memahami perkembangan topik serta menjadi dasar yang kuat untuk penelitian lanjutan.

3. HASIL DAN PEMBAHASAN

Kemiskinan didefinisikan sebagai standar tingkat hidup yang rendah,yaitu adanya suatu tingkat kekurangan materi pada sejumlah atau golongan orang dibandingkan dengan standar hidup yang berlaku dalam masyarakat bersangkutan (Agustiya et al., 2024).

Kemiskinan juga merupakan permasalahan yang senantiasa dihadapi oleh manusia. Masalah kemiskinan ini telah ada sejak awal Sejarah kemanusiaan dan dampaknya dapat melibatkan seluruh aspek kehidupan manusia. Meskipun terkadang kehadirannya tidak disadari sebagai masalah oleh sebagian individu, bagi masyarakat yang berada dalam kondisi miskin, dapat merasakan dan mengalami bagaimana hidup dengan keterbatasan ekonomi. Dalam konteks yang lebih luas, kemiskinan dapat diartikan sebagai standar tingkat hidup yang rendah, mengindikasikan tingkat kekurangan materi pada sejumlah individu atau kelompok yang tidak sebanding dengan standar umum kehidupan dalam Masyarakat (Sazwati et al., 2024).

Regresi linear adalah teknik analisis statistik yang digunakan untuk memodelkan hubungan linier antara satu atau lebih variabel independen dan variabel dependen. Dengan menggunakan regresi linear, kita dapat memprediksi nilai variabel dependen berdasarkan nilai variabel independen yang telah diketahui (Pustaka, 2024)

Pada tahap preprocessing data, fokus utama adalah menemukan dan memperbaiki berbagai elemen data yang dapat mengakibatkan ketidaklengkapan, kebenaran, relevansi, ketidakakuratan, atau bahkan kehilangan informasi. Integrasi perubahan data disesuaikan dengan kebutuhan penelitian yang relevan karena fokus penelitian penulis untuk analisis prediksi regresi adalah rentang waktu lima tahun terakhir, yaitu dari 2018 hingga 2022 (Laelatul Azizah et al., 2024)

Prediksi merupakan dugaan atau prediksi mengenai terjadinya suatu kejadian atau peristiwa di waktu yang akan datang. Prediksi bisa bersifat kualitatif (tidak berbentuk angka) maupun kuantitatif (berbentuk angka). Prediksi kualitatif sulit dilakukan untuk memperoleh hasil yang baik karena variabelnya sangat relatif sifatnya. Prediksi kuantitatif dibagi dua yaitu: prediksi tunggal (point prediction) dan prediksi selang (interval prediction). Prediksi tunggal terdiri dari satu nilai, sedangkan prediksi selang terdiri dari beberapa nilai, berupa suatu selang (interval) yang dibatasi oleh nilai batas bawah prediksi batas bawah dan batas atas prediksi tinggi (Novianty et al., 2021).

Data mining adalah teknologi komputer yang mengklasifikasikan, mengkategorikan dan memprediksi jumlah data yang besar (Riandari et al., 2022). Data mining dalam arti luasnya yaitu proses memperoleh pengetahuan yang tersembunyi dan berpotensi berharga dari informasi yang luas, tidak sempurna dan data acak. Definisi tersebut mengandung makna sebagai berikut: (1) data sumbernya harus nyata, masif, dan berisik. (2) Apa yang ditemukan bermakna pengetahuan bagi penggunanya. (3) Pengetahuan yang diperoleh harus dapat diterima, dapat dimengerti dan hasilnya diharapkan dapat diterapkan (Yin et al., 2024).

3.1. Algoritma Linear Regression

Regresi linear adalah teknik analisis statistik yang digunakan untuk memodelkan hubungan linier antara satu atau lebih variabel independen dan variabel dependen. Dengan menggunakan regresi linear, kita dapat memprediksi nilai variabel dependen berdasarkan nilai variabel independen yang telah diketahui (Pustaka, 2024).

Untuk melihat hubungan antara dua variabel, gunakan regresi linier sederhana. Hubungan ini dapat berupa hubungan sebab akibat, hubungan korelasi, atau hubungan fungsional. Variabel penyebab dan variabel akibat biasanya digambarkan dengan huruf X dan Y. Secara umum, rumus untuk persamaan linier regresi adalah sebagai berikut:

$$Y = a + bX \quad [9]$$

Y = Variabel dependen
a = Konstanta
b = Koefisien
X = Variabel Independen

$$a = \frac{(\Sigma y)(\Sigma x^2) - (\Sigma x)\,\Sigma xy)}{n.(\Sigma x^2) - (\Sigma x)^2}$$

$$b = \frac{n.(\Sigma xy) - (\Sigma x) - (\Sigma y)}{n.(\Sigma x^2) - (\Sigma x)^2}$$

Sumber Referensi (Laelatul Azizah et al., 2024)

Pengambilan data untuk tinjauan sistematis dalam penelitian ini dilakukan dengan merujuk pada artikel jurnal berjudul “Prediksi Tingkat Kemiskinan Di Provinsi Jawa Barat Menggunakan Algoritma Regresi Linear” (Laelatul Azizah et al., 2024) yang diterbitkan oleh Novi Laelatul Azizah 1, Nana Suarna 2, Willy Prihartono 3 pada tahun 2024 di JATI (Jurnal Mahasiswa Teknik Informatika). Artikel ini dipilih karena memiliki relevansi yang signifikan dengan topik penelitian, yaitu Prediksi Tingkat Kemiskinan, serta menawarkan perspektif yang mendalam berdasarkan kajian teoritis dan

INFORMATIKA
Jurnal Informatika, Manajemen dan Komputer, Vol. 16 No. 2, Desember 2024
eISSN : 2580-3042
pISSN : 1979-0694

empiris.

Metode dalam artikel tersebut dianalisis untuk memahami pendekatan yang digunakan, termasuk *Linear Regression* untuk prediksi tingkat kemiskinan. Data utama yang diambil mencakup Jumlah penduduk miskin, nama kota, serta tahun yang mendukung eksplorasi masalah penelitian.

Dengan mengacu pada artikel, penelitian ini mampu mengidentifikasi kerangka teori, metode yang relevan, atau temuan empiris yang akan digunakan sebagai landasan untuk membangun argumen dan analisis lebih lanjut.

3.1.1. Algoritma Linear Regression

Pada tahap preprocessing data, fokus utama adalah menemukan dan memperbaiki berbagai elemen data yang dapat mengakibatkan ketidaklengkapan, kebenaran, relevansi, ketidakakuratan, atau bahkan kehilangan informasi. Integrasi perubahan data disesuaikan dengan kebutuhan penelitian yang relevan karena fokus penelitian penulis untuk analisis prediksi regresi adalah rentang waktu lima tahun terakhir, yaitu dari 2018 hingga 2022.

Tabel 1: Preprocessing

Id	Nama Kabupaten Kota	Jumlah Penduduk Miskin	Tahun
1.	Kabupaten Bogor	415	2018
2.	Kabupaten Sukabumi	166,3	2018
3.	Kabupaten Cianjur	221,6	2018
4.	Kabupaten Bandung	246,1	2018
5.	Kabupaten Garut	241,3	2018
6.	Kabupaten Tasikmalaya	172,4	2018
7.	Kabupaten Ciamis	85,7	2018
8.	Kabupaten Kuningan	131,2	2018
…	…	…	…
129.	Kota Bandung	109,8	2022
130.	Kota Cirebon	31,5	2022
131.	Kota Bekasi	137,4	2022
132.	Kota Depok	64,4	2022
133.	Kota Cimahi	31,2	2022
134.	Kota Tasikmalaya	87,1	2022
135.	Kota Banjar	12,7	2022

3.1.2. Data Mining

Pada tahap ini, peneliti akan membangun model untuk memprediksi jumlah penduduk miskin dengan menggunakan model regresi linear yang ada di RapidMiner versi 10.1. Selanjutnya, akan menggunakan dataset 135 baris data dengan tahun (integer) dan label untuk jumlah penduduk miskin. Untuk membuat model linier regresi, beberapa operator digunakan, seperti pembagian data, yang digunakan untuk membagi data menjadi dua, yaitu data latihan dan data pengujian; operator linear regresi juga digunakan; dalam penelitian ini, percobaan egresi linier dilakukan dengan menggunakan 80% data pelatihan dan 20% data pengujian.

3.1.3. Hasil

Hasil evaluasi pada tahapan regresi linear sederhana menunjukkan bahwa kinerja model prediksi tingkat kemiskinan di Provinsi Jawa Barat dapat diukur melalui beberapa metrik evaluasi. Dalam hal ini, nilai RMSE (Kesalahan Rata-rata Kuadrat) mencapai 65,837 dengan nilai standar yang sangat rendah, yaitu +/- 0,000. Selain itu, Kesalahan Rata-rata Absolut (MAE) menghasilkan nilai sebesar 53,156 dengan nilai standar +/- 38,845. Sementara itu, Kesalahan Rata-rata Relatif Absolut (MARE) menunjukkan angka sebesar 66,47% dengan nilai standar +/- 80,57%. Hasil ini menandakan bahwa model regresi linear sederhana mampu memberikan prediksi yang sangat baik, terutama karena nilai-nilai RMSE, MAE, dan MARE mendekati 0 atau nilai aktual. dapat disimpulkan bahwa model ini memiliki kinerja yang baik dalam meramalkan tingkat kemiskinan di Provinsi Jawa Barat.

3.1.4. Knowledge

Hasil jumlah penduduk miskin di Provinsi Jawa Barat menunjukkan bahwa pada tahun 2018, jumlah penduduk miskin tertinggi di Provinsi tersebut mencapai 415.000 jiwa. Hasil prediksinya meningkat menjadi 239.429 jiwa, tetapi kemudian terjadi penurunan. Pada tahun 2019, jumlah penduduk miskin sebelumnya 235.200 jiwa, dan hasil prediksinya menurun menjadi 213.718 jiwa. Namun, pada tahun 2022, jumlah penduduk miskin sebelumnya 94 justru meningkat menjadi 226.909 jiwa meningkat. Hasil prediksi tersebut dapat dilihat pada Gambar 2.

Row No.	tahun	jumlah_pen...	predictio...	nama_kabu...	id
16	2021	269.200	243.588	3	85
1	2018	415	239.429	0	1
20	2022	94	226.909	6	115
9	2020	262.800	224.137	4	59
17	2021	96.600	216.490	6	88
6	2019	235.200	213.718	4	32
21	2022	266.100	208.844	8	117
22	2022	147.100	199.812	9	118
10	2020	247.900	188.007	8	63
2	2018	131.200	176.202	7	8
23	2022	155.300	172.715	12	121
3	2018	129.300	158.138	9	10

ExampleSet (27 examples, 3 special attributes, 2 regular attributes)

Gambar 2: Hasil Regresi Linear Sumber Referensi Hasil Analisis Peneliti Software RapidMiner

Berdasarkan hasil analisis grafik bar, terdapat perkembangan terkait prediksi jumlah penduduk

miskin di Provinsi Jawa Barat. Pada tahun 2018, jumlah penduduk miskin di Provinsi tersebut sebesar 415.000 jiwa. Jumlah tersebut meningkat hingga mencapai 266.100 jiwa pada tahun 2022. jumlah penduduk miskin di Provinsi tersebut kembali menurun menjadi 226.909 jiwa.

3.2. Algoritma K-Nearest Neighbor

Algoritma K-NN atau K-Nearest Neighbor adalah metode klasifikasi pada data mining berdasarkan jarak terdekat dengan data uji. Tahapan metode K-NN ini dengan mencari pola dan informasi yang menarik dari data dengan atribut k, yaitu jumlah tetangga yang dijadikan acuan pada data testing terhadap data training. Algoritma K-NN berjalan dengan menentukan data training sebagai index terbanyak sebagai bahan perhitungan dengan data testing (Duwo Jiwo Saputro et al., 2024). Proses metode K-NN sebagai berikut:

a. Menyiapkan data presentase kemiskinan kabupaten/kota Provinsi Jawa Barat yang menjadi data training dan data testing
b. Menentukan nilai k
c. Menghitung jarak (D) dari data testing ke data training menggunakan rumus Euclidean Distance
$$Eucidean\ Distance = \sqrt{\sum_{i=1}^{n}(X_i - Y_i)^2}$$
d. Mengurutkan data (D) menurut rangking terkecil
e. Klasifikasi data testing mayoritas

Pengambilan data untuk tinjauan sistematis dalam penelitian ini dilakukan dengan merujuk pada artikel jurnal berjudul "Analisa Data Mining Dalam Memprediksi Masyarakat Kurang Mampu Menggunakan Metode K-Nearest Neighbor" (Nurdin, 2024) yang diterbitkan oleh Nurdin1 1, Mela Rizki 2, Maryana 3 pada tahun 2024 di JITET (Jurnal Informatika dan Teknik Elektro Terapan). Artikel ini dipilih karena memiliki relevansi yang signifikan dengan topik penelitian, yaitu Prediksi Tingkat Kemiskinan, serta menawarkan perspektif yang mendalam berdasarkan kajian teoritis dan empiris.

Metode dalam artikel tersebut dianalisis untuk memahami pendekatan yang digunakan, termasuk *K-Nearest Neighbor* untuk prediksi masyarakat kurang mampu. Data utama yang diambil mencakup data kependudukan, pekerjaan, penghasilan dan kondisi rumah yang mendukung eksplorasi masalah penelitian.

Dengan mengacu pada artikel, penelitian ini mampu mengidentifikasi kerangka teori, metode yang relevan, atau temuan empiris yang akan digunakan sebagai landasan untuk membangun argumen dan analisis lebih lanjut.

3.2.1. Variabel dan Dataset yang digunakan

Data yang digunakan dalam penelitian ini diperoleh dari Dinas Sosial Kabupaten Aceh Tengah. Variabel atau atribut penelitian yang digunakan sebagai dasar kebutuhan dari metode yang digunakan. Untuk meineintuikan nilai bobot, peineiliti meimasuikkan nilai sampeil peirhituingan, bobot yang dimiliki masing- masing atribuit ataui kriteiria akan dijuimlahkan uintuik meineintuikan kateigori data teirseibuit "Mampui" ataui "Tidak Mampui". Beirikuit bobot nilai seitiap kriteiria.

Tabel 2: Bobot nilai setiap kriteria

Kode	Nama Kriteria	Nilai	Bobot
X1	Pekerjaan	Tidak bekerja	4
		Buruh, petani	3
		Pedagang, wiraswasta pensiunan, pegawai.	2
		PNS, Polri, TNI	1
X4	Penghasilan	<500.000	4
		500.000 - 1.000.000	3
		1.000.000 - 2.500.000	2
		>2.500.000	1
X5	Kondisi Rumah	Bambu / Triplek	4
		Kayu	3
		Batako	2
		Batu bata	1
		Beton	0

3.2.2. Penerapan Metode *K-Nearest Neighbor*

Dalam penelitian ini menggunakan data training sejumlah 20 data, data training ini diharapkan mampu menjadikan sistem yang akan dibangun memiliki hasil perhitungan tepat dan akurat.

INFORMATIKA
Jurnal Informatika, Manajemen dan Komputer, Vol. 16 No. 2, Desember 2024
eISSN : 2580-3042
pISSN : 1979-0694

Tabel 3: Data *training*

Nama kepala keluarga	X1	X2	X3	X4	X5	Label
xxxxxxxx	3	4	44	3	3	Kurang mampu
xxxxxxxx	3	3	49	3	3	Kurang mampu
xxxxxxxx	3	2	55	2	3	Kurang mampu
xxxxxxxx	3	3	40	3	2	Kurang mampu
xxxxxxxx	3	4	41	3	2	Kurang mampu
xxxxxxxx	3	4	39	3	3	Kurang mampu
xxxxxxxx	2	4	46	2	0	Mampu
xxxxxxxx	3	3	37	3	3	Kurang mampu
xxxxxxxx	3	5	43	3	3	Kurang mampu
xxxxxxxx	3	3	40	3	3	Kurang mampu
xxxxxxxx	3	1	33	3	3	Kurang mampu
xxxxxxxx	2	3	29	2	2	Mampu
……	..	..	..	..	..	……
xxxxxxxx	3	4	50	3	3	Kurang mampu

3.2.3. Hasil

a. Variabel yang diguinakan pada penelitian ini adalah data Kepala Keluarga yaitu: Jumlah pekerjaan, Juimlah tanggungan, Umur, Penghasilan dan Kondisi rumah.
b. Penelitian ini menggunakan data training dan data testing dengan rasio perbandingan 70:30 dari dataset yang berjumlah 309 data yaitu data training sebanyak 216 data, dan data testing sebanyak 93 data.
c. Hasil peinelitian ini meinghasilkan sebuah output prediksi metode K-Nearest Neighbor dengan menggunakan 2 class, yaitu: Mampui dan Tidak Mampu.
d. nilai akurasi yang diperoleih menunjukkan Algoritma K-Nearest Neighbor dengan menggunakan pendekatan *Euclidean Distance* dan K=3 didapatkan tingkat keakurasian algoritma sebesar 86,02%.
e. Berdasarkan penelitian prediksi Masyarakat Kurang Mampu di Kecamatan Pegasing, menggunakan algoritma K-Nearest Neighbor yang menghasilkan nilai performa algoritma yang cuikuip baik yaitu: *accuracy* sebesar 86,02%, precission sebesar 72,22%, recall sebesar 61,90% dan F1-Score sebesar 66,04%

3.3. Algoritma Random Forest

Untuk Random Forest adalah sebuah metode yang memiliki konsep pembuatan sejumlah besar pohon keputusan dimana semua pohon berkolerasi dan bertindak sebagai model ansambel. Prediksi kelas dan keputusan diletakkan di setiap pohon keputusan dengan dasar hasil maksimum (Sandi et al., 2023).

$$\hat{y}^{(t)} = \frac{1}{N_t} \sum_{i \in \text{Leaf}(t)} y_i$$

Di mana:

- $\hat{y}^{(t)}$: prediksi dari pohon ke-t.
- N_t: jumlah data di daun pohon t.
- y_i: nilai aktual dari data pada daun tersebut.

Pengambilan data untuk tinjauan sistematis dalam penelitian ini dilakukan dengan merujuk pada artikel jurnal berjudul "Analisa Prediksi Kesejahteraan Masyarakat Nelayan Lombok Timur Menggunakan Algoritma Random Forest" (Sandi et al., 2023) yang diterbitkan oleh Arnila Sandi 1, Kusrini 2, Kusnawi 3 pada tahun 2023 di Infotek : Jurnal Informatika dan Teknologi. Artikel ini dipilih karena memiliki relevansi yang signifikan dengan topik penelitian, yaitu Prediksi Tingkat Kemiskinan, serta menawarkan perspektif yang mendalam berdasarkan kajian teoritis dan empiris.

Metode dalam artikel tersebut dianalisis untuk memahami pendekatan yang digunakan, termasuk *Random Forest* untuk prediksi kesejahteraan masyarakat nelayan. Data utama yang diambil mencakup Jumlah penduduk miskin, nama kota, serta tahun yang mendukung eksplorasi masalah penelitian.

Dengan mengacu pada artikel, penelitian ini mampu mengidentifikasi kerangka teori, metode yang relevan, atau temuan empiris yang akan digunakan sebagai landasan untuk membangun argumen dan analisis lebih lanjut.

3.3.1. Pembahasan

Proses pengolahan data pada penelitian ini menggunakan software Rapidminer dengan Jumlah data masyarakat nelayan yang terkumpul adalah 1855 data dengan atribut yang digunakan sebanyak 8 (delapan) atribut, antara lain: Nomor Anggota, Nama, Dusun, Pendidikan, Anggota Keluarga, Sumur, Pekerjaan dan Rumah. Dari delapan atribut tersebut dapat digunakan sebagai komponen penentu keputusan akhir dari keaadaan masyarakat nelayan.

Tabel 4: Atribut yang digunakan

INFORMATIKA
Jurnal Informatika, Manajemen dan Komputer, Vol. 16 No. 2, Desember 2024
eISSN : 2580-3042
pISSN : 1979-0694

No	Atribut
1	No Anggota
2	Nama
3	Dusun
4	Pendidikan
5	Anggota Keluarga
6	Sumur
7	Pekerjaan
8	Perumahan

Berikut adalah hasil dari pengolahan data menggunakan rapidminer dengan metode klasifikasi dan algoritma random forest. Pengujian dilakukan dengan K-Fold 10 cross validation.

Tabel 5: Hasil pengolahan data menggunakan *Algoritma Random Forest*

Cross Validation	Hasil Pengolaha Algoritma Random Forest			
	Accurasi	Precission	Recall	AUC
	%	%	%	%
k-2	93,37	100	93,37	0,597
k-3	93,37	100	93,37	0,701
k-4	93,37	100	93,37	0,694
k-5	93,37	100	93,37	0,735
k-6	93,32	100	93,32	0,663
k-7	93,37	100	93,37	0,694
k-8	93,37	100	93,37	0,667
k-9	93,37	100	93,37	0,723
k-10	93,37	100	93,37	0,726

Dari hasil pengujian menggunakan k-fold cross validation hasil akurasi tertinggi berada di k-5 yaitu 93,37% untuk nilai accurasi dan 0,735 untuk nilai AUC. Dari tabel ini dapat dibuktikan bahwa algoritma random forest sangat baik untuk mengkalsifikasikan data masyarakat nelayan, sehingga nilai yang diperoleh dapat digunakan untuk memprediksi tingkat kesejahteraan masyarakat nelayan di Lombok Timur NTB.

3.3.2. Hasil

Dari hasil penelitian yang telah dilakukan, metode klasifikasi dengan algoritma random forest dalam memprediksi tingkat kesejahteraan masyarakat nelayan di Lombok Timur NTB, menggunakan 10 kali pengujian dengan k-fold cross validation mendapatkan nilai akurasi sebesar 93,37% dari k- 5. Sedangkan untuk nilai AUC diperoleh 0,735%. Dari nilai akurasi yang diperoleh dapat disimpulkan untuk algoritma random forest sangat akurat digunakan untuk menganalisis serta memprediksi tingkat kesejahteraan masyarakat nelayan di Lombok Timur Nusa Tenggara Barat.

4. KESIMPULAN

Peneliti pertama menggunakan Algoritma *Linear Regression* untuk memprediksi tingkat kemiskinan di Jawa Barat berdasarkan data dari 2002 hingga 2022. Hasil pengujian menggunakan RapidMiner menunjukkan nilai RMSE sebesar 65,837, MAE sebesar 53,156, dan MARE sebesar 66,47%. Model ini mampu memberikan prediksi yang cukup akurat untuk mendukung perencanaan kebijakan pemerintah dalam penanggulangan kemiskinan.

Peneliti kedua menggunakan Algoritma *K-Nearest Neighbor* (KNN) untuk memprediksi masyarakat kurang mampu di Kecamatan Pegasing. Berdasarkan lima atribut utama (jenis pekerjaan, jumlah tanggungan, umur, penghasilan, dan kondisi rumah), model menghasilkan akurasi sebesar 86,02%, recall 61,90%, precision 72,22%, dan F1-score 66,04%. Penelitian ini menunjukkan bahwa KNN dapat memberikan prediksi yang cukup baik untuk membantu pemerintah dalam menyalurkan bantuan sosial secara tepat sasaran.

Peneliti ketiga menggunakan Algoritma *Random Forest* untuk memprediksi kesejahteraan masyarakat nelayan di Lombok Timur. Dari 1.855 data yang dianalisis, hasil pengujian dengan *K-Fold Cross Validation* menunjukkan nilai akurasi tertinggi sebesar 93,37% dan AUC sebesar 0,735. Algoritma ini dinilai sangat efektif dalam mengklasifikasikan data kesejahteraan masyarakat nelayan ke dalam kategori sejahtera dan tidak sejahtera.

Karakteristik data menentukan metode yang kita gunakan seperti *Linear Regression* untuk memodelkan hubungan linier antara satu atau lebih variabel independen dan variabel dependen. Sedangkan untuk model non-linear bisa digunakan salah satu dari metode seperti *K-Nearest Neighbor* dan *Random Forest* bergantung pada karakteristik data. Untuk data dengan pola sederhana dan linear, *Linear Regression* memberikan hasil yang efisien. Namun, untuk data dengan pola yang lebih kompleks dan non-linear, metode *Random Forest* mendapatkan tingkat akurasi yang lebih tinggi dari metode *K-Nearest Neighbor* (KNN).

Dengan demikian masing-masing metode memiliki keunggulan dan kelemahan tergantung dari jenis dan karakteristik data yang diolah untuk mendapatkan fleksibilitas yang lebih tinggi dalam menangkap hubungan variabel yang ada.

5. REFERENSI

Agustiya, K., Wulandary, D., Nufus, N. F. B., & Hasanah, H. (2024). Kontribusi Dinas Sosial dalam Upaya Pengentasan Kemiskinan di Kabupaten Jember. *Jurnal Pengabdian Mandiri*, *3*(2), 193–200. https://bajangjournal.com/index.php/JPM/article/view/7478

Duwo Jiwo Saputro, A., Darmawan, A., & Nurina Sari, B. (2024). Klasifikasi Persentase Kemiskinan Di Jawa Barat Menggunakan Data Mining Algoritma K-Nearest Neighbor (Knn). *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(4), 2718–2723. https://doi.org/10.36040/jati.v7i4.7178

Faisal, M., Utami, W. S., & Parmica, S. (2023). Implementasi Algoritma K-Nearest Neighbor (KNN) Dalam Memprediksi Indeks Kemiskinan. *Journal Sensi*, *9*(1), 11–23. https://doi.org/10.33050/sensi.v9i1.2616

Irawati, M., & Pakereng, M. A. I. (2023). Analisis Pengaruh Jumlah Pengangguran Terhadap Jumlah Kemiskinan Menggunakan Metode Regresi Linier (Studi Kasus: Kota Salatiga). *Jurnal EMT KITA*, *7*(2), 401–408. https://doi.org/10.35870/emt.v7i2.1013

Laelatul Azizah, N., Suarna, N., & Prihartono, W. (2024). Prediksi Tingkat Kemiskinan Di Provinsi Jawa Barat Menggunakan Algoritma Regresi Linear. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(6), 3377–3381. https://doi.org/10.36040/jati.v7i6.8201

Novianty, D., Palasara, N. D., & Qomaruddin, M. (2021). Algoritma Regresi Linear pada Prediksi Permohonan Paten yang Terdaftar di Indonesia. *Jurnal Sistem Dan Teknologi Informasi (Justin)*, *9*(2), 81. https://doi.org/10.26418/justin.v9i2.43664

Nurdin, N. (2024). Analisa Data Mining Dalam Memprediksi Masyarakat Kurang Mampu Menggunakan Metode K-Nearest Neighbor. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(2), 1090–1098. https://doi.org/10.23960/jitet.v12i2.4131

Pustaka, T. (2024). *Prediksi jumlah pengangguran di jawa barat dengan menggunakan algoritma regresi liniear*. *8*(5), 10170–10176.

Riandari, F., Sihotang, H. T., & Husain, H. (2022). Forecasting the Number of Students in Multiple Linear Regressions. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *21*(2), 249–256. https://doi.org/10.30812/matrik.v21i2.1348

Sandi, A., Kusrini, K., & Kusnawi, K. (2023). Analisa Prediksi Kesejahteraan Masyarakat Nelayan Lombok Timur Menggunakan Algoritma Random Forest. *Infotek : Jurnal Informatika Dan Teknologi*, *6*(2), 238–248. https://doi.org/10.29408/jit.v6i2.10104

Sazwati, A., Pratama, D., Anam, K., Wahyudin, E., & Rifa'i, A. (2024). Penerapan Data Mining Untuk Mengestimasi Persentase Penduduk Miskin Di Jawa Barat Menggunakan Regresi Linier Berganda. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *8*(1), 1103–1108. https://doi.org/10.36040/jati.v8i1.8214

Yin, X., Zuo, Y., & Fu, G. (2024). Design of intelligent detection method for electricity transmission line equipment defect based on data mining algorithm. *International Journal of Thermofluids*, *24*(August). https://doi.org/10.1016/j.ijft.2024.100814