

Implementasi Algoritma K-Nearest Neighbor (KNN) Dalam Memprediksi Indeks Kemiskinan

Muhammad Faisal*¹, Wiranti Sri Utami², Sarah Parmica³

¹Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Raharja
Tangerang

²Program Studi Teknik Informatika, Fakultas Teknik, Universitas Cendekia Abditama,
Tangerang

³Program Studi Komputer Akutansi, Fakultas Sains dan Teknologi, Universitas Raharja,
Tangerang

e-mail: *muhammad.faisal@raharja.info, wirantisutami@uca.ac.id,
sarah.parmica@raharja.info

Abstrak

Kemisikinan merupakan masalah yang harus dihadapi oleh Pemerintah, kemiskinan juga dapat berpengaruh terhadap tindak kriminal. Oleh sebab itu diperlukan perhatian khusus untuk menekan angka kemiskinan, Pemerintah telah melakukan upaya dalam menekan angka kemiskinan, diantaranya dengan memberikan berbagai macam bantuan kepada rakyat miskin berdasarkan data yang diperoleh. Selain itu Pemerintah juga perlu memperhatikan indeks kemiskinan pada setiap provinsi, hal ini bertujuan untuk mengetahui informasi indeks kemiskinan dalam waktu tertentu. Penelitian ini membahas tentang prediksi indeks kemiskinan menggunakan metode K-Nearest Neighbor (KNN) dalam memprediksi indeks kemiskinan disetiap Provinsi dengan menggunakan data yang diperoleh dari Badan Pusat Statistik (BPS).

Kata kunci—Data mining, Indeks, K-Nearest Neighbor (KNN)

Abstract

Poverty is a problem that must be faced by the Government, poverty can also affect crime. Therefore, special attention is needed to reduce the poverty rate, the Government has made efforts to reduce the poverty rate, including by providing various kinds of assistance to the poor based on the data obtained. In addition, the government also needs to pay attention to the poverty index in each province, this aims to find out information on the poverty index in a certain time. This study discusses the prediction of the poverty index using the K-Nearest Neighbor (KNN) method in predicting the poverty index in each province using data obtained from the Central Statistics Agency (BPS).

Keywords—Data Mining, Index, K-Nearest Neighbor

1. PENDAHULUAN

Indeks kedalaman kemiskinan biasa juga disebut dengan istilah Poverty Gap Index-P1 merupakan ukuran rata-rata kesenjangan pengeluaran masing-masing penduduk miskin terhadap garis miskin. Semakin tinggi nilai indeks maka semakin jauh rata-rata pengeluaran penduduk dari garis kemiskinan. Indeks kemiskinan merupakan suatu masalah ketimpangan sosial yang menjadi perhatian khusus bagi pemerintah, salah satu cara untuk menanggulangi masalah tersebut adalah memberikan subsidi bagi rakyat yang terdampak dalam kategori miskin baik berupa kebutuhan makanan, pendidikan, kesehatan, dll. Maka dari itu diperlukan suatu metode penelitian yang dapat memprediksi perkembangan indeks kemiskinan di setiap Provinsi.

Provinsi di Negara Kesatuan Republik Indonesia (NKRI) terdiri dari beberapa pulau, tentunya terdapat ketimpangan social pada setiap Provinsi, hal tersebut dapat diamati dari data yang didapatkan dari Badan Pusat Statistik (BPS). Berdasarkan data yang didapatkan tersebut penulis mencoba melakukan penelitian untuk memprediksi indeks kemiskinan ditahun berikutnya dengan menggunakan data mining dan metode K-Nearest Neighbor (KNN). Prediksi dalam menentukan indeks kemiskinan dapat membantu pemerintah dalam upaya memberikan bantuan yang tepat sasaran, Oleh sebab itu hal ini akan memberikan informasi kepada pemerintah dalam memberikan bantuan baik berupa bantuan Pendidikan, Sembako, Bantuan Langsung Tunai (BLT), Bantuan Kesehatan, dll.

Penelitian yang dilakukan ini dengan menerapkan data mining menggunakan metode KNN diharapkan dapat 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 pendekatan 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). Algoritma KNN merupakan algoritma yang digunakan dalam dunia Machine Learning untuk proses Klasifikasi.

Penelitian ini akan fokus terhadap prediksi indeks kemiskinan di setiap provinsi, dengan melakukan pengamatan terhadap data yang diperoleh dari Badan Pusat Statistik (BPS), dan penerapan data mining menggunakan metode K-Nearest Neighbor (KNN) pendekatan lainnya untuk mendukung proses penelitian ini dengan menggunakan metode kuantitatif.

2. METODE PENELITIAN

Penelitian ini menggunakan data yang diperoleh dari Badan Pusat Statistik (BPS) data tersebut merupakan indeks kemiskinan pada setiap Provinsi di Indonesia. Data tersebut dapat dilihat pada Tabel 1, dibawah ini.

Table 1. Indeks Kedalaman Kemiskinan (P1)

Provinsi	Indeks Kedalaman Kemiskinan (P1) Menurut Provinsi dan Daerah (Persen)																	
	Perkotaan						Perdesaan						Perkotaan+Perdesaan					
	2020			2021			2020			2021			2020			2021		
	Se ms 1	Se ms 2	T h n	Se ms 1	Se ms 2	T h n	Se ms 1	Se ms 2	T h n	Se ms 1	Se ms 2	T h n	Se ms 1	Se ms 2	T h n	Se ms 1	Se ms 2	T h n

Aceh	1,55	1,61	-	1,79	-	-	3,28	2,46	-	3,40	-	-	2,72	2,85	-	2,86	-	-
SUMUT	1,48	1,54	-	1,50	-	-	1,55	1,67	-	1,54	-	-	0,92	0,99	-	1,04	-	-
Riau	0,97	1,02	-	1,01	-	-	1,25	1,53	-	1,09	-	-	1,14	1,32	-	1,06	-	-
Jambi	1,71	1,77	-	2,20	-	-	0,80	0,89	-	0,85	-	-	1,10	1,18	-	1,29	-	-
SUMSEL	2,02	2,38	-	2,19	-	-	2,12	2,19	-	2,30	-	-	2,08	2,26	-	2,26	-	-
...
Papua	0,62	0,82	-	0,85	-	-	8,37	9,34	-	7,52	-	-	6,16	6,90	-	5,60	-	-

Menurut Provinsi dan Daerah (Source BPS)

2.1 Literatur Review

Untuk mendukung penelitian ini maka penulis melakukan studi literature review tentang penulisan sebelumnya yang berkaitan dengan penulisan ini, Berikut ini literature review yang berkaitan dengan penelitian ini:

1. Penelitian yang dilakukan oleh Sekar, dkk, dengan judul “Penerapan Algoritma K-Nearest Neighbor untuk Prediksi Kelulusan Siswa”. Penelitian ini membahas mengenai kelulusan siswa yang berpengaruh terhadap penilaian akreditasi terhadap Sekolah SMK Anak Bangsa. Oleh sebab itu dilakukan penelitian untuk memprediksi tingkat kelulusan Siswa menggunakan penerapan K-NN [1].
2. Selanjutnya penelitian yang dilakukan oleh Muhammad Naja, dkk, dengan judul “Penerapan Algoritma K-Nearest Neighbor (KNN) Untuk Memprediksi Penyakit Stroke”. Penelitian ini membahas mengenai gangguan fungsi otak terhadap aliran darah ke otak yang mengakibatkan timbulnya penyakit stroke. Penyakit ini merupakan penyakit mematikan nomor 3 didunia, oleh sebab itu dilakukan penelitian menggunakan metode K-Nearest Neighbor dalam memprediksi penyakit stroke [2].
3. Kemudian penelitian yang dilakukan oleh Muhammad Yusuf, dkk, dengan judul “Penerapan Algoritma K-Nearest Neighbor (KNN) dalam memprediksi dan Menghitung Tingkat Akurasi Cuaca di Indonesia”. Penelitian ini membahas mengenai negara Indonesia merupakan negara yang rawan terhadap terjadinya bencana. Terdapat berbagai macam bencana yang sering terjadi di Indonesia antara lain banjir, longsor, gempa, dll. Oleh sebab itu dilakukan penelitian menggunakan metode KNN untuk memprediksi dan menghitung tingkat akurasi data dan cuaca di Indonesia [3].
4. Selanjutnya penelitian yang dilakukan oleh Bagus, dan Fahrur, dengan judul “Prediksi Penjualan Sepatu Menggunakan Metode K-Nearest Neighbor”. Penelitian ini membahas tentang obral sepatu murah dalam mengklasifikasi potensi pelanggan baru, maka dilakukan penelitian menggunakan metode KNN dalam memprediksi pelanggan baru [4].
5. Kemudian penelitian yang dilakukan Abdul Ghani, dkk, dengan judul “Penerapan Algoritma K-Nearest Neighbor”. Penelitian ini membahas tentang proses penjualan yang sangat penting pada sebuah perusahaan agar perusahaan tersebut tetap dapat beroperasi. Berdasarkan hal tersebut penelitian ini membuat suatu program untuk memprediksi penjualan berbasis web dengan menggunakan metode KNN [5].

2. 1.1 Data Mining

Data mining adalah proses penggalian informasi dan pola data yang sangat besar [6]. Data mining merupakan kumpulan data hasil dari proses ekstrak data, analisa data, dan statistika data. Data mining adalah rangkaian data yang dapat diolah dan diproses dengan cara menggali atau menambang suatu serangkaian data yang mengandung informasi untuk menghasilkan suatu pengetahuan yang diambil melalui proses manual.

2. 1.2 Prediksi

Prediksi merupakan proses untuk memperkecil kesalahan, sehingga dapat meminimalisir antara kejadian yang sebenarnya [7]. Prediksi dapat menentukan kebutuhan dalam kurun waktu minggu, bulan, bahkan tahun hal ini berkaitan dengan dukungan jumlah data atau prediksi waktu yang dianalisa sehingga dapat diperhitungkan dalam memprediksi data pada tahun mendatang.

2. 2 K-Nearest Neighbor (KNN)

K-Nearest Neighbor (KNN) merupakan salah satu dari algoritma klasifikasi data mining, algoritma ini digunakan untuk mencari atau mendapatkan prediksi dimasa yang akan datang [8]. KNN merupakan algoritma dari data mining yang paling sederhana. Tujuan dari algoritma ini untuk mencari hasil dari klasifikasi jarak terdekat pada setiap objek. Dalam menentukan jarak tersebut awal mula data dibagi menjadi dua bagian yaitu data training dan data testing, setelah mendapat data training dan data testing selanjutnya dihitung jarak dari tiap-tiap data testing (Euclidean Distance) terhadap data training. Rumus dari KNN untuk menghitung antar jarak data testing (Euclidean Distance) dapat dilihat pada keterangan dibawah ini:

$$d_1(x,y) = \sum_{i=1}^N |x_i - y_i|$$

2. 3 Mean

Untuk membantu proses penelitian ini dalam mengolah data peneliti menggunakan nilai rata-rata atau Mean. Mean merupakan nilai rata-rata yang mewakili himpunan atau kelompok dari suatu data. Mean didapatkan dengan menjumlahkan seluruh data individu yang ada atau yang berada didalam suatu kelompok. Proses atau cara mencari nilai Mean dapat dilihat pada rumus dibawah ini.

$$\chi = \frac{\sum X_1}{n} = \frac{1}{n}(X_1 + X_2 + \dots + X_n)$$

3. HASIL DAN PEMBAHASAN

Data yang digunakan pada penelitian ini merupakan data indeks kemiskinan pada tahun 2020 yang diperoleh dari BPS. Data tersebut merupakan indeks kemiskinan di setiap Provinsi di Indonesia, total Provinsi sebanyak 38 Provinsi. Sebelum data yang akan di teliti masuk kedalam proses penerapan data mining, data tersebut terlebih dahulu dibersihkan, gunanya untuk memfilter data yang akan digunakan dalam proses penerapan data mining. Data tersebut dapat dilihat pada Tabel 2, dibawah ini.

Table 2. Indeks Kedalaman Kemiskinan yang telah melalui proses pemberishan data

Provinsi	Indeks Kedalaman Kemiskinan (P1) Menurut Provinsi dan Daerah (Persen)
----------	---

	2020		2020		2020	
	Sems 1 (Maret)	Sems 2 (Sept)	Sems 1 (Maret)	Sems 2 (Sept)	Sems 1 (Maret)	Sems 2 (Sept)
Aceh	1,55	1,61	3,28	3,46	2,72	2,85
Sumatera Utara	1,48	1,54	1,55	1,67	1,51	1,6
Sumatera Barat	0,74	0,8	1,07	1,17	0,92	0,99
Riau	0,97	1,02	1,25	1,53	1,14	1,32
Jambi	1,71	1,77	0,8	0,89	1,1	1,18
Sumatera Selatan	2,02	2,38	2,12	2,19	2,08	2,26
Bengkulu	2,44	2,49	2,38	2,52	2,4	2,51
Lampung	1,22	1,3	2,2	2,48	1,9	2,11
Kep. Bangka Belitung	0,44	0,64	0,79	0,94	0,6	0,77
Kep. Riau	1,02	1,16	1,84	1,89	1,11	1,21
.....
Papua	0,62	0,82	8,37	9,34	6,16	6,9

Selanjutnya data dilakukan analisa yaitu dengan menetapkan skala normalisasi. Skala normalisasi didapatkan dari proses perhitungan nilai Mean atau nilai rata-rata. Langkah tersebut menetapkan nilai rata-rata indeks kemiskinan dari masing-masing provinsi. Nilai tersebut dapat dilihat pada Tabel 3, dibawah ini.

Table 3. Nilai skala normalisasi

Rendah	1 s/d 3
Sedang	4 s/d 6
Tinggi	7 s/d 10

Skala normalisasi diatas menggunakan proses perhitungan Mean, dimana data indeks kemiskinan akan dicari nilai rata-rata dan hasil nilai akhir akan didapat skala yang telah ditetapkan seperti gambar diatas. Misalnya Provinsi Aceh memiliki nilai rata-rata 3,443 Nilai tersebut dibulatkan menjadi 3 sesuai dengan Skala normalisasi sehingga Provinsi Aceh masuk kedalam kategori Rendah. Adapun proses perhitungannya adalah :

$$\begin{aligned} \text{Aceh} &= \frac{(1,55+1,61)+(3,28+3,46)+(2,72+2,85)}{6} = 3,433 \\ \text{SUMUT} &= \frac{(1,48+1,54)+(1,55+1,67)+(1,51+1,6)}{6} = 2,065 \\ \text{SUMBAR} &= \frac{(0,74+0,8)+(1,07+1,17)+(0,92+0,99)}{6} = 1,295 \\ \text{Riau} &= \frac{(0,97+1,02)+(1,25+1,53)+(1,14+1,32)}{6} = 1,555 \\ &\dots\dots\dots \\ \text{Papua} &= \frac{(0,62+0,82)+(8,37+9,34)+(6,16+6,9)}{6} = 6,763 \end{aligned}$$

Langkah berikutnya mengikuti proses perhitungan seperti langkah diatas. untuk mencari nilai rata-rata pada masing-masing Provinsi. Sehingga hasil akhir dapat dilihat pada Table 4, dibawah ini.

Table 4. Nilai Rata-rata

Provinsi	Indeks Kedalaman Kemiskinan (P1) Menurut Provinsi dan Daerah (Persen)						Rata-rata	Ket
	Perkotaan		Perdesaan		Perkotaan+Perdesaan			
	Semester 1 (Mar)	Semester 2 (Sept)	Semester 1 (Mar)	Semester 2 (Sept)	Semester 1 (Mar)	Semester 2 (Sept)		
Aceh	1,55	1,61	3,28	3,46	2,72	2,85	3	Rendah
Sumatera Utara	1,48	1,54	1,55	1,67	1,51	1,6	2	Rendah
Sumatera Barat	0,74	0,8	1,07	1,17	0,92	0,99	1	Rendah
Riau	0,97	1,02	1,25	1,53	1,14	1,32	2	Rendah
Jambi	1,71	1,77	0,8	0,89	1,1	1,18	2	Rendah
Sumatera Selatan	2,02	2,38	2,12	2,19	2,08	2,26	3	Rendah
Bengkulu	2,44	2,49	2,38	2,52	2,4	2,51	3	Rendah
Lampung	1,22	1,3	2,2	2,48	1,9	2,11	2	Rendah
Kep. Bangka Belitung	0,44	0,64	0,79	0,94	0,6	0,77	1	Rendah
Kep. Riau	1,02	1,16	1,84	1,89	1,11	1,21	2	Rendah
.....
Papua	0,62	0,82	8,37	9,34	6,16	6,9	7	Tinggi

Setelah mendapatkan nilai rata-rata proses selanjutnya adalah menentukan klasifikasi menggunakan algoritma K- Nearest Neighbor (KNN) dengan melakukan normalisasi data training dan data uji. Proses untuk mencari nilai normalisasi data training terlebih dahulu menentukan nilai maksimal dan nilai minimal dari table data indeks Provinsi kemiskinan. Nilai tersebut dapat dilihat pada Table 5, dibawah ini.

Table 5. Nilai Normalisasi Maksimal dan Minimal

Provinsi	Perkotaan	Perdesaan	Perkotaan+Perdesaan
	2020	2020	2020

	Semester 1 (Maret)	Semester 2 (Sept)	Semester 1 (Maret)	Semester 2 (Sept)	Semester 1 (Maret)	Semester 2 (Sept)
MAX	2,54	2,85	9,29	2,74	6,16	6,9
MIN	0,38	0,43	0	0	0,52	0,61

Setelah mendapatkan nilai Maksimal dan Minimal langkah selanjutnya adalah melakukan perhitungan nilai normalisasi data training. Proses tersebut dapat dilihat pada perhitungan nilai dibawah ini.

$$\begin{aligned} \text{Aceh (S1 / Perkotaan)} &= (1,55-0,38) : (2,54-0,38) = 0,542 \\ \text{Aceh (S2 / Perkotaan)} &= (1,61-0,43) : (2,85-0,43) = 0,488 \\ \text{Aceh (S1 / Perdesaan)} &= (3,28-0) : (9,29-0) = 0,353 \\ \text{Aceh (S2 / Perdesaan)} &= (3,46-0) : (9,74-0) = 0,355 \\ \text{Aceh (S1 / Kota \& Desa)} &= (2,72-0,52) : (6,16-0,52) = 0,390 \\ \text{Aceh (S2 / Kota \& Desa)} &= (2,85-0,61) : (6,9-0,61) = 0,356 \end{aligned}$$

Langkah berikutnya menentukan masing-masing nilai normalisasi data training pada setiap Provinsi. Nilai normalisasi data training dapat dilihat pada Table 6, dibawah ini.

Table 6. Nilai Normalisasi Data Training

Normalisasi Data Training	Semester 1 (Maret)	Semester 2 (Sept)	Semester 1 (Maret)	Semester 2 (Sept)	Semester 1 (Maret)	Semester 2 (Sept)
A1	0,542	0,488	0,353	0,355	0,390	0,356
A2	0,509	0,459	0,167	0,171	0,176	0,157
A3	0,167	0,153	0,115	0,120	0,071	0,060
A4	0,273	0,244	0,135	0,157	0,110	0,113
A5	0,616	0,554	0,086	0,091	0,103	0,091
A6	0,759	0,806	0,228	0,225	0,277	0,262
A7	0,954	0,851	0,256	0,259	0,333	0,302
A8	0,389	0,360	0,237	0,255	0,245	0,238
A9	0,028	0,087	0,085	0,097	0,014	0,025
A10	0,296	0,302	0,198	0,194	0,105	0,095
A11	0,097	0,099	0	0	0,012	0,010
A12	0,296	0,310	0,160	0,167	0,108	0,107
...
A34	0,111	0,161	0,901	0,959	1	1

Setelah mendapatkan nilai normalisasi data training dari masing-masing Provinsi proses selanjutnya menetapkan data uji, data tersebut diambil secara acak. Penulis telah menetapkan 3 (tiga) data yang akan di uji. Data tersebut dapat dilihat pada Table 7, dibawah ini.

Table 7. Data Uji

Provinsi	Perkotaan (2020)		Perdesaan (2020)		Perkotaan+Perdesaan (2020)	
	Semester 1 (Maret)	Semester 2 (Sept)	Semester 1 (Maret)	Semester 2 (Sept)	Semester 1 (Maret)	Semester 2 (Sept)
Jawa Barat	1,02	1,18	1,49	1,63	1,13	1,28
Maluku	0,82	0,94	5,55	5,95	3,47	3,76
Papua	0,62	0,82	8,37	9,34	6,16	6,9

Setelah menetapkan data uji kemudian mencari nilai maksimal dan minimal dari data yang akan diuji. Nilai data maksimal dan minimal dapat dilihat pada Table 8, dibawah ini.

Table 8. Nilai Maksimal dan Minimal

Provinsi	Perkotaan (2020)		Perdesaan (2020)		Perkotaan+Perdesaan (2020)	
	Semester 1 (Maret)	Semester 2 (Sept)	Semester 1 (Maret)	Semester 2 (Sept)	Semester 1 (Maret)	Semester 2 (Sept)
MAX	1,02	1,18	8,37	9,34	6,16	6,9
MIN	0,62	0,82	1,49	1,63	1,13	1,28

Setelah mendapatkan nilai maksimal dan minimal proses selanjutnya melakukan perhitungan normalisasi nilai uji. Proses tersebut dapat dilihat pada perhitungan nilai dibawah ini.

$$\begin{aligned} \text{Jabar (S1 / Perkotaan)} &= (1,02-0,62) : (1,02-0,62) = 1 \\ \text{Jabar (S2 / Perkotaan)} &= (1,18-0,82) : (1,18-0,82) = 1 \\ \text{Jabar (S1 / Perdesaan)} &= (1,49-1,49) : (8,37-1,49) = 0 \\ \text{Jabar (S2 / Perdesaan)} &= (1,63-1,63) : (9,34-1,63) = 0 \\ \text{Jabar (S1 / Kota&Desa)} &= (1,13-1,13) : (6,16-1,13) = 0,062 \\ \text{Jabar (S2 / Kota&Desa)} &= (1,28-1,28) : (6,9-1,28) = 0,082 \end{aligned}$$

$$\begin{aligned} \text{Maluku (S1 / Perkotaan)} &= (0,82-0,62) : (1,02-0,62) = 0,500 \\ \text{Maluku (S2 / Perkotaan)} &= (0,94-0,82) : (1,18-0,82) = 0,333 \\ \text{Maluku (S1 / Perdesaan)} &= (5,55-1,49) : (8,37-1,49) = 1 \\ \text{Maluku (S2 / Perdesaan)} &= (5,95-1,63) : (9,34-1,63) = 1 \\ \text{Maluku (S1 / Kota&Desa)} &= (3,47-1,13) : (6,16-1,13) = 0 \\ \text{Maluku (S2 / Kota&Desa)} &= (3,76-1,28) : (6,9-1,28) = 0 \end{aligned}$$

$$\begin{aligned} \text{Papua (S1 / Perkotaan)} &= (0,62-0,62) : (1,02-0,62) = 0 \\ \text{Papua (S2 / Perkotaan)} &= (0,82-0,82) : (1,18-0,82) = 0 \\ \text{Papua (S1 / Perdesaan)} &= (8,37-1,49) : (8,37-1,49) = 1 \\ \text{Papua (S2 / Perdesaan)} &= (9,34-1,63) : (9,34-1,63) = 1 \\ \text{Papua (S1 / Kota\&Desa)} &= (6,16-1,13) : (6,16-1,13) = 1 \\ \text{Papua (S2 / Kota\&Desa)} &= (6,9-1,28) : (6,9-1,28) = 1 \end{aligned}$$

Selanjutnya dilakukan proses perhitungan normalisasi data uji pada masing-masing data yang telah ditentukan. Proses tersebut dapat dilihat pada table 9, dibawah ini.

Table 9. Nilai Normalisasi Data Uji

Normalisasi Uji	Semester 1 (Maret)	Semester 2 (Sept)	Semester 1 (Maret)	Semester 2 (Sept)	Semester 1 (Maret)	Semester 2 (Sept)
JABAR	1	1	0	0	0,062	0,082
MALUKU	0,500	0,333	1	1	0	0
PAPUA	0	0	1	1	1	1

Setelah mendapatkan nilai normalisasi data uji proses selanjutnya adalah melakukan proses perhitungan Eulidean Distance pada masing-masing Provinsi, dimana setiap Provinsi dilambangkan dengan Variable, Variable tersebut berjumlah 34 Variable, sesuai dengan jumlah Provinsi yang di teliti berdasarkan data yang di peroleh dari BPS. Proses perhitungan nilai Euclidean Distance sebagai berikut.

$$A1 \text{ (JABAR)} = (0,542-1)^2+(0,488-1)^2+(0,353-0)^2+(0,355-0)^2+(0,390-0,062)^2+(0,356-0,082)^2 = 0,952$$

$$A1 \text{ (Maluku)} = (0,542-0,500)^2+(0,488-0,333)^2+(0,353-1)^2+(0,355-1)^2+(0,390-0)^2+(0,356-0)^2 = 0,370$$

$$A1 \text{ (Papua)} = (0,542-0)^2+(0,488-0)^2+(0,353-1)^2+(0,355-1)^2+(0,390-1)^2+(0,356-1)^2 = 1,467$$

$$A2 \text{ (JABAR)} = (0,509-1)^2+(0,459-1)^2+(0,167-0)^2+(0,171-0)^2+(0,176-0,062)^2+(0,157-0,082)^2 = 0,781$$

$$A2 \text{ (Maluku)} = (0,509-0,500)^2+(0,459-0,333)^2+(0,167-1)^2+(0,171-1)^2+(0,176-0)^2+(0,157-0)^2 = 0,715$$

$$A2 \text{ (Papua)} = (0,509-0)^2+(0,459-0)^2+(0,167-1)^2+(0,171-1)^2+(0,176-1)^2+(0,157-1)^2 = 1,800$$

$$A3 \text{ (JABAR)} = (0,167-1)^2+(0,153-1)^2+(0,115-0)^2+(0,120-0)^2+(0,071-0,062)^2+(0,060-0,082)^2 = 1,200$$

$$A3 \text{ (Maluku)} = (0,167-0,500)^2+(0,153-0,333)^2+(0,115-1)^2+(0,120-1)^2+(0,071-0)^2+(0,060-0)^2 = 0,929$$

$$A3 \text{ (Papua)} = (0,167-0)^2+(0,153-0)^2+(0,115-1)^2+(0,120-1)^2+(0,071-1)^2+(0,060-1)^2 = 1,831$$

$$A4 \text{ (JABAR)} = (0,273-1)^2+(0,244-1)^2+(0,135-0)^2+(0,157-0)^2+(0,110-0,062)^2+(0,113-0,082)^2 = 1,071$$

$$A4 \text{ (Maluku)} = (0,273-0,500)^2+(0,244-0,333)^2+(0,135-1)^2+(0,157-1)^2+(0,110-0)^2+(0,113-0)^2 = 0,815$$

$$A4 \text{ (Papua)} = (0,273-0)^2+(0,244-0)^2+(0,135-1)^2+(0,157-1)^2+(0,110-1)^2+(0,113-1)^2 = 1,781$$

$$A5 \text{ (JABAR)} = (0,616-1)^2+(0,554-1)^2+(0,086-0)^2+(0,091-0)^2+(0,103-0,062)^2+(0,091-0,082)^2 = 0,604$$

$$A5 \text{ (Maluku)} = (0,616-0,500)^2+(0,554-0,333)^2+(0,086-1)^2+(0,091-1)^2+(0,103-0)^2+(0,091-0)^2 = 0,889$$

$$A5 \text{ (Papua)} = (0,616-0)^2+(0,554-0)^2+(0,086-1)^2+(0,091-1)^2+(0,103-1)^2+(0,091-1)^2 = 1,995$$

.....

.....

$$A34 \text{ (JABAR)} = (0,111-1)^2+(0,161-1)^2+(0,901-0)^2+(0,959-0)^2+(1-0,062)^2+(1-0,082)^2 = 2,225$$

$$A34 \text{ (Maluku)} = (0,111-0,500)^2+(0,161-0,333)^2+(0,901-1)^2+(0,959-1)^2+(1-0)^2+(1-0)^2 = 1,017$$

$$A34 \text{ (Papua)} = (0,111-0)^2+(0,161-0)^2+(0,901-1)^2+(0,959-1)^2+(1-1)^2+(1-1)^2 = 0,223$$

Hasil dari proses perhitungan nilai Euclidean Distance dapat dilihat pada table 10, dibawah ini.

Tabel 10. Euclidean Distance

Normalisasi Uji	Jawa Barat	Maluku	Papua
A1	0,952	0,370	1,467
A2	0,781	0,715	1,800
A3	1,200	0,929	1,831
A4	1,071	0,815	1,781
A5	0,604	0,889	1,995
A6	0,526	0,776	1,868
A7	0,528	0,845	1,915
A8	0,981	0,567	1,603
A9	1,342	1,063	1,893
A10	1,030	0,762	1,758
A11	1,278	1,127	1,994
A12	1,014	0,788	1,782
A13	0,778	0,646	1,743
A14	0,633	0,668	1,786
A15	0,968	0,551	1,603
A16	1,124	0,851	1,792
A17	1,358	1,098	1,929
A18	0,557	0,946	1,972
A19	1,381	0,246	1,024
A20	1,196	0,900	1,802
A21	1,137	0,957	1,889

A22	1,233	1,014	1,901
A23	1,277	0,881	1,748
A24	1,310	0,977	1,827
A25	1,171	0,836	1,752
A26	0,841	0,466	1,563
A27	1,255	0,740	1,607
A28	0,871	0,547	1,621
A29	1,619	0,626	1,160
A30	0,577	0,728	1,846
A31	1,541	0,335	0,936
A32	1,363	0,998	1,801
A33	2,095	0,901	0,436
A34	2,225	1,017	0,223

Setelah mendapatkan nilai Euclidean Distance selanjutnya menentukan nilai Normalisasi K, nilai tersebut dapat dilihat pada table 11, dibawah ini.

Tabel 11. Normalisasi Nilai K

Normalisasi Uji	Jawa Barat	Maluku	Papua
A1		Rendah	
A2			
A3			
A4			
A5	Rendah		
A6	Rendah		
A7	Rendah		
A8			
A9			
A10			
A11			
A12			
A13			
A14			
A15			
A16			
A17			
A18	Rendah		
A19		Tinggi	Tinggi
A20			
A21			
A22			
A23			
A24			
A25			
A26		Rendah	
A27			
A28		Rendah	

A29			Rendah
A30	Rendah		
A31		Tinggi	Tinggi
A32			
A33			Tinggi
A34			Tinggi

Tabel 12. Tabel Prediksi

Data Uji	Keterangan Asli	Keterangan Prediksi	Rendah	Tinggi
JABAR	Rendah	Rendah	5	0
MALUKU	Tinggi	Rendah	3	2
PAPUA	Tinggi	Tinggi	1	4

4. KESIMPULAN

Dari hasil penelitian yang dilakukan diatas pemodelan dengan penerapan data mining menggunakan metode K-Nearest Neighbor (KNN) berdasarkan data yang didapatkan dari Badan Pusat Statistik (BPS). Maka dapat dilihat kategori dengan Provinsi Jawa Barat memiliki nilai K=5 dengan kategori rendah, sedangkan Maluku memiliki nilai K=3 dengan kategori rendah dan K=2 dengan kategori tinggi, dan Papu memiliki nilai K=1 dengan kategori rendah dan K=4 dengan kategori tinggi.

5. SARAN

Dari hasil penelitian yang sudah dibahas ini penulis menyarankan agar diadakan penelitian lebih lanjut guna membantu program pemerintah dalam menekan angka kemiskinan dimasing-masing Provinsi. Terdapat banyak sekali metode yang dapat menyelesaikan masalah ini salah satu contoh adalah dengan menggunakan Metode *Decision Support System* (DSS)

DAFTAR PUSTAKA

- [1] Sekar, R,R., Sundari, R, A., Dedi, S. 2019. *Penerapan Algoritma K-Nearest Neighbor Untuk Prediksi Kelulusan Siswa*. Prosiding Seminar Nasional Riset Information Science (Senaris), Hal. 670-676.
- [2] Naja, M, M., Harliana, Sukerti, S., Herdian, R,M. 2022.*Penerapan Algoritma K-Nearest Neighbor(KNN) Untuk Memprediksi Penyakit Stroke*. Jurnal Ilmiah Intech, No. 1, Vol. 4, 130-140.
- [3] Yusuf, M, R., Valensyah, M, A., Gunawan, W. 2021. *Penerapan Algoritma K-Nearest Neighbor (KNN) Dalam Memprediksi dan Menghitung Tingkat Akurasi Data Cuaca di Indonesia*. Jurnal Teknik dan Sains, No. 2, Vol. 2, 11-16.
- [4] Hardiyanto, B., Rozi, F. 2020. *Prediksi Penjualan Sepatu Menggunakan Metode K-Nearest Neighbor*. Jurnal JOEICT, No. 02, Vol. 04, 13-18.

- [5] Ghani, A, M., Auliasari, K., Santi, F, W. 2020. *Penerapan Algoritma K-Nearest Neighbor Untuk Prediksi Penjualan Berbasis Web Pada PT. Wika Industry Energy*. Jurnal JATI, No. 2, Vol. 4, 1-6.
- [6] M. Arhami and M. Nasir., 2020, *Data Mining Algoritma dan Implementasi*. Penerbit Andi.
- [7] Erfan Hasmin Hasmin, Nurul Aini.,2020, *Data Mining For Inventory Forecasting Using Double Exponential Smoothing Method*. IEEE Access, Conference Paper. International Conference on Cybernetics and Intelligent System (ICORIS).
- [8] MA. Lusiandro, SM Nasution, C. Setianingsih., 2020, *Implementation of the Advanced Traffic Management System using k-Nearest Neighbor Algorithm*, IEEE Access, Conference Paper. International Conference on Information Technology Systems and Innovation (ICITSI).