

P-ISSN: 2774-4574 ; E-ISSN: 2774-4582
TRILOGI, 2(3), September-Desember 2021 (319-323)
©2021 Lembaga Penerbitan, Penelitian,
dan Pengabdian kepada Masyarakat (LP3M)
Universitas Nurul Jadid Paiton Probolinggo

JURNAL **TRILOGI**
Ilmu Teknologi, Kesehatan, dan Humaniora

IMPLEMENTASI K-NEAREST NEIGHBOR (KNN) UNTUK KLASIFIKASI MASYARAKAT PRA SEJAHTERA DESA SAPIKEREP KECAMATAN SUKAPURA

Ahmad Khairi

Universitas Nurul Jadid, Indonesia

Achmad Fais Ghozali

Universitas Nurul Jadid, Indonesia

Ach Darul Nur Hidayah

Universitas Nurul Jadid, Indonesia

Abstract

Poverty is a very complex general problem that exists in several parts of Indonesia, this can be seen from the low level of education, health and quality of life of the community. Poverty is also one of the obstacles for someone to get a decent life or a prosperous life. The government has an important role in the welfare of its citizens from poverty. The government has made several efforts to alleviate poverty, including the Family Hope Program (PKH), Cash Direct Assistance (BLT) and several other programs. However, in practice, there are still beneficiaries who are not on target, so that there are families with the prosperous category who still receive assistance and disadvantaged families do not receive assistance, as in the village of Sapikerep, Sukapura, Probolinggo. Several criteria to determine the welfare of the community that have been determined by the local government are assets/property, type of floor, type of wall, status of land or building, source of lighting, source of drinking water, income, and source of cooking energy. From the problems above, an appropriate analysis is needed to determine a decision. Therefore, in this study, we will analyze the data obtained with the indicators mentioned above by using a data mining algorithm. The K-Nearest Network (K-NN) algorithm is an algorithm that can be used to predict or classify data depending on the type of data in the existing data set. K-NN performs a classification on a data based on the value of k that has been set previously. The value of k in K-NN must use an odd value if it is used for the different classification process if it is used to predict the value of k on K-NN can be an odd or even number. The k values used to test the performance of K-NN are $k=3$, $k=5$, and $k=7$, where the results of the accuracy of each k value on the K-NN used have different accuracy values for the accuracy of $k=5$ and $k=7$ which has the same accuracy value. The accuracy value of $k=3$ on K-NN is 97.36% while the accuracy value of $k=5$ and $k=7$ on K-NN is 98.68%.

Keywords: Poverty; K-Nearest Neighbor; prosperous

Abstrak

Kemiskinan merupakan masalah umum yang sangat kompleks yang terdapat di beberapa bagian negara Indonesia, hal ini dapat diketahui dari rendahnya pendidikan, kesehatan dan kualitas hidup masyarakat. Kemiskinan juga merupakan salah satu penghambat bagi seseorang untuk memperoleh kehidupan yang layak atau kehidupan yang sejahtera. Pemerintah memiliki peranan penting dalam mensejahterakan warga negaranya dari kemiskinan. Pemerintah telah melakukan beberapa upaya untuk mengentaskan kemiskinan diantaranya dengan adanya Program Keluarga Harapan (PKH), Bantuan Langsung Tunai (BLT) dan beberapa program lainnya. Namun dalam pelaksanaannya masih terdapat penerima bantuan yang tidak tepat sasaran sehingga terdapat keluarga dengan kategori sejahtera masih mendapatkan bantuan dan keluarga tidak sejahtera tidak memperoleh bantuan, seperti yang terdapat di desa Sapikerep Sukapura Probolinggo. Beberapa kriteria untuk menentukan masyarakat sejahtera yang telah ditentukan oleh pemerintah setempat diantaranya yakni aset/harta benda, jenis lantai, jenis dinding, status lahan atau bangunan, sumber penerangan, sumber air minum, penghasilan, serta sumber energi memasak. Dari permasalahan di atas dibutuhkan suatu analisa yang tepat untuk menentukan suatu keputusan. Oleh sebab itu, dalam penelitian ini akan melakukan analisa terhadap data yang diperoleh dengan indikator yang telah disebutkan di atas dengan menggunakan suatu algoritma datamining. Algoritma K-Nearest Network (K-NN) merupakan algoritma yang bisa digunakan untuk melakukan prediksi ataupun klasifikasi terhadap suatu data tergantung dari jenis data pada kumpulan data yang ada. K-NN melakukan klasifikasi pada suatu data berdasarkan nilai k yang telah ditetapkan sebelumnya. Nilai k pada K-NN harus menggunakan nilai ganjil jika digunakan untuk proses klasifikasi beda halnya jika digunakan untuk melakukan prediksi nilai k pada K-NN dapat berupa bilangan ganjil ataupun genap. Nilai k yang digunakan untuk menguji kinerja algoritma K-NN yakni k-3, k-5, dan k-7, dimana hasil akurasi masing-masing nilai k pada K-NN yang digunakan memiliki hasil akurasi yang berbeda terkecuali nilai akurasi k-5 dan k-7 yang memiliki nilai akurasi yang sama. Nilai akurasi k-3 pada K-NN sebesar 97.36% sedangkan nilai akurasi k-5 dan k-7 pada K-NN sebesar 98.68%.

Katakunci: Kemiskinan; K-Nearest Neighbor; sejahtera

1 Pendahuluan

Kesejahteraan sosial merupakan kondisi dimana kebutuhan material, sosial serta spriritual warga negara harus terpenuhi untuk dapat memiliki hidup yang layak serta dapat mengembangkan diri sehingga mampu melaksanakan fungsi sosialnya. Masyarakat yang tidak sejahtera pada umumnya dikategorikan masyarakat kurang mampu hingga dikategorikan miskin sesuai dengan kondisi masing-masing masyarakat.

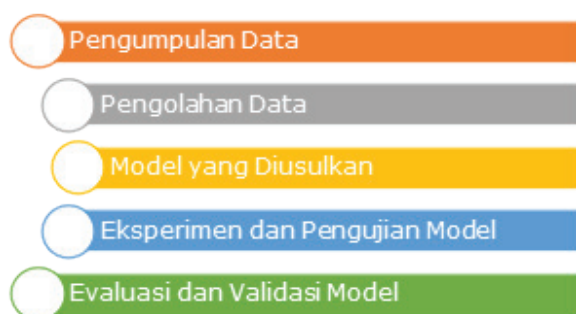
Kemiskinan merupakan masalah umum yang sangat kompleks yang terdapat di beberapa bagian negara Indonesia, hal ini dapat diketahui dari rendahnya pendidikan, kesehatan dan kualitas hidup masyarakat. Kemiskinan juga merupakan salah satu penghambat bagi seseorang untuk memperoleh kehidupan yang layak atau kehidupan yang sejahtera. Pemerintah memiliki peranan penting dalam mensejahterakan warga negaranya dari kemiskinan. Pemerintah telah melakukan beberapa upaya untuk mengentaskan kemiskinan diantaranya dengan adanya Program Keluarga Harapan (PKH), Bantuan Langsung Tunai (BLT) dan beberapa program lainnya. Namun

dalam pelaksanaannya masih terdapat penerima bantuan yang tidak tepat sasaran sehingga terdapat keluarga dengan kategori sejahtera masih mendapatkan bantuan dan keluarga tidak sejahtera tidak memperoleh bantuan, seperti yang terdapat di desa Sapikerep Sukapura Probolinggo yang merupakan salah satu desa yang saat ini masih banyak warganya tergolong kurang mampu dari segi ekonomi hingga rumah yang layak untuk dihuni menjadi perhatian pemerintah desa dalam meningkatkan masyarakat yang sejahtera. Beberapa kriteria untuk menentukan masyarakat sejahtera yang telah ditentukan oleh pemerintah desa diantaranya yakni aset/harta benda, jenis lantai, jenis dinding, status lahan atau bangunan, sumber penerangan, sumber air minum, penghasilan, serta sumber energi memasak. Dari permasalahan di atas dibutuhkan suatu analisa yang tepat untuk menentukan suatu keputusan. Oleh sebab itu, dalam penelitian ini akan melakukan analisa terhadap data yang diperoleh dengan indikator yang telah disebutkan di atas dengan menggunakan suatu algoritma datamining. Algoritma datamining yang digunakan dalam penelitian ini yakni

menggunakan Algoritma K-Nearest Network (K-NN) merupakan algoritma yang bisa digunakan untuk melakukan prediksi ataupun klasifikasi terhadap suatu data tergantung dari jenis data pada kumpulan data yang ada. K-NN melakukan klasifikasi pada suatu data berdasarkan nilai k yang telah ditetapkan sebelumnya. Nilai k pada K-NN harus menggunakan nilai ganjil jika digunakan untuk proses klasifikasi beda halnya jika digunakan untuk melakukan prediksi nilai k pada K-NN dapat berupa bilangan ganjil ataupun genap. Perhitungan jarak yang digunakan pada K-NN menggunakan *Euclidean Distance*.

2 Metode

Penelitian ini menggunakan metode penelitian terapan deskriptif. Metode ini bertujuan untuk melakukan penerapan, melakukan pengujian serta melakukan evaluasi terhadap suatu teori algoritma dalam proses memecahkan suatu permasalahan. Adapun tahapan dari penelitian ini yakni sebagai berikut:



Pengumpulan data dilakukan dengan observasi di desa Sapikerep yang dibantu oleh perangkat desa dalam melakukan pendataan terhadap masyarakat tidak sejahtera serta pencocokan data dengan data yang telah ada di desa tersebut.

Pengolahan data awal dilakukan untuk memastikan data yang akan digunakan tidak terdapat *noise*. *Noise* yang dimaksud yakni data yang tidak lengkap atau data yang memiliki nilai kosong (atribut yang tidak memiliki nilai). Selanjutnya menentukan model yang akan digunakan. Model yang digunakan dalam penelitian ini yakni menggunakan algoritma K-NN untuk melakukan klasifikasi terhadap data yang telah diperoleh.

Eksperimen dan Pengujian algoritma K-NN terhadap data yang diperoleh dilakukan untuk mengetahui tingkat akurasi model yang digunakan. Hasil dari eksperimen yang dilakukan

selanjutnya dievaluasi dengan memilih hasil terbaik dari proses klasifikasi telah dilakukan.

Algoritma K-NN merupakan algoritma yang bisa digunakan untuk melakukan prediksi ataupun klasifikasi terhadap suatu data tergantung dari jenis data pada kumpulan data yang ada. K-NN melakukan klasifikasi pada suatu data berdasarkan nilai k yang telah ditetapkan sebelumnya. Nilai k pada K-NN harus menggunakan nilai ganjil jika digunakan untuk proses klasifikasi beda halnya jika digunakan untuk melakukan prediksi nilai k pada K-NN dapat berupa bilangan ganjil ataupun genap. Perhitungan jarak yang digunakan pada K-NN menggunakan *Euclidean Distance*. Berikut persamaan yang digunakan.

$$d(X_i, X_j) = \sqrt{\sum_r^n = 1 (a_r(x_i) - a_r(x_j))^2}$$

Keterangan:

$d(X_i, X_j)$: Jarak Euclidean

(x_i) : record ke-i (baris)

(x_j) : record ke-j (kolom)

(a_r) : data ke-r

i, j : 1,2,3...n

Sedangkan evaluasi menggunakan *k-Fold Cross Validation* dimana cara ini akan membagi seluruh data secara acak berdasarkan nilai k yang telah ditentukan yang nantinya masing-masing bagian akan dilakukan proses klasifikasi dan validasi dari hasil model yang telah digunakan menggunakan *Confusion Matrix* untuk menentukan kinerja dari model yang digunakan. Adapun tabel *Confusion Matrix* seperti berikut:

Tabel 1. *Confusion Matrix*

	Ya	Tidak
Ya	TP	FN
Tidak	FP	TN

Untuk mengetahui akurasi dari suatu model pada *Confusion Matrix* menggunakan persamaan berikut:

$$Akurasi = \frac{TP + TN}{TP + FP + FN + TN}$$

Keterangan:

TP : Jumlah prediksi positif yang diklasifikasikan benar (positif)

- TN : Jumlah prediksi negatif yang diklasifikasikan benar (negatif)
- FP : Jumlah prediksi positif yang di klasifikasikan salah (negatif)
- FN : Jumlah prediksi negatif yang di klasifikasikan salah (positif)

Jumlah data yang digunakan dalam penelitian ini sebanyak 681 data valid dari 920 data kepala keluarga yang terdapat di desa Sapikerep Probolinggo. Data tersebut merupakan data yang telah dilakukan pengolahan data awal dengan menghilangkan data *noise*.

Adapun atribut yang digunakan dalam penelitian ini meliputi:

Tabel 2. Daftar Atribut

Nama Atribut	Isi
Tidak memiliki lahan atau rumah sendiri	Ya/tidak
Tidak mempunyai aset barang atau harta	Ya/tidak
Jenis dinding dari bambu atau tembok sederhana	Ya/tidak
Jenis lantai tanah atau sederhana	Ya/tidak
Sumber penerangan tidak menggunakan listrik atau menggunakan daya listrik dibawah 900 wat	Ya/tidak
Omset usaha <Rp.1.000.00 atau tidak memiliki usaha	Ya/tidak
Sumber air minum dari sumur	Ya/tidak
Sumber energi memasak dari tabung gas	Ya/tidak
Status	Prasejahtera/ Non Prasejahtera

3 Hasil dan Pembahasan

A. Persiapan

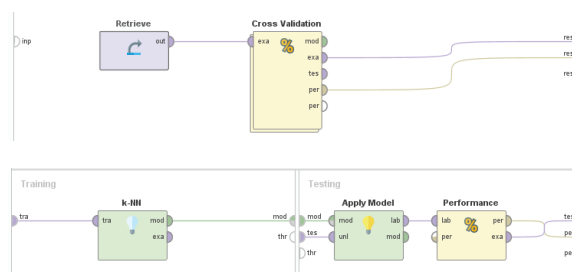
Pengumpulan data dilakukan dengan observasi di desa Sapikerep yang dibantu oleh perangkat desa dalam melakukan pendataan terhadap masyarakat tidak sejahtera serta pencocokan data dengan data yang telah ada di desa tersebut. Data yang digunakan dalam penelitian ini menggunakan data sebanyak 681 data yang diperoleh dari hasil observasi dan validasi yang telah dilakukan di lapangan serta data terkumpul dalam file excel. *Tools*

yang digunakan dalam penelitian ini untuk menguji algoritma K-NN terhadap data yang diperoleh menggunakan Rapidminer Studio.

Jumlah atribut yang digunakan pada dataset yang telah ada sebanyak 9 dan 1 atribut berlaku sebagai *label* dari masing-masing atribut pada data yang ada. Validasi yang digunakan pada penelitian ini menggunakan *10-fold cross validation*, dimana seluruh data yang ada akan dipartisi menjadi 10 bagian yang selanjutnya masing-masing bagian dilakukan proses klasifikasi terhadap bagian yang lain.

B. Pembahasan dan Hasil

Implementasi dari algoritma K-NN terhadap dataset yang telah ada dapat dilihat pada gambar berikut:



Gambar 1. Penerapan Algoritma K-NN dan k-Fold Cross Validation pada Tools Rapidminer

Gambar diatas adalah proses pengolahan data serta penerapan algoritma K-NN serta penggunaan *Cross Validation*. *Cross Validation* digunakan untuk mempartisi data yang digunakan untuk memperoleh akurasi dari algoritma K-NN yang digunakan dalam penelitian ini. Nilai *fold cross validation* yang digunakan dalam proses tersebut menggunakan 10, jadi data yang ada akan dibagi menjadi 10 bagian yang nantinya masing-masing bagian akan dilakukan proses klasifikasi terhadap bagian lainnya. Nilai *k* yang diuji pada K-NN yakni menggunakan *k-3, k-5, dan k-7*. Perhitungan jarak *k* yang digunakan menggunakan *Euclidean Distance*. Adapun hasil dari masing-masing nilai *k* pada algoritma K-NN yakni sebagai berikut:

Tabel 3. Akurasi K-NN berdasarkan nilai *k*

Nilai K	Akurasi
K = 3	97.36%
K = 5	98.68%
K = 7	98.68%

Berdasarkan tabel 3 diatas, hasil klasifikasi algoritma K-NN terhadap data yang ada menunjukkan nilai akurasi terbesar dari hasil klasifikasi algoritma K-NN diperoleh pada nilai $k=5$ dan $k=7$ dengan hasil akurasi 98.68%, sedangkan $k=3$ memperoleh nilai akurasi sebesar 97.36%. Adapun untuk mengetahui akurasi dari setiap nilai k yang digunakan dapat dilihat pada 2 tabel *Confusion Matrix* berikut:

Tabel 4. *Confusion Matrix* nilai $k=3$

	True Prasejahtera	True Non Prasejahtera	Class Precision
Pred. Prasejahtera	309	14	95.67%
Pred. Non Prasejahtera	9	368	98.88%
Class Recall	98.72%	96.20%	

Tabel 4. *Confusion Matrix* nilai $k=5$ dan $k=7$

	True Prasejahtera	True Non Prasejahtera	Class Precision
Pred. Prasejahtera	304	0	100.00%
Pred. Non Prasejahtera	9	368	97.61%
Class Recall	97.12%	100.00%	

4 Kesimpulan

Penelitian ini melakukan pemodel terhadap dataset masyarakat prasejahtera dengan menggunakan algoritma K-NN. Banyak data yang digunakan pada penelitian ini sebanyak 681 data dengan 9 atribut yang terdapat didalam dataset tersebut. Nilai k yang digunakan untuk menguji kinerja algoritma K-NN yakni $k=3$, $k=5$, dan $k=7$, dimana hasil akurasi masing-masing nilai k pada K-NN yang digunakan memiliki hasil akurasi yang berbeda terkecuali nilai akurasi $k=5$ dan $k=7$ yang memiliki nilai akurasi yang sama. Nilai akurasi $k=3$ pada K-NN sebesar 97.36% sedangkan nilai akurasi $k=5$ dan $k=7$ pada K-NN sebesar 98.68%. Berdasarkan hasil tersebut algoritma K-NN sangat sesuai untuk digunakan dalam proses klasifikasi masyarakat prasejahtera yang terdapat di desa sapikerep probolinggo, karena algoritma K-NN memiliki tingkat akurasi yang baik dalam proses klasifikasi terhadap data yang digunakan.

5 Referensi

- Arifin, Zainal, Wali Jafar Shudiq, and Saidatul Maghfiroh. 2019. "Penerapan Metode Knn (K-Nearest Neighbor) Dalam Sistem Pendukung Keputusan Pen-Erimaan KIP (Kartu Indonesia Pintar) Berbasis Web Dan MySQL." 27-34.
- Gorunescu, Florin. 2011. *Data Mining: Concepts, Models and Techniques*.
- Han, Jiawei, Micheline Kamber, and Jian Pei. 2011. *Data Mining Concepts and Techniques Third Edition*. 3rd ed. 225 Wyman Street, Waltham, USA: Morgan Kaufmann.
- Jasri, Moh, and Mochammad Faid. 2017. "Analisis Pemilihan Algoritma Klasifikasi Terbaik Dengan Studi Kasus Dalam Penentuan Bidang Minat Dan Bakat Santri Pada Usia 13-15 Di Pondok Pesantren Nurul Jadid." 1-6.
- Pawening, Ratri Enggar, Wali Jafar Shudiq, and Wahyuni. 2020. "Klasifikasi Kualitas Jeruk Lokal Berdasarkan Tekstur Dan Bentuk Menggunakan Metode K-Nearest Neighbor (K-NN)." 1(1):10-17.
- Rivki, Muhammad, and Adam Mukharil Bachtiar. 2017. "Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia." *Jurnal Sistem Informasi* 13(1):31. doi: 10.21609/jsi.v13i1.500.
- Rizal, Fathur, Andi Wijaya, and Urip Rahman Hidayat. 2020. "Penerapan Algoritma Backpropagation Untuk Klasifikasi Jenis Buah Rambutan Berdasarkan Fitur Tekstur Daun." *Jurnal Aplikasi Teknologi Informasi Dan Manajemen (JATIM)* 1(2).
- Santra, A. K., and C. Josephine Christy. 2012. *Genetic Algorithm and Confusion Matrix for Document Clustering*.
- Supriana, I. Wayan, and Luh Gede Astuti. 2019. "Implementasi K-Nearest Neighbor Pada Penentuan Keluarga Miskin Bagi Dinas Sosial Kabupaten Tabanan." 120-29.
- Windarti, Mariana, and Agustinus Suradi. 2019. "Perbandingan Kinerja 6 Algoritma Klasifikasi Data Mining Untuk Prediksi Masa Studi Mahasiswa." 1(1).