

Classification of Receiving Electricity Subsidy Assistance in Blang Panyang Village Using the K-NN (K-Nearest Neighbor) Method

Miftahul Jannah^{1*}, Cut Syahira Salsabila², Nur Faiza³, Mutasar⁴

^{1,2,3} Universitas Malikussaleh, Indonesia

⁴ Universitas Islam Kebangsaan Indonesia, Indonesia

*Corresponding Author Email: [*miftahul.200170150@mhs.unimal.ac.id](mailto:miftahul.200170150@mhs.unimal.ac.id)

ABSTRAK

Received: 29 December 2023
Revised: 31 December 2023
Accepted: 31 December 2023
Available online: 1 January 2024

Kata Kunci:

Euclidean Distance, K-Nearest Neighbors, Penerimaan subsidi listrik

Program subsidi listrik adalah salah satu program penanggulangan kemiskinan dengan memberikan dana bantuan subsidi listrik kepada rumah tangga miskin dan tidak mampu yang dibayar oleh Pemerintah Indonesia kepada PT. PLN (persero). Pemerintah menerapkan kebijakan subsidi listrik tepat sasaran, subsidi harus benar-benar dinikmati oleh orang-orang miskin. Tujuan penelitian ini adalah untuk melakukan pengujian algoritma K-Nearest Neighbors dalam memprediksi penerimaan bantuan subsidi listrik. Dalam dataset penerima bantuan yang digunakan dalam penelitian ini, terdapat 45 record atau tuple dengan empat atribut (kondisi rumah, penghasilan, pekerjaan dan jumlah ampere). Prediksi kategori data baru dilakukan dengan menggunakan tahapan perhitungan manual Euclidean Distance dari tiga nilai K yang berbeda. Hasilnya menunjukkan bahwa dengan K=15, K=30 dan K=45 data baru (46) memiliki kategori "Tidak Layak" dengan tingkat akurasi sebesar 100%. Kemudian dengan K=45, K=30 dan K=45 data baru (D46) memiliki kategori "Layak" dengan tingkat akurasi sebesar 66,6%.

ABSTRACT

Keywords:

Euclidean Distance, K-Nearest Neighbors, Electricity subsidy acceptance

The electricity subsidy program is one of the poverty reduction programs by providing electricity subsidy assistance funds to poor and disadvantaged households paid by the Government of Indonesia to PT PLN (Persero). The government implements a targeted electricity subsidy policy, the subsidy must be truly enjoyed by the poor. The purpose of this research is to test the K-Nearest Neighbors algorithm in predicting the receipt of electricity subsidy assistance. In the dataset of beneficiaries used in this study, there are 45 records or tuples with four attributes (house condition, income, occupation and number of amperes). The prediction of new data categories is done by using the manual calculation stage of Euclidean Distance from three different K values. The results show that with K=15, K=30 and K=45 the new data (46) has an "Ineligible" category with an accuracy rate of 100%. Then with K=45, K=30 and K=45 the new data (D46) has a "Viable" category with an accuracy rate of 66.6%.

1. INTRODUCTION

Program subsidi listrik adalah salah satu program penanggulangan kemiskinan dengan memberikan dana bantuan subsidi listrik kepada rumah tangga miskin dan tidak mampu yang dibayar oleh Pemerintah Indonesia kepada PT. PLN (persero). Berdasarkan permasalahan di atas, maka hal inilah yang menjadi alasan penulis mengajukan judul penelitian yaitu "KLASIFIKASI PENERIMAAN BANTUAN SUBSIDI LISTRIK MENGGUNAKAN METODE KNN (K-NEAREST NEIGHBOR)". Dalam adanya sistem tersebut diharapkan dapat membantu dalam pengambilan keputusan untuk menentukan calon penerima subsidi dengan cepat, dan tepat. Tujuan pemberian subsidi listrik untuk membantu masyarakat yang lemah, miskin, marjinal dan kaum kecil secara sosio ekonomi sehingga mereka dapat lebih mandiri dan dapat memenuhi kebutuhan dasar hidupnya. Di dalam Undang-Undang Peraturan Menteri Energi dan Sumber Daya Mineral Republik Indonesia Nomor 29 Tahun 2016 tentang

Mekanisme Pemberian Subsidi Tarif Tenaga Listrik Untuk Rumah Tangga. Subsidi listrik merupakan sejumlah dana yang dibayar oleh Pemerintah Indonesia kepada PT. PLN (Persero) Dengan adanya subsidi tersebut, diharapkan ketersediaan listrik dapat terpenuhi, kelangsungan penyediaan listrik dapat berjalan stabil, serta memberi kesempatan kepada pelanggan yang kurang mampu dan masyarakat yang belum terjangkau pelayanan PT.PLN untuk dapat ikut dapat menikmati energi listrik.

Pemberian klasifikasi penerimaan Subsidi listrik pada penelitian ini menggunakan metode K-Nearest Neighbor (KNN). K-Nearest Neighbor adalah metode klasifikasi dengan mencari jarak terdekat antara data yang akan dievaluasi dengan K tetangga (*neighbor*) terdekatnya dalam data pelatihan.

Penentuan nilai terbaik dapat ditentukan dengan optimasi parameter, misalnya dengan menggunakan K-Fold Cross Validation yang merupakan salah satu metode yang digunakan untuk mengetahui rata-rata keberhasilan dari suatu sistem

dengan cara melakukan perulangan dengan mengacak atribut masukan sehingga sistem tersebut teruji untuk beberapa atribut input yang acak.

Tujuan penelitian ini adalah untuk melakukan pengujian K-Fold Cross Validation pada algoritma K-Nearest Neighbors dalam memprediksi penerimaan dana bantuan desa dengan menggunakan kriteria-kriteria yang sudah ditentukan untuk mengetahui penerimaan subsidi listrik layak atau tidak layak untuk menerima bantuan berikut.

2. RESEARCH METHODS

Penelitian ini membahas mengenai implementasi data mining menggunakan algoritma K-Nearest Neighbor (KNN) dengan studi kasus pengujian KFold Cross Validation dalam klasifikasi penerimaan penerimaan Subsidi listrik. Dataset yang digunakan dalam penelitian ini memiliki 45 record atau tuple dengan 4 atribut atau kriteria, yaitu kondisi rumah, penghasilan, pekerjaan dan jumlah ampere. Tabel 1 berikut ini merupakan bobot nilai dari setiap kriteria yang digunakan:

Tabel 1. Bobot nilai setiap kriteria

KRITERIA	NILAI	BOBOT
KONDISI RUMAH	BAMBU	4
	TRIPLEK	3
	BATAKO	2
	BATU BATA	1
	BETON	0
PENGHASILAN	<500.000	4
	500.000-1.000.000	3
	1.000.001-3.000.000	2
	3.000.001-5.000.000	1
	>5.000.000	0
PEKERJAAN	BURUH	4
	PETANI	3
	PNS	2
	WIRAUSAHA	1
	PENGUSAHA	0
JUMLAH AMPERE	2	4
	4	3
	6	2
	8	1
	10	0

Bobot yang dimiliki masing-masing atribut atau kriteria akan dijumlahkan untuk menentukan kategori data tersebut “Layak” atau “Tidak layak”. Dalam pengujian peneliti menggunakan program Microsoft Excel 2010 sebagai alat bantu perhitungan manual.

3. RESULT AND DISCUSSION

Dalam perhitungan manual ini dataset yang telah diperoleh akan dihitung menggunakan metode K-NN secara manual untuk mengetahui kualitas data yang dimiliki apakah cukup baik untuk digunakan sebagai data latih. Metode K-NN adalah salah satu metode klasifikasi yang sering digunakan, bertujuan untuk mengklasifikasikan objek baru berdasarkan data pembelajaran atau data latih yang jaraknya paling dekat dengan objek baru tersebut.

Dalam dataset penerimaan subsidi listrik yang digunakan dalam penelitian ini, terdapat 45 record atau tuple dengan 4 atribut (kondisi rumah, penghasilan, pekerjaan dan jumlah Ampere). Untuk melakukan perhitungan manual metode KNN terlebih dahulu harus ditentukan jumlah data tetangga terdekat yang dinotasikan dengan K, dimana nilai K ditentukan secara bebas. Dalam penelitian ini ditentukan 3 nilai K, yaitu K=15, K=30, K=45. Langkah selanjutnya yaitu menghitung Euclidean Distance sebagai teknik pencarian tetangga terdekat dan berfungsi untuk menguji ukuran yang bisa digunakan sebagai interpretasi kedekatan jarak antara dua objek. Rumus perhitungan euclidean distance sebagai berikut.

$$d_e = \sqrt{\sum_{k=1}^m (fd_{i,k} - k_j)^2} \quad (1)$$

Keterangan:

de : jarak euclidean

fdi : data training

kj : data testing

m : jumlah data penelitian

Berikut ini merupakan sampel perhitungan manual untuk mencari nilai euclidean distance antara data baru (D45) ke data pertama (D1)

$$D(X_i, X_j) = \sqrt{\sum_{n=1}^n (ar(X_i) - ar(X_j))^2}$$

$$D(1,70) = \sqrt{(3-1)^2 + (3-2)^2 + (4-1)^2 + (4-2)^2}$$

$$D(1,70) = 4,242640687$$

Jarak antara data baru (D46) dengan data kedua dan seterusnya dihitung menggunakan cara yang sama. Hasil perhitungan Euclidean Distance seperti pada Tabel 2 berikut ini:

Tabel 2. Hasil perhitungan Euclidean Distance

NO	Home	Income	Work	Ampere	Hasil
1	3	3	4	4	4,242640687
2	2	2	4	4	3,741657387
3	1	1	4	3	3,31662479
4	3	4	4	4	4,582575695
5	3	2	4	4	4,123105626
6	1	2	4	4	3,605551275
7	3	2	4	4	4,123105626
....
....
44	2	2	2	2	1,414213562
45	4	4	3	2	4,123105626

Setelah diketahui jarak antara data baru dengan setiap data latih, maka data diurutkan berdasarkan euclidean distance terkecil ke terbesar. Semakin kecil euclidean distance berarti jarak antar data semakin dekat. Hasil pengurutan akan diberi peringkat sesuai dengan nilai K. Lalu tentukan kategori atau label dari data yang telah di peringkat tersebut. Tabel 3 merupakan data yang telah diurutkan dengan nilai K=15, sehingga data diberi peringkat 1 sampai 15.

Tabel 3. Peringkat data dengan K=15

NO.	KR	PH	PJ	JA	ED	KET	PERINGKAT
46	1	2	1	2	0	Tidak Layak	1
9	1	1	2	2	1,414213562	Tidak Layak	2
26	2	2	2	2	1,414213562	Tidak Layak	3
27	2	1	1	2	1,414213562	Tidak Layak	4
40	2	1	1	2	1,414213562	Tidak Layak	5
42	1	1	2	2	1,414213562	Tidak Layak	6
44	2	2	2	2	1,414213562	Tidak Layak	7
21	1	1	2	2	1,414213562	Tidak Layak	8
13	2	1	1	3	1,732050808	Tidak Layak	9
15	1	3	2	3	1,732050808	Tidak Layak	10
19	2	2	2	3	1,732050808	Tidak Layak	11
39	1	1	2	3	1,732050808	Tidak Layak	12
43	2	1	2	2	1,732050808	Tidak Layak	13
22	2	3	2	3	2	Tidak Layak	14
33	1	0	1	2	2	Tidak Layak	15

Dengan melihat kategori dari 15 tetangga terdekat tersebut maka akan dapat diprediksi kategori dari data baru berdasarkan kategori tetangga terdekat yang paling banyak. Dari Tabel 3 diatas diketahui bahwa 15 data memiliki kategori “Tidak layak”. Maka dapat disimpulkan bahwa data baru (D46) dengan K=15 memiliki kategori “Tidak layak”. Tabel 4 merupakan data yang telah diurutkan dengan nilai K=30, sehingga data diberi peringkat 1 sampai 30.

Tabel 4. Peringkat data dengan K=30

NO.	KR	PH	PJ	JA	ED	KET	PERINGKAT
9	1	1	2	2	1,414213562	Tidak Layak	1
21	1	1	2	2	1,414213562	Tidak Layak	2
....
....
11	2	2	4	3	3,31662479	Layak	25
6	1	2	4	4	3,605551275	Layak	26
17	3	3	3	4	3,605551275	Layak	27
2	2	2	4	4	3,741657387	Layak	28
31	2	2	4	4	3,741657387	Layak	29
5	3	2	4	4	4,123105626	Layak	30

Dengan melihat kategori dari 30 tetangga terdekat tersebut maka akan dapat diprediksi kategori dari data baru berdasarkan kategori tetangga terdekat yang paling banyak.

Dari Tabel 4 diatas diketahui bahwa 15 data memiliki kategori “Layak”, sedangkan 15 data memiliki kategori “Tidak layak”. Maka dapat disimpulkan bahwa data baru (D46) dengan K=30 memiliki kategori “Tidak layak dan layak”. Tabel 5 merupakan data yang telah diurutkan dengan nilai K=45, sehingga data diberi peringkat 1 sampai 45.

Tabel 5. Peringkat data dengan K=45

NO.	KR	PH	PJ	JA	ED	KET	PERINGKAT
9	1	1	2	2	1,414213562	Layak	1
21	1	1	2	2	1,414213562	Layak	2
....
....
4	3	4	4	4	4,582575695	Layak	40
34	4	4	3	4	4,582575695	Layak	41
35	4	4	3	4	4,582575695	Layak	42
38	4	4	4	2	4,69041576	Layak	43
24	4	4	4	4	5,099019514	Layak	44
29	4	4	4	4	5,099019514	Layak	45

Dengan melihat kategori dari 45 tetangga terdekat tersebut maka akan dapat diprediksi kategori dari data baru berdasarkan kategori tetangga terdekat yang paling banyak. Dari Tabel 5 diatas diketahui bahwa 30 data memiliki kategori “Layak”, sedangkan 15 data memiliki kategori “Tidak layak”. Maka dapat disimpulkan bahwa data baru (D46) dengan K=45 memiliki kategori “Layak”.

Dari percobaan perhitungan penentuan penerimaan dana bantuan desa menggunakan perhitungan manual, kategori data baru dan akurasi KNN dengan nilai K yang digunakan dapat dilihat pada Tabel 6 berikut:

Tabel 6. Kategori data baru dan akurasi K-NN dari berbagai nilai K

K	15	30	45
Kategori	tidak layak	tidak layak/layak	layak
Akurasi	100%	50%	66,6%

Hasil penelitian menunjukkan bahwa nilai K=15 merupakan nilai K yang paling optimal dengan tingkat akurasi sebesar 100 % dan kategori data baru “Tidak Layak”. Sedangkan untuk K=30 dan K=45 memiliki tingkat akurasi sebesar 66,6% dengan kategori data baru “Layak”. Perhitungan ini menggunakan confusion matrix untuk menentukan prosentase data yang sesuai dengan kenyataan dibandingkan jumlah keseluruhan data yang ada. Nilai K yang digunakan dalam penelitian ini seluruhnya adalah nilai kelipatan 15 mulai dari K=15 sampai dengan K=45. Penentuan hasil klasifikasi untuk nilai K lebih dari 1 digunakan metode hasil terbanyak atau mayoritas hasil klasifikasi.

4. CONCLUSION

Algoritma K-NN digunakan dalam penelitian ini untuk melakukan klasifikasi terhadap dataset penerimaan bantuan subsidi listrik. Penelitian dilakukan melalui tahapan

perhitungan manual untuk mengetahui hasil kelayakan atau ketidaklayakan dari sebuah data baru. Hasilnya menunjukkan bahwa dengan $K=15$ memiliki kategori “Tidak layak” dan $K=30$ data baru (D46) memiliki kategori “Tidak layak dan layak” dengan tingkat akurasi sebesar 50%. Kemudian dengan $K=45$ data baru (D46) memiliki kategori “Layak” dengan tingkat akurasi sebesar 66,6%. Diharapkan untuk penelitian selanjutnya menggunakan data yang lebih banyak agar hasil lebih optimal, serta membandingkan dengan algoritma lainnya. Hasil penelitian ini direkomendasikan untuk dipergunakan sebagai pertimbangan dalam menentukan penerima bantuan subsidi listrik.

REFERENCES

- [1] Dinata, R.K., Adek, R.T., Hasdyna, N., Retno, S. (2023). K-nearest neighbor classifier optimization using purity. *AIP Conference Proceedings*. 2431(1).
- [2] Duca, A. L., Bacciu, C., & Marchetti, A. (2017, June). A K-nearest neighbor classifier for ship route prediction. In *OCEANS 2017-Aberdeen* (pp. 1-6). IEEE.
- [3] Gunawan, A. A. N., Supardi, I. W., Wendri, N., Negara, A. A. N. F. C., Bagaskara, A. A. N., Patriawan, P., & Dharmawan, I. B. G. (2024). Digital Mammography Data Transformation Model to Improve the Accuracy of Benign and Malignant Detection in Breast Cancer. *International Journal of Intelligent Systems and Applications in Engineering*, 12(1), 647-665.
- [4] Hasanah, R.L., Hasan, M., Pangesti, W.E., Wati, F.F., Gata, W. (2019). *KLASIFIKASI PENERIMA DANA BANTUAN DESA MENGGUNAKAN METODE KNN (K-NEAREST NEIGHBOR)*. *Jurnal TECHNO Nusa Mandiri*. 16(1): 1–6.
- [5] Hasdyna, N., Retno, S. (2022). Machine Learning Approach to Determine the Drug-Prone Areas in Lhokseumawe City, Indonesia. *International Journal of Multidisciplinary Research and Analysis*. 5(9): 2354-2464.
- [6] Kesarwani, A., Chauhan, S. S., & Nair, A. R. (2020, June). Fake news detection on social media using k-nearest neighbor classifier. In *2020 international conference on advances in computing and communication engineering (ICACCE)* (pp. 1-4). IEEE.
- [7] Labuda, N., Seeliger, J., Gedrande, T., & Kozak, K. (2017). Selecting adaptive number of nearest neighbors in k-nearest neighbor classifier apply diabetes data. *J. Math. Stat. Sci*, 2017, 1-13.
- [8] Muharrom, M. (2023). Analisis Komparasi Algoritma Data Mining Naive Bayes, K-Nearest Neighbors dan Regresi Linier Dalam Prediksi Harga Emas. *Bulletin of Information Technology (BIT)*, 4(4), 430-438.
- [9] Retno, S., Dinata, R.K., Hasdyna, N. (2023). Evaluasi model data chatbot dalam natural language processing menggunakan k-nearest neighbor. *Jurnal CoSciTech (Computer Science and Information Technology)*. 4(1): 146-153.
- [10] Xie, T., Chen, L., Yi, B., Li, S., Leng, Z., Gan, X., & Mei, Z. (2023). Application of the Improved K-Nearest Neighbor-Based Multi-Model Ensemble Method for Runoff Prediction. *Water*, 16(1), 69.