



Algoritma *K-Nearest Neighbor* Untuk Menentukan Kelayakan Keluarga Penerima Bantuan Pangan Non Tunai (Studi Kasus : Kelurahan Karya Jaya)

Sastri Yani¹, Fithri Selva Jumeilah², Muhamad Kadafi³

^{1,2,3}Program Studi Sistem Informasi, UIN Raden Fatah Palembang, Indonesia
Email: ¹sastriyani_1635400127@radenfatah.ac.id, ²fithriselvajumeilah@radenfatah.ac.id,
³kadafi_uin@radenfatah.ac.id

Abstract

Non-cash Food Assistance (BPNT) is food social assistance in the form of non-cash. In its implementation, this program still encounters a number of obstacles, one of which is in the sub-optimal distribution of aid in several regions, including Karya Jaya Village. This is because the Ministry of Social Affairs is not optimal in determining BPNT recipients. One way to solve this problem is by utilizing one of the data mining concepts, namely the classification technique with the *K-Nearest Neighbor* algorithm. Where KPM data previously only accumulated can be used as useful information, one of which is to predict the eligibility of BPNT recipients in the next period. The results of this research are in the form of information on the results of predictions of appropriate KPM as BPNT recipients in 2021 and Local Environmental Units (SLS) which are the most receiving regions. This information can be used as evaluation material for the Ministry of Social Affairs in determining the more targeted BPNT recipients. The prediction results of BPNT recipients in Karya Jaya Village in 2021 are 511 recipients with an accuracy rate of 75.79%, 76.17% Precision, 89.24% Recall, and 82.19% F-measure. And it can be seen that the most BPNT recipient categories are in SLS RW 005, namely 74 recipients. Where there are variables that most influence, namely *sta_kis*.

Keywords: Feasibility, BPNT, Data Mining, *K-Nearest Neighbor*

1. PENDAHULUAN

Salah satu perkara mendasar yang menjadi pusat perhatian pemerintah diberbagai negara, termasuk Indonesia ialah tingkat kemiskinan yang tinggi. Untuk membantu meminimalkan tingkat kemiskinan tersebut, pemerintah mengadakan beberapa program bantuan, yaitu salah satunya program bantuan sosial untuk rakyat miskin dalam bentuk non tunai atau yang biasa disebut dengan Bantuan Pangan Non Tunai (BPNT) [1].



BPNT ini ialah bantuan sosial pangan dalam bentuk non tunai yang bersumber dari pemerintah, kemudian diserahkan kepada Keluarga Penerima Manfaat (KPM) melalui mekanisme account elektronik pada setiap bulannya yang akan digunakan hanya untuk membeli bahan pangan berupa telur dan beras di pedagang bahan pangan atau *e-warong* [2]. Dalam pelaksanaannya BPNT masih menemui sejumlah kendala salah satunya dalam hal penyaluran bantuan yang belum optimal di beberapa daerah, termasuk Kelurahan Karya Jaya. Dalam menjalankan program BPNT, banyak warga yang mengeluh karena tidak mendapat bantuan, sedangkan ada beberapa warga yang dianggap mampu justru mendapatkan bantuan. Sehingga timbul masalah pembagian bantuan sosial yang tidak merata dan tepat sasaran di lingkungan Kelurahan Karya Jaya. Hal ini disebabkan karena kurang optimalnya dalam penentuan penerima BPNT oleh Kementerian Sosial.

Salah satu cara untuk mengatasi permasalahan diatas ialah dengan memanfaatkan salah satu konsep *data mining* yaitu teknik klasifikasi. *Data Mining* ialah disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang kita miliki. Klasifikasi ialah proses untuk memperoleh model atau aturan yang dapat mengklasifikasikan data baru yang belum pernah dipelajari dengan mempelajari sekumpulan data yang lama [3]. Teknik yang digunakan adalah dengan memanfaatkan data penerima BPNT lama yang nantinya akan digunakan sebagai *data training* yang digunakan untuk menguji data baru apakah data baru berhak menerima BPNT berdasarkan kemiripan atribut yang ada di data lama dengan data baru.

Sehingga dalam penelitian ini akan dilakukan proses pengolahan data menggunakan *data mining* untuk mengklasifikasi layak dan tidak layaknya menerima BPNT. Penggunaan algoritma yang tepat dapat meningkatkan keakuratan keputusan yang diambil. Metode klasifikasi algoritma *K-Nearest Neighbor* merupakan salah satu metode pengklasifikasian data yang memiliki konsistensi yang kuat, dengan cara mencari kasus dengan menentukan nilai jarak pada pengujian *data testing* dan *data training* berdasarkan nilai terkecil dari nilai ketetanggaan terdekat [4]. Algoritma ini memiliki beberapa kelebihan yaitu bahwa dia tangguh terhadap *training* data yang noise, lebih efektif dalam melakukan *training* data yang besar dan dapat menghasilkan data yang lebih akurat. Metode ini dapat melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut [5].

Berdasarkan latar belakang di atas penulis tertarik untuk melakukan penelitian dengan judul "Algoritma *K-Nearest Neighbor* Untuk Menentukan Kelayakan

Keluarga Penerima Bantuan Pangan Non Tunai (Studi Kasus : Kelurahan Karya Jaya)”.
2. METODOLOGI PENELITIAN

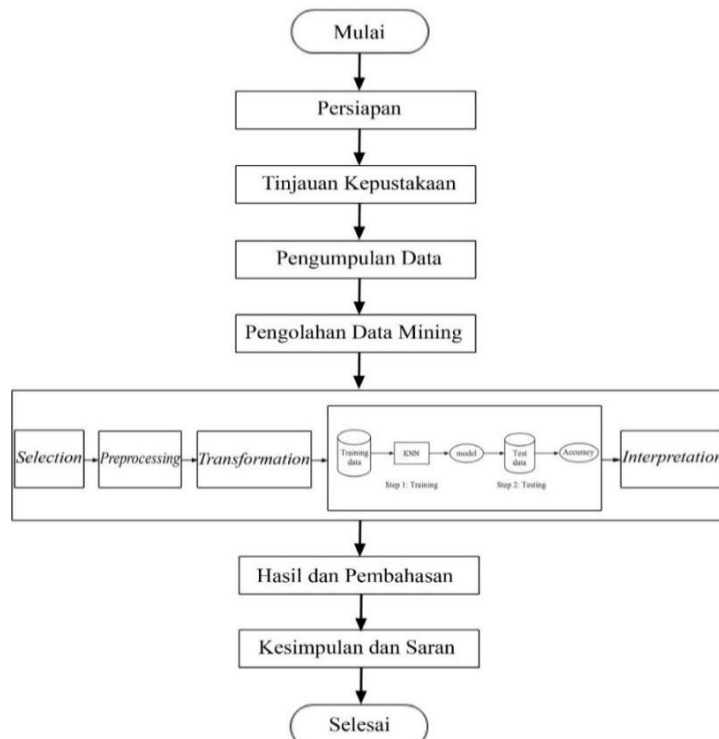
2.1 Metode Penelitian

Metode penelitian yang digunakan dalam penelitian ini ialah metode penelitian kuantitatif dengan memanfaatkan metode data mining dan perhitungan algoritma K-Nearest Neighbor.

2.2 Bahan Penelitian

Dalam penelitian ini bahan yang digunakan untuk diolah menjadi acuan adalah data dari daftar kolektif hasil kuisisioner PPLS (Program Perlindungan Sosial) Kelurahan Karya Jaya pada tahun 2017-2020.

2.3 Tahapan Penelitian



Gambar 1. Tahapan penelitian

Tahapan penelitian yang dituangkan dalam diagram alir diatas menggambarkan proses penelitian yang akan dilaksanakan, yaitu :

- a. Persiapan
Pada tahap ini menentukan objek yang akan diteliti yaitu data dari daftar kolektif hasil kuisisioner PPLS (Program Perlindungan Sosial) Kelurahan Karya Jaya pada tahun 2017-2020 yang akan menjadi bahan untuk klasifikasi kelayakan penerima Bantuan Pangan Non Tunai (BPNT) Kelurahan Karya Jaya.
- b. Tinjauan Kepustakaan
Pada tahap ini dilakukan telaah terhadap kriteria dalam klasifikasi penerima Bantuan Pangan Non Tunai (BPNT).
- c. Pengumpulan Data
Proses pengumpulan data dilakukan dengan cara mengambil data dari daftar kolektif hasil Kuisisioner PPLS (Program Perlindungan Sosial) tahun 2017-2020.
- d. Pengolahan Data mining
Data yang telah dikumpulkan akan diolah sesuai dengan tahapan Knowledge Discovery in Databases (KDD).
- e. Hasil dan Pembahasan
Pada tahap ini akan dijelaskan proses hasil dari data mining yang dilakukan dengan algoritma *K-Nearest Neighbor* (KNN).
- f. Kesimpulan dan Saran
Membuat kesimpulan dari hasil penelitian dan saran untuk penelitian selanjutnya.

3. HASIL DAN PEMBAHASAN

3.1 Hasil

Setelah melakukan analisa dengan tahapan data mining untuk menghasilkan klasifikasi kelayakan terhadap penerima BPNT pada Kelurahan Karya Jaya dengan menggunakan algoritma *K-Nearest Neighbor*, maka hasil yang dicapai oleh peneliti ialah mengetahui bahwa dengan teknik klasifikasi maka dapat dilihat hasil prediksi KPM (Keluarga Penerima Manfaat) yang layak sebagai penerima BPNT pada tahun 2021. Dimana prediksi BPNT Kelurahan Karya Jaya pada tahun 2021 ialah sebanyak 511 penerima dengan tingkat Akurasi 75,79%, *Precision* 76,17%, *Recall* 89,24%, *F-measure* 82,19%, Serta dapat

dilihat kategori penerima BPNT terbanyak ada pada SLS (Satuan Lingkungan Setempat) RW 005 yaitu 74 penerima.

Pada Gambar 2 dapat dilihat bahwa dari semua variabel yang digunakan terdapat variabel yang paling mempengaruhi dalam proses klasifikasi yaitu variabel `sta_kis` yang artinya 100% KPM (Keluarga Penerima Manfaat) yang terdata sebagai penerima Kartu Indonesia Sehat (KIS) akan mendapat peluang yang besar menjadi penerima BPNT.

```
> varImp(data_knn)
ROC curve variable importance

  only 20 most important variables shown (out of 26)

      Importance
sta_kis      100.000
sta_rastra   99.581
Jumlah_ART   61.823
sta_pkh      46.056
sta_bangunan 39.378
daya         31.017
lantai       24.482
sumber_airminum 21.066
fasbab       20.849
dinding      17.985
buang_tinja  17.878
ada_sepeda   16.231
kondisi_atap 13.778
cara_peroleh_airminum 11.995
kondisi_dinding 9.771
Jumlah_Keluarga 8.438
ada_lemari_es 8.089
ada_motor    5.249
ada_tabung_gas 4.557
sumber_penerangan 3.457
> |
```

Gambar 2. Variabel Yang Paling Mempengaruhi

3.2 Pembahasan

Tahapan yang digunakan dalam pengolahan *data mining* pada penelitian ini, yaitu mengikuti tahapan *Knowledge Discovery in Database* (KDD) sebagai berikut :

3.2.1 Data Selection

Dataset yang digunakan ialah daftar Program Perlindungan Sosial Kelurahan Karya Jaya pada tahun 2017-2020 terdiri atas 71 variabel dan 3081 record. Sebelum dilakukan proses data mining pada proses ini dataset akan di reparasi yaitu menghilangkan variabel yang tidak dipakai ketika proses data mining.

Variabel yang dipakai dalam proses klasifikasi penerima BPNT ini menggunakan Variabel dari Badan Pusat Statistik, Pemilihan variabel untuk proses klasifikasi penerima BPNT dapat dilihat pada Tabel 1.

Tabel 1. Keterangan Variabel

No	VARIABEL	KETERANGAN	TIPE DATA
1	Jumlah_Keluarga	Jumlah Keluarga yang masuk tanggungan kepala rumah tangga	<i>Integer</i>
2	Jumlah_ART	Jumlah anggota rumah tangga	<i>Integer</i>
3	sta_bangunan	Status penguasaan bangunan tempat tinggal yang ditempati	Ordinal
4	luas_lantai	Luas Lantai ... m^2	<i>Integer</i>
5	lantai	Jenis lantai terluas	Ordinal
6	dinding	Jenis dinding terluas	Ordinal
7	kondisi_dinding	Jika jenis dinding terluas berkode 1 atau 2, kondisi dinding	Ordinal
8	Atap	Jenis atap terluas	Ordinal
9	kondisi_atap	Jika jenis atap terluas berkode 1,2,3,4, atau 5, kondisi atap	Ordinal
10	sumber_airminum	Sumber air minum	Ordinal
11	cara_peroleh_air_minum	Cara memperoleh air minum	Ordinal
12	sumber_penerangan	Sumber penerangan utama	Ordinal
13	Daya	Jika listrik PLN ($8a=1$), daya terpasang	Ordinal
14	bb_masak	Bahan bakar/energy utama untuk memasak	Ordinal
15	Fasbab	Penggunaan fasilitas tempat buang air besar	Ordinal
16	buang_tinja	Tempat pembuangan akhir tinja	Ordinal
17	ada_mobil	Apakah rumah tangga memiliki sendiri asset mobil	Ordinal
18	ada_motor	Apakah rumah tangga memiliki sendiri asset motor	Ordinal
19	ada_sepeda	Apakah rumah tangga memiliki sendiri asset sepeda	Ordinal
20	ada_lemari_es	Apakah rumah tangga memiliki sendiri asset lemari es	Ordinal
21	ada_tabung_gas	Apakah rumah tangga memiliki sendiri asset tabung gas	Ordinal

No	VARIABEL	KETERANGAN	TIPE DATA
22	sta_pkh	Apakah rumah tangga menjadi peserta program pkh	Ordinal
23	sta_rastra	Apakah rumah tangga menjadi peserta program rastra	Ordinal
24	sta_kis	Apakah rumah tangga menjadi peserta program kis	Ordinal
25	sta_asuransi	Apakah rumah tangga menjadi peserta program asuransi	Ordinal
26	sta_jamsostek	Apakah rumah tangga menjadi peserta program jamsostek	Ordinal
27	KETERANGAN	Keterangan : 1. Penerima 2. Bukan Penerima	Ordinal

Dalam penelitian ini pada awalnya terdapat 71 variabel kemudian setelah dilakukan seleksi data menjadi 27 variabel.

3.2.2 Pre-Processing

Dalam *dataset* penerima BPNT yang digunakan dalam penelitian ini, terdapat 3081 *record* atau *tuple* sebelum dilakukan proses *pre-processing* dan setelah dilakukan *pre-processing* menjadi 2748 *record* atau *tuple*.

3.2.3 Transformation

Pada tahap ini data yang telah dipilih akan diubah sehingga sesuai untuk proses *data mining*. Adapun *attribute* yang melalui proses transformasi sebagai berikut :

- luas_lantai
Pada *attribute* luas_lantai data akan diubah dari nilai *integer* menjadi ordinal sebagai berikut :
 $\leq 8 \text{ m}^2/\text{orang} = 1$;
 $> 8 \text{ m}^2/\text{orang} = 2$.

Tabel 2. Transformasi Data luas_lantai

Data Awal luas_lantai	Jumlah_ART	Luas_lantai/orang	Konversi Data Ordinal
14 m^2	3	4,67 m^2	1
45 m^2	2	22,5 m^2	2

2. jumlah_keluarga

Pada *attribute* jumlah_keluarga tipe data akan diubah dari *integer* menjadi ordinal sebagai berikut :

Tabel 3. Transformasi Data jumlah_keluarga

Data Awal	Konversi Data Ordinal
1	1
2	2
3	3
4	4

3. jumlah_ART

Pada *attribute* jumlah_ART data akan diubah menjadi ordinal sebagai berikut:

Tabel 4. Transformasi Data jumlah_ART

Data Awal	Konversi Data Ordinal
3	2
4	3
5	3
6	4
7	4
8	5
9	5

3.2.4 *K-Nearest Neighbor*

Berikut ini contoh perhitungan manual *K-Nearest Neighbor* dengan menggunakan data sample pada tahun 2017-2020 sebagai *data training* dan 2021 sebagai *data testing*.

Adapun langkah-langkah dari perhitungan *K-Nearest Neighbor* :

1. Penentuan nilai K. Pada penelitian ini nilai K yang digunakan adalah K=5.
2. Hitung jarak antar data training dan data uji (test) dengan menggunakan perhitungan Euclidean Distance sebagai berikut :

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

Keterangan :

$d(x,y)$ = Jarak Euclidean x_i = Nilai data latih yang ke- i
 y_i = Nilai data uji yang ke- i_n = Banyaknya data
 i = 1,2,3,...n

Tabel 5. *Data Training*

No	Jumlah_Keluarga	Jumlah_ART	sta_bangunan	luas_lantai	lantai	dinding	kondisi_dinding	...	sta_rastra	sta_kis	sta_asuransi	sta_jamsostek	KETERANGAN
1	1	2	2	2	8	3	2	...	3	2	4	2	PENERIMA
2	1	4	1	2	8	3	2	...	3	1	4	2	BUKAN PENERIMA
3	1	3	1	2	8	3	2	...	3	1	4	2	PENERIMA
4	1	2	5	2	8	3	2	...	3	2	4	2	BUKAN PENERIMA
5	1	3	1	1	6	1	2	...	4	2	4	2	BUKAN PENERIMA
6	1	3	3	2	8	3	2	...	3	1	4	2	PENERIMA
7	1	3	1	1	8	3	2	...	3	1	4	2	PENERIMA
8	2	4	2	1	8	3	2	...	3	2	4	2	PENERIMA
9	2	4	1	1	8	3	2	...	3	2	4	2	PENERIMA
10	1	2	1	2	8	3	2	...	3	1	4	2	PENERIMA
...
2050	1	3	1	1	8	3	2	...	3	1	4	2	PENERIMA

Tabel 6. *Data Testing*

No	Jumlah_Keluarga	Jumlah_ART	sta_bangunan	luas_lantai	lantai	dinding	kondisi_dinding	...	sta_rastra	sta_kis	sta_asuransi	sta_jamsostek	KETERANGAN
1	1	3	1	2	8	3	2	...	3	2	4	2	?

$$\begin{aligned}
 d(1,1) &= \sqrt{(1-1)^2 + (2-3)^2 + (2-1)^2 + \dots + (4-4)^2 + (2-2)^2} \\
 &= \sqrt{(0)^2 + (-1)^2 + (1)^2 + \dots + (0)^2 + (0)^2} \\
 &= \sqrt{0 + 1 + 1 + \dots + 0 + 0} \\
 &= \sqrt{45} \\
 &= 6.70820393249937
 \end{aligned}$$

Jarak antara *data training* dan data uji (*test*) seterusnya dihitung menggunakan cara yang sama. Hasil perhitungan *Euclidean distance* secara manual seperti pada Tabel 7.

Tabel 7. Hasil Perhitungan *Euclidean Distance*

No	Jumlah_Keluarga	Jumlah_ART	sta_bangunan	...	sta_kis	sta_asuransi	sta_jamsostek	KETERANGAN	<i>Euclidean Distance</i>
1	1	2	2	...	2	4	2	PENERIMA	6.70820393249937
2	1	4	1	...	1	4	2	BUKAN PENERIMA	7.41619848709566
3	1	3	1	...	1	4	2	PENERIMA	6.6332495807108
4	1	2	5	...	2	4	2	BUKAN PENERIMA	8.94427190999916
5	1	3	1	...	2	4	2	BUKAN PENERIMA	7.61577310586391
6	1	3	3	...	1	4	2	PENERIMA	7.34846922834953
7	1	3	1	...	1	4	2	PENERIMA	6.48074069840786
8	2	4	2	...	2	4	2	PENERIMA	3.74165738677394
9	2	4	1	...	2	4	2	PENERIMA	6.78232998312527
10	1	2	1	...	1	4	2	PENERIMA	8.42614977317636
...
2050	1	3	1	...	2	4	2	BUKAN PENERIMA	6.85565460040104

- Setelah diketahui jarak antara data baru dengan data latih, maka data diurutkan berdasarkan *euclidean distance* terkecil ke terbesar. Semakin kecil *euclidean distance* berarti jarak antar data semakin dekat.

Tabel 8. Hasil Perhitungan *Euclidean Distance* Yang Telah Diurutkan

No	Jumlah_Keluarga	Jumlah_ART	sta_bangunan	...	sta_kis	sta_asuransi	sta_jamsostek	KETERANGAN	<i>Euclidean Distance</i>
531	1	3	1	...	2	4	2	BUKAN PENERIMA	1.73205080756888
181 2	1	4	1	...	1	4	2	PENERIMA	1.73205080756888
274	1	2	1	...	2	4	2	BUKAN PENERIMA	2.00000000000000
743	1	2	1	...	2	4	2	BUKAN PENERIMA	2.00000000000000
142 6	1	2	1	...	2	4	2	BUKAN PENERIMA	2.00000000000000
526	3	3	1	...	1	4	2	PENERIMA	2.44948974278318
661	1	2	1	...	2	4	2	BUKAN PENERIMA	2.44948974278318
108 0	1	2	1	...	2	4	2	BUKAN PENERIMA	2.44948974278318
141 6	1	2	3	...	2	4	2	PENERIMA	2.44948974278318
202 8	1	5	1	...	2	4	2	BUKAN PENERIMA	2.64575131106459
...
195 7	1	3	1	...	2	4	2	BUKAN PENERIMA	11.13552872566

- Menentukan kelompok data hasil uji berdasarkan label mayoritas dari k tetangga terdekat. Karena nilai k = 5, maka yang diambil 5 terkecil.

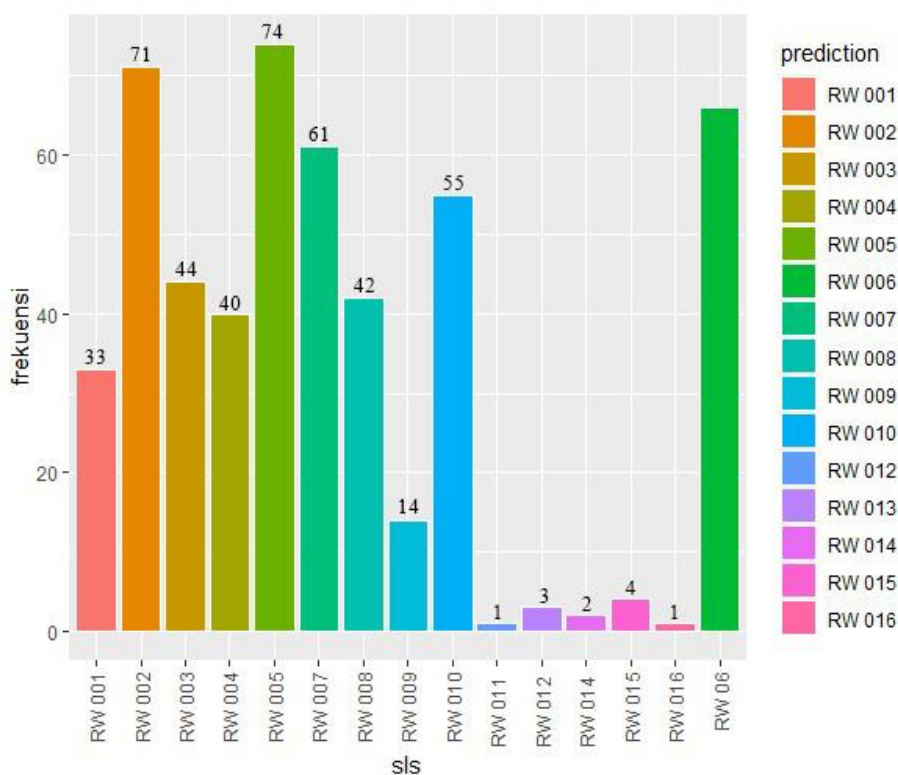
Tabel 9. Peringkat Data dengan K=5

No	Jumlah_Keluarga	Jumlah_ART	sta_bangunan	...	sta_kis	sta_asuransi	sta_jamsostek	KETERANGAN	<i>Euclidean Distance</i>
----	-----------------	------------	--------------	-----	---------	--------------	---------------	------------	---------------------------

531	1	3	1	...	2	4	2	BUKAN PENERIMA	1.73205080756888
181	1	4	1	...	1	4	2	PENERIMA	1.73205080756888
274	1	2	1	...	2	4	2	BUKAN PENERIMA	2.00000000000000
743	1	2	1	...	2	4	2	BUKAN PENERIMA	2.00000000000000
142	1	2	1	...	2	4	2	BUKAN PENERIMA	2.00000000000000

4.3 Implementasi RStudio

Hasil prediksi dengan RStudio dapat dilihat pada Gambar 3. Pada grafik tersebut dapat dilihat bahwa hasil prediksi penerima Bantuan Pangan Non Tunai ada pada SLS RW 005 dengan hasil prediksi 74 KPM.



Gambar 3. Grafik Hasil Prediksi Berdasarkan Kategori SLS

5. KESIMPULAN

Berdasarkan hasil dan analisis data yang telah dipaparkan, maka dapat disimpulkan bahwa :

1. Teknik klasifikasi data mining yaitu dengan menggunakan algoritma K-Nearest Neighbor dapat digunakan untuk prediksi penerima BPNT pada tahun 2021.
2. K optimal untuk melakukan pemodelan adalah $k=5$ dengan accuracy 76,91%.
3. Prediksi penerima BPNT Kelurahan Karya Jaya pada tahun 2021 adalah sebanyak 511 penerima dengan kategori penerima BPNT terbanyak ada pada SLS RW 005 yaitu 74 penerima.
4. Tingkat akurasi dari hasil Prediksi penerima Bantuan Pangan Non Tunai (BPNT) Kelurahan Karya Jaya pada tahun 2021 adalah 75,79%.
5. Variabel yang paling mempengaruhi proses klasifikasi adalah `sta_kis`.

DAFTAR PUSTAKA

- [1] P. Maharani, "Pedoman Umum Bantuan Pangan Nontunai 2019," pp. 1–174, 2019, [Online]. Available: <https://www.kemsos.go.id/uploads/topics/15767284433221.pdf>.
- [2] C. A. Sugianto and F. R. Maulana, "Algoritma Naïve Bayes Untuk Klasifikasi Penerima Bantuan Pangan Non Tunai (Studi Kasus Kelurahan Utama)," *Techno.Com*, vol. 18, no. 4, pp. 321–331, 2019, doi: 10.33633/tc.v18i4.2587.
- [3] F. S. Jumeilah, "Klasifikasi Opini Masyarakat Terhadap Jasa Ekspedisi JNE dengan Naïve Bayes," *J. Sist. Inf. Bisnis*, vol. 8, no. 1, p. 92, 2018, doi: 10.21456/vol8iss1pp92-98.
- [4] F. Liantoni, "Klasifikasi Daun Dengan Perbaikan Fitur Citra Menggunakan Metode K-Nearest Neighbor," *J. Ultim.*, vol. 7, no. 2, pp. 98–104, 2016, doi: 10.31937/ti.v7i2.356.
- [5] N. Bhatia and Vandana, "Survey of Nearest Neighbor Techniques," vol. 8, no. 2, pp. 302–305, 2010, [Online]. Available: <http://arxiv.org/abs/1007.0085>.