

Penerapan Metode Algoritma K-Nearest Neighbor Menggunakan Rapidminer Studio pada Klasifikasi Status Sosial Ekonomi Studi Kasus : Kelurahan Kapuk Muara Rt 010 Rw 04

Ridlwanaul Fitra¹, Ibnu Rusdi^{*2}

^{1,2}Informatika, Universitas Nusa Mandiri

E-mail: ¹ridlwanaulf@gmail.com, ^{*2}ibnu.ibr@nusamandiri.ac.id

Abstrak

Proses kebijakan sosial ekonomi pemerintah terkait pembagian bantuan sosial pemerintah selalu dibarengi dengan permasalahan. Hal ini dikarenakan tidak adanya pembaruan informasi di masyarakat, proses pembaruan informasi dilakukan beberapa tahun sekali dan itupun dilakukan pemerintah melalui sistem yang terkadang berbeda dengan kenyataannya. Oleh karena itu diharapkan dengan adanya penelitian ini dapat mengatasi permasalahan terkait bantuan sosial dengan melakukan klasifikasi status sosial ekonomi masyarakat agar pembagian bantuan sosial bisa tepat sasaran serta pembaruan informasi dapat dilakukan supaya pendataan status sosial ekonomi dalam masyarakat bisa maksimal. Penggunaan Metode Algoritma K-Nearest Neighbour dapat memberikan hasil akurasi yang tepat dengan mencari kasus dan menghitung jarak antara kasus baru dengan kasus lama berdasarkan bobot. Ditambah dengan Aplikasi RapidMiner Studio yang bersifat terbuka bisa digunakan sebagai alat uji keakuratan data. Hasil dari penelitian ini diharapkan dapat digunakan sebagai acuan untuk memaksimalkan pendataan status sosial ekonomi masyarakat sehingga segala kebijakan sosial ekonomi dapat tersampaikan kepada yang membutuhkan.

Kata Kunci: K-Nearest Neighbour, Klasifikasi, Status Sosial Ekonomi, RapidMiner Studio

1. PENDAHULUAN

untuk digunakan mengklasifikasi sampel data baru ke dalam kelas-kelas yang sejenis(Sartika and Indra 2017). Indonesia merupakan salah satu negara berkembang yang sedang berjuang mengatasi masalah kemiskinan. Pada tahun 2017, persentase penduduk Indonesia yang hidup dalam kemiskinan tercatat sebesar 10.1%(Wijayaningrum 2018). Perlunya mengetahui perbedaan status ekonomi masyarakat agar nantinya jika ada bantuan untuk masyarakat kurang mampu, pendataan warga kurang mampu, dan lainnya diharapkan dapat tepat sasaran sehingga akan benar-benar membantu masyarakat yang membutuhkan(Dewi, Obert, and Gusmana 2018). Status sosial ekonomi menggambarkan tentang kondisi seseorang atau suatu masyarakat yang ditinjau dari segi ekonomi, gambaran itu seperti tingkat pendidikan, pendapatan, dan pekerjaan(Indrawati 2015). Saat ini proses pengumpulan data status sosial ekonomi masyarakat yang pernah dilakukan di Lingkungan Kelurahan Kapuk Muara RT. 010 RW.004 belum begitu maksimal. Berdasarkan penjelasan di atas maka penulis mempunyai tujuan dalam melakukan penelitian ini yaitu untuk mengetahui klasifikasi status sosial ekonomi masyarakat Lingkungan Kelurahan Kapuk Muara RT.010, RW.004 dengan memanfaatkan metode Algoritma K-Nearest Neighbor beserta pengujian keakuratan datanya memanfaatkan aplikasi RapidMiner Studio. Metode klasifikasi Algoritma K-Nearest Neighbor merupakan salah satu metode pengklasifikasian data yang memiliki konsistensi yang kuat, dengan cara mencari kasus dengan menghitung kedekatan antara kasus baru dengan kasus lama berdasarkan pencocokan

bobot(Buulolo 2020). Diharapkan dengan diterapkannya Algoritma K-Nearest Neighbor tersebut dapat membantu mengklasifikasikan status ekonomi masyarakat serta memperbaiki pendataannya.

Rumusan masalah adalah bagaimana mengklasifikasikan status sosial ekonomi Masyarakat dari jumlah data yang banyak dan bagaimana menguji akurasi dan kinerja model klasifikasi status sosial ekonomi masyarakat?

Tujuan Penelitian adalah untuk memaksimalkan klasifikasi dan pendataan status sosial ekonomi masyarakat dengan Algoritma K-Nearest Neighbor dan untuk menguji dan memastikan keakuratan data dengan menggunakan software RapidMiner studio..

2. METODE PENELITIAN

2.1. Data Mining

Data mining adalah proses menganalisa data dari perspektif atau penilaian yang berbeda dan menyimpulkannya menjadi informasi-informasi penting yang dapat dipakai untuk meningkatkan keuntungan serta memperkecil biaya pengeluaran dan secara teknis, data mining dapat disebut sebagai proses untuk menemukan korelasi atau pola dari ratusan atau ribuan field dari sebuah relasional database yang besar(Yulianti, Utami, and Hikmah 2019). Data mining sering disebut dengan KDD (Knowledge Discovery in Database) yang berarti proses penemuan pengetahuan melalui pencarian pola yang teratur atau hubungan dalam set data berjumlah besar dengan menggunakan metode atau teknik tertentu(Wanto 2020).

2.2. Klasifikasi Data Mining

Klasifikasi adalah tindakan untuk memberikan kelompok pada setiap keadaan(Sarasati et al. 2020). Klasifikasi pada data mining adalah suatu metode pembelajaran untuk memprediksi nilai dari sekelompok ttribute dalam menggambarkan dan membedakan kelas data atau konsep yang bertujuan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui(Ardiyansyah, Rahayuningsih, and Maulana 2018).

2.3. Algoritma K-Nearest Neighbor (K-NN)

Algoritma K-Nearest Neighbor (K-NN) adalah suatu metode yang menggunakan algoritma supervised dimanahasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada K-NN(Yahya and Puspita Hidayanti 2020). Kelebihan dari metode Algoritma K-Nearest Neighbor (K-NN) yaitu Algoritma K-Nearest Neighbor (K-NN) yang terbukti mencapai hasil akurasi yang baik dan sesuai dengan perhitungan yang diterapkan dalam sebuah aplikasi(Ernawati and Wati 2018). Langkah-langkah cara perhitungan K-Nearest Neighbor berdasarkan Euclidean distance adalah(Yahya and Puspita Hidayanti 2020):

1. Tentukan parameter k.
2. Menghitung jarak antara data yang akan dievaluasi dengan semua data pelatihan.

Untuk menghitung jaraknya menggunakan rumus Euclidean Distance :

$$di = \sqrt{\sum_{i=1}^n (P_i - Q_i)^2}$$

$$di = \sqrt{(p_x - q_x)^2 + (p_y - q_y)^2 + \dots + (p_n - q_n)^2}$$

dimana:

di = jarak variable ke- i

i = variable data ($i = x, y, \dots, n$)

$n = \text{dimensi data}$

$p = \text{data uji } p = (P_X, P_Y, \dots, P_n)$

$q = \text{sampel data } q = (Q_X, Q_Y, \dots, Q_n)$

3. Mengurutkan jarak yang terbentuk dan tentukan jarak terdekat sampai ke $-k$.
4. Memasangkan kelas atau kategori yang bersesuaian.
5. Mencari jumlah terbanyak dari tetangga paling dekat, dan tetapkanlah kategori tersebut sebagai kategori dari data yang dicari.

2.4. RapidMiner

RapidMiner adalah platform perangkat lunak ilmu data yang dikembangkan oleh perusahaan ternama sama dengan yang menyediakan lingkungan terintegrasi untuk persiapan data, pembelajaran mesin, pembelajaran dalam penambangan teks, dan analisis prediktif (Nur Ariesanto Ramdhan, Abdul Khamid, and Agyztia Premana. S. Kom 2020). RapidMiner mempunyai fitur GUI (Graphic User Interface) untuk merancang sebuah pipeline analitis. GUI ini akan menghasilkan file XML (Extensible Markup Language) yang mendefinisikan proses analitis keinginan dibaca oleh RapidMiner untuk melakukan analisis secara otomatis (C Dennis et al. 2013).

2.5. Klasifikasi Status Sosial Ekonomi

Status sosial ekonomi dalam suatu keluarga menggambarkan kondisi seseorang yang ditinjau dari segi ekonomi dengan gambaran tingkat pendidikan, pekerjaan dan pendapatan. Kondisi sosial ekonomi dalam masyarakat ditandai dengan adanya saling kenal mengenal antar satu dengan yang lainnya, paguyuban, sifat kegotong-royongan dan kekeluargaan.

Tolak ukur dari status sosial ekonomi individu bisa dari beberapa unsur kepentingan dalam kehidupan manusia dan juga statusnya didalam masyarakat diantaranya profesi, kekerabatan, jabatan dan agama yang dianutnya (Anon n.d.).

Secara umum dalam masyarakat dunia baik yang sederhana maupun yang kompleks, pola interaksi atau pergaulan hidup antar individu menunjuk pada perbedaan status derajat dan kedudukan kriteria dalam membedakan status pada masyarakat yang kecil biasanya sangat sederhana, karena di samping jumlah warga yang relative lebih sedikit, juga orang-orang semua yang dianggap tinggi statusnya tidak begitu banyak jumlah dan ragamnya (Juariyah 2010).

2.6. Metode Analisis Data

Metode penelitian yang digunakan adalah analisa data kualitatif dan analisa data kuantitatif. Analisa data kualitatif merupakan sebuah analisi data yang dipergunakan apabila data yang terkumpul tidak dapat diangkat dalam artian hanya berupa uraian kata menjadi suatu masalah (Cendekia et al. 2019). Sedangkan analisi data kuantitatif merupakan sebuah analisis data yang dipergunakan apabila kesimpulan-kesimpulan yang diperoleh berupa angka-angka dan juga dalam perhitungan dipergunakan rumus yang ada hubungannya dengan analisis penulisan (Cendekia et al. 2019).

Populasi dan Sampel diambil dari data warga Kelurahan Kapuk Muara RT.010, RW.004. Dari Populasi sebanyak 300 di Kelurahan Kapuk Muara RT.010, RW.004 didapatkan 34 Sampel keluarga yang diambil dari hasil pengisian kusioner.

Preproses data dilakukan, karena data yang didapatkan masih dalam keadaan berantakan dan harus diolah lagi terlebih dahulu. Setelah data ditransformasi atau diolah, kemudian masuk kepada tahap analisis data.

Dalam hal ini akan dipergunakan metode analisis K-Nearest Neighbor dengan membuat data training dan data testing.

1. Data Training

Tabel 1. Data Training

No	Penghasilan (Juta)	Jumlah tanggungan (orang)	Status tempat tinggal (point)	Status ekonomi
1	3,8	4	1	Cukup
2	2,9	3	1	Kurang
3	5,6	6	4	Cukup
4	2,8	1	4	Cukup
5	3,4	5	4	Cukup
6	2,6	4	4	Kurang
7	4,3	5	4	Cukup
8	6,5	4	4	Lebih
9	3,2	4	4	Cukup
10	2,7	4	4	Cukup
11	2	3	4	Kurang
12	3,2	4	1	Cukup
13	3,5	5	1	Cukup
14	3,55	3	4	Cukup
15	3,4	4	1	Cukup
16	3,2	7	4	Kurang
17	2,8	3	1	Cukup
18	3,3	5	4	Cukup
19	2	3	1	Kurang
20	2,6	3	1	Kurang

2. Data Testing

Tabel 2. Data Testing

No	Penghasilan (Juta)	Jumlah tanggungan (orang)	Status tempat tinggal (point)	Status ekonomi
1	3,1	4	1	?

3. HASIL DAN PEMBAHASAN

3.1. Perhitungan Menggunakan Algoritma K-Nearest Neighbor

1. Parameter K yang digunakan adalah 8
2. Hitung jarak antara data uji dengan sampel data yang ada.

Tabel 3. Hasil Perhitungan Ecludian Distanc

No	Penghasilan (Juta)	Jumlah tanggungan (orang)	Status tempat tinggal (point)	Status ekonomi	Jarak
1	3,8	4	1	Cukup	0,7
2	2,9	3	1	Kurang	1,01
3	5,6	6	4	Cukup	4,38

4	2,8	1	4	Cukup	4,25
5	3,4	5	4	Cukup	3,17
6	2,6	4	4	Kurang	3,04
7	4,3	5	4	Cukup	3,38
8	6,5	4	4	Lebih	4,53
9	3,2	4	4	Cukup	3,001
10	2,7	4	4	Cukup	3,02
11	2	3	4	Kurang	3,34
12	3,2	4	1	Cukup	0,1
13	3,5	5	1	Cukup	1,07
14	3,55	3	4	Cukup	1,09
15	3,4	4	1	Cukup	0,3
16	3,2	7	4	Kurang	4,24
17	2,8	3	1	Cukup	1,04
18	3,3	5	4	Cukup	3,16
19	2	3	1	Kurang	1,48
20	2,6	3	1	Kurang	1,118

3. Data yang telah selesai dihitung diurutkan berdasarkan nilai jarak dari nilai yang terkecil sampai nilai terbesar.

Tabel 4. Pengurutan jarak dari yang terkecil sampai terbesar

No	Penghasilan (Juta)	Jumlah tanggungan (orang)	Status tempat tinggal (point)	Status ekonomi	Jarak
12	3,2	4	1	Cukup	0,1
15	3,4	4	1	Cukup	0,3
1	3,8	4	1	Cukup	0,7
2	2,9	3	1	Kurang	1,01
17	2,8	3	1	Cukup	1,04
13	3,5	5	1	Cukup	1,07
14	3,55	3	4	Cukup	1,09
20	2,6	3	1	Kurang	1,118
19	2	3	1	Kurang	1,48
9	3,2	4	4	Cukup	3,001
10	2,7	4	4	Cukup	3,02
6	2,6	4	4	Kurang	3,04
18	3,3	5	4	Cukup	3,16
5	3,4	5	4	Cukup	3,17
11	2	3	4	Kurang	3,34
7	4,3	5	4	Cukup	3,38
16	3,2	7	4	Kurang	4,24
4	2,8	1	4	Cukup	4,25
3	5,6	6	4	Cukup	4,38
8	6,5	4	4	Lebih	4,53

4. Tetapkan tetangga terdekat berdasarkan jarak minimum ke-k, jika ditetapkan nilai k=8, maka diambil 8 jarak terdekat yang terlibat.

Tabel 5. Pengambilan data sejumlah nilai k

No	Penghasilan (Juta)	Jumlah tanggungan (orang)	Status tempat tinggal (point)	Status ekonomi	Jarak
----	--------------------	---------------------------	-------------------------------	----------------	-------

12	3,2	4	1	Cukup	0,1
15	3,4	4	1	Cukup	0,3
1	3,8	4	1	Cukup	0,7
2	2,9	3	1	Kurang	1,01
17	2,8	3	1	Cukup	1,04
13	3,5	5	1	Cukup	1,07
14	3,55	3	4	Cukup	1,09
20	2,6	3	1	Kurang	1,118

5. Berdasarkan hasil tersebut menunjukkan bahwa dari 8 tetangga terdekat terdapat 6 status ekonomi cukup, 2 status ekonomi kurang. Maka dapat disimpulkan bahwa data baru atau data testing warga dengan penghasilan 3,1 Juta, Jumlah tanggungan 4 orang dan status tempat tinggal adalah sewa (1 point) termasuk kedalam status ekonomi “Cukup”

Tabel 6. Hasil akhir

No	Penghasilan (Juta)	Jumlah tanggungan (orang)	Status tempat tinggal (point)	Status ekonomi
1	3,1	4	1	Cukup

3.2. Pengujian dengan menggunakan RapidMiner Studio

Berikut tampilan pengujian dengan RapidMiner Studio:



Gambar 1. Desain RapidMiner untuk menjalankan program

Dari desain tersebut dibuat agar proses uji data dengan rapid miner dapat dijalankan, didalam desain tersebut diantaranya adalah data training, K-NN, data testing, Apply model, dan Performance (Classification).

Row No.	Status ekon.	prediction	confidence	No	Penghasilan
1	1	cukup	0.73	0.203	0.006

Gambar 2. Hasil akurasi

Setelah desain program sebelumnya dijalankan didapatkan hasil seperti diatas diuraikan bahwa warga dengan penghasilan 3,1 Juta, Jumlah tanggungan 4 orang dan status tempat tinggal adalah sewa (1 point) termasuk kedalam status ekonomi “Cukup”.

4. KESIMPULAN

Berdasarkan hasil pembahasan yang telah peneliti lakukan mengenai klasifikasi status sosial ekonomi menggunakan metode Algoritma K-Nearest Neighbor dapat disimpulkan bahwa :

1. Metode Algoritma K-Nearest Neighbor

Metode Algoritma K-Nearest Neighbor dapat digunakan untuk klasifikasi status sosial ekonomi di masyarakat. Dapat kita lihat hasil dari perhitungan Algoritma K-Nearest Neighbor memang akurat dengan melihat data terdekatnya. Kekurangan dalam metode ini adalah Algoritma K-Nearest Neighbor perlu menentukan nilai dari parameter K, dan pembelajaran oleh jarak tidak jelas mengenai jenis jarak dan atribut apa yang harus digunakan untuk memperoleh hasil terbaik serta biaya komputasi yang cukup tinggi karena diperlukan perhitungan jarak dari setiap data training dengan data testingnya.

2. RapidMiner Studio

Penggunaan RapidMiner Studio dalam proses uji keakuratannya memang cepat dalam penentuan hasil. sebagai alat uji keakuratan juga diperoleh hasil yang akurat dengan diperolehnya prediksi status ekonomi cukup dan tidak terdapat missing pada hasil pengujiannya . namun kekurangannya dibutuhkan spesifikasi laptop atau PC dengan RAM minimal 8 Gb agar software ini dapat berjalan dengan lancar.

3. Hasil perhitungan

Setelah didapatkan status sosial ekonomi maka hasil tersebut dapat digunakan sebagai acuan apakah keluarga tersebut layak mendapat bantuan sosial ketika pembagian bantuan sosial sedang berlangsung sehingga dapat mengurangi rasa kecemburuan antar-warga..

DAFTAR PUSTAKA

- [1] Anon. n.d. “Pengertian Status Sosial Ekonomi : Tingkatan, Tolak Ukur Dan Faktor Yang Mempengaruhi Sosial Ekonomi.” Retrieved April 28, 2021 (<https://www.pelajaran.co.id/2020/12/pengertian-status-sosial-ekonomi.html>).
- [2] Ardiyansyah, Panny Agustia Rahayuningsih, and Reza Maulana. 2018. “Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid Miner.” *Jurnal Khatulistiwa Informatika* VI(1):20–28.
- [3] Buulolo, E. 2020. *Data Mining Untuk Perguruan Tinggi*. Deepublish.
- [4] C Dennis, Aprilia, Donny Aji Baskoro, Lia Ambarwati, and I. Wayan Simri Wicaksana. 2013. *Belajar Data Mining Dengan RapidMiner*. edited by R. Sanjaya. Jakarta: Gramedia Pustaka Utama.
- [5] Cendekia, M. S., M. S. Dr. Drs. Ismail Nurdin, M. S. Dra. Sri Hartati, and A. I. 228/JTI/2019. 2019. *METODOLOGI PENELITIAN SOSIAL. MEDIA SAHABAT CENDEKIA*.
- [6] Dewi, Riski Fadilla Karunia, Obert, and Roman Gusmana. 2018. “Implementasi Metode K-Nearest Neighbor (KNN) Dalam Pengelompokan Status Ekonomi Warga.” *Journal of Big Data Analytic and Artificial Intelligence* 4(1):15–22.

- [7] Ernawati, Siti, and Risa Wati. 2018. "Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel." *Jurnal Khatulistiwa Informatika* VI(1):64–69.
- [8] Indrawati, Endang Sri. 2015. "Status Sosial Ekonomi Dan Intensitas Komunikasi Keluarga Pada Ibu Rumah Tangga Di Panggung Kidul Semarang Utara." *Jurnal Psikologi Undip* 14(1):52–57. doi: 10.14710/jpu.14.1.52-57.
- [9] Juariyah, Basrowi dan Siti. 2010. "Analisis Kondisi Sosial Ekonomi Dan Tingkat Pendidikan Masyarakat Desa Srigading, Kecamatan Labuhan Maringgai, Kabupaten Lampung Timur." *Jurnal Ekonomi & Pendidikan* 7(April):58–81.
- [10] Nur Ariesanto Ramdhan, S. K. M. K., S. T. M. T. Abdul Khamid, and M. K. Agyztia Premana. S. Kom. 2020. *Modul Belajar Data Mining Dan RapidMiner*. Penerbit Lakeisha.
- [11] Sarasati, Fajar, Lia Dwi Cahyanti, Annida Purnamawati, and Riyan Latifahul Hasanah. 2020. "Prediksi Klasifikasi Pembangunan Merek Kosmetik Dengan Metode Enbag K-Logres Berdasarkan Keterlibatan Pengguna Facebook." *Indonesian Journal on Software Engineering (IJSE)* 6(1):116–25. doi: 10.31294/ijse.v6i1.8107.
- [12] Sartika, Dewi, and Dana Indra. 2017. "Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, Dan Decision Tree Pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian." *Jurnal Teknik Informatika Dan Sistem Informasi* 1(2):151–61.
- [13] Wanto, A. 2020. *Data Mining : Algoritma Dan Implementasi*. Yayasan Kita Menulis.
- [14] Wijayaningrum, V. N. 2018. "Optimasi Aturan Fuzzy Menggunakan Particle Swarm Optimization Untuk Penentuan Prioritas Rumah Tangga Miskin." *Sentia* 2018 (June).
- [15] Yahya, Yahya, and Winda Puspita Hidayanti. 2020. "Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Efektivitas Penjualan Vape (Rokok Elektrik) Pada 'Lombok Vape On.'" *Infotek: Jurnal Informatika Dan Teknologi* 3(2):104–14. doi: 10.29408/jit.v3i2.2279.
- [16] Yulianti, Yulianti, Dwi Yuni Utami, and Noer Hikmah. 2019. "Implementasi Data Mining Menentukan Game Android Paling Diminati Dengan Algoritma Apriori." *Paradigma - Jurnal Komputer Dan Informatika* 21(1):29–34. doi: 10.31294/p.v21i1.4941.