

Penerapan Seleksi Fitur Untuk Klasifikasi Penerima Bantuan Sosial Pangkalan Sesai Menggunakan Metode K-Nearest Neighbor

Muhammad Fauzan*, Siska Kurnia Gusti, Jasril, Pizaini

Fakultas Sains dan Teknologi, Prodi Teknik Informatika, Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Email: ¹fauzanfaturas@gmail.com, ²siskakurniagusti@uin-suska.ac.id, ³jasril@uin-suska.ac.id, ⁴pizaini@uin-suska.ac.id

Email Penulis Korespondensi: fauzanfaturas@gmail.com

Submitted: 12/08/2023; Accepted: 21/09/2023; Published: 25/09/2023

Abstrak–Ketidakmampuan memenuhi kondisi kebutuhan dasar manusia adalah bagaimana kemiskinan didefinisikan. Untuk menanggulangi masalah tersebut, pemerintah Indonesia menyalurkan Bantuan Sosial salah satunya yaitu Kartu Indonesia Pintar (KIP) yang bertujuan untuk menyekolahkan anak berusia 7-18 tahun yang kurang mampu secara gratis. Tetapi dalam pemberian bantuan di Kelurahan Pangkalan Sesai, petugas pemberi bantuan kerap mengalami kesulitan dikarenakan banyaknya calon penerima bantuan yang mendaftar, lalu data memiliki banyak ketentuan, dan juga waktu para petugas yang relatif singkat. Untuk menyalurkan bantuan sosial ini tentunya harus tepat sasaran. Oleh karena itu penelitian ini bertujuan untuk mencari nilai akurasi terhadap data calon penerima bantuan Kartu Indonesia Pintar (KIP) untuk meningkatkan hasil proses verifikasi data. Untuk mengatasi masalah tersebut penelitian ini menggunakan algoritma K-Nearest Neighbor (K-NN) dan juga seleksi fitur Information Gain untuk mengurangi atribut yang kurang berpengaruh. Data yang digunakan adalah data KIP tahun 2023 yang berjumlah 1998 data dalam bentuk excel dan memiliki 33 atribut. Setelah dilakukannya cleaning dan seleksi fitur Information Gain, data tersebut kini berjumlah 1675 data dan atribut yang dipilih dari hasil seleksi fitur adalah 5 atribut. Hasil klasifikasi terbaik pada penelitian ini didapatkan dengan rasio 7:3 dan 8:2 dan nilai $k = 5$ yang mendapatkan nilai akurasi terbaik sebesar 98,21%, sedangkan nilai akurasi terendah didapatkan menggunakan rasio 9:1 dengan nilai k yang sama saat tidak menggunakan Information Gain sebesar 89,82%.

Kata Kunci: Pangkalan Sesai; Klasifikasi; K-NN; Information Gain; KIP

Abstract–The inability to fulfill basic human needs is how poverty is defined. To address this issue, the Indonesian government implements various social assistance programs, one of which is Kartu Indonesia Pintar (KIP), aimed at providing free education to children aged 7-18 who are economically disadvantaged. However, in the distribution of aid in the Pangkalan sesai sub-district, distributing officers often face challenges due to the high number of eligible recipients applying, complex data requirements, and limited time for the officers. Distributing this social assistance accurately is crucial. Therefore, this research aims to determine the accuracy value for the data of potential recipients of the Kartu Indonesia Pintar (KIP) to enhance the data verification process's outcomes. To tackle this issue, the research employs the K-Nearest Neighbor (K-NN) algorithm and also employs feature selection using Information Gain to reduce less influential attributes. The data used consists of 1998 records of KIP beneficiaries from the 2023 in excel format, with 33 attributes. After performing data cleaning and Information Gain-based feature selection, the dataset is reduced to 1675 records, with 5 selected attributes. The best classification result in this study is achieved with ratios of 7:3 and 8:2, and a value of $k = 5$, yielding the highest accuracy of 98,21%. The lowest accuracy is obtained using a ratio of 9:1 with the same k value when not using Information Gain, resulting in an accuracy of 89,82%.

Keywords: Sesai Base; Classification; K-NN; Information Gain; KIP

1. PENDAHULUAN

Kemiskinan didefinisikan sebagai ketidakmampuan untuk memenuhi kebutuhan dasar manusia. Data Terpadu Kesejahteraan Sosial (DTKS) yaitu suatu data yang menjadi sumber acuan untuk pelaksanaan penyelenggaraan kesejahteraan sosial. Dalam hal ini data tersebut berisikan data pemerlu bantuan sosial dan pelayanan sosial [1]. Dengan adanya Kartu Indonesia Pintar (KIP), siswa yang memiliki keterbatasan finansial dapat melanjutkan pendidikan ke tingkat sekolah menengah dan lebih tinggi [2]. Kartu Indonesia Pintar (KIP) merupakan sebuah kartu yang dirancang khusus untuk keluarga yang memiliki tingkat kemiskinan atau keterbatasan finansial, dengan tujuan memberikan kesempatan kepada anak-anak dalam kelompok usia 7 sampai 18 tahun untuk mendapatkan pendidikan tanpa biaya. Mereka yang mendapatkan Kartu Indonesia Pintar (KIP) ini mendapatkan uang secara tunai untuk pembayaran sekolah yang disimpan dalam fitur KIP negara. Tujuan dari program ini adalah untuk memberikan dukungan kepada siswa yang berasal dari keluarga dengan pendapatan terbatas agar mereka dapat mengakses pendidikan yang layak, mencegah angka putus sekolah, serta memenuhi kebutuhan pendidikan mereka [3]. Kebijakan Kartu Indonesia Pintar (KIP) merupakan sebuah program yang diselenggarakan oleh pemerintah, didasarkan pada Undang-Undang Nomor 32 Tahun 2004 dan Peraturan Pemerintah Nomor 25 Tahun 2005. Program ini bertujuan untuk memastikan timbulnya generasi berkualitas serta memastikan generasi muda menerima pendidikan yang memadai. Program ini juga menjamin bahwa seluruh warga Indonesia memiliki hal atas pendidikan dan pembelajaran tanpa ada pengecualian. Termasuk Masyarakat yang berada dalam kategori ekonomi tinggi maupun rendah, baik di perkotaan maupun di perdesaan [4].

Pangkalan Sesai adalah sebuah desa atau kelurahan yang terletak di wilayah Kecamatan Dumai Barat, Kota Dumai, Provinsi Riau. Terdiri dari 16 RT dan berpenduduk sebanyak 9.265 jiwa atau 2.252 KK. Dengan luas wilayah 13,82 km² maka kepadatan penduduk rata-rata 670 jiwa/km². Ditinjau dari sarana Pendidikan Pangkalan Sesai memiliki 3 TK, 2 SD, 1 SMP dan 1 Universitas. Oleh karena itu, dalam pemberian bantuan sosial bagi anak-anak yang kurang mampu harus tepat sasaran agar anak-anak milenial dapat bersekolah dan belajar pada anak

umumnya. Dalam pemberian bantuan, petugas pemberi bantuan kerap kali mengalami kesulitan penerimaan bantuan tersebut [5]. Hal tersebut disebabkan oleh beberapa faktor yaitu banyaknya penerima bantuan KIP yang mendaftar, penentuan calon penerima bantuan KIP yang memiliki banyak ketentuan parameter tertentu dan waktu para petugas yang relatif singkat. Sehingga diperlukan teknik mengkategorikan calon penerima KIP dengan klasifikasi agar memudahkan petugas penyeleksi bantuan. Oleh sebab itu, algoritma K-Nearest Neighbor (K-NN) digunakan dalam penelitian ini untuk memudahkan para petugas penyeleksi bantuan KIP. Algoritma K-Nearest Neighbor (K-NN) adalah metode yang bisa diterapkan untuk melakukan prediksi atau klasifikasi data berdasarkan jenis data yang ada [6]. Metode K-NN juga berdasarkan prinsip mengklasifikasikan objek berdasarkan data Latihan yang memiliki jarak paling dekat dengan objek tersebut [7]. Dalam penerapan ini, dilakukan perhitungan jarak antara data uji dan data pelatihan. Kemudian, data yang memiliki jarak terdekat diambil berdasarkan mayoritas anggota kelompok K, sesuai dengan nilai K yang telah ditentukan sebelumnya dalam proses klasifikasi [8]. Lalu untuk meningkatkan akurasi dibutuhkan penerepan seleksi fitur. Seleksi fitur dipandang sebagai metode pengolahan awal data yang penting dalam tiap pengenalan pola [9]. Dalam penelitian ini, digunakan metode seleksi fitur Information Gain untuk mengidentifikasi nilai gain paling tinggi. Information Gain adalah pendekatan yang bertujuan untuk menghindari karakteristik atau atribut yang redundan, dan Information Gain mencerminkan properti yang akan dieksploitasi [10] dan juga dapat mempengaruhi pada tingkat nilai akurasi.

Terdapat beberapa jurnal terkait yang dijadikan sebagai referensi dalam penelitian ini. Referensi penelitian yang pertama yaitu penelitian yang dilakukan oleh S. R. Cholil [11], penelitian tersebut menghasilkan nilai akurasi dari metode K-NN dengan hasil akurasi sebesar 90.5% dari 89 data calon penerima beasiswa siswa SMA. Rujukan penelitian kedua adalah penelitian yang dilakukan V. Verdian [12] yang menggunakan penerapan algoritma K-NN untuk menentukan atau memutuskan kelayakan siswa penerima beasiswa berprestasi. Data yang dipakai dalam penelitian itu yaitu sebesar 182 data yang menghasilkan akurasi sebesar 94%, nilai presisi sebesar 91% dan recall 91%. Akurasi tersebut didapatkan dengan menggunakan $K=7$. Lalu rujukan penelitian ketiga yaitu penelitian yang dilakukan W. Suci [13], penelitian tersebut menggunakan metode K-NN untuk klasifikasi data bantuan sosial pada Desa Sindangpano. Penelitian tersebut memperoleh hasil akurasi sebesar 99.57%, presisi sebesar 99.50%, dan BNPT sebesar 99.62%.

Rujukan penelitian selanjutnya yaitu riset dengan menerapkan seleksi fitur Information Gain dan algoritma K-NN. Penelitian rujukan pertama yaitu penelitian yang dilakukan oleh S. H. A. AINI [14] yang mana penelitian ini melakukan klasifikasi yang menerapkan metode K-NN dan proses seleksi fitur Information Gain untuk mengklasifikasikan penyakit jantung. Hasil dari riset ini mengungkapkan bahwa akurasi paling optimal tercapai pada tingkat 92,31%. Hasil ini diperoleh ketika dilakukan pengujian pada distribusi kelas yang seimbang, dengan penggunaan 6 fitur atau atribut dengan nilai k sebesar 25. Sementara itu, pada pengujian dengan distribusi kelas yang tidak seimbang, digunakan 4 fitur dan nilai k sebesar 35. Rujukan penelitian kedua yaitu yang penelitian yang dilakukan oleh M. R. Hasibuan [15] yang mana penelitian ini menginvestigasi klasifikasi penyakit gagal ginjal dengan menerapkan strategi seleksi fitur Information Gain dan metode K-NN. Hasil dari penelitian tersebut mengungkapkan bahwa nilai tertinggi diperoleh dari 4 atribut atau fitur yang terpilih, dengan menggunakan nilai k dari 2 dan 4. Hasil ini menghasilkan tingkat akurasi mencapai 97,7%. Ketika metode Information Gain digunakan tingkat akurasi mencapai 96,8%, sedangkan pada saat tanpa menggunakan penerapan Information Gain, akurasi turun menjadi 79,9%. Sumber rujukan penelitian ketiga adalah penelitian yang dilakukan oleh T. H. Putrisanni [16], penelitian ini bertujuan untuk melakukan klasifikasi terhadap tafsir Al-Quran agar mempermudah umat muslim untuk mencari topik tertentu di dalam Al-Quran. Hasil dalam pengujian penelitian ini menunjukkan tingkat akurasi sebesar 64,10%, presisi sebesar 63%, dan recall sebesar 62,68%. Metode yang digunakan adalah metode K-NN dan Information Gain dengan parameter $k = 17$ dan dataset memiliki pembagian perbandingan sebesar 1:9. Dan rujukan penelitian terakhir yaitu diambil dari salah satu jurnal internasional yang ditulis oleh N. Hasdyna [17] yang meneliti tentang dataset identifikasi kaca yang didapatkan dari UCI Machine Learning Repository dengan memakai metode K-NN yang menghasilkan nilai akurasi terbaik saat melakukan reduksi pertama setelah tidak memakai reduksi dengan menggunakan Information Gain sebesar 77,93%, F-Measure sebesar 37,46%, dan nilai MCC sebesar 16,06%.

Penelitian ini melakukan perhitungan menggunakan nilai K yang paling optimal menggunakan algoritma K-Nearest Neighbor untuk klasifikasi penerima bantuan sosial Pangkalan Sesai. Tujuan penelitian ini untuk mengetahui strategi apa yang akan dipakai dalam penyaluran bantuan dan perbandingan nilai akurasi terhadap data yang menggunakan Information Gain dan tidak menggunakan Information Gain. Hasil penelitian ini semoga bisa digunakan sebagai dasar acuan dalam penyaluran bantuan, salah satunya adalah Kartu Indonesia Pintar (KIP) yang dikelola oleh Dinas Sosial Daerah Kota Dumai.

2. METODOLOGI PENELITIAN

2.1 Langkah-Langkah Penelitian

Langkah-langkah dalam riset ini melibatkan penerapan metode K-Nearest Neighbor, yang dijelaskan melalui alur proses dalam flowchart berikut ini :



Gambar 1. Flowchart Alur Penelitian

Dari flowchart di atas menjelaskan beberapa tahapan, yaitu:

1. Dataset adalah data penelitian yang didapat dari Dinas Sosial Kota Dumai yang berjumlah 1998 data
2. Proses Cleaning Data atau pembersihan data dilakukan dengan membersihkan data yang tidak berguna atau mengganggu untuk mengurangi noise dan mencegah terjadinya masalah kualitas data. Data setelah di cleaning menjadi 1675 data
3. Dalam riset ini, seleksi fitur dilakukan dengan memanfaatkan metode Information Gain. Data pelatihan yang diambil merupakan informasi tentang siswa yang menerima KIP. Rumus yang digunakan dalam seleksi fitur Information Gain untuk menghitung nilai gain adalah sebagai berikut :

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * \text{Entropy}(S_i) \quad (1)$$

4. Pendekatan K-Nearest Neighbor dimanfaatkan untuk mengklasifikasikan data yang akan digunakan. Proses klasifikasi ini berdasarkan contoh dasar yang tidak secara langsung menyajikan pola eksplisit dalam kategori, tetapi bergantung pada label kategori yang terasosiasi dengan dokumen pelatihan [18]. Klasifikasi dataset dilakukan menggunakan tools Rapid Miner dan Python
5. Langkah terakhir yaitu pengujian dan penilaian hasil metode klasifikasi K-Nearest Neighbor dilakukan dengan membandingkan penggunaan Information Gain dan tanpa menggunakan Information Gain.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Pengumpulan Data

Proses pengumpulan data ini diperoleh langsung dari Dinas Sosial Kota Dumai Provinsi Riau melalui wawancara secara tatap muka dengan salah satu pegawai bidang Pemberdayaan Sosial dan Penanganan Fakir mengenai Kartu Indonesia Pintar (KIP). Dalam riset ini, informasi data yang digunakan berasal dari Data Terpadu Kesejahteraan Sosial (DTKS) yang diperoleh dari Kelurahan Pangkalan Sesai. Data ini disajikan dalam bentuk excel. Data yang digunakan merupakan data Kartu Indonesia Pintar (KIP) tahun 2023, yang terdiri dari 33 karakteristik atau atribut yang akan digunakan untuk menilai kelayakan calon penerima bantuan KIP. Dataset ini terdiri dari 384 entri yang menerima bantuan KIP, dan 1291 entri yang tidak menerima bantuan KIP. Data ini dikelompokkan menjadi dua kategori atau label, yaitu “YES” untuk penerima bantuan kip dan “NO” untuk mereka yang tidak menerima bantuan KIP. Di bawah ini terdapat penjelasan mengenai atribut-atribut data yang diambil dari Dinas Sosial Kota Dumau, yang disajikan dalam Tabel 1.

Tabel 1. Keterangan Atribut Data KIP

Variabel	Kode	Keterangan	Deskripsi
Jumlah ART	AT1	Jumlah Anggota Rumah Tangga	Seluruh jumlah anggota rumah tangga di dalam 1 rumah
Sta Bangunan	AT2	Status penguasaan bangunan tempat tinggal yang di tempati	1. Milik sendiri 2. Kontrak/sewa 3. Bebas sewa 4. Dinas 5. Lainnya
Sta Lahan	AT3	Status lahan yang ditempati	1. Milik sendiri 2. Milik orang lain 3. Tanah negara

Variabel	Kode	Keterangan	Deskripsi
Lantai	AT4	Jenis lantai yang digunakan	4. Lainnya 1. Marmet granit 2. Keramik 3. Parket/virnil/permadani 4. Ubil/tegel/teraso 5. Kayu/papan kualitas tinggi 6. Semestara/bata 7. Babmbu 8. Kayu/papan 9. Tanah 10. Lainnya
Dinding	AT5	Jenis dinding yang digunakan	1. Tembok 2. Plasteran anyaman bambu 3. Kayu 4. Anyaman bambu 5. Batang kayu 6. Bambu 7. Lainnya
Kondisi dinding	AT6	Kondisi atap yang digunakan	1. Bagus/kualitas tinggi 2. Jelek/kualitas rendah
Atap	AT7	Jenis atap yang digunakan	1. Beton/genteng beton 2. Genteng keramik 3. Gentel metal 4. Genteng tanah liat 5. Asbes 6. Seng 7. Sirap 8. Bambu 9. Jerami/ijuk/daun 10. Lainnya
Kondisi Atap	AT8	Kondisi atap yang digunakan	1. Bagus/kualitas tinggi 2. Jelek/kualitas rendah
Jumlah Kamar	AT9	Berapa banyak kamar tidur yang dimiliki keluarga	Menyesuaikan jumlah kamar pada masing-masing rumah
Sumber Air Minum	AT10	Sumber air minum	1. Air kemasan bermerek 2. Air isi ulang 3. Leding meteran 4. Leding eceran 5. Sumur bor/pompa 6. Sumur terlindung 7. Sumur tak terlindung 8. Mata air terlindung 9. Mata air tak terlindung 10. Air sungai/danau/waduk 11. air hujan 12. lainnya
Sumber Penerangan	AT11	Keluarga memiliki sumber penerangan utama di rumah	1. Listrik PLN 2. Listrik Non PLN 3. Bukan listrik
Daya	AT12	Keluarga memiliki daya yang terpasang di rumah	1. 450 watt 2. 2.900 watt 3. 1.300 watt 4. 2.200 watt 5. >2.200 watt 6. Tanpa meteran
BB Masak	AT13	Bahan bakar untuk memasak	1. Listrik 2. Gas >3kg 3. Gas 3kg 4. Gas kota/biogas 5. Minyak tanah

Variabel	Kode	Keterangan	Deskripsi
Fas BAB	AT14	Fasilitas penggunaan tempat buang air besar di rumah	6. Briket
			7. Arang
			8. Kayu bakar
			9. Tidak memasak di rumah
Kloset	AT15	Jenis kloset di rumah	1. Sendiri
			2. Bersama
			3. Tidak ada
			1. Leher angsa
Buang Tinja	AT16	Keluarga memiliki pembuangan air tinja	2. Plengsengan
			3. Cemplung/cubluk
			4. Tidak pakai
			1. Tangki
Tabung Gas	AT17	Keluarga memiliki tabung gas 5,5 kg atau lebih	2. Plengsengan
			3. Lubang tanah
			4. Kolam/sawah/sungai/danau/laut
			5. Pantai/tanah lapang/kebun
Lemari Es	AT18	Keluarga memiliki kulkas atau lemari es	1. Ya
			2. Tidak
AC	AT19	Keluarga memiliki AC	1. Ya
			2. Tidak
Pemanas Air	AT20	Memiliki pemanas air	1. Ya
			2. Tidak
Telepon	AT21	Memiliki telepon rumah (PTSN)	1. Ya
			2. Tidak
TV	AT22	Keluarga memiliki televisi	1. Ya
			2. Tidak
Emas	AT23	Keluarga memiliki emas atau perhiasan (senilai 10 gram emas)	1. Ya
			2. Tidak
Komputer/Laptop	AT24	Keluarga memiliki komputer atau laptop	1. Ya
			2. Tidak
Sepeda	AT25	Keluarga memiliki sepeda	1. Ya
			2. Tidak
Sepeda Motor	AT26	Keluarga memiliki kendaraan roda 2	1. Ya
			2. Tidak
Mobil	AT27	Keluarga memiliki mobil	1. Ya
			2. Tidak
Perahu	AT28	Keluarga memiliki perahu	1. Ya
			2. Tidak
Motor Tempel	AT29	Keluarga memiliki motor tempel	1. Ya
			2. Tidak
Kapal	AT30	Keluarga memiliki kapal	1. Ya
			2. Tidak
Aset Tak Bergerak	AT31	Keluarga memiliki aset tak bergerak	a. Lahan :
			1. Ya
			2. Tidak
			b. Rumah ditempat lain :
Rumah Lain	AT32	Keluarga memiliki aset rumah lain	3. Ya
			4. Tidak
Sta ART Usaha	AT33	Anggota rumah tangga memiliki usaha sendiri/bersama	1. Ya
			2. Tidak
Status	Label/Class	Yes yaitu menerima bantuan KIP, sedangkan NO yaitu tidak menerima bantuan KIP	YES
			NO

Pada Tabel 2 ini memperlihatkan data sampel calon penerima bantuan Kartu Indonesia Pintar (KIP).

Tabel 2. Data Sampel Calon Penerima KIP

NO	AT1	AT2	AT3	AT4	AT5	...	AT29	AT30	AT31	AT32	AT33	STATUS
D1	4	2	2	2	1	...	2	2	2	4	0	YES
D2	4	1	1	2	1	...	2	2	1	4	2	YES
D3	6	2	2	6	1	...	2	2	2	4	2	NO
D4	1	NUL	NUL	NUL	NUL	...	NUL	NUL	NUL	NUL	NUL	NO
		L	L	L	L	...	L	L	L	L	L	
D5	6	1	1	2	1	...	2	2	1	4	2	NO
D6	1	NUL	NUL	NUL	NUL	...	NUL	NUL	NUL	NUL	NUL	YES
		L	L	L	L	...	L	L	L	L	L	
D7	3	1	1	2	1	...	2	2	2	4	0	YES
D8	4	1	1	6	1	...	2	2	1	4	2	NO
D9	4	1	1	2	1	...	2	2	1	4	2	YES
D10	2	1	1	2	1	...	2	2	1	4	1	NO
...
D199	4	1	1	8	3	...	2	2	2	4	2	YES
8												

3.2 Praproses Data

3.2.1 Cleaning Data

Dataset yang awalnya berjumlah 1998 data, lalu pada penelitian ini dataset tersebut melakukan proses cleaning data, ada terdapat 323 data NULL yang dihapus dari dataset tersebut, sehingga dataset berjumlah 1675 data. Terdapat 384 warga yang berhak menerima KIP dan 1291 warga yang tidak berhak menerima KIP setelah dilakukannya proses cleaning pada data. Dataset tersebut memiliki 2 label yaitu “YES” ditujukan bagi penerima KIP dan “NO” bagi yang tidak mendapatkan KIP. Data sampel hasil proses cleaning data dapat dilihat pada Tabel 3.

Tabel 3. Hasil Cleaning Data

NO	AT1	AT2	AT3	AT4	AT5	...	AT29	AT30	AT31	AT32	AT33	STATUS
D1	4	2	2	2	1	...	2	4	2	4	0	YES
D2	4	1	1	2	1	...	2	4	1	4	2	YES
D3	6	2	2	6	1	...	2	4	2	4	2	NO
D4	6	1	1	2	1	...	2	4	1	4	2	NO
D5	3	1	1	2	1	...	2	4	2	4	0	NO
D6	4	1	1	6	1	...	2	4	1	4	2	YES
D7	4	1	1	2	1	...	2	4	1	4	2	YES
D8	2	1	1	2	1	...	2	4	1	4	1	NO
D9	4	1	1	2	1	...	2	4	1	4	2	YES
D10	1	1	1	6	1	...	2	4	1	4	2	NO
...
D1675	4	1	1	8	3	...	2	4	2	4	2	YES

3.2.2 Seleksi Fitur Information Gain

Setelah dilakukannya cleaning terhadap data, langkah selanjutnya adalah menggunakan seleksi fitur Information Gain. Untuk menyeleksi fitur-fitur atau atribut yang paling berpengaruh terhadap data calon penerima KIP. Adapun langkah pertamanya yaitu mencari nilai entropy pada semua atribut data. Penjelasan Rumus :

$$Entropy(S) = \sum_{i=1}^n -p_i \cdot \log_2 p_i \quad (2)$$

S : Himpunan Kasus

n : Jumlah Pastisi S

pi : Proporsi Si terhadap S

Adapun cara pengerjaannya sebagai berikut.

Total data (kasus) : 1675

Jumlah penerima bantuan KIP : 384

Jumlah yang tidak menerima bantuan KIP : 1291

$$Entropy(\text{total}) = \left(-\frac{384}{1675} \times \log_2 \left(\frac{384}{1675}\right)\right) + \left(-\frac{1291}{1675} \times \log_2 \left(\frac{1291}{1675}\right)\right) = 0,77670812$$

Lalu setelah mencari nilai entropy, kemudian langkah selanjutnya yaitu mencari nilai gain, rumusnya sebagai berikut.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (1)$$

Keterangan :

S = himpunan kasus

A = fitur

n = jumlah partisi atribut A

|S_i| = proporsi S_i terhadap S

|S| = jumlah kasus dalam S

Lalu setelah melakukan perhitungan mencari nilai gain, penelitian ini kemudian menggunakan tools Rapid Miner untuk melanjutkan perhitungan keseluruhan atribut. Hasil dari nilai gain atribut dapat dilihat pada tabel 4.

Tabel 4. Nilai Gain Atribut

Atribut	Nilai Gain
AT1	0.195
AT9	0.013
AT26	0.011
AT18	0.009
AT13	0.007
...	...
AT21	0.00

Lalu setelah mencari nilai gain pada setiap atribut, atribut yang memiliki nilai gain tertinggi akan digunakan untuk proses klasifikasi K-NN, penelitian ini akan menggunakan 5 atribut yang memiliki nilai gain paling besar. Berikut adalah tabel data yang telah menggunakan seleksi fitur Information Gain.

Tabel 5. Data Seleksi Fitur Information Gain

NO	AT1	AT9	AT13	AT18	AT26	STATUS
D1	4	2	3	3	3	YES
D2	4	3	3	3	3	YES
D3	6	1	8	3	4	NO
D4	6	3	3	3	3	NO
D5	3	2	3	3	3	NO
D6	4	2	5	3	3	YES
D7	4	3	3	3	3	YES
D8	2	2	3	3	3	NO
D9	4	3	3	3	3	YES
D10	1	3	3	4	4	NO
...
D1675	4	2	2	3	3	YES

3.3 Hasil Pemodelan Klasifikasi

3.3.1 Klasifikasi Manual K-Nearest Neighbor

Cara menghitung algoritma K-NN adalah yang pertama dengan menghitung jarak pada data yaitu dengan rumus, yaitu :

$$\sqrt{(a_1 - b_1)^2 + (a_2 - b_2)^2 + \dots + (a_n - b_n)^2} \quad (3)$$

Dimana a sebagai data uji dan b sebagai data uji. Contoh perhitungan manual pada data KIP sebagai penelitian ini dapat dihitung dari sampel 10 data pertama pada Tabel 6.

Tabel 6. Data Sampel

AT1	AT9	AT13	AT18	AT26	Status
4	2	3	3	3	YES
4	3	3	3	3	YES
6	1	8	3	4	NO
6	3	3	3	3	NO
3	2	3	3	3	NO
4	2	5	3	3	YES
4	3	3	3	3	YES
2	2	3	3	3	NO
4	3	3	3	3	?
1	3	3	4	4	?

Contoh perhitungan jarak metode K-NN adalah :

- 1). $((4-2)^2+(2-2)^2+(3-3)^2+(3-3)^2+(3-3)^2) = 2$
- 2). $((4-2)^2+(3-2)^2+(3-3)^2+(3-3)^2+(3-3)^2) = 2,3$
- 3). $((6-2)^2+(1-2)^2+(8-3)^2+(3-3)^2+(4-3)^2) = 6,5$
- 4). $((6-2)^2+(3-2)^2+(3-3)^2+(3-3)^2+(3-3)^2) = 4,1$
- 5). $((3-2)^2+(2-2)^2+(3-3)^2+(3-3)^2+(3-3)^2) = 1$
- 6). $((4-2)^2+(2-2)^2+(5-3)^2+(3-3)^2+(3-3)^2) = 2,8$
- 7). $((4-2)^2+(3-2)^2+(3-3)^2+(3-3)^2+(3-3)^2) = 2,2$
- 8). $((2-2)^2+(2-2)^2+(3-3)^2+(3-3)^2+(3-3)^2) = 0$

Lalu setelah mendapatkan nilai jarak pada data sampel, langkah selanjutnya yaitu menghitung nilai k. Penelitian ini menggunakan perhitungan di Ms. excel dengan rumus/formula seperti berikut :

=IF(baris satu nilai jarak <=SMALL(blok semua kolom jarak;nilai k);baris satu atribut kelas,"")

Jika nilainya bukan yang terkecil, maka kolom pada Ms. Excel akan kosong dan jika nilainya adalah nilai terkecil maka akan muncul atribut kelasnya. Contoh penentuan nilai k pada Ms. Excel dapat lihat pada Tabel 7.

Tabel 7. Penentuan Nilai K

K1	K3	K5	K7
YES	YES	YES	YES
	YES	YES	
			NO
NO	NO	NO	NO
			YES
		YES	YES
NO	NO	NO	NO

3.3.2 Klasifikasi K-Nearest Neighbor Pada Rapid Miner

Hasil klasifikasi metode K-NN dapat dilihat pada tabel berikut.

Tabel 8. Hasil Klasifikasi K-NN

Menggunakan Information Gain	Rasio	Akurasi
	7:3	98.21%
	8:2	98.21%
	9:1	97.01%
Tidak Menggunakan Information Gain	7:3	91.04%
	8:2	92.84%
	9:1	89.82%

Dari tabel di atas, dapat dilihat bahwa klasifikasi metode K-NN mendapatkan akurasi terbesar dengan menggunakan seleksi fitur Information Gain, yang mana pada tahap pertama dengan pembagian data 7:3 mendapatkan akurasi 98,21%, lalu pada tahap percobaan kedua dengan pembagian data 8:2 tidak terjadi peningkatan terhadap nilai akurasi, dan pada percobaan ketiga menggunakan pembagian data 9:1 terjadi penurunan nilai akurasi sebanyak 1,20% dengan nilai akurasi yaitu 97,01%. Pada tahap percobaan klasifikasi K-NN dengan tidak menggunakan seleksi fitur Information Gain, tahap ini mengalami penurunan nilai akurasi yang cukup jauh, percobaan pertama dengan pembagian data sebanyak 7:3 mendapatkan akurasi sebesar 91,04%, lalu percobaan kedua dengan pembagian data 8:2 mengalami kenaikan akurasi sebanyak 1,80% dengan nilai akurasi yaitu 92,84%, dan pada tahap percobaan ketiga dengan pembagian data 9:1 mengalami penurunan nilai akurasi sebanyak 3,02% dengan nilai akurasi yaitu 89,82%.

3.3.3 Klasifikasi K-Nearest Neighbor Pada Python

Klasifikasi algoritma K-NN pada Python menggunakan k = 5 dan rasio 8:2 menghasilkan diagram confusion matrix yang mendapatkan 393 pada label “YES” tetapi ada 8 label “NO” dan juga terdapat 4 label “NO” dibagian label “YES” lalu 98 di bagian label “NO”. Hasil tersebut tertera pada gambar 3.

393	4
8	98

Gambar 3. Diagram Confusion Matrix pada Python

Lalu pada hasil klasifikasi di python juga mendapatkan akurasi sebesar 0.98 atau 98%, presisi label “YES” sebesar 96% dan label “NO” sebesar 98% dan recall yang didapatkan sebesar 0.92 atau 92% pada bagian label “YES” dan 0.99 atau 99% pada bagian label “NO”. Hasil klasifikasi tertera pada Gambar 4.

	precision	recall	f1-score	support
NO	0.98	0.99	0.98	397
YES	0.96	0.92	0.94	106
accuracy			0.98	503
macro avg	0.97	0.96	0.96	503
weighted avg	0.98	0.98	0.98	503

Gambar 4. Hasil Klasifikasi K-NN Pada Python

4. KESIMPULAN

Dari hasil klasifikasi yang diperoleh dalam penelitian ini, terlihat bahwa penggunaan seleksi fitur Information Gain menghasilkan akurasi yang lebih tinggi menggunakan jarak $k = 5$ sebesar 98,21% pada pembagian data atau rasio 7:3 dan 8:2, dibandingkan dengan klasifikasi K-NN yang tidak menggunakan seleksi fitur Information Gain hanya mendapatkan nilai akurasi tertinggi pada rasio 8:2 sebesar 92,84%. Lalu perbandingan nilai klasifikasi pada tools Python juga mengalami sedikit perbedaan nilai yang mendapatkan akurasi sebesar 98% dengan menggunakan rasio 8:2 dan jarak nilai k yang sama. Kesimpulannya adalah seleksi fitur Information Gain mempengaruhi nilai tingkat akurasi karena seleksi memilih fitur yang terbaik dan spesifik sehingga nilai akurasi dapat meningkat.

REFERENCES

- [1] E. V. Manoppo, N. A. Laoh, I. Pemerintahan, D. Negeri, P. Provinsi, dan S. Utara, “Strategi Pemanfaatan Data Terpadu Kesejahteraan Sosial (DTKS) Dalam Penyaluran Bantuan Sosial RS-RTLH Oleh Dinas Sosial Provinsi Sulawesi Utara Indonesia sebagai negara demokrasi memiliki tujuan salah satunya yakni untuk memajukan serta wajib menjamin b,” vol. 4, no. April, hal. 25–39, 2022.
- [2] S. J. Ontak, E. Israel, dan A. Kamal, “Politeknik Negeri Nusa Utara Sistem Pendukung Keputusan Penerima Kartu Indonesia Pintar (KIP) Dengan Metode Simple Additive Weighting (SAW) (Studi Kasus: SD GMIST Petra Nagha),” J. Ilm. Behongang, vol. 5, no. 1, hal. 1–6, 2022.
- [3] V. A. Saputri, “Implementasi Kebijakan Kartu Indonesia Pintar (Kip) Dalam Upaya Mengurangi Anak Putus Sekolah Di Smp N 1 Poncowarno,” Spektrum Anal. Kebijak. Pendidik., vol. 10, no. 1, hal. 99–111, 2021, doi: 10.21831/sakp.v10i1.17349.
- [4] S. Syaefuddin, L. Yuliani, dan L. Oktiawanti, “Pemberian Kartu Indonesia Pintar (KIP) Dalam Upaya Meningkatkan Motivasi Belajar Bagi Peserta Didik di PKBM GEMA Pada Pendidikan Kesetaraan Program Paket B dan C Kecamatan Tawang Kota Tasikmalaya,” Comm-Edu (Community Educ. Journal), vol. 2, no. 2, hal. 147, 2019, doi: 10.22460/comm-edu.v2i2.2611.
- [5] C. R. Raharya, N. Hidayat, dan E. Santoso, “Penentuan Penerimaan Beasiswa Menggunakan Metode Modified K-Nearest Neighbor | Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer,” J. Pengemb. Teknol. Inf. dan Ilmu Komput., vol. 2, no. 11, hal. 4984–4990, 2018, [Daring]. Tersedia pada: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/3158>
- [6] A. Khairi, “Implementasi K-Nearest Neighbor (KNN) untuk Klasifikasi Masyarakat Pra Sejahtera Desa Sapikerap Kecamatan Sukarapu,” J. TRILOGI, vol. 2, no. 3, hal. 319–323, 2021, [Daring]. Tersedia pada: <https://ejournal.unuja.ac.id/index.php/trilogi/article/view/2878>
- [7] A. Firdaus, “Aplikasi Algoritma K-Nearest Neighbor pada Analisis Sentimen Omicron Covid-19,” J. Ris. Stat., hal. 85–92, 2022, doi: 10.29313/jrs.v2i2.1148.
- [8] S. Margareta, I. Arwani, dan D. E. Ratnawati, “Implementasi Algoritma K-Nearest Neighbor Pada Database Menggunakan Bahasa SQL,” vol. 4, no. 7, hal. 2043–2052, 2020, [Daring]. Tersedia pada: <http://j-ptiik.ub.ac.id>
- [9] A. D. Ghani, N. Salman, dan Mustikasari, “Algoritma k-Nearest Neighbor Berbasis Backward Elimination Pada Client Telemarketing,” Pros. Semin. Ilm. Sist. Inf. dan Teknol. Inf., vol. 8, no. 2, hal. 141–150, 2019.
- [10] I. K. Hasan, R. Resmawan, dan J. Ibrahim, “Perbandingan K-Nearest Neighbor dan Random Forest dengan Seleksi Fitur Information Gain untuk Klasifikasi Lama Studi Mahasiswa,” Indones. J. Appl. Stat., vol. 5, no. 1, hal. 58, 2022, doi: 10.13057/ijas.v5i1.58056.
- [11] J. Homepage, S. R. Cholil, T. Handayani, R. Prathivi, dan T. Ardianita, “IJCIT (Indonesian Journal on Computer and Information Technology) Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa,” IJCIT (Indonesian J. Comput. Inf. Technol., vol. 6, no. 2, hal. 118–127, 2021.
- [12] V. Alvian, D. Hidayatullah, A. Nilogiri, H. Azizah, dan A. Faruq, “Klasifikasi Siswa Berprestasi Menggunakan Metode K-Nearest Neighbor (KNN) Pada SMA Negeri 2 Situbondo Classification Of Achieving Students Using K-Nearest Neighbor (KNN) Method At SMA Negeri 2 Situbondo,” J. Smart Teknol., vol. 3, no. 6, hal. 2774–1702, 2022, [Daring]. Tersedia pada: <http://jurnal.unmuhjembar.ac.id/index.php/JST>
- [13] W. Suci dan F. M. Basysyar, “Klasifikasi Data Bantuan Sosial pada Desa Sindangpano dengan Menggunakan Algoritma K-Nearest Neighbor,” vol. 5, no. 2, hal. 167–174, 2022.
- [14] S. H. A. Aini, Y. A. Sari, dan A. Arwan, “Seleksi Fitur Information Gain untuk Klasifikasi Penyakit Jantung Menggunakan Kombinasi Metode K-Nearest Neighbor dan Naïve Bayes,” J. Pengemb. Teknol. Inf. dan Ilmu Komput., vol. 2, no. 9, hal. 2546–2554, 2018, [Daring]. Tersedia pada: <http://j-ptiik.ub.ac.id>



- [15] M. R. Hasibuan, “Pemilihan Fitur dengan Information Gain untuk Klasifikasi Penyakit Gagal Ginjal menggunakan Metode Modified K-Nearest Neighbor (MKNN),” vol. 3, no. 11, hal. 3659–875, 2019, [Daring]. Tersedia pada: <http://j-ptiik.ub.ac.id>
- [16] S. N. Salsabila, B. N. Sari, R. Mayasari, I. Komputer, dan U. S. Karawang, “KLASIFIKASI ULASAN PENGGUNA APLIKASI DISCORD MENGGUNAKAN,” vol. 9, no. 2, hal. 383–392, 2023.
- [17] N. Hasdyna, E. B. Nababan, dan S. Effendi, “Dimension Reduction in Datasets Using Information Gain To Enhance K-NN Performance,” *Int. J. Trend Res. Dev.*, vol. 6, no. 3, hal. 2394–9333, 2019, [Daring]. Tersedia pada: www.ijtrd.com
- [18] R. P. Fitrianti, A. Kurniawati, dan D. Agusten, “A-27 Implementasi Algoritma K-Nearest Neighbor Terhadap Analisis Sentimen Review Restoran Dengan Teks Bahasa Indonesia,” *Semin. Nas. Apl. Teknol. Inf.*, hal. 1907–5022, 2019.