

Klasifikasi Tingkat Pendidikan Orang Tua Berdasarkan Tingkat Kelulusan menggunakan Algoritma K-NN

Hasda Surya Putra^{1*}, Vincentius Anggarda P.M², Umami Kultsum³, Dwi Hartanti⁴

^{1,2,3,4}Teknik Informatika

Universitas Duta Bangsa Surakarta

^{1*}202020238@mhs.udb.ac.id, ²202021153@gmail.com, ³202020932@mhs.udb.ac.id · ⁴dwhartanti@udb.ac.id

Abstrak— Klasifikasi tingkat pendidikan orang tua berdasarkan tingkat kelulusan merupakan topik yang penting dalam bidang pendidikan dan sosiologi. Memahami tingkat pendidikan orang tua dapat memberikan wawasan tentang faktor-faktor sosial dan ekonomi yang mempengaruhi keberhasilan pendidikan anak-anak. Metode yang digunakan dalam penelitian ini adalah algoritma K-NN. Algoritma K-NN memanfaatkan prinsip dasar bahwa data yang serupa cenderung berkumpul bersama. Dalam konteks ini, tingkat kelulusan orang tua digunakan sebagai fitur untuk mengklasifikasikan tingkat pendidikan mereka.

Kata kunci— K-Nearest Neighbour, Klasifikasi, Tingkat Pendidikan Orang tua, RapidMiner

Abstract— The classification of parents' educational level based on graduation level is an important topic in the field of education and sociology. Understanding the level of education of parents can provide insight into the social and economic factors that influence children's educational success. The method used in this study is the K-NN algorithm. The K-NN algorithm exploits the basic principle that similar data tends to cluster together. In this context, the pass rate of parents is used as a feature to classify their level of education. **Keywords**— K-Nearest Neighbor, Classification, Parent level of Education, RapidMiner

I. PENDAHULUAN

Tingkat pendidikan orang tua merupakan faktor penting dalam mempengaruhi perkembangan dan kesuksesan anak-anak mereka dalam bidang pendidikan. Dalam analisis sosial ekonomi, klasifikasi tingkat pendidikan orang tua dapat memberikan wawasan yang berharga dalam memahami pola-pola sosial dan ekonomi dalam suatu populasi. Selain itu, informasi ini juga dapat digunakan dalam perencanaan pendidikan dan pengambilan kebijakan publik yang lebih efektif.

Salah satu metode yang dapat digunakan untuk mengklasifikasikan tingkat pendidikan orang tua berdasarkan tingkat kelulusan mereka adalah algoritma K-Nearest Neighbors (K-NN). Algoritma K-NN adalah salah satu algoritma klasifikasi yang populer dalam bidang pembelajaran mesin. Metode ini berdasarkan prinsip dasar bahwa data yang serupa cenderung berkumpul bersama.

Dalam konteks ini, tingkat kelulusan orang tua dapat digunakan sebagai fitur untuk mengklasifikasikan tingkat pendidikan mereka.

Algoritma K-NN akan membandingkan tingkat kelulusan orang tua yang akan diklasifikasikan dengan tingkat kelulusan orang tua pada data pelatihan yang sudah ada. Dengan mempertimbangkan k-neighbors terdekat (k tetangga terdekat), algoritma ini akan mengklasifikasikan tingkat pendidikan orang tua yang baru berdasarkan mayoritas tingkat pendidikan tetangga terdekatnya.

Penelitian ini bertujuan untuk menerapkan algoritma K-NN dalam mengklasifikasikan tingkat pendidikan orang tua berdasarkan tingkat kelulusan mereka. Dengan menggunakan dataset yang terdiri dari tingkat kelulusan dan tingkat pendidikan orang tua yang dikumpulkan dari berbagai sumber, penelitian ini akan membagi dataset menjadi data pelatihan dan data uji. Data pelatihan akan digunakan untuk melatih model K-NN, sedangkan data uji akan digunakan untuk mengevaluasi kinerja model.

Diharapkan bahwa penelitian ini akan memberikan pemahaman yang lebih baik tentang hubungan antara tingkat pendidikan orang tua dan

tingkat kelulusan mereka. Hasil klasifikasi ini dapat memberikan wawasan yang berharga dalam analisis sosial ekonomi dan dapat digunakan dalam perencanaan pendidikan serta pengambilan kebijakan publik yang lebih akurat dan efektif.

Dalam penelitian ini, akan dibahas langkah-langkah implementasi algoritma K-NN, pengolahan data, dan evaluasi kinerja model. Selain itu, juga akan dibahas potensi dan manfaat dari hasil klasifikasi tingkat pendidikan orang tua berdasarkan tingkat kelulusan menggunakan algoritma K-NN dalam konteks sosial ekonomi dan pendidikan..

II. METODOLOGI PENELITIAN

Penelitian ini menggunakan metode kualitatif dengan menggunakan algoritma K-Nearest Neighbor untuk menentukan hasil klasifikasi berupa tingkat 12 xcel 12 ikan orang tua berdasarkan tingkat kelulusan.

Algoritma K-Nearest Neighbor (KNN) merupakan sebuah metode yang digunakan untuk mengklasifikasikan obyek baru berdasarkan (K) tetangga terdekatnya. KNN termasuk algoritma supervised learning, yang mana hasil dari query instance yang baru, diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Kelas yang memiliki kemunculan terbanyak akan menjadi hasil klasifikasi.

Rumus untuk melakukan penghitungan kedekatan antara kedua dua kasus adalah sebagai berikut :

$$\frac{\sum_{i=1}^n f(T_i, S_i) * w_i}{w_i} \quad (1)$$

Dimana :

T : kasus baru

S : kasus dalam penyimpanan

n : jumlah atribut tiap kasus

i : atribut individu 1 s/d n

f : fungsi kesamaan atribut i antara kasus T dan S

w : bobot pada atribut yang ke i

Urutan Algoritma KNN

1. Menentukan parameter K (jumlah tetangga paling dekat)

2. Menghitung kuadrat jarak euclidian (euclidean distance) masing-masing obyek terhadap data sampel yang diberikan dengan rumus :

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

3. Mengurutkan objek-objek tersebut ke dalam kelompok yang mempunyai jarak euclid terkecil
4. Mengumpulkan kategori Y (klasifikasi nearest neighbor)
5. Dengan menggunakan kategori mayoritas, maka dapat diprediksikan nilai query distance yang telah dihitung.

RapidMiner merupakan 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).

III. HASIL DAN PEMBAHASAN

Pengujian K-Nearest Neighbor Pada proses ini adalah data graduation yang akan dibandingkan dengan data training yang sudah ada. Data Set yang digunakan sebanyak 200 data yang dapat dilihat pada table berikut:

Tabel 1. Data Latih *Graduation Rate*

ACT composite score	SAT total score	parent level of education	parent income	High school gpa	college gpa	years to graduate
22	1625	high school	40999	3	3.1	7
29	2090	associate's degree	75817	4	3.4	5
30	2188	bachelor's degree	82888	4	3.9	3
33	2151	associate's degree	93518	4	3.7	5
29	2050	associate's degree	79153	4	3.4	6
28	1976	master's degree	100048	3.8	3.5	4
29	2097	some college	46883	4	3.4	6
30	1976	some college	67379	3.7	3.4	5
27	2072	bachelor's degree	02424	3.9	3.9	4
32	2246	bachelor's degree	56793	4	3.6	3
...
27	1881	high school	50248	3.4	3.1	4

Dengan menggunakan data latih :

Tabel 2. Data Testing *Graduation Rate*

DATA GRADUATION RATE						
ACT composite score	SAT total score	parent level of education	parent income	High school gpa	college gpa	years to graduate
33	2314	master	10042 8	4	3.8	4

A. Perhitungan Manual dengan Excel

Mengubah data *Training* dengan menormalisasikannya. Menggunakan rumus:

$$= (\text{baris 1 atribut jumlah MK} - \text{MIN}(\text{blok kolom jumlah MK})) / (\text{MAX}(\text{blok kolom jumlah MK}) - \text{MIN}(\text{blok kolom jumlah MK}))$$

Tabel 3. Hasil perhitungan normalisasi

DATA GRADUATION RATE					
ACT composite score	SAT total score	parent income	high school gpa	college gpa	years to graduate
0.000	0.000	0.182	0.1	0.308	0.667
0.636	0.638	0.578	1.1	0.538	0.333
0.727	0.772	0.659	1.1	0.923	0.000
1.000	0.722	0.780	1.1	0.769	0.333
0.636	0.583	0.616	1.1	0.538	0.500
0.545	0.481	0.854	0.9	0.615	0.167
0.636	0.647	0.249	1.1	0.538	0.500
0.727	0.481	0.482	0.8	0.538	0.333
0.455	0.613	0.881	1	0.923	0.167
0.909	0.852	0.361	1.1	0.692	0.000
...
1.000	0.945	0.858	1.1	0.846	0.167

Menghitung Euclidean Distance atau jarak.

Jarak Euclidean diformulasikan untuk mencari jarak antara 2 titik dalam ruang dua dimensi. Berikut rumusnya:

$$= \text{SQRT}(\sum(\text{baris 1 Data Uji} - \text{Dasta latih})^2)$$

Tabel 4. Hasil perhitungan Euclidean Distance

Distance
1.973
0.654
0.422
0.300
0.727
0.718
0.894
0.802
0.651
0.562
...
1.275

Menghitung nilai k atau nilai tetangga terdekat.

Tahapan ini adalah menentukan atau pemilihan nilai distance terkecil yang masuk peringkat sejumlah nilai k. Jika hasilnya bukan nilai terkecil maka kolom akan kosong dan jika nilai adalah nilai terkecil maka muncul atribut kelas nya. Berikut rumus yang digunakan:

$$= \text{IF}(\text{Baris satu nilai Distance} \leq \text{SMALL}(\text{Blok semua kolom Distance, Nilai k}), \text{Baris satu atribut Kelas, ""})$$

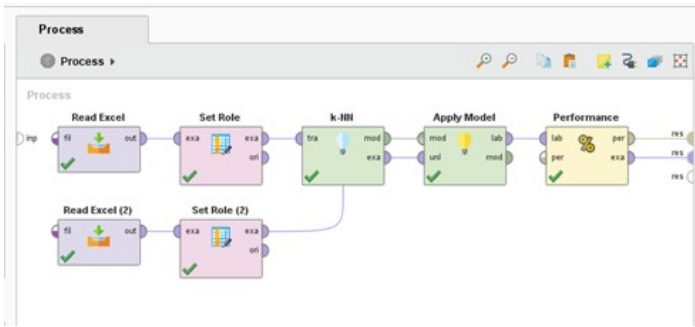
Tabel 5. Pengambilan data sejumlah nilai k

K=1	K=5	K=9
...

associate's degree

B. Pengujian dengan menggunakan RapidMiner

Berikut tampilan pengujian dengan RapidMiner :



Gambar 1. Desain Pengujian K-NN dengan Rapidminer

Gambar 2. Hasil dari pengujian K-NN

Hasil Pengujian pada Gambar 2 dengan nilai K=5, kolom parental level of education dan prediction (parenta level of education) merupakan prediksi tingkat pendidikan orang tua berdasarkan tingkat kelulusannya. Hasil Performace vektor dapat dilihat pada gambar 3

accuracy: 62.00%

	true high school	true associate's d.	true bachelor's de.	true master's degr.	true some college	true some high sc.	class precision
pred. high school	24	0	1	3	4	5	64.86%
pred. associate's ...	5	21	2	2	2	3	60.00%
pred. bachelor's d.	0	4	13	3	1	2	56.52%
pred. master's de.	0	0	2	9	0	0	81.82%
pred. some college	7	8	5	3	41	4	60.29%
pred. some high s.	3	1	1	0	5	16	61.54%
class recall	61.54%	61.76%	54.17%	45.00%	77.36%	53.33%	

Gambar 3. Hasil Performace Vector

Hasil pengujian dari gambar 3 dapat dilihat akurasi adalah 62,00%. Terdapat Perhitungan kedekatan kasus lama pada data training dengan kasus baru pada data testing, diketahui dari 200 record data,

- 39 data disklasifikasikan true high school,
- 34 data diklasifikasikan true associate's degree,
- 24 data diklasifikasikan true bachelor's degree,
- 20 data diklasifikasikan true master's degree,
- 53 data diklasifikasikan true some college, dan
- 30 data diklasifikasikan true some high school.

Dari hasil pengujian dan analisa dapat dilihat bahwa dari 200 data yang digunakan sesuai dengan prediksi. Tingkat akurasi penerapan algoritma k-nearest neighbor ini sebesar 62.00% jika K nya 5

IV. 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 tingkat pendidikan orang tua berdasarkan tingkat kelulusannya. 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 pengujian akurasi pada data kelulusan dan menghitung klasifikasi tingkat pendidikan orang
3. Hasil perhitungan

Pada data kelulusan tersebut didapatkan bahwa tingkat pendidikan orang tua pada akurasi yang telah ditetapkan yakni 62%. Maka hasil dari tingkat Pendidikan orang tua berdasarkan tingkat kelulusan ialah

1. 64.86% High School
2. 60.00% Associate's Degree
3. 56.52% Bachelor's Degree
4. 81.82% Master's Degree
5. 60.29% Some College
6. 61.54% Some Highschool

Dengan membandingkan akurasi yang didapat maka tingkat Pendidikan pada orang tua pada data tersebut ialah Highschool dengan 64.86% dan Master's Degree dengan 81.82%

REFERENSI

- [1] Ardiyansyah, Panny Agustia Rahayuningsih, and Reza Maulana. 2018. "Analisis Perbandingan Algoritma Klasifikasi Data Mining Untuk Dataset Blogger Dengan Rapid
- [2] 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.
- [3] 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.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] Wijayaningrum, V. N. 2018. "Optimasi Aturan Fuzzy Menggunakan Particle Swarm Optimization Untuk Penentuan Prioritas Rumah Tangga Miskin." *Sentia* 2018 (June)
- [8] 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.
- [9] Ernawati, Siti, and Risa Wati. 2018. "Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel." *Jurnal Khatulistiwa Informatika* VI(1):64–69.
- [10] Buulolo, E. 2020. *Data Mining Untuk Perguruan Tinggi*. Deepublish.