JIKO (JURNAL INFORMATIKA DAN KOMPUTER)
Juni 2025, Volume: **9**, No. **2** | Pages 352–360
doi: 10.26798/jiko.v9i2.1605
e-ISSN : 2477-3964 – p-ISSN : 2477-4413

ARTICLE

Komparasi Kinerja Label-Encoding dengan One-Hot-Encoding pada Algoritma K-Nearest Neighbor menggunakan Himpunan Data Campuran

(Studi Kasus : Kelulusan Mahasiswa Program Sarjana)

Performance Comparison of Label-Encoding with One-Hot-Encoding Methode on K-Nearest Neighbor Algorithm with Mixed Type Data Set

(Case Study: Graduation of Undergraduate Students)

Mohammad Guntara[1] dan Femi Dwi Astuti[2]

[1]Teknik Komputer, Fakultas Teknologi Informasi, Universitas Teknologi Digital Indonesia, Yogyakarta, Indonesia
[2]Informatika, Fakultas Teknologi Informasi, Universitas Teknologi Digital Indonesia, Yogyakarta, Indonesia

(Disubmit 17-12-24; Diterima 10-03-25; Dipublikasikan online pada 20-06-25)

Abstrak

Algoritma K-Nearest Neighbor (KNN) adalah salah satu algoritma dalam *data mining*, bekerja didasarkan pada pengukuran jarak antara tupel pada data-uji dan masing-masing data-latih untuk memutuskan luaran klasifikasi akhir. Pada algoritma ini jenis data pada tiap atribut harus berupa numerik atau kontinyu, karena dihitung jarak setiap atribut yang sama untuk setiap tupel dengan tupel yang akan dicari kelasnya. Akan tetapi pada realitanya himpunan-data yang akan diolah tidak selalu berupa numerik, tetapi dapat berupa kategorikal baik nominal maupun ordinal.Untuk data numerik dapat langsung diolah tanpa proses transformasi(pengodean), Pada jenis data nominal pengodean ke numerik dilakukan dengan metode label umumnya berupa nomor urut, namun hal ini dirasa kurang tepat mengingat nomor urut sebetulnya berhirarki atau memiliki kuantitas, dimana angka yang lebih besar memiliki kuantitas yang lebih besar pula, sehingga memungkinkan terjadinya bias dalam pengodean. Untuk itulah digunakan pengodean dengan metode *one-hot-encoding* (OHE) dimana setiap item data pada suatu atribut dikonversi ke bit 1 dan 0 yang menjadikan antar obyek pada himpunan-data setara. Untuk mengetahui sejauh mana akurasi transformasi OHE dibanding transformasi label diimplementasikan pada algoritma KNN untuk prediksi kelulusan mahasiswa dimana terdapat 2 jenis atribut yakni numerik dan nominal. Berdasarkan pengujian 2 metode pengodean tersebut diketahui bahwa akurasi transformasi dengan OHE pada berbagai nilai neighbor lebih tinggi dibanding metode label, sedangkan kecepatan proses kedua metode pengodean relatif sama Pada metode OHE akurasi tertinggi terjadi pada neighbor 1, sedangkan metode Label pada neighbor sama dengan 7. Dengan adanya akurasi yang lebih baik untuk OHE maka dalam memprediksi kelulusannyapun akan semakin baik.

Kata kunci: one-hot-encoding; label; pengodean; KNN

Abstract

The K-Nearest Neighbor (KNN) algorithm is one of the algorithms in data mining, working based on measuring the distance between the tuples on the test-data and each data-training to decide on the final classification output. In this algorithm, the type of data on each attribute must be numeric or continuous, because the distance of each attribute is equal to that of each tuple with the tuple to be searched for its class. However, in reality, the data sets to be processed are not always numerical, but can be categorical, both nominal and ordinal. For numerical data can be directly processed without a transformation process (encoding), In this type of nominal data, encoding to numerical is carried out by the label method, generally in the form of a sequence number, but

How to Cite: M. Guntara *et al.*, "Komparasi Kinerja Label-Encoding dengan One-Hot-Encoding pada Algoritma K-Nearest Neighbor menggunakan Himpunan Data Campuran (Studi Kasus : Kelulusan Mahasiswa Program Sarjana)", *JIKO (JURNAL INFORMATIKA DAN KOMPUTER)*, Volume: **9**, No.**2**, Pages 352–360, Juni 2025, doi: 10.26798/jiko.v9i2.1605.

this is considered inappropriate considering that the sequence number is actually hierarchical or has a quantity, where a larger number has a larger quantity, so that it allows bias in encoding. For this reason, encoding with the one-hot-encoding (OHE) method is used where each data item in an attribute is converted to bits 1 and 0 which make between objects in the data set equivalent. To find out the extent of the accuracy of the OHE transformation compared to the label transformation, it is implemented in the KNN algorithm to predict student graduation where there are 2 types of attributes, namely numeric and nominal. Based on the testing of the 2 coding methods, it is known that the transformation accuracy with OHE at various neighbor values is higher than that of the label method, while the process speed of the two coding methods is relatively the same In the OHE method the highest accuracy occurs in neighbor 1, while the Label method on the neighbor is equal to 7. With better accuracy for OHE, it will be better in predicting graduation.

KeyWords: one-hot-encoding; label; encoding; KNN

1. Pendahuluan

Algoritma K-Nearest Neighbor(KNN) adalah salah satu algoritma dalam *data mining* yang digunakan dalam klasifikasi dan prediksi numerik [1] . KNN bekerja dengan membandingkan tupel uji (*data testing*) dengan k-tupel pelatihan (*data training*) yang terdekat dalam ruang n-dimensi, menggunakan metrik jarak seperti jarak Euclidean atau Manhattan Distance. Dalam klasifikasi,tupel yang tidak dikenal diberi kelas yang paling umum di antara k tetangga terdekatnya, sementara dalam prediksi numerik, nilai rata-rata dari label bernilai nyata dari k tetangga terdekat digunakan[2]. Tupel pelatihannya adalah dijelaskan oleh n atribut. Setiap tupel mewakili sebuah titik dalam ruang berdimensi n. Dengan cara ini, semua tupel pelatihan disimpan dalam ruang pola berdimensi-n. Ketika diberikan tupel yang tidak diketahui, pengklasifikasi k-tetangga-terdekat (k-*neighbor* mencari ruang pola untuk k tupel pelatihan yang paling dekat dengan tupel yang tidak diketahui tersebut. Kedekatan didefinisikan dalam metrik jarak, seperti jarak Euclidean . Jarak Euclidean antara dua titik atau tupel ditunjukkan dengan persamaan [3]:

$$x_1 = (x_{11} - x_{12}, ..., x_{1n}) \tag{1}$$

dan

$$x_2 = (x_{21} - x_{22}, ..., x_{2n}) \tag{2}$$

sehingga , jarak antara tupel latih dengan tupel uji adalah

$$dist\,(X_1, X_2) = \sqrt{\sum_{i-1}^{n} (x_{1i} - x_{2i})^2} \tag{3}$$

Sementara itu himpunan-data (*dataset*) yang tersedia tidak selalu memiliki atribut numerik, tetapi berupa kategorikal atau kualitatif, sehingga harus atribut jenis ini harus dikonversi ke bentuk numerik. Jenis data atau atribut kategorikal memiliki 2 kategori : nominal dan ordinal[3]. Atribut nominal terdiri dari sekumpulan nilai diskrit terbatas tanpa hubungan antar nilai, sedangkan atribut ordinal terdiri dari sekumpulan nilai diskrit terbatas dengan urutan peringkat atau hirarki antar nilai. Transformasi ke numerik untuk atribut nominal dapat menggunakan pengodean dengan metode *one-hot-encoding* (OHE) , sementara jenis ordinal lebih tepat menggunakan pengodean metode label[4].

Pengodengan dengan metode One-Hot-Encoding merupakan sebuah teknik dalam mengubah bentuk data kategorik menjadi bentuk biner, dengan panjang sesuai dengan banyaknya kategori data yang berbeda . OHE dilakukan dengan cara mengubah nilai setiap kelas menjadi nilai biner, dengan indeks kelas yang bersangkutan diberi nilai 1 dan yang lain diberi nilai 0[5].

Secara struktur tabel, pengodean label tidak mengubah cacah kolom, karena pengodean dilakukan secara menggunakan numerik secara sekuen sedangkan pengodean OHE akan menambah cacah kolom sesuai item data yang dikodekan. Pengodean OHE ini menggunakan biner 0 dan 1. Nilai 1 bila sesuai dengan atribut-nya dan 0 untuk yang tidak sesuai, sehingga untuk jenis data nominal memiliki strata kode yang sama tidak berhirarki atau nilai berbeda untuk tiap item data yang dikodekan

Sebagai obyek untuk implementasi kedua metode pengodean di atas yaitu himpunan-data kelulusan mahasiswa perguruan tinggi dengan 3 kelas : lulus- tepat-waktu, terlambat, dan keluar (termasuk [?]). Disebut

lulus-tepat-waktu untuk program sarjana bila masa studi maksimal 4 tahun [6] , disebut terlambat bila dinyatakan lulus lebih dari 4 tahun sampai dengan 7 tahun, dan keluar bila masa studi lebih dari 7 tahun untuk program sarjana [7].

Adapun atribut himpunan-data yang akan diolah terdapat 2 jenis data, Untuk data berjenis numerik seperti : index prestasi (IP), jumlah sks, masa studi, jumlah cuti, tanpa konversi dapat digunakan untuk menghitung jarak pada metode KNN , akan tetapi untuk data yang berjenis kategorikal harus di konversi ke ke numerik dengan 2 metode di atas. Berdasarkan jenis data yang ada pada atribut kelulusan di atas yang berjenis kategorikal yakni : program studi, jenis kelamin, dan provinsi asal berjenis nominal sehingga pengodean dilakukan menggunakan OHE.

Metode KNN diimplementasikan dengan data numerik untuk memprediksi kelulusan mahasiswa dengan nilai K=5 hingga K=11 diperoleh hasil performa terbaik pada nilai K=11 dengan nilai akurasi sebesar 76%, presisi 70% dan recall 77% untuk klasifikasi tepat waktu dan tidak tepat waktu [8].

Metode KNN ini dikombinasikan dengan Metode Simple Additive Weighting untuk pengambilan keputusan seleksi penerimaan Paskibraka dengan atribut 8 kriteria utama menghasilkan akurasi 82,15% [9]. Teknik transformasi OHE dapat digunakan untuk mengklasifikasi risiko penularan COVID-19 menggunakan Algoritma K-Means. Metode OHE digunakan untuk transformasi data nominal dengan riset ini berupa 3 klaster (C1,C2,C3) untuk kabupaten/kota di Provinsi Riau dimana cacah klaster terbaik diuji menggunakan Silhouette Coefficient[9][10].

Metode OHE juga dapat diimplementasikan untuk mendeteksi peristiwa kebocoran data orang dalam organisasi terkait dengan keamanan siber.Hasil eksperimen menunjukkan kekokohan algoritma pembelajaran mesin *decission tree* dan *random forrest* dibandingkan dengan algoritma terapan lainnya. Pendekatan ini menunjukkan kinerja yang lebih baik dalam mendeteksi peristiwa kebocoran data orang dalam selama periode sensitif sementara orang dalam berencana untuk melakukan serangan dan meninggalkan lingkungan kerja organisasi sesudahnya. [11].

Metode OHE ini juga digunakan untuk membandingkan model *machine learning* dan *deep learning* untuk *pre-processing* datanya. *Convolutional Neural Network (CNN)* dapat dijalankan dengan baik untuk semua model, dengan akurasi 95,97%. Tetapi CNN dengan *Word Embedding* berkinerja lebih baik daripada menggunakan OHE(96,34%)[12] .

Dari berbagai referensi tersebut terlihat belum ada yang membandingkan kinerja 2 metode pengodean,sehingga pada riset ini dicoba untuk membandingkan kinerja yang dalam hal ini dari segi akurasi dan kecepatan prosesnya untuk algoritma KNN. Sebagai obyek adalah kelulusan mahasiswa program Sarjana, dengan parameter index prestasi mahasiswa selama 4 semester awal dan parameter lainnya yakni provinsi asal, program studi,dan jenis kelamin. Metode yang memiliki kinerja terbaik dapat digunakan untuk memprediksi kelulusan lebih akurat yang merupakan modal awal untuk melakukan pembinaan,sehingga dapat di minimalisir yang terlambat lulus atau keluar.

2. Metode

2.1 Persiapan dan transformasi data

Untuk menyelesaikan permasalahan terkait dengan penggunaan implementasi metode K-Nearest Neighbor (KNN) diperlukan persiapan data sebagai berikut [13] :

1. Pengumpulan Data. Pengumpulan data pada penelitian ini dilakukan secara langsung oleh *back office* dengan mengakses data pada server https://sia.utdi.ac.id. Dari *database server* dihasilkan himpunan data yang masih mentah dengan sekitar 27 atribut dan banyak terjadi redundansi karena setiap semester disimpan dalam 1 baris dan status kelulusan masih berupa tanggal lulus.
2. *Data Preprocessing*. *Data Preprocessing* bertujuan untuk mengubah data mentah menjadi data yang berkualitas sehingga data layak untuk diolah pada tahapan selanjutnya. *Data Preprocessing* dilakukan dengan

   (a) Berdasarkan data mentah dimana seorang mahasiswa akan direkam dalam 1 record per semester,

sehingga akan banyak redudansi data sejumlah semester yang ditempuh. Untuk itu perlu diakumulasi dalam 1 record dengan index prestasi dibuat rerata

(b) Mengeliminasi kolom/atribut yang tidak relevan untuk digunakan dalam proses komputasi yaitu : alamat, kota, pekerjaan, penghasilan.

3. Transformasi data.Transformasi data adalah tahapan di mana data diubah ke dalam bentuk yang sesuai dengan kebutuhan untuk siap diproses[?]. Pada data mentah sebagai kelas atau label berupa tanggal lulus sehingga perlu ditransformasikan ke 3 kelas : Tepat (lulus tepat waktu maksimal 4 tahun), Lambat (lulus lebih dari 4 tahun dan maksimal 7 tahun) , Keluar (untuk mahasiswa yang dinyatakan "keluar/*drop out*" atau masa studi lebih dari 7 tahun)

2.2 Implementasi Metode

Untuk *Dataset* yang sudah ditransformasikan perlu di konversi ke format CSV agar dapat diakses oleh sistem Python. Atribut yang digunakan dalam komputasi adalah : Nim, Prodi,Jenis,Provinsi,IPK,masa_studi, dan Status. Sebagai kelas adalah Status yang terdiri atas 3 jenis : Tepat, Lambat, Keluar. Dari *dataset* ini 85% digunakan untuk *data training* dan selebihnya merupakan *data testing*.

2.3 Diagram Aktivitas

Diagram aktivitas untuk metode KNN terlihat seperti Gambar 1 [14]

Gambar 1. Diagram Aktivitas Metode KNN

Penjelasan untuk diagram aktivitas pada Gambar 1 sebagai berikut.

1. Menentukan jumlah tetangga terdekat yang disimbolkan dengan nilai parameter K. Nilai pada parameter K yang akurat untuk algoritma ini tergantung pada *data training* yang digunakan.
2. Menghitung jarak tupel latih dengan tupel uji sesuai persamaan 3
3. Melakukan pengurutan dari hasil perhitungan no 2 secara urut naik (urut dari nilai rendah ke nilai tinggi);
4. Mengumpulkan data pada kategori Y (Klasifikasi tetangga terdekat berdasarkan nilai K);
5. Kategori Y yang paling banyak muncul menjadi hasil akhir dari klasifikasi.

2.4 Evaluasi dan Analisis data

Evaluasi untuk kinerja dilakukan dengan menggunakan *Confusion Matriks* untuk menghitung akurasi dan kecepatan proses kedua metode. *Confusion Matriks* ini digunakan untuk menghitung sejauh mana akurasi dari kedua metode tersebut sedangkan dan kecepatan proses (dalam milidetik) dihitung dari saat proses transformasi data sampai hasil komputasi dengan algoritma KNN selesai. Hasil dari kedua metode pengodean ini akan dianalisis akurasi dan kecepatannya dan dituangkan baik dalam bentuk tabel maupun grafik, yang untuk selanjutnya disimpulkan pola yang terjadi untuk kedua metode tersebut.

3. Hasil

Berdasarkan Tabel 1 terlihat bahwa sumber data masih memiliki atribut yang berjenis nominal. Pada algoritma KNN dikarenakan akan dihitung jarak antar obyek, maka harus dikonversi menjadi numerik yakni atribut Prodi, Jenis, dan Provinsi.

Tabel 1. Dataset Sumber Kelulusan S1 (740 record)

Id	NIM	Prodi	Jenis	Provinsi	IPK	Masa_Studi	Status
1	155410001	Informatika	P	Nusa Tenggara Barat	2.72	4.56	LAMBAT
2	155410002	Informatika	L	Jawa Timur	3.66	4.56	LAMBAT
3	155410003	Informatika	L	D.I. Yogyakarta	3.68	3.54	TEPAT
4	155410004	Informatika	L	D.I. Yogyakarta	3.51	4.05	TEPAT
5	155410005	Informatika	P	D.I. Yogyakarta	3.71	3.54	TEPAT
…	…	…	…	…	..	…	…
735	165610132	Sistem Informasi	P	D.I. Yogyakarta	2.97	1.51	TEPAT
736	165610133	Sistem Informasi	L	Sumatera Barat	0.00	-118.37	KELUAR
737	165610134	Sistem Informasi	L	Sumatera Selatan	0.00	-118.37	KELUAR
738	165610135	Sistem Informasi	L	Nusa Tenggara Barat	3.19	-118.37	KELUAR
739	165610136	Sistem Informasi	P	Bengkulu	0.74	-118.37	KELUAR
740	165610137	Sistem Informasi	L	DKI Jakarta	0.00	-118.37	KELUAR

Hasil transformasi atribut nominal ke jenis label terdapat pada Gambar 2, sedangkan transformasi ke jenis *one-hot-encoding* terdapat pada Gambar 3,

id	prodi	jenis	provinsi	ipk	masa_studi	status
0	0	1	20	2.72	4.56	LAMBAT
1	0	0	10	3.66	4.56	LAMBAT
2	0	0	4	3.68	3.54	TEPAT
3	0	0	4	3.51	4.05	TEPAT
4	0	1	4	3.71	3.54	TEPAT
..	..	..	..	...	...	...
735	1	0	30	0.00	-118.37	KELUAR
736	1	0	31	0.00	-118.37	KELUAR
737	1	0	20	3.19	-118.37	KELUAR
738	1	1	3	0.74	-118.37	KELUAR
739	1	0	5	0.00	-118.37	KELUAR

Gambar 2. Dataset hasil transformasi data Alphanumerik ke Numerik tipe Label

Setelah melalui proses pengolahan menggunakan algoritma KNN didapat hasil komputasi berupa tabel Akurasi dan Kecepatan komputasi untuk setiap cacah neighbor (1 sampai dengan 10). Untuk tipe Label terdapat pada Gambar 4, sedangkan tipe OHE terlihat pada Gambar 5.

	Prodi_Informatika	Prodi_Sistem Informasi	Sex_L	Sex_P	Prop_Bali
0	1	0	0	1	0
1	1	0	1	0	0
2	1	0	1	0	0
3	1	0	1	0	0
4	1	0	0	1	0
..	...	...	...	...	...
735	0	1	1	0	0
736	0	1	1	0	0
737	0	1	1	0	0
738	0	1	0	1	0
739	0	1	1	0	0

lanjutan

	Prop_Sumatera Utara	Prop_Timor Timur	ipk	masa_studi	status
0	0	0	2.72	4.56	LAMBAT
1	0	0	3.66	4.56	LAMBAT
2	0	0	3.68	3.54	TEPAT
3	0	0	3.51	4.05	TEPAT
4	0	0	3.71	3.54	TEPAT
..	...	...	...	...	...
735	0	0	0.00	-118.37	KELUAR
736	0	0	0.00	-118.37	KELUAR
737	0	0	3.19	-118.37	KELUAR
738	0	0	0.74	-118.37	KELUAR
739	0	0	0.00	-118.37	KELUAR

[740 rows x 41 columns]

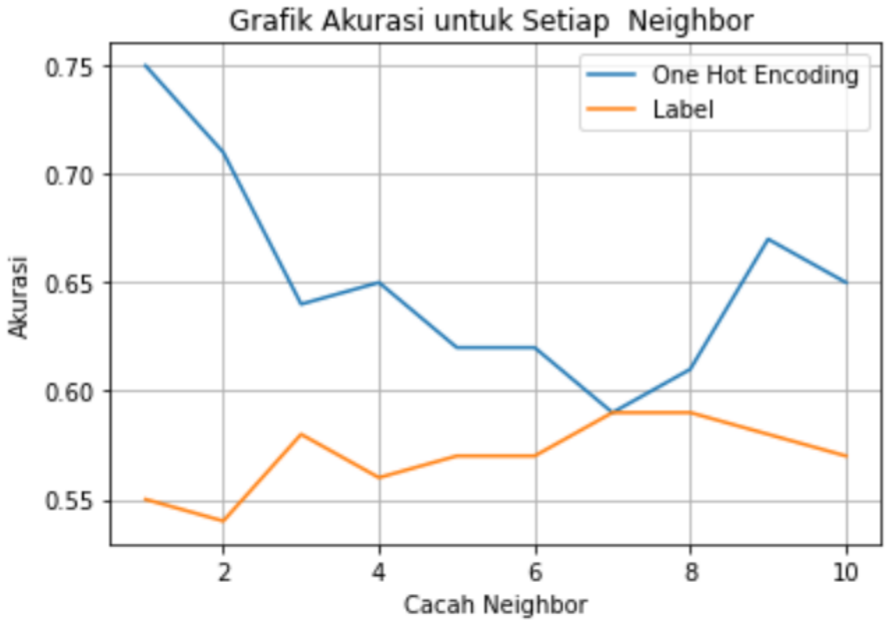
Gambar 3. Dataset hasil transformasi data Alphanumerik ke Numerik tipe One Hot Encoding (OHE)

Neighbor	Akurasi	Kecepatan (mdetik)
1	0.55	78.11
2	0.54	78.11
3	0.58	93.76
4	0.56	109.36
5	0.57	124.98
6	0.57	124.98
7	0.59	140.61
8	0.59	156.23
9	0.58	156.23
10	0.57	171.86

Gambar 4. Tabel hasil pengujian akurasi dan kecepatan pengodean Label

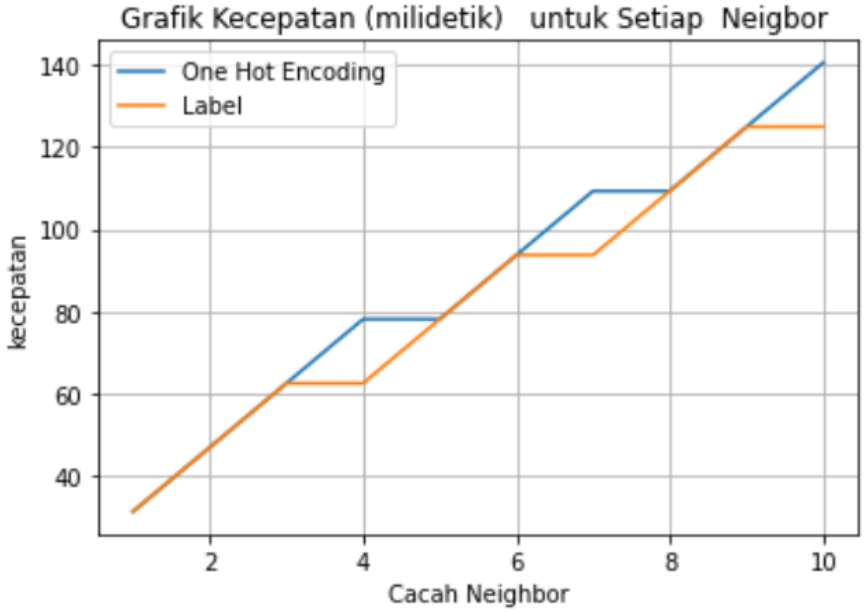
Neighbor	Akurasi	Kecepatan (mdetik)
1	0.75	46.87
2	0.71	62.50
3	0.64	62.50
4	0.65	78.12
5	0.62	93.75
6	0.62	93.75
7	0.59	109.37
8	0.61	125.00
9	0.67	125.00
10	0.65	140.62

Gambar 5. Tabel haisl pengujian akurasi dan Kecepatan untuk pengodean OHE



Gambar 6. Grafik Akurasi untuk pengodean Label dan OHE

Untuk mempermudah komparasi kinerja tipe Label dan OHE dibuat grafik, yakni grafik dengan sumbu x berupa Cacah Neighbor, dan sumbu y berupa Akurasi untuk kedua tipe terlihat pada Gambar 6, sedangkan dari segi komparasi kecepatan terlihat pada Gambar 7.



Gambar 7. Grafik Kecepatan untuk pengodean Label dan OHE

4. Pembahasan

Berdasarkan sumber data sudah siap di olah (*clean data*) seperti pada Tabel 1 dimana himpunan data ini masih memiliki 3 atribut yang berisi data nominal yakni Prodi, Jenis, dan Provinsi. Jenis dari ketiga atribut tersebut adalah nominal bukan merupakan hirarki atau peringkat,sehingga harus di transformasi-kan ke bentuk numerik, agar dapat diolah menggunakan algoritma KNN.

Gambar 2 menunjukkan bagaimana ke tiga atribut yang bersifat nominal di konversi menjadi data numerik dimulai dari 0, 1, dan seterusnya sesuai cacah atribut yang dikodekan dan pada Gambar 3 menunjukkan

bagaimana jenis data nominal menjadi numerik berupa angka 1 dan 0.

Cacah atribut hasil dari transformasi data metode label akan sama dengan cacah atribut himpunan data seperti terlihat pada tabel Gambar 2. Sedangkan pada metode one_hot_encoding (OHE) cacah atribut menjadi jauh lebih banyak. Hal ini dikarenakan cacah atribut/kolom dipengaruhi cacah obyek yang di kode-kan. Misal untuk Provinsi yang semula hanya 1 kolom menjadi 35 kolom sesuai cacah provinsi tersebut seperti terlihat pada tabel Gambar 3.

Berdasarkan hasil komputasi menggunakan algoritma KNN, dengan berbagai nilai neighbor (dari 1 sampai dengan 10) terlihat bahwa untuk **data label** akan memiliki akurasi tertinggi pada pada cacah *neighbor* 7 dengan akurasi 0.5, sedangkan pada OHE akurasi tertinggi pada cacah neighbor 1 dengan akurasi 0.75. Hal ini terlihat pada grafik Gambar 6.

Adapun untuk kecepatan proses seperti terlihat pada Gambar 7 semakin banyak cacah *neighbor* akan semakin lamban, namun keduanya memiliki sedikit perbedaan pada cacah *neighbor* 4, 6 dan 10.

5. Simpulan

Berdasarkan pembahasan terdahulu dapat disimpulkan: Akurasi transformasi dengan metode *one-hot-encoding* pada berbagai nilai *neighbor* lebih tinggi dibanding metode label, kecuali pada nilai *neighbor* 7, keduanya memiliki akurasi yang sama. Kecepatan proses kedua metode transformasi data relatif sama kecuali pada *neighbor* 4, 7, dan 10 terdapat sedikit perbedaan dimana metode Label sedikit lebih cepat. Pada metode OHE akurasi tertinggi terjadi pada *neighbor* 1, sedangkan metode Label pada *neighbor* 7. Dari segi jumlah kolom, maka metode Label lebih sederhana atau sesuai dengan kolom semula sedangkan metode OHE akan jauh lebih banyak tergantung dari obyek data yang dikodekan. Pada riset ini akurasi lebih rendah dibanding dibandingkan dengan [15] yakni 96,2%, menggunakan algoritma yang sama, untuk memprediksi kelulusan, dimana pada [15] obyek risetnya adalah Program Diploma dengan atribut: ip semester dan ip komulatif . Sementara itu pada riset ini terdapat 1 atribut numerik ip komulatif dan 3 atribut tipe kategorikal yang harus dikodekan yakni : jenis kelamin, asal propinsi, dan program studi. Dengan demikian barangkali adanya 3 atribut ini tidak siginifikan mendukung akurasi.

Berkenaan dengan pembahasan terdahulu maka disarankan Perlu adanya kajian lebih lanjut sejauh mana presisi, *recall*, dan *F1-score* pada kedua metode pengodean tersebut. Untuk mengolah *dataset* yang besar perlu metode untuk menentukan cacah *neighbor* yang paling optimal agar tidak perlu menguji tiap cacah *neighbor*.

Pustaka

[1] E. E. Hussein, A. Derdour, B. Zerouali, A. Almaliki, Y. J. Wong, M. Ballesta-de los Santos, P. Minh Ngoc, M. A. Hashim, and A. Elbeltagi, "Groundwater quality assessment and irrigation water quality index prediction using machine learning algorithms," *Water*, vol. 16, no. 2, 2024. [Online]. Available: https://www.mdpi.com/2073-4441/16/2/264

[2] M. A. Harriz and H. Setiyowati, " Komparasi Algoritma Decission Tree dan KNN dalam Mengklasifikasi Daerah Berdasarkan Produksi Listrik," *JIKO(JURNAL INFORMATIKA DAN KOMPUTER*, vol. 7 No 2, p. 168, 2023.

[3] M. C. Jia Wei Han, *Data Mining : Comcept and Tecniques (Requirements for Cluster Analysis).* USA: Morgan Kaufmann Publisher, 2012.

[4] J. Brownly. Machine learning mastery. [Online]. Available: https://machinelearningmastery.com/one-hot-encoding-for-categorical-data/

[5] N. Y. M. Sofyan Irwanto, Fitra A. Bachtiar, "Klasifikasi aktivitas manusia menggunakan algoritme computed input weight extreme learning machine dengan reduksi dimensi principal component analysis," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 9, no. 6, pp. 1195–1202, 2022.

[6] Rektor, *Peraturan Rektor No :L.05.1/001/UTDI/PR/IX/2022 tentang Pendidikan dan Pembelajaran UTDI.* UTDI, 2022.

[7] Kemdikbud, *Permendikbud no 3 tahun 2020 tentang Standar Nasional Pendidikan Tinggi.* Kemdikbud, 2020.

[8] J. A. Samudraa, S. Anraeni, and Hermana, "Penerapan metode k-nearest neighbor untuk memprediksi tingkat kelulusan mahasiswa berbasis web pada Fakultas Ilmu Komputer UMI," *Buletin Sistem Informasi dan Teknologi Islam*, vol. 1 No 4, pp. 230–237, 2020.

[9] A. J. T, D. Yanosma, and K. A. , "Implementasi metode k-nearest neighbor (KNN) dan simple additive weighting (saw) dalam pengambilan keputusan seleksi penerimaan anggota paskibraka," *Pseudocode*, vol. 2 no.3, pp. 98–112, 2016.

[10] Silviana, R. Kurniawan, A. Nazir, E. Budianita, F. Syafria, and S. K. Gusti, "Pengklasteran risiko covid-19 di riau menggunakan teknik one hot encoding dan algoritma k-means clustering," *JURNAL INFORMASI DAN KOMPUTER*, vol. 10 no.2, pp. 154–162, 2022.

[11] T. Al-Shehari and R. A. Alsowail, "An insider data leakage detection using one-hot encoding, synthetic minority oversampling and machine learning techniques," *Entropy*, vol. 23, no. 10, pp. 1257–1258, 2021.

[12] S. Bagui, D. Nandi, S. Bagui, and R. J. White), "Machine learning and deep learning for phishing email classification using one-hot encodin," *Journal of Computer Science-Univ. of Wes Florida*, vol. 17, pp. 610–623, 2021.

[13] S. Mulyati, S. M. Husein, and Ramdhan, "Rancang bangun aplikasi data mining prediksi kelulusan ujian nasional menggunakan algoritma (knn) k-nearest neighbor dengan metode euclidean distance pada smpn 2 pagedangan," *Jurnal Teknik Informatika (JIKA) Universitas Muhammadiyah Tangerang*, pp. . 65–73, Januari 2020.

[14] S. R. Cholil, T. Handayani, R. Prathivi, and R. Ardianita, "Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa," *IJCIT*, vol. 6, no. 2, pp. 118–127, 2021.

[15] Kartarina and N. K. S. N. luh Putu Juniarti, "Analisis metode k-nearest neighbors (k-nn) dan naive bayes dalam memprediksi kelulusan mahasiswa," *JTIM : Jurnal Teknologi Informasi dan Multimedia*, vol. 3, no. 2, pp. 106–112, 2021.