Jurnal Rekayasa Sistem Informasi dan Teknologi
Volume 1, No 2 – November 2023
e-ISSN : 3025-888X

K-NEAREST NEIGHBOR (K-NN) UNTUK PENANGANAN MISSING VALUEPADA DATA UMKM

Wahyu Sudrajat[1], Idahm Cholid[2]
[12]Universitas Multi Data Palembang
wahyu.sudrajat@mdp.ac.id

Received: 28-09- 2023	Revised: 20-10-2023	Approved: 28-10-2023

ABSTRAK

Missing Value dalam data masih menjadi permasalahan dalam analisis data diberbagai penelitian, terutama di bidang bisnis. Permasalahan tersebut tentu akan mempengaruhi keputusan dalam penentuan keputusan bisnis yang salah atau kurang tepat. Tujuan dari penelitian ini adalah melakukan pre processing data UMKM sehingga didapatkan data yang berkualitas. Diaman data UMKM tersebut dapat digunakan untuk proses pengelompokan dalam rangka pengembangan UMKM di Provinsi Sumatera Selatan. Tahapan penelitian yang dilakukan adalah Collecting Data phase, Data Preprocessing phase, Simulation of Missing Value phase, Modelling phase, Performance Evaluation phase. Pada penelitina ini metode K-Nearest Neighbor (K-NN) digunakan untuk menginputasi data yang hilang. Dengan metode yang sama juga dilakukan pengujian terhadap akurasi dari hasil imputasi yang telah dilakukan. Pada pengujian 798 data UMKM, terdapat 23 pada atribut omzet dan 62 pada atribut asset yang memiliki data tidak lengkap atau missing value. Hasil evaluasi menunjukkan bahwa semakin kecil jumlah k maka akurasi akan semakin baik, yaitu jumlah k = 9 nilai akurasi sebesar 98,12%, jumlah k = 19 nilai akurasi sebesar 97,50% dan jumlah k = 39 nilai akurasi sebesar 96,88%. Tidak adanya aturan terhadap jumlah k merupakan kelemahan dari metode K-NN, diharapkan kedepan dapat digunakan metode untuk menemukan jumlah k yang sesuai.

Kata kunci— *K-NN, UMKM, Missing Value, Clustering*

PENDAHULUAN

Missing data atau missing value adalah satu kondisi dimana terdapat nilai yang tidak lengkap atau kosong pada satu atau beberapa kriteria [1]. Missing Value merupakan hal yang tidak diinginkan dalam *machine learning* dan data mining karena missing value akan menimbulkan banyak masalah. Missing value terjadi pada set data nyata [2] oleh karena beberapa sebab, misalnya: kerusakan peralatan, non-respon dalam survey, tidak cukup data, kerusakan gambar, pengukuran yang salah, salah dalam memasukkan data, atau kesalahan eksperimen dalam prosedur laboratorium. Missing value dapat dikategorika menjadi tiga jenis, yaitu: hilang sepenuhnya secara acak, hilang tidak secara acak, dan hilang secara acak. Missing value merupakan sumber masalah kualitas data [3]. Missing Value dapat menurunkan kualitas model dan bahkan menyebabkan wawasan yang salah[4]. Sehingga sangat penting untuk menangani missing value pada pemrosesan data guna mendapatkan informasi yang dapat dijadikan dasar dalam pengambilan keputusan.

Keseriusan masalah ini tergantung pada berapa banyak data yang hilang, pola kehilangan data dan mekanisme yang mendasarinya [5]. Ada tiga cara untuk mengatasi missing value. Cara pertama dan yang paling tidak efektif [5] adalah dengan menghapus baris-baris dengan nilai nol. Yang kedua meliputi berbagai teknik imputasi seperti rata-rata atau median substitusi, yang dianggap tradisional. Lebih lanjut solusi tingkat lanjut dari kategori ini adalah beberapa imputasi,

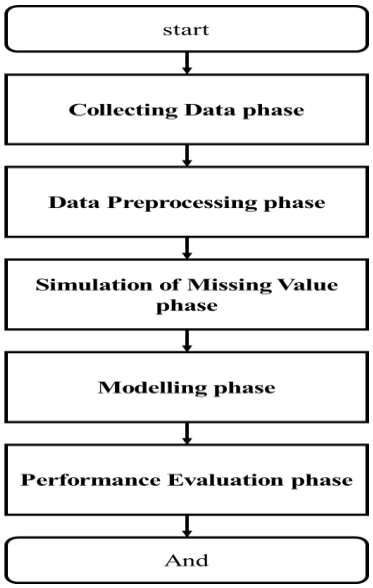
kemungkinan maksimum atau maksimalisasi ekspektasi [6]. Yang ketiga berfokus pada machine learning prediktif yang cenderung memberikan hasil yang baik [7].

Missing value terjadi pada data UMKM yang dikumpulkan pada saat pendataan terhadap penilaian usah yang dilakukan oleh setiap UMKM. Missing value yang sering terjadi pada saat pendataan UMKM adalah hilang nya beberapa nilai pada atribut penilaian tertentu. Pada saat dikonfirmasi mengenai data yang atributnya tidak memiliki nilai yang lengkap, sebagian besar pelaku UMKM menjawab tidak bisa menilai. Contoh kasus pelaku tidak bisa menilai aset atau modal usaha, karena tidak semua pelaku UMKM memiliki administrasi keuangan yang baik. Bentuk dari missing value bisa bermacam-macam, data yang paling sering ditemukan adalah kosong/NaN, 0 dan -. Terdapat banyak cara yang dapat digunakan untuk mengatasi hal missing value. Pada penelitian ini akan digunakan metode mechine learning yaitu K Nearest Neighbor Imputation (KKNI) untuk memprediksi nilai atau data yang hilang dari data UMKM. Tujuan dari penelitian ini adalah sebagai proses pre-procesing, agar nantinya UMKM dapat di kelompokan sehingga dapat dilakukan pendekatan yang tepat dalam pengembangan UMKM yang ada. Pada penelitian ini akan menggunakan metode Root Squared Error (RMSE) untuk menguji keakuratan prediksi missing value.

METODE PENELITIAN

Kerangka Kerja Penelitian

Kerangka kerja penelitian merupakan merupakan gambaran proses yang sistematis, sehingga tujuan dari penelitian dapat dicapai dan berjalan dengan baik. Kerangka kerja dalam penelitian ini ditunjukkan pada gambar 1.



Gambar 1. Tahapan Penelitian

Berikut ini adalah penjelasan dari masing-masing tahapan dari kerangka kerja penelitian, yaitu:

a. Collecting Data phase

Pegumpulan data merupakan tahapan pertama yang di lakukan dalam penelitian ini. Data yang dikumpulkan merupakan data sekunder dari data UMKM yang dikumpulkan pada tahun 2017 oleh salah satu instansi yang berada pada salah satu kabupaten kota di wilayah Pemerintah Daerah Sumatera Selatan. Terdapat 15 atribut yang di gunakan dalam penelitian ini, yaitu: nama UMKM, kategori industry, Alamat, kecamatan, NPWP, bidang usaha, asal bahan baku, hasil produksi, nama PIC, No Telp, jumlah tenaga kerja, modal usaha, omzet dan asest yang dimiliki. Terdapat 798 record data UMKM yang didapatakan untuk penelitian ini.

b. Data Preprocessing phase

Tahapan ini dilakukan preprocessing data yang diperoleh dengan menghilangkan data yang duplikat. Penelitian ini hanya berfokus pada data UMKM. Tidak semua atribut dari data UMKM digunakan seluruhnya, dalam hal ini hanya akan digunakan 4 atribut dari 15 atribut data yang didapatkan. Atribut yang digunakan diantaranya adalah nama UMKM, omzet, asset dan jenis Industri. Data UMKM yang akan digunakan ditunukkan pada tabel 1.

Tabel 1 Data UMKM

NO	NAMA	OMZET	ASSET	JENIS INDUSTRI
1	PABRIK CINCAU	1,080,000,000	60,000,000	Kecil
2	PABRIK TAHU	50,000,000	20,000,000	Mikro
3	PABRIK ROTI	24,000,000		Mikro
4	WARUNG KECIL	144,000,000		Mikro
5	WARUNG TIMAH	12,600,000		Mikro
6	WARUNG WAHYU	288,000,000	16,000,000	Mikro
7	WARUNG YURIA	57,600,000	1,000,000	Mikro
8	TOKO BAJU ALIKHA		10,000,000	Mikro
9	Maya DM	96,000,000		Mikro
10	Abdul Karim	216,000,000	20,000,000	Mikro
11	Bambang Irawan	120,000,000	1,000,000	Mikro
12	Sutikno	108,000,000	10,000,000	Mikro
13	Sri warni	72,000,000	500,000	Mikro
14	Umy Kalsum	19,000,000	32,000,000	Mikro
15	Dewi Mudmainah	13,000,000	30,000,000	Mikro
...	...	...	...	...
789	Husin	79,200,000		Mikro
790	Rita	12,000,000	2,000,000	Mikro
791	Sopian		2,000,000	Mikro
792	Dan	144,000,000	700,000	Mikro
793	Boti	378,000,000	300,000,000	Kecil
794	Diana	2,500,000		Mikro
795	Masa Rimen	18,000,000		Mikro
796	Yus Parida	18,000,000		Mikro
797	Ismono	270,000,000	7,000,000	Mikro
798	Said Hermawan	216,000,000	10,000,000	Mikro

c. Simulation of Missing Value phase

Pada tahapaan ini dilakukan simulasi missing value. Pada simulasi ini dilakukan penerapan metode K-NN dari data UMKM yang ditemukan. Jumlah data yang digunakan dalam simulasi adalah sebanyak 798 record data UMKM. Statistik data UMKM ditunjukan pada tabel 2, dimana dapat dilihat jumlah nilai yang hilang untuk masing-masing atribut dari data UMKM yang akan digunakan. Nilai dengan missing value terbanyaka adalah untuk atribut asset sebanyak 62 record yang hilang atau tidak lengkap. Sedangkan untuk atribut modal sebanyak 32 record yang tidak lengkap atau hilang.

Tabel 2 Statistik Data UMKM

No.	Atribut	Missing	Jumlah Minimum	Jumlah Maksimum
1	Nama	0	Nina aminah	Iskandar
4	Omzet	23	50.000	2.147483647
5	Asset	62	100.000	2.000.000.000

d. Modelling phase

Data UMKM yang telah dikumpulkan dan telah dilakukan pendataan terhadap record yang memiliki nilai missing value selanjutnya diterapkan model pendekatan dalam mengatasi missing value. Pada penelitian ini digunakan metode K-NN untuk menginput data yang hilang atau tidak legkap yang telah diidentifikasi sebelumnya sesuai dengan tahapaan yang telah dijelaskan. Pada tahapan ini akan digunakan tools rapidminer dalam melakukan pemodelan dengan KNN dari data UMKM yang didapatkan.

e. Performance Evaluation phase

Setelah model data didapatkan, langkah selanjutnya adalah evaluasi performance. Hal ini dilakukan untuk menguji keakuratan data dari hasil penerapan metode K-NN yang telah dilakukan. Dalam melakukan evaluasi digunakan metode *Root Mean Squared Error* (RMSE) dimana nilai error terkecil akan menjadi yang terbaik.

K-Nearest Neighbor (K-NN)

Metode K-NN merupakan metode imputasi yang paling populer untuk menyelesaikan masalah Missing Value. Metode ini sederhana dan efektif dalam masalah imputasi missing value, karena paling banyak digunakan untuk menyelesaikan banyak masalah prediksi [8]. Keunggulan metode imputasi dengan KNN dapat digunakan untuk meramalkan 2 jenis data yaitu data diskrit (nilai modus) dan data kontinu (nilai rata-rata), Imputasi dengan KNN tidak membutuhkan pembentukan model peramalan untuk setiap kriteria data yang memiliki data yang hilang (missing value), Kelemahan imputasi dengan KNN adalah pada saat mencari pengamatan yang paling sesuai dengan pengamatan yang memiliki missing value, imputasi dengan KNN akan mencari seluruh data training atau dataset [9]. Urutan langkah dalam proses pencarian missing value dengan KNN adalah sebagai berikut:

a. Langkah 1: Tentukan K

Menentukan jumlah centroids (K) secara acak, dimana k merupakan jumlah observasi terdekat yang akan digunakan. Tidak ada metode khusus dalam menentukan nilai k dalam metode K-NN. Jika nilai k terlalu kecil, maka akan terdapat banyak noise yang mengurangi tingkat akurasi dalam klasifikasi, namun jika terlalu besar juga juga dapat menyebabkan kesalahan dalam membatasi nilai yang diambil dan secara tidak langsung mempengaruhi akurasi [10]

b. Langkah 2: Menghitung jarak Euclidian antara contoh data yang missing value dan data yang lengkap

Menghitung jarak antara euclidean yang memiliki missing value pada observasi ke-j dengan observasi lain yang tidak memiliki missing value pada variabel tersebut sesuai dengan perhitungan jarak Euclidean pada rumus 1.

$$d_{(x,y)} = \sqrt{\sum_{j=1}^{s}(x_j - y_j)^2} \quad \text{.....(1)}$$

Diaman:

$d_{(x,y)}$ = Jarak Euclidian
j = data testing
s = jumlah atribut
x_{aj} = nilai dari atribut ke-j yang mengandung data yang hilang
y_{bj} = nilai dari atribut ke-j lainnya yang berisi
data yang lengkap

c. Langkah 3: mencari k berdasarkan jarak Euclidian minimum

Nilai j pada k observasi terpendek akan digunakan pada proses imputasi untuk observasi yang memiliki missing value.

d. Langkah 4: Hitung nilai k

Observasi yang paling dekat akan mendapatkan nilai tertinggi.

e. Langkah 5: Menghitung nilai rata-rata pada k

menghitung nilai rata-rata pada k observasi terpendek yang tidak memiliki missing value dengan menggunakan rumus 2.

$$x_j = \frac{\sum_{k=1}^{k} w_k v_k}{\sum_{k=1}^{k} w_k} \quad \text{.....(2)}$$

Dimana:
x_j = Weight Mean Estimation
K = jumlah parameter k yang digunakan
w_k = nilai observasi tetangga terdekat K
v_k = nilai dalam data lengkap pada atribut yang mengandung data hilang berdasarkan parameter k

f. Langkah 6: melakukan proses imputasi nilai missing value

Nilai missing value dengan menggunakan nilai rata-rata yang diperoleh pada langkah 5

Root Mean Squared Error (RMSE)

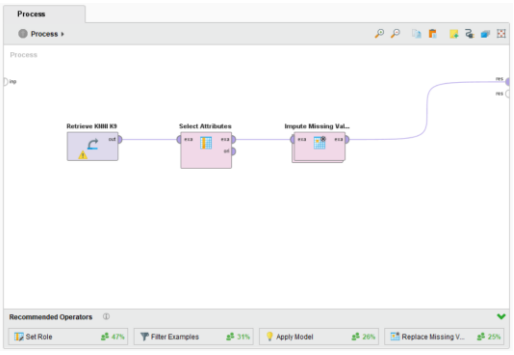
RMSE adalah salah satu dari beberapa metode untuk mengukur keakuratan hasil prediksi atau untuk mengevaluasi teknik prediksi. RMSE menyatakan nilai rata-rata dari jumlah kuadrat hasil prediksi [11]. Nilai terkecil dari perhitungan RMSE menunjukkan bahwa distribusi nilai yang diperoleh dari hasil prediksi mendekati distribusi nilai observasi. Perhitungan RMSE dapat dilihat pada rumus 3.

$$RMSE = \sqrt{\frac{\sum(y_t - \hat{y}_t)^2}{n}} \quad \text{.....(3)}$$

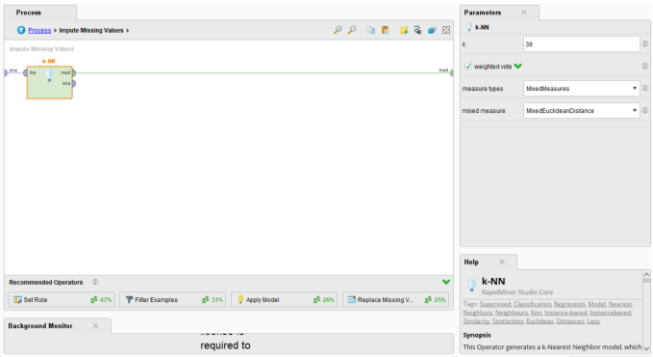
HASIL DAN PEMBAHASAN

Penerapan K-NN terhadap Data UMKM dengan RapidMiner

Data UMKM selanjutnya dilakukan proses imputation dengan menggunakan software Rapidminer 9.10. Data UMKM yang telah dikumpulkan dilakukan pengelolaan sederhana pada Microsoft excel sehingga diperoleh file dalam bentuk excel. Gambar 2 merupakan model imputation missing value pada rapidminer. Dari data UMKM, atribut yang tidak lengkap atau missing value adalah omzet dan azet. Atribut tersebut yang selanjutnya akan dikenakan metode KNN imputation.



Gambar 2 Model Imputation Missing Value



Gambar 3 KNN pada Rapidminer

Gambar 3 menunjukkan penerapan KNN pada model imputation missing value. Pada penelitian ini dilakukan dua kali percobaan, yaitu percobaan 1 dengan jumlah k = 19 dan percobaan ke 2 dengan jumlah k = 29. Gambar 4 menunjukkan data dari masing-masing atribut yang tidak lengkap (missing value) telah terisi (missing = 0) hal ini berbeda dengan hasil sebelumnya yang ditunjukkan pada tabel 2 statistik data UMKM, dimana data UMKM masih terdapat data yang tidak lengkap.

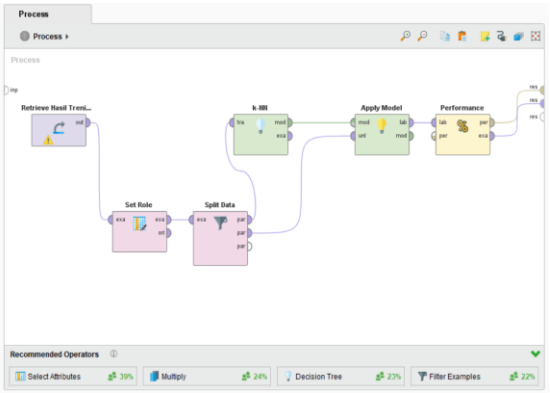
Jurnal Rekayasa Sistem Informasi dan Teknologi
Volume 1, No 2 – November 2023
e-ISSN : 3025-888X

Name	Type	Missing	Statistics	Filter (4 / 4 attributes): Search for Attributes	
Label JENIS INDUSTRI	Polynominal	0	Least Menengah (5)	Most Mikro (758)	Values Mikro (758), Kecil (35), ...
NAMA	Polynominal	0	Least nina aminah (1)	Most Iskandar (4)	Values Iskandar (4), Johan (4), ...
OMZET	Real	0	Min 50000	Max 24200000000	Average 136523668.374
ASSET	Real	0	Min 100000	Max 50000000000	Average 111751070.884

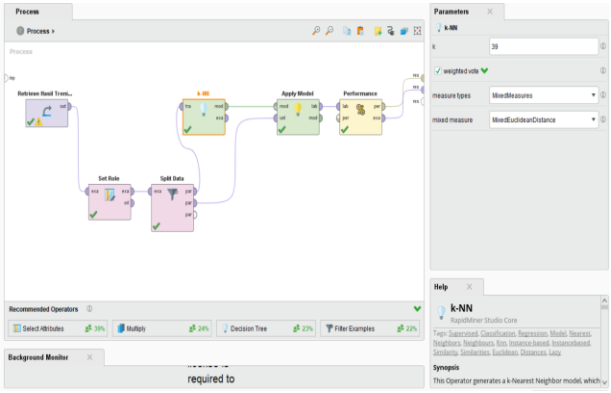
Gambar 4 Hasil Imputation Missing Value dengan Metode KNN

Pengujian Hasil Imputasi KNN dengan RapidMiner

Data UMKM yang telah dilakan perbaik, dengan menerapkan metode KNN untuk mengatasi missing value, maka langkah selanjutnya yang tidak kalah penting adalah melakukan pengujian atau evaluasi terhadap hasil yang didapatkaan. Dalam pengujian akan dilihat tingkat keakurasian dari proses imputasi yang dilakukan terhadap label yang sebelumnya telah ditentukan. Gambar 5 menunjukkan model pengujian data UMKM menggunakan rapidminer. Dalam pengujian ini digunakan splid data dimana dari jumlah data yang ada 80% sebagai data set dan 20% adalah data testing.



Gambar 5 Model Evaluasi Data UMKM pada Rapidminer



Gambar 6 Penentuan Jumlah k pada Model KNN

Gambar 6 merupakan penentuan jumlah k pada model KNN dalam evaluasi hasil imputation data UMKM. Dalam hal ini kita akaan melihat tingkat keakurasian data dari proses imputation yang telah dilakukan sebelumnya. Pada penelitian ini dilakukan 3 kali percobaan, percobaan pertama jumlah k = 9, percobaan kedua jumlah k = 19 dan percobaan ketiga dengan jumlah k = 39. Nilai k ditentukan secara acak. Gambar 7 merupakan hasil pengujian yang dilakukan terhadap data UMKM dengan menggunakan metode KNN. Data awal jenis industri pada record 18 adalah menengah, namun setelah dilakukan pengujian KNN prediksi menjadi kecil, hal ini didukung karena nilai confidence kecil adalah sebesar 0,717 lebih besar dari nilai confidence mikro dan menengah. artinya ada perbedaan dari nilai awal terhadap nilai setelah dilakukan proses imputation dengan metode KNN.

Row No.	Jenis Industri	prediction(Jenis Industri)	confidence(Kecil)	confidence(Mikro)	confidence(Menengah)
17	Mikro	Mikro	0	1	0
18	Menengah	Kecil	0.717	0.205	0.078
19	Mikro	Mikro	0	1	0

Gambar 7 Hasil Pengujian dengan KNN pada Rapidminer

Gambar 8 menunjukkan tingkat akurasi data UMKM setaleh dilakukan Metode Imputation KNN, diperoleh tingkaat akurasi sebesar 96,88% untuk nilai k sebesar 39.

accuracy: 96.88%

	true Kecil	true Mikro	true Menengah	class precision
pred. Kecil	3	0	1	75.00%
pred. Mikro	4	152	0	97.44%
pred. Menengah	0	0	0	0.00%
class recall	42.86%	100.00%	0.00%	

Gambar 8 Akurasi hasil penerapan Metode KNN pada Rapidminer

Analisis Hasil

Penerapan metode imputation KNN pada data UMKM yang mengalami data tidak lengkap atau missing value berhasil dilakukan. Percobaan terhadap missing value dengan imputation KNN melalui dua percobaan pada nilai k yang berbeda yaitu percobaan pertama dengan nilai k = 19 dan percobaan kedua dengan nilai k = 29. Nilai tidak lengkap telah berhasil diselesaikan. Hasil imputation dari dua percobaan tersebut tidak terlalu banyak berbeda nilainya.

Hasil dari *pre processing* yang dilakukan selanjutnya diukur tingkat akurasi dengan menggunakan metode yang sama. Hasil evaluasi akurasi pada percobaan pertama dengan jumlah k = 9 menunjukkan jumlah k = 9 nilai akurasi sebesar 98,12%, jumlah k = 19 nilai akurasi sebesar 97,50% dan jumlah k = 39 nilai akurasi sebesar 96,88%. Tidak ada perbedaan ketika dilakukan pada data UMKM untuk jumlah k = 29.

Tabel 3 Hasil Evaluasi Data UMKM dengan menggunakan Metode KNN

No.	Metode K-NN Imputation	Pengujian dengan K-NN		
		K = 9	K = 19	K = 39
1	Jumlah K = 19	99,88%	97,50 %	96,88%
2	Jumlah K = 29	99,88%	97,50 %	96,88%

Tabel 3 menunjukkan hasil evaluasi data UMKM dengan menggunakan Metode KNN, dimana jika dilihat secara nilai terhadap pengujian dengan nilai k = 9, nilai k = 19 dan nilai k = 39 adalah semakin kecil nilai k maka semakin baik hasil akurasi nya.

KESIMPULAN

Missing Value adalah masalah yang sangat penting sebelum data digunakan dalam proses pengolahan untuk pengambilan keputusan. Jika masalah ini tidak diselesaikan dengan baik, hasil akhir atau hasil analisis dapat terpengaruh. Pada pengujian 798 data UMKM, terdapat 23 pada atribut omzet dan 62 pada atribut asset yang memiliki data tidak lengkap atau missing value. Setelah dilakukan simulasi dengan metode KNN dan melakukan evaluasi didapatkan kesimpulan sebagai berikut:

1. Data UMKM dengan data tidak lengkap atau missing value pada atribut asset dan omzet dilakukan pre processing dengan metode imputation KNN sehingga menjadi lengkap dan siap digunakan dalam analisis.
2. Hasil evaluasi menunjukkan bahwa semakin kecil jumlah k maka akurasi akan semakin baik, yaitu jumlah k = 9 nilai akurasi sebesar 98,12%, jumlah k = 19 nilai akurasi sebesar 97,50% dan jumlah k = 39 nilai akurasi sebesar 96,88%.
3. KNN sangat tergantung dalam penentuan jumlah k, percobaan menunjukkan semakin kecil jumlah k tingkat akurasi semakin baik.

SARAN

Penerapan metode KNN masih memiliki kelemahan dalam penentuan nilai K, untuk itu pada penelitian selanjutnya perlu digunakan metode penentuan k agar hasil yang diperoleh lebih akurat. Pada penelitian selanjutnya perlu dilakukan perbandingan metode yang telah dilakukan dengan menggunakan metode statistic sehingga diperoleh pandangan metode yang lebih baik.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Kepala LPPM Universitas Multi Data Palembang yang telah memberi dukungan sehingga penelitian ini dapat terlaksana.

DAFTAR PUSTAKA

[1] A. Fadlil, Herman, and D. Praseptian M, "K Nearest Neighbor Imputation Performance on Missing Value Data Graduate User Satisfaction," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 6, no. 4, pp. 570–576, 2022, doi: 10.29207/resti.v6i4.4173.

[2] Y. Zhang, "Fairness-aware Missing Data Imputation," no. Tsrml, 2022.

[3] R. Atiq, F. Fariha, M. Mahmud, S. S. Yeamin, K. I. Rushee, and S. Rahim, "A Comparison of Missing Value Imputation Techniques on Coupon Acceptance Prediction," *Int. J. Inf. Technol. Comput. Sci.*, vol. 14, no. 5, pp. 15–25, 2022, doi: 10.5815/ijitcs.2022.05.02.

[4] R. D. Camino, C. A. Hammerschmidt, and R. State, "Improving Missing Data Imputation with Deep Generative Models," 2019.

[5] E. Perkowski, "Impact of Ensemble Machine Learning Methods on Handling Missing Data," pp. 1–8, 1996.

[6] F. Sciences, "Working With Missing Values," *J. Marriage Fam.*, vol. 67, no. November, pp. 1012–1028, 2005.

[7] D. Bertsimas, C. Pawlowski, and Y. D. Zhuo, "From predictive methods to missing data imputation: An optimization approach," *J. Mach. Learn. Res.*, vol. 18, pp. 1–39, 2018.

[8] Z. A. Nadzurah, I. Amelia Ritahani, and A. Nurul, "Performance Analysis of Machine Learning Algorithms for Missing Value Imputation," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 6, 2018.

[9] G. E. A. P. A. Batista and M. C. Monard, "A study of k-nearest neighbour as an imputation method," *Front. Artif. Intell. Appl.*, vol. 87, no. May 2014, pp. 251–260, 2002.

[10] N. D. Irawan, W. Wijono, and O. Setyawati, "Perbaikan Missing value Menggunakan Pendekatan Korelasi Pada Metode K-Nearest Neighbor," *J. Infotel*, vol. 9, no. 3, 2017, doi: 10.20895/infotel.v9i3.286.

[11] I. J. Fadillah and S. Muchlisoh, "PERBANDINGAN METODE HOT-DECK IMPUTATION DAN METODE KNNI DALAM MENGATASI MISSING VALUES Penerapan Pada Data Susenas Maret Tahun 2017," vol. 2017, no. March, pp. 275–285, 2017.