

Analisis Kinerja Algoritma K-Nearest Neighbor Imputation (KNNI) Untuk Missing Value Pada Klasifikasi Data Mining

Miraati Laia

Fakultas Ilmu Komputer dan Teknologi Informasi, Program Studi Teknik Informatika, Universitas Budi Darma, Medan, Indonesia

Email: Miralaia150@gmail.com

Email Penulis Korespondensi: Miralaia150@gmail.com

Abstrak—Pada saat ini banyak peneliti menggunakan dataset untuk diteliti, didalam dataset ada banyak informasi penting yang disediakan sehingga memudahkan peneliti mengolah data tersebut, akan tetapi ada kendala saat hendak mengolah data yaitu beberapa nilai data hilang, atau bahkan rusak (tidak sesuai) dengan nilai lainnya. Jumlah atribut serta sampel data yang ada di dataset tidak terhingga yang membuat peneliti kesulitan dalam menggali informasi yang penting berdasarkan tujuan peneliti masing-masing. Nilai yang hilang atau rusak tersebut dikenal sebagai missing value dan hal tersebut sering terjadi saat peneliti mengolah data yang diambil dari dataset. Pada saat peneliti melakukan klasifikasi terhadap sebuah dataset, tentunya sulit untuk di klasifikasi karena jika terdapat data yang hilang maka proses klasifikasi tidak dapat dilakukan dengan sepenuhnya. Tentu saja hal tersebut akan sangat berpengaruh pada hasil yang didapatkan dan juga hasil akurasi dari proses. Permasalahan hilangnya data (missing value) tersebut harus diselesaikan, salah satu cara untuk menyelesaikan permasalahan tersebut dengan menggunakan algoritma K-Nearest Neighbor Imputation (K-NNI). Algoritma KNNI adalah salah satu algoritma data mining yang dapat mengembalikan permasalahan missing value dengan melakukan pengenalan pola data-data terdekat dari data yang hilang. Adapun hasil akhir setelah diterapkan KNNI dalam memperbaiki missing value yaitu memperoleh nilai bobot (weight) pada setiap K observasi terdekat sebesar 51,6161 pada K5. Dataset yang mengalami missing value telah disesuaikan berdasarkan nilai bobot serta implementasi algoritma KNNI dengan aplikasi rapidminer. Dengan mengembalikan data menggunakan algoritma KNNI ini diharapkan dapat membantu menyelesaikan permasalahan tersebut sehingga hasil yang didapatkan dari proses klasifikasi lebih akurat dan hasil akurasi besar.

Kata Kunci: Dataset; Data Mining; Missing Value; K-NNI; Klasifikasi

Abstract—At this time many researchers use datasets for research, in the dataset there is a lot of important information provided to make it easier for researchers to process the data, but there are obstacles when trying to process the data, namely some data values are lost, or even damaged (incompatible) with other values. The number of attributes and data samples in the dataset is unlimited which makes it difficult for researchers to find important information based on the goals of each researcher. The missing or damaged values are known as missing values and this often happens when researchers process data taken from datasets. When researchers classify a dataset, of course it is difficult to classify because if there is missing data, the classification process cannot be carried out completely. Of course this will greatly affect the results obtained and also the results of the accuracy of the process. The problem of missing data (missing value) must be resolved, one way to solve this problem is by using the K-Nearest Neighbor Imputation (K-NNI) algorithm. The KNNI algorithm is one of the data mining algorithms that can restore the missing value problem by performing pattern recognition of the nearest data from the missing data. The final result after applying KNNI in fixing the missing value is to obtain a weight value for each K closest observation of 51.6161 on K5. Datasets that experience missing values have been adjusted based on the weight values and the implementation of the KNNI algorithm with the rapidminer application. By returning data using the KNNI algorithm, it is hoped that it can help solve these problems so that the results obtained from the classification process are more accurate and the results of greater accuracy.

Keywords: Datasets; Mining Data; Missing Value; K-NNI; Classification

1. PENDAHULUAN

Salah satu hal yang sangat diharapkan ketika melakukan pengumpulan data adalah menghasilkan data yang lengkap. Namun ketika terdapat data tidak lengkap atau data hilang di dalamnya, maka sebelumnya harus diklarifikasi penyebabnya. Secara umum ketidaklengkapan data terjadi karena adanya factor-faktor tertentu yang menyebabkan data menjadi tidak lengkap.

Pada penelitian, data yang tidak lengkap akan mempengaruhi hasil yang didapatkan. Hal tersebut dikarenakan tidak maksimalnya proses yang dilakukan pada penelitian tersebut. Penelitian bidang komputer juga memerlukan data, pada umumnya penelitian bidang komputer menggunakan data set sebagai data yang digunakan untuk proses pada penelitian tersebut.

Data set merupakan kumpulan dari informasi data-data yang tersimpan dalam waktu lama dan menjadi tumpukan data yang besar. Tidak sering data set yang digunakan pada penelitian data yang tersaji tidak terpenuhi secara keseluruhan. Tentu saja hal tersebut sangat berpengaruh terhadap proses yang akan dilakukan nantinya. Data set pada umumnya memiliki data yang lengkap, namun sering juga ditemukan terdapat data yang hilang (*missing value*) pada data set tersebut. *Missing value* pada data set akan menjadi permasalahan yang penting dan harus dilakukan penanganan pada penelitian.

Permasalahan *missing value* tersebut dapat diselesaikan dengan menggunakan teknik data mining. Data mining merupakan proses penggalian informasi dari kumpulan data yang sudah tersimpan di data *warehouse*. Data mining adalah serangkaian proses mendapatkan pengetahuan atau pola dari kumpulan data. Data mining akan memecahkan masalah dengan menganalisa data yang telah ada dalam basis data. Dan klasifikasi merupakan salah satu cara yang dapat dilakukan untuk menyelesaikan permasalahan pada data mining.

Klasifikasi adalah proses mencari identitas umum di antara entitas-entitas yang berbeda dan mengklasifikasikannya ke dalam class yang sesuai. Klasifikasi pada data yang berjumlah besar dan kompleks jika dilakukan secara manual akan sulit dan membutuhkan waktu yang lama. Hal ini dapat diatasi dengan pembangunan suatu *tools* klasifikasi karena dengan adanya *tools* klasifikasi data yang berjumlah besar dan kompleks tersebut akan lebih mudah dan cepat dikelompokkan sesuai dengan kelasnya masing-masing. Salah satu metode pada klasifikasi yang dapat digunakan untuk menyelesaikan permasalahan *missing value* adalah algoritma K-Nearest Neighbor Imputation (K-NNI).

Algoritma K-Nearest Neighbor Imputation adalah sebuah sistem yang menggunakan algoritma pembelajaran terawasi dan bertujuan menemukan pola data baru dengan menghubungkan pola data yang ada dengan data yang baru. Ada dua jenis algoritma NN yaitu INN dan KNN. INN atau Nearest Neighbor adalah pendekatan yang melakukan klasifikasi pada 1 data terdekat, sedangkan KNN adalah pendekatan yang melakukan klasifikasi pada K data terdekat, dengan $K > 1$.

KNNI merupakan pendekatan yang digunakan untuk mengidentifikasi objek berdasarkan informasi tertentu yang merupakan jarak terdekat ke objek. Pada klasifikasi, KNNI bekerja dengan menghitung jarak antara data baru (data testing) dengan data yang sudah diketahui kelasnya (data training) menggunakan jarak euclidian.

Beberapa penelitian yang telah dilakukan berkaitan dengan Algoritma KNNI salah satunya dilakukan oleh Euis Sartika pada tahun 2018 dengan judul Analisis Metode K Nearest Neighbor Imputation (KNNI) Untuk Mengatasi Data Hilang Pada Estimasi Data Survey mendapatkan hasil bahwasannya algoritma KNNI tersebut dapat digunakan untuk mengembalikan data yang hilang[1].

Pada tahun 2019, Dese Narfa Firmansyah, dkk melakukan sebuah penelitian dengan hasil penelitian menunjukkan bahwa algoritme Self-Organizing Maps (SOM) dengan konfigurasi optimal, yaitu $\alpha=0,1$; $\eta=0,2$; epoch=160; dan jumlah neuron=2x2, berhasil membentuk kelompok-kelompok (cluster) dengan Silhouette Coefficient tertinggi sebesar 0,351. Nilai tersebut mengindikasikan adanya struktur yang lemah dalam kelompok-kelompok tersebut. Proporsi area yang dikelompokkan menjadi 1:37. Selain itu, penggunaan metode K-Nearest Neighbors (KNN) untuk mengisi data yang hilang juga terbukti berhasil. Pengisian data kosong dengan KNN menggunakan parameter $K=2$ telah meningkatkan nilai Silhouette Coefficient sebesar 3,4% jika dibandingkan dengan hasil clustering tanpa pengisian data kosong menggunakan KNN. Ini menunjukkan bahwa penerapan KNN secara efektif membantu mengatasi permasalahan data yang tidak lengkap pada dataset PMKS[2].

Penelitian lainnya yang dilakukan oleh Iman Jihad Fadillah dan Siti Muchlisoh pada tahun 2019 yang berjudul Perbandingan Metode Hot-Deck Imputation Dan Metode KNNI Dalam Mengatasi *Missing Values* didapatkan hasil bahwasannya Algoritma KNNI dapat melakukan simulasi terhadap data yang hilang dan dari sisi akurasi algoritma KNNI menghasilkan ketepatan estimator yang lebih akurat serta konsisten[3].

Di tahun 2019 terdapat penelitian yang dilakukan oleh Moch. Lutfi dan Mochamad Hasyim dengan judul Penanganan Data *Missing Value* Pada Kualitas Produksi Jagung Dengan Menggunakan Metode K-NN Imputation Pada Algoritma C4.5 didapatkan hasil K-NN Imputation dapat dipergunakan untuk mengembalikan data yang hilang dari data set dan juga meningkatkan akurasi dari hasil klasifikasi[4].

2. METODOLOGI PENELITIAN

Pada metodologi penelitian dijabarkan tahapan-tahapan yang dilakukan dalam penelitian. Metodologi penelitian terdiri dari beberapa tahapan yang terkait secara sistematis. Tahapan ini diperlukan untuk mempermudah dalam melakukan penelitian. Sebelum membuat kerangka penelitian, penulis terlebih dahulu menganalisa topik yang akan diteliti. Dibawah ini merupakan alur dari tahapan penelitian serta proses pengumpulan data yang dilakukan:

1. Tahap Identifikasi masalah

Pada tahap ini merupakan cara dari penulis untuk dapat menduga, memperkirakan dan menguraikan apa yang sedang menjadi masalah, ditentukan terlebih dahulu sebelum sampai pada tahap pembahasan selanjutnya agar pembahasan suatu masalah dapat terarah atau fokus pada suatu tujuan.

2. Tahap Pengumpulan Data

Pada tahap ini dilakukan pengumpulan data yang berhubungan dengan aktivitas mencari data untuk memperoleh informasi yang dibutuhkan oleh penulis.

3. Tahap Studi Literatur

Pada tahapan ini, dilakukan pemahaman terhadap objek yang akan diteliti, dengan membaca berbagai sumber referensi seperti, buku-buku, jurnal, maupun sumber bacaan lainnya.

4. Tahap Analisis Dan Metode K-Nearest Neighbor Imputation KNNI

Pada tahapan analisis ini, dilakukan pemahaman terhadap proses untuk memecahkan masalah. Pada tahap metode K-Nearest Neighbor Imputation KNNI, proses penelitian dimana data diproses untuk menanggapi perumusan masalah.

5. Tahap Implementasi dan Pengujian Sistem

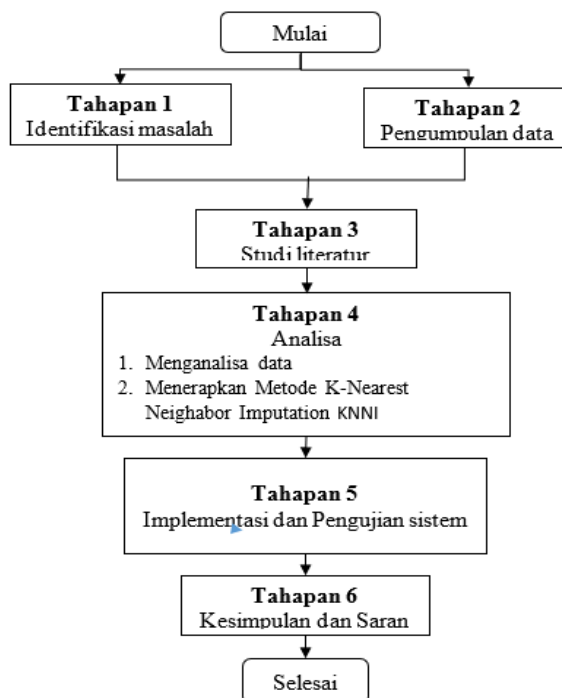
Merupakan tahap penerapan dari proses analisa dan perancangan sistem, dimana data akan diproses kedalam perangkat lunak sistem (*source code*), tujuannya untuk mengetahui apakah sistem dapat berjalan sesuai kebutuhan. Setelah rancangan sistem selesai dibuat, maka tahap selanjutnya adalah tahap pengujian. Tahap pengujian

diperlukan untuk mengetahui apakah sistem yang dibuat dapat berjalan dengan baik atau diperlukannya perbaikan pada sistem tersebut.

6. Kesimpulan dan Saran

Tahapan akhir dari penelitian ini adalah penarikan kesimpulan berdasarkan hasil yang diperoleh dari tahapan-tahapan sebelumnya, mengumpulkan semua data, bahan, serta hasil pengujian yang dituangkan kedalam bentuk laporan penelitian.

Dari tahapan diatas dapat dilihat pada gambar 1 berikut ini.



Gambar 1. Flowchart Tahapan Penelitian

2.1 Data Mining

Data mining merupakan analisis dari peninjauan kumpulan data untuk menemukan hubungan yang tidak diduga dan meringkas data dengan cara berbeda dengan cara yang berbeda dengan sebelumnya, yang dapat dipahami dan bermanfaat bagi pemilik data[5][6][7][8]. Data mining merupakan bidang dari beberapa bidang keilmuan yang menyatukan teknik dari pembelajaran mesin, pengenalan pola, statistic, database, dan visualisasi untuk penanganan permasalahan pengambilan informasi dari database yang besar[9][10][11].

2.2 Teori Missing Value

Missing Value adalah informasi yang tidak tersedia untuk sebuah objek (kasus). Missing value terjadi karena informasi untuk sesuatu tentang objek tidak diberikan, sulit dicari, atau memang informasi tersebut tidak ada. Missing value pada dasarnya tidak bermasalah bagi keseluruhan data, apa lagi jika jumlahnya hanya sedikit, misalnya 1 % dari seluruh data. Menurut defenisinya, missing value adalah hilangnya beberapa data yang telah diperoleh. Dalam dunia data science, missing value erat kaitannya dalam proses perselisihan data (data wrangling) sebelum nantinya akan dilakukan analisis dan prediksi data[12][13][14][15].

2.3 Metode K-Nearest Neighbor Imputation (K-NNI)

Metode imputasi KNN merupakan salah satu metode untuk mengatasi data hilang pada data multivariat yang paling mudah dan populer. Algoritma K-Nearest Neighbor Imputation telah banyak digunakan untuk diimplementasikan dalam imputasi *missing* data terutama dala dataset dengan *missing* data lebih dari satu variabel. Algoritma KNN-I menggunakan observasi yang mirip/serupa dengan observasi yang memiliki nilai missing[16][17]. Langkah-langkah pengerjaan imputasi missing data dengan metode KNNI adalah sebagai berikut[18][19][20]:

1. Menentukan nilai k, yaitu banyaknya observasi terdekat yang akan digunakan ;
2. Menghitung jarak antara observasi yang mengandung missing data pada variabel ke-j dengan observasi lainnya yang tidak mengandung missing data pada variabel yang bersesuaian dengan menggunakan rumus:

$$d(x_a, x_b) = \sqrt{\sum_{j=1}^m (x_{aj} - x_{bj})^2} \quad (1)$$

Keterangan :

$d(x_a, x_b)$: jarak antar observasi target x_a , dan observasi x_b

x_{aj} , :nilai pengamatanke – j pada observasi target $x_a, j = 1,2,\dots,m$

x_{bj} , :nilai pengamatanke – j pada observasi lainnya $x_b, j = 1,2,\dots,m$

3. Menentukan nilai K observasi terdekat berdasarkan nilai jarak terkecil. Nilai variabel pada K observasi terdekat ini yang akan digunakan untuk proses imputasi pada observasi yang mendukung nilai missing.
4. Menghitung bobot (weight) pada setiap K observasi terdekat. Observasi yang paling dekat akan mendapatkan bobot yang paling besar.
5. Menghitung nilai rata-rata pada K observasi terdekat yang tidak mengandung nilai missing data dengan prosedur weighted mean estimation yaitu dengan rumus:

$$x_j = \frac{1}{\sum_{k=1}^K W_k} \sum_{k=1}^K W_k v_{kj} \tag{2}$$

Keterangan :

$$W = \sum_{k=1}^K W_k = 1$$

Dengan W_k : bobot observasi tetangga terdekat ke-k yang dirumuskan :

$$W_k = \frac{1}{d(x, v_k)^2} \tag{3}$$

6. Melakukan proses imputasi missing data pada observasi yang mengandung missing data dengan menggunakan nilai rata-rata yang diperoleh pada tahap 5.

3. HASIL DAN PEMBAHASAN

Pada tahap ini salah satu hal yang sangat diharapkan ketika melakukan pengumpulan data adalah menghasilkan data yang lengkap. Namun ketika terdapat data tidak lengkap atau data hilang didalamnya, maka sebelumnya harus diklasifikasikan penyebabnya. Data set merupakan kumpulan dari informasi data-data yang tersimpan dalam waktu lama dan menjadi tumpukan data yang besar. Data set pada umumnya memiliki data yang lengkap, namun sering juga ditemukan terdapat data yang hilang (missing value) pada data set tersebut missing value akan menjadi permasalahan yang penting dan harus dilakukan penanganan pada penelitian. Algoritma K-nearest Neighbor Imputation adalah sebuah sistem yang menggunakan algoritma pembelajaran terawasi dan bertujuan menemukan pola data baru dengan menghubungkan algoritma pembelajaran terawasi dan bertujuan menemukan pola data baru dengan menghubungkan pola data yang ada dengan data yang baru. Ada dua jenis algoritma NN yaitu INN dan KNN. INN atau Nearest Neighbor adalah pendekatan yang melakukan klasifikasi pada 1 data terdekat, sedangkan KNN adalah pendekatan yang melakukan klasifikasi pada K data terdekat, dengan $K > 1$.

Maka permasalahan missig value tersebut dapat diselesaikan dengan menggunakan teknik data mining. Data mining adalah proses penggalian informasi dari kumpulan data yang sudah tersimpan di data warehouse. Dan klasifikasi merupakan salah satu cara untuk menyelesaikan permasalahan pada data mining. Klasifikasi pada data yang berjumlah besar dan kompleks jika dilakukan secara manual akan sulit dan membutuhkan waktu yang lama. Hal ini dapat diatasi dengan pembangunan suatu tools klasifikasi karena adanya tools klasifikasi data yang berjumlah besar dan kompleks tersebut akan lebih mudah dan cepat dikelompokkan sesuai dengan kelasnya masing-masing.

Berdasarkan uraian diatas bahwa algoritma K-Nearest Neighbor KNNI dapat digunakan untuk menyelesaikan permasalahan pada missing value.

3.1 Penerapan Algoritma K-Nearest Neighbor Imputation

Sebelum dilakukan penelitian, perlu kiranya mengetahui data set yang akan digunakan pada penelitian. Dimana pada dataset tersebut memiliki nilai yang tidak lengkap (*missing value*). Adapun data tersebut dapat dilihat tabel 2 berikut ini:

Tabel 2. Data set

No	Age	Cp	Trtbps	chol	Fbs	restecg	thalachh	exng	oldpeak	slp	caa	Thall	output
1	63	3	145	233	1	0	150		2,3	0	1	1	1
2		2	130	250	0	1	187	1	3,5	0		2	1
3	41	1		204	0	0	172	0	1,4	2	0	2	1
4	56	1	120	236	0	1	178		0,8	2	1	2	1
5	57		120	354	1	1	163	1		2		2	1
6		0		192	0	1	148	0	0,4	1	1	1	1
7	56	1	140	294	0	0	153	0	1,3	1	0	2	1
8	44	1	120	263	0	1	173			2	0	3	0
9	52		172	199	1	1	162	1	0,5	2	1	3	0
10		2		168	0	1	174	0	1,6	2		2	0
11	54	0	140	239	1		160		1,2	2	1	2	0
12	48	2	130	275	0	1	139	0		2	1	2	0

Penanganan *missing data* dengan *KNN* diawali dengan menentukan sejumlah tetangga terdekat atau observasi terdekat yang disimbolkan dengan *K*, kemudian menghitung jarak terkecil dari setiap observasi yang tidak mengandung *missing data*. Langkah-langkah imputasi *missing data* dengan metode *KNN* adalah sebagai berikut:

1. Tentukan parameter *K*, *K* adalah jumlah observasi terdekat atau tetangga terdekat yang akan digunakan.

Jumlah $K = 5$

2. Menghitung jarak antara observasi yang mengandung *missing data* dengan observasi lengkap pada variabel ke-*j* yang tidak mengandung *missing data* dengan variabel *j* lainnya yang bersesuaian dengan rumus jarak euclidian yaitu[5]:

$$d(x_a, x_b) = \sqrt{\sum_{j=1}^m (x_{aj} - x_{bj})^2}$$

jarak objek observasi

$$\begin{aligned} d(x_2, x_2) &= \sqrt{(0 - 63)^2 + (2 - 3)^2 + (130 - 145)^2 + (250 - 233)^2 +} \\ &= \sqrt{(0 - 1)^2 + (0 - 1)^2 + (187 - 150)^2 + (1 - 0)^2 + (3,5 - 2,5)^2} \\ &= \sqrt{(0 - 0)^2 + (0 - 1)^2 + (2 - 1)^2 + (1 - 1)^2} = 76,5441 \end{aligned}$$

Tabel 3. Perhitungan Jarak

No	Objek	Jarak
1	Objek 1	76,5441
2	Objek 2	14,6872
3	Objek 3	59,3909
4	Objek 4	121,5
5	Objek 5	147,6570
6	Objek 6	79,6042
7	Objek 7	49,2061
8	Objek 8	87,8350
9	Objek 9	154,2809
10	Objek 10	62,3240
11	Objek 11	72,4724

Tabel 4. Data Hasil Perhitungan Jarak

No	Data perhitungan jarak terdekat	Jarak
1	Objek 2	14,6872
2	Objek 7	49,2061
3	Objek 3	59,3909
4	Objek 10	62,3240
5	Objek 11	72,4724

3. Menentukan nilai *K* observasi terdekat berdasarkan nilai jarak terkecil. Nilai variabel pada *K* observasi terdekat ini yang akan digunakan untuk proses imputasi pada observasi yang mendukung nilai *missing*.

$$\begin{aligned} R_1, R_2 &= \sqrt{\frac{12}{11} * (2 - 3)^2 + (145 - 130)^2 + (223 - 250)^2 +} \\ &= \sqrt{(1 - 0)^2 + (0 - 1)^2 + (150 - 187)^2 + (2,3 - 3,5)^2 +} \\ &= \sqrt{(0 - 2,5)^2 + (0 - 0)^2} = 48,2627 \end{aligned}$$

Tabel 5. K Jarak Observasi

No	Objek	Jarak
1	Objek 1	48,2627
2	Objek 2	36,5213
3	Objek 3	37,8318
4	Objek 4	124,7637
5	Objek 5	41,2967
6	Objek 6	6,8118
7	Objek 7	45,5012
8	Objek 8	45,1468
9	Objek 9	69,3764
10	Objek 10	13,3495

No	Objek	Jarak
11	Objek 11	46,0325

Berdasarkan dengan hasil pengamatan nilai K Observasi dengan K=5. Maka didapatkan data hasil observasi.

Tabel 6. Data Hasil K Observasi

No	Data perhitungan jarak terdekat observasi	Jarak
1	Objek 6	6,8118
2	Objek 10	13,3495
3	Objek 2	36,5213
4	Objek 3	37,8318
5	Objek 8	45,1468

4. Menghitung bobot (weight) pada setiap K observasi terdekat. Observasi yang paling dekat akan mendapatkan bobot yang paling besar.

$$\begin{aligned}
 K5 &= \frac{1}{(6,8118)} * 14,6872 + \frac{1}{(13,3495)} * 49,2061 + \frac{1}{(36,5213)} * 59,3909 + \frac{1}{(37,8318)} * 62,3240 + \frac{1}{(45,1468)} * 72,4724 \\
 &= \frac{1}{(6,8118)} + \frac{1}{(13,3495)} + \frac{1}{(36,5213)} + \frac{1}{(37,8318)} + \frac{1}{(45,1468)} \\
 &= \frac{14,6872}{1} + \frac{49,2061}{1} + \frac{59,3909}{1} + \frac{62,3240}{1} + \frac{72,4724}{1} \\
 &= \frac{258,0806}{5} = 51,6161
 \end{aligned}$$

Pada pengujian penerapan metode K-nearest Neighbor Imputation KNNI menggunakan Rapidminer. Adapun hasil pengujian metode K-nearest Neighbor Imputation KNNI dalam pencarian data yang hilang. Dalam tahapan ini penulis menampilkan bentuk hasil tampilan yang disediakan oleh K-nearest Neighbor Imputation KNNI seperti yang terlihat pada gambar 2 berikut.

Gambar 2. Tahapan Hasil

4. KESIMPULAN

Sebagai penutup pembahasan dalam penulisan tugas akhir, penulis mengambil kesimpulan-kesimpulan sekaligus memberikan saran kepada sipembaca maupun yang ingin mengembangkan kembali penerapan data mining untuk pencarian data yang hilang dengan metode K-nearest Neighbor Imputation KNNI. Penerapan data mining untuk pencarian data yang hilang dengan metode K-Nearset Neighbor Imputation KNNI merupakan proses untuk menghasilkan *knowledge* baru berupa perbandingan antara faktor-faktor yang mempengaruhi pencarian data yang hilang tersebut. Hasil dari data mining dengan menggunakan metode K-Nearest Neighbor Imputation KNNI merupakan sebuah susunan urutan kegiatan yang saling mendukung didalam proses. Dengan menggunakan *tools* rapid miner algoritma K-Nearest Neighbor Imputation KNN dapat dilakukan pengujian.

REFERENCES

- [1] E. Sartika, "Analisis metode k nearest neighbor imputation (knni) untuk mengatasi data hilang pada estimasi data survey," *Tedc*, vol. 12, no. 3, pp. 219-227, 2018.
- [2] D. N. Firmansyah, S. Adinugroho, and B. Rahayudi, "Kesejahteraan sosial menggunakan Algoritme Self-Organizing Maps dengan perbaikan Missing Value K-Nearest Neighbors," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 7, pp. 7205-7213, 2019.

- [3] I. J. Fadillah and S. Muchlisoh, "Perbandingan Metode Hot-Deck Imputation Dan Metode Knni Dalam Mengatasi Missing Values," *Semin. Nas. Off. Stat.*, vol. 2019, no. 1, pp. 275–285, 2020, doi: 10.34123/semnasoffstat.v2019i1.101.
- [4] Moch. Lutfi and Mochamad Hasyim, "Penanganan Data Missing Value Pada Kualitas Produksi Jagung Dengan Menggunakan Metode K-Nn Imputation Pada Algoritma C4.5," *J. Resist. (Rekayasa Sist. Komputer)*, vol. 2, no. 2, pp. 89–104, 2019, doi: 10.31598/jurnalresistor.v2i2.427.
- [5] F. O. Lusiana, I. Fatma, and A. P. Windarto, "Estimasi Laju Pertumbuhan Penduduk Menggunakan Metode Regresi Linier Berganda Pada BPS Simalungun," *J. Informatics Manag. Inf. Technol.*, vol. 1, no. 2, pp. 79–84, 2021, [Online]. Available: <https://hostjournals.com/>
- [6] Z. Nabila, A. Rahman Isnain, and Z. Abidin, "Analisis Data Mining Untuk Clustering Kasus Covid-19 Di Provinsi Lampung Dengan Algoritma K-Means," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 2, p. 100, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTISI>
- [7] S. Al Syahdan and A. Sinar, "Data Mining Penjualan Produk Dengan Metode Apriori Pada Indomaret Galang Kota," *J. Nas. Komputasi dan Teknol. Inf.*, vol. 1, no. 2, 2018, doi: 10.32672/jnkti.v1i2.771.
- [8] A. Wanto *et al.*, *Data Mining: Algoritma dan Implementasi*. Yayasan kita menulis, 2020.
- [9] A. Z. Siregar, "Implementasi Metode Regresi Linier Berganda Dalam Estimasi Tingkat Pendaftaran Mahasiswa Baru," *Kesatria J. Penerapan Sist. Inf. (Komputer dan Manajemen)*, vol. 2, no. 3, pp. 133–137, 2021, [Online]. Available: <https://tunasbangsa.ac.id/pkm/index.php/kesatria/article/view/73>
- [10] S. S. S, A. T. Purba, V. Marudut, M. Siregar, T. Komputer, and P. B. Indonesia, "SISTEM PENDUKUNG KEPUTUSAN KELAYAKAN PEMBERIAN PINJAMAN," vol. 3, pp. 25–30, 2020, doi: 10.37600/tekinkom.v3i1.131.
- [11] B. S. Pranata and D. P. Utomo, "Penerapan Data Mining Algoritma FP-Growth Untuk Persediaan Sparepart Pada Bengkel Motor (Study Kasus Bengkel Sinar Service)," *Bull. Inf. Technol.*, vol. 1, no. 2, pp. 83–91, 2020.
- [12] L. Mutawalli and W. Murniati, "PENERAPAN KNNIMPUTER DALAM MENGOLAH DATA MISSING VALUE UNTUK MEMBANTU MENINGKATKAN AKURASI SUPPORT VECTOR MACHINE KLASIFIKASI PENYAKIT TIROID," *J. Inform. Teknol. dan Sains*, vol. 4, no. 4, pp. 386–390, 2022.
- [13] M. Lutfi and M. Hasyim, "Penanganan Data Missing Value Pada Kualitas Produksi Jagung Dengan Menggunakan Metode K-Nn Imputation Pada Algoritma C4. 5," *J. Resist. (Rekayasa Sist. Komputer)*, vol. 2, no. 2, pp. 89–104, 2019.
- [14] B. K. KHOTIMAH, "Optimasi Pemodelan Hybrid Naive Bayes Berbobot Dengan Menggunakan Algoritma Genetika Pada Data Missing Value." UNIVERSITAS AIRLANGGA, 2020.
- [15] R. A. Maula *et al.*, "Handling Missing Value dengan Pendekatan Regresi pada Dataset Akuakultur Berukuran Kecil," *J. Rekayasa Elektr.*, vol. 18, no. 3, 2022.
- [16] I. J. Fadillah and C. D. Puspita, "PEMANFAATAN METODE WEIGHTED K-NEAREST NEIGHBOR IMPUTATION (WEIGHTED KNNI) UNTUK MENGATASI MISSING DATA," in *Seminar Nasional Official Statistics*, 2020, vol. 2020, no. 1, pp. 511–518.
- [17] A. F. Sallaby and A. Azlan, "Analysis of Missing Value Imputation Application with K-Nearest Neighbor (K-NN) Algorithm in Dataset," *IJICS (International J. Informatics Comput. Sci.)*, vol. 5, no. 2, pp. 141–144, 2021.
- [18] D. N. F. M. DESY NUR FITRIANA MURJITO, "PERBANDINGAN METODE IMPUTASI: METODE MEAN DAN METODE K NEAREST NEIGHBOR (KNN) UNTUK MENGATASI DATA HILANG PADA DATA SURVEI," 2021.
- [19] A. Fadlil, "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.
- [20] F. R. A. Widodo, "PENANGANAN DATA MISSING VALUE DENGAN KNN-IMPOTASI PADA ALGORITMA K-NEAREST NEIGHBOUR UNTUK KLASIFIKASI DATA PELANGGAN INDIHOME PT TELKOM." Universitas Yudharta, 2021.