

KOMPARASI METODE MEAN DAN KNN IMPUTATION DALAM MENGATASI MISSING VALUE PADA DATASET KECIL

Fandi Yulian Pamuji¹, Ahmad Rofiqul Muslikh², Rizza Muhammad Arief³, Delviana Muti⁴

^{1,2,3,4}Jurusan Sistem Informasi Fakultas Teknologi Informasi Universitas Merdeka Malang

¹fandi.pamuji@unmer.ac.id, ²rofickachmad@unmer.ac.id, ³rizza@unmer.ac.id,

⁴delviana.muti@student.unmer.ac.id

Abstrak

Missing value pada dataset yang kecil akan mengakibatkan berkurangnya data yang dapat digunakan untuk pembelajaran sehingga prediksi hasil klasifikasi dari data tersebut akan berkurang. Metode Imputasi sebagai solusi metode yang paling umum digunakan untuk menangani masalah dataset yang tidak lengkap. Metode Imputasi proses di mana beberapa teknik statistik digunakan untuk mengganti data yang hilang dengan nilai pengganti. Tujuan penelitian ini dengan kinerja klasifikasi yang dapat dipertahankan dengan metode imputasi missing value, karena metode ini dapat menghindari berkurangnya jumlah dataset yang digunakan dalam proses klasifikasi pada dataset dan meningkatkan kinerja klasifikasi pada dataset yang tidak ideal terutama untuk jumlah dataset yang kecil. Berdasarkan hasil eksperimen yang telah dilakukan dari penelitian ini yaitu bahwa pengujian metode imputasi Mean dan KNN Imputation dengan metode klasifikasi mampu menangani data kosong dengan jumlah missing value sedikit maupun banyak dengan menghasilkan nilai accuracy mencapai kinerja prediksi yang lebih besar dibandingkan dengan menggunakan missing value nilai 0. Kemudian untuk dataset Hepatitis nilai Accuracy tinggi menggunakan metode imputasi KNN Imputasi dengan nilai 0,79 menggunakan metode Logistic Regression dan dataset Ginjal Kronis nilai Accuracy tinggi menggunakan metode imputasi Mean dengan nilai 0,97 dengan menggunakan metode Naïve Bayes. Hal tersebut menunjukkan bahwa proses metode imputasi terhadap nilai kosong di setiap column dataset kecil pada tahap data preprocessing memberikan pengaruh terhadap nilai Accuracy metode Mean dan KNN Imputation pada metode klasifikasi.

Kata kunci : Metode Imputasi, Missing Value, Mean, KNN Imputation

1. Pendahuluan

Data Mining sebagai tugas penting dalam menganalisis data dan menantang untuk banyak masalah dalam kehidupan sehari-hari. Untuk melakukan analisis data mining berisi beberapa proporsi data tidak lengkap yang memiliki satu atau lebih nilai atribut yang hilang ((SENASIF) & 2022, 2022). Ada banyak alasan tidak lengkap dataset, yang timbul dari berbagai sumber, sistem database itu sendiri, memasukkan data yang tidak tepat, salah, dan sebagainya. Masalah klasifikasi menjadi lebih berat ketika mempunyai sebuah dataset dengan jumlah kecil dan memiliki nilai yang kosong pada beberapa atribut dengan sistem (Ramadhan & Pamuji, 2022).

Missing value pada dataset yang kecil akan mengakibatkan berkurangnya data yang dapat digunakan untuk pembelajaran sehingga prediksi hasil klasifikasi dari data tersebut akan berkurang (Pamuji & Ramadhan, 2021). Metode Imputasi sebagai solusi metode yang paling umum digunakan untuk menangani masalah dataset yang tidak lengkap. Metode Imputasi proses di mana beberapa teknik statistik digunakan untuk mengganti data

yang hilang dengan nilai pengganti (Pamuji et al., 2021). Beberapa teknik statistik untuk menangani missing value data seperti Mean dan KNN Imputation yang telah diterapkan untuk metode missing value pada penelitian ini untuk meningkatkan kinerja klasifikasi pada dataset yang tidak ideal terutama untuk jumlah dataset yang kecil (Pamuji & Soeleman, 2020).

Dalam metode imputasi Mean prosesnya mengisi data yang memiliki nilai yang hilang dan mendapatkan data yang lengkap dengan jumlah dari perhitungan rata-rata dari data yang tidak hilang. Metode imputasi KNN Imputation prosesnya menentukan jumlah tetangga terdekat yang disimbolkan dengan k, kemudian menghitung jarak terkecil dari setiap observasi yang tidak mengandung missing data pada dataset kecil (Misdrum et al., 2023).

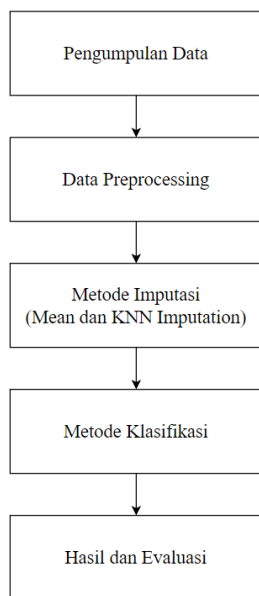
Masalah yang ada pada penelitian ini yaitu dataset yang kecil dan terdapat nilai yang kosong pada umumnya jika dilakukan penghapusan data maka akan terjadi penurunan kinerja akurasi yang disebabkan oleh pengurangan jumlah data maka solusi yang dilakukan proses imputasi missing value

menggunakan Mean dan KNN Imputation (Martinez et al., 2022). Dalam penelitian ini menggunakan metode klasifikasi meliputi metode Random Forest, KNN dan Naive Bayes kemudian untuk dataset yang digunakan menggunakan 2 dataset data public dari repository UCI (Dinh et al., 2021). Random Forest KNN dan Naive Bayes merupakan metode algoritma data mining yang digunakan untuk melakukan klasifikasi sebuah dataset. Tujuan penelitian ini dengan kinerja klasifikasi yang dapat dipertahankan dengan metode imputasi missing value, karena metode ini dapat menghindari berkurangnya jumlah dataset yang digunakan dalam proses klasifikasi pada dataset dan meningkatkan kinerja klasifikasi pada dataset yang tidak ideal terutama untuk jumlah dataset yang kecil (Misdrum et al., 2020).

2. Metodologi Penelitian

Penelitian ini yang digunakan dataset adalah data kecil dari kumpulan repository UCI. Data kecil ini diolah terlebih dahulu menggunakan metode imputasi Mean dan KNN Imputation kemudian diuji menggunakan metode klasifikasi Random Forest, KNN dan Naive Bayes. Penelitian ini akan dilakukan beberapa tahapan seperti pada Gambar 1 dibawah ini.

2.1 Alur Penelitian



Gambar 1. Alur Penelitian

Alur penelitian ini yang pertama melakukan pengumpulan data dari kumpulan repository UCI Machine Learning Repository berjumlah 2 dataset yang terdiri dari 2 Class. Langkah berikutnya Data Preprocessing melakukan perubahan dataset yang sesuai dalam proses data mining. Tahapan normalization yang kedua untuk merubah nilai kosong dengan menggunakan metode missing value Mean dan KNN Imputation agar dataset tersebut

valuennya lengkap kemudian yang ketiga merubah tipe data text pada class menjadi nilai class dalam rentang 1 dan 2 untuk dataset kecil. Langkah berikutnya menggunakan metode imputasi Mean dan KNN Imputation untuk dataset Hepatitis dan Ginjal Kronis. Langkah berikutnya menggunakan metode klasifikasi Random Forest, KNN dan Naive Bayes. Langkah terakhir hasil dan evaluasi melakukan pengujian accuracy terhadap dataset Hepatitis dan Ginjal Kronis berdasarkan proses metode imputasi dan metode klasifikasi dengan mengukur metode mana yang lebih tinggi tingkat any (Muti & Pamuji, 2023).

2.2 Pengumpulan Data

Teknik pengumpulan data public yang dilakukan dengan mempersiapkan data kecil dari kumpulan repository UCI Machine Learning Repository berjumlah 2 dataset yang terdiri dari 2 Class. Data yang telah dikumpulkan pada Tabel 1 dibawah ini sebagai berikut:

Tabel 1. Dataset Kecil

Dataset	Instance	Atribute	Missing Value
Hepatitis	155	20	132
Ginjal Kronis	382	25	825

Pada Tabel 1 merupakan 2 dataset yang memiliki missing value dan memiliki 2 Class di setiap masing-masing dataset dan jumlah data kemudian jumlah fitur nya berbeda-beda tergantung dari datasetnya dari kumpulan repository UCI Machine Learning Repository <https://archive.ics.uci.edu/datasets>. Untuk dataset hepatitis merupakan jumlah instance 155 data, jumlah atribut 20 dan jumlah missing value 132 kemudian untuk dataset Ginjal Kronis merupakan jumlah datanya lebih banyak dari dataset Hepatitis dengan jumlah instance 382 data, jumlah atribut 25 dan jumlah missing value 825.

Tabel 2. Data Awal Hepatitis (Missing Value)

Varices	Bilirubin	Alk Phosphate	Sgot	Albumin	Prottime
2					
2	0,104166667	280	98	0,130555556	40
2	0,041666667		60		
2	0,041666667	67	28	0,168055556	
2	0,173611111		48	0,0875	73
2	0,138888889	127	182	0	
2	0,083333333	167	242	0,127083333	
1	0,0625	135	55		41
1	0,118055556	165	64	0,088888889	
2					

Pada Tabel 2 merupakan data awal dari dataset Hepatitis yang mempunyai missing value berjumlah 132.

Tabel 3. Data Awal Ginjal Kronis (Missing Value)

age	bp	sg	al	su	rbc
48	80	1.020	1	0	
7	50	1.020	4	0	
62	80	1.010	2	3	1
48	70	1.005	4	0	1
51	80	1.010	2	0	1
60	90	1.015	3	0	
68	70	1.010	0	0	
24	75	1.015	2	4	1
52	100	1.015	3	0	1
48	80	1.020	1	0	

Pada Tabel 3 merupakan data awal dari dataset Ginjal Kronis yang mempunyai missing value berjumlah 825.

2.3 Data Preprocessing

Tahapan pada data transformation yang pertama melakukan perubahan dataset yang sesuai dalam proses data mining. Tahapan normalization yang kedua untuk merubah nilai kosong dengan menggunakan metode missing value Mean dan KNN Imputation agar dataset tersebut valuennya lengkap kemudian yang ketiga merubah tipe data text pada class menjadi nilai class dalam rentang 1 dan 2 untuk dataset kecil (Sanlialp & Ozturk, 2020). Terakhir dataset tersebut kemudian split 80% data training dan 20% data testing sebelum dilakukan ke metode klasifikasi Random Forest, KNN dan Naive Bayes (Lin & Tsai, 2020).

2.4 Metode Imputasi (Mean)

Dengan cara ini metode imputasi (Mean) dmenghitung rata-rata nilai yang tidak hilang dalam atribut beberapa data dan kemudian mengganti nilai yang hilang di setiap atribut secara terpisah dan terpisah dari yang lain hanya dapat digunakan dengan data numerik (Amin et al., 2019). Pada Tabel 4 Data Hepatitis (Mean) dan Tabel 5 Data Ginjal Kronis (Mean) merupakan cara menentukan imputasi dengan nilai Mean, dimana nilai atribut yang kosong ditandai dengan simbol NaN, kemudian tanda NaN diganti dengan nilai rata-rata dari banyaknya nilai dalam satu kolom dibagi dengan jumlah nilai dalam kolom tersebut (Nilashi et al., 2019).

Tabel 4. Data Hepatitis (Mean)

Varices	Bilirubin	Alk Phosphate	Sgot	Albumin	Protine
2	0,071338384	104,8692	85	0,146365741	62
2	0,104166667	280	98	0,130555556	40
2	0,041666667	104,8692	60	0,146365741	62
2	0,041666667	67	28	0,168055556	62
2	0,173611111	104,8692	48	0,0875	73
2	0,138888889	127	182	0	62
2	0,083333333	167	242	0,127083333	62
1	0,0625	135	55	0,146365741	41
1	0,118055556	165	64	0,088888889	62
2	0,071338384	104,8692	85	0,146365741	62

Tabel 5. Data Ginjal Kronis (Mean)

age	bp	sg	al	su	rbc
48	80	1.020	1	0	1
7	50	1.020	4	0	1
62	80	1.010	2	3	1
48	70	1.005	4	0	1
51	80	1.010	2	0	1
60	90	1.015	3	0	1
68	70	1.010	0	0	1
24	75	1.015	2	4	1
52	100	1.015	3	0	1
48	80	1.020	1	0	1

2.5 Metode Imputasi (KNN Imputation)

Dengan cara ini metode imputasi (KNN Imputation) menghitung kNN untuk menemukan k sampel terdekat sebagai gantinya dan memasukkan elemen yang hilang ini hanya dapat digunakan dengan data numerik (Li et al., 2019). Pada Tabel 6 Data Hepatitis (KNN Imputation) dan Tabel 7 Data Ginjal Kronis (KNN Imputation) merupakan cara menentukan imputasi dengan nilai KNN Imputation, dimana nilai atribut yang kosong ditandai dengan simbol NaN, kemudian tanda NaN diganti dengan pendekatan k tetangga terdekat untuk mengisi value yang hilang dalam kolom tersebut (Wei et al., 2018).

Tabel 6. Data Hepatitis (KNN Imputation)

Varices	Bilirubin	Alk Phosphate	Sgot	Albumin	Protine
2	0,111111111	161,6666667	117	0,129398148	55
2	0,104166667	280	98	0,130555556	40
2	0,041666667	145,6666667	60	0,10162037	22
2	0,041666667	67	28	0,168055556	82
2	0,173611111	133,3333333	48	0,0875	73
2	0,138888889	127	182	0	23
2	0,083333333	167	242	0,127083333	39
1	0,0625	135	55	0,128009259	41
1	0,118055556	165	64	0,088888889	43
2	0,111111111	161,6666667	117	0,129398148	55

Tabel 7. Data Ginjal Kronis (KNN Imputation)

age	bp	sg	al	su	rbc
48	80	1020	1	0	1
7	50	1020	4	0	0
62	80	1010	2	3	1
48	70	1005	4	0	1
51	80	1010	2	0	1
60	90	1015	3	0	0
68	70	1010	0	0	0
24	75	1015	2	4	1
52	100	1015	3	0	1
48	80	1020	1	0	1

2.6 Metode Klasifikasi

Logistic Regression merupakan metode statistika yang digunakan untuk menganalisis data yang mendeskripsikan antara variabel respon dengan satu atau lebih variabel prediksi. Variabel respon dari Logistic Regression bersifat dikotomi yang hanya bernilai 1 (ya) dan 0 (tidak), sehingga variabel respon yang dihasilkan akan mengikuti distribusi Bernoulli dengan fungsi probabilitas persamaan sebagai berikut (Liu et al., 2016):

$$p_i = \frac{1}{1 + e^{-(w^T x_i + b)}} \tag{1}$$

K-Nearest Neighbor merupakan algoritma yang digunakan untuk klasifikasi, estimasi dan prediksi. K-Nearest Neighbor basis pembelajaran dimana data training disimpan sehingga klasifikasi untuk data yang belum terklasifikasi dapat ditemukan dengan cara membandingkannya pada catatan data training. Ada beberapa perhitungan pada metode KNN, perhitungan yang paling umum digunakan adalah perhitungan berdasarkan k jarak terdekat (Purwar & Singh, 2015). Rumus K-NN menggunakan persamaan sebagai berikut:

$$dis(x_1, x_2) = \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2} \quad (2)$$

Naïve Bayes adalah pengklasifikasian statistika yang digunakan untuk memprediksi probabilitas keanggotaan suatu class. Naïve bayes memiliki akurasi yang tinggi saat diaplikasikan ke dalam database dengan data yang besar. Metode Naïve Bayes pendekatan statistik untuk melakukan inferensi induksi pada persoalan klasifikasi (Xu et al., 2019). Rumus Naïve Bayes menggunakan persamaan sebagai berikut:

$$P(C|X) = \frac{P(X|C)P(c)}{P(x)} \quad (3)$$

2.7 Hasil dan Evaluasi

Pada tahapan ini dilakukan evaluasi terhadap tingkat akurasi dari masing-masing metode untuk melihat kinerja setiap metode yang digunakan. Pada penelitian ini metode klasifikasi Random Forest, KNN dan Naive Bayes akan dievaluasi dan divalidasi menggunakan alat ukur Confusion Matrix dengan persamaannya sebagai berikut:

Tabel 8. Confusion Matrix

Classification	Predicted Class	
	Class = Yes	Class = No
Class = Yes	A (True Positive-YP)	B (True Negative-FN)
Class = No	C (True Positive-FP)	D (True Negative-TN)

Keterangan dari tabel confusion matrix tersebut adalah:

- A (True Positive-TP): proporsi benar dalam dataset kategori benar.
- B (False Negative-FN): proporsi salah dalam dataset kategori salah.
- C (False Positive-FP): proporsi salah dalam dataset kategori benar.
- D (True Negative-TN): proporsi benar dalam dataset kategori salah.

Nilai Akurasi (acc) merupakan proporsi jumlah prediksi yang benar. Persamaannya adalah sebagai berikut:

$$acc = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Dataset Hepatitis dengan jumlah class 1 = 31 dan 2 = 124 berikut ini hasil prediksi model dibawah ini.

Nilai TP = 25
 Nilai TN = 115
 Nilai FP = 6
 Nilai FN = 5

Dataset Ginjal Kronis dengan jumlah class 1 = 145 dan 2 = 237 berikut ini hasil prediksi model dibawah ini.

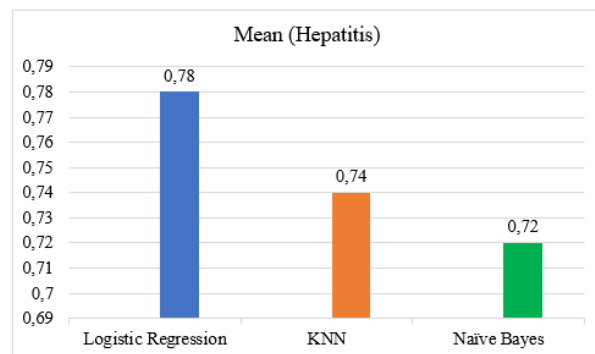
Nilai TP = 100
 Nilai TN = 220
 Nilai FP = 45
 Nilai FN = 25

3. Hasil dan Pembahasan

Tahapan proses missing value melakukan proses data preprocessing dengan menggunakan metode Mean dan KNN Imputation untuk mengisi nilai value yang kosong pada dataset kemudian tahapan normalization dilakukan untuk merubah nilai class dalam rentang nilai 1 dan 2 kemudian di proses menggunakan metode klasifikasi. Dari hasil metode klasifikasi dataset dengan missing value metode Mean dan KNN Imputation tersebut akan diambil nilai Accuracy untuk di analisa nantinya dengan menggunakan metode Logistic Regression, KNN dan Naïve Bayes.

3.1 Metode Imputasi Mean (Hepatitis)

Pada evaluasi penelitian ini hasil nilai Accuracy data kecil (Hepatitis) menggunakan metode imputasi Mean pada Gambar 2 sebagai berikut:

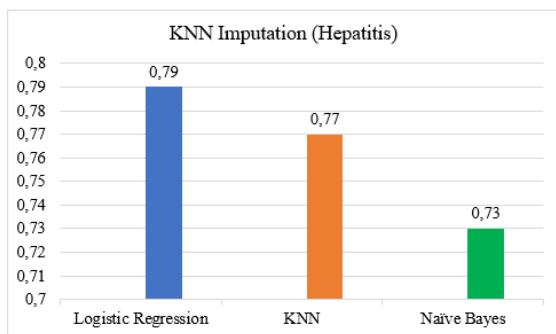


Gambar 2. Metode Imputasi Mean (Hepatitis)

Dari Gambar 2 hasil tertinggi dari nilai Accuracy pada dataset Hepatitis menggunakan metode Imputasi Mean adalah metode Klasifikasi Logistic Regression dengan nilai tertinggi Accuracy = 0,78. Penanganan missing value pada dataset Hepatitis menggunakan metode imputasi Mean dapat meningkatkan nilai akurasi pada metode Logistic Regression.

3.2 Metode Imputasi KNN Imputation (Hepatitis)

Pada evaluasi penelitian ini hasil nilai Accuracy dari data kecil (Hepatitis) menggunakan metode imputasi KNN Imputation pada Gambar 3 sebagai berikut:

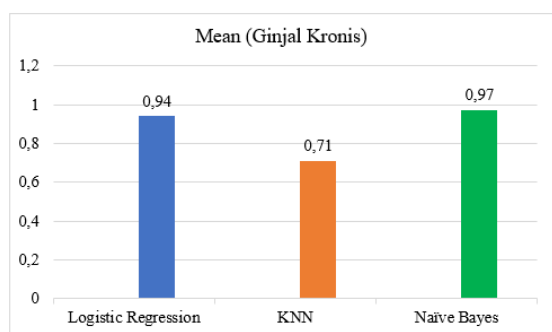


Gambar 3. Metode Imputasi KNN Imputation (Hepatitis)

Dari Gambar 3 hasil tertinggi dari nilai Accuracy pada dataset Hepatitis menggunakan metode Imputasi KNN Imputation adalah metode Klasifikasi Logistic Regression dengan nilai tertinggi Accuracy = 0,79. Penanganan missing value pada dataset Hepatitis menggunakan metode imputasi KNN Imputation dapat meningkatkan nilai akurasi pada metode Logistic Regression.

3.3 Metode Imputasi Mean (Ginjal Kronis)

Pada evaluasi penelitian ini hasil nilai Accuracy dari data kecil (Ginjal Kronis) menggunakan metode imputasi Mean pada Gambar 4 sebagai berikut:

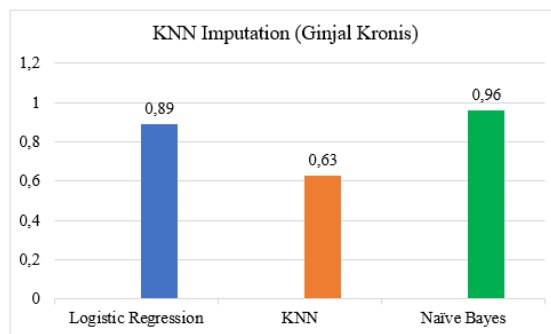


Gambar 4. Metode Imputasi Mean (Ginjal Kronis)

Dari Gambar 4 hasil tertinggi dari nilai Accuracy pada dataset Ginjal Kronis menggunakan metode Imputasi Mean adalah metode Klasifikasi Naive Bayes dengan nilai tertinggi Accuracy = 0,97. Penanganan missing value pada dataset Ginjal Kronis menggunakan metode imputasi Mean dapat meningkatkan nilai akurasi pada metode Naive Bayes.

3.4 Metode Imputasi KNN Imputation (Ginjal Kronis)

Pada evaluasi penelitian ini hasil nilai Accuracy dari data kecil (Ginjal Kronis) menggunakan metode imputasi KNN Imputation pada Gambar 5 sebagai berikut:



Gambar 5. Metode Imputasi KNN Imputation (Ginjal Kronis)

Dari Gambar 5 hasil tertinggi dari nilai Accuracy pada dataset Ginjal Kronis menggunakan metode Imputasi KNN Imputation adalah metode Klasifikasi Naive Bayes dengan nilai tertinggi Accuracy = 0,96. Penanganan missing value pada dataset Ginjal Kronis menggunakan metode imputasi KNN Imputation dapat meningkatkan nilai akurasi pada metode Naive Bayes.

4. Kesimpulan dan Saran

Berdasarkan hasil eksperimen yang telah dilakukan dari penelitian ini yaitu bahwa pengujian metode imputasi Mean dan KNN Imputation dengan metode klasifikasi mampu menangani data kosong dengan jumlah missing value sedikit maupun banyak dengan menghasilkan nilai accuracy mencapai kinerja prediksi yang lebih besar dibandingkan dengan menggunakan missing value nilai 0. Kemudian untuk dataset Hepatitis nilai Accuracy tinggi menggunakan metode imputasi KNN Imputasi dengan nilai 0,79 menggunakan metode Logistic Regression dan dataset Ginjal Kronis nilai Accuracy tinggi menggunakan metode imputasi Mean dengan nilai 0,97 dengan menggunakan metode Naive Bayes. Hal tersebut menunjukkan bahwa proses metode imputasi terhadap nilai kosong di setiap column dataset kecil pada tahap data preprocessing memberikan pengaruh terhadap nilai Accuracy metode Mean dan KNN Imputation pada metode klasifikasi. Penelitian ini yang dilakukan hanya menggunakan 2 dataset kecil yang terdapat missing value, menggunakan metode imputasi Mean dan KNN Imputation kemudian 3 metode klasifikasi Logistic Regression, KNN dan Naive Bayes untuk menangani missing value data kecil. Saran penelitian selanjutnya menggunakan metode imputasi dan metode klasifikasi yang lain selain pada penelitian

ini untuk meningkatkan prediksi hasil Accuracy pada dataset yang terdapat missing value.

Daftar Pustaka:

- (SENASIF), F. P.-S. N. S. I., & 2022, undefined. (2022). Pengujian Metode SMOTE Untuk Penanganan Data Tidak Seimbang Pada Dataset Binary. *Jurnalfti.Unmer.Ac.Id*, 2022(September), 3200–3208. <https://jurnalfti.unmer.ac.id/index.php/senasif/article/view/403>
- Amin, M. S., Chiam, Y. K., & Varathan, K. D. (2019). Identification of significant features and data mining techniques in predicting heart disease. *Telematics and Informatics*, 36, 82–93. <https://doi.org/10.1016/j.tele.2018.11.007>
- Dinh, D. T., Huynh, V. N., & Sriboonchitta, S. (2021). Clustering mixed numerical and categorical data with missing values. *Information Sciences*, 571, 418–442. <https://doi.org/10.1016/j.ins.2021.04.076>
- Li, L., Zhang, J., Wang, Y., & Ran, B. (2019). Missing value imputation for traffic-related time series data based on a multi-view learning method. *IEEE Transactions on Intelligent Transportation Systems*, 20(8), 2933–2943. <https://doi.org/10.1109/TITS.2018.2869768>
- Lin, W. C., & Tsai, C. F. (2020). Missing value imputation: a review and analysis of the literature (2006–2017). *Artificial Intelligence Review*, 53(2), 1487–1509. <https://doi.org/10.1007/s10462-019-09709-4>
- Liu, Z. G., Pan, Q., Dezert, J., & Martin, A. (2016). Adaptive imputation of missing values for incomplete pattern classification. *Pattern Recognition*, 52, 85–95. <https://doi.org/10.1016/j.patcog.2015.10.001>
- Martinez, E. S., Maldonado, S. V., Wu, A. S., McMahan, R. P., Liu, X., & Oakley, B. (2022). Effects of imputation strategy on genetic algorithms and neural networks on a binary classification problem. *GECCO 2022 - Proceedings of the 2022 Genetic and Evolutionary Computation Conference*, 1272–1280. <https://doi.org/10.1145/3512290.3528863>
- Misdram, M., Muljono, Purwanto, & Noersongko, E. (2023). Gradually Generative Adversarial Networks Method for Imbalanced Datasets. *International Journal of Advanced Computer Science and Applications*, 14(4), 51–58. <https://doi.org/10.14569/IJACSA.2023.0140408>
- Misdram, M., Noersongko, E., Syukur, A., Faculty, P., Muljono, M., Agus Santoso, H., & Ignatius Moses Setiadi, D. R. (2020). Analysis of imputation methods of small and unbalanced datasets in classifications using naïve bayes and particle swarm optimization. *Proceedings - 2020 International Seminar on Application for Technology of Information and Communication: IT Challenges for Sustainability, Scalability, and Security in the Age of Digital Disruption, ISemantic 2020*, 115–119. <https://doi.org/10.1109/iSemantic50169.2020.9234225>
- Muti, D., & Pamuji, F. Y. (2023). Analisis Perbandingan Metode Naïve Bayes dan K-NN dalam Klasifikasi Kelayakan Keluarga Terdaftar DTKS Penerimaan Bantuan Sosial di Desa Dubesi. *September*, 3753–3761.
- Nilashi, M., Ahmadi, H., Shahmoradi, L., Ibrahim, O., & Akbari, E. (2019). A predictive method for hepatitis disease diagnosis using ensembles of neuro-fuzzy technique. *Journal of Infection and Public Health*, 12(1), 13–20. <https://doi.org/10.1016/j.jiph.2018.09.009>
- Pamuji, F. Y., Dwi, S., & Putri, A. (2021). Komparasi Metode SMOTE dan ADASYN Untuk Penanganan Data Tidak Seimbang MultiClass. 331–338.
- Pamuji, F. Y., & Ramadhan, V. P. (2021). Komparasi Algoritma Random Forest dan Decision Tree untuk Memprediksi Keberhasilan Immunotherapy. *Jurnal Teknologi Dan Manajemen Informatika*, 7(1), 46–50. <https://doi.org/10.26905/jtmi.v7i1.5982>
- Pamuji, F. Y., & Soeleman, M. A. (2020). Improved number detection for low resolution image using the canny algorithm. *Proceedings - 2020 International Seminar on Application for Technology of Information and Communication: IT Challenges for Sustainability, Scalability, and Security in the Age of Digital Disruption, ISemantic 2020*, 638–642. <https://doi.org/10.1109/iSemantic50169.2020.9234190>
- Purwar, A., & Singh, S. K. (2015). Hybrid prediction model with missing value imputation for medical data. *Expert Systems with Applications*, 42(13), 5621–5631. <https://doi.org/10.1016/j.eswa.2015.02.050>
- Ramadhan, V. P., & Pamuji, F. Y. (2022). *Jurnal Teknologi dan Manajemen Informatika Analisis Perbandingan Algoritma Forecasting dalam Prediksi Harga Saham LQ45 PT Bank Mandiri Sekuritas (BMRI)*. 8(1), 39–45.
- Sanlialp, I., & Ozturk, M. M. (2020). Investigating the Impact of Code Refactoring Techniques on Energy Consumption in Different Object-Oriented Programming Languages. In *Lecture Notes on Data Engineering and Communications Technologies* (Vol. 43, Issue Icaime). https://doi.org/10.1007/978-3-030-36178-5_12
- Wei, R., Wang, J., Su, M., Jia, E., Chen, S., Chen, T., & Ni, Y. (2018). Missing Value Imputation

Approach for Mass Spectrometry-based Metabolomics Data. *Scientific Reports*, 8(1), 1–10. <https://doi.org/10.1038/s41598-017-19120-0>

Xu, X., Chen, W., & Sun, Y. (2019). Over-sampling

algorithm for imbalanced data classification. *Journal of Systems Engineering and Electronics*, 30(6), 1182–1191. <https://doi.org/10.21629/JSEE.2019.06.12>

