

**PENANGANAN DATA MISSING VALUE DENGAN
KNN-IMPUTASI PADA ALGORITMA K- NEAREST
NEIGHBOUR UNTUK KLASIFIKASI DATA
PELANGGAN INDIHOME PT TELKOM**



SKRIPSI

**Diajukan Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer**

Oleh:

**Faizhal Rustyanto Ardy Widodo
201769040052**

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS YUDHARTA PASURUAN
2021**

PERNYATAAN PENULIS

JUDUL : PENANGANAN DATA MISSING VALUE
DENGAN KNN-IMPOTASI PADA ALGORITMA K-
NEAREST NEIGHBOUR UNTUK KLASIFIKASI
DATA PELANGGAN INDIHOME PT TELKOM

NAMA : FAIZHAL RUSTYANTO ARDY WIDODO

NIM : 201769040052

“Saya menyatakan dan bertanggung jawab dengan sebenarnya bahwa Skripsi ini adalah hasil karya saya sendiri kecuali cuplike dan ringkasa yang masing-masing telah saya jelaskan sumbernya. Jika pada waktu selanjutnya ada pihak lain yang mengklaim bahwa Skripsi ini sebagai karyanya, yang disertai bukti-bukti yang cukup, maka saya bersedia untuk dibatalkan gelar Sarjana Komputer saya beserta segala hak dan kewajiban yang melekat pada gelar tersebut”.

Pasuruan, 07 Agustus 2021



Faizhal Rustyanto Ardy Widodo
Penulis

PERSETUJUAN SKRIPSI

JUDUL : PENANGANAN DATA MISSING VALUE
DENGAN KNN-IMPUTASI PADA
ALGORITMA K-NEAREST NEIGHBOUR
UNTUK KLASIFIKASI DATA PELANGGAN
INDIHOME PT TELKOM

NAMA : FAIZHAL RUSTYANTO ARDY WIDODO

NIM : 201769040052

Skripsi ini telah diperiksa dan disetujui

Pasuruan, 07 Agustus 2021



Muhammad Imron Rosadi,
S.Kom, M. Kom.,
NIP.Y.0690213121

Kaprodi,

Pembimbing

A handwritten signature in black ink, appearing to read 'Moch. Lutfi', is written over a faint circular stamp.

Moch. Lutfi, S.Kom, M. Kom.,
NIP.Y.06916030004

PENGESAHAN SKRIPSI

JUDUL : PENANGANAN DATA MISSING VALUE
DENGAN KNN-IMPUTASI PADA ALGORITMA
K-NEAREST NEIGHBOUR UNTUK KLASIFIKASI
DATA PELANGGAN INDIHOME PT TELKOM

NAMA : FAIZHAL RUSTYANTO ARDY WIDODO

NIM : 201769040052

Skripsi ini telah diujikan dan dipertahankan di depan Dewan Penguji pada Sidang Skripsi tanggal, 07 Agustus 2021. Menurut pandangan kami, Skripsi ini memadai dari segi kualitas untuk tujuan penganugerahan gelar

Sarjana Komputer (S.Kom)
Pasuruan, 07 Agustus 2021

Pembimbing,



Moch. Lutfi, S.Kom, M. Kom.,
NIP. Y. 06916030004
Penguji Utama,

Penguji Pendamping,

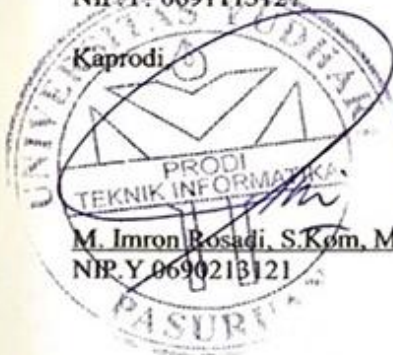


Cahya Bagus Sanjaya, S.Kom., M.Kom
NIP. Y. 0691113127

M. Imron Rosadi, S.Kom, M. Kom.,
NIP. Y. 0690213121

Kaprodi

Dekan Fakultas Teknik,



M. Imron Rosadi, S.Kom, M. Kom.,
NIP. Y. 0690213121

Misbach Munir, ST., MT.
NIP. Y. 0690201015

bsttrak

Data mining adalah salah satu cabang keilmuan yang banyak dipakai dalam menggali informasi dari suatu data yang disajikan. Salah satu permasalahan yang sering muncul dalam penggalan informasi itu adalah adanya missing value pada data. Hal ini dapat membuat tingkat keakuratan dari informasi yang didapat menjadi berkurang. Ada beberapa cara untuk mengatasi hal ini. Salah satunya adalah dengan melakukan pengisian terhadap Missing value tersebut. Cara ini disebut proses imputasi. KNN Imputation adalah salah satu metoda imputasi yang dikembangkan dari metode klasifikasi KNN (K-Nearest Neighbour). Metoda ini adalah salah satu metoda imputasi berbasis statistik. Pada intinya metoda ini akan mencoba mengisi missing value dari suatu data dengan mengambil nilai dari tetangga terdekatnya. Pengukuran terhadap performansi metoda ini terhadap proses klasifikasi, yang merupakan salah satu proses dalam data mining akan dianalisis dari akurasi saat data diklasifikasi setelah diimputasi. Diharapkan setelah mengalami proses imputasi dengan metoda ini, penggalan informasi terhadap suatu data yang memiliki missing value menjadi lebih baik dan akurat dari sebelumnya.

Kata Kunci : imputasi, KNN Imputation, missing value, klasifikasi

Abstract

Data mining is one of the branches of science that is widely used in extracting information from the data presented. One of the problems that often arise in extracting information is the Missing value in the data. This can reduce the accuracy of the information obtained. There are a few ways to get through this. One of them is to fill in the missing value. This method is called the imputation process. KNN Imputation is one of the imputation methods developed from the KNN (K-Nearest Neighbor) classification method. This method is one of the statistical-based imputation methods. In essence, this method will try to fill in the missing value from a data by taking the value from its nearest neighbor. The measurement of the performance of this method on the classification process, which is one of the processes in data mining, will be analyzed from accuracy when the data is classified after imputation. It is hoped that after experiencing the imputation process with this method, extracting information on a data that has a missing value will be better and more accurate than before.

Keywords: imputation, KNN Imputation, missing value, classification

Kata Pengantar

Puji syukur penulis panjatkan ke hadirat Allah SWT yang telah melimpahkan rahmat, taufiq dan hidayah-Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul : “PENANGANAN DATA MISSING VALUE DENGAN KNN-IMPUTASI PADA ALGORITMA K-NEAREST NEIGHBOUR UNTUK KLASIFIKASI DATA PELANGGAN INDIHOME PT TELKOM”. Skripsi ini diajukan untuk memenuhi salah satu syarat dalam menyelesaikan Program Sarjana (S-1) pada Program Sarjana Fakultas Teknik Universitas Yudharta Pasuruan. Penulis menyadari bahwa penyusunan skripsi ini tidak akan terwujud tanpa adanya bantuan dan dorongan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis mengucapkan terima kasih kepada yang terhormat:

1. Dr. H. Kholid Murtadlo, S.E, M.E. selaku Rektor Universitas Yudharta Pasuruan.
2. Misbach Munir, ST., MT. selaku Dekan Fakultas Teknik Universitas Yudharta Pasuruan.
3. Muhammad Imron Rosadi, S.Kom, M. Kom., selaku Ketua Program Studi Informatika Fakultas Teknik Universitas Yudharta Pasuruan
4. Moch. Lutfi, S.Kom, M. Kom. selaku Dosen Pembimbing yang telah meluangkan waktu untuk memberikan bimbingan, saran, motivasi dan pengarahan kepada penulis.
5. Seluruh pegawai PT Telkom Malang atas bantuannya memberikan informasi yang dibutuhkan penulis.
6. Kedua Orang tua tercinta yang telah memberikan Doa restu nasehat serta dorongan semangat yang sangat berarti bagi penulis.
7. Semua pihak yang telah membantu dalam penulisan skripsi yang tidak dapat disebutkan satu-persatu.

Semoga Allah SWT senantiasa melimpahkan rahmat dan hidayah-Nya kepada semua pihak yang telah membantu penulis. penulis menyadari bahwa dalam penulisan skripsi ini masih banyak kekurangan dan masih jauh dari sempurna, maka dari itu dengan kerendahan hati penulis mengharapkan saran dan kritik yang bersifat membangun dari semua pihak demi perbaikan dan penyempurnaan di masa mendatang.

Pasuruan, 07 Agustus 2021

Faizhal Rustyanto Ardy Widodo

Daftar Isi

PERNYATAAN PENULIS	ii
PERSETUJUAN SKRIPSI.....	iii
PENGESAHAN SKRIPSI	iv
Abstrak	v
<i>Abstract</i>	vi
Kata Pengantar	vii
Daftar Isi.....	ix
Daftar Gambar	xii
Daftar Tabel.....	xiii
BAB I.....	1
PENDAHULUAN.....	1
1.1 Latar belakang	1
1.2 Rumusan Masalah	4
1.3 Batasan Masalah.....	4
1.4 Tujuan.....	5
1.5 Manfaat Penelitian.....	5
BAB II.....	7
TINJAUAN PUSTAKA.....	7
2.1 Penelitian Terdahulu.....	7

2.2	Landasan Teori	16
a.	K- Nearest Neighbour	16
b.	K- Nearest Neighbour Imputation (KNNI)	18
c.	<i>Cross Validation</i>	19
d.	Bahasa Pemrograman R	20
BAB III.....		23
METODE PENELITIAN		23
3.1	Instrumen Penelitian.....	23
3.1.1	Bahan.....	23
3.1.2	Peralatan	23
3.2	Metode Penelitian.....	24
3.3	Metode Pemilihan Data	29
3.4	Desain Penelitian	33
3.4.1	Pemahaman Data (<i>Data Understanding</i>)	33
3.4.2	Pengolahan Data (<i>Data Preparation</i>).....	33
3.4.3	Pemodelan (<i>Modelling</i>)	34
BAB IV		35
ANALISIS DAN PEMBAHASAN.....		35
4.1	Pembahasan	35
4.1.1	<i>Data Selection</i>	35
4.1.2	<i>Preprocessing</i>	38
4.1.3	Transformasi.....	43

4.2	Hasil	48
4.2.1	Implementasi <i>K-Nearest Neighbor</i>	48
4.2.2	Hasil Akurasi	53
BAB V		59
KESIMPULAN DAN SARAN		59
5.1	Kesimpulan	59
5.2	Saran	59
Daftar Pustaka		61

Daftar Gambar

Diagram Penelitian (1) 2.1.....	8
Diagram Penelitian (2) 2.2.....	10
Diagram Penelitian (3) 2.3.....	12
Diagram metode penelitian 3.1.....	25
Diagram penelitian 3.2.....	33
Diagram Imputasi 4.1.....	41
Grafik 4.2 Knn (1).....	55
Grafik 4.3 Knn (2).....	57

Daftar Tabel

Tabel 2.1 Penelitian Terdahulu	15
Tabel 3.1 Data Primer	31
Tabel 4.1 Dataset	37
Tabel 4.2 Missing Value	38
Tabel 4.3 Dataset Imputasi	43
Tabel 4.4 Dataset Training	46
Tabel 4.5 Dataset Test	47
Tabel 4.6 Dataset tes	49
Tabel 4.7 Dataset training tes	49
Tabel 4.8 K (1)	53
Tabel 4.9 Akurasi (1)	54
Tabel 4.10 Confusion Matrix	55
Tabel 4.11 Akurasi (2)	56
Tabel 4.12 Confusion Matrix (2)	57