

**PENGARUH BALANCING DATA DAN NORMALISASI DATA  
PADA PERFORMA K-NEAREST NEIGHBOR DALAM  
KLASIFIKASI KUALITAS WINE**

**SKRIPSI**

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**INDRA PUTRA GEMILANG**

**19.11.2780**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2022**

**PENGARUH BALANCING DATA DAN NORMALISASI DATA  
PADA PERFORMA K-NEAREST NEIGHBOR DALAM  
KLASIFIKASI KUALITAS WINE**

**SKRIPSI**

untuk memenuhi salah satu syarat mencapai derajat Sarjana  
Program Studi Informatika



disusun oleh

**INDRA PUTRA GEMILANG**

**19.11.2780**

Kepada

**FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA**

**2022**

**HALAMAN PERSETUJUAN**

**SKRIPSI**

**PENGARUH BALANCING DATA DAN NORMALISASI DATA PADA  
PERFORMA K-NEAREST NEIGHBOR DALAM KLASIFIKASI  
KUALITAS WINE**

yang disusun dan diajukan oleh

**INDRA PUTRA GEMILANG**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 19 Desember 2022

**Dosen Pembimbing,**

**Prof. Dr. Kusriani, M.Kom**  
**NIK. 190302106**

**HALAMAN PENGESAHAN**

**SKRIPSI**

**PENGARUH BALANCING DATA DAN NORMALISASI DATA PADA  
PERFORMA K-NEAREST NEIGHBOR DALAM KLASIFIKASI  
KUALITAS WINE**

yang disusun dan diajukan oleh

**INDRA PUTRA GEMILANG**

**19.11.2780**

Telah dipertahankan di depan Dewan Penguji  
pada tanggal 19 Desember 2022

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Prof. Dr. Kusrini, M.Kom**  
**NIK. 190302106**

**Dina Maulina, M.Kom**  
**NIK. 190302250**

**Mardhiya Hayaty, S.T., M.Kom**  
**NIK. 190302108**

Skrripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 19 Desember 2022

**DEKAN FAKULTAS ILMU KOMPUTER**

**Hanif Al Fatta, S.Kom., M.Kom.**  
**NIK. 190302096**

## HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : **INDRA PUTRA GEMILANG**  
NIM : **19.11.2780**

Menyatakan bahwa Skripsi dengan judul berikut:

**PENGARUH BALANCING DATA DAN NORMALISASI DATA PADA PERFORMA K-NEAREST NEIGHBOR DALAM KLASIFIKASI KUALITAS WINE**

Dosen Pembimbing : **Kusrini, Prof., Dr., M.Kom**

1. Karya tulis ini adalah benar-benar **ASLI** dan **BELUM PERNAH** diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan **gagasan, rumusan dan penelitian SAYA sendiri**, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab **SAYA**, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini **SAYA** buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka **SAYA** bersedia menerima **SANKSI AKADEMIK** dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 19 Desember 2022

Yang Menyatakan,



Indra Putra Gemilang

## KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah SWT yang telah melimpahkan banyak rahmat dan nikmat sehingga penulis dapat menyelesaikan penyusunan skripsi ini, penulisan skripsi ini dilakukan dalam rangka memenuhi salah satu syarat untuk mencapai gelar sarjana komputer program studi informatika di Universitas Amikom Yogyakarta. Skripsi ini dapat diselesaikan berkat adanya dukungan ilmu, saran maupun kritik yang di terima penulis oleh karena itu peneliti ingin mengucapkan terima kasih kepada:

1. Kedua orang tua dan keluarga penulis yang selalu mendoakan dan memberikan dukungan dalam pembuatan skripsi ini.
2. Ibu Prof. Dr. Kusrini, M.Kom selaku dosen pembimbing skripsi, atas waktu yang telah diberikan untuk membimbing serta memberikan masukan kepada penulis pada pengerjaan skripsi ini.
3. Seluruh dosen dan staff karyawan Universitas Amikom Yogyakarta
4. Teman-teman yang telah memberikan dukungan pada pembuatan skripsi ini.

Penulis berharap bahwa dibuatnya skripsi ini dapat bermanfaat bagi pembaca, penulis menyadari bahwa masih ada kekurangan dalam pembuatan skripsi ini, semoga yang menjadi kekurangan dalam skripsi ini dapat disempurnakan.

Yogyakarta, 30 November 2022

Penulis

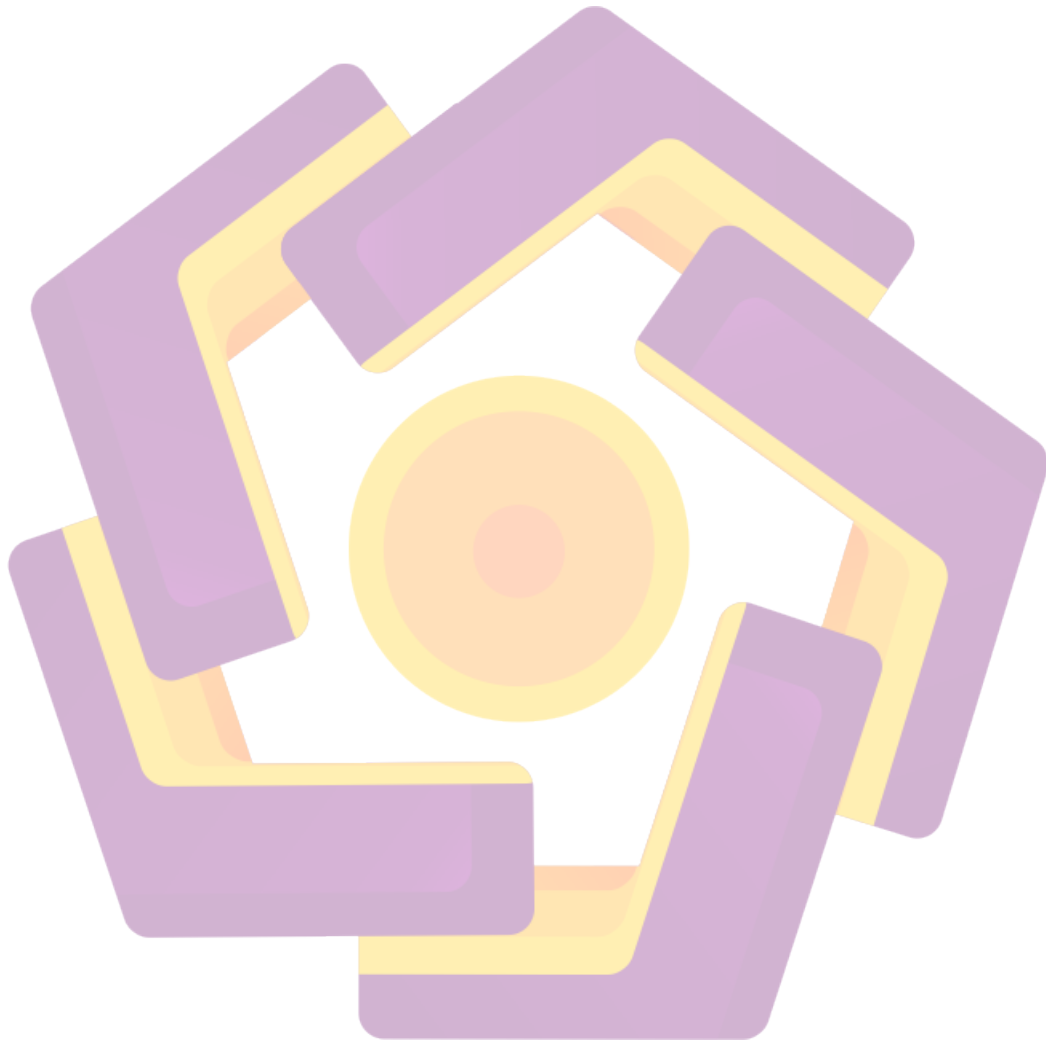
## DAFTAR ISI

HALAMAN JUDUL .....	i
HALAMAN PERSETUJUAN.....	ii
HALAMAN PENGESAHAN .....	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI .....	iv
KATA PENGANTAR .....	v
DAFTAR ISI.....	vi
DAFTAR TABEL.....	ix
DAFTAR GAMBAR .....	x
DAFTAR LAMBANG DAN SINGKATAN .....	xi
DAFTAR ISTILAH .....	xii
INTISARI .....	xiii
ABSTRACT.....	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah .....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah .....	2
1.4 Tujuan Penelitian .....	2
1.5 Manfaat Penelitian .....	3
1.6 Sistematika Penulisan .....	3
BAB II TINJAUAN PUSTAKA .....	4
2.1 Studi Literatur .....	4
2.2 Dasar Teori .....	15
2.2.1 <i>Wine</i> .....	15
2.2.2 <i>Data Mining</i> .....	16
2.2.3 Klasifikasi .....	17
2.2.4 <i>Supervised Learning</i> .....	18
2.2.4.1 <i>K-Nearest Neighbors</i> .....	19
2.2.5 Data Normalisasi.....	20

2.2.5.1	<i>Min-max Normalization</i> .....	20
2.2.6	<i>Imbalance Data</i> .....	21
2.2.6.1	SMOTE.....	22
2.2.7	Matriks Evaluasi .....	22
2.2.7.1	Confusion matrix .....	22
<b>BAB III METODE PENELITIAN</b> .....		25
3.1	Alur Penelitian .....	25
3.1.1	Dataset.....	26
3.1.2	<i>Preprocessing</i> .....	27
3.1.3	Implementasi Algoritma .....	27
3.1.4	Evaluasi.....	27
3.1.5	Analisis statistik .....	27
3.2	Alat Dan Bahan.....	28
3.3	Penjelasan dan Perhitungan Manual Metode.....	28
3.3.1	Algoritma <i>K-Nearest Neighbor</i> .....	28
3.3.2	Algoritma <i>Min-Max</i> .....	30
3.3.3	Algoritma SMOTE.....	32
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....		33
4.1	Pengumpulan Data.....	33
4.2	<i>Preprocessing</i> .....	33
4.3	<i>Split data</i> .....	35
4.4	Min-max Normalization .....	35
4.5	SMOTE.....	36
4.6	Kombinasi <i>Min-max</i> dan SMOTE .....	37
4.7	K-NN tanpa <i>Min-max</i> dan SMOTE.....	38
4.8	K-NN dengan <i>Min-max</i> .....	40
4.9	K-NN dengan SMOTE .....	42
4.10	K-NN dengan Min-max dan SMOTE.....	44

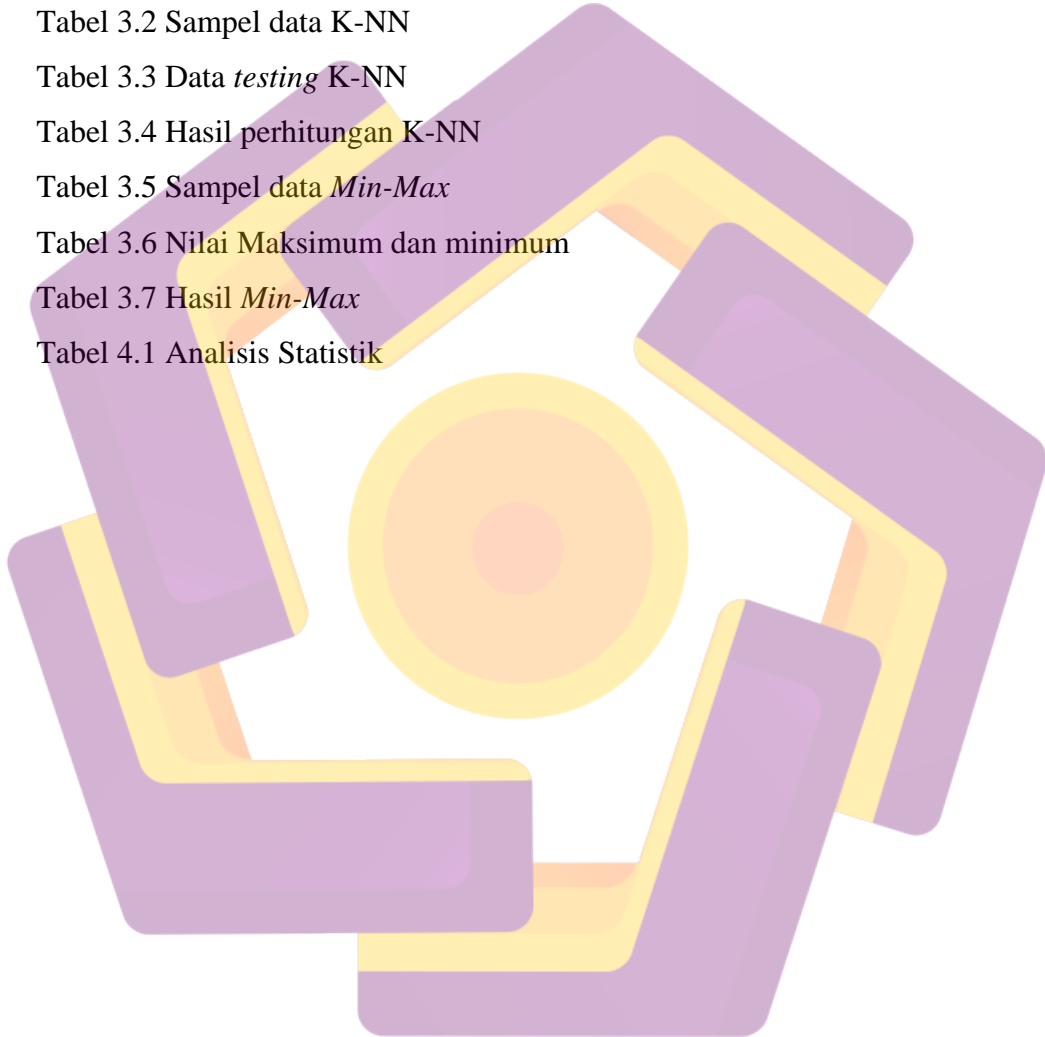


4.11 Analisis Statistik .....	48
BAB V PENUTUP .....	50
5.1 Kesimpulan .....	50
5.2 Saran .....	51
REFERENSI .....	52



## DAFTAR TABEL

Tabel 2.1. keaslian penelitian	7
Tabel 2.2 <i>Confusion matrix</i>	23
Tabel 3.1. kebutuhan perangkat lunak dan perangkat keras	28
Tabel 3.2 Sampel data K-NN	28
Tabel 3.3 Data <i>testing</i> K-NN	29
Tabel 3.4 Hasil perhitungan K-NN	29
Tabel 3.5 Sampel data <i>Min-Max</i>	30
Tabel 3.6 Nilai Maksimum dan minimum	31
Tabel 3.7 Hasil <i>Min-Max</i>	32
Tabel 4.1 Analisis Statistik	48

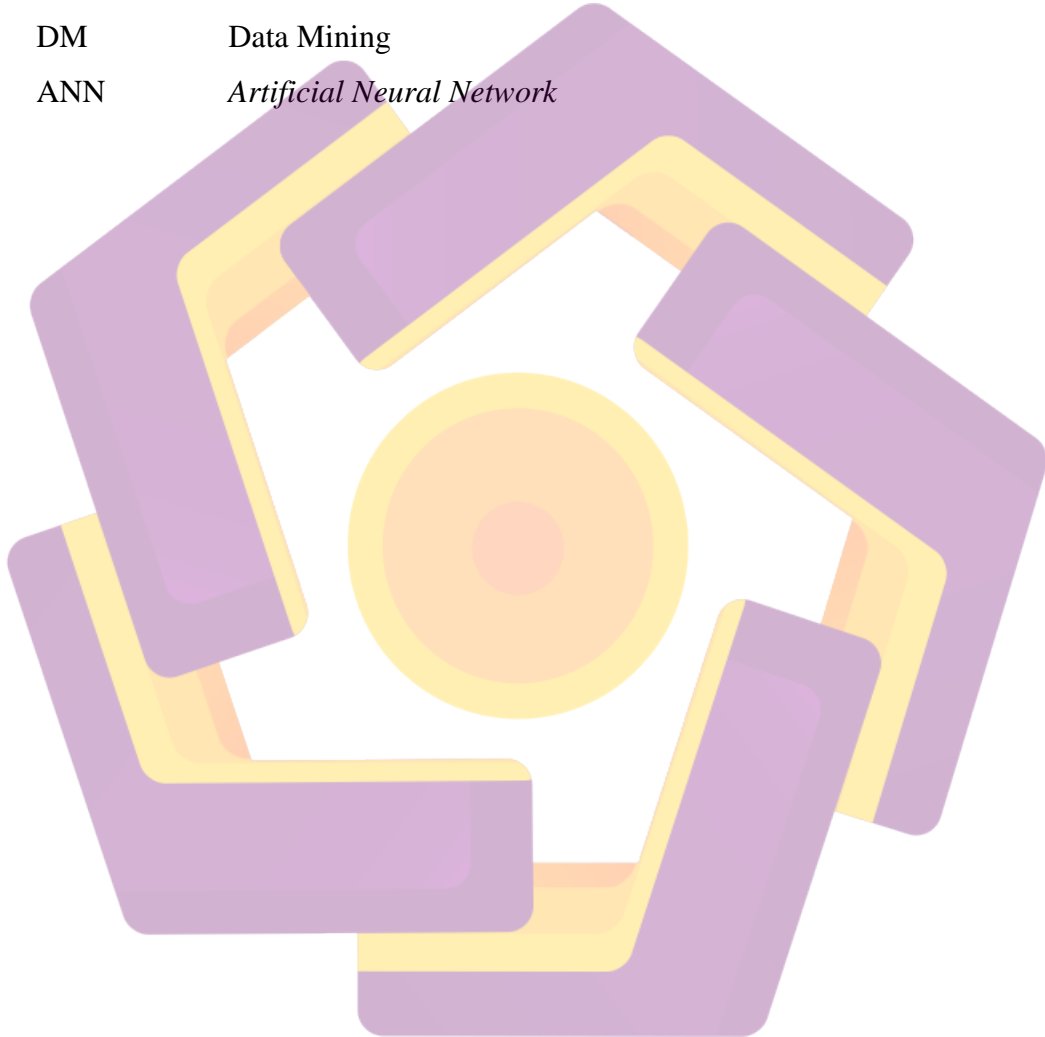


## DAFTAR GAMBAR

Gambar 2.1. Alur pemrosesan data	17
Gambar 3.1. Alur penelitian	25
Gambar 4.1. Dataset <i>red wine quality</i>	33
Gambar 4.2. <i>Import data</i>	34
Gambar 4.3. Hasil <i>isna()</i> dan <i>sum()</i>	34
Gambar 4.4. Mengubah <i>quality</i> dari 1-10 menjadi 1 dan 0	35
Gambar 4.5. <i>Split data</i>	35
Gambar 4.6 <i>Min-max Normalization</i>	36
Gambar 4.7 SMOTE	37
Gambar 4.8 <i>Min-max</i> dan SMOTE	38
Gambar 4.9 K-NN tanpa <i>Min-max</i> dan SMOTE	39
Gambar 4.10 <i>Confusion Matrix</i> K-NN tanpa <i>Min-max</i> dan SMOTE	40
Gambar 4.11 Hasil evaluasi K-NN tanpa <i>Min-max</i> dan SMOTE	40
Gambar 4.12 K-NN dengan <i>Min-max</i>	41
Gambar 4.13 <i>Confusion Matrix</i> K-NN dengan <i>Min-max</i>	42
Gambar 4.14 hasil evaluasi K-NN dengan <i>Min-max</i>	42
Gambar 4.15 K-NN dengan SMOTE	43
Gambar 4.16 <i>Confusion Matrix</i> K-NN dengan SMOTE	44
Gambar 4.17 Hasil evaluasi K-NN dengan SMOTE	44
Gambar 4.18 Proses balancing data	45
Gambar 4.19 proses normalisasi data	45
Gambar 4.20 Mencari jumlah <i>K-Neighbor</i>	46
Gambar 4.21 <i>Confusion Matrix</i> K-NN dengan <i>Min-max</i> dan SMOTE	46
Gambar 4.22 Hasil evaluasi K-NN dengan <i>Min-max</i> dan SMOTE	47

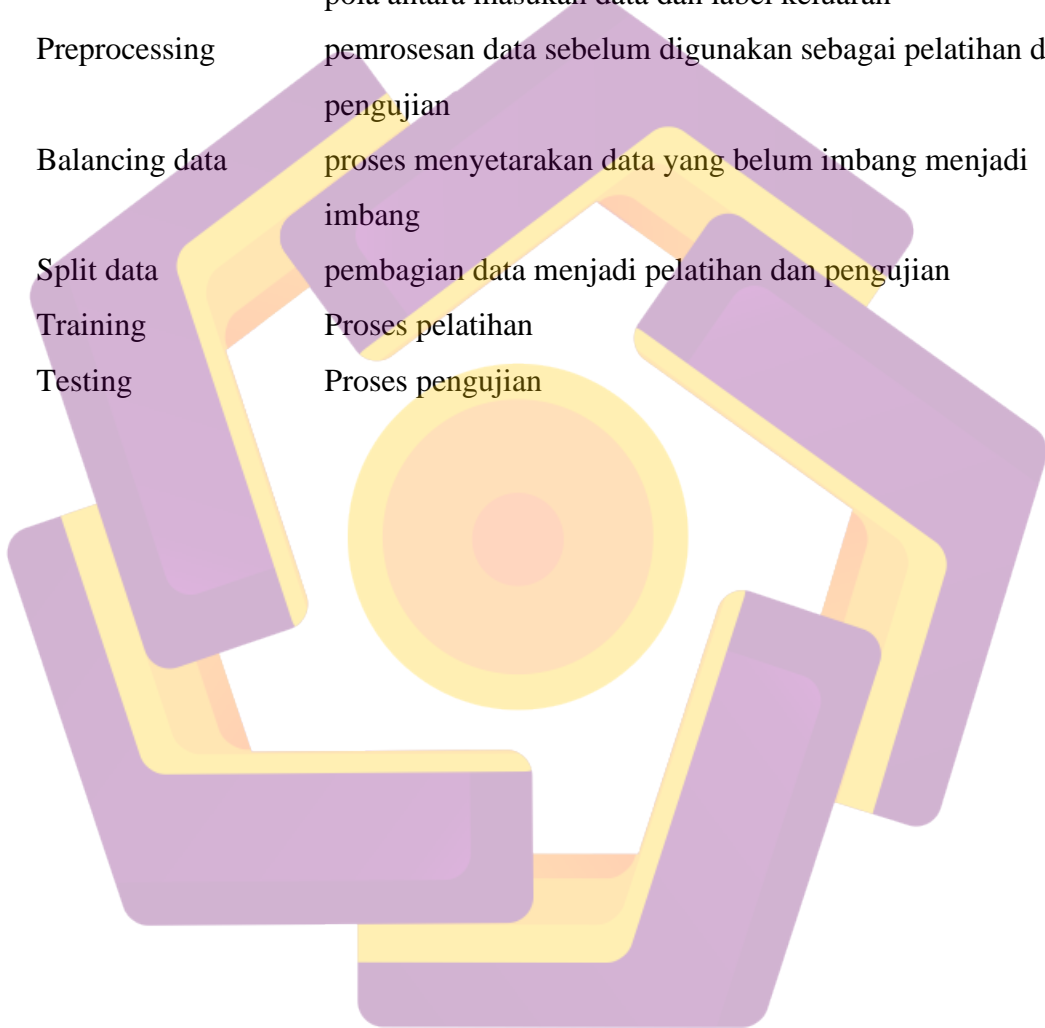
## DAFTAR LAMBANG DAN SINGKATAN

K-NN	K-Nearest Neighbors
SVM	<i>Support Vector Machine</i>
ELM	<i>Extreme Learning Machine</i>
SMOTE	<i>synthetic minority over-sampling technique</i>
DM	Data Mining
ANN	<i>Artificial Neural Network</i>



## DAFTAR ISTILAH

Data mining	proses pengumpulan data
Machine Learning	Salah satu cabang dari kecerdasan buatan
supervised learning	pendekatan dalam pembuatan kecerdasan untuk mengenali pola antara masukan data dan label keluaran
Preprocessing	pemrosesan data sebelum digunakan sebagai pelatihan dan pengujian
Balancing data	proses menyetarakan data yang belumimbang menjadiimbang
Split data	pembagian data menjadi pelatihan dan pengujian
Training	Proses pelatihan
Testing	Proses pengujian



## INTISARI

*K-Nearest neighbors* atau yang biasa disebut (K-NN) merupakan algoritma *supervised learning* dalam *machine learning*. K-NN biasanya digunakan untuk klasifikasi dan regresi. Sebelum data bisa digunakan sebagai klasifikasi data harus di proses terlebih dahulu dan proses ini disebut *pre-processing* ada banyak cara teknik *pre-processing* data diantaranya adalah normalisasi data dan *balancing data*. Normalisasi data adalah proses untuk menyamakan rentang nilai dari yang berbeda menjadi sama sedangkan *balancing data* adalah proses menyamakan data kelas yang tidakimbang menjadiimbang.

Data yang akan digunakan untuk penelitian ini adalah data kualitas *red wine* yang di dapatkan dari *website* Kaggle.com dan data ini bersifat publik. Data set ini memiliki 1599 data dan memiliki 12 atribut, untuk atribut yang digunakan untuk klasifikasi bernama *quality* yang berisi nilai 1-10 semakin besar nilainya semakin tinggi kualitas *wine* tersebut. Penelitian ini berfokuskan pada pengaruh proses *pre-processing data* antara *balancing data* dan normalisasi data pada performa *K-Nearest Neighbors* dalam mengklasifikasi data dan menentukan jumlah *K-Neighbors* yang di perlukan untuk mencapai akurasi tinggi.

Metode yang mendapatkan akurasi tertinggi pada penelitian ini adalah metode K-NN dengan Min-Max dengan hasil akurasi sebanyak 80.75% dengan  $k = 17$ , yang kedua yaitu metode K-NN dengan Min-Max dan SMOTE dengan Akurasi sebanyak 79.75% dengan  $k = 19$ , ketiga yaitu metode K-NN saja dengan akurasi sebanyak 75% dengan  $k = 23$  dan terakhir yaitu metode K-NN dengan SMOTE dengan akurasi sebanyak 74.5% dengan  $k = 23$

**Kata Kunci:** *K-Nearest neighbors*, Normalisasi data, *Balancing data*

## **ABSTRACT**

*K-Nearest neighbor or commonly called (K-NN) is a supervised learning algorithm in machine learning K-NN is usually used for classification and regression. Before data can be used as data classification it must be processed first and this process is called pre-processing. There are many ways of data pre-processing techniques including data normalization and data balancing. Data normalization is the process of equating the range of values from different to equal, while data balancing is the process of equating data classes that are not balanced to be balanced.*

*The data that will be used for this research is red wine quality data obtained from the Kaggle.com website and this data is public. This data set has 1599 data and has 12 attributes, for the attribute used for the classification named quality which contains values 1-10 the greater the value the higher the quality of the wine. This study focuses on the influence of the data pre-processing process between data balancing and data normalization on the performance of K-Nearest Neighbors in classifying data and determining the number of K-Neighbors needed to achieve high accuracy.*

*The method with the highest accuracy in this study is K-NN with Min-Max with an accuracy of 80.75% with  $k = 17$ , the second is K-NN with Min-Max and SMOTE with an accuracy of 79.75% with  $k = 19$ , the third is the K-NN method only with an accuracy of 75% with  $k = 23$  and the last is the K-NN with SMOTE with an accuracy of 74.5% with  $k = 23$*

**Keyword:** *K-Nearest neighbors, Data normalization, Data balancing*