

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

# KLASIFIKASI STATUS MESIN POMPA AIR MENGUNAKAN ALGORITMA C4.5

## TUGAS AKHIR

Disusun Sebagai Salah Satu Syarat  
Untuk Memperoleh Gelar Sarjana Teknik  
Pada Jurusan Teknik Informatika

Oleh

**MUKHTIA KHAIRI ANGGI**

**NIM. 11950115142**



UIN SUSKA RIAU

FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU  
PEKANBARU  
2025



**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

# LEMBAR PERSETUJUAN

## KLASIFIKASI STATUS MESIN POMPA AIR MENGUNAKAN ALGORITMA C4.5

### TUGAS AKHIR

Oleh

**MUKHTIA KHAIRI ANGGI**

**NIM. 11950115142**

Telah diperiksa dan disetujui sebagai Laporan Tugas Akhir  
di Pekanbaru, pada tanggal 07 Januari 2025

Pembimbing I,

**Dr. Lestari Handayani, S.T., M.Kom.**

**NIP. 19811113 200710 2 003**

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**LEMBAR PENGESAHAN**

**KLASIFIKASI STATUS MESIN POMPA AIR  
MENGUNAKAN ALGORITMA C4.5**

Oleh

**MUKHTIA KHAIRI ANGGI**

**NIM. 11950115142**

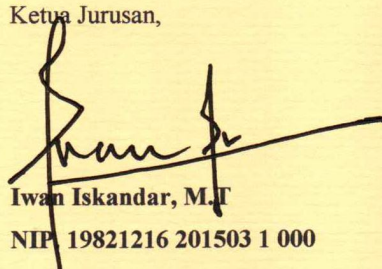
Telah dipertahankan di depan sidang dewan penguji  
sebagai salah satu syarat untuk memperoleh gelar Sarjana Teknik  
pada Universitas Islam Negeri Sultan Syarif Kasim Riau

Pekanbaru, 07 Januari 2025

Mengesahkan,

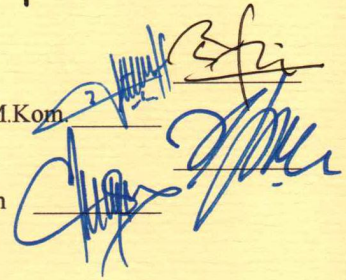
Ketua Jurusan,

  
**Dr. Hartono, M.Pd**  
**NIP. 19640301 199203 1 003**

  
**Iwan Iskandar, M.T**  
**NIP. 19821216 201503 1 000**

**DEWAN PENGUJI**

|              |                                      |
|--------------|--------------------------------------|
| Ketua        | : Yelfi Vitriani, S.Kom, MMSI        |
| Pembimbing I | : Dr. Lestari Handayani, S.T., M.Kom |
| Penguji I    | : Prof. Dr. Okfalisa, S.T., M.Sc     |
| Penguji II   | : Fadhilah Syafria, S.T., M.Kom      |



## LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL

Tugas Akhir yang tidak diterbitkan ini terdaftar dan tersedia di Perpustakaan Universitas Islam Negeri Sultan Syarif Kasim Riau adalah terbuka untuk umum dengan ketentuan bahwa hak cipta pada penulis. Referensi kepustakaan diperkenankan dicatat, tetapi pengutipan atau ringkasan hanya dapat dilakukan seizin penulis dan harus disertai dengan kebiasaan ilmiah untuk menyebutkan sumbernya.

Penggandaan atau penerbitan sebagian atau seluruh Tugas Akhir ini harus memperoleh izin dari Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau. Perpustakaan yang meminjamkan Tugas Akhir ini untuk anggotanya diharapkan untuk mengisi nama, tanda peminjaman dan tanggal pinjam.

### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

**SURAT PERNYATAAN**

Saya yang bertanda tangan dibawah ini :

Nama : Mukhtia Khairi Anggi  
 NIM : 11950115142  
 Tempat/Tgl.Lahir : Payakumbuh, 9 Januari 1999  
 Jurusan/Semester : Teknik Informatika/XI  
 Fakultas : Sains dan Teknologi  
 Lulus Munaqasah : 07 Januari 2025  
 Judul Skripsi : KLASIFIKASI STATUS MESIN POMPA AIR  
 MENGGUNAKAN ALGORITMA C4.5

Menyatakan dengan sebenar-benarnya bahwa :

1. Penulisan Skripsi dengan judul sebagaimana tersebut di atas adalah hasil pemikiran dan penelitian saya sendiri.
2. Semua kutipan pada karya tulis saya ini sudah disebutkan sumbernya.
3. Oleh karena itu Skripsi saya ini, saya nyatakan bebas dari plagiat.
4. Apabila dikemudian hari terbukti terdapat plagiat dalam penulisan skripsi saya tersebut, maka saya bersedia menerima sanksi sesuai peraturan perundang-undangan.

Demikian surat pernyataan ini saya buat dengan penuh kesadaran dan akal sehat.

Pekanbaru, 7 Januari 2025

Yang membuat pernyataan



**MUKHTIA KHAIRI ANGGI**

NIM : 11950115142

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## LEMBAR PERSEMBAHAN

Alhamdulillah Rabbil Aalamin, puji syukur atas karunia Allah SWT yang telah memberikan kesehatan, kelancaran, dan kemudahan sehingga laporan tugas akhir ini dapat terselesaikan dengan baik.

Tugas akhir ini saya persembahkan untuk orang tua dan keluarga yang telah memberikan dukungan, doa, serta motivasi dalam setiap langkah menyelesaikan laporan tugas akhir ini.

Halaman persembahan ini juga ditujukan kepada dosen pembimbing yaitu Ibu Dr. Lestari Handayani, S.T., M.Kom yang telah memberikan masukan, saran, serta motivasi, sehingga laporan tugas akhir ini dapat terselesaikan.

Serta teman-teman yang telah bersama-sama berjuang dalam menyelesaikan laporan tugas akhir ini dan semua pihak yang turut membantu dalam proses penulisan laporan tugas akhir ini.

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## ABSTRAK

Pompa air digunakan sebagai salah satu cara untuk mendapatkan air secara mekanik melalui saluran pipa. Namun, kerusakan pompa seperti pompa mati secara mendadak dan kerusakan lainnya dapat menyebabkan masalah terhadap kehidupan masyarakat. Tujuan penelitian ini yaitu untuk melakukan klasifikasi pada status mesin pompa air tersebut menggunakan algoritma C4.5 agar dapat mengetahui kelas dari status mesin pompa air yang terdiri dari normal, recovery, dan broken. Pada penelitian ini juga menerapkan metode SMOTE untuk mengatasi jumlah data yang tidak seimbang karena jumlah data broken sangat berbeda dengan jumlah data recovering dan normal serta dilakukan beberapa skenario pengujian. Hasil yang didapatkan menunjukkan bahwa algoritma C4.5 mampu melakukan klasifikasi dan mendapatkan akurasi rata-rata sebesar 98.76%.

Kata kunci: C4.5, Klasifikasi, Pompa Air, SMOTE

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## ABSTRACT

Water pumps are used as one way to mechanically obtain water through pipelines. However, pump damage such as sudden pump failure and other damage can cause problems to people's lives. The purpose of this research is to classify the status of the water pump engine using the C4.5 method in order to determine the class of the water pump engine status consisting of normal, recovery, and broken. This research also applies the SMOTE method to overcome the unbalanced amount of data because the amount of broken data is very different from the amount of recovering and normal data and several of test scenarios are carried out. The results obtained show the C4.5 method is able to classify and get the average accuracy of 98.76%.

Keywords: C4.5, Classification, SMOTE, Water Pump



## KATA PENGANTAR

*Assalammu'alaikum wa rohmatullohi wa barokatuh.*

*Alhamdulillah robbil'alamin*, tak henti-hentinya kami ucapkan kehadiran Allah *Subhanahu wa ta'ala*, yang dengan rahmat dan hidayah-Nya kami mampu menyelesaikan Tugas Akhir ini dengan baik. Tidak lupa bershalawat kepada Nabi dan Rasul-Nya, Nabi Muhammad *Sholallohu 'alaihi wa salam*, yang telah membimbing kita sebagai umatnya menuju jalan kebaikan.

Tugas Akhir ini disusun sebagai salah satu syarat untuk mendapatkan gelar sarjana pada jurusan Teknik Informatika Universitas Islam Negeri Sultan Syarif Kasim Riau. Banyak sekali pihak yang telah membantu kami dalam penyusunan laporan ini, baik berupa bantuan materi ataupun berupa motivasi dan dukungan kepada kami. Semua itu tentu terlalu banyak bagi kami untuk membalasnya, namun pada kesempatan ini kami hanya dapat mengucapkan terima kasih kepada:

1. Bapak Prof. Dr. Hairunas, M.Ag, selaku Rektor Universitas Islam Negeri Sultan Syarif Kasim Riau.
2. Bapak Dr. Hartono, M.Pd, selaku Dekan Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
3. Bapak Iwan Iskandar, M.T, selaku Ketua Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
4. Bapak Syarifuddin, M.Ag, selaku Dosen Penasihat Akademik yang telah memberikan arahan serta bimbingan selama ini.
5. Ibu Dr. Lestari Handayani, S.T., M.Kom, selaku Dosen Pembimbing yang telah memberikan waktu untuk membimbing dalam rangka penyelesaian tugas akhir ini.
6. Ibu Prof. Dr. Okfalisa, S.T., M.Sc, selaku Dosen Penguji 1 Laporan Tugas Akhir Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

7. Ibu Fadhilah Syafria, S.T., M.Kom, selaku Dosen Penguji 2 Laporan Tugas Akhir dan Koordinator Tugas Akhir Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
8. Bapak dan Ibu Dosen Teknik Informatika yang telah memberikan ilmu dan motivasinya.
9. Orang tua saya yaitu Bapak Mukhlis dan Ibu Mardiaty dan keluarga tercinta yang telah memberikan dukungan moril, doa dan kasih sayang.
10. Seluruh pihak yang belum kami cantumkan, terima kasih atas dukungannya, baik material maupun spiritual.

Kami menyadari bahwa dalam penulisan laporan ini masih banyak kesalahan dan kekurangan, oleh karena itu kritik dan saran yang sifatnya membangun sangat kami harapkan untuk kesempurnaan laporan ini. Akhirnya kami berharap semoga laporan ini dapat memberikan sesuatu yang bermanfaat bagi siapa saja yang membacanya.

*Wassalamu'alaikum wa rohmatullohi wa barokatuh.*

Pekanbaru, Januari 2025

Penulis

UIN SUSKA RIAU



**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## DAFTAR ISI

|                                                  |      |
|--------------------------------------------------|------|
| LEMBAR PERSETUJUAN .....                         | ii   |
| LEMBAR PENGESAHAN .....                          | iii  |
| LEMBAR HAK ATAS KEKAYAAN INTELEKTUAL .....       | iv   |
| LEMBAR PERNYATAAN .....                          | v    |
| LEMBAR PERSEMBAHAN .....                         | vi   |
| ABSTRAK .....                                    | vii  |
| ABSTRACT .....                                   | viii |
| KATA PENGANTAR .....                             | ix   |
| DAFTAR ISI .....                                 | xi   |
| DAFTAR GAMBAR .....                              | xiv  |
| DAFTAR TABEL .....                               | xv   |
| DAFTAR RUMUS .....                               | xvi  |
| BAB 1 PENDAHULUAN .....                          | 1    |
| 1.1 Latar Belakang .....                         | 1    |
| 1.2 Rumusan Masalah .....                        | 2    |
| 1.3 Batasan Masalah .....                        | 3    |
| 1.4 Tujuan Penelitian .....                      | 3    |
| 1.5 Manfaat Penelitian .....                     | 3    |
| BAB 2 KAJIAN PUSTAKA .....                       | 4    |
| 2.1 <i>Knowledge Discovery in Database</i> ..... | 4    |
| 2.2 <i>Data Mining</i> .....                     | 5    |
| 2.3 Klasifikasi .....                            | 5    |
| 2.4 Algoritma C4.5 .....                         | 6    |



**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau

State Islamic University of Sultan Syarif Kasim Riau

|                                          |                                                                |    |
|------------------------------------------|----------------------------------------------------------------|----|
| 2.5                                      | <i>Synthetic Minority Oversampling Technique (SMOTE)</i> ..... | 8  |
| 2.6                                      | <i>Pre-processing</i> .....                                    | 10 |
| 2.6.1                                    | <i>Missing Value</i> .....                                     | 10 |
| 2.6.2                                    | <i>Normalisasi</i> .....                                       | 11 |
| 2.6.3                                    | <i>Imbalance</i> .....                                         | 12 |
| 2.7                                      | <i>Confusion Matrix</i> .....                                  | 13 |
| 2.7.1                                    | <i>Accuracy</i> .....                                          | 13 |
| 2.7.2                                    | <i>Precision</i> .....                                         | 14 |
| 2.7.3                                    | <i>Recall</i> .....                                            | 14 |
| 2.8                                      | <i>Mesin Pompa Air</i> .....                                   | 14 |
| 2.9                                      | <i>Penelitian Terkait</i> .....                                | 15 |
| <b>BAB 3 METODOLOGI PENELITIAN</b> ..... |                                                                | 19 |
| 3.1                                      | <i>Pengumpulan Data</i> .....                                  | 19 |
| 3.1.1                                    | <i>Studi Literatur</i> .....                                   | 19 |
| 3.1.2                                    | <i>Dataset</i> .....                                           | 20 |
| 3.2                                      | <i>Preprocessing</i> .....                                     | 20 |
| 3.2.1                                    | <i>Penanganan Missing Value</i> .....                          | 20 |
| 3.2.2                                    | <i>Normalisasi Data</i> .....                                  | 21 |
| 3.2.3                                    | <i>Penanganan Imbalance Data</i> .....                         | 21 |
| 3.3                                      | <i>Pembagian Data</i> .....                                    | 21 |
| 3.3.1                                    | <i>Data Latih</i> .....                                        | 21 |
| 3.3.2                                    | <i>Data Uji</i> .....                                          | 22 |
| 3.4                                      | <i>Pelatihan</i> .....                                         | 22 |
| 3.5                                      | <i>Pengujian</i> .....                                         | 22 |
| 3.6                                      | <i>Evaluasi</i> .....                                          | 22 |



### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

© Hak cipta milik UIN Suska Riau  
Statistik Islamic University of Sultan Syarif Kasim Riau

|                                       |    |
|---------------------------------------|----|
| BAB 4 PEMBAHASAN .....                | 23 |
| 4.1 Data Selection .....              | 23 |
| 4.2 Preprocessing .....               | 24 |
| 4.2.1 Penanganan Missing Values ..... | 24 |
| 4.2.2 Normalisasi Data .....          | 27 |
| 4.2.3 Penanganan Imbalance Data ..... | 28 |
| 4.3 Pembagian Data .....              | 30 |
| 4.4 Pelatihan .....                   | 31 |
| 4.5 Pengujian .....                   | 34 |
| 4.6 Analisis Hasil .....              | 41 |
| BAB 5 PENUTUP .....                   | 42 |
| 5.1 Kesimpulan .....                  | 42 |
| 5.2 Saran .....                       | 42 |
| DAFTAR PUSTAKA .....                  | 43 |
| LAMPIRAN .....                        | 46 |
| DAFTAR RIWAYAT HIDUP .....            | 48 |

UIN SUSKA RIAU



## Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## DAFTAR GAMBAR

|                                                                         |    |
|-------------------------------------------------------------------------|----|
| Gambar 1 Tahapan KDD .....                                              | 4  |
| Gambar 2 Metodologi Penelitian .....                                    | 19 |
| Gambar 3 Dataset <i>Pump Sensor</i> .....                               | 23 |
| Gambar 4 Jumlah <i>Missing Value</i> Sebelum Dilakukan Penanganan ..... | 24 |
| Gambar 5 Grafik Batang <i>Imbalance Data</i> .....                      | 28 |
| Gambar 6 Proses Pembagian Data .....                                    | 30 |
| Gambar 7 Tahapan Metode C4.5 .....                                      | 31 |
| Gambar 8 Proses pelatihan metode C4.5 .....                             | 32 |
| Gambar 9 Menampilkan Pohon Keputusan .....                              | 32 |
| Gambar 10 Tampilan Pohon Keputusan .....                                | 33 |
| Gambar 11 Aturan Keputusan.....                                         | 34 |
| Gambar 18 Hasil Pengujian Data Balance .....                            | 34 |
| Gambar 19 Confusion Matrix Data Balance .....                           | 35 |
| Gambar 20 Hasil Pengujian Balance 10% .....                             | 35 |
| Gambar 21 Confusion Matrix Balance 10% .....                            | 36 |
| Gambar 22 Hasil Pengujian Balance 25%.....                              | 36 |
| Gambar 23 Confusion Matrix Balance 25% .....                            | 37 |
| Gambar 24 Hasil Pengujian Balance 50% .....                             | 37 |
| Gambar 25 Confusion Matrix Balance 50% .....                            | 38 |
| Gambar 26 Hasil Pengujian Balance 75% .....                             | 38 |
| Gambar 27 Confusion Matrix Balance 75% .....                            | 39 |
| Gambar 28 Hasil Pengujian Balance 100% .....                            | 39 |
| Gambar 29 Confusion Matrix Balance 100% .....                           | 40 |
| Gambar 30 Grafik Perbandingan Precision, Recall, F1-Score .....         | 40 |
| Gambar 31 Grafik Perbandingan Akurasi.....                              | 41 |



**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## DAFTAR TABEL

|                                                                   |    |
|-------------------------------------------------------------------|----|
| Tabel 1 <i>Confusion Matrix</i> .....                             | 13 |
| Tabel 2 Penelitian Terkait .....                                  | 15 |
| Tabel 3 Data Sebelum Penanganan <i>Missing Value</i> .....        | 25 |
| Tabel 4 Hasil Penanganan <i>Missing Value</i> .....               | 26 |
| Tabel 5 Hasil Normalisasi Data .....                              | 27 |
| Tabel 6 Hasil Penanganan data <i>imbalance</i> dengan SMOTE ..... | 29 |



**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## DAFTAR RUMUS

|                         |    |
|-------------------------|----|
| (1) Rumus Satu.....     | 7  |
| (2) Rumus Dua .....     | 7  |
| (3) Rumus Tiga .....    | 7  |
| (4) Rumus Empat .....   | 7  |
| (5) Rumus Lima .....    | 9  |
| (6) Rumus Enam .....    | 14 |
| (7) Rumus Tujuh .....   | 14 |
| (8) Rumus Delapan ..... | 14 |



# BAB 1

## PENDAHULUAN

### 1.1 Latar Belakang

Pompa air termasuk salah satu komponen krusial dalam sistem penyediaan air. pompa air merupakan sebuah alat yang berfungsi untuk mengalirkan zat berbentuk cairan menggunakan pipa dari suatu tempat ke tempat lainnya dengan menyuplai energi cairan secara terus-menerus (Lubis et al., 2019). Kegagalan atau kerusakan pada mesin pompa air dapat menyebabkan berbagai masalah, termasuk penurunan pasokan air, kegagalan operasional, dan biaya perbaikan yang tinggi (Harahap et al., 2023). Oleh karena itu, penting untuk melakukan pemantauan secara terus-menerus terhadap status dan kinerja mesin pompa air guna mendeteksi dini potensi masalah.

Beberapa penelitian tentang mesin pompa air telah dilakukan diantaranya yaitu Vinaya dkk yang membahas tentang analisis kerusakan pompa. Pada penelitiannya, dilakukan klasifikasi terhadap kerusakan pompa air untuk mengidentifikasi kondisi pompa yang tidak normal, seperti ketidakseimbangan, ketidaksejajaran, dan kerusakan bantalan. Adapun hasil akurasi klasifikasi yang diperoleh pada penelitian tersebut yaitu 90% dalam mendiagnosis kondisi pompa air menggunakan *Fast Fourier Transform* (FFT) dan *Fuzzy Logic Inference* (FIS) (Vinaya et al., 2019).

Penelitian sebelumnya yang dilakukan oleh Lestari Handayani memprediksi kegagalan sistem pada mesin pompa air dan memperkirakan *Remaining Useful Life* (RUL) dari pompa air berdasarkan data sensor pompa air menggunakan metode *Hidden Semi-Markov Model* (HSSM) dan *Hidden Markov Model* (HMM) (Handayani, 2022). Pada penelitian tersebut permasalahan yang terjadi yaitu dalam satu tahun terdapat 7 kegagalan sistem pada mesin pompa air di daerah kecil yang jauh dari kota besar. Oleh sebab itu, beberapa keluarga menghadapi masalah hidup

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

yang signifikan sebagai akibat dari kegagalan tersebut. Berdasarkan hal tersebut, Keir Halewood bekerja di KVH Consulting di Great Torrington, Inggris memberikan informasi sistem ini selama diskusi data dan menyimpulkan bahwa penyebab kegagalan sistem yang paling mungkin adalah pompa air di stasiun pemompaan.

Berdasarkan penelitian tersebut belum dilakukan proses klasifikasi pada status mesin pompa air, sehingga pada penelitian ini akan dilakukan klasifikasi agar dapat mengetahui kelas dari status mesin pompa air yang terdiri dari normal, *recovering*, dan *broken*. Ada berbagai teknik yang bisa diterapkan untuk melakukan klasifikasi. Salah satunya adalah algoritma C4.5 yang terkenal karena kemudahan interpretasi dan visualisasi pohon yang dihasilkan. Algoritma C4.5 banyak digunakan karena memiliki tingkat akurasi yang dapat diterima dan efisien dalam menangani atribut bertipe diskret dan numerik serta dapat menghasilkan pohon keputusan yang mudah diterapkan. Syamsul Bahri dalam penelitiannya tentang klasifikasi penyakit pada anak telah menerapkan algoritma C4.5 dengan perbandingan kinerja algoritma C4.5 dan Naïve Bayes. Hasil perbandingan menunjukkan bahwa algoritma C4.5 yang terbaik dengan akurasi 90,00% diikuti oleh algoritma Naïve Bayes dengan akurasi sebesar 89,98% (Bahri et al., 2018).

Dataset yang digunakan pada penelitian ini terdapat ketidakseimbangan data (*Imbalance data*). *Imbalance data* terjadi karena perbedaan dalam jumlah data antara kelas mayoritas (kelas dengan jumlah anggota yang sangat banyak) dan kelas minoritas (kelas dengan jumlah anggota yang sangat sedikit) (Thabtah et al., 2020). Pada dataset ini, jumlah data normal dan *recovering* sangat banyak, sedangkan data *broken* hanya berjumlah 7 data. Oleh karena itu, perlu dilakukan penanganan terhadap ketidakseimbangan data agar mendapatkan performa yang lebih baik.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka penulis mengambil rumusan masalah sebagai berikut:

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1. Bagaimana hasil performa algoritma C4.5 dalam klasifikasi status mesin pompa air?
2. Bagaimana perbandingan performa algoritma C4.5 sebelum dan sesudah data *imbalance* diatasi?

### 1.3 Batasan Masalah

Pada penelitian ini diperlukan batasan masalah agar tetap fokus, terstruktur, dan terarah. Berikut batasan-batasan masalah pada penelitian ini:

1. Studi kasus dalam penelitian ini bersumber dari situs web Kaggle.
2. Data yang digunakan satu tahun dengan jumlah data sensor yaitu 52 dengan 3 status mesin yaitu normal, *recovering*, dan *broken*.
3. Penanganan ketidakseimbangan kelas dengan metode SMOTE

### 1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini sebagai berikut:

1. Mengimplementasikan algoritma C4.5 untuk klasifikasi status mesin pompa air dan mengetahui hasil performanya.
2. Untuk mengetahui perbandingan performa algoritma C4.5 sebelum dan sesudah data *imbalance* diatasi.

### 1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini sebagai berikut:

1. Penelitian tugas akhir ini diharapkan dapat memberi informasi berdasarkan hasil klasifikasi sehingga bermanfaat untuk pemeliharaan prediktif mesin tersebut.
2. Menambah kajian tentang kasus dan metode yang memberi kontribusi dalam bidang data mining.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

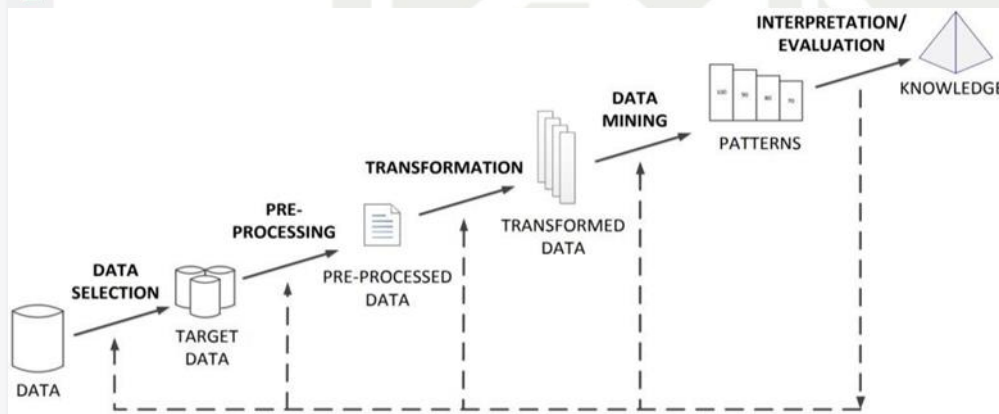
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## BAB 2

### KAJIAN PUSTAKA

#### 2.1 *Knowledge Discovery in Database*

*Knowledge Discovery in Database (KDD)* adalah penemuan pengetahuan yang tersimpan dalam database berukuran besar, data *warehouse*, web, atau tempat penyimpanan informasi lainnya. Sebagian besar orang mengartikan KDD sama dengan data mining. Sebagai suatu rangkaian proses, KDD dapat dibagi menjadi beberapa tahap yang diilustrasikan di gambar ini (Ha et al., 2011):



Gambar 1 Tahapan KDD

Tahapan KDD ada lima, yaitu:

1. *Data Selection*, merupakan tahap pemilihan data yang relevan untuk proses analisa yang akan dilakukan nantinya.
2. *Pre-processing*, merupakan tahap untuk menghilangkan noise dan data yang tidak konsisten.
3. *Transformation*, merupakan tahap transformasi data menjadi bentuk yang layak untuk proses mining.
4. *Data Mining*, merupakan tahap penting untuk menerapkan metode agar pola pada data bisa didapatkan.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

5. *Interpretation/Evaluation*, merupakan tahap untuk memvisualisasikan pengetahuan yang telah didapat/tahap untuk mengidentifikasi pola yang benar-benar penting (merepresentasikan pengetahuan) berdasarkan pengujian tertentu.

## 2.2 Data Mining

*Data mining* adalah proses menganalisis dan mengekstrak dari sejumlah data besar yang disimpan dalam *database*, gudang data, atau *repository* lainnya untuk mendapatkan suatu fakta, sesuatu yang baru, dan memiliki manfaat sehingga dapat ditemukan pola-pola tertentu dalam data tersebut. *Data mining* berfungsi untuk memperoleh informasi yang berguna bagi pengguna untuk menambah pengetahuan. *Data mining* memiliki empat fungsi dasar sebagai berikut (Normah et al., 2022):

1. Fungsi Prediksi

Proses menemukan pola untuk memprediksi variabel lain dengan menggunakan beberapa variabel.

2. Fungsi Deskripsi

Proses menemukan karakteristik penting pada setiap data di *database*.

3. Fungsi Klasifikasi

Klasifikasi merupakan proses menemukan model yang menggambarkan konsep dari suatu data. Proses yang digunakan untuk memprediksi data.

4. Fungsi Asosiasi

Proses menemukan relasi dalam nilai atribut dari sekumpulan nilai pada tiap atribut di dalam *database*.

## 2.3 Klasifikasi

Klasifikasi adalah proses menemukan model (fungsi) yang menggambarkan dan mencirikan kelas atau konsep data, dengan tujuan dapat menggunakan model untuk memprediksi kelas objek yang klasifikasinya tidak diketahui. Model turunan berdasarkan analisis kumpulan data training. Contoh data latih yaitu objek data



#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

yang label kelasnya diketahui. Model ini kemudian digunakan untuk memprediksi label kelas objek yang kelasnya tidak diketahui.

Klasifikasi terdiri dari dua proses. Proses pertama (learning) adalah membangun model klasifikasi berdasarkan data latih. Proses kedua (classification) adalah menentukan keakuratan model dengan menggunakan data uji, apakah dapat diterima atau tidak. Jika keakuratan model dapat diterima, model tersebut digunakan untuk mengklasifikasi data baru.

## 2.4 Algoritma C4.5

Algoritma C4.5 adalah program yang menyumbangkan kumpulan data berlabel dan menghasilkan pohon keputusan sebagai output. Pohon keputusan tindak lanjut ini kemudian diperiksa terhadap data uji yang tidak terlihat untuk menghitung generalisasinya. C4.5 adalah program yang digunakan untuk membuat basis taksonomi menggunakan pohon keputusan dari kumpulan data tertentu (Eko Cahyo & Nilogiri, 2022).

Algoritma C4.5 merupakan pengembangan dari algoritma inti ID3 dan dirancang oleh Quinlan. C4.5 adalah salah satu algoritma pembelajaran yang banyak digunakan. Algoritma C4.5 membangun pohon keputusan dari serangkaian data pelatihan yang mirip dengan algoritma ID3, menggunakan konsep *entropy* informasi. C4.5 juga dikenal sebagai klasifikasi statistik.

Perbaikan pada ID3 yang termasuk dalam C4.5 meliputi kemampuan untuk menangani data persisten, kemampuan untuk memutuskan pohon keputusan (*pruning*), dan kemampuan untuk mengatasi data yang hilang. Keunggulan lain dari algoritma ini adalah dapat menggambarkan sugesti dalam bentuk pohon keputusan sehingga prosedur prediksi dapat diamati. Penggunaannya telah ditemukan di berbagai bidang, termasuk kedokteran (diagnosis penyakit), ilmu komputer (struktur data), psikologi (teori pengambil keputusan), dll.

Langkah-langkah algoritma C4.5 untuk membangun *decision tree* adalah sebagai berikut:

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

1. Pemilihan atribut pembagi (*root*):

- a. Hitung *entropy* untuk semua data terhadap komposisi kelas:

$$E(s) = - \sum_{i=1}^m p(\omega_i|s) \log_2 p(\omega_i|s) \quad (1)$$

dengan ketentuan:

$p(\omega_i|s)$  = proporsi kelas ke-I dalam data yang di proses di node s

$m$  = jumlah nilai berbeda dalam data

- b. Hitung *gain* untuk setiap atribut:

$$G(s, j) = E(s) - \sum_{i=1}^n p(v_i|s) \times E(s_i) \quad (2)$$

dengan ketentuan:

$p(v_i|s)$  = proporsi nilai v muncul pada kelas dalam node s

$E(s_i)$  = entropy komposisi nilai v dari kelas ke-j dalam data ke-I di node s

$n$  = jumlah nilai berbeda dalam *node*

- c. Hitung *split info* untuk atribut diskrit:

$$SplitInfo(s, j) = \sum_{i=1}^k p(v_i|s) \log_2 p(v_i|s) \quad (3)$$

dengan ketentuan:

$p(v_i|s)$  = proporsi nilai v muncul pada kelas dalam node s

$k$  = jumlah pembagian atau pemecah

- d. Hitung *gain ratio* untuk atribut diskrit:

$$GainRatio(s, i) = \frac{Gain(s, i)}{SplitInfo(s, j)} \quad (4)$$

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- e. Bandingkan nilai-nilai *gain* atribut kontinu dengan *gain ratio* atribut diskrit. Atribut dengan nilai tertinggi terpilih menjadi atribut pembagi (*root*).
2. Ulangi seluruh proses untuk setiap cabang hingga seluruh kasus untuk cabang tersebut memiliki kelas yang sama.
3. Partisi (pembagian) rekursif dihentikan apabila memenuhi salah satu kondisi berikut:
  - a. Seluruh data di partisi mempunyai kelas yang sama
  - b. Tidak ada lagi atribut yang tersisa untuk membagi data
  - c. Partisi tidak memiliki data lagi

### 2.5 *Synthetic Minority Oversampling Technique (SMOTE)*

Kondisi ketidakseimbangan kelas (*class imbalance*) akan menyebabkan hasil akhir proses klasifikasi yang buruk dan selalu tidak optimal (Buda et al., 2018). Terdapat dua pendekatan untuk mengatasi ketidakseimbangan kelas, yaitu (Weiss, 2013):

1. Pendekatan level data  
Mencakup berbagai metode resampling untuk mengatasi ketidakseimbangan kelas dengan mengeliminasi beberapa data dari kelas mayoritas (*undersampling*) atau menambahkan beberapa data menggunakan hasil dari proses generated atau duplikat data ke kelas minoritas (*oversampling*).
2. Pendekatan tingkat algoritma  
Metode utamanya adalah menyesuaikan operasi algoritma yang ada untuk membuat klasifikasi lebih konduktif terhadap kelas minoritas.

Metode *resampling* dilakukan dengan teknik *undersampling*, *oversampling*, dan gabungan (*hybrid*). *Oversampling* bekerja pada kelas minoritas dengan menyeimbangkan data dengan teknik acak. Namun, dengan adanya replikasi pengamatan akan menyebabkan terjadinya *overfitting*. *Overfitting* adalah suatu keadaan dimana data yang digunakan untuk pelatihan itu adalah yang terbaik.



#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Sehingga apabila dilakukan tes dengan menggunakan data yang berbeda dapat mengurangi akurasi (hasil yang dibuat tidak sesuai yang diharapkan).

*Overfitting* dapat terjadi ketika beberapa batasan didasarkan pada sifat khusus yang tidak membuat perbedaan pada data. Selain itu duplikasi data minor yang berlebihan juga dapat mengakibatkan terjadinya *overfitting*. Untuk mengatasi hal tersebut teknik populer yang digunakan adalah *synthetic minority oversampling technique* (SMOTE).

*Synthetic minority oversampling technique* (SMOTE) adalah salah satu turunan dari metode *oversampling*. SMOTE pertama kali diperkenalkan oleh Nithes V. Chawla. Pendekatan ini bekerja dengan membuat replikasi dari data minoritas. Replikasi tersebut dikenal dengan data sintesis (*synthetic data*).

SMOTE bekerja untuk menangani *overfitting*, yaitu dengan memanfaatkan ketetanggaan terdekat (KNN) untuk setiap data di kelas minoritas. Metode SMOTE bekerja dengan mencari *k nearest neighbors* (yaitu ketetanggaan terdekat data sebanyak *k*) untuk setiap data di kelas minoritas, setelah itu dibuat data sintesis sebanyak persentase duplikasi data minor (*percentage oversampling*, *N%*) yang diinginkan dan *k-nearest neighbors* yang dipilih secara acak.

Secara umum rumus untuk menentukan data sintesis sebagai berikut:

$$X_{syn} = X_i + (X_{knn} - X_i) \times t \quad (5)$$

Keterangan :

$X_{syn}$  = data sintesis yang akan diciptakan

$X_i$  = data yang akan direplikasi

$X_{knn}$  = data yang memiliki jarak terdekat dari data yang akan direplikasi

$t$  = bilangan acak antara 0 dan 1

## 2.6 Pre-processing

Pra-pemrosesan data adalah proses awal untuk mengumpulkan dan menyiapkan data untuk memungkinkan pemrosesan dan penggunaan data mentah awal pada tahap berikutnya (Muzakir et al., 2022). Data *preprocessing* merupakan tahapan untuk melakukan pembersihan pada data yang sudah diambil. Data preprocessing berfungsi sebagai dasar untuk analisis data yang valid (Prasetya et al., 2023).

Pemilihan dan proses persiapan data harus dilakukan dengan hati-hati tanpa mengabaikan proses pembersihan data yang tepat. Jika dataset tidak baik, maka akan berdampak pada model, rentan terhadap *overfitting* dan bias dalam mentransfer dataset ke model (Mishra et al., 2020). Untuk menghindari tingkat kesalahan yang tinggi, data yang hilang harus diatasi sebelum menggunakan model pembelajaran. Pada *preprocessing* dilakukan penanganan *missing value*, normalisasi dan *imbalance* data.

### 2.6.1 Missing Value

Tidak adanya satu atau lebih entri dalam matriks yang berisi data eksperimen disebut sebagai data yang hilang. Beberapa faktor dapat menyebabkan data hilang, seperti nilai di luar rentang instrumen, instrumen tidak berfungsi pada waktu tertentu, kegagalan komunikasi pengontrol digital dan instrumen, beberapa sensor digunakan untuk tujuan yang sama tetapi pada tingkat pengambilan sampel yang berbeda, atau kesalahan instrumental selama proses pengambilan data.

Untuk mengatasi nilai yang hilang, kita harus menemukannya dan kemudian memilih cara terbaik untuk membersihkannya. Mendeteksi nilai yang hilang adalah tantangan untuk menangani data yang hilang secara otomatis. Dataset dapat menggunakan berbagai opsi, seperti "?" dan "nan", untuk menampilkan nilai yang hilang.

Ada dua cara untuk menangani nilai yang hilang. Pertama, kolom keseluruhan dari nilai yang hilang dihapus. Kekurangan metode ini ialah akan berdampak pada performa data secara keseluruhan jika datanya sedikit. Namun, metode ini dapat

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

digunakan jika datanya banyak (Istiana & Mustafiril, 2023). Kedua dengan melakukan imputasi pada nilai yang tidak ada. Metode ini adalah yang terbaik karena memungkinkan proses analisis lanjutan dilakukan tanpa menghilangkan nilai yang tidak ada. Metode ini bertujuan untuk mengisi nilai yang tidak ada dengan nilai yang diperkirakan. Metode imputasi memberikan keuntungan bahwa data dapat dipertimbangkan atau dianalisis dari pengumpulan data sebelumnya yang memiliki nilai yang hilang. Nilai pengganti dapat dihitung dengan menghitung median dan rata-rata, atau nilai yang merupakan bagian dari dataset, seperti angka 0 (Fadila & Muchlisoh, 2022). Berikut adalah cara mengatasi missing value dengan teknik imputasi:

- Melakukan pengecekan terhadap data yang mengandung missing value.
- Menghitung nilai rata-rata untuk data yang terdapat nilai yang kosong menggunakan rumus sebagai berikut:

$$\bar{x} = \frac{\sum_{i=1}^N xi}{N} = \frac{x_1 + x_2 + x_n}{N}$$

Keterangan:

$\bar{x}$  = Rata – rata (mean)

$\sum_{i=1}^N xi$  = Nilai atribut ke  $i$

$N$  = Jumlah dari seluruh data

- Mengganti data yang mengandung missing value dengan nilai rata-rata yang telah dihitung.

#### 2.6.2 Normalisasi

Dalam pemetaan atau penskalaan, salah satu teknik *pre-processing* adalah normalisasi. Pada tahap ini, skala data diubah menjadi skala yang lebih kecil. Skala data yang baru dapat membantu klasifikasi karena dapat menghapus fitur yang memiliki tingkat noise dan relevansi yang rendah (Henderi, 2021). Ada beberapa metode dalam normalisasi yaitu *Min-max normalization* dan *Z-score normalization*. *Z-score normalization* merupakan bagian dari teknik statistika yang dapat digunakan pada big data. *Z-score* biasanya digunakan dalam data mining

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

untuk menemukan data yang merupakan bagian dari *outlier* data. Metode normalisasi *Z-score* menghasilkan hasil dari nilai rata-rata dan standar deviasi data. Berikut adalah tahapan untuk menghitung *Z-score normalization*:

- Menghitung nilai mean atau rata-rata pada data yang akan dinormalisasi menggunakan persamaan (1)
- Menghitung nilai standar deviasi dengan persamaan sebagai berikut:

$$\sigma = \frac{\sqrt{\sum f_i(x_i - \bar{x})^2}}{\sum f_i}$$

Keterangan:

$\sigma$  = Standar deviasi

$x_i$  = Nilai tengah

$\bar{x}$  = Nilai rata – rata (Mean)

$f_i$  = Frekuensi

- Selanjutnya menghitung hasil *Z-score* dengan rumus berikut:

$$n_i^1 = \frac{n_i - \bar{A}}{\sigma_A}$$

Keterangan:

$n_i^1$  = Hasil *Z-score normalization*

$n_i$  = data yang akan dinormalisasi

$\bar{A}$  = nilai rata-rata

$\sigma_A$  = standar deviasi

Penelitian tahun 2021 oleh Henderi, dkk menemukan bahwa hasil penelitian dengan menggunakan algoritma *Z-score* untuk proses normalisasi data menunjukkan bahwa tingkat akurasi yang diperoleh dari proses klasifikasi telah meningkat.

### 2.6.3 Imbalance

*Imbalance* data (data tidak seimbang) merupakan keadaan dimana terjadi perbedaan jumlah data yang mengakibatkan terjadinya kelas mayoritas (kelas dengan anggota sangat banyak) dan kelas minoritas (kelas dengan anggota sangat sedikit). Ketika



dataset yang diteliti mengalami ketidakseimbangan yang tinggi, performansi klasifikasi algoritma terpengaruh.

Ketidakseimbangan data dapat menyebabkan pembuatan model yang tidak sesuai. Selain itu, ketidakseimbangan data dapat menyebabkan *overfitting* dan penurunan akurasi klasifikasi. Data akan dianggap sebagai *noise* atau *outlier* jika data berasal dari kelompok minoritas. Data tidak seimbang dapat diselesaikan dengan penambahan data sintesis pada kelas minoritas dengan metode *oversampling*, ataupun pengurangan data pada kelas mayoritas dengan metode *undersampling*. Salah satu jenis metode *resampling* adalah *oversampling*.

## 2.7 Confusion Matrix

*Confusion matrix* adalah tabel yang menunjukkan jumlah data testing yang diklasifikasikan dengan benar dan jumlah data testing yang diklasifikasikan dengan salah. Contoh *confusion matrix* untuk klasifikasi biner sebagai berikut:

**Tabel 1 Confusion Matrix**

|                  |         | Kelas Prediksi |         |
|------------------|---------|----------------|---------|
|                  |         | Positif        | Negatif |
| Kelas Sebenarnya | Positif | TP             | FN      |
|                  | Negatif | FP             | TN      |

*True Positive* (TP) adalah jumlah record positif yang diklasifikasikan sebagai positif, *False Positive* (FP) adalah jumlah record negatif yang diklasifikasi sebagai positif, *False Negative* (FN) adalah jumlah record positif yang diklasifikasikan sebagai negatif, *True Negative* (TN) adalah jumlah record negatif yang diklasifikasi sebagai negatif. Hasil klasifikasi yang berbentuk confusion matrix digunakan untuk menghitung *accuracy*, *precision*, dan *recall*.

### 2.7.1 Accuracy

Accuracy merupakan persentase jumlah *record* data yang diklasifikasi secara benar setelah dilakukan pengujian pada hasil klasifikasi. Untuk menghitung *accuracy* digunakan persamaan berikut:

$$Accuracy = \frac{TP + TN}{TP + FN + FP + TN}$$



(6)

### 2.7.2 Precision

*Precision* adalah persentase kasus yang diprediksi positif yang juga benar positif pada data sebenarnya. *Precision* dapat diartikan sebagai kecocokan antara bagian data yang digunakan dengan informasi yang dibutuhkan. Untuk menghitung *precision* digunakan persamaan berikut:

$$Precision = \frac{TP}{TP + FP} \quad (7)$$

### 2.7.3 Recall

*Recall* merupakan persentase kasus positif yang diprediksi positif secara benar. Untuk menghitung *recall* digunakan persamaan berikut:

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

## 2.8 Mesin Pompa Air

Air merupakan komponen yang sangat penting bagi kehidupan manusia. Tingkat pertumbuhan penduduk yang tinggi berpengaruh terhadap tingkat ketersediaan air. Oleh karena itu, pemanfaatan air yang beragam terutama untuk memenuhi kebutuhan sehari-hari menjadikan permintaan pasokan air yang semakin tinggi (Pramono & Andana, 2019). Pompa air merupakan sebuah alat pendukung terhadap ketersediaan air. Pompa merupakan mesin yang berfungsi untuk mengalirkan air dari tempat rendah ke tempat yang tinggi (Kesuma & Handoko, 2021).

Pompa air dapat digunakan sebagai alat dalam menyediakan pasokan air yang cukup sehingga dapat didistribusikan ke masyarakat yang membutuhkan (Arifin et al., 2020). Oleh karena itu, kesehatan mesin pada pompa air perlu dijaga agar tidak terjadinya kerusakan yang akan berdampak pada proses dan kinerja. Pemeliharaan

#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

kesehatan mesin perlu dilakukan dengan tepat agar kondisi mesin dapat beroperasi sesuai dengan prosedur.

## 2.9 Penelitian Terkait

Penelitian terkait berisi deskripsi terkini terkait topik penelitian yang dilakukan, data-data penunjang terkait dengan motivasi penelitian, metodologi penelitian, dan teori/konsep yang digunakan.

**Tabel 2 Penelitian Terkait**

| No | Judul                                                         | Penulis                                            | Tahun | Hasil Penelitian                                                                                                                                                                                                                                                                                                                                                                                     |
|----|---------------------------------------------------------------|----------------------------------------------------|-------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1  | <i>Predictive Maintenance with Hidden Semi-Markov Model</i>   | Lestari Handayani                                  | 2022  | Penelitian ini menunjukkan bahwa pendekatan berbasis data menggunakan metode Markov dapat digunakan secara efektif dalam mendiagnosis dan memprediksi kesehatan mesin pompa air. Metode <i>Hidden Semi-Markov Models</i> (HSMM) dengan topologi 2 (HSMM-FB T2) memiliki kinerja terbaik dalam mendeteksi kesalahan dan memberikan estimasi <i>Remaining Useful Life</i> (RUL) pada sistem pompa air. |
| 2  | Perbandingan Algoritma Naïve Bayes dan C4.5 Untuk Klasifikasi | Syamsul Bahri, Dwi Marisa Midyanti, Rahmi Hidayati | 2018  | Penelitian ini melakukan perbandingan algoritma Naïve Bayes dan C4.5 untuk klasifikasi penyakit anak. Dari hasil perbandingan menunjukkan bahwa algoritma terbaik dengan akurasi sebesar 90.00% yaitu                                                                                                                                                                                                |



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

|    |                                                                                   |                                                                                 |      |                                                                                                                                                                                                                                                                                                                                    |
|----|-----------------------------------------------------------------------------------|---------------------------------------------------------------------------------|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|    | Penyakit Anak                                                                     |                                                                                 |      | Decision Tree C4.5. Sementara algoritma Naïve Bayes memperoleh tingkat akurasi sebesar 89.58%.                                                                                                                                                                                                                                     |
|    | Klasifikasi Jamur dapat Dikonsumsi dan Beracun Menggunakan Model Bayesian Network | Marselia Ghanyyu Wahdini, Nurul Fuady Adhalia H, Armin Lawi                     | 2022 | Berdasarkan hasil penelitian ini menunjukkan atribut yang mempengaruhi jamur dapat dikonsumsi dan beracun yaitu : <i>odor</i> , <i>spore-print color</i> , <i>stalk-root</i> , dan <i>cap-shape</i> , dimana data <i>training</i> menghasilkan akurasi sebesar 98,66% dan data <i>testing</i> menghasilkan akurasi sebesar 97,97%. |
| 4  | Analisis Perbandingan Algoritma C4.5 dan CART untuk Klasifikasi Penyakit Stroke   | Suryani, Desvita Rahmadani, Ali Alamuddin Muzafar, Abdul Hamid, Rahmatul Annisa | 2022 | Pada penelitian ini dilakukan percobaan pertama menghasilkan tingkat akurasi algoritma C4.5 sebesar 96% algoritma CART dengan akurasi 95,76% dan percobaan kedua algoritma C4.5 menghasilkan tingkat akurasi sebesar 95,76% dan algoritma CART menghasilkan tingkat akurasi sebesar 95,11%.                                        |
| 16 | <i>Fault Diagnosis of water Pump Based on Acoustic</i>                            | Anindita Adikaputri Vinaya, Qurrotin Ayunina                                    | 2019 | Data diproses menggunakan <i>Fuzzy Inference System</i> (FIS) dan <i>Fast Fourier Transform</i> (FFT) untuk mengekstrak fitur dalam domain waktu dan                                                                                                                                                                               |



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

|                                                                                                      |                                                                                                              |      |                                                                                                                                                                                                                  |
|------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------|------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <i>Emission Signal Using Fast Fourier Transform Technique and Fuzzy Logic Inference</i>              | Maulida Okta Arifianti, Nicky Yessica, Dhany Arifianto, Aulia Siti Aisjah                                    |      | frekuensi. Hasil eksperimen menunjukkan klasifikasi sebesar 90%, yang menunjukkan potensi penggunaan FIS dalam analisis dalam analisis emisi akustik untuk mendeteksi berbagai kategori kerusakan pada pompa air |
| Klasifikasi Algoritma C4.5 Dalam Penerapan Tingkat Kepuasan Siswa Terhadap Media Pembelajaran Online | Kiki Aidi Saputra, Jaya Tata Hardinata, Muhammad Ridwan Lubis, Sundari Retno Andani, Ilham Syahputra Saragih | 2020 | Penelitian ini menghasilkan 20 <i>rules</i> dan tingkat akurasi sebesar 86,67% dimana faktor yang paling mempengaruhi adalah sajian pembelajaran dengan nilai <i>gain</i> sebesar 0,353960518.                   |
| Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan   | Senna Hendrian                                                                                               | 2018 | Berdasarkan penelitian ini mendapatkan nilai akurasi sebesar 98,80%. Nilai untuk <i>precision</i> sebesar 98,02% dan untuk nilai <i>sensitivity</i> atau <i>recall</i> sebesar 99,00%.                           |



#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

|                                                                                 |              |      |                                                                                                                                                                                                                                                                                                                                                                                 |
|---------------------------------------------------------------------------------|--------------|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Penerapan Data Mining Untuk Memprediksi Transaksi Nasabah Dengan Algoritma C4.5 | Desi Marlina | 2021 | Hasil dari penelitian ini mendapatkan tingkat akurasi untuk data <i>training</i> sebesar 65,00% dengan <i>mean precision</i> sebesar 55,86%, <i>classification error</i> sebesar 35%, dan <i>recall</i> sebesar 55,43%. Sedangkan untuk data <i>testing</i> mendapatkan tingkat akurasi sebesar 87,88% dengan <i>precision</i> sebesar 87,66% dan <i>recall</i> sebesar 87,44%. |
|---------------------------------------------------------------------------------|--------------|------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|

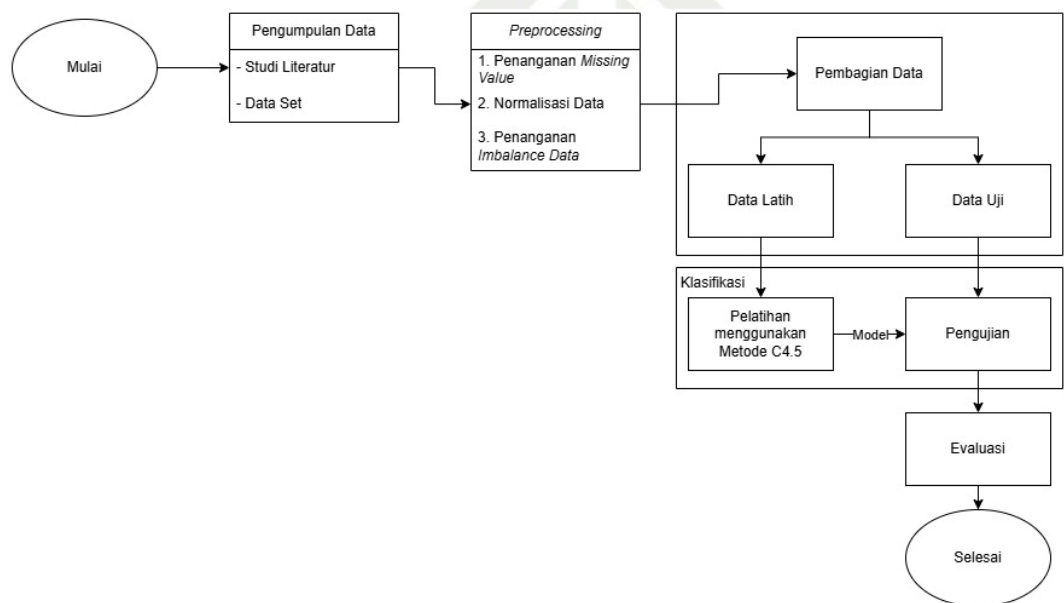
#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## BAB 3

### METODOLOGI PENELITIAN

Metodologi penelitian adalah prosedur atau tahapan yang digunakan untuk melakukan penelitian. Tujuan metodologi penelitian adalah untuk mencapai hasil yang diinginkan. Berikut adalah tahapan-tahapan penelitian :



Gambar 2 Metodologi Penelitian

### 3.1 Pengumpulan Data

Pada tahap ini akan menjelaskan bagaimana proses mengumpulkan data serta referensi yang digunakan dalam pengumpulan data, penjelasannya sebagai berikut.

#### 3.1.1 Studi Literatur

Studi literatur mencakup pencarian literatur yang relevan dengan topik penelitian saat ini. Dalam tahap ini, sumber-sumber tertulis seperti artikel jurnal ilmiah, buku, tesis, laporan penelitian, dan lainnya harus dikumpulkan, ditelaah kemudian dibuat perbandingan yang diharapkan dapat melengkapi dari penelitian sebelumnya. Adapun studi literatur yang dilakukan dalam penelitian ini berupa



teori-teori tentang klasifikasi, mesin pompa air, *preprocessing*, metode yang digunakan yaitu Random forest dan evaluasi model.

### 3.1.2 Dataset

Data yang digunakan pada penelitian ini adalah data sekunder yang bersumber dari web *Kaggle* yaitu dataset *pump sensor* data tahun 2018 yang berjudul “*Pump sensor data for predictive maintenance*”. Data yang digunakan terdiri time stamp data, data sensor, dan machine status dengan jumlah data sebanyak 220.320 dengan jumlah sensor sebanyak 52 buah dan 3 status mesin yaitu normal, recovering dan broken. Dataset yang digunakan pada penelitian ini merujuk pada penelitian sebelumnya tentang prediksi sebelum kerusakan terhadap kondisi mesin pompa air menggunakan dataset *pump sensor*.

## 3.2 Preprocessing

Tahapan yang dilakukan selanjutnya adalah *pre-processing* data, yang dimaksud dengan *pre-processing* adalah metode untuk mempersiapkan dan membersihkan data sebelum diproses oleh model atau algoritma. Berikut adalah beberapa teknik *pre-processing* dalam penelitian ini :

### 3.2.1 Penanganan Missing Value

Dalam penelitian ini bentuk penanganan *missing value* diantaranya menggunakan teknik imputasi dengan mengganti nilai dengan rata-rata kolom atau nilai median untuk kolom yang memiliki *missing value* dalam jumlah sedikit dan menghapus baris/kolom yang mengandung *missing value* dengan jumlah yang banyak. adapun tahap dalam melakukan *missing value* sebagai berikut :

1. Melakukan pengecekan terhadap dataset yang digunakan.
2. Untuk kolom atau baris yang terdapat banyak *missing value*, maka dilakukan penghapusan pada kolom/baris tersebut.
3. Untuk data yang terdapat beberapa *missing value*, lakukan teknik imputasi dmenggunakan persamaan (1) yaitu mengganti nilai yang kosong dengan rata-rata.



#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

### 3.2.2 Normalisasi Data

Pada tahap ini dilakukan normalisasi data menggunakan metode *Z-score normalization*. Sehingga dengan dilakukan normalisasi pada data yang digunakan pada penelitian ini dapat meningkatkan hasil klasifikasi pada pengujian model nantinya. Adapun tahapan dalam melakukan normalisasi menggunakan *Z-score normalization* sebagai berikut :

1. Menghitung nilai mean atau rata-rata pada data yang akan dinormalisasi menggunakan persamaan (1)
2. Menghitung nilai standar deviasi dengan persamaan (7)
3. Selanjutnya menghitung hasil *Z-Score* dengan persamaan (8).

### 3.2.3 Penanganan *Imbalance Data*

Menangani *imbalance data* yaitu langkah yang diambil untuk mengatasi masalah ketidakseimbangan proporsi antara kelas atau kategori yang berbeda dalam sebuah dataset. Pada penelitian ini menggunakan teknik *oversampling* yaitu menambahkan data sintesis ke kelas minoritas sehingga kuantitas datanya sebanding dengan kelas mayoritas. Teknik *oversampling* yang digunakan dalam penelitian ini yaitu *Synthetic Minority Oversampling Technique (SMOTE)*

## 3.3 Pembagian Data

Pada penelitian ini data dibagi menjadi dua bagian yaitu data pelatihan dan data pengujian. Proporsi pembagian data pada penelitian ini menggunakan perbandingan yaitu 80:20.

### 3.3.1 Data Latih

Pada penelitian ini, data pelatihan digunakan dengan komposisi 80%. Pembuatan data pelatihan dilakukan dengan menggunakan komposisi yang telah ditetapkan dari jumlah total data. Sehingga data pelatihan yang diperoleh digunakan untuk pemodelan klasifikasi.



#### Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

### 3.3.2 Data Uji

Pada penelitian ini, data pengujian merupakan sisa dari data pelatihan dengan komposisi 20%. Hasil data pengujian digunakan untuk menentukan pemodelan klasifikasi yang terbaik dengan tingkat akurasi.

### 3.4 Pelatihan

Pada tahap data mining, penulis akan menerapkan algoritma C4.5. Langkah-langkah dalam menerapkan algoritma C4.5 yaitu:

1. Mengumpulkan data dan dibagi dua yakni data latih dan data uji. Sebagai pembelajaran algoritma C4.5.
2. Menghitung nilai *entropy* total dan *entropy* setiap atribut pada data latih dengan algoritma C4.5.
3. Menghitung nilai *gain* pada setiap atribut pada data latih.
4. Membangun akar pohon *gain* tertinggi sebagai akar pohon keputusan.
5. Mengulangi setiap proses langkah 2 dan 3 apabila belum mencapai iterasi maksimal atau akurasi 100%.

### 3.5 Pengujian

Pada langkah ini dilakukan beberapa skenario pengujian untuk melihat pengaruhnya terhadap performa model yang telah dibuat. Pada penelitian ini pengujian dilakukan terhadap kelas *broken* yang telah dilakukan *oversampling* dengan persentase 10%, 25%, 50%, 75%, dan 100% dari kelas *recovering*.

### 3.6 Evaluasi

Evaluasi merupakan tahap mengacu pada proses di mana suatu pengujian dinilai untuk menilai sejauh mana mencapai tujuan yang telah ditetapkan. Pada tahap evaluasi dilakukan perhitungan akurasi dengan menggunakan *confusion matrix* yang akan menghasilkan nilai *Accuracy*, *Precision*, dan *Recall*.

**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## **BAB 5**

### **PENUTUP**

#### **5.1 Kesimpulan**

Berdasarkan penelitian yang telah dilakukan maka dapat ditarik beberapa kesimpulan diantaranya:

1. Klasifikasi status mesin pompa air menggunakan algoritma C4.5 sudah berhasil dilakukan.
2. Pada klasifikasi status mesin pompa air menggunakan algoritma C4.5 dengan kelas yang tidak seimbang memiliki akurasi yang tinggi tapi belum bisa mengklasifikasikan status mesin ke dalam tiga kelas
3. Pada Klasifikasi status mesin pompa air menggunakan algoritma C4.5 Metode C4.5 dengan distribusi kelas yang sudah seimbang menggunakan metode SMOTE mendapatkan akurasi rata-rata sebesar 98,76% yang bisa mengklasifikasi mesin ke dalam tiga status mesin yaitu normal, recovering, dan broken.
4. Atribut yang paling berpengaruh pada klasifikasi status mesin pompa air yaitu Sensor\_10

#### **5.2 Saran**

Dari penelitian yang telah dilakukan maka peneliti menyarankan untuk melakukan penelitian dengan algoritma klasifikasi lainnya dan mencoba menambahkan teknik preprocessing yang lain.

Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.

b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

- ## DAFTAR PUSTAKA
- Arifin, Z., Tamamy, A. J., & Islahu, N. (2020). Perancangan Mesin Pompa Air Tenaga Surya untuk Mengurangi Konsumsi Listrik Skala Rumahan. *Jurnal Nasional Teknik Elektro*, 9(2), 79. <https://doi.org/10.25077/jnte.v9n2.758.2020>
- Bahri, S., Marisa Midyanti, D., Hidayati, R., Sistem Komputer, J., & Mipa, F. (2018). Perbandingan Algoritma Naive Bayes dan C4.5 Untuk Klasifikasi Penyakit Anak. *Seminar Nasional Aplikasi Teknologi Informasi (SNATi)*, 11–2018.
- Buda, M., Maki, A., & Mazurowski, M. A. (2018). A systematic study of the class imbalance problem in convolutional neural networks. *Neural Networks*, 106, 249–259. <https://doi.org/10.1016/j.neunet.2018.07.011>
- Eko Cahyo, A., & Nilogiri, A. (2022). Klasifikasi Gangguan Autisme Pada Anak Menggunakan Algoritma C4.5 Dengan Teknik Random Forest. *National Multidisciplinary Sciences*, 1(6), 846–851. <https://doi.org/10.32528/nms.v1i6.241>
- Fadila, L. M. A., & Muchlisoh, S. (2022). Perbandingan Kinerja Metode Hybrid KNNI-GA dan MissForest Dalam Menangani Missing Values. *Seminar Nasional Official Statistics*, 2022(1), 553–562. <https://doi.org/10.34123/semnasoffstat.v2022i1.1315>
- Ha, J., Kambe, M., & Pe, J. (2011). Data Mining: Concepts and Techniques. In *Data Mining: Concepts and Techniques*. <https://doi.org/10.1016/C2009-0-61819-5>
- Handayani, L. (2022). *INSA Centre Val de Loire DOCTORAL SCHOOL Mathematics , Computer Science , Theoretical Physics and Systems Engineering ( MIPTIS ) Lestari HANDAYANI Predictive Maintenance with Hidden Semi-Markov Model*.



1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

- Harahap, D. M., Azmi, Z., & Elfutriani, E. (2023). Implementasi Metode Certainty Factor Dalam Mendeteksi Kerusakan Pada Pompa Pengolahan Air PDAM Tirtanadi. *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, 2(4), 618. <https://doi.org/10.53513/jursi.v2i4.5426>
- Henderi, H. (2021). Comparison of Min-Max normalization and Z-Score Normalization in the K-nearest neighbor (kNN) Algorithm to Test the Accuracy of Types of Breast Cancer. *IJIIS: International Journal of Informatics and Information Systems*, 4(1), 13–20. <https://doi.org/10.47738/ijis.v4i1.73>
- Istiana, N., & Mustafiril, A. (2023). Perbandingan Metode Klasifikasi pada Data dengan Imbalance Class dan Missing Value. *Jurnal Informatika*, 10(2), 101–108. <https://doi.org/10.31294/inf.v10i2.15540>
- Kesuma, H., & Handoko, K. (2021). Sistem Pakar Untuk Mendeteksi Kerusakan Pompa Utama Elektrik Pemadam Gedung Bertingkat Berbasis Web. *Comasie*, 05(03), 97–106.
- Lubis, Z., Lungguk, A., Saputra, N., Winata, S., Annisa, A., Muhazzir, B., Satria, M., & Sri, W. (2019). Kontrol Mesin Air Otomatis Berbasis Arduino Dengan Smartphone. *Cetak) Buletin Utama Teknik*, 14(3), 1410–4520.
- Mishra, P., Biancolillo, A., Roger, J. M., Marini, F., & Rutledge, D. N. (2020). New data preprocessing trends based on ensemble of multiple preprocessing techniques. *TrAC - Trends in Analytical Chemistry*, 132, 116045. <https://doi.org/10.1016/j.trac.2020.116045>
- Muzakir, A., Syaputra, H., & Panjaitan, F. (2022). A Comparative Analysis of Classification Algorithms for Cyberbullying Crime Detection: An Experimental Study of Twitter Social Media in Indonesia. *Scientific Journal of Informatics*, 9(2), 133–138. <https://doi.org/10.15294/sji.v9i2.35149>
- Normah, Rifai, B., Vambudi, S., & Maulana, R. (2022). Analisa Sentimen Perkembangan Vtuber Dengan Metode Support Vector Machine Berbasis



Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:

- a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
- b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.

2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

SMOTE. *Jurnal Teknik Komputer AMIK BSI*, 8(2), 174–180.  
<https://doi.org/10.31294/jtk.v4i2>

Pramono, G. E., & Andana, R. (2019). Analisis Kebutuhan Air Dan Pembangunan Sistem Pengairan Menggunakan Pompa Air Di Kampung Cengal Desa Karacak. *Abdi Dosen : Jurnal Pengabdian Pada Masyarakat*, 3(3), 216.  
<https://doi.org/10.32832/abdidos.v3i3.331>

Prasetya, M. R. A., Priyatno, A. M., & Nurhaeni. (2023). Penanganan Imputasi Missing Values pada Data Time Series dengan Menggunakan Metode Data Mining. *Jurnal Informasi Dan Teknologi*, 5(2), 52–62.  
<https://doi.org/10.37034/jidt.v5i2.324>

Thabtah, F., Hammoud, S., Kamalov, F., & Gonsalves, A. (2020). Data imbalance in classification: Experimental evaluation. *Information Sciences*, 513, 429–441. <https://doi.org/10.1016/j.ins.2019.11.004>

Vinaya, A. A., Arifianti, Q. A. M. O., Yessica, N., Arifianto, D., & Aisjah, A. S. (2019). Fault Diagnosis of Water Pump Based on Acoustic Emission Signal Using Fast Fourier Transform Technique and Fuzzy Logic Inference. *2019 International Conference on Engineering, Science, and Industrial Applications, ICESI 2019*, 1–6. <https://doi.org/10.1109/ICESI.2019.8863023>

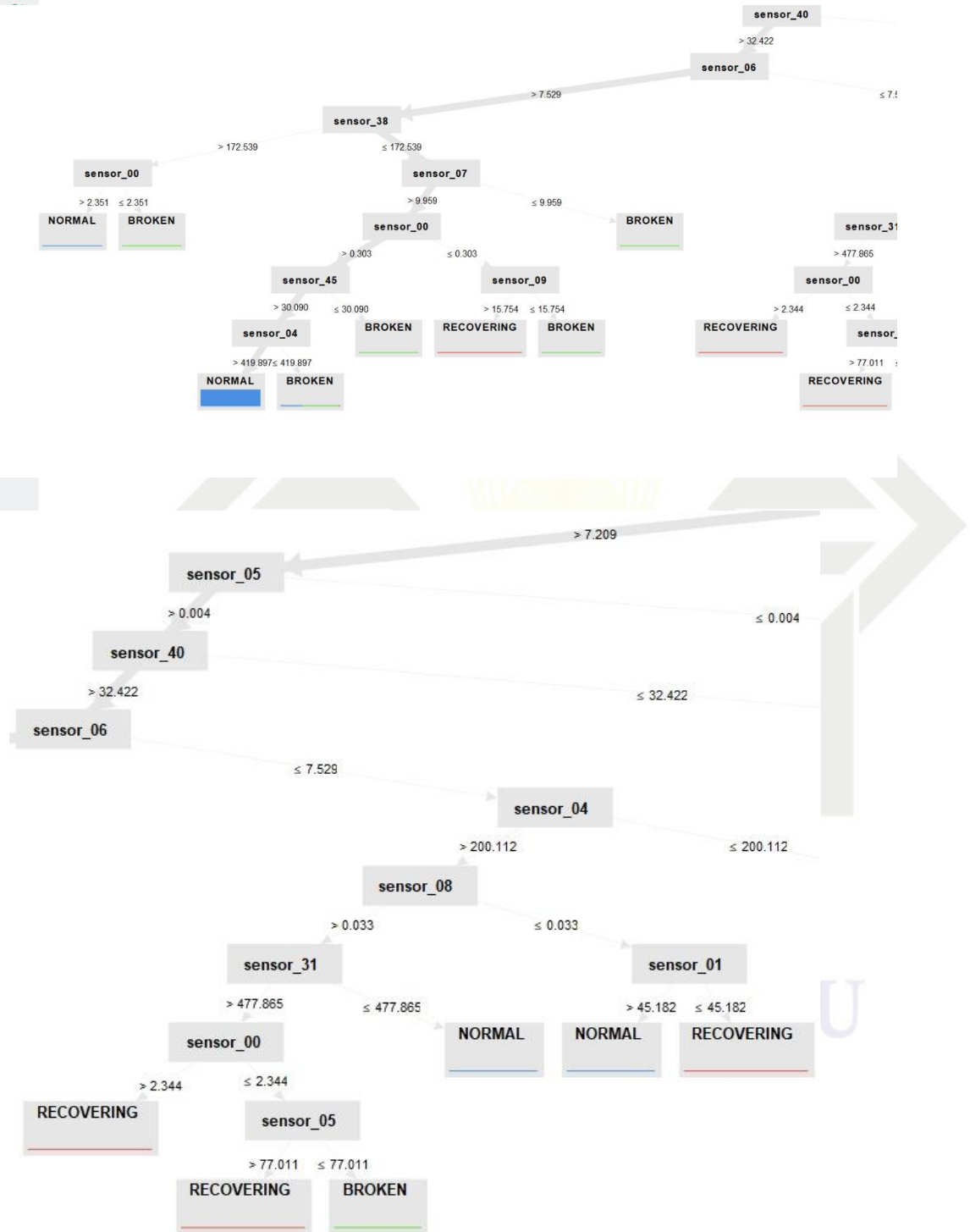
Weiss, G. M. (2013). Foundations of imbalanced learning. *Imbalanced Learning: Foundations, Algorithms, and Applications*, 13–41.  
<https://doi.org/10.1002/9781118646106.ch2>

UIN SUSKA RIAU

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

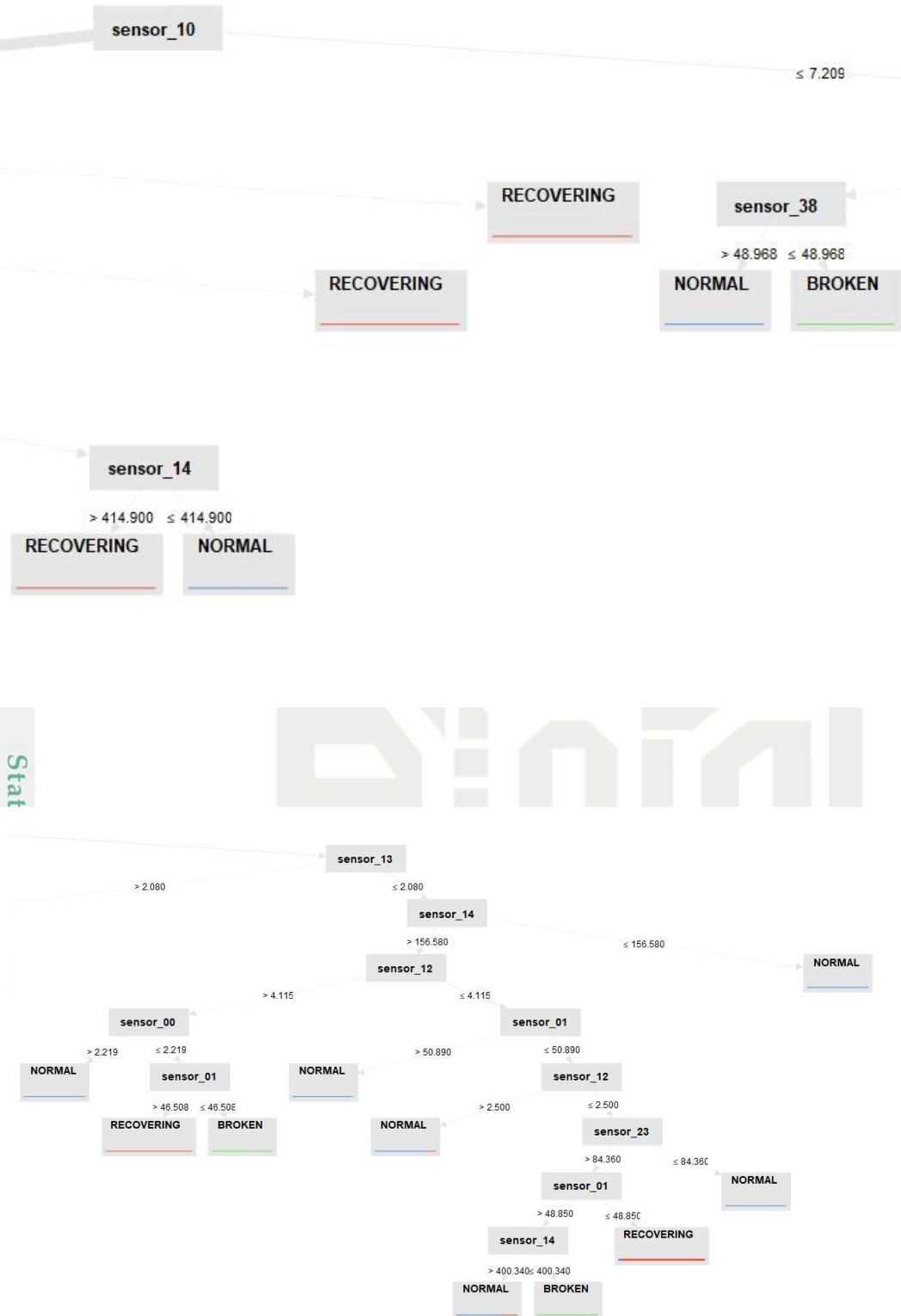
## LAMPIRAN

### TAMPILAN POHON KEPUTUSAN



## Hak Cipta Dilindungi Undang-Undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.





**Hak Cipta Dilindungi Undang-Undang**

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
  - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
  - b. Pengutipan tidak merugikan kepentingan yang wajar UIN Suska Riau.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

## DAFTAR RIWAYAT HIDUP

Nama : Mukhtia Khairi Anggi

Jenis Kelamin : Laki-Laki

Agama : Islam

Kewarganegaraan : Indonesia

Alamat : Jalan Bangun Karya Panam

Email : [11950115142@students.uin-suska.ac.id](mailto:11950115142@students.uin-suska.ac.id)

### Informasi Pendidikan

Tahun 2005-2006 : TK Asiyah Pangkalan

Tahun 2005-2012 : SDN 01 Pangkalan

Tahun 2012-2015 : MTsN Payakumbuh

Tahun 2015-2018 : SMA N 02 Payakumbuh

Tahun 2018-2023 : S1 Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Islam Negeri Sultan Syarif Kasim Riau.

UIN SUSKA RIAU