

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 METODE *RANDOM FOREST*

TUGAS AKHIR

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

Oleh

YOLANDA

NIM. 12050120358



UIN SUSKA RIAU

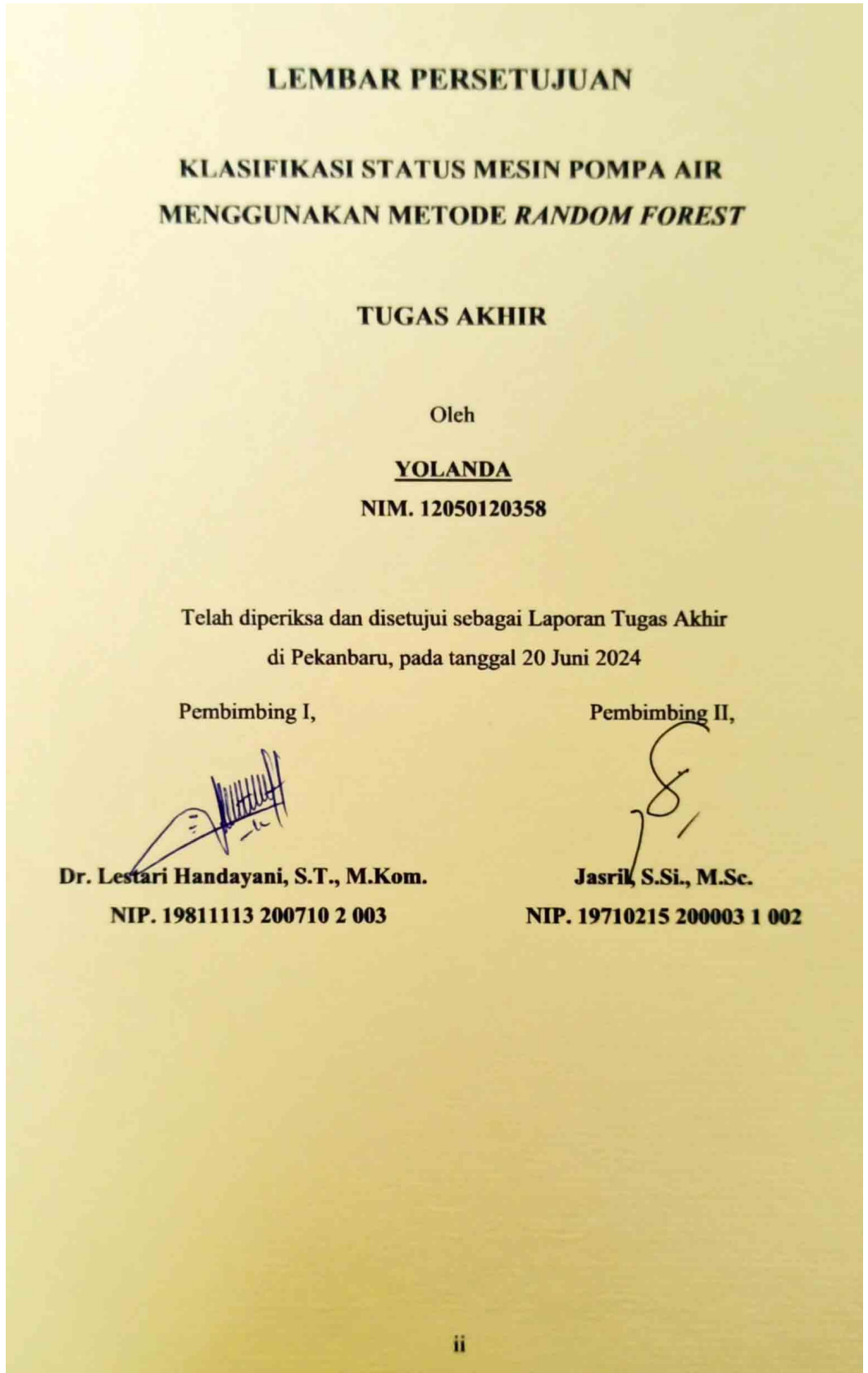
**FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SULTAN SYARIF KASIM RIAU**

PEKANBARU

2024

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.

LEMBAR PENGESAHAN

**KLASIFIKASI STATUS MESIN POMPA AIR MENGGUNAKAN
METODE *RANDOM FOREST***

Oleh

YOLANDA
NIM. 12050120358

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, 20 Juni 2024
Mengesahkan,
Ketua Jurusan,



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



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

DEWAN PENGUJI

| | |
|---------------|---------------------------------------|
| Ketua | : Iwan Iskandar, M.T. |
| Pembimbing I | : Dr. Lestari Handayani, S.T., M.Kom. |
| Pembimbing II | : Jasril, S.Si., M.Sc. |
| Penguji I | : Suwanto Sanjaya, S.T., M.Kom. |
| Penguji II | : Fitri Insani, S.T., M.Kom. |



iii

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 : Yolanda
NIM : 12050120358
Tempat /Tanggal Lahir : Sawahlunto Sijunjung, 24 Januari 2002
Fakultas : Sains dan Teknologi
Prodi : Teknik Informatika
Judul Skripsi : Klasifikasi Status Mesin Pompa Air Menggunakan Metode
Random Forest

Menyatakan dengan sebenar-benarnya bahwa :

1. Penulisan Skripsi dengan judul “Klasifikasi Status Mesin Pompa Air Menggunakan Metode *Random Forest*” adalah hasil pemikiran dan penelitian saya sendiri.
2. Semua kutipan pada karya ilmiah saya 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-undang.

Demikianlah surat pernyataan ini saya buat dengan penuh kesadaran dan tanpa paksaan dari pihak manapun juga.

Pekanbaru, 26 Juni 2024
Yang membuat pernyataan,



Yolanda
NIM. 12050120358

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.

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan didalam daftar pustaka.

Pekanbaru, 20 Juni 2024

Yang membuat pernyataan,

YOLANDA

12050120358

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.

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

Tugas Akhir ini dipersembahkan untuk kedua orang tua saya yang telah membesarkan saya dari kecil serta berkat doa mereka saya bisa sampai dititik ini yaitu menulis tugas akhir

Tidak lupa ucap terimakasih kepada saudara – saudara saya yang telah mendukung dalam bentuk semangat apapun dalam mengerjakan tugas akhir ini

Tidak lupa kepada teman – teman yang selalu mendukung dalam keadaan apapun

Terimakasih semuanya.



UIN SUSKA RIAU

ABSTRAK

Terjadi kerusakan pada mesin pompa air dapat menyebabkan masalah yang besar salah satunya dapat mengakibatkan kesalahan mekanis pada mesin pompa. Tujuan dari penelitian ini adalah menghitung akurasi dalam proses klasifikasi menggunakan metode *Random forest* (RF) serta membandingkan akurasi data *imbalance* dan setelah ditangani *imbalance* data. Pada penelitian ini menggunakan data sekunder yang bersumber dari web Kaggle yaitu dataset pump sensor data yang terdiri dari 52 kolom sensor dan 3 status mesin yaitu normal, *recovering*, dan *broken*. Data ini memiliki masalah ketidakseimbangan data (*imbalancing data*) sehingga untuk mengatasi tersebut digunakan metode *Synthetic Minority Oversampling Technique* (SMOTE). Pengujian pada penelitian ini dilakukan pada data yang telah ditangani *outlier* dan tanpa penanganan *outlier* dengan data *balance* dan *imbalance* menggunakan rasio perbandingan dan kombinasi dari parameter RF. Hasil penelitian ini mendapatkan akurasi rata-rata sebesar 99,97% pada klasifikasi menggunakan data tanpa penanganan *outlier* dan 80,29% menggunakan data dengan penanganan *outlier*. Sehingga dapat disimpulkan metode *Random forest* efektif digunakan dalam proses klasifikasi multiclass.

Kata kunci: Mesin pompa air, *Random forest*, *Imbalancing data*, *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.

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

Damage to the water pump engine can cause major problems, one of which can result in mechanical faults in the pump engine. The purpose of this research is to calculate accuracy in the classification process using the Random forest (RF) method and compare the accuracy of imbalance data and after handling imbalance data. This research uses secondary data sourced from the Kaggle web, namely the pump sensor data dataset which consists of 52 sensor columns and 3 engine statuses, namely normal, recovering, and broken. This data has the problem of imbalancing data, so to overcome this, the Synthetic Minority Oversampling Technique (SMOTE) method is used. Testing in this study was carried out on data that had been handled by outliers and without handling outliers with balance and imbalance data using comparison ratios and combinations of RF parameters. The results of this study obtained an average accuracy of 99.97% on classification using data without outlier handling and 80.29% using data with outlier handling.. So it can be concluded that the Random forest method is effectively used in the multiclass classification process.

Keywords: *water pump machine, Random forest, Imbalancing data, SMOTE*

KATA PENGANTAR

Assalamu '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 Dains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau.
4. Ibu Liza Afriyanti, M.Kom selaku Penasihat Akademik yang telah memberikan arahan serta bimbingan selama ini
5. Ibu Lestari Handayani, S.T., M.Kom selaku Pembimbing 1 yang telah memberikan waktu untuk membimbing dalam rangka penyelesaian tugas akhir ini.
6. Bapak Jasril, S.Si., M.Sc. selaku Pembimbing 2 yang memberikan bimbingan untuk menyelesaikan 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.

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. Bapak Suwanto Sanjaya, S.T., M.Kom. Sebagai Penguji 1 Laporan Tugas Akhir Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau
8. Ibu Fitri Insani, S.T., M.Kom. Sebagai Penguji 2 Laporan Tugas Akhir Jurusan Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Sultan Syarif Kasim Riau
9. Ibu Fadhillah Syafria, S.T., M.Kom. selaku Koordinator Tugas Akhir Jurusan Teknik Informatika yang telah membantu proses administrasi tugas akhir.
10. Seluruh dosen Teknik Informatika yang telah memberikan ilmu dan arahan yang bermanfaat.
11. Orang tua yaitu Bapak Marlis dan Ibu Maryam, serta saudara-saudara yaitu Andra, Jepriadi, Weni Andola, dan Yosi Arista yang selalu memberikan do'a, serta motivasi untuk menyelesaikan tugas akhir ini.
12. Kepada sahabat yaitu Dilla Rahmiati dan Maiyorahma Digna yang selalu bersama-sama dalam menyelesaikan tugas akhir ini
13. 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, 20 Juni 2024

Penulis



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..... | xvii |
| DAFTAR RUMUS | xviii |
| BAB 1 PENDAHULUAN | 1 |
| 1.1 Latar Belakang..... | 1 |
| 1.2 Rumusan Masalah | 3 |
| 1.3 Batasan Masalah..... | 3 |
| 1.4 Tujuan Penelitian..... | 4 |
| 1.5 Manfaat Penelitian..... | 4 |
| BAB 2 KAJIAN PUSTAKA..... | 5 |
| 2.1 Mesin Pompa Air..... | 5 |
| 2.2 <i>Pre-processing</i> | 5 |
| 2.2.1 <i>Missing values</i> | 5 |
| 2.2.2 <i>Outliers</i> | 7 |

- 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.

| | | |
|--|---|-----------|
| 2.2.3 | Normalisasi | 8 |
| 2.2.4 | <i>Imbalance</i> | 10 |
| 2.3 | <i>Data mining</i> | 12 |
| 2.4 | Klasifikasi..... | 12 |
| 2.5 | <i>Random forest</i> | 13 |
| 2.5.1 | Tahapan pembuatan pohon keputusan pada algoritma <i>Random forest</i> 14 | |
| 2.5.2 | Algoritma <i>Random forest</i> | 15 |
| 2.5.3 | Parameter <i>Random forest</i> | 17 |
| 2.5.4 | Keuntungan <i>Random forest</i> | 18 |
| 2.6 | Pembagian Data..... | 18 |
| 2.7 | Evaluasi | 19 |
| 2.8 | Penelitian Terkait..... | 20 |
| BAB 3 METODOLOGI PENELITIAN | | 33 |
| 3.1 | Pengumpulan Data..... | 33 |
| 3.1.1 | Studi Literatur | 33 |
| 3.1.2 | Data set..... | 34 |
| 3.2 | <i>Pre-processing</i> Data | 34 |
| 3.2.1 | Penanganan <i>Missing Value</i> | 34 |
| 3.2.2 | Penanganan <i>Outlier</i> | 35 |
| 3.2.3 | Normalisasi Data..... | 35 |
| 3.2.4 | Penanganan <i>imbalance</i> data | 36 |
| 3.3 | Pembagian data..... | 36 |
| 3.3.1 | Data Latih..... | 36 |
| 3.3.2 | Data Uji | 36 |



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.

| | | |
|-----------------------------------|--|-----|
| 3.4 | Pelatihan | 36 |
| 3.5 | Pengujian | 37 |
| 3.5.1 | Skenario Pengujian | 37 |
| 3.6 | Evaluasi | 40 |
| BAB 4 PEMBAHASAN | | 41 |
| 4.1 | Deskripsi Data | 41 |
| 4.2 | <i>Preprocessing Data</i> | 42 |
| 4.2.1 | Penanganan <i>Missing Value</i> | 42 |
| 4.2.2 | Penanganan <i>Outlier</i> | 45 |
| 4.2.3 | Normalisasi Data | 47 |
| 4.2.4 | <i>Imbalancing Data</i> | 50 |
| 4.3 | Pembagian data | 52 |
| 4.4 | Pelatihan | 53 |
| 4.5 | Pengujian | 58 |
| 4.5.1 | Pengujian setelah ditangani <i>outlier</i> | 59 |
| 4.5.2 | Pengujian tanpa penanganan <i>outlier</i> | 70 |
| 4.6 | Analisis Hasil | 80 |
| BAB 5 PENUTUP | | 86 |
| 5.1 | Kesimpulan | 86 |
| 5.2 | Saran | 87 |
| DAFTAR PUSTAKA | | 88 |
| LAMPIRAN | | 96 |
| DAFTAR RIWAYAT HIDUP | | 128 |

DAFTAR GAMBAR

| | |
|--|----|
| Gambar 2. 1 <i>imbalance</i> dataset | 10 |
| Gambar 2. 2 Ilustrasi metode <i>Random forest</i> | 14 |
| Gambar 3. 1 Tahapan Medodologi Penelitian | 33 |
| Gambar 4. 1 <i>missing value</i> sebelum ditangani..... | 42 |
| Gambar 4. 2 grafik sebelum ditangani missing value sensor_51 | 45 |
| Gambar 4. 3 grafik setelah ditangani missing value sensor_51 | 45 |
| Gambar 4. 4 Statistik deskriptif data setelah normalisasi dengan penanganan <i>outlier</i> | 47 |
| Gambar 4. 5 Statistik deskriptif data setelah normalisasi tanpa penanganan <i>outlier</i> | 48 |
| Gambar 4. 6 grafik <i>imbalancing</i> data..... | 50 |
| Gambar 4. 7 Proses Pembagian Data | 52 |
| Gambar 4. 8 jumlah data pelatihan dan pengujian data <i>imbalance</i> | 53 |
| Gambar 4. 9 jumlah data pelatihan dan pengujian 100% data <i>balance</i> | 53 |
| Gambar 4. 10 jumlah data pelatihan dan pengujian 10% <i>balance</i> | 53 |
| Gambar 4. 11 jumlah data pelatihan dan pengujian 25% data <i>balance</i> | 53 |
| Gambar 4. 12 <i>Flowchart</i> pelatihan metode RF..... | 54 |
| Gambar 4. 13 Library pelatihan model | 54 |
| Gambar 4. 14 kombinasi parameter RF | 54 |
| Gambar 4. 15 <i>Instance</i> model Random Forest..... | 55 |
| Gambar 4. 16 Prediksi model..... | 55 |
| Gambar 4. 17 Visualisasi Pohon | 56 |
| Gambar 4. 18 visualisasi pohon cabang I model random forest | 57 |
| Gambar 4. 19 visualisasi pohon cabang II model random forest..... | 57 |
| Gambar 4. 20 visualisasi pohon cabang III model random forest..... | 57 |
| Gambar 4. 21 Visualisasi Prediksi Akhir Random Forest | 58 |
| Gambar 4. 22 Encoding kelas status mesin..... | 59 |
| Gambar 4. 23 hasil encoding status mesin | 59 |

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.

| | |
|---|----|
| Gambar 4. 24 laporan klasifikasi perbandingan 90:10..... | 61 |
| Gambar 4. 25 laporan klasifikasi perbandingan 80:20..... | 61 |
| Gambar 4. 26 laporan klasifikasi perbandingan 70:30..... | 62 |
| Gambar 4. 27 Proses SMOTE 100% | 63 |
| Gambar 4. 28 grafik 100% data <i>balance</i> | 63 |
| Gambar 4. 29 laporan klasifikasi perbandingan 90:10 data 100% <i>balance</i> | 64 |
| Gambar 4. 30 laporan klasifikasi perbandingan 80:20 data 100% <i>balance</i> | 64 |
| Gambar 4. 31 laporan klasifikasi perbandingan 70:30 data 100% <i>balance</i> | 65 |
| Gambar 4. 32 Proses SMOTE 25% | 65 |
| Gambar 4. 33 grafik 25% data <i>balance</i> | 65 |
| Gambar 4. 34 laporan klasifikasi perbandingan 90:10 data 25% <i>balance</i> | 66 |
| Gambar 4. 35 laporan klasifikasi perbandingan 80:20 data 25 25% <i>balance</i> | 67 |
| Gambar 4. 36 laporan klasifikasi perbandingan 70:30 data 25 25% <i>balance</i> | 67 |
| Gambar 4. 37 Proses SMOTE 10% | 67 |
| Gambar 4. 38 grafik 10% data <i>balance</i> | 68 |
| Gambar 4. 39 laporan klasifikasi perbandingan 90:10 data 10% <i>balance</i> | 69 |
| Gambar 4. 40 laporan klasifikasi perbandingan 80:20 data 10% <i>balance</i> | 69 |
| Gambar 4. 41 laporan klasifikasi perbandingan 70:30 data 10% <i>balance</i> | 70 |
| Gambar 4. 42 laporan klasifikasi perbandingan 90:10 data <i>imbalance</i> tanpa proses <i>outlier</i> | 72 |
| Gambar 4. 43 laporan klasifikasi perbandingan 80:20 data <i>imbalance</i> tanpa proses <i>outlier</i> | 72 |
| Gambar 4. 44 laporan klasifikasi perbandingan 70:30 data <i>imbalance</i> tanpa proses <i>outlier</i> | 73 |
| Gambar 4. 45 laporan klasifikasi perbandingan 90:10 data 100% <i>balance</i> tanpa <i>outlier</i> | 74 |
| Gambar 4. 46 laporan klasifikasi perbandingan 80:20 data 100% <i>balance</i> tanpa <i>outlier</i> | 75 |
| Gambar 4. 47 laporan klasifikasi perbandingan 70:30 data <i>balance</i> 100% tanpa <i>outlier</i> | 75 |

| | |
|---|----|
| Gambar 4. 48 laporan klasifikasi perbandingan 90:10 data <i>balance</i> 25% tanpa outlier | 76 |
| Gambar 4. 49 laporan klasifikasi perbandingan 80:20 data <i>balance</i> 25% tanpa outlier | 77 |
| Gambar 4. 50 laporan klasifikasi perbandingan 70:30 data <i>balance</i> 25% tanpa outlier | 77 |
| Gambar 4. 51 laporan klasifikasi perbandingan 90:10 data <i>balance</i> 10% tanpa outlier | 79 |
| Gambar 4. 52 laporan klasifikasi perbandingan 80:20 data <i>balance</i> 10% tanpa outlier | 80 |
| Gambar 4. 53 laporan klasifikasi perbandingan 70:30 data <i>balance</i> 10% tanpa outlier | 80 |

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 Imputasi Rata-Rata..... | 6 |
| (2) Rumus Quartil Satu (Q1)..... | 8 |
| (3) Rumus Quartil Tiga (Q3)..... | 8 |
| (4) Rumus <i>Interquartil Range</i> (IQR) | 8 |
| (5) Rumus Batas Atas | 8 |
| (6) Rumus Batas Bawah..... | 8 |
| (7) Rumus Standar Deviasi | 9 |
| (8) Rumus <i>Z-score</i> | 9 |
| (9) Rumus <i>imbalance Ratio</i> (IR)..... | 11 |
| (10) Rumus Presentase Kelas Minoritas | 11 |
| (11) Rumus Akurasi Klasifikasi..... | 13 |
| (12) Rumus Gini..... | 14 |
| (13) Rumus Gini Split..... | 14 |
| (14) Rumus Mtry Satu..... | 15 |
| (15) Rumus Mtry Dua | 15 |
| (16) Rumus Mtry Tiga | 15 |
| (17) Rumus <i>Accuracy</i> | 19 |
| (18) Rumus <i>Precision</i> | 20 |
| (19) Rumus <i>Recall</i> | 20 |
| (20) Rumus <i>F1-score</i> | 20 |

UIN SUSKA RIAU

- Hak Cipta Dilindungi Undang-Undang
- 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.
 - 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.

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Mesin pompa air digunakan sebagai alat untuk menyalurkan air dari sumber air (sumur) ke tempat penyimpanan air sementara seperti bak mandi dan tangki air (Zhang et al., 2019). Terjadi kerusakan pada mesin pompa air dapat menyebabkan masalah yang besar salah satunya dapat mengakibatkan kesalahan mekanis pada mesin pompa. (Ahmad et al., 2020).

Beberapa penelitian telah dilakukan tentang mesin pompa diantaranya penelitian yang dilakukan oleh Vinaya, dkk tentang klasifikasi kondisi empat pompa air yang berbeda yang terdiri dari pompa normal, ketidakseimbangan, ketidaksejajaran, dan kerusakan bantalan. Penelitian ini menggunakan *Fuzzy Inference System* (FIS) dan *Fast Fourier Transform* (FFT). Hasil penelitian menunjukkan akurasi klasifikasi sebesar 90% (Vinaya et al., 2019). Penelitian yang dilakukan oleh Minzheng Jiang, dkk menggunakan metode *Random forest* melakukan diagnosis kesalahan pada pompa *submersible screw*, dimana menggunakan 10 parameter diantaranya termasuk parameter variabel kontinu dan kategori seperti daya aktif, produksi cairan, tekanan minyak, tekanan casing, serta jenis pasir, lilin, tipe pompa, dan nomor sumur. Hasil dari penelitian tersebut mendapatkan akurasi sebesar 92.86% (Jiang et al., 2020). Penelitian lainnya yang dilakukan oleh Lestari Handayani untuk memprediksi kondisi mesin pompa air sebelum terjadi kerusakan pada dataset *pump sensor* menggunakan metode *Hidden Semi-Markov Model* (Handayani, 2022).

Pada penelitian (Handayani, 2022) bentuk permasalahan yang terjadi dalam penelitian tersebut yaitu sistem pemompaan air di daerah tersebut mati sebanyak 7 kali dalam setahun sehingga menimbulkan permasalahan yang serius bagi keluarga tersebut. Menurut seorang konsultan bernama Kaer Halewood, ia memberikan

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.

mesin yaitu normal, *recovering* dan *broken*. Selain itu dalam penanganan *outlier* menggunakan metode *Interquartile Range (IQR)*.

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini sebagai berikut:

1. Dapat menghitung akurasi metode *Random forest* dalam klasifikasi status mesin pompa air
2. Dapat membandingkan akurasi metode *Random forest* setelah dan sebelum ditangani *imbalance* data

1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini sebagai berikut:

1. Sistem bisa mengklasifikasi data mesin pompa air menjadi beberapa kelas
2. Menambah kajian tentang kasus dan metode yang memberi kontribusi pada penelitian tersebut.

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 Mesin Pompa Air

Pompa adalah suatu mesin atau alat yang digunakan untuk mengangkat air dari dataran ke pegunungan atau untuk meningkatkan tekanan air dari air rendah agar air menjadi tinggi serta aliran pada sistem jaringan transmisi. Hal ini dicapai dengan menciptakan tekanan rendah pada saluran masuk atau sisi air dan tekanan tinggi pada saluran keluar atau keluar pompa (Khairuddin Tampubolon, Alinur, Elazhari, Ardi Ermawy, 2021).

Mesin pompa air mengalir air dari sumur ke bak mandi, tandon air, dan tempat lain yang membutuhkan air. (Arifin et al., 2020). Dengan adanya mesin pompa air sehingga dapat memudahkan aktifitas yang di lakukan masyarakat untuk keperluan rumah tangganya.

2.2 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). *Pre-processing* memiliki beberapa tahapan diantaranya mengelola *missing value*, mengelola *duplicated* data, mengelola data *outliers*, *normalization* data, *feature encoding*, dan *class imbalance*.

2.2.1 Missing values

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

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 berbeda, atau kesalahan instrumental selama proses pengambilan data (Mishra et al., 2020).

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 (Bilal et al., 2022).

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 digunakan jika datanya banyak (Prabiantissa Citra Nurina, 2021). 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 :

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

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

Keterangan :

\bar{x} = Rata – rata (Mean)

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

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.

N = Jumlah dari seluruh data

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

2.2.2 Outliers

Outlier adalah jenis pengamatan di mana titik pengamatan menyimpang dari pola data dan jauh dari pusat data. *Outlier* biasanya dapat disebabkan oleh kesalahan manusia, kesalahan instrumen, perilaku curang, perubahan dalam perilaku sistem atau kesalahan sistem, dan penyimpangan alami dalam populasi. Ada kemungkinan bahwa adanya *outlier* pada data akan menyebabkan kesalahan dalam hasil analisis data dan ketidakakuratan hasil penelitian. Oleh karena itu, salah satu cara untuk mengatasi masalah ini adalah dengan mengidentifikasi adanya *outlier* pada data (Sari et al., 2021).

Ada beberapa alasan mengapa data *outlier* muncul. Yang pertama adalah kesalahan dalam meng-entri data, yang kedua adalah kesalahan dalam mengidentifikasi nilai yang hilang dalam program komputer, dan yang terakhir adalah *outlier* yang berasal dari populasi yang kami sampelkan, tetapi memiliki nilai yang tidak normal (Pradana & Astika, 2019).

Keadaan *outlier* adalah ketika nilai parameter ekstrim muncul dalam kumpulan data. Masalah ini perlu ditangani karena pencilan atau *outlier* dapat menyebabkan kesalahan klasifikasi model, bias dalam estimasi parameter, atau hasil data di bawah standar. Metode *Interquartile Range* (IQR) dapat digunakan untuk mengatasi *outlier* (Dinata et al., 2023). Setelah nilai *outlier* terdeteksi, nilai tersebut dapat diganti dengan mean dari masing-masing kolom yang terdeteksi *outlier* (Syukron et al., 2020). Berikut adalah tahapan untuk menghitung IQR (Irianto et al., 2022) :

1. Menentukan data yang akan diatasi *outlier*
2. Menentukan nilai Q_3 (kuartil 3) dan nilai Q_1 (kuartil 1) dengan persamaan 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.

$$Q1 = \frac{25}{100}(n) \quad (2)$$

$$Q3 = \frac{75}{100}(n) \quad (3)$$

3. Menghitung nilai IQR (*Interquartile Range*) dengan rumus sebagai berikut :

$$IQR = Q3 - Q1 \quad (4)$$

Keterangan :

IQR : Nilai *interquartil range*

Q3 : Nilai kuartil 3

Q1 : Nilai kuartil 1

4. Selanjutnya menghitung nilai batas atas dan batas bawah dengan rumus sebagai berikut :

$$Batas atas = Q3 + 1,5 * IQR \quad (5)$$

$$Batas bawah = Q1 - 1,5 * IQR \quad (6)$$

5. Langkah terakhir yaitu membandingkan data dengan nilai batas atas dan bawah untuk mengidentifikasi *outlier*, setelah itu mengganti nilai *outlier* dengan *mean*.

2.2.3 Normalisasi

Dalam pemetaan atau penskalaan, salah satu teknik *pre-processing* adalah normalisasi. Pada tahap ini, skala data diubah menjadi skala yang lebih kecil. Skala

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.

data yang baru dapat membantu klasifikasi karena dapat menghapus fitur yang memiliki tingkat noise dan relevansi yang rendah (Riskha Chairunisa et al., 2020).

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 untuk menemukan data yang merupakan bagian dari *outlier* data (Henderi, 2021). Metode normalisasi *Z-score* menghasilkan hasil dari nilai rata-rata dan standar deviasi data. Berikut adalah tahapan untuk menghitung *Z-score normalization* (Gde Agung Brahmama Suryanegara et al., 2021):

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

$$\sigma = \frac{\sqrt{\sum f_i(x_i - \bar{x})^2}}{\sum f_i} \quad (7)$$

Keterangan :

σ = Standar deviasi

x_i = Nilai tengah

\bar{x} = Nilai rata-rata (mean)

f_i = Frekuensi

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

$$n_i^1 = \frac{n_i - \bar{A}}{\sigma_A} \quad (8)$$

Dimana :

n_i^1 = Hasil *Z-score* normalization

n_i = data yang akan dinormalisasi

\bar{A} = nilai rata-rata

Hak Cipta Dilindungi Undang-Undang

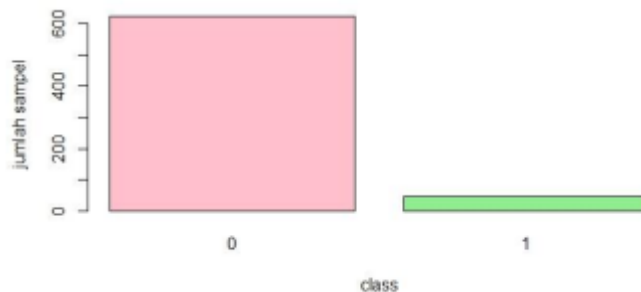
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan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

σ_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 (Henderi, 2021).

2.2.4 Imbalance

Salah satu komponen paling berpengaruh dalam kinerja prediksi klasifikasi adalah ketidakseimbangan kelas. Model pembelajaran bias yang dihasilkan oleh pengklasifikasi cenderung tidak akurat untuk kelas minoritas dibandingkan dengan kelas mayoritas. Ini disebabkan oleh fakta bahwa sebagian besar pembelajaran pengklasifikasi, termasuk pohon keputusan, *Neural Network backpropagation*, *Support Vector Machines*, dan lainnya, dirancang dengan asumsi bahwa distribusi kelas relatif seimbang. Jika nilai variabel target memiliki jumlah sampel yang jauh lebih kecil daripada nilainya di kelas lain, data dianggap tidak seimbang. Seperti gambar berikut (Prasetya, 2022)



Gambar 2.1 imbalance dataset

Nilai variabel target yang lebih kecil daripada sampelnya menyebabkan ketidakseimbangan (Prasetya, 2022). Tingkat ketidakseimbangan kelas dapat dihitung dengan dua cara yaitu rasio ketidakseimbangan (*imb*) seperti yang ditunjukkan dalam Persamaan (1) atau persentase kelas minoritas (*%minoritas*) seperti yang ditunjukkan dalam Persamaan (2). Di mana *M* dan *m* adalah jumlah instance di kelas mayoritas dan minoritas (Vuttipittayamongkol et al., 2021).

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.

$$\text{imb} = \frac{M}{m} \quad (9)$$

$$\text{minority}(\%) = \frac{m}{M} * 100 \quad (10)$$

Mengatasi kumpulan data yang tidak seimbang dapat dilakukan secara algoritmik atau di tingkat data. Proses mengubah algoritma pembelajaran mesin untuk menangani ketidakseimbangan data dikenal sebagai pendekatan pada tingkat algoritmik. Algoritma seperti *Neural Network KMeans*, *Naive Bayes*, *Random forest*, dan C4.5 biasanya dimodifikasi.

Dalam pendekatan tingkat data, *resampling* diperlukan untuk mengurangi ketidakseimbangan kelas. Dua metode pengambilan sampel yang umum digunakan pada tingkat data adalah *random oversampling* (ROS) dan *random undersampling* (RUS) (Arifiyanti & Wahyuni, 2020). Selain itu teknik *oversampling heuristik* tingkat lanjut yang disebut *Synthetic Minority Oversampling Technique (SMOTE)*.

2.2.4.1 Synthetic Minority Oversampling Technique (SMOTE)

Minority Oversampling Technique (SMOTE) adalah metode *oversampling* yang bekerja dengan replikasi data secara acak untuk meningkatkan jumlah kelas positif, sehingga jumlah data positif sama dengan data negatif. Algoritma SMOTE menggunakan duplikasi data sintetis pada kelas yang kecil untuk menggunakannya. Ini dilakukan dengan mencari tetangga terdekat k untuk kelas positif, kemudian membuat duplikasi data sintetis sebanyak persentase yang diinginkan antara kelas yang dipilih secara acak dan positif. Metode ini memungkinkan penilai untuk mengatasi ketidakseimbangan data (Churn & Kombinasi, 2021).

Rumus perhitungan SMOTE dapat dilihat pada persamaan 12 berikut :

$$X_{syn} = X_i + (X_{knn} - X_i) \times t \quad (11)$$

Keterangan :

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.

- X_{syn} = Pengamatan baru hasil pembangkitan
- X_i = Pengamatan ke-i
- X_{knn} = x terdekat dari xi
- t = Bilangan acak antara 0 dan 1

Untuk menghindari *overfitting*, metode sampling SMOTE menambahkan kelas minoritas dengan menghasilkan data buatan atau sintesis berdasarkan k-tetangga terdekat (*k-nearestneighbor*) antar kelas minoritas (Saputro et al., 2020).

2.3 Data mining

Data mining adalah disiplin ilmu yang mempelajari cara mengumpulkan informasi atau menemukan pola dalam data. Teknologi ini pertama kali muncul di bidang ilmu komputer pada tahun 1990. *Data mining* digunakan untuk mencari pengetahuan dalam basis data yang besar, karena itu sering disebut *Knowledge Discovery Database (KDD)*. Ini dilakukan dengan mengidentifikasi pola yang relevan atau menarik dari data yang ada dalam basis data dan menghasilkan data (Nur, 2022).

Data mining adalah proses mengekstraksi dan mengidentifikasi informasi yang relevan dari berbagai database yang sangat besar dengan menggunakan matematika, statistik, kecerdasan buatan, dan *machine learning* (Yuli Mardi, 2019).

Salah satu jenis dari kelompok data mining adalah klasifikasi, dimana Pada *data mining*, klasifikasi digunakan untuk mengelompokkan item berdasarkan kategori. Kategori ini dapat merujuk pada item yang telah digolongkan sebelumnya, seperti gaji rendah, sedang, dan tinggi, dan curah hujan rendah, sedang, dan tinggi, antara lain (Dewi, 2020).

2.4 Klasifikasi

Salah satu metode pengolahan data adalah klasifikasi, yang menggabungkan data ke dalam kelompok atau kelas yang telah ditentukan. Metode *supervised learning* yang menggunakan klasifikasi data yang dilabelkan untuk menghasilkan

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.

aturan yang mengklasifikasikan data uji ke dalam kelompok atau kelas yang telah ditentukan (Setio et al., 2020).

Untuk mengevaluasi kinerja algoritma pembelajaran mesin, indikator penilaian sangat penting. Akurasi dan *area under the curve* (AUC) adalah dua indikator penilaian di bidang klasifikasi. Dimana akurasi dapat diukur dengan persamaan berikut:

$$Akurasi = \frac{TP + TN}{TP + FP + TN + FN} \times 100 \quad (12)$$

True *Positive* (TP) adalah jumlah sampel positif yang diprediksi benar, False *Positive* (FP) adalah banyaknya sampel positif yang diprediksi salah, True *Negative* (TN) adalah jumlah sampel negatif yang diprediksi dengan benar, dan False *Negative* (FN) adalah jumlah sampel negatif yang diprediksi dengan salah (Yoga Religia et al., 2021).

Ada banyak algoritma yang dapat digunakan untuk melakukan proses klasifikasi teks, seperti *Support Vector Machine (SVM)*, *Naive Bayes*, *Random forest (RF)*, *k-Nearest Neighbor (KNN)*, *Decision Tree*, dan *Artificial Neural Networks (ANN)* (Akromunnisa et al., 2019). Metode *Random forest* dapat menangani banyak sampel data dan mengklasifikasi data yang tidak lengkap.

2.5 Random forest

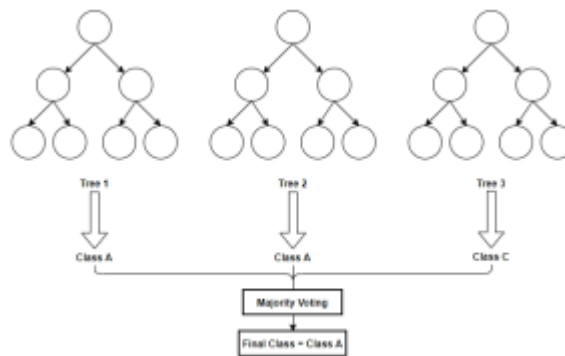
Random forest adalah teknik pembelajaran *ensemble* yang menggunakan beberapa pohon keputusan dari berbagai subset data dan melakukan voting terhadap hasil dari beberapa pohon keputusan untuk menghasilkan output dari pohon random (Sun et al., 2020). Dengan kata lain, *Random forest* terdiri dari sejumlah pohon keputusan, atau pohon keputusan. Kumpulan pohon keputusan ini digunakan untuk mengklasifikasikan data ke dalam suatu kelas.

Random forest adalah salah satu metode *data mining* CART (*Classification and Regression Tree*), yang tidak memerlukan asumsi. Konsep pohon keputusan (*decision tree*) digunakan dalam metode ini. Model ini terdiri dari banyak pohon

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 membentuk sebuah kumpulan pohon seperti hutan. Untuk melakukan ini, metode *aggregating bootstrap* (bagging) dan pemilihan fitur acak digunakan. Gambar berikut menunjukkan pohon keputusan dan pengambilan keputusan dengan hutan random (Iman & Wahyu, 2021).



Gambar 2. 2 Ilustrasi metode *Random forest*

2.5.1 Tahapan pembuatan pohon keputusan pada algoritma *Random forest*

Berikut adalah tahapan dalam pembuatan pohon keputusan

1. Proses pemecahan dan pelabelan simpul. Pada proses ini dilakukan pemecahan pada simpul akar menjadi dua cabang. Proses pemecahan simpul ini dilakukan dengan menggunakan indeks gini:

$$Gini(T) = 1 - \sum_{j=1}^n \pi_j^2 \quad (13)$$

Pada dasarnya, indeks GINI dan *entropy* informasi adalah beberapa metode untuk mengukur kebaikan suatu fitur (juga disebut atribut). Dalam kebanyakan kasus, metode ini dianggap menghasilkan pohon yang sebanding.

2. Setelah itu kumpulan data T dibagi menjadi dua himpunan bagian T1 dan T2, masing-masing dengan ukuran N1 dan N2, dan indeks pada split data berisi contoh dari kelas n, maka indeks gini(T) dirumuskan dengan:

$$Gini\ split(T) = \frac{n_1}{n} Gini(T)_1 + \frac{n_2}{n} Gini(T)_2 \quad (14)$$

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.

Proses tersebut terus dilakukan pada tiap variabel hingga diperoleh $Gini_{split}$ terkecil. Selanjutnya proses pelabelan kelas diperoleh dari hasil $Gini_{split}$ terkecil.

3. Setelah cabang terbentuk, ulangi langkah 1 dan 2. Namun jika cabang telah mencapai maksimal cabang yang diperbolehkan, daun akan terbentuk dengan nilai mayoritas dari nilai data.

Random forest membutuhkan dua parameter yaitu jumlah pohon yang akan digunakan dan jumlah variabel independen yang digunakan untuk proses pencabangan. Untuk mencapai nilai yang optimal, kedua parameter ini harus dipenuhi. Untuk mendapatkan nilai $mtry$ dapat menggunakan rumus berikut (Perdana et al., 2021):

$$mtry_1 = \frac{1}{2} |\sqrt{p}| \tag{15}$$

$$mtry_2 = |\sqrt{p}| \tag{16}$$

$$mtry_3 = 2 \times |\sqrt{p}| \tag{17}$$

Dimana:

$mtry$: banyaknya variabel independent untuk setiap split

p : banyaknya variabel independent

2.5.2 Algoritma *Random forest*

Algoritma *Random forest* untuk klasifikasi yaitu sebagai berikut (Lestari & Sirodj, 2022):

1. Membuat suatu bootstrap sample yaitu melakukan penarikan sampel acak berukuran N pengamatan dari data training dengan pengembalian (replacement).

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.

Menggunakan sampel bootstrap (BS_j) sebagai data training untuk membangun pohon ke- j ($j = 1, 2, \dots, J$), setiap pohon tumbuh secara maksimal tidak dilakukan pemangkasan. pembangunan pohon klasifikasi dilakukan dengan menerapkan random future selection dengan mengulangi langkah-langkah untuk setiap node terminal pohon, sampai ukuran node minimum tercapai.

- (i) Pilih variabel m secara acak dari variabel prediktor p dengan $m_{try} < p$. Menggunakan persamaan (13), (14), dan (15).
 - (ii) Pilih variabel/split terbaik diantara semua kemungkinan m_{try} fitur dari proses (i) sebagai penyekat.
 - (iii) Pisahkan *node* menjadi dua *sub - node* berdasarkan solusi split yang diperoleh dari proses (ii).
3. Ulangi langkah 1-3 sebanyak j kali sehingga terbentuk sebuah hutan yang terdiri dari j pohon. Tiap pohon klasifikasi akan menghasilkan satu suara hasil prediksi sehingga akan didapatkan j buah suara.
 4. selanjutnya menentukan hasil prediksi akhir yaitu dengan cara mengaggregasi atau menggabungkan hasil prediksi setiap pohon dengan menggunakan pemungutan suara terbanyak (majority vote).

Random forest adalah pengklasifikasi yang terdiri dari kumpulan klasifikasi pohon-terstruktur $\{h(x, \Theta_k), k = 1, \dots\}$ di mana $\{\Theta_k\}$ adalah distribusi vektor acak independen dan setiap pohon memberikan satu unit suara untuk kelas paling populer pada input x . *Random forest* adalah kombinasi pohon prediktor di mana setiap pohon bergantung pada nilai vektor acak yang dijadikan sampelindependen dengan distribusi yang sama untuk semua pohon di hutan (Aryani & Wijayanto, 2021).

Untuk mencapai keputusan akhir, RF melakukan klasifikasi dengan menggunakan metode kelompok dengan kemunculan mayoritas dari berbagai pohon. Pada algoritma RF, set data pelatihan didefinisikan sebagai $S = \{(x_i, y_j), i = 1, 2, \dots, N; j = 1, 2, \dots, M\}$ dengan x sebagai sampel dan y sebagai variabel fitur S . Jumlah sampel pelatihan adalah N , dan setiap sampel memiliki variabel fitur M (Noga Religia et al., 2021).

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.

Untuk klasifikasi, algoritma RF dibangun dalam tiga langkah:

- (1) mengumpulkan himpunan bagian pelatihan k ,
- (2) membuat setiap model pohon keputusan, dan
- (3) menggabungkan k pohon ke dalam model RF. Algoritma RF memiliki hasil performa yang baik dan waktu eksekusi yang cepat, sehingga dapat diterapkan pada data yang tidak seimbang dalam jumlah besar.

2.5.3 Parameter *Random forest*

Random forest mempunyai beberapa parameter dimana dapat digunakan sesuai kebutuhan. Parameter ini terdapat beberapa pilihan dimana Ketika ditentukan, akan melakukan proses pembelajaran sesuai kebutuhan dan keinginan pengguna. Berikut beberapa parameter yang ada pada *Random forest*.

a. *n_estimator*

n_estimator merupakan jumlah pohon yang ada dalam suatu forest. Nilai pada *n_estimator* dapat diubah sesuai dengan kebutuhan atau keinginan pengguna.

b. *Criterion*

Criterion digunakan untuk mengukur kualitas dari split. Kriteria yang didukung yakni “gini” untuk ketidakmurnian gini, dan “entropy” digunakan dalam perolehan informasi.

c. *Max_depth*

Max_depth merupakan kedalaman maksimum dalam suatu pohon. Jika tidak ada, maka node akan diperluas sampai semua daun murni atau sampai semua daun berisi kurang dari sampel *min_samples_splits*.

d. *Min_samples_split*

Min_samples_split merupakan jumlah sampel minimal yang diperlukan untuk memisahkan node internal.

e. *min_samples_leaf*

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.

Jumlah minimum sampel yang diperlukan berada pada simpul daun. Titik perpecahan pada kedalaman apa pun hanya akan dipertimbangkan jika *min_samples_leaf* masing-masing di cabang kiri dan kanan.

f. *Max_features*

Max_features merupakan jumlah fitur yang harus dipertimbangkan disaat mencari split terbaik. *Max_features* memiliki nilai-nilai, antara lain auto, log2, dan sqrt. Pencarian split tidak akan berhenti sampai setidaknya satu partisi valid dari node sampel ditemukan.

2.5.4 Keuntungan *Random forest*

Algoritma RF memiliki banyak keunggulan diantaranya

1. Menghasilkan tingkat kesalahan yang relatif kecil
2. Kinerja klasifikasi yang luar biasa
3. Kemampuan untuk menangani jumlah besar data pelatihan
4. Metode yang efektif untuk memperkirakan data yang hilang menggunakan sampel pelatihan dan subset yang dipilih secara acak dari variabel input pada setiap node
5. Mengatasi masalah utama *overfitting* pohon keputusan dimana menyimpan dan menyesuaikan data pelatihan dengan ketat. Beberapa pohon keputusan dibuat untuk mengatasi masalah ini dan membuat model yang lebih umum di RF (Agustiani et al., 2022).

2.6 Pembagian Data

Pada proses ini data dibagi menjadi dua bagian yaitu data *training* dan data *testing*. Proses pembentukan data *training* yaitu dengan cara menggunakan komposisi yang telah ditentukan dari jumlah seluruh data. Sedangkan data *testing* merupakan sisa komposisi data yang telah digunakan pada data *training* (Siti Aisyah, Sri Wahyuningsih, 2021). Ada tiga pendekatan pembagian jumlah data baik untuk data latih maupun data uji dengan perbandingan 90:10, 80:20, 70:30 (Widiastiwi & Ernawati, 2021).

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.

2.7 Evaluasi

Tahap selanjutnya adalah evaluasi setelah hasil pengujian dikumpulkan. Sebuah evaluasi dilakukan untuk mengukur performa model yang dibuat. Matriks klasifikasi *confusion matrix* adalah alat pengukuran performa yang digunakan dalam penelitian ini. Tabel *confusion matrix* menunjukkan jumlah prediksi yang benar dan salah pada set data yang diberikan untuk menjelaskan kinerja model pembelajaran mesin (Martias et al., 2023).

Confusion Matrix merupakan sebuah cara untuk mempresentasikan hasil daripada akurasi dari model yang sudah dibuat. *Confusion Matrix* merangkum performa dalam mengklasifikasi sesuai dengan banyaknya kategori yang diklasifikasi dari nilai benar dari suatu kelas dari objek yang diprediksi (Aldean et al., 2022).

Table 2. 1 Confusion Matrix

| <i>Confusion Matrix</i> | <i>Predicted Positive</i> | <i>Predicted Negative</i> |
|-------------------------|---------------------------|---------------------------|
| <i>Actual Positive</i> | <i>TP</i> | <i>TN</i> |
| <i>Actual Negatif</i> | <i>FP</i> | <i>FN</i> |

TP = jumlah data positif yang terklasifikasi benar

TN = jumlah data positif yang terklasifikasi salah

FP = jumlah data positif yang terklasifikasi salah

FN = jumlah data negatif yang terklasifikasi salah.

Berikut adalah rumus yang digunakan untuk perhitungan *confusion matrix*:

1. Accuracy

Accuracy adalah persentase prediksi yang benar yang dilakukan oleh model.

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

(18)

2. Precision

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.

Precision merupakan sebuah definisi dari ratio antara *True Positive* dan *Predicted Positive*

$$Precision = \frac{TP}{TP+FP} \quad (19)$$

3. *Recall*

Recall adalah rasio antara *True Positives* dan *Actual Positives*.

$$Recall = \frac{TP}{TP+FN} \quad (20)$$

4. *F1-Score*

F1-score adalah hasil rata-rata dari *precision* dan *recall*.

$$f1 - score = 2x \frac{precision \times recall}{precision + recall} \quad (21)$$

2.8 Penelitian Terkait

Tabel berikut berisikan jurnal yang berkaitan dengan tugas akhir.

Table 2. 2 Penelitian terkait

| No | Penulis | Judul | Hasil | Kelebihan dan Kekurangan |
|----|--|--|---|---|
| 1. | Yihan Wang, Hongmei Liu (Wang & Liu, 2019) | <i>Centrifugal pump fault diagnosis based on MEEMD-PE Time-frequency information entropy</i> | Metode diagnosis kesalahan pada pompa sentrifugal menggunakan Entropi Informasi Frekuensi Waktu Modified Ensemble Empirical Mode Decomposition-Permutation Entropy (MEEMD-PE) dan | Kelebihan dari penelitian ini adalah mendapatkan akurasi tinggi dengan metode yang diusulkan mencapai akurasi klasifikasi 100%. Kekurangan dari penelitian adalah kurang membahas secara rinci |

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.

| No | Penulis | Judul | Hasil | Kelebihan dan Kekurangan |
|----|---|--|---|--|
| 1. | | <i>and Random forest</i> | klasifikasi <i>Random forest</i> . Mendapatkan akurasi 100% menggunakan metode <i>Random forest</i> . | mengenai potensi keterbatasan atau tantangan dari metode diagnosis kesalahan yang diusulkan untuk pompa sentrifugal. |
| 2. | Qin Hu, Xiao-Sheng Si, Qing-Hua Zhang, Ai-Song Qin (Hu et al., 2020) | <i>A rotating machinery fault diagnosis method based on multi-scale dimension less indicators and random forests</i> | diagnosis kerusakan untuk mesin berputar menggunakan Multi-Scale Dimensionless Indicators (MSDIs) dan klasifikasi <i>Random forest</i> (RF). klasifikasi kerusakan menggunakan RF, menghasilkan akurasi rata-rata sebesar 95,58%. | Kelebihan penelitian adalah kemampuan yang lebih kuat dalam menyelesaikan masalah klasifikasi kerusakan. Kekurangannya adalah kompleksitas komputasi yang meningkat seiring dengan peningkatan jumlah faktor skala pada MSDIs. |
| 3. | Sayed Ali Sharaf, Tatweer Petroleum, Patrick Bangert, Algorithmica | <i>Beam Pump Dynamometer Card Classification Using</i> | Pompa batang penghisap (rod pumping) adalah metode pengangkatan buatan yang paling tua dan paling banyak digunakan. Penggunaan | Kelebihan dari penelitian ini adalah kemampuannya untuk meningkatkan efisiensi surveilans dan diagnosis pompa batang penghisap |

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.

| No | Penulis | Judul | Hasil | Kelebihan dan Kekurangan |
|----|--|---|--|---|
| a | Technologies, Mohamed Fardan, Khalil Alqassab, Mohamed Abubakr, Mahmood Ahmed, dan Tatweer Petroleum (Sharaf et al., 2019) | <i>Machine Learning</i> | teknik machine learning untuk pengenalan pola dapat membantu mengotomatisasi proses interpretasi visual, meningkatkan efisiensi, dan mengurangi aktivitas pemeliharaan akibat diagnosis dini yang terlewatkan | (rod pumping) dalam industri minyak dan gas. Kekurangannya adalah Kekurangan dari penelitian ini tidak disebutkan secara eksplisit dalam teks yang disediakan. Oleh karena itu, kutipan spesifik untuk keterbatasan penelitian tidak tersedia |
| 4. | Zainib Noshad, Nadeem Javaid, Tanzila Saba, Zahid Wadud, Muhammad Qaiser Saleem, Mohammad Eid Alzahrani dan Osama E. Sheta | <i>Fault Detection in Wireless Sensor Networks through the Random forest Classifier</i> | Penelitian ini membahas deteksi kesalahan dalam jaringan sensor nirkabel menggunakan klasifikasi <i>Random forest</i> . Studi mengevaluasi kinerja enam klasifikasi (SVM, RF, SGD, MLP, CNN, dan PNN) untuk deteksi kesalahan dalam jaringan sensor nirkabel. Hasilnya | Kelebihan dari penelitian ini adalah melakukan perbandingan kinerja enam klasifikasi yang berbeda, memberikan wawasan yang berharga tentang klasifikasi mana yang paling efektif dalam mendeteksi kesalahan dalam WSN. Kekurangannya adalah tidak membahas skalabilitas teknik deteksi kesalahan yang |

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.

| No | Penulis | Judul | Hasil | Kelebihan dan Kekurangan |
|------------------------|---|---|--|--|
| a milik UIN Suska Riau | (Noshad et al., 2019) | | menunjukkan bahwa RF memiliki kinerja lebih baik daripada klasifikasi lainnya, sementara SVM menempati peringkat kedua. | diusulkan, terutama dalam implementasi WSN berskala besar. |
| 5. | Nugraha Listiana Hanun; Achmad Udin Zailani (Zailani & Hanun, 2020) | Penerapan Algoritma Klasifikasi <i>Random forest</i> Untuk Penentuan Kelayakan Pemberian Kredit Di Koperasi Mitra Sejahtera | Penelitian ini menggunakan model <i>Random forest</i> untuk menganalisis kelayakan pemberian kredit. Data uji diambil dari laporan kredit Koperasi Mitra Sejahtera tahun 2016 sebanyak 20 sampel. Evaluasi model dilakukan dengan menggunakan confusion matrix dan kurva ROC. Hasil evaluasi menunjukkan bahwa Algoritma <i>Random forest</i> mampu menganalisis kredit dengan akurasi sebesar 87,88%. Model Pohon Keputusan Algoritma | Kelebihan dari penelitian ini adalah menggunakan Algoritma <i>Random forest Classification</i> untuk menganalisis kelayakan kredit, yang telah terbukti efektif dalam memprediksi debitur yang bermasalah dan yang tidak bermasalah dengan tingkat akurasi sebesar 87,88%. Penelitian ini menggunakan 20 sampel dari laporan kredit Koperasi Mitra Sejahtera pada tahun 2016, sehingga memberikan set data yang spesifik dan relevan untuk dianalisis. |

| No | Penulis | Judul | Hasil | Kelebihan dan Kekurangan |
|----|--|---|--|---|
| a | miilik UIN Suska Riau | | <p><i>Random forest</i> digunakan untuk memprediksi kelayakan pemberian kredit pada Koperasi Mitra Sejahtera. Metode penelitian menggunakan model CRISP-DM yang terdiri dari enam tahap. Tahap-tahap tersebut meliputi pemahaman bisnis, pemahaman data, persiapan data, pemodelan, evaluasi, dan penerapan model. Tahap evaluasi melibatkan pengujian dengan Cross validation, confusion matrix, dan kurva ROC.</p> | <p>Kekurangan dari penelitian ini adalah hanya menggunakan 20 sampel dari laporan kredit Koperasi Mitra Sejahtera pada tahun 2016, yang mungkin tidak sepenuhnya mewakili keragaman situasi kredit dan peminjam. Selain itu, penelitian ini tidak memberikan pembahasan rinci tentang potensi bias atau keterbatasan dari Algoritma Klasifikasi <i>Random forest</i>, yang dapat mempengaruhi akurasi dan generalisasi hasil.</p> |
| 6. | Siska Devella; Yohannes; Firda Novia Rahmawati | Implementasi <i>Random forest</i> Untuk Klasifikasi | <p>Hasil penelitian ini menunjukkan bahwa implementasi metode SIFT dan klasifikasi menggunakan <i>Random forest</i> berhasil dalam</p> | <p>Kelebihan penelitian ini adalah berhasil mengimplementasikan metode SIFT dan klasifikasi <i>Random forest</i> untuk mengklasifikasikan</p> |

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.

| No | Penulis | Judul | Hasil | Kelebihan dan Kekurangan |
|------------------------|------------------------|--|--|--|
| a milik UIN Suska Riau | (Devella et al., 2020) | Motif Songket Palembang Berdasarkan SIFT | mengklasifikasi motif songket Palembang. Metode SIFT melibatkan deteksi ekstrema skala, lokalisasi keypoint, penugasan orientasi, dan deskripsi keypoint. <i>Random forest</i> digunakan untuk klasifikasi motif songket dan hasil evaluasi menunjukkan tingkat akurasi, presisi, dan <i>recall</i> yang tinggi. Dengan menggunakan 115 citra dari setiap motif, penelitian ini mencapai akurasi keseluruhan sebesar 92.98%. Saran untuk penelitian selanjutnya termasuk menambah jumlah citra motif songket dan menggunakan metode ekstraksi fitur dan klasifikasi lainnya. | motif songket Palembang secara akurat. Selain itu, penelitian ini mencapai akurasi, presisi, dan <i>recall</i> yang tinggi dalam klasifikasi motif, dengan akurasi keseluruhan sebesar 92,98%. Kekurangan dari penelitian ini adalah tidak semua objek gambar dapat diklasifikasikan secara penuh oleh metode <i>Random forest</i> karena adanya perbedaan antara motif pada dataset pelatihan dan motif pada data pengujian. Selain itu, penelitian ini tidak mengeksplorasi dampak potensial dari kondisi pencahayaan yang berbeda-beda terhadap klasifikasi motif, yang dapat mempengaruhi akurasi dan ketahanan model 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.

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.

| No | Penulis | Judul | Hasil | Kelebihan dan Kekurangan |
|----|--|---|--|---|
| 7. | Zahoor Ahmad, Akhand Rai, Andrei S. Maliuk, dan Jong-Myon Kim (Ahmad et al., 2020) | <i>Discriminant Feature Extraction for Centrifugal Pump Fault Diagnosis</i> | Hasil penelitian menunjukkan bahwa strategi perawatan berbasis kondisi (CBM) sangat menguntungkan untuk diagnosis kerusakan pada pompa sentrifugal (CP). Penelitian ini menemukan bahwa 39% kegagalan pompa disebabkan oleh kerusakan segel mekanik. | Kelebihan penelitian ini adalah pengembangan metode ekstraksi fitur diskriminan yang mempertimbangkan sinyal getaran pompa sentrifugal (CP) menggunakan fungsi cross-korelasi. Kekurangannya adalah kurangnya penekanan pada pengujian dan validasi metode di lapangan yang dapat mempengaruhi generalisasi hasil penelitian ini dalam situasi nyata. |
| 8. | Mauricio Barrios Castellanos, Alberto Luiz Serpaa, Jorge Luiz Biazussi, William Monte Verde, Natache do Socorro Dias | <i>Fault identification using a chain of decision trees in an electrical submersible pump</i> | Penelitian ini bertujuan untuk mendeteksi dan mengklasifikasikan kegagalan incipient dalam sistem pemompaan. Penelitian lain juga menunjukkan bahwa analitika canggih dapat digunakan untuk | Kelebihan dari penelitian ini adalah menggunakan pendekatan yang inovatif dengan menggunakan rantai pohon keputusan untuk mendeteksi dan mengklasifikasikan kegagalan dalam sistem pemompaan. |

| No | Penulis | Judul | Hasil | Kelebihan dan Kekurangan |
|----|---|--|--|---|
| a | Arrifano Sassim (Barrios Castellanos et al., 2020) | <i>operating in a liquid-gas flow</i> | pengambilan keputusan berbasis data dalam manajemen operasi pompa submersible listrik serta penggunaan algoritma pembelajaran mesin dan analitika prediktif dapat membantu dalam mendeteksi kegagalan incipient dalam sistem pemompaan | Kekurangannya adalah kurangnya pembahasan mengenai pengujian dan validasi algoritma yang diusulkan. |
| 9. | Minzheng Jiang, Tiancai ChengID, Kangxing Dong, Shufan Xu, Yulong Geng (Jiang et al., 2020) | <i>Fault diagnosis method of submersible screw pump based on random forest</i> | Metode diagnosis kesalahan pompa sekrup submersible berdasarkan hutan acak telah diusulkan dalam penelitian ini. Telah diverifikasi bahwa tingkat akurasi diagnosis kesalahan adalah 92,86%. Metode diagnosis kesalahan pompa sekrup submersible berdasarkan hutan acak ini telah dibandingkan | Kelebihan dari penelitian ini adalah berhasil mencapai tingkat akurasi diagnosis kesalahan pompa sekrup submersible sebesar 92,86%, menunjukkan keunggulan dalam mengidentifikasi bentuk kesalahan dengan tepat. Kekurangannya adalah kurangnya pembahasan mengenai faktor-faktor lain yang dapat mempengaruhi akurasi |

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.

| No | Penulis | Judul | Hasil | Kelebihan dan Kekurangan |
|-----|---|---|--|---|
| | | | dengan metode diagnosis lainnya | diagnosis kesalahan pompa sekrup submersible |
| 10. | Amiratul Diyana Amirruddin, Farrah Melissa Muharama, Mohd Hasmadi Ismailb, Mohd Firdaus Ismailc, Ngai Paing Tand, dan Daljit Singh Karamd (Amirruddin et al., 2020) | <i>Hyperspectral remote sensing for assessment of chlorophyll l sufficiency levels in mature oil palm (Elaeis guineensis) based on frond numbers: Analysis of decision tree and random forest</i> | Klasifikasi tentang kandungan klorofil pada kelapa sawit dewasa dapat dibedakan secara efektif menggunakan kelas-kelas yang diusulkan (rendah, sedang, dan tinggi) dan hasil klasifikasi diungguli oleh metode RF dengan akurasi sebesar | Kelebihan penelitian ini adalah membandingkan kinerja klasifikasi menggunakan decision tree dan <i>random forest</i> , dengan menunjukkan bahwa <i>random forest</i> memiliki kinerja yang lebih baik Kekurangannya adalah tidak menyebutkan secara eksplisit keterbatasan-keterbatasan yang ditemui selama penelitian |

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.

| No | Penulis | Judul | Hasil | Kelebihan dan Kekurangan |
|-----|--|---|---|--|
| 11. | Yulia Aryani, Arie Wahyu Wijayanto (Aryani & Wijayanto, 2021) | Klasifikasi Pengemba lian Radar dari Ionosfer Mengguna kan SVM, <i>Naive Bayes</i> dan <i>Random forest</i> | hasil percobaan menunjukkan prediksi dari setiap metode klasifikasi dengan tingkat akurasi dan prediksi yang berbeda- beda. Metode <i>Random forest</i> memberikan hasil akurasi, presisi, dan <i>recall</i> terbaik, dengan tingkat akurasi mencapai 90.57% dan presisi sebesar 94.12%. | kelebihan dari penelitian ini adalah memberikan perbandingan yang komprehensif tentang metode klasifikasi seperti SVM, <i>Naive Bayes</i> , dan <i>Random forest</i> untuk data return radar ionosfer, yang memberikan wawasan tentang keefektifan masing- masing metode. Kelemahan penelitian ini adalah penelitian ini menggunakan metode <i>Split Validation</i> yang sederhana untuk validasi data. |
| 12. | Lei Yafei , Jiang Wanlu, Niu Hongjie, Shi Xiaodong dam Yang Xukang (Yafei et al., 2021) | <i>Fault Diagnosis of Axial Piston Pump Based on Extreme- Point</i> | Membahas tentang diagnosis kesalahan dan pembuatan model untuk pompa piston aksial, mendapatkan akurasi sebesar 97,14% menggunakan metode <i>Random forest</i> . Data yang digunakan dalam penelitian ini tersedia | Kelebihan penelitian ini adalah Metode yang diusulkan telah diuji dan berfungsi dengan tingkat akurasi dan presisi yang tinggi, menunjukkan kehandalan dalam mendiagnosis kesalahan pada pompa piston aksial. |

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.

| No | Penulis | Judul | Hasil | Kelebihan dan Kekurangan |
|------------------------|--|---|---|---|
| a milik UIN Suska Riau | | <i>Symmetric Mode Decomposition and Random forests</i> | dari Situs Web Pusat Data Case Western Reserve University. | Kekurangannya adalah kurang membahas secara rinci mengenai potensi keterbatasan atau tantangan dari metode yang diusulkan |
| 13. | Bakhtawer, Bushra Naz, Sanam Narejo, Naseer U Din, Waqar Ahmed (Bakhtawer, Bushra Naz, Sanam Narejo, Naseer U Din, 2022) | <i>Soil Classification & Prediction of Crop Status with Supervised Learning Algorithm : Random forest</i> | Penelitian ini menghasilkan pengembangan Sistem Manajemen Tanaman (CMS) yang menggunakan algoritma supervised learning, <i>Random forest</i> , untuk klasifikasi tanah dan prediksi status tanaman. | Kelebihan penelitian ini adalah menggunakan algoritme <i>Supervised Learning</i> yaitu <i>Random forest</i> , untuk klasifikasi tanah dan prediksi status tanaman. Para penulis mengklaim akurasi 96,5% untuk metode klasifikasi tanah mereka, yang menunjukkan keefektifan sistem yang dikembangkan. Kelemahan penelitian ini adalah ketergantungan sistem pada berbagai sensor untuk pengumpulan data dapat menimbulkan masalah potensial terkait akurasi, pemeliharaan, dan kalibrasi sensor, yang |

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.

| No | Penulis | Judul | Hasil | Kelebihan dan Kekurangan |
|------------------------|-------------------------------------|---|--|--|
| a milik UIN Suska Riau | | | | dapat berdampak pada efektivitas sistem secara keseluruhan. |
| 14. | Lestari Handayani (Handayani, 2022) | <i>Predictive Maintenance with Hidden Semi-Markov Model</i> | Studi ini menerapkan <i>Hidden Semi-Markov Model</i> (HSMM) dan <i>Hidden Markov Model</i> (HMM) pada <i>Prognostics and Health Management</i> (PHM) dengan menggunakan dataset sensor pompa air untuk memprediksi kegagalan. Hasil penelitian menunjukkan bahwa metode HSMM-FB berkinerja lebih baik dalam mendeteksi kesalahan dan memberikan durasi alarm dibandingkan dengan metode HMM dan HSMM-Vit-FB. Studi ini juga menyoroti pentingnya menggunakan data pelatihan yang tepat | Kelebihan Dari penelitian ini adalah menyajikan beberapa keuntungan dari pendekatan yang diusulkan untuk diagnosis dan prognosis mesin pompa air. Pendekatan ini menggunakan <i>Hidden Semi-Markov Model</i> (HSMM) dan <i>Hidden Markov Model</i> (HMM) untuk memperkirakan kesehatan suatu sistem dengan menggunakan metode <i>Markov</i> dalam bentuk multi-pengamatan Kekurangan dari penelitian ini adalah , dataset yang digunakan dalam penelitian ini mengandung nilai yang hilang, dan pendekatan untuk menangani data yang hilang tidak dibahas |

| No | Penulis | Judul | Hasil | Kelebihan dan Kekurangan |
|-----|--|---|---|--|
| | | | untuk hasil decoding yang akurat. | secara ekstensif. Meskipun penelitian ini menyebutkan perlunya mengisi nilai yang hilang dan menangani <i>outlier</i> |
| 15. | Fahrezal Zubedi, Bagus Sartono, Khairil Anwar Notodiputro (Zubedi et al., 2022) | <i>Implementation of winsorizing and random oversampling on data containing outliers and unbalanced data with the random forest classification method</i> | Penelitian ini bertujuan untuk membandingkan kinerja hutan acak sebelum dan setelah data diterapkan pada teknik winsorizing dan <i>random oversampling</i> untuk memprediksi IDH di kabupaten/kota di Indonesia bagian timur. Hasil klasifikasi hutan acak setelah penanganan data yang mengandung <i>outlier</i> dan data yang tidak seimbang memiliki kinerja yang lebih baik dalam hal akurasi dan nilai kappa, yaitu 96,43% dan 93,41%. | Kelebihan penelitian ini adalah penggunaan teknik winsorizing dan <i>random oversampling</i> untuk menangani data yang mengandung <i>outlier</i> dan tidak seimbang, sehingga meningkatkan kinerja metode klasifikasi hutan acak. Kekurangannya adalah tidak secara eksplisit membahas potensi keterbatasan atau tantangan yang dihadapi selama implementasi teknik winsorizing dan <i>random oversampling</i> . |

Hak Cipta Dilindungi Undang-Undang

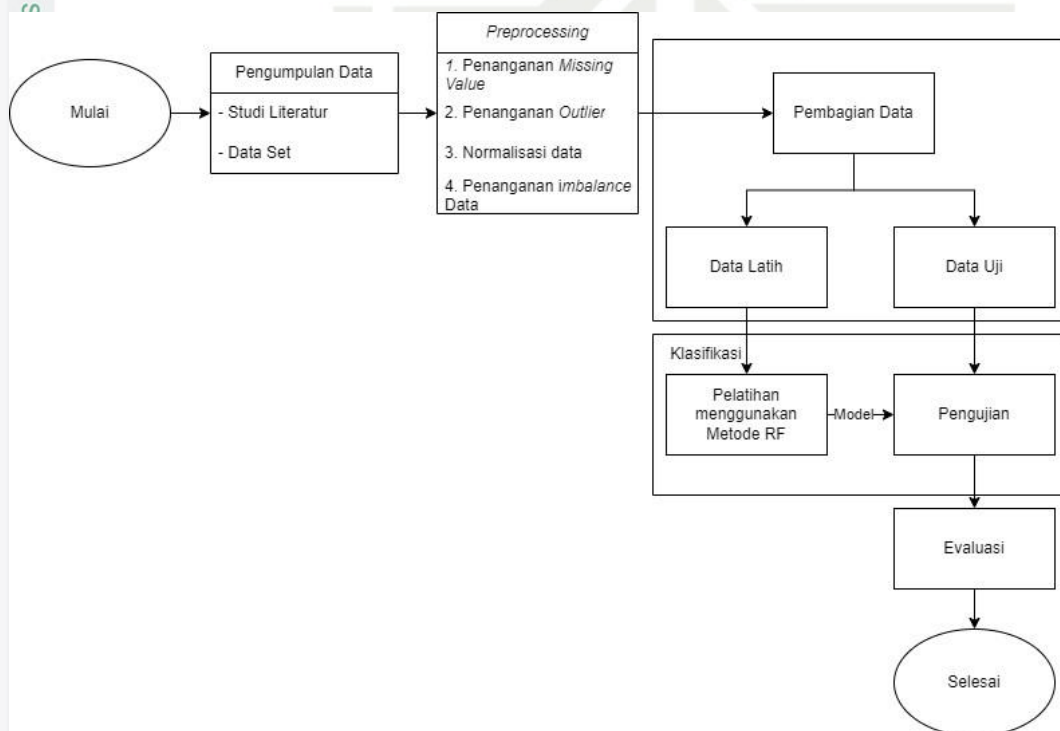
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah.
 - b. Pengutipan tidak merugikan kepentingan 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. 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 3. 1 Tahapan Medodologi 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

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.

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 Data set

Data yang digunakan pada penelitian ini adalah data sekunder yang bersumber dari web Kaggle yaitu *dataset pump sensor data* tahun 2018 yang bisa ditemukan melalui alamat web <https://www.kaggle.com/datasets/nphantawee/pump-sensor-data>. 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*. Data set yang digunakan pada penelitian ini merujuk pada penelitian sebelumnya tentang prediksi sebelum kerusakan terhadap kondisi mesin pompa air menggunakan data set *pump sensor* (Handayani, 2022).

3.2 Pre-processing Data

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

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. Untuk data yang terdapat beberapa *missing value*, lakukan teknik imputasi dmenggunakan persamaan (1) yaitu mengganti nilai yang kosong dengan rata-rata.

3.2.2 Penanganan *Outlier*

Dalam penelitian ini data yang mengandung *outlier* akan diganti dengan nilai mean dari setiap kolom yang terdapat *outlier* jika rentang nilainya jauh berbeda dibandingkan nilai data lainnya. Pengecekan *outlier* pada penelitian ini menggunakan metode IQR (*Interquartile Range*) dengan tahapan sebagai berikut:

1. Menentukan data yang akan diatasi *outlier*
2. Menentukan nilai Q_3 (kuartil 3) dengan persamaan (2) dan nilai Q_1 (kuartil 1) dengan persamaan (3)
3. Menghitung nilai IQR (*Interquartile Range*) dengan persamaan (4)
4. Selanjutnya menghitung nilai batas atas dengan persamaan (5) dan batas bawah dengan persamaan (6)
5. Langkah terakhir yaitu membandingkan nilai data dengan nilai batas atas dan bawah untuk mengidentifikasi outlier, lalu mengganti nilai *outlier* dengan mean.

3.2.3 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-Szore dengan persamaan (8).

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.

3.2.4 Penanganan *imbalance data*

Dalam penelitian ini akan dilakukan penanganan terhadap data yang *imbalance* menggunakan teknik *over sampling* dengan metode SMOTE yang berguna untuk meningkatkan hasil klasifikasi yang lebih bagus.

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 3 perbandingan yaitu 90:10, 80:20, 70:30. Persentase data pelatihan lebih besar daripada data uji coba. Data pengujian yang dapat digunakan dari sepuluh persen hingga tiga puluh persen dari data sebelumnya.

3.3.1 Data Latih

Pada penelitian ini, data pelatihan digunakan dengan komposisi 90%, 80%, dan 70%. Pembuatan data pelatihan dilakukan dengan menggunakan komposisi yang telah ditetapkan dari jumlah total data. Sehingga data pelatihan yang diperoleh digunakan untuk pemodelan klasifikasi.

3.3.2 Data Uji

Pada penelitian ini, data pengujian merupakan sisa dari data pelatihan dengan komposisi 10%, 20%, dan 30%. Hasil data pengujian digunakan untuk menentukan pemodelan klasifikasi yang terbaik dengan tingkat akurasi.

3.4 Pelatihan

Pada tahap pelatihan ini akan menggunakan algoritma *Random forest* dalam melakukan klasifikasi status mesin pompa air terhadap data yang telah dibersihkan. Berikut tahapan-tahapan dalam melakukan pelatihan dengan klasifikasi dengan *random forest* adalah sebagai berikut:

1. Buat suatu bootstrap sample atau pengambilan sampel dengan replacement (pengembalian) dari suatu ukuran dari gugus data.
2. Pilih m ($m \leq p$) variabel secara random dari p variable $m \leq p$, menggunakan persamaan (15), (16), dan (17).

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. Setelah dilakukan pemilihan m secara random, maka pohon ditumbuhkan tanpa pruning (pemangkasan).
4. Langkah 1-3 dilakukan sebanyak kali hingga terbentuk suatu forest sebanyak pohon.
5. Proses penentuan suatu kelas dilakukan dengan *majority vote*.
6. Setelah diperoleh nilai mtry optimal kemudian dilakukan prediksi dengan data testing

3.5 Pengujian

Pada tahap ini hasil model dari pelatihan akan melakukan pengujian, pada penelitian ini ada beberapa bentuk pengujian yang dilakukan terhadap model yang dibentuk sebagai berikut :

1. Pengujian terhadap data sebelum dan sesudah diatasi *imbalance* menggunakan komposisi perbandingan 90:10, 80:20, dan 70:30. Metode *imbalance* data yang digunakan adalah SMOTE. Pengujian ini dilakukan terhadap data yang telah ditangani *outlier* dan tanpa penanganan *outlier*.
2. Pengujian selanjutnya yaitu terhadap data sebelum dan sesudah diatasi *imbalance* dengan kombinasi parameter Metode RF yaitu *n_estimators*, *max_dept*, *min_samples_split*, *min_samples_leaf*. Pada pengujian ini terdapat data yang telah ditangani *outluer* dan tanpa di tangani *outlier*.

3.5.1 Skenario Pengujian

Skenario pengujian pada penelitian ini dilakukan sebanyak dua kali, pengujian pertama dilakukan terhadap data *imbalance* dan data *balance* dengan presentase 100%, 25%, dan 10%. Pengujian tersebut dilakukan terhadap data yang telah ditangani *outlier*. Pengujian kedua dilakukan terhadap data tanpa penanganan *outlier* pada data *imbalance* dan *balance* dengan presentase data 100%, 25%, dan 10%. Kedua pengujian menggunakan kombinasi parameter dari metode RF yaitu *n_estimator*, *max_depth*, *min_samples_split*, dan *min_samples_leaf*. Dengan rasio perbandingan 90:10, 80:20, dan 70:30.

Berikut pada tabel 3.1 adalah tabel skenario dari pengujian yang akan dilakukan dipenelitian ini.

Tabel 3. 1 skenario pengujian

| Rasio Perbandingan | $N_{estimator}$ r | Max_dept h | $Min_samples_split$ it | $Min_samples_level$ af |
|--------------------|------------------------|--------------------|-------------------------------|-------------------------------|
| 90:10 | 50 | 10 | 2 | 5 |
| | | | 5 | 10 |
| | | | 5 | 5 |
| | | 20 | 2 | 10 |
| | | | 5 | 5 |
| | | | 5 | 10 |
| | 100 | 10 | s2 | 5 |
| | | | 5 | 10 |
| | | | 5 | 5 |
| | | 20 | 2 | 10 |
| | | | 5 | 5 |
| | | | 5 | 10 |
| 80:20 | 50 | 10 | 2 | 5 |
| | | | 5 | 10 |
| | | 20 | 2 | 5 |
| | 5 | | 10 | |
| | 5 | | 5 | |

- 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.

| Rasio Perbandingan | $N_{estimator}$ r | Max_depth h | $Min_samples_split$ it | $Min_samples_leaf$ af | |
|----------------------|------------------------|---------------------|-------------------------------|------------------------------|----|
| milik UIN Suska Riau | 100 | 10 | 2 | 5 | |
| | | | | 10 | |
| | | | 5 | 5 | |
| | | | 10 | | |
| | | 20 | 2 | 5 | |
| | | | | 10 | |
| | 5 | | 5 | | |
| | | | | 10 | |
| | 70:30 | 50 | 10 | 2 | 5 |
| | | | | | 10 |
| | | | | 5 | 5 |
| | | | | 10 | |
| 20 | | | 2 | 5 | |
| | | | | 10 | |
| | | 5 | 5 | | |
| | | | | 10 | |
| 100 | | 10 | 2 | 5 | |
| | | | | 10 | |
| | | | 5 | 5 | |
| | | | 10 | | |
| | 20 | 2 | 5 | | |
| | | | 10 | | |
| 5 | | 5 | | | |
| | | | 10 | | |

- Hak Cipta Dilindungi Undang-Undang**
- 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.
 - 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.

3.6 Evaluasi

Pada tahap ini dilakukan evaluasi untuk mengetahui performa algoritma yang telah digunakan dalam penelitian ini yaitu *Random forest*. Pada tahap evaluasi ini menggunakan *confusion matrix*. Dimana dalam *confusion matrix* menghitung nilai *accuracy* menggunakan persamaan (18), *precision* menggunakan persamaan (19), *recall* menggunakan persamaan (20), dan *f1-score* menggunakan persamaan (21).

Accuracy berfungsi untuk persentase prediksi yang benar yang dilakukan pada klasifikasi status mesin pompa air dalam 3 kelas status mesin yaitu normal, *recoervering*, dan *broken*. *Precision* dalam dataset *pump sensor* berfungsi untuk menentukan ratio antara True Positive dan Predicted Positive mesin pompa air dalam proses klasifikasi. *Recall* digunakan untuk menentukan rasio antara True Positives dan Actual Positives pada data mesin pompa air. Sedangkan *F1-score* digunakan untuk mendapatkan hasil rata-rata dari *precision* dan *recall* pada dataset *pump sensor*. Berikut pada tabel 3.2 adalah tabel *confusion matrix multiclass* :

Tabel 3. 2 confusion matrix multiclass

| <i>Multiclass Confusion matrix</i> | | Prediksi | | |
|------------------------------------|---------------------|----------|---------------------|---------------|
| | | Normal | <i>Recoervering</i> | <i>Broken</i> |
| Aktual | Normal | TP | FN | FN |
| | <i>Recoervering</i> | FP | TN | TN |
| | <i>Broken</i> | FP | TN | TN |

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. Metode *Random forest* efektif digunakan dalam proses klasifikasi multicalss dibuktikan pada penelitian ini menggunakan data *pump sensor* mendapatkan akurasi rata-rata sebesar 99,97% pada klasifikasi menggunakan data tanpa penanganan *outlier* dan 80,29% menggunakan data dengan penanganan *outlier* yang mengklasifikasi mesin ke dalam tiga status mesin yaitu normal, *recovering*, dan *broken*.
2. Performa metode *Random forest* dalam klasifikasi status mesin pompa air menggunakan data *imbalance* mendapatkan rata-rata precision sebesar 0.00% dimana sistem belum bisa mengklasifikasikan data ke tiga status mesin pompa air. Sedangkan menggunakan data *balance* 10% mendapatkan rata-rata precision sebesar 49%, data *balance* 25% sebanyak 50%, dan data *balance* 100% sebesar 68% dimana sistem mampu mengklasifikasikan data kedalam tiga status mesin yaitu normal, *recovering*, dan *broken*.
3. Dalam proses klasifikasi pada penelitian ini mendapatkan akurasi terbaik saat menggunakan data tanpa penanganan *outlier* dibandingkan dengan data yang telah ditangani *outlier*
4. Pada penelitian ini menggunakan dataset *pump_sensor* semakin besar presentase data *balance* yang digunakan pada data yang telah ditangani *outlier* maka semakin menurun akurasi yang didapatkan, sebaliknya semakin besar presentase data *balance* tanpa penanganan *outlier* semakin naik akurasi yang didapatkan pada proses klasifikasi.

5.2 Saran

Pada penelitian ini metode yang digunakan untuk penanganan *outlier* adalah IQR, terdapat metode lainnya dalam penanganan *outlier* sehingga disarankan untuk menggunakan metode penanganan *outlier* yang lainnya. Selain itu, nilai *outlier* yang terdeteksi pada penelitian ini diganti menggunakan nilai mean, teknik tersebut kurang cocok untuk digunakan pada data penelitian ini disarankan menggunakan teknik 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.
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 PUSTAKA

- Agustiani, S., Tajul Arifin, Y., Junaidi, A., Khotimatul Wildah, S., & Mustopa, A. (2022). Klasifikasi Penyakit Daun Padi menggunakan Random Forest dan Color Histogram. *Jurnal Komputasi*, 10(1). <https://doi.org/10.23960/komputasi.v10i1.2961>
- Ahmad, Z., Rai, A., Maliuk, A. S., & Kim, J. M. (2020). Discriminant feature extraction for centrifugal pump fault diagnosis. *IEEE Access*, 8, 165512–165528. <https://doi.org/10.1109/ACCESS.2020.3022770>
- Akromunnisa, K., Hidayat, R., Informatika, J. T., & Adisucipto, J. L. (2019). K-NEAREST NEIGHBOR. 4(1), 69–75.
- Aldean, M. Y., Paradise, P., & Setya Nugraha, N. A. (2022). Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 di Twitter Menggunakan Metode Random Forest Classifier (Studi Kasus: Vaksin Sinovac). *Journal of Informatics, Information System, Software Engineering and Applications (INISTA)*, 4(2), 64–72. <https://doi.org/10.20895/inista.v4i2.575>
- Amirruddin, A. D., Muharam, F. M., Ismail, M. H., Ismail, M. F., Tan, N. P., & Karam, D. S. (2020). Hyperspectral remote sensing for assessment of chlorophyll sufficiency levels in mature oil palm (*Elaeis guineensis*) based on frond numbers: Analysis of decision tree and random forest. *Computers and Electronics in Agriculture*, 169(December 2019), 105221. <https://doi.org/10.1016/j.compag.2020.105221>
- Arifin, Z., Tamamy, A. J., & Islahu, N. (2020). Perancangan Mesin Pompa Air Tenaga Surya untuk Mengurangi Konsumsi Listrik Skala Rumah. *Jurnal Nasional Teknik Elektro*, 9(2), 79. <https://doi.org/10.25077/jnte.v9n2.758.2020>
- Arifiyanti, A. A., & Wahyuni, E. D. (2020). SMOTE : METODE PENYEIMBANG KELAS PADA KLASIFIKASI DATA MINING. XV, 34–39.
- Aryani, Y., & Wijayanto, A. W. (2021). Klasifikasi Pengembalian Radar dari

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.

Ionosfer Menggunakan SVM , Naïve Bayes dan Random Forest Classification of Radar Returns from the Ionosphere Using SVM , Naïve Bayes and Random Forest. 10(28). <https://doi.org/10.34010/komputika.v10i2.4347>

Bakhtawer, Bushra Naz, Sanam Narejo, Naseer U Din, W. A. (2022). Soil Classification & Prediction of Crop Status with Supervised Learning Algorithm: Random Forest. *International Journal of Innovations in Science and Technology*, 4(4), 1011–1022. <https://doi.org/10.33411/ijist/2022040403>

Barrios Castellanos, M., Serpa, A. L., Biazussi, J. L., Monte Verde, W., & do Socorro Dias Arrifano Sassim, N. (2020). Fault identification using a chain of decision trees in an electrical submersible pump operating in a liquid-gas flow. *Journal of Petroleum Science and Engineering*, 184(February 2019), 106490. <https://doi.org/10.1016/j.petrol.2019.106490>

Bilal, M., Ali, G., Iqbal, M. W., Anwar, M., Malik, M. S. A., & Kadir, R. A. (2022). Auto-Prep: Efficient and Automated Data Preprocessing Pipeline. *IEEE Access*, 10(June), 107764–107784. <https://doi.org/10.1109/ACCESS.2022.3198662>

Churn, C., & Kombinasi, M. (2021). *Penanganan Ketidakseimbangan Data pada Prediksi*. 6(1), 31–37.

Devella, S., Yohannes, Y., & Rahmawati, F. N. (2020). Implementasi Random Forest Untuk Klasifikasi Motif Songket Palembang Berdasarkan SIFT. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 7(2), 310–320. <https://doi.org/10.35957/jatisi.v7i2.289>

Dewi, A. O. P. (2020). Big Data di Perpustakaan dengan Memanfaatkan Data Mining. *Anuva: Jurnal Kajian Budaya, Perpustakaan, Dan Informasi*, 4(2), 223–230. <https://doi.org/10.14710/anuva.4.2.223-230>

Dinata, N. A. S., Abdurrahman, G., & Fitriyah, N. Q. (2023). Perbandingan Optimasi Algoritma Random Forest Menggunakan Teknik Boosting Terhadap Kasus Klasifikasi Churn Pelanggan Di Industri Telekomunikasi. *Jurnal*

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.

Aplikasi Sistem Informasi Dan Elektronika, 5(1), 28–37.

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>

Gde Agung Brahmana Suryanegara, Adiwijaya, & Mahendra Dwifebri Purbolaksono. (2021). Peningkatan Hasil Klasifikasi pada Algoritma Random Forest untuk Deteksi Pasien Penderita Diabetes Menggunakan Metode Normalisasi. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(1), 114–122. <https://doi.org/10.29207/resti.v5i1.2880>

Handayani, L. (2022). *Predictive Maintenance with Hidden Semi-Markov Model*.

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/ijiis.v4i1.73>

He, Q., Si, X. S., Zhang, Q. H., & Qin, A. S. (2020). A rotating machinery fault diagnosis method based on multi-scale dimensionless indicators and random forests. *Mechanical Systems and Signal Processing*, 139, 106609. <https://doi.org/10.1016/j.ymssp.2019.106609>

Iman, Q., & Wahyu, A. (2021). *Klasifikasi Rumah Tangga Penerima Beras Miskin (Raskin)/ Beras Sejahtera (Rastra) di Provinsi Jawa Barat Tahun 2017 dengan Metode Random Forest dan Support Vector Machine Classification of Poor Rice (Raskin)/ Prosperous Rice (Rastra) Recipient Hou*. 09(2), 178–184. <https://doi.org/10.26418/justin.v9i2.44137>

Irianto, M. R., Maududie, A., & Arifin, F. N. (2022). Implementation of K-Means Clustering Method for Trend Analysis of Thesis Topics (Case Study: Faculty of Computer Science, University of Jember). *Berkala Sainstek*, 10(4), 210.



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.

<https://doi.org/10.19184/bst.v10i4.29524>

- Jiang, M., Cheng, T., Dong, K., Xu, S., & Geng, Y. (2020). Fault diagnosis method of submersible screw pump based on random forest. *PLoS ONE*, *15*(11 November). <https://doi.org/10.1371/journal.pone.0242458>
- Khairuddin Tampubolon, Alinur, Elazhari, Ardi Ermawy, R. S. M. (2021). Penyuluhan Tentang Mengenal Mesin Pompa Air dan Cara Perawatannya di Serikat Tolong Menolong Nurul Iman (STMNI) Kelurahan Timbang Deli Kecamatan Medan Amplas. *Jurnal PKM*, *1*(2), 1–8.
- Lestari, T. S., & Sirodj, D. A. N. (2022). Klasifikasi Penipuan Transaksi Kartu Kredit Menggunakan Metode Random Forest. *Jurnal Riset Statistika*, *1*(2), 160–167. <https://doi.org/10.29313/jrs.v1i2.525>
- Martias, M. F., Jasril, J., Sanjaya, S., Handayani, L., & Yanto, F. (2023). Klasifikasi Citra Daging Sapi dan Daging Babi Menggunakan CNN Arsitektur EfficientNet-B6 dan Augmentasi Data. *Jurnal Sistem Komputer Dan Informatika (JSON)*, *4*(4), 642. <https://doi.org/10.30865/json.v4i4.6195>
- 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*. *9*(2), 133–138. <https://doi.org/10.15294/sji.v9i2.35149>
- Nguyen, J. M., Jézéquel, P., Gillois, P., Silva, L., Azzouz, F. Ben, Lambert-Lacroix, S., Juin, P., Campone, M., Gaultier, A., Moreau-Gaudry, A., & Antonioli, D. (2021). Random forest of perfect trees: concept, performance, applications and perspectives. *Bioinformatics*, *37*(15), 2165–2174. <https://doi.org/10.1093/bioinformatics/btab074>

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.

- Noshad, Z., Javaid, N., Saba, T., Wadud, Z., Saleem, M. Q., Alzahrani, M. E., & Sheta, O. E. (2019). Fault detection in wireless sensor networks through the random forest classifier. *Sensors (Switzerland)*, *19*(7), 1–21. <https://doi.org/10.3390/s19071568>
- Nur, M. (2022). Data Mining Untuk Memprediksi Kelulusan Mahasiswa Jurusan Teknik Informatika UIN Syarif Hidayatullah Jakarta Menggunakan Metode Klasifikasi C4. 5. *Repository.Uinjkt.Ac.Id*. [https://repository.uinjkt.ac.id/dspace/handle/123456789/65006%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/65006/1/MUHAMMAD NUR-FST.pdf](https://repository.uinjkt.ac.id/dspace/handle/123456789/65006%0Ahttps://repository.uinjkt.ac.id/dspace/bitstream/123456789/65006/1/MUHAMMAD%20NUR-FST.pdf)
- Perdana, A. Y., Latuconsina, R., & Dinimaharawati, A. (2021). Prediksi Stunting Pada Balita Dengan Algoritma Random Forest. *E-Proceeding of Engineering*, *8*(5), 6650–6656.
- Prabiantissa Citra Nurina. (2021). 1818-5505-1-Pb. *Seminar Nasional Teknik Elektro, Sistem Informasi, Dan Teknik Informatika*, *1*(1), 263–268.
- Pradana, R., & Astika, I. B. P. (2019). Pengaruh Ukuran Perusahaan, Penerapan Good Corporate Governance, dan Pengungkapan Corporate Social Responsibility pada Nilai Perusahaan. *E-Jurnal Akuntansi*, *28*(3), 1920. <https://doi.org/10.24843/eja.2019.v28.i03.p18>
- Prasetya, J. (2022). Penerapan Klasifikasi Naive Bayes dengan Algoritma Random Oversampling dan Random Undersampling pada Data Tidak Seimbang Cervical Cancer Risk Factors. *Leibniz: Jurnal Matematika*, *2*(2), 11–22. <https://doi.org/10.59632/leibniz.v2i2.173>
- Riska Chairunisa, Adiwijaya, & Widi Astuti. (2020). Perbandingan CART dan Random Forest untuk Deteksi Kanker berbasis Klasifikasi Data Microarray. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *4*(5), 805–812. <https://doi.org/10.29207/resti.v4i5.2083>
- Saputro, K. E., Fadli, S., Saputro, K. E., & Fadli, S. (2020). *K-means-SMOTE untuk*

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.

menangani ketidakseimbangan kelas dalam klasifikasi penyakit diabetes dengan C4.5, SVM, dan naive Bayes K-means-SMOTE for handling class imbalance in the classification of diabetes. 8(November 2019), 89–93. <https://doi.org/10.14710/jtsiskom.8.2.2020.89-93>

Sari, P. P., Herdiani, E. T., & Sunusi, N. (2021). Outlier Detection Using Minimum Vector Variance Algorithm with Depth Function and Mahalanobis Distance. *Jurnal Matematika, Statistika Dan Komputasi*, 17(3), 418–427. <https://doi.org/10.20956/j.v17i3.12629>

Savargiv, M., Masoumi, B., & Keyvanpour, M. R. (2021). A new random forest algorithm based on learning automata. *Computational Intelligence and Neuroscience*, 2021. <https://doi.org/10.1155/2021/5572781>

Setio, P. B. N., Saputro, D. R. S., & Bowo Winarno. (2020). Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika*, 3, 64–71.

Sharaf, S. A., Bangert, P., Fardan, M., Alqassab, K., Abubakr, M., & Ahmed, M. (2019). Beam pump dynamometer card classification using machine learning. *SPE Middle East Oil and Gas Show and Conference, MEOS, Proceedings, 2019-March*. <https://doi.org/10.2118/194949-ms>

Siti Aisyah, Sri Wahyuningsih, F. D. T. A. (2021). Peramalan Jumlah Titik Panas Provinsi Kalimantan Timur Menggunakan Metode Radial Basis Function Neural Network. *Jambura Journal of Probability and Statistics*, 2(2), 64–74. <https://doi.org/10.34312/jjps.v2i2.10292>

Sun, D., Wen, H., Wang, D., & Xu, J. (2020). Journal Pre-proof. *Geomorphology*, 107201. <https://doi.org/10.1016/j.geomorph.2020.107201>

Syukron, M., Santoso, R., & Widiharih, T. (2020). Perbandingan Metode Smote Random Forest Dan Smote Xgboost Untuk Klasifikasi Tingkat Penyakit Hepatitis C Pada Imbalance Class Data. *Jurnal Gaussian*, 9(3), 227–236. <https://doi.org/10.14710/j.gauss.v9i3.28915>

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.

- 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>
- Vuttipittayamongkol, P., Elyan, E., & Petrovski, A. (2021). On the class overlap problem in *imbalanced* data classification. *Knowledge-Based Systems*, 212, 106631. <https://doi.org/10.1016/j.knosys.2020.106631>
- Wang, Y., & Liu, H. (2019). Centrifugal pump fault diagnosis based on MEEMD-PE Time-frequency information entropy and Random forest. *Proceedings of 2019 11th CAA Symposium on Fault Detection, Supervision, and Safety for Technical Processes, SAFEPROCESS 2019*, 932–937. <https://doi.org/10.1109/SAFEPROCESS45799.2019.9213261>
- Widiastiwi, Y., & Ernawati, I. (2021). Klasifikasi Penyakit Batu Ginjal Menggunakan Algoritma Decision Tree C4 . 5 Dengan Membandingkan Hasil Uji Akurasi. *Jurnal IKRA-ITH INFORMATIKA*, 5(2), 128.
- Yafei, L., Wanlu, J., Hongjie, N., Xiaodong, S., & Xukang, Y. (2021). Fault Diagnosis of Axial Piston Pump Based on Extreme-Point Symmetric Mode Decomposition and Random Forests. *Shock and Vibration*, 2021. <https://doi.org/10.1155/2021/6649603>
- Yoga Religia, Agung Nugroho, & Wahyu Hadikristanto. (2021). Klasifikasi Analisis Perbandingan Algoritma Optimasi pada Random Forest untuk Klasifikasi Data Bank Marketing. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(1), 187–192. <https://doi.org/10.29207/resti.v5i1.2813>
- Yuli Mardi. (2019). Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . *Jurnal Edik Informatika. Jurnal Edik Informatika*, 2(2), 213–219.

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.

Zailani, A. U., & Hanun, N. L. (2020). Penerapan Algoritma Klasifikasi Random Forest Untuk Penentuan Kelayakan Pemberian Kredit Di Koperasi Mitra Sejahtera. *Infotech: Journal of Technology Information*, 6(1), 7–14. <https://doi.org/10.37365/jti.v6i1.61>

Zhang, H., Chen, H., Guo, Y., Wang, J., Li, G., & Shen, L. (2019). Sensor fault detection and diagnosis for a water source heat pump air-conditioning system based on PCA and preprocessed by combined clustering. *Applied Thermal Engineering*, 160(October 2017), 114098. <https://doi.org/10.1016/j.applthermaleng.2019.114098>

Zubedi, F., Sartono, B., & Notodiputro, K. A. (2022). Implementation of Winsorizing and random oversampling on data containing *outliers* and *unbalanced* data with the random forest classification method. *Jurnal Natural*, 22(2), 108–116. <https://doi.org/10.24815/jn.v22i2.25499>

LAMPIRAN

Lampiran I Hasil pengujian data *imbalance* dengan penanganan *outlier*

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_Samples_split</i> | <i>Min_Samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>Recall</i> | <i>F1-score</i> |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|
| 90:10 | 50 | 10 | 2 | 5 | 93,55% | 93,55% | 100% | 96,67% |
| | | | | 10 | 93,55% | 93,55% | 100% | 96,67% |
| | | | 5 | 5 | 93,55% | 93,55% | 100% | 96,67% |
| | | | | 10 | 93,55% | 93,55% | 100% | 96,67% |
| | | 20 | 2 | 5 | 93,55% | 93,55% | 100% | 96,67% |
| | | | | 10 | 93,55% | 93,55% | 100% | 96,67% |
| | | | 5 | 5 | 93,55% | 93,55% | 100% | 96,67% |
| | | | | 10 | 93,55% | 93,55% | 100% | 96,67% |
| | 100 | 10 | 2 | 5 | 93,55% | 93,55% | 100% | 96,67% |
| | | | | 10 | 93,55% | 93,55% | 100% | 96,67% |
| | | | 5 | 5 | 93,55% | 93,55% | 100% | 96,67% |
| | | | | 10 | 93,55% | 93,55% | 100% | 96,67% |
| | | 20 | 2 | 5 | 93,55% | 93,55% | 100% | 96,67% |

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_Samples_split</i> | <i>Min_Samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>Recall</i> | <i>F1-score</i> |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|
| | Suska Riau | | 5 | 10 | 93,55% | 93,55% | 100% | 96,67% |
| | | | | 5 | 93,55% | 93,55% | 100% | 96,67% |
| | | | | 10 | 93,55% | 93,55% | 100% | 96,67% |
| 80:20 | 50 | 10 | 2 | 5 | 93,60% | 93,59% | 100% | 96,69% |
| | | | | 10 | 93,60% | 93,59% | 100% | 96,69% |
| | | | | 5 | 93,60% | 93,59% | 100% | 96,69% |
| | | 20 | 2 | 5 | 93,60% | 93,59% | 100% | 96,69% |
| | | | | 10 | 93,60% | 93,59% | 100% | 96,69% |
| | | | | 5 | 93,60% | 93,59% | 100% | 96,69% |
| | 100 | 10 | 2 | 5 | 93,60% | 93,59% | 100% | 96,69% |
| | | | | 10 | 93,60% | 93,59% | 100% | 96,69% |
| | | | | 5 | 93,60% | 93,59% | 100% | 96,69% |
| | | 20 | 2 | 5 | 93,60% | 93,59% | 100% | 96,69% |
| | | | | 10 | 93,60% | 93,59% | 100% | 96,69% |
| | | | | 5 | 93,60% | 93,59% | 100% | 96,69% |

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_Samples_split</i> | <i>Min_Samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>Recall</i> | <i>F1-score</i> |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|
| | Suska Riau | | 5 | 10 | 93,60% | 93,59% | 100% | 96,69% |
| | | | | 5 | 93,60% | 93,59% | 100% | 96,69% |
| | | | | 10 | 93,60% | 93,59% | 100% | 96,69% |
| 70:30 | 50 | 10 | 2 | 5 | 93,51% | 93,50% | 100% | 96,64% |
| | | | | 10 | 93,51% | 93,50% | 100% | 96,64% |
| | | | | 5 | 93,51% | 93,50% | 100% | 96,64% |
| | | 20 | 2 | 5 | 93,51% | 93,50% | 100% | 96,64% |
| | | | | 10 | 93,51% | 93,50% | 100% | 96,64% |
| | | | | 5 | 93,51% | 93,50% | 100% | 96,64% |
| | 100 | 10 | 2 | 5 | 93,51% | 93,50% | 100% | 96,64% |
| | | | | 10 | 93,51% | 93,50% | 100% | 96,64% |
| | | | | 5 | 93,51% | 93,50% | 100% | 96,64% |
| | | 20 | 2 | 5 | 93,51% | 93,50% | 100% | 96,64% |
| | | | | 10 | 93,51% | 93,50% | 100% | 96,64% |
| | | | | 5 | 93,51% | 93,50% | 100% | 96,64% |

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_Samples_split</i> | <i>Min_Samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>Recall</i> | <i>F1-score</i> |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|
| | | | 5 | 10 | 93,51% | 93,50% | 100% | 96,64% |
| | | | | 5 | 93,51% | 93,50% | 100% | 96,64% |
| | | | | 10 | 93,51% | 93,50% | 100% | 96,64% |

Seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber: tingkatan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan penting yang wajar UIN Suska Riau. memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.



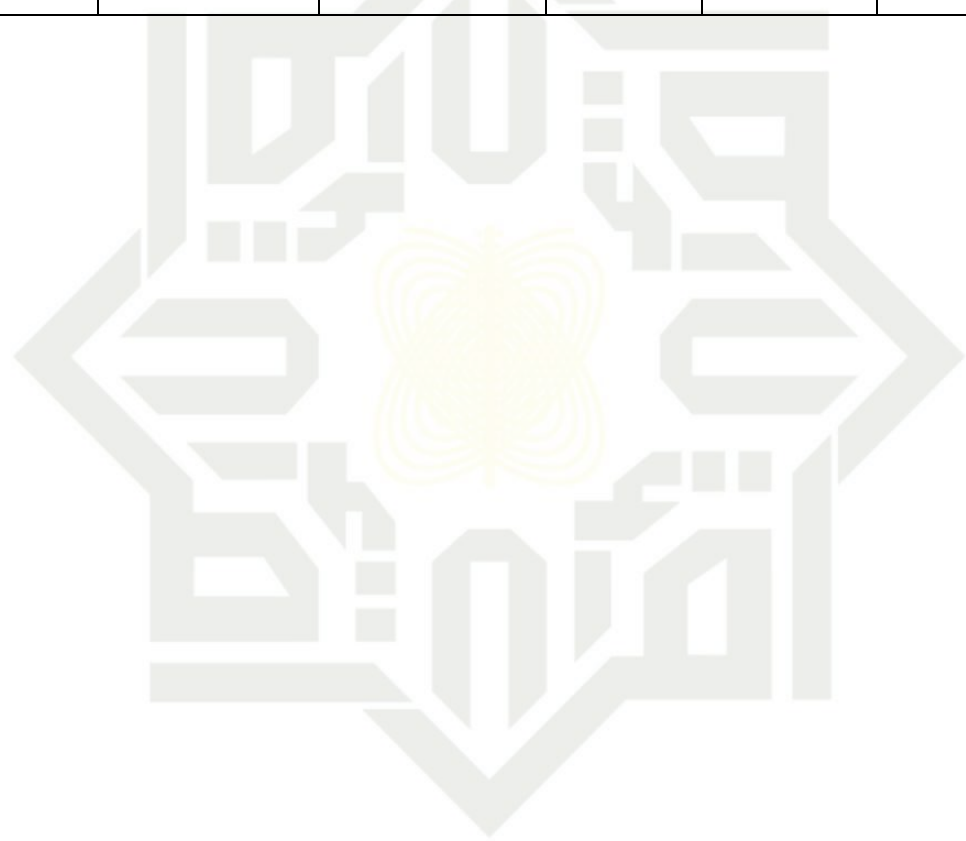
Lampiran I Hasil pengujian 100% data *balance* dengan penanganan *outlier*

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_Samples_split</i> | <i>Min_Samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>Recall</i> | <i>F1-score</i> |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|
| 90:10 | 50 | 10 | 2 | 5 | 42,8% | 45,27% | 42,8% | 32,67% |
| | | | | 10 | 42,8% | 45,27% | 42,8% | 32,67% |
| | | | 5 | 5 | 42,8% | 45,27% | 42,8% | 32,67% |
| | | | | 10 | 42,8% | 45,27% | 42,8% | 32,67% |
| | | 20 | 2 | 5 | 42,8% | 45,27% | 42,8% | 32,67% |
| | | | | 10 | 42,8% | 45,27% | 42,8% | 32,67% |
| | 5 | | 5 | 42,8% | 45,27% | 42,8% | 32,67% | |
| | | | 10 | 42,8% | 45,27% | 42,8% | 32,67% | |
| | 100 | 10 | 2 | 5 | 42,8% | 45,27% | 42,8% | 32,67% |
| | | | | 10 | 42,8% | 45,27% | 42,8% | 32,67% |
| | | | 5 | 5 | 42,8% | 45,27% | 42,8% | 32,67% |
| | | | | 10 | 42,8% | 45,27% | 42,8% | 32,67% |
| 20 | | 2 | 5 | 42,8% | 45,27% | 42,8% | 32,67% | |
| | | | 10 | 42,8% | 45,27% | 42,8% | 32,67% | |
| | | 5 | 5 | 42,8% | 45,27% | 42,8% | 32,67% | |

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_Samples_split</i> | <i>Min_Samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>Recall</i> | <i>F1-score</i> | | | | |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|--------|--------|--------|--------|
| 80:20 | 50 | | | 10 | 42,8% | 45,27% | 42,8% | 32,67% | | | | |
| | | | | 10 | 5 | 42,78% | 45,44% | 42,78% | 32,71% | | | |
| | | | | | 10 | 42,78% | 45,44% | 42,78% | 32,71% | | | |
| | | | | | 5 | 42,78% | 45,44% | 42,78% | 32,71% | | | |
| | | | | | 10 | 42,78% | 45,44% | 42,78% | 32,71% | | | |
| | | | | 20 | 5 | 42,78% | 45,44% | 42,78% | 32,71% | | | |
| | | | | | 10 | 42,78% | 45,44% | 42,78% | 32,71% | | | |
| | | | | | 5 | 42,78% | 45,44% | 42,78% | 32,71% | | | |
| | | | | | 10 | 42,78% | 45,44% | 42,78% | 32,71% | | | |
| | | | | 100 | 10 | | 2 | 5 | 42,78% | 45,44% | 42,78% | 32,71% |
| | | | | | | | 10 | 42,78% | 45,44% | 42,78% | 32,71% | |
| | | | | | | | 5 | 42,78% | 45,44% | 42,78% | 32,71% | |
| | | | | | 20 | | 2 | 5 | 42,78% | 45,44% | 42,78% | 32,71% |
| | | | | | | | 10 | 42,78% | 45,44% | 42,78% | 32,71% | |
| | | | | | | | 5 | 42,78% | 45,44% | 42,78% | 32,71% | |

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_Samples_split</i> | <i>Min_Samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>Recall</i> | <i>F1-score</i> | |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|--------|
| 70:30 | 50 | 10 | 2 | 10 | 42,78% | 45,44% | 42,78% | 32,71% | |
| | | | | 5 | 42,96% | 45,53% | 42,96% | 32,91% | |
| | | | 10 | 5 | 42,96% | 45,53% | 42,96% | 32,91% | |
| | | | 10 | 5 | 42,96% | 45,53% | 42,96% | 32,91% | |
| | | | 20 | 2 | 5 | 42,96% | 45,53% | 42,96% | 32,91% |
| | | | 20 | 10 | 42,96% | 45,53% | 42,96% | 32,91% | |
| | | 100 | 10 | 2 | 5 | 42,96% | 45,53% | 42,96% | 32,91% |
| | | | | | 10 | 42,96% | 45,53% | 42,96% | 32,91% |
| | | | 10 | 5 | 5 | 42,96% | 45,53% | 42,96% | 32,91% |
| | | | | | 10 | 42,96% | 45,53% | 42,96% | 32,91% |
| | | | 20 | 2 | 5 | 42,96% | 45,53% | 42,96% | 32,91% |
| | | | | | 10 | 42,96% | 45,53% | 42,96% | 32,91% |
| | | 20 | 5 | 5 | 42,96% | 45,53% | 42,96% | 32,91% | |
| | | | | 5 | 42,96% | 45,53% | 42,96% | 32,91% | |

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_Samples_split</i> | <i>Min_Samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>Recall</i> | <i>F1-score</i> |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|
| | | | | 10 | 42,96% | 45,53% | 42,96% | 32,91% |



UIN SUSKA RIAU

hak cipta milik
 ta Diindungi Unda
 rang mengutip seb
 engutipan hanya ur
 engutipan tidak me
 rang mengumumka
 g
 seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 ntangan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan
 apentingan yang wajar UIN Suska Riau.
 mperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Suska Riau

State Islamic University of Sultan Syar

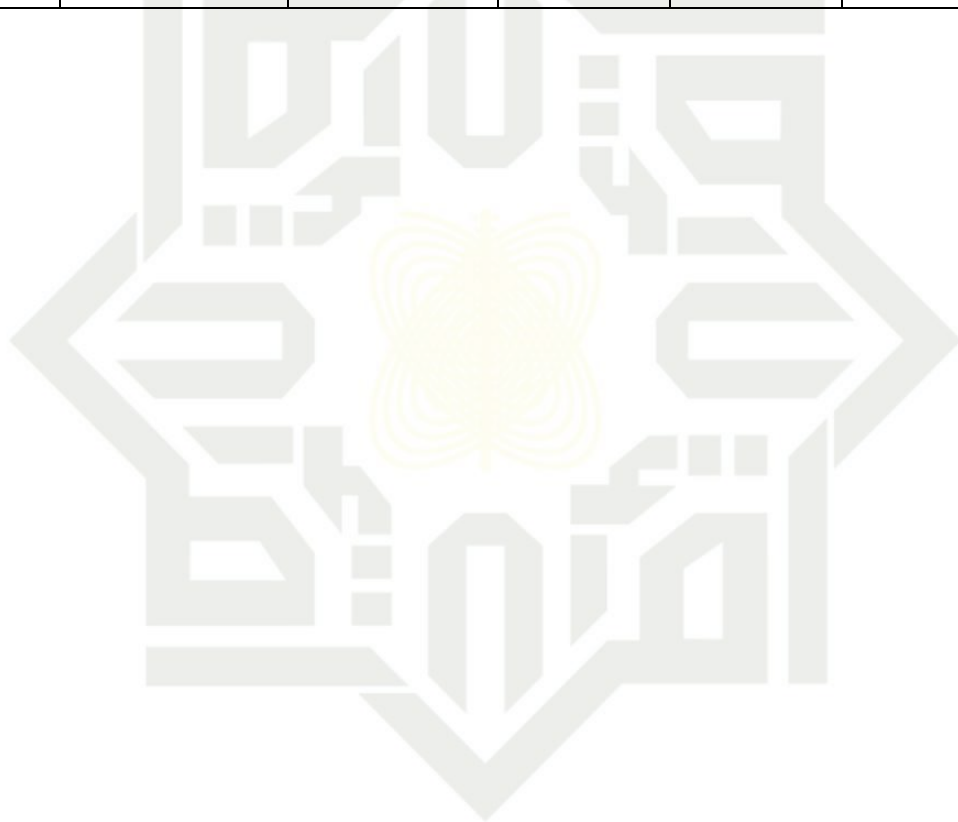
Lampiran III Hasil pengujian 25% data *balance* dengan penanganan *outlier*

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_Samples_split</i> | <i>Min_samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>recall</i> | <i>F1-score</i> |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|
| 90:10 | 50 | 10 | 2 | 5 | 92,08% | 84,79% | 92,08% | 88.29% |
| | | | | 10 | 92,08% | 84,79% | 92,08% | 88.29% |
| | | | 5 | 5 | 92,08% | 84,79% | 92,08% | 88.29% |
| | | 20 | 2 | 5 | 92,08% | 84,79% | 92,08% | 88.29% |
| | | | | 10 | 92,08% | 84,79% | 92,08% | 88.29% |
| | | | 5 | 5 | 92,08% | 84,79% | 92,08% | 88.29% |
| | 100 | 10 | 2 | 5 | 92,08% | 84,79% | 92,08% | 88.29% |
| | | | | 10 | 92,08% | 84,79% | 92,08% | 88.29% |
| | | | 5 | 5 | 92,08% | 84,79% | 92,08% | 88.29% |
| | | 20 | 2 | 5 | 92,08% | 84,79% | 92,08% | 88.29% |
| | | | | 10 | 92,08% | 84,79% | 92,08% | 88.29% |
| | | | 5 | 5 | 92,08% | 84,79% | 92,08% | 88.29% |

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_Samples_split</i> | <i>Min_samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>recall</i> | <i>F1-score</i> | | | |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|--------|--------|--------|
| 80:20 | 50 | | | 10 | 92,08% | 84,79% | 92,08% | 88.29% | | | |
| | | | | 2 | 5 | 91,94% | 84,53% | 91,94% | 88,08% | | |
| | | | | | 10 | 91,94% | 84,53% | 91,94% | 88,08% | | |
| | | | | 5 | 5 | 91,94% | 84,53% | 91,94% | 88,08% | | |
| | | | | | 10 | 91,94% | 84,53% | 91,94% | 88,08% | | |
| | | | | 20 | 2 | 5 | 91,94% | 84,53% | 91,94% | 88,08% | |
| | | | | | | 10 | 91,94% | 84,53% | 91,94% | 88,08% | |
| | | | | | 5 | 5 | 91,94% | 84,53% | 91,94% | 88,08% | |
| | | | | | | 10 | 91,94% | 84,53% | 91,94% | 88,08% | |
| | | | | 100 | 10 | 2 | 5 | 91,94% | 84,53% | 91,94% | 88,08% |
| | | | | | | | 10 | 91,94% | 84,53% | 91,94% | 88,08% |
| | | | | | | 5 | 5 | 91,94% | 84,53% | 91,94% | 88,08% |
| | | | | | | | 10 | 91,94% | 84,53% | 91,94% | 88,08% |
| | | | | | 20 | 2 | 5 | 91,94% | 84,53% | 91,94% | 88,08% |
| | | | | | | | 10 | 91,94% | 84,53% | 91,94% | 88,08% |
| | | | | | | 5 | 5 | 91,94% | 84,53% | 91,94% | 88,08% |
| | | | | | | | 5 | 91,94% | 84,53% | 91,94% | 88,08% |

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_Samples_split</i> | <i>Min_samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>recall</i> | <i>F1-score</i> | | | |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|--------|--------|--------|
| 70:30 | 50 | | | 10 | 91,94% | 84,53% | 91,94% | 88,08% | | | |
| | | | | 2 | 5 | 91,9% | 84,45% | 91,9% | 88,02% | | |
| | | | | | 10 | 91,9% | 84,45% | 91,9% | 88,02% | | |
| | | | | 5 | 5 | 91,9% | 84,45% | 91,9% | 88,02% | | |
| | | | | | 10 | 91,9% | 84,45% | 91,9% | 88,02% | | |
| | | | | 20 | 2 | 5 | 91,9% | 84,45% | 91,9% | 88,02% | |
| | | | | | | 10 | 91,9% | 84,45% | 91,9% | 88,02% | |
| | | | | | 5 | 5 | 91,9% | 84,45% | 91,9% | 88,02% | |
| | | | | | | 10 | 91,9% | 84,45% | 91,9% | 88,02% | |
| | | | | 100 | 10 | 2 | 5 | 91,9% | 84,45% | 91,9% | 88,02% |
| | | | | | | | 10 | 91,9% | 84,45% | 91,9% | 88,02% |
| | | | | | | 5 | 5 | 91,9% | 84,45% | 91,9% | 88,02% |
| | | | | | 20 | 2 | 5 | 91,9% | 84,45% | 91,9% | 88,02% |
| | | | | | | | 10 | 91,9% | 84,45% | 91,9% | 88,02% |
| | | | | | | 5 | 5 | 91,9% | 84,45% | 91,9% | 88,02% |

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_Samples_split</i> | <i>Min_samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>recall</i> | <i>F1-score</i> |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|
| | | | | 10 | 91,9% | 84,45% | 91,9% | 88,02% |



UIN SUSKA RIAU

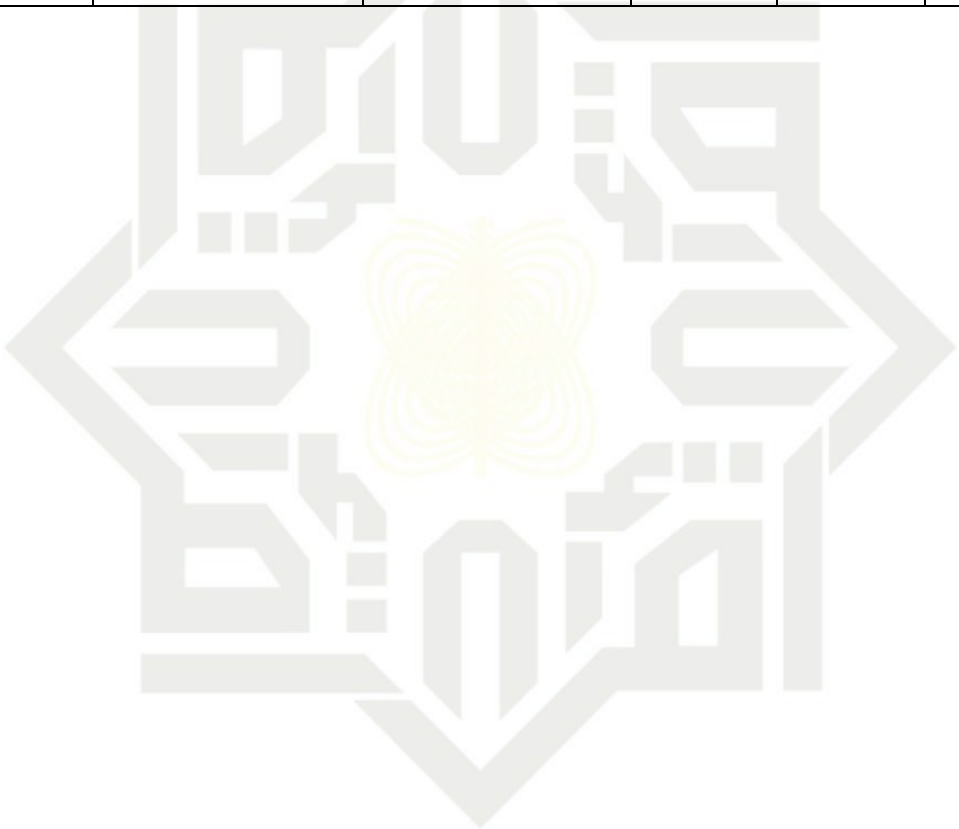
Lampiran IV Hasil pengujian 10% data *balance* dengan proses *outlier*

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_samples_split</i> | <i>Min_Samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>recall</i> | <i>F1-score</i> | | |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|--------|--------|
| 90:10 | 50 | 10 | 2 | 5 | 92,59% | 85,72% | 92,59% | 89,02% | | |
| | | | | 10 | 92,59% | 85,72% | 92,59% | 89,02% | | |
| | | | 5 | 5 | 92,59% | 85,72% | 92,59% | 89,02% | | |
| | | | | 10 | 92,59% | 85,72% | 92,59% | 89,02% | | |
| | | | 20 | 2 | 5 | 92,59% | 85,72% | 92,59% | 89,02% | |
| | | | | | 10 | 92,59% | 85,72% | 92,59% | 89,02% | |
| | | 5 | | 5 | 92,59% | 85,72% | 92,59% | 89,02% | | |
| | | | | 10 | 92,59% | 85,72% | 92,59% | 89,02% | | |
| | | 100 | | 10 | 2 | 5 | 92,59% | 85,72% | 92,59% | 89,02% |
| | | | | | | 10 | 92,59% | 85,72% | 92,59% | 89,02% |
| | | | 5 | | 5 | 92,59% | 85,72% | 92,59% | 89,02% | |
| | | | | | 10 | 92,59% | 85,72% | 92,59% | 89,02% | |
| | 20 | | 2 | | 5 | 92,59% | 85,72% | 92,59% | 89,02% | |
| | | | | | 10 | 92,59% | 85,72% | 92,59% | 89,02% | |
| | | | 5 | 5 | 92,59% | 85,72% | 92,59% | 89,02% | | |
| | | | | 5 | 92,59% | 85,72% | 92,59% | 89,02% | | |

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_samples_split</i> | <i>Min_Samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>recall</i> | <i>F1-score</i> | | | |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|--------|--------|--------|
| 80:20 | 50 | | | 10 | 92,59% | 85,72% | 92,59% | 89,02% | | | |
| | | | | 5 | 92,85% | 86,22% | 92,85% | 89,41% | | | |
| | | | | 10 | | 10 | 92,85% | 86,22% | 92,85% | 89,41% | |
| | | | | | | 5 | 92,85% | 86,22% | 92,85% | 89,41% | |
| | | | | | | 10 | 92,85% | 86,22% | 92,85% | 89,41% | |
| | | | | | | 5 | 92,85% | 86,22% | 92,85% | 89,41% | |
| | | | | 20 | | 2 | 92,85% | 86,22% | 92,85% | 89,41% | |
| | | | | | | 10 | 92,85% | 86,22% | 92,85% | 89,41% | |
| | | | | | | 5 | 92,85% | 86,22% | 92,85% | 89,41% | |
| | | | | | | 10 | 92,85% | 86,22% | 92,85% | 89,41% | |
| | | | | 100 | | 10 | 2 | 92,85% | 86,22% | 92,85% | 89,41% |
| | | | | | | | 10 | 92,85% | 86,22% | 92,85% | 89,41% |
| | | | | | | | 5 | 92,85% | 86,22% | 92,85% | 89,41% |
| | | | | | | 20 | 2 | 92,85% | 86,22% | 92,85% | 89,41% |
| | | | | | | | 10 | 92,85% | 86,22% | 92,85% | 89,41% |
| | | | | | | | 5 | 92,85% | 86,22% | 92,85% | 89,41% |

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_samples_split</i> | <i>Min_Samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>recall</i> | <i>F1-score</i> | | |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|--------|--------|
| 70:30 | 50 | | | 10 | 92,85% | 86,22% | 92,85% | 89,41% | | |
| | | | | 5 | 92,92% | 86,34% | 92,92% | 89,51% | | |
| | | | | 10 | 10 | 92,92% | 86,34% | 92,92% | 89,51% | |
| | | | | | 5 | 92,92% | 86,34% | 92,92% | 89,51% | |
| | | | | | 10 | 92,92% | 86,34% | 92,92% | 89,51% | |
| | | | | | 5 | 92,92% | 86,34% | 92,92% | 89,51% | |
| | | | | 20 | 2 | 92,92% | 86,34% | 92,92% | 89,51% | |
| | | | | | 10 | 92,92% | 86,34% | 92,92% | 89,51% | |
| | | | | | 5 | 92,92% | 86,34% | 92,92% | 89,51% | |
| | | | | | 10 | 92,92% | 86,34% | 92,92% | 89,51% | |
| | | | | 100 | 10 | 2 | 92,92% | 86,34% | 92,92% | 89,51% |
| | | | | | | 10 | 92,92% | 86,34% | 92,92% | 89,51% |
| | | | | | 20 | 5 | 92,92% | 86,34% | 92,92% | 89,51% |
| | | | | | | 10 | 92,92% | 86,34% | 92,92% | 89,51% |
| | | | | | | 5 | 92,92% | 86,34% | 92,92% | 89,51% |
| | | | | | | 5 | 92,92% | 86,34% | 92,92% | 89,51% |

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_samples_split</i> | <i>Min_Samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>recall</i> | <i>F1-score</i> |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|
| | | | | 10 | 92,92% | 86,34% | 92,92% | 89,51% |



UIN SUSKA RIAU

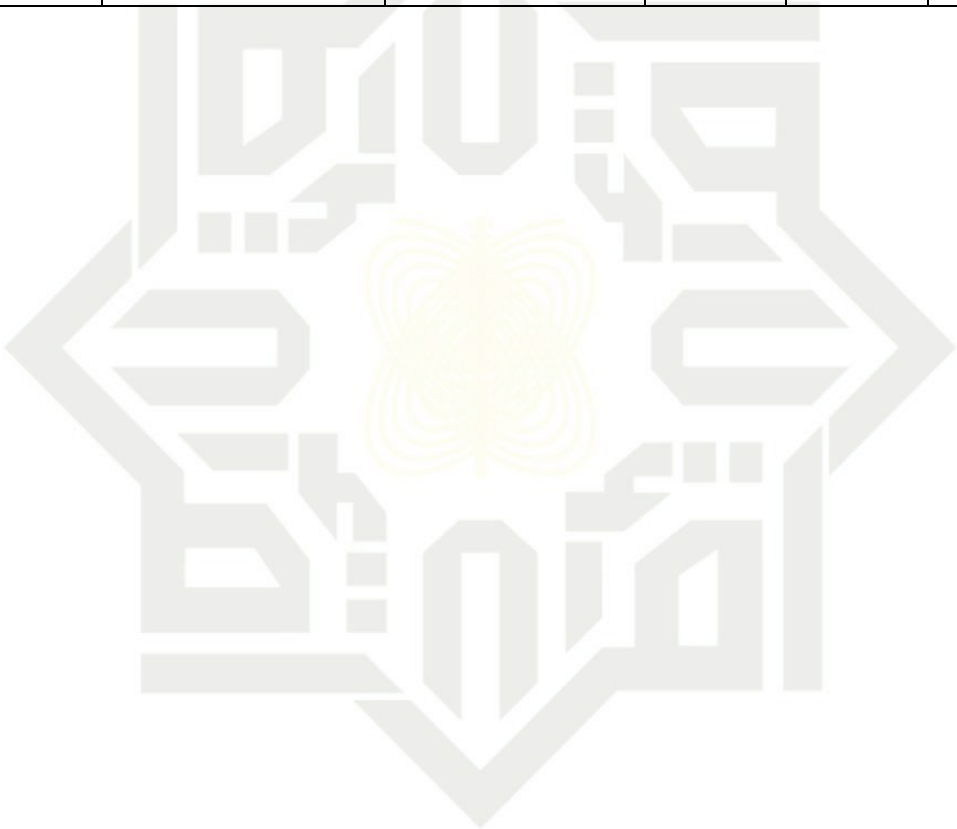
Lampiran V Hasil pengujian data *imbalance* tanpa proses *outlier*

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_dept h</i> | <i>Min_samples_split</i> | <i>Min_samples_leaf f</i> | <i>accuracy y</i> | <i>precision n</i> | <i>recall</i> | <i>F1-score</i> |
|--------------------|--------------------|-------------------|--------------------------|---------------------------|-------------------|--------------------|---------------|-----------------|
| 90:10 | 50 | 10 | 2 | 5 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | 10 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | 5 | 5 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | 10 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | 20 | 2 | 5 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | 10 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | 5 | 5 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | 10 | 99,99% | 99,99% | 99,99% | 99,99% |
| | 100 | 10 | 2 | 5 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | 10 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | 5 | 5 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | 10 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | 20 | 2 | 5 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | 10 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | 5 | 5 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | 5 | 99,99% | 99,99% | 99,99% | 99,99% |

| Rasio Perbandingan | $N_{estimator}$ | Max_dept h | $Min_samples_split$ | $Min_samples_leaf$ f | $accuracy$ y | $precision$ n | $recall$ | $F1-score$ |
|--------------------|-----------------|--------------------|-----------------------|-----------------------------|-------------------|--------------------|----------|------------|
| 80:20 | 50 | 10 | 2 | 10 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | 5 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | 5 | 10 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | 2 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | 10 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | 5 | 99,99% | 99,99% | 99,99% | 99,99% |
| | 100 | 10 | 2 | 5 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | 10 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | | 5 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | 5 | 10 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | | 2 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | 10 | 99,99% | 99,99% | 99,99% | 99,99% |
| | 5 | 20 | 2 | 99,99% | 99,99% | 99,99% | 99,99% | |
| | | | 5 | 99,99% | 99,99% | 99,99% | 99,99% | |

| Rasio Perbandingan | $N_{estimator}$ | Max_dept h | $Min_samples_split$ | $Min_samples_leaf$ f | $accuracy$ y | $precision$ n | $recall$ | $F1$ - $score$ |
|--------------------|-----------------|--------------------|-----------------------|-----------------------------|-------------------|--------------------|----------|-------------------|
| 70:30 | 50 | 10 | 2 | 10 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | | 5 | 99,99% | 99,99% | 99,99% | 99,99% |
| | 5 | | 10 | 99,98% | 99,98% | 99,98% | 99,98% | |
| | | | 5 | 99,99% | 99,99% | 99,99% | 99,99% | |
| | 20 | 2 | 5 | 99,99% | 99,99% | 99,99% | 99,99% | |
| | | | 10 | 99,98% | 99,98% | 99,98% | 99,98% | |
| | | 5 | 5 | 99,99% | 99,99% | 99,99% | 99,99% | |
| | | | 10 | 99,98% | 99,98% | 99,98% | 99,98% | |
| | 100 | 10 | 2 | 5 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | | 10 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | 5 | 5 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | 10 | | 5 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | 10 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | 20 | 2 | 5 | 99,99% | 99,99% | 99,99% | 99,99% |
| | 5 | | | 99,99% | 99,99% | 99,99% | 99,99% | |

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_dept h</i> | <i>Min_samples_split</i> | <i>Min_samples_leaf</i> <i>f</i> | <i>accuracy</i> <i>y</i> | <i>precision</i> <i>n</i> | <i>recall</i> | <i>F1-score</i> |
|--------------------|--------------------|-------------------|--------------------------|-------------------------------------|-----------------------------|------------------------------|---------------|-----------------|
| | | | | 10 | 99,98% | 99,98% | 99,98% | 99,98% |



UIN SUSKA RIAU

Lampiran VI Hasil pengujian data *balance* 100% tanpa penanganan *outlier*

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_samples_split</i> | <i>Min_samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>recall</i> | <i>F1-score</i> |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|
| 90:10 | 50 | 10 | 2 | 5 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | | 10 | 99,96% | 99,96% | 99,96% | 99,96% |
| | | | 5 | 5 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | | 10 | 99,96% | 99,96% | 99,96% | 99,96% |
| | | 20 | 2 | 5 | 100% | 100% | 100% | 100% |
| | | | | 10 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | 5 | 5 | 100% | 100% | 100% | 100% |
| | | | | 10 | 99,99% | 99,99% | 99,99% | 99,99% |
| | 100 | 10 | 2 | 5 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | | 10 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | 5 | 5 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | | 10 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | 20 | 2 | 5 | 100% | 100% | 100% | 100% |
| | | | | 10 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | 5 | 5 | 100% | 100% | 100% | 100% |
| | | | | 5 | 100% | 100% | 100% | 100% |

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_samples_split</i> | <i>Min_samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>recall</i> | <i>F1-score</i> |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|
| 80:20 | 50 | 10 | 2 | 10 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | 5 | 99,97% | 99,97% | 99,97% | 99,97% |
| | 5 | | 10 | 99,98% | 99,98% | 99,98% | 99,98% | |
| | | | 5 | 99,97% | 99,97% | 99,97% | 99,97% | |
| | 20 | 2 | 5 | 100% | 100% | 100% | 100% | |
| | | | 10 | 99,99% | 99,99% | 99,99% | 99,99% | |
| | | | 5 | 100% | 100% | 100% | 100% | |
| | | 5 | 10 | 99,99% | 99,99% | 99,99% | 99,99% | |
| | | | 5 | 99,98% | 99,98% | 99,98% | 99,98% | |
| | | | 10 | 99,98% | 99,98% | 99,98% | 99,98% | |
| | 10 | 10 | 2 | 5 | 100% | 100% | 100% | 100% |
| | | | | 10 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | 5 | 100% | 100% | 100% | 100% | |
| | | 20 | 2 | 5 | 100% | 100% | 100% | 100% |
| | | | | 10 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | 5 | 100% | 100% | 100% | 100% | |

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_samples_split</i> | <i>Min_samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>recall</i> | <i>F1-score</i> | | |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|--------|--------|
| 70:30 | Suska Riau | 50 | | 10 | 99,99% | 99,99% | 99,99% | 99,99% | | |
| | | | | 5 | 99,97% | 99,97% | 99,97% | 99,97% | | |
| | | | | 10 | 2 | 10 | 99,96% | 99,96% | 99,96% | 99,96% |
| | | | | | 5 | 5 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | | | 10 | 99,96% | 99,96% | 99,96% | 99,96% | |
| | | | | | 5 | 99,99% | 99,99% | 99,99% | 99,99% | |
| | | | | 20 | 2 | 5 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | | 10 | 99,99% | 99,99% | 99,99% | 99,99% | |
| | | | | | 5 | 99,99% | 99,99% | 99,99% | 99,99% | |
| | | | | | 10 | 99,99% | 99,99% | 99,99% | 99,99% | |
| | | | | 10 | 2 | 5 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | | | | 10 | 99,96% | 99,96% | 99,96% | 99,96% |
| | | | | | | 5 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | | | 5 | 10 | 99,96% | 99,96% | 99,96% | 99,96% |
| | | | | | | 5 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | | | 10 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | 20 | 2 | 5 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | | | 10 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | | 5 | 99,99% | 99,99% | 99,99% | 99,99% | |

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_samples_split</i> | <i>Min_samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>recall</i> | <i>F1-score</i> |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|
| | | | | 10 | 99,99% | 99,99% | 99,99% | 99,99% |



UIN SUSKA RIAU

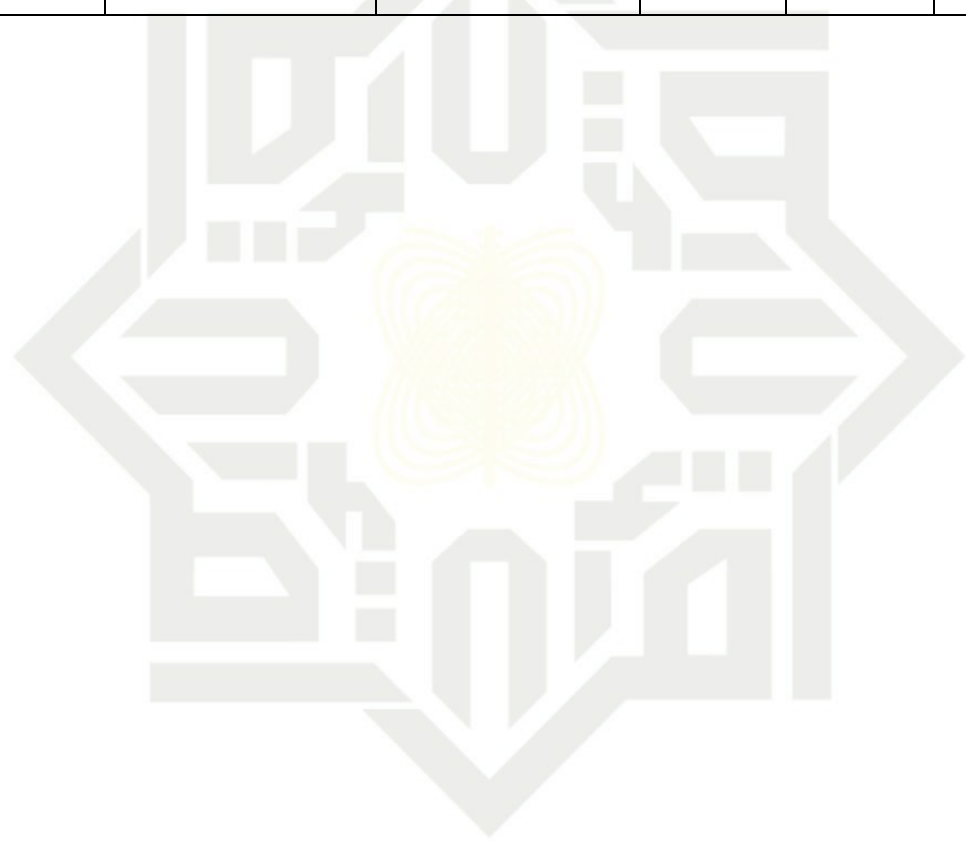
Lampiran VII Hasil pengujian data *balance* 25% tanpa penanganan *outlier*

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_samples_split</i> | <i>Min_samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>recall</i> | <i>F1-score</i> |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|
| 90:10 | 50 | 10 | 2 | 5 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | | 10 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | 5 | 5 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | 20 | 2 | 5 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | | 10 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | 5 | 5 | 99,98% | 99,98% | 99,98% | 99,98% |
| | 100 | 10 | 2 | 5 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | | 10 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | 5 | 5 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | 20 | 2 | 5 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | | 10 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | 5 | 5 | 99,98% | 99,98% | 99,98% | 99,98% |

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_samples_split</i> | <i>Min_samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>recall</i> | <i>F1-score</i> |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|
| 80:20 | 50 | 10 | 2 | 10 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | | 5 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | | 10 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | | 5 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | | 10 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | | 5 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | 10 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | | 5 | 99,99% | 99,99% | 99,99% | 99,99% |
| | | | | 10 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | | 5 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | | 10 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | | 5 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | | 10 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | | 5 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | | 10 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | | 5 | 99,98% | 99,98% | 99,98% | 99,98% |

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_samples_split</i> | <i>Min_samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>recall</i> | <i>F1-score</i> | | |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|--------|--------|
| 70:30 | Suska Riau | 50 | | 10 | 99,98% | 99,98% | 99,98% | 99,98% | | |
| | | | | 5 | 99,97% | 99,97% | 99,97% | 99,97% | | |
| | | | | 10 | 10 | 99,96% | 99,96% | 99,96% | 99,96% | |
| | | | | | 5 | 99,97% | 99,97% | 99,97% | 99,97% | |
| | | | | | 10 | 99,96% | 99,96% | 99,96% | 99,96% | |
| | | | | | 5 | 99,98% | 99,98% | 99,98% | 99,98% | |
| | | | | 20 | 2 | 99,96% | 99,96% | 99,96% | 99,96% | |
| | | | | | 5 | 99,98% | 99,98% | 99,98% | 99,98% | |
| | | | | | 10 | 99,96% | 99,96% | 99,96% | 99,96% | |
| | | | | | 5 | 99,96% | 99,96% | 99,96% | 99,96% | |
| | | | | 100 | 10 | 2 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | | | | 5 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | | | | 10 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | | | 20 | 2 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | | | | 10 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | | | | 5 | 99,98% | 99,98% | 99,98% | 99,98% |

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_samples_split</i> | <i>Min_samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>recall</i> | <i>F1-score</i> |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|
| | | | | 10 | 99,97% | 99,97% | 99,97% | 99,97% |



UIN SUSKA RIAU

hak cipta milli
 ta Diindungi Unda
 rang mengutip seb
 engutipan hanya ur
 engutipan tidak me
 rang mengumumka
 g
 seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 ntingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan
 apentingan yang wajar UIN Suska Riau.
 mperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

Suska Riau

State Islamic University of Sultan Syar

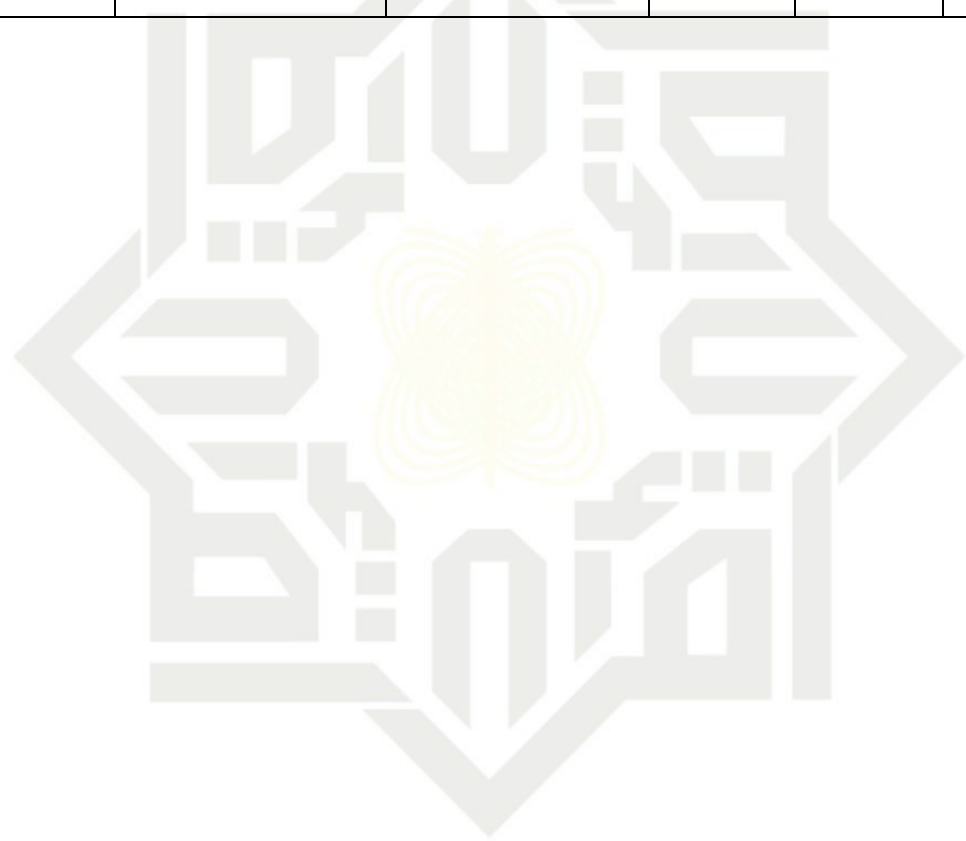
Lampiran VIII Hasil pengujian data *balance* 10% tanpa penanganan *outlier*

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_samples_split</i> | <i>Min_samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>recall</i> | <i>F1-score</i> |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|
| 90:10 | 50 | 10 | 2 | 5 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | | 10 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | 5 | 5 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | 20 | 2 | 5 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | | 10 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | 5 | 5 | 99,98% | 99,98% | 99,98% | 99,98% |
| | 100 | 10 | 2 | 5 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | | 10 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | 5 | 5 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | 20 | 2 | 5 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | | 10 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | 5 | 5 | 99,98% | 99,98% | 99,98% | 99,98% |

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_samples_split</i> | <i>Min_samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>recall</i> | <i>F1-score</i> |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|
| 80:20 | 50 | 10 | 2 | 10 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | | 5 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | 10 | 99,96% | 99,96% | 99,96% | 99,96% | |
| | | 20 | 2 | 5 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | | 10 | 99,96% | 99,96% | 99,96% | 99,96% |
| | | | 5 | 99,98% | 99,98% | 99,98% | 99,98% | |
| | 100 | 10 | 2 | 5 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | | 10 | 99,96% | 99,96% | 99,96% | 99,96% |
| | | | 5 | 99,98% | 99,98% | 99,98% | 99,98% | |
| | | 20 | 2 | 5 | 99,98% | 99,98% | 99,98% | 99,98% |
| | | | | 10 | 99,97% | 99,97% | 99,97% | 99,97% |
| | | | 5 | 99,98% | 99,98% | 99,98% | 99,98% | |

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_samples_split</i> | <i>Min_samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>recall</i> | <i>F1-score</i> | | |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|--------|--------|
| 70:30 | 50 | 10 | 2 | 10 | 99,97% | 99,97% | 99,97% | 99,97% | | |
| | | | | 5 | 99,97% | 99,97% | 99,97% | 99,97% | | |
| | | | 10 | 5 | 99,97% | 99,97% | 99,97% | 99,97% | | |
| | | | 10 | 10 | 99,96% | 99,96% | 99,96% | 99,96% | | |
| | | | 20 | 2 | 5 | 99,98% | 99,98% | 99,98% | 99,98% | |
| | | | | 2 | 10 | 99,97% | 99,97% | 99,97% | 99,97% | |
| | | 5 | | 5 | 99,98% | 99,98% | 99,98% | 99,98% | | |
| | | 5 | | 10 | 99,97% | 99,97% | 99,97% | 99,97% | | |
| | | 100 | 10 | 2 | 5 | 99,98% | 99,98% | 99,98% | 99,98% | |
| | | | | | 10 | 99,96% | 99,96% | 99,96% | 99,96% | |
| | | | | 5 | 5 | 99,98% | 99,98% | 99,98% | 99,98% | |
| | | | 20 | 2 | 5 | 99,96% | 99,96% | 99,96% | 99,96% | |
| | | | | | 10 | 99,98% | 99,98% | 99,98% | 99,98% | |
| | | | | 5 | 2 | 99,97% | 99,97% | 99,97% | 99,97% | |
| | | | | | 5 | 5 | 99,98% | 99,98% | 99,98% | 99,98% |

| Rasio Perbandingan | <i>N_estimator</i> | <i>Max_depth</i> | <i>Min_samples_split</i> | <i>Min_samples_leaf</i> | <i>accuracy</i> | <i>precision</i> | <i>recall</i> | <i>F1-score</i> |
|--------------------|--------------------|------------------|--------------------------|-------------------------|-----------------|------------------|---------------|-----------------|
| | | | | 10 | 99,97% | 99,97% | 99,97% | 99,97% |



UIN SUSKA RIAU

hak cipta milli
 ta Diindungi Unda
 rang mengutip seb
 engutipan hanya ur
 engutipan tidak me
 rang mengumumka
 seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber:
 tingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan
 penting yang wajar UIN Suska Riau.
 mperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin UIN Suska Riau.

DAFTAR RIWAYAT HIDUP

Data Pribadi

| | |
|----------------------|--|
| Nama Lengkap | : Yolanda |
| Tempat/Tanggal Lahir | : Sawahlunto Sijunjung/ 24 Januari 2002 |
| Jenis Kelamin | : Perempuan |
| Agama | : Islam |
| Kewarganegaraan | : Indonesia |
| Alamat | : Jorong Sungai Lansek, Kecamatan Sitiung, Kabupaten Dharmasraya |
| No. HP | : 082246009284 |
| Email | : 12050120358@students.uin-suska.ac.id |



Pendidikan

| | |
|-----------|--|
| 2006-2008 | : TK Bukit Candi Padang Roco |
| 2008-2014 | : SDN 03 Sitiung |
| 2014-2017 | : SMPN Unggul Dharmasraya |
| 2017-2020 | : SMAN 1 Pulau Punjung |
| 2020-2024 | : S1 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.