

Pendekatan *Hybrid: Naïve Bayes* dan *Decision Tree* untuk Prediksi Kerusakan Mesin pada Industri Manufaktur PT X

Iin Rohmatika Aulia^{1*}, Arief Hermawan¹, Donny Avianto²

¹Magister Teknologi Informasi, Universitas Teknologi Yogyakarta

²Teknik Informatika, Fakultas Sains & Teknologi, Universitas Teknologi Yogyakarta

*Email: iin.6240211012@student.uty.ac.id

Info Artikel

Kata Kunci :

Klasifikasi, *Naïve Bayes*, *Decision Tree*, Prediktif, Maintenance.

Keywords :

Classification, *Naïve Bayes*, *Decision Tree*, *Predictif*, *Maintenance*.

Tanggal Artikel

Dikirim : 8 Januari 2025

Direvisi : 26 Mei 2025

Diterima : 31 Mei 2025

Abstrak

Perkembangan teknologi sistem informasi banyak dirasakan di setiap sector ekonomi. PT X merupakan perusahaan manufaktur di bidang percetakan, dimana produktivitas dipengaruhi dari efisiensi mesin. Optimasi produktivitas mesin dapat dilakukan dengan prediktif maintenance. Penelitian ini bertujuan untuk mengembangkan teknik *data mining* dalam prediktif kerusakan mesin produksi. Fokus utama penelitian adalah untuk mengklasifikasi kerusakan mesin berdasarkan data historis pada PT X. Model klasifikasi yang akan dikembangkan menggunakan algoritma model *Naïve Bayes* dan *Decision Tree*. Dalam klasifikasi ada 2 label keputusan yaitu tingkat resiko (tinggi, sedang rendah) dan kegiatan preventif (Ya,Tidak) Evaluasi dilakukan dengan menilai akurasi dan efektivitas setiap model. Hasil uji klasifikasi preventif dengan model *Naïve Bayes* memiliki nilai akurasi 97,90 %, sedangkan dengan model *Decision Tree* memiliki nilai akurasi 77%. Hasil uji klasifikasi tingkat resiko dengan model *Naïve Bayes* nilai akurasi 98% sedangkan dengan model *Decision Tree* nilai akurasinya 100%. hasil uji menunjukan untuk label preventif dengan 2 kelas lebih baik menggunakan model *Naïve Bayes* sedangkan label tingkat resiko dengan 3 kelas lebih baik menggunakan model *Decision Tree*. Hasil uji ini dapat dijadikan acuan Perusahaan X khususnya divisi maintenance dalam melakukan penjadwalan prediktif maintenance. Metode ini juga dapat diterapkan pada Perusahaan lain jika memiliki data historis kerusakan mesin, memiliki mesin dengan jenis operasional yang relevan, dan memiliki tujuan dan klasifikasi yang sesuai.

Abstract

The advancement of information system technology has significantly impacted all economic sectors. PT X, a manufacturing company in the printing industry, experiences productivity fluctuations that are strongly influenced by machine efficiency. Optimizing machine productivity can be achieved through predictive maintenance. This study aims to develop data mining techniques for predicting machine failures in production. The primary focus is to classify machine failures based on historical data from PT X. The classification models employed are the Naïve Bayes algorithm and the Decision Tree algorithm. Two classification labels are used: risk level (high, medium, low) and preventive action (Yes, No). Evaluation was conducted by measuring the accuracy and effectiveness of each model. The classification results for the preventive action label showed that the Naïve Bayes model achieved an accuracy of 97.90%, while the Decision Tree model reached 77%. For the risk level label, the Naïve Bayes model achieved 98% accuracy, and the Decision Tree model achieved 100%. The findings indicate that the Naïve Bayes model is more suitable for binary classifications such as preventive actions, while the Decision Tree model performs better in multi-class classifications such as risk levels. These results can serve as a reference for PT X's maintenance division in scheduling predictive maintenance. Moreover, the method can be applied to other companies, provided they have historical machine failure data, machines with similar operational characteristics, and compatible classification objectives.

1. PENDAHULUAN

PT X merupakan perusahaan manufaktur bidang percetakan. Ada beberapa faktor yang mempengaruhi profit dalam Perusahaan manufaktur bidang percetakan, antara lain produktivitas, efisiensi dan efektivitas [1]. Adapun yang menunjang produktivitas dalam Perusahaan antara lain mesin [2]. industri percetakan besaran produktivitasnya tergantung dengan dua factor utama yaitu produktivitas mesin dan tenaga kerja. Produktivitas mesin dipengaruhi oleh *downtime* mesin. Semakin sedikit *downtime* mesin maka produk yang dihasilkan semakin banyak. Semakin sedikit *downtime* juga artinya semakin sedikit waktu terbuang (*waste time*). Maka Perusahaan harus mengupayakan zero atau sedikit waktu *downtime* mesin[3][4]. *Downtime* mesin dalam PT X dapat dikelompokkan dengan 2 kelompok secara penanganan yaitu preventif dan tidak preventif [5]. Secara umum predictive maintenance adalah perawatan yang bersifat prediksi, dalam hal ini merupakan evaluasi dari perawatan berkala (Preventive Maintenance)[6]. secara resiko terbagi menjadi 3 kategori yaitu : resiko rendah, resiko sedang, dan resiko tinggi [6][7]. PT X memiliki data historis *downtime* mesin. Dataset ini akan dimanfaatkan untuk memprediksi kerusakan mesin kedepannya. Kerusakan mesin yang dapat diprediksi, sehingga dapat melakukan penanganan preventif. Oleh karena itu, maka diharapkan *waste time* akan berkurang dan akan meningkatkan produktivitas PT X. Dataset yang didapat untuk mengetahui nilai tambah atau pengetahuan yang ada didalamnya, maka perlu dilakukan Knowledge Discovery in Database (KDD), dalam hal ini akan dilakukan *Data mining* [8][9].

Data mining adalah proses otomatis untuk menemukan pola, hubungan, dan informasi berguna dari dataset besar menggunakan teknik statistik, pembelajaran mesin (*machine learning*), dan algoritma komputasional lainnya [10]. Tujuan utamanya adalah untuk menganalisis data dengan cara yang mengungkapkan pola-pola tersembunyi yang bisa digunakan untuk memprediksi tren atau mendukung keputusan berbasis data. *Data mining* biasanya mencakup beberapa teknik seperti klasifikasi, klustering, asosiasi, dan analisis prediktif [11]. *Data mining* adalah sebuah proses eksplorasi dan analisis data besar untuk menemukan pola yang tidak terlihat sebelumnya dan untuk meningkatkan pemahaman serta pengambilan keputusan yang berbasis data. Proses ini menggunakan algoritma kompleks untuk mengekstrak informasi yang tersembunyi dalam data besar yang tidak dapat ditemukan melalui analisis tradisional [12]. Dalam proses menuju *data mining* terdapat tahapan-tahapan sebelum dan sesudah *data mining* yaitu : *Data Selection*, *Preprocessing*, *Data Mining*, Interpretasi & Evaluasi, Knowledge. Pada tahap *data mining* terdapat beberapa Teknik seperti klasifikasi, klustering, regresi dan sebagainya [13]. Dalam penelitian ini akan menggunakan algoritma *Naïve Bayes* yang merupakan salah satu Teknik dari klasifikasi.

Algoritma *Naïve Bayes* adalah salah satu algoritma klasifikasi yang berbasis pada teori probabilitas *Bayes*, dengan asumsi bahwa fitur-fitur yang digunakan dalam model bersifat independen satu sama lain (meskipun dalam kenyataannya sering tidak) [14]. Meskipun asumsi independensi ini sangat sederhana dan tidak selalu berlaku dalam data nyata, Naive Bayes tetap efektif dalam banyak aplikasi, terutama di bidang klasifikasi teks, seperti spam filtering atau analisis sentiment. Kelebihan utama algoritma *Naïve Bayes* antara lain : Kecepatan dan Efisiensi, Kinerja Baik pada Data Besar, Berguna untuk Masalah Klasifikasi Multikelas, Kemampuan dalam Menangani Missing Data, Toleransi terhadap Noise, Kinerja yang Kompetitif untuk Teks dan Data Kategorikal.[15][16]

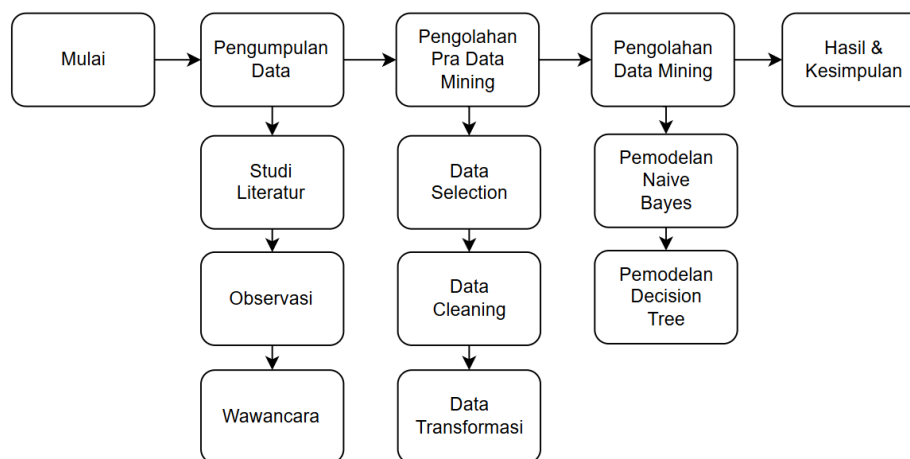
Algoritma *Decision Tree* bersifat sangat kuat, populer, berbasis logika, dan mudah dipahami. Hal yang menarik dari *Decision Tree* adalah penggunaan struktur pohon (*tree*) yang berfungsi untuk merepresentasikan aturan yang terbentuk dari hasil klasifikasi. Dalam tree, atribut direpresentasikan oleh sebuah node, dan kelas direpresentasikan oleh daun (*leaf*). Setiap pohon memiliki akar (*root*) yaitu node yang berada di paling atas [17]. Integrasi algoritma pembelajaran mesin dengan mesin produksi meningkatkan kemampuan pemeliharaan prediktif, mengurangi waktu henti, dan secara signifikan meningkatkan efisiensi produksi secara keseluruhan. Dalam menjalankan Teknik klasifikasi dengan algoritma *Naïve Bayes* dan *Decision Trees* [18]. Penelitian ini akan menggunakan aplikasi *machine learning RapidMiner* dalam pengolahan data, *RapidMiner* adalah sebuah platform perangkat lunak yang digunakan untuk analisis data dan pengolahan data besar dengan berbagai algoritma *machine learning*. *RapidMiner* menyediakan antarmuka grafis yang memudahkan pengguna untuk melakukan proses analisis data tanpa menulis kode secara langsung. Platform ini mendukung berbagai teknik data mining, termasuk klasifikasi, regresi, klustering, dan asosiasi.

Menurut Dyah & Fajar [19] “bahwa dengan metode algoritma *Naïve Bayes* berdasarkan data dari history kerusakan mesin dengan jumlah data training 754 record dan data testing 189 record, metode ini dapat memprediksi mesin mana saja yang perlu dilakukan preventive maintenance dengan tingkat akurasi 93,65%. Sehingga dapat membantu pihak maintenance department sebagai acuan untuk menyusun jadwal untuk melakukan kegiatan preventive maintenance”. Menurut Elda & Ika [20] dalam keseluruhan, model *Naïve Bayes* menunjukkan kinerja yang cukup baik dalam memprediksi keberadaan atau tidaknya penyakit anemia, meskipun terdapat kecenderungan untuk lebih baik dalam memprediksi kelas tanpa anemia daripada kelas dengan anemia. Penelitian ini bertujuan untuk mengembangkan teknik *data mining* dalam pemeliharaan prediktif mesin produksi dengan memanfaatkan database

kerusakan mesin. Fokus utama penelitian adalah untuk mengidentifikasi pola-pola kerusakan mesin berdasarkan data historis kerusakan pada PT X yang ada. Model prediktif akan dikembangkan menggunakan algoritma *Naïve Bayes Classifier* dan *Decision Trees* untuk memprediksi kemungkinan kerusakan mesin di masa depan [21]. Evaluasi akan dilakukan untuk menilai akurasi dan efektivitas setiap model dalam mendeteksi kerusakan lebih awal. Hasil penelitian diharapkan dapat menjadi bahan guna meningkatkan efisiensi pemeliharaan dan mengurangi biaya *downtime* produksi.

2. METODE PENELITIAN

Berikut adalah bagan alur metode penelitian ini ditunjukkan pada Gambar 1.



Gambar 1. Bagan Metode Penelitian

Penelitian ini dilakukan disebuah Perusahaan percetakan yang terletak di Kabupaten Klaten, Provinsi Jawa Tengah. Penelitian ini dilakukan pada November 2024. Secara garis besar tahapan-tahapan pada penelitian ini terbagi menjadi 3 yaitu 1. Pengumpulan Data, 2. Pengolahan Data, 3. Pengolahan Data Mining

2.1. Pengumpulan Data

Dalam pengumpulan data, penelitian ini menggunakan 3 metode :

1. Studi Literatur

Studi literatur adalah tahap awal dalam proses penelitian yang bertujuan untuk mengkaji, menganalisis, dan menyintesis penelitian-penelitian sebelumnya yang relevan dengan topik yang sedang diteliti. Dalam konteks penelitian mesin produksi percetakan, studi literatur melibatkan tinjauan terhadap berbagai sumber ilmiah yang membahas teori, temuan penelitian, serta metodologi yang berkaitan dengan kinerja mesin percetakan, pemeliharaan mesin, optimasi proses produksi, dan teknologi percetakan. Tujuan dari studi literatur ini adalah untuk mendapatkan pemahaman yang lebih dalam mengenai isu-isu yang relevan dengan mesin produksi percetakan, mengidentifikasi gap penelitian yang ada, serta memberikan dasar teori yang kuat bagi pengembangan penelitian lebih lanjut.

2. Observasi

Observasi dalam konteks ilmiah merupakan metode pengumpulan data yang dilakukan melalui pengamatan langsung terhadap objek atau fenomena yang sedang diteliti, dengan tujuan untuk mendapatkan informasi yang akurat dan sistematis. Dalam penelitian atau studi yang berhubungan dengan mesin produksi percetakan, observasi berfungsi untuk menganalisis kinerja, keandalan, serta kondisi operasional mesin dalam proses produksi. Dalam hal ini, observasi tidak hanya melibatkan pengamatan terhadap aspek fisik mesin, tetapi juga terhadap variabel-variabel yang memengaruhi efisiensi dan efektivitas proses produksi, seperti kecepatan operasional, kualitas cetakan, serta kemungkinan adanya kerusakan atau masalah teknis lainnya.

3. Wawancara

Wawancara adalah salah satu metode pengumpulan data dalam penelitian yang melibatkan interaksi langsung antara peneliti dan informan untuk memperoleh informasi yang lebih mendalam mengenai fenomena yang sedang diteliti. Dalam konteks penelitian mesin produksi percetakan, wawancara berfungsi untuk memperoleh informasi yang lebih mendalam mengenai kondisi operasional mesin, tantangan yang dihadapi operator atau

teknisi mesin, serta praktik pemeliharaan dan perawatan yang dilakukan di lapangan. Wawancara ini memberikan gambaran kualitatif yang berguna untuk mengidentifikasi masalah yang mungkin tidak terdeteksi melalui data kuantitatif, seperti kerusakan mesin, efisiensi produksi, atau prosedur pemeliharaan yang dilakukan oleh operator.

2.2. Pengolahan Data Pra Data Mining

Data yang digunakan dalam penelitian ini terdapat 5 kolom attribute yaitu jenis mesin, hambatan, waktu hambatan, jenis hambatan, dan PIC. Adapun jumlah data masing-masing parameter terdapat 2500 baris. Terdapat 2 tahapan sebelum masuk kedalam tahapan inti yaitu Data Mining, Adapun 2 tahapan ini adalah *Data Selection* dan *Transformation* [22].

(1) Data Selection

Pada tahap ini data akan diseleksi dengan cara melihat kecenderungan data/kesesuaian data topik/ judul penelitian yang akan diteliti, Adapun data yang akan diteliti diperoleh dari database file excel yang diperoleh dalam pengumpulan data di Tabel 1 [23].

Tabel 1. Data Selection

Mesin	Jenis	Hambatan	Jam Mulai	Jam Selesai	Selisih	Jenis
30 B	EKSTERNAL	LAIN-LAIN (TEST PRINT(KOORDINATOR))	05-09-2024 15:40:29	05-09-2024 15:41:07	1	KOORDINATOR
D300	EKSTERNAL	LAIN-LAIN (TUNGGU INFORMASI (GANTI ORDER))	01-09-2024 01:14:09	01-09-2024 02:07:38	53	KOORDINATOR
D300	EKSTERNAL	LAIN-LAIN (TUNGGU INFORMASI (GANTI KERTAS))	01-09-2024 02:16:05	01-09-2024 02:23:26	7	KOORDINATOR
D300	EKSTERNAL	LAIN-LAIN (TUNGGU INFORMASI (GANTI ORDER))	02-09-2024 08:21:20	02-09-2024 09:22:00	61	KOORDINATOR
D300	EKSTERNAL	LAIN-LAIN (TEST PRINT(KOORDINATOR))	02-09-2024 10:05:53	02-09-2024 10:34:02	29	KOORDINATOR
D300	EKSTERNAL	LAIN-LAIN (BON PLATE) - AUS : 1	02-09-2024 10:34:45	02-09-2024 10:45:24	11	REPRO
D300	EKSTERNAL	LAIN-LAIN (TEST PRINT(KOORDINATOR))	02-09-2024 11:23:55	02-09-2024 11:29:14	6	KOORDINATOR
D300	EKSTERNAL	SPLICER (REM KURANG OPTIMAL)	02-09-2024 11:30:17	02-09-2024 11:52:07	22	TEKNIK
D300	EKSTERNAL	FOLDER (LIPATAN)	02-09-2024 11:53:03	02-09-2024 17:29:35	336	TEKNIK
D300	EKSTERNAL	LAIN-LAIN (TUNGGU ACC PEMESAN)	02-09-2024 17:32:01	02-09-2024 17:34:41	2	KOORDINATOR
D300	EKSTERNAL	LAIN-LAIN (TUNGGU ACC PEMESAN)	02-09-2024 17:36:47	02-09-2024 18:04:40	28	KOORDINATOR
D300	EKSTERNAL	LAIN-LAIN (TUNGGU INFORMASI (GANTI KERTAS))	02-09-2024 19:31:33	02-09-2024 19:55:08	24	KOORDINATOR
D300	EKSTERNAL	LAIN-LAIN (TUNGGU ACC PEMESAN)	02-09-2024 20:04:15	02-09-2024 20:29:43	25	KOORDINATOR
D300	EKSTERNAL	FOLDER (LIPATAN)	02-09-2024 20:30:01	02-09-2024 22:32:58	122	TEKNIK
D300	EKSTERNAL	LAIN-LAIN (TEST PRINT(KOORDINATOR))	03-09-2024 16:01:37	03-09-2024 16:10:43	9	KOORDINATOR
D300	EKSTERNAL	PRINTING UNIT (MOTOR ROLL AIR RUSAK)	03-09-2024 16:11:25	03-09-2024 16:43:44	32	TEKNIK
D300	EKSTERNAL	STUCKER (STACKER MACET)	03-09-2024 16:49:00	03-09-2024 17:05:43	16	TEKNIK
D300	EKSTERNAL	LAIN-LAIN (TUNGGU ACC PEMESAN)	03-09-2024 17:54:28	03-09-2024 20:21:16	147	KOORDINATOR
D300	EKSTERNAL	LAIN-LAIN (TUNGGU ACC PEMESAN)	03-09-2024 20:41:48	03-09-2024 21:58:09	77	KOORDINATOR
D300	EKSTERNAL	LAIN-LAIN (TEST PRINT(KOORDINATOR))	03-09-2024 21:58:28	03-09-2024 22:12:43	14	KOORDINATOR
D300	EKSTERNAL	LAIN-LAIN (TUNGGU INFORMASI (GANTI ORDER))	03-09-2024 22:13:40	03-09-2024 22:50:42	37	KOORDINATOR
D300	EKSTERNAL	PRINTING UNIT (ROLL)	03-09-2024 23:16:40	03-09-2024 23:27:39	11	TEKNIK
D300	EKSTERNAL	LAIN-LAIN (TEST PRINT(KOORDINATOR))	03-09-2024 23:30:21	03-09-2024 23:48:53	18	KOORDINATOR
D300	EKSTERNAL	LAIN-LAIN (TUNGGU INFORMASI (GANTI ORDER))	03-09-2024 23:49:16	04-09-2024 00:18:10	29	KOORDINATOR
D300	EKSTERNAL	FOLDER (LIPATAN)	04-09-2024 00:18:30	04-09-2024 03:12:07	174	TEKNIK

(2) *Data Processing/cleaning*

Dalam tahapan preprocessing yaitu menghilangkan noise data yang tidak konsisten, data *error* atau data kosong. Dalam tahap ini dilakukan beberapa hal. Pertama, penambahan atribut part untuk memberikan jumlah kelas yang terukur. Kedua, menghilangkan atribut Pic/jenis dikarenakan terdapat data yang kosong atau tidak diisi Ketika proses perbaikan mesin (Tabel 2).

Tabel 2. Data Cleaning

Mesin	Hambatan	Selisih	Jenis
D30 B	LAIN-LAIN (TEST PRINT(KOORDINATOR))	1	KOORDINATOR
D300	LAIN-LAIN (TUNGGU INFORMASI (GANTI ORDER))	53	KOORDINATOR
D300	LAIN-LAIN (TUNGGU INFORMASI (GANTI KERTAS))	7	KOORDINATOR
D300	LAIN-LAIN (TUNGGU INFORMASI (GANTI ORDER))	61	KOORDINATOR
D300	LAIN-LAIN (TEST PRINT(KOORDINATOR))	29	KOORDINATOR
D300	LAIN-LAIN (BON PLATE) - AUS : 1	11	REPRO
D300	LAIN-LAIN (TEST PRINT(KOORDINATOR))	6	KOORDINATOR
D300	SPLICER (REM KURANG OPTIMAL)	22	TEKNIK
D300	FOLDER (LIPATAN)	336	TEKNIK
D300	LAIN-LAIN (TUNGGU ACC PEMESAN)	2	KOORDINATOR
D300	LAIN-LAIN (TUNGGU ACC PEMESAN)	28	KOORDINATOR
D300	LAIN-LAIN (TUNGGU INFORMASI (GANTI KERTAS))	24	KOORDINATOR
D300	LAIN-LAIN (TUNGGU ACC PEMESAN)	25	KOORDINATOR
D300	FOLDER (LIPATAN)	122	TEKNIK
D300	LAIN-LAIN (TEST PRINT(KOORDINATOR))	9	KOORDINATOR
D300	PRINTING UNIT (MOTOR ROLL AIR RUSAK)	32	TEKNIK
D300	STUCKER (STACKER MACET)	16	TEKNIK
D300	LAIN-LAIN (TUNGGU ACC PEMESAN)	147	KOORDINATOR
D300	LAIN-LAIN (TUNGGU ACC PEMESAN)	77	KOORDINATOR
D300	LAIN-LAIN (TEST PRINT(KOORDINATOR))	14	KOORDINATOR
D300	LAIN-LAIN (TUNGGU INFORMASI (GANTI ORDER))	37	KOORDINATOR
D300	PRINTING UNIT (ROLL)	11	TEKNIK

(3) *Tranformation data*

Data akan berubah atau Bersatu menjadi bentuk yang tepat untuk menambang dengan ringkasan performa atau operasi agresi. Adapun dalam tahap ini data akan diubah menjadi bentuk yang sesuai untuk proses data mining.

data yang telah melalui proses sebelumnya akan ditransporasikan agar sesuai dengan algoritma yang dipakai yaitu *Naïve Bayes classifier*. Dalam tahapan ini attribute yang akan dipakai akan diberi label mengikuti kondisi data-data pada attribute tersebut (Tabel 3).

- Klasifikasi *attribute problem*
Berdasarkan hasil dataset yang siap diolah terdapat jenis problem dari *break down* mesin.
- Klasifikasi *attribute part*
Attribute part dikelompokkan menjadi 8 kategori :
- Klasifikasi attribute 558ingkat resiko
Tingkat resiko dikelompokkan menjadi 3 kategori yaitu rendah, sedang, tinggi
- Klasifikasi *attribute preventive maintenance*
- *Attribute preventive maintenance* didasarkan pada 2 keputusan : “Ya” dan “Tidak”.

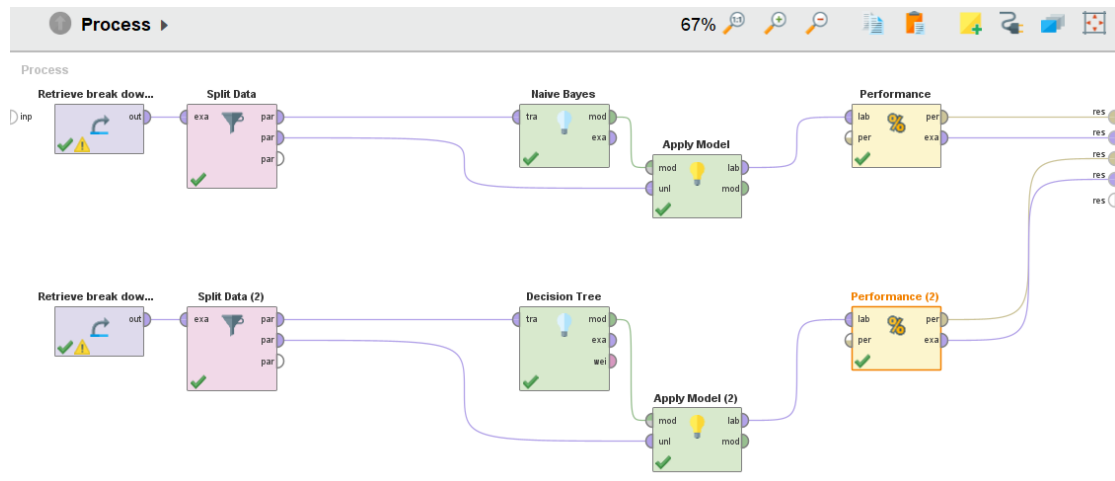
Tabel 3. Transfromasi Data

Mesin	Hambatan	PART	WASTE TIME	tindakan representativ	tingkat resiko
D30 B	LAIN-LAIN (TEST PRINT(KOORDINATOR))	LAIN-LAIN	1	tidak	rendah
D300	LAIN-LAIN (TUNGGU INFORMASI (GANTI ORDER))	LAIN-LAIN	53	tidak	rendah
D300	LAIN-LAIN (TUNGGU INFORMASI (GANTI KERTAS))	LAIN-LAIN	7	tidak	rendah
D300	LAIN-LAIN (TUNGGU INFORMASI (GANTI ORDER))	LAIN-LAIN	61	tidak	sedang
D300	LAIN-LAIN (TEST PRINT(KOORDINATOR))	LAIN-LAIN	29	tidak	rendah
D300	LAIN-LAIN (BON PLATE) - AUS : 1	LAIN-LAIN	11	ya	rendah
D300	LAIN-LAIN (TEST PRINT(KOORDINATOR))	LAIN-LAIN	6	tidak	rendah
D300	SPLICER (REM KURANG OPTIMAL)	SPLICER	22	tidak	rendah
D300	FOLDER (LIPATAN)	FOLDER	336	tidak	tinggi
D300	LAIN-LAIN (TUNGGU ACC PEMESAN)	LAIN-LAIN	2	tidak	rendah
D388 A	LAIN-LAIN (BERSIH BERSIH MESIN (HARI SENIN) (KOORDINATOR))	LAIN-LAIN	74	ya	sedang
D388 A	LAIN-LAIN (TUNGGU INFORMASI (GANTI KERTAS))	LAIN-LAIN	18	tidak	rendah
D388 A	LAIN-LAIN (MESIN ERROR)	LAIN-LAIN	22	tidak	rendah
D388 A	LAIN-LAIN (TEST PRINT(KOORDINATOR))	LAIN-LAIN	27	tidak	rendah
D388 A	LAIN-LAIN (TUNGGU INFORMASI (GANTI KERTAS))	LAIN-LAIN	13	tidak	rendah
D388 A	LAIN-LAIN (BON PLATE) - AUS : 1	LAIN-LAIN	21	ya	rendah
D388 A	LAIN-LAIN (BERSIH BERSIH MESIN (HARI SENIN) (KOORDINATOR))	LAIN-LAIN	17	ya	rendah
D388 A	LAIN-LAIN (TEST PRINT(KOORDINATOR))	LAIN-LAIN	50	tidak	rendah
D388 A	LAIN-LAIN (TEST PRINT(KOORDINATOR))	LAIN-LAIN	13	tidak	rendah
D388 B	SPLICER (SPLICER RUSAK)	SPLICER	94	tidak	tinggi
D388 B	SPLICER (SPLICER RUSAK)	SPLICER	12	tidak	rendah

Pada label status akan dijadikan sebagai proses untuk melakukan proses eksekusi pada program *RapidMiner* sehingga harus digolongkan menjadi 2 kriteria yaitu YA dan TIDAK. Pada attribute dengan tingkat resiko rendah akan diberi label preventive maintenance YA dan TIDAK tergantung dengan ketentuan dari divisi maintenance. Jika kerusakan mesin dengan durasi lebih dari 60 menit dan kerusakan mesin berulang dengan kerusakan yang sama lebih dari 5 kali maka akan diberi label YA, jika tidak memenuhi ketentuan tersebut maka akan diberi label TIDAK. Sedangkan pada attribute dengan tingkat resiko sedang dan tinggi akan diberi label preventive maintenance YA. Setelah hasil proses klasifikasi tersebut, maka langkah selanjutnya agar dapat diolah dengan menggunakan software *data mining* yaitu *RapidMiner*. Data yang sudah dilakukan pada proses tahapan diatas ini adalah data yang sudah siap atau sudah bersih dari data *redundancy* dan data yang *loss* (kosong). Pada Tabel 3 merupakan dataset yang siap diolah untuk dilakukan proses data mining.

2.3. Pengolahan Data Mining

Pengolahan *data mining* dengan model klasifikasi *Naïve Bayes* dan *Decision Tree*, Pada Gambar 2 adalah pemodelan *Naïve Bayes* pada aplikasi *RapidMiner*. Design proses dengan rapid miner tersebut dalam *retrieve* data terdapat dua kali jenis data yaitu preventive dan tingkat resiko, dilakukan secara berulang. Pada operator split data dibagi 90% menjadi data training dan 10% menjadi data *testing*. Rasio data training dan data *testing* dipilih 90 : 10 bertujuan untuk melatih dan evaluasi model cepat. Data berjumlah besar juga cocok untuk penggunaan rasio ini, hasilnya menunjukkan bahwa rasio 90:10 dapat memberikan performa yang sangat baik dalam hal sensitivitas, spesifisitas, dan akurasi, tergantung pada model dan dataset yang digunakan [24].



Gambar 2. Pemodelan *Naïve Bayes* dan *Decision Tree*

3. HASIL DAN PEMBAHASAN

Hasil dari pengolahan menggunakan aplikasi rapid miner dengan pemodelan seperti pada bab sebelumnya, didapatkan *output* dari *RapidMiner* sebagai berikut :

3.1. Preventive mesin (ya & tidak)

Pada Gambar 3 adalah hasil dari pengolahan untuk hal preventive mesin :

accuracy: 77.33%

	true tidak	true ya	class precision
pred. tidak	232	68	77.33%
pred. ya	0	0	0.00%
class recall	100.00%	0.00%	

Gambar 3. Hasil Uji Preventif Model *Decision Tree*

Berdasar Gambar 3 dapat diketahui bahwa hasil akurasi prediksi yang tepat sebesar 77,33%, sedangkan prediksi yang tidak tepat sebesar 22,77%. Dimana 300 data testing terdapat 232 data mesin yang dapat dilakukan *preventive maintenance* berhasil diprediksi dengan tepat dan sebanyak 68 data tidak dapat diprediksi dengan tepat. *Precision* adalah jumlah data yang *true positive* (jumlah data positif yang dikenali secara benar sebagai positif) dibagi dengan jumlah data yang dikenali sebagai positif). Dari hasil pengujian nilai *precision* yaitu 77,33%. *Recall* adalah jumlah data yang *true positive* dibagi dengan jumlah data yang sebenarnya positif (*true positive + true negative*). Dari hasil pengujian nilai *recall* yaitu 0,00%.

accuracy: 98.33%

	true tidak	true ya	class precision
pred. tidak	228	1	99.56%
pred. ya	4	67	94.37%
class recall	98.28%	98.53%	

Gambar 4. Hasil Uji Preventif Model *Naive Bayes*

Berdasar Gambar 4 dapat diketahui bahwa hasil akurasi prediksi yang tepat sebesar 99,56%, sedangkan prediksi yang tidak tepat sebesar 0,44%. Dimana 300 data testing terdapat 395 data mesin yang dapat dilakukan *preventive maintenance* berhasil diprediksi dengan tepat dan sebanyak 5 data tidak dapat diprediksi dengan tepat. *Precision* adalah jumlah data yang *true positive* (jumlah data positif yang dikenali secara benar sebagai positif) dibagi dengan jumlah data yang dikenali sebagai positif). Dari hasil pengujian nilai *precision* yaitu 98,33%. *Recall* adalah jumlah data yang *true positive* dibagi dengan jumlah data yang sebenarnya positif (*true positive + true negative*). Dari hasil pengujian nilai *recall* yaitu 98,53%.

3.2. Tingkat Resiko (Tinggi, Sedang, Rendah)

Berikut hasil dari pengolahan untuk hal tingkat resiko :

accuracy: 100.00%

	true rendah	true sedang	true tinggi	class precision
pred. rendah	245	0	0	100.00%
pred. sedang	0	16	0	100.00%
pred. tinggi	0	0	38	100.00%
class recall	100.00%	100.00%	100.00%	

Gambar 5. Hasil Uji Tingkat Resiko Model *Decision Tree*

Berdasar Gambar 5 dapat diketahui bahwa hasil akurasi prediksi yang tepat sebesar 100%, sedangkan prediksi yang tidak tepat sebesar 0%. Dimana 300 data *testing* terdapat 300 data mesin yang dapat diketahui Tingkat resiko berhasil diprediksi dengan tepat dan sebanyak 0 data tidak dapat diprediksi dengan tepat. *Precision* adalah jumlah data yang true positive (jumlah data positif yang dikenali secara benar sebagai positif) dibagi dengan jumlah data yang dikenali sebagai positif). Dari hasil pengujian nilai precision yaitu 100%. *Recall* adalah jumlah data yang true positive dibagi dengan jumlah data yang sebenarnya positif (*true positive + true negative*). Dari hasil pengujian nilai *recall* yaitu 100%.

accuracy: 97.32%

	true rendah	true sedang	true tinggi	class precision
pred. rendah	244	2	0	99.19%
pred. sedang	0	13	4	76.47%
pred. tinggi	1	1	34	94.44%
class recall	99.59%	81.25%	89.47%	

Gambar 6. Hasil Uji Tingkat Resiko Model Naive Bayes

Berdasar Gambar 6 dapat diketahui bahwa hasil akurasi prediksi yang tepat sebesar 94,44%, sedangkan prediksi yang tidak tepat sebesar 5,66%. Dimana 300 data *testing* terdapat 296 data mesin yang dapat diketahui Tingkat resiko berhasil diprediksi dengan tepat dan sebanyak 22 data tidak dapat diprediksi dengan tepat. *Precision* adalah jumlah data yang *true positive* (jumlah data positif yang dikenali secara benar sebagai positif) dibagi dengan jumlah data yang dikenali sebagai positif). Dari hasil pengujian nilai precision yaitu 97,32%. *Recall* adalah jumlah data yang *true positive* dibagi dengan jumlah data yang sebenarnya positif (*true positive + true negative*). Dari hasil pengujian nilai *recall* yaitu 89,47%.

Berdasarkan hasil evaluasi model, dapat disimpulkan bahwa model *Naïve Bayes* menunjukkan performa yang unggul dalam prediksi preventif dengan dua kelas (ya & tidak). Model ini mencatatkan nilai akurasi sebesar 97,50%, presisi 93,87%, dan *recall* 97,07%, yang menunjukkan bahwa model ini sangat akurat dalam mengidentifikasi kondisi yang membutuhkan tindakan preventif. Tingginya nilai *recall* memperlihatkan bahwa model berhasil mendeteksi hampir semua kasus positif yang sebenarnya, sedangkan nilai presisi yang tinggi menunjukkan bahwa prediksi positif yang diberikan oleh model sebagian besar memang benar. Hal ini menjadikan *Naïve Bayes* lebih andal untuk skenario deteksi awal di mana kesalahan dalam melewatkan kasus positif (*false negative*) harus diminimalkan. Hal ini dapat disebabkan karena *Naïve Bayes* merupakan metode yang sederhana cocok dengan label biner [24].

Sebaliknya, model *Decision Tree* menunjukkan performa yang lebih rendah dalam konteks prediksi preventif. Dengan nilai akurasi sebesar 77%, meskipun memiliki presisi dan *recall* yang sama tinggi dengan *Naïve Bayes* (93,87% dan 97,07% secara berurutan), hasil ini mengindikasikan bahwa model ini membuat lebih banyak kesalahan dalam klasifikasi secara keseluruhan. Hal ini dapat disebabkan oleh kecenderungan *Decision Tree* untuk overfitting terhadap data pelatihan, terutama ketika jumlah kelas terbatas [25]. Maka dari itu, meskipun nilai presisi dan *recall* tampak kompetitif, akurasi keseluruhan model menjadi penentu bahwa *Naïve Bayes* lebih cocok digunakan dalam konteks preventif.

Namun demikian, ketika beralih pada prediksi tingkat risiko yang terdiri dari tiga kelas (rendah, sedang, tinggi), performa kedua model menunjukkan dinamika yang berbeda. Meskipun model *Naïve Bayes* tetap

mencatatkan nilai akurasi yang tinggi, yakni sebesar 97,55%, serta presisi 97,22%, nilai *recall*-nya turun menjadi 91,3%. Nilai *recall* yang lebih rendah ini menunjukkan bahwa meskipun prediksi positif yang dibuat sebagian besar benar (tingginya presisi), masih terdapat sejumlah data risiko yang tidak berhasil terklasifikasi dengan benar oleh model (*false negative*). Dalam konteks klasifikasi multikelas, hal ini menjadi kelemahan karena keterwakilan setiap kelas sangat penting untuk akurasi keseluruhan sistem.

Sementara itu, model *Decision Tree* memiliki hasil nilai akurasi 100%, nilai presisi 100% dan nilai *recall* 100%. *Decision Tree* menunjukkan keunggulan dalam menangani klasifikasi risiko yang memiliki tiga kategori. Kemampuannya untuk membentuk aturan keputusan yang jelas dan terstruktur berdasarkan nilai-nilai atribut memungkinkan *Decision Tree* lebih efektif dalam mengidentifikasi serta membedakan antara kelas risiko tinggi, sedang, dan rendah. Dengan struktur cabang yang eksplisit, *Decision Tree* dapat menangani perbedaan antar kelas secara lebih fleksibel, terutama jika terdapat ketidakseimbangan data di antara kategori tersebut. Hal ini membuatnya lebih adaptif dalam kasus prediksi kompleks yang memerlukan pembeda yang rinci antara berbagai tingkat risiko.

4. KESIMPULAN

Hasil evaluasi model dengan *Naïve Bayes* untuk prediksi preventif memiliki nilai akurasi 97,50% nilai presisi 93,87% dan nilai *recall* 97,07%. Hasil evaluasi model dengan *Decision Trees* untuk prediksi preventif memiliki nilai akurasi 77% nilai presisi 93,87% dan nilai *recall* 97,07%. Dalam uji preventif dengan 2 kelas (ya & tidak) model *Naïve Bayes* lebih baik hasilnya dari model *Decision Trees*. Hasil evaluasi model dengan *Naïve Bayes* untuk prediksi tingkat resiko memiliki nilai akurasi 97,55% nilai presisi 97,22% dan nilai *recall* 91,3%. Hasil evaluasi model dengan *Decision Tree* untuk prediksi tingkat resiko memiliki nilai akurasi 100% nilai presisi 100% dan nilai *recall* 100%. Dalam uji prediktif tingkat resiko dengan 3 kelas (tinggi, sedang dan rendah) model *Decision Tree* lebih baik. Perbandingan antara 2 model *Naïve Bayes* dan *Decision Trees* disimpulkan bahwa model *Naïve Bayes* lebih baik digunakan uji preventive dengan memiliki 2 kelas (ya & tidak), dibandingkan dengan *Decision Tree* lebih baik digunakan uji tingkat resiko dengan memiliki kelas 3 (rendah, sedang & tinggi). Dalam pemilihan model yang tepat harus disesuaikan dengan struktur data dan jumlah kelas yang digunakan dalam prediksi. Untuk klasifikasi biner seperti prediksi preventif (ya atau tidak), model *Naïve Bayes* lebih efektif karena kesederhanaannya dan kemampuannya untuk menangani distribusi probabilitas antar fitur. Namun, dalam klasifikasi multikelas seperti penilaian tingkat risiko (rendah, sedang, tinggi), *Decision Tree* memberikan hasil yang lebih representatif dan menyeluruh. Hasil uji ini dapat dijadikan acuan perusahaan khususnya divisi maintenance dapat melakukan penjadwalan prediktif maintenance, sehingga produktivitas perusahaan dapat meningkat Evaluasi akan dilakukan untuk menilai akurasi dan efektivitas model dalam mendeteksi kerusakan lebih awal. Hasil penelitian diharapkan dapat meningkatkan efisiensi pemeliharaan dan mengurangi biaya *downtime* produksi.

DAFTAR PUSTAKA

- [1] M. F. J. Pratiwi, S. Yuniarti, S. Sunardi, and E. Y. Prihantono, "The operational process efficiency and profitability on food and beverage companies in Indonesia," *J. Penelit.*, vol. 20, no. 2, pp. 84–94, 2023, doi: 10.26905/jp.v20i2.12751.
- [2] E. D. Putri and H. Herlina, "Analisis Kelayakan Investasi Mesin Printing Percetakan Offset untuk Meningkatkan Kapasitas Hasil Output pada PT. XYZ," *J. Tek. Ind. Terintegrasi*, vol. 7, no. 2, pp. 921–932, 2024, doi: 10.31004/jutin.v7i2.27850.
- [3] D. D. Candra and W. Setiafindari, "Analisis efektivitas kinerja mesin printing SM-52 untuk produk buku tahunan," *Jumantara J. Manaj. dan Teknol. Rekayasa*, vol. 3, no. 1, pp. 33–42, 2024, doi: 10.28989/jumantara.v3i1.1769.
- [4] Arif Rahman and S. Perdana, "ANALISIS PRODUKTIVITAS MESIN PERCETAKAN PERFECT BINDING DENGAN METODE OEE DAN FMEA Arif Rahman dan Surya Perdana," *J. Ilm. Tek. Ind.*, vol. 7, no. 1, pp. 34–42, 2019.
- [5] P. Studi, T. Industri, F. Teknik, O. E. Effectiveness, and S. Tools, "Menurunkan *Downtime* Pada Mesin Slitter Dengan Menggunakan Metode Quality Control Circle (QCC) dan Overall Equipment Effectiveness (OEE) Di Perusahaan Manufaktur Makanan Abstrak," vol. 4, pp. 291–307, 2022.
- [6] A. Haseng and M. G. Sono, "Analisis Perawatan Preventif Dalam Penggantian Komponen Mesin Untuk Meminimumkan *Downtime* Metode Reliability Centered Maintenance (Rcm) Di Pt . Pln (Persero) Ulp Banggai," no. 1, pp. 1–11, 2024.

- [7] I. A. Marie, L. Hakim, D. Sugiarto, and W. Septiani, "Analisis Data Gangguan Kerusakan Mesin Produksi Menggunakan Teknik Association Rules," *J. Ilm. Tek. Ind.*, vol. 7, no. 1, pp. 43–52, 2019, doi: 10.24912/jitiuntar.v7i1.5035.
- [8] I. H. Witten, E. Frank, C. Hall, and J. Pal, *Data mining Practical Machine learning Tools and Techniques*. Morgan Kaufmann, 2016.
- [9] I. Gede Iwan Sudipa *et al.*, *DATA MINING*. [Online]. Available: www.globaleksekutifteknologi.co.id
- [10] Ardiyansyah, P. A. Rahayuningsih, and R. Maulana, "Analisis Perbandingan Algoritma Klasifikasi *Data mining* Untuk Dataset Blogger Dengan Rapid Miner," *J. Khatulistiwa Inform.*, vol. VI, no. 1, pp. 20–28, 2018.
- [11] Sekar Setyaningtyas, B. Indarmawan Nugroho, and Z. Arif, "Tinjauan Pustaka Sistematis: Penerapan *Data mining* Teknik Clustering Algoritma K-Means," *J. Teknoif Tek. Inform. Inst. Teknol. Padang*, vol. 10, no. 2, pp. 52–61, 2022, doi: 10.21063/jtif.2022.v10.2.52-61.
- [12] D. Muchlinski, "Machine learning and deep learning," *Elgar Encycl. Technol. Polit.*, pp. 114–118, 2022, doi: 10.4337/9781800374263.machine.learning.deep.
- [13] G. Mariscal, Ó. Marbán, and C. Fernández, "A survey of *data mining* and knowledge discovery process models and methodologies," *Knowl. Eng. Rev.*, vol. 25, no. 2, pp. 137–166, 2010, doi: 10.1017/S0269888910000032.
- [14] M. I. Kamil and A. P. Wibowo, "Implementation of the Naive Bayes Classifier Algorithm for Classifying Toddler Nutritional Status," vol. 8, no. 2, pp. 421–427, 2024.
- [15] N. Agustina, D. H. Citra, W. Purnama, C. Nisa, and A. R. Kurnia, "Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 1, pp. 47–54, 2022, doi: 10.57152/malcom.v2i1.195.
- [16] A. Nugroho and Y. Religia, "Analisis Optimasi Algoritma Klasifikasi Naive Bayes menggunakan Genetic Algorithm dan Bagging," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 3, pp. 504–510, 2021, doi: 10.29207/resti.v5i3.3067.
- [17] D. A. Mukhsinin, M. Rafliansyah, S. A. Ibrahim, R. Rahmadden, and D. Wulandari, "Implementasi Algoritma *Decision Tree* untuk Rekomendasi Film dan Klasifikasi Rating pada Platform Netflix," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 4, no. 2, pp. 570–579, 2024, doi: 10.57152/malcom.v4i2.1255.
- [18] Islamia Nuraini, Arnowo Hari Wibowo, Muhammad Asrol, Taufik, and Dyah Lang Wilis, "ANALISIS PEMELIHARAAN PREDIKTIF KENDARAAN OPERASIONAL MENGGUNAKAN ALGORITMA *NAÏVE BAYES* DAN *DECISION TREE*," *J. Publicuho*, vol. 7, no. 2, pp. 591–603, May 2024, doi: 10.35817/publicuho.v7i2.412.
- [19] M. Y. Saputra and I Wayan Santiyasa, "Optimizing *Machine learning* Performance with The Naive Bayes Classifier Process in Smart Farming," *J. Nas. Pendidik. Tek. Inform.*, vol. 13, no. 2, pp. 382–393, 2024, doi: 10.23887/janapati.v13i2.76926.
- [20] D. A. Muthia, "Opinion Mining Pada Review Buku Menggunakan Algoritma Naive Bayes," *J. Tek. Komput.*, vol. 2, no. 1, pp. 1–8, 2016.
- [21] M. Iqbal, S. Miskiyah, S. L. Sham, S. Anwar, and M. H. Fuad, "Perbandingan Metode *Decision Tree* dan Naive Bayes Pada Tingkat Penjualan Minuman Kopi di Kopi Pawon Nusantara," *J. Insa. J. Inf. Syst. Manag. Innov.*, vol. 4, no. 1, pp. 27–34, 2024, doi: 10.31294/jinsan.v4i1.3682.
- [22] S. Pramanik, D. Banik, D. Bhattacharjee, M. Nasipuri, M. K. Bhowmik, and G. Majumdar, "Suspicious-Region Segmentation From Breast Thermogram Using DLPE-Based Level Set Method," *IEEE Trans. Med. Imaging*, vol. 38, no. 2, pp. 572–584, 2019, doi: 10.1109/TMI.2018.2867620.
- [23] N. Afkhami-Jeddi, T. Hartman, and A. Tajdini, "Fast conformal bootstrap and constraints on 3d gravity," *J. High Energy Phys.*, vol. 2019, no. 5, pp. 1–31, 2019, doi: 10.1007/JHEP05(2019)087.
- [24] I. O. Muraina, "Ideal Dataset Splitting Ratios in *Machine learning* Algorithms: General Concerns for Data Scientists and Data Analysts," *7th Int. Mardin Artuklu Sci. Res. Conf.*, no. February, pp. 496–504, 2022, [Online]. Available: https://www.researchgate.net/publication/358284895_IDEAL_DATASET_SPLITTING_RATIOS_IN_MACHINE_LEARNING_ALGORITHMS_GENERAL_CONCERNS_FOR_DATA_SCIENTISTS_AND_DATA_ANALYSTS
- [25] A. Adam, M. I. Shapiai, Z. Ibrahim, M. Khalid, and L. Jau, "Development of a hybrid Artificial Neural Network - Naive Bayes classifier for binary classification problem of imbalanced datasets," *ICIC Express Lett.*, vol. 5, pp. 3171–3175, 2011.
- [26] M. Bramer, "Using J-pruning to reduce overfitting in classification trees," *Knowledge-Based Syst.*, vol. 15, no. 5, pp. 301–308, 2002, doi: [https://doi.org/10.1016/S0950-7051\(01\)00163-0](https://doi.org/10.1016/S0950-7051(01)00163-0).