



Implementasi Algoritma Naive Bayes Pada Sistem *Daily Report* Gde (PT Geo Dipa Energi) Untuk Klasifikasi Kerusakan Peralatan

Imam Subekti¹, Nahar Mardiyantoro², Nulngafan³

^{1,2,3} Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Sains Al-Qur'an, Indonesia
imamsubekti@gmail.com^{1*}, nahar@unsiq.ac.id², affan@unsiq.ac.id³

Abstrak

PT Geo Dipa Energi (Persero) sebagai perusahaan yang bergerak di bidang energi panas bumi memerlukan sistem pelaporan yang akurat untuk mendukung kegiatan pemeliharaan dan pengawasan peralatan pembangkit listrik. Tingginya volume laporan harian teknisi serta variasi deskripsi kerusakan menyebabkan proses identifikasi tingkat kerusakan secara manual menjadi kurang efisien dan berisiko menimbulkan keterlambatan analisis. Penelitian ini bertujuan mengimplementasikan algoritma *Multinomial Naive Bayes* pada sistem *Daily Report Instrumentation and Control* (I&C) untuk melakukan klasifikasi otomatis tingkat kerusakan peralatan ke dalam tiga kategori, yaitu Ringan, Sedang, dan Berat. Data yang digunakan berjumlah 4.408 entri laporan harian periode Juni 2022 hingga Mei 2025 yang diperoleh melalui studi literatur, observasi, wawancara, dan dokumentasi. Data teks diproses melalui tahapan *case folding*, tokenisasi, *stopword removal*, dan stemming, kemudian dibobotkan menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) dengan kombinasi unigram dan bigram. Mengingat distribusi kelas yang tidak seimbang, data latih diseimbangkan menggunakan metode *Random Oversampling* sebelum dilatih dengan algoritma *Naive Bayes*. Evaluasi model menggunakan metrik akurasi, *precision*, *recall*, F1-Score, dan Macro F1-Score. Hasil pengujian menunjukkan model memperoleh akurasi sebesar 96,49% dan Macro F1-Score sebesar 0,82, melampaui target minimal 70% yang ditetapkan. Meskipun demikian, *precision* kelas Sedang tercatat rendah yaitu 0,33 akibat pergeseran *decision boundary* yang ditimbulkan oleh *Random Oversampling* berbasis replikasi identik, sehingga sistem lebih tepat digunakan sebagai alat bantu pendukung keputusan yang tetap memerlukan validasi teknisi. Sistem *Daily Report* berbasis *website* yang dikembangkan mampu meningkatkan efektivitas pengelolaan laporan pemeliharaan serta mendukung keandalan operasional pembangkit listrik di PT Geo Dipa Energi (Persero).

Kata Kunci: Naive Bayes, Daily Report, Klasifikasi Teks, Kerusakan Peralatan, TF-IDF, Macro F1-Score, Data Tidak Seimbang.

1. Pendahuluan

PT Geo Dipa Energi (Persero) merupakan perusahaan yang bergerak di bidang energi panas bumi serta memiliki peran strategis dalam penyediaan energi bersih di Indonesia. Aktivitas operasional seperti pemeliharaan, pengawasan, dan pengelolaan peralatan pembangkit listrik membutuhkan sistem pelaporan yang akurat agar kondisi peralatan dapat dipantau secara konsisten. Volume laporan harian yang diterima setiap hari menimbulkan kendala dalam proses identifikasi jenis kerusakan karena banyaknya variasi deskripsi dari teknisi lapangan. Situasi tersebut memicu keterlambatan analisis, ketidaktepatan prioritas perbaikan, serta peningkatan risiko *downtime* operasional. Ketergantungan pada proses manual memperlihatkan kebutuhan akan sistem yang dapat mengelola laporan secara cepat dan terstruktur.

Penelitian sebelumnya menunjukkan efektivitas *Naive Bayes* dalam tugas klasifikasi teks. Penelitian yang dilakukan oleh Handayani dan Susanto (2023) memperoleh akurasi sebesar 92,4%, presisi 90,8%, serta *recall* 91,2% dalam klasifikasi laporan kerja berbasis teks. Penelitian yang dilaksanakan oleh Aininnisa dan Nuryana (2023) juga menunjukkan bahwa algoritma *Naive Bayes* dapat diterapkan untuk mengklasifikasikan jenis pelaporan pada sistem helpdesk berbasis web, sedangkan penelitian yang dilaksanakan oleh Putra dkk. (2021) menerapkan *Naive Bayes* untuk memprediksi risiko kerusakan mesin industri berbasis web. Temuan-temuan tersebut memperlihatkan kemampuan metode *Naive Bayes* dalam mengolah dokumen teknis maupun laporan operasional. Namun demikian, penelitian-penelitian tersebut belum mencakup penerapan pada sektor energi panas bumi yang memiliki karakteristik laporan berbeda, seperti istilah teknis khusus, struktur penulisan yang variatif, serta pola kerusakan yang bergantung pada kondisi peralatan pembangkit. Ketidakesesuaian ini menunjukkan bahwa hasil penelitian sebelumnya belum cukup memadai untuk diterapkan secara langsung pada lingkungan operasional PT Geo Dipa Energi.

Kemajuan teknologi memberikan peluang untuk mengotomatisasi proses klasifikasi laporan kerusakan melalui penerapan metode komputasi cerdas. Algoritma *Naive Bayes* dipilih karena mampu mengolah data berbasis teks dengan metode probabilistik yang sederhana namun akurat dalam mengenali pola kata teknis. Sistem *Daily Report* GDE (PT Geo Dipa Energi) berpotensi memperoleh manfaat signifikan melalui pengelompokan otomatis laporan

teknisi ke dalam kategori kerusakan ringan, sedang, atau berat. Informasi hasil klasifikasi dapat disajikan dalam bentuk dashboard interaktif yang memudahkan teknisi, supervisor, dan manajemen dalam menentukan tindakan korektif serta prioritas kerja.

Data *Daily Report Instrumentation and Control* (I&C) yang dihimpun selama periode Juni 2022 hingga Mei 2025 berjumlah 4.772 entri, yang setelah melalui tahap pembersihan data tersisa 4.408 entri valid. Distribusi label pada data tersebut sangat tidak seimbang, yaitu kelas Ringan sebanyak 4.272 data (96,91%), kelas Sedang sebanyak 133 data (3,02%), dan kelas Berat hanya 3 data (0,07%). Ketimpangan distribusi ini menjadi tantangan tersendiri dalam membangun model klasifikasi yang mampu mengenali kelas minoritas secara memadai, sehingga diperlukan strategi penyeimbangan data serta metrik evaluasi yang tepat, seperti Macro F1-Score, agar performa model tidak hanya diukur dari akurasi global yang cenderung bias terhadap kelas mayoritas.

Sebagian besar data *Daily Report* tersedia dalam bentuk teks tidak terstruktur yang memerlukan teknik *Natural Language Processing* (NLP) dan Text Mining agar dapat diproses oleh algoritma machine learning. Data teks laporan diproses melalui tahapan *case folding*, tokenisasi, stopword removal, dan stemming, kemudian direpresentasikan secara numerik menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) dengan kombinasi unigram dan bigram guna menangkap konteks lokal antarkata, seperti frasa "tidak normal" atau "gagal total" yang maknanya berbeda apabila katanya diperlakukan terpisah. Algoritma *Multinomial Naive Bayes* dipilih karena efektif dalam menangani data teks berdimensi tinggi, memiliki kompleksitas komputasi yang rendah, serta cocok digunakan pada sistem klasifikasi berbasis frekuensi kata atau TF-IDF, sehingga sesuai diterapkan untuk mengklasifikasikan laporan harian teknisi berbasis teks pada sistem *Daily Report Instrumentation and Control* (I&C).

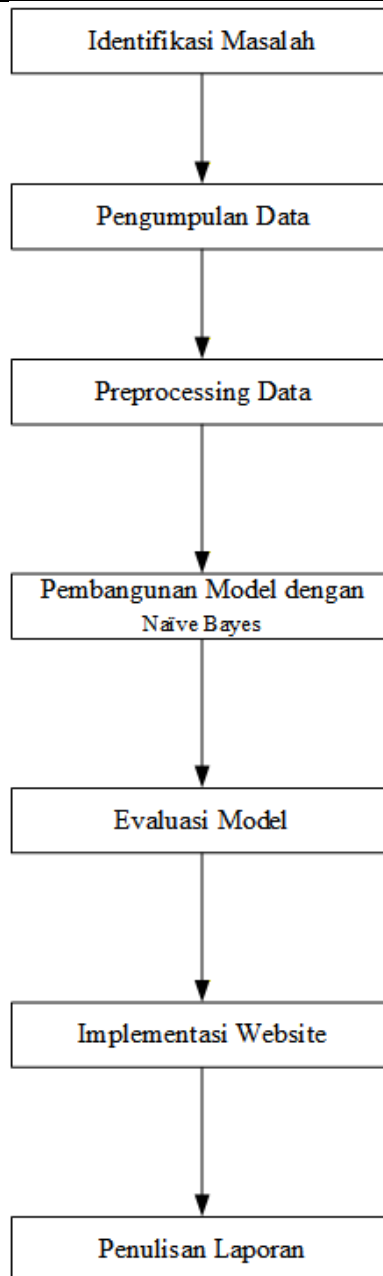
Berdasarkan uraian tersebut, penelitian ini berfokus pada pengembangan sistem klasifikasi kerusakan peralatan berbasis algoritma *Naive Bayes* yang terintegrasi dengan data aktual *Daily Report GDE* (PT Geo Dipa Energi), dengan target Macro F1-Score minimal 70% serta nilai precision dan recall setiap kelas di atas 65% sebagai pendukung pengambilan keputusan operasional..

2. Metode Penelitian

a. Jenis Penelitian

Penelitian ini termasuk dalam penelitian terapan (*applied research*) dengan pendekatan kuantitatif. Penelitian terapan dipilih karena hasil penelitian tidak hanya menghasilkan konsep atau kajian teoritis, tetapi berupa produk sistem berbasis *website* yang dapat diterapkan secara langsung pada lingkungan kerja GDE (PT Geo Dipa Energi). Pendekatan kuantitatif digunakan karena penelitian ini melibatkan pengolahan data terstruktur dan tidak terstruktur dalam bentuk numerik, penerapan algoritma *machine learning*, serta pengukuran kinerja model secara objektif menggunakan metrik akurasi, *precision*, *recall*, dan *f1-score*. Hasil pengujian model dianalisis secara kuantitatif untuk menilai tingkat ketepatan dan keandalan sistem dalam mengklasifikasikan tingkat kerusakan peralatan ke dalam kategori Ringan, Sedang, dan Berat.

b. Alur Penelitian



Gambar 2.1 Alur Penelitian

Penelitian ini dimulai dengan identifikasi masalah, yaitu bagaimana mengimplementasikan algoritma *Multinomial Naive Bayes* berbasis teks pada sistem *Daily Report Instrumentation and Control (I&C)* untuk melakukan klasifikasi otomatis tingkat kerusakan peralatan. Selanjutnya dilakukan pengumpulan data melalui studi literatur, observasi lapangan, wawancara, dan dokumentasi laporan harian teknisi. Data yang diperoleh kemudian melalui tahap data *cleaning*, *preprocessing* teks, serta pembobotan fitur menggunakan TF-IDF. Pada tahap pengembangan model, algoritma *Naive Bayes* dibangun dan dilatih menggunakan data latih yang telah diseimbangkan dengan *Random Oversampling*. Tahap berikutnya adalah evaluasi model menggunakan data uji untuk mengukur akurasi, *precision*, *recall*, dan *f1-score*, sehingga dapat diidentifikasi kekuatan dan kelemahan model. Hasil klasifikasi selanjutnya diimplementasikan ke dalam sistem *Daily Report* berbasis *website*, dan tahap akhir adalah penyusunan laporan yang mendokumentasikan keseluruhan proses penelitian, dari pendahuluan, landasan teori, metodologi, hasil dan pembahasan, hingga kesimpulan dan saran.

c. Metode Pengumpulan Data

Metode pengumpulan data yang digunakan dalam penelitian ini meliputi studi literatur, observasi lapangan, wawancara, dan dokumentasi. Studi literatur dilakukan dengan menelaah jurnal ilmiah, buku referensi, serta hasil penelitian terdahulu yang berkaitan dengan algoritma *Naive Bayes* dan teknik klasifikasi teks guna memperoleh

landasan teori yang kuat. Observasi lapangan dilakukan secara langsung di PT Geo Dipa Energi (Persero) untuk memahami proses pelaporan harian teknisi yang sedang berjalan, khususnya dalam pencatatan kondisi dan kerusakan peralatan. Wawancara dilakukan dengan teknisi, supervisor, dan staf terkait untuk memperoleh informasi mengenai kebutuhan pengguna serta kriteria klasifikasi laporan yang diharapkan agar sistem yang dibangun sesuai dengan kondisi nyata di lapangan. Metode dokumentasi digunakan untuk mengumpulkan data laporan harian teknisi yang tersimpan dalam bentuk arsip digital maupun fisik. Dataset mentah yang digunakan dalam penelitian ini berjumlah 4.772 data laporan harian dengan periode pengambilan data mulai dari Juni 2022 hingga Mei 2025. Setelah melalui proses pembersihan data, jumlah data yang valid dan siap digunakan untuk pelatihan dan pengujian model berjumlah 4.408 data.

d. Metode Analisis Data

Metode analisis data dalam penelitian ini dilakukan melalui beberapa tahapan sistematis untuk membangun model klasifikasi laporan kerusakan peralatan menggunakan algoritma *Naive Bayes*. Tahap pertama adalah data *preprocessing*, yaitu pembersihan dan normalisasi data teks laporan harian teknisi yang meliputi *case folding*, tokenisasi, *stopword removal*, dan *stemming* menggunakan *library* PySastrawi. Tahap berikutnya adalah pembobotan fitur menggunakan metode TF-IDF dengan kombinasi unigram dan bigram, yang mengubah data teks hasil preprocessing menjadi representasi numerik. Dataset kemudian dibagi menjadi data latih (80%) dan data uji (20%) sebelum ekstraksi fitur dilakukan untuk menghindari data *leakage*. Mengingat distribusi kelas pada data latih sangat tidak seimbang, sebelum tahap pelatihan model dilakukan proses penyeimbangan data menggunakan metode *Random Oversampling* pada data latih, yang bekerja dengan menambah jumlah sampel pada kelas minoritas melalui replikasi identik. Algoritma *Naive Bayes* kemudian diterapkan pada data latih yang telah seimbang untuk membentuk model klasifikasi, yang selanjutnya digunakan untuk memprediksi kategori laporan pada data uji. Tahap akhir adalah evaluasi model menggunakan metrik akurasi, *precision*, *recall*, f1-score, dan confusion matrix untuk menilai tingkat keandalan dan efektivitas model.

3. Hasil dan Pembahasan

a. Pengolahan Data

Daily Report Instrumentation and Control (I&C) berasal dari dokumentasi kegiatan harian teknisi yang memuat informasi mengenai tanggal pelaksanaan pekerjaan, uraian aktivitas, jenis pekerjaan, peralatan yang ditangani, status pekerjaan, serta catatan kondisi peralatan. Dataset mentah awal berjumlah 4.772 entri yang masih mengandung ketidakkonsistenan penulisan, perbedaan penggunaan huruf kapital, keberadaan entri hari libur dan kegiatan non-I&C, serta baris kosong tanpa uraian pekerjaan. Tahapan data cleaning mencakup penghapusan baris tanpa uraian pekerjaan, penghapusan entri hari libur dan kegiatan non-operasional, penghapusan kegiatan yang tidak berkaitan dengan *Instrumentation and Control (I&C)*, serta normalisasi penulisan kolom Hari. Setelah seluruh tahapan cleaning selesai, jumlah data yang valid dan siap digunakan adalah 4.408 entri dari total 4.772 entri awal.

Proses pelabelan data dilakukan berdasarkan aturan yang dirancang melalui konsultasi dengan teknisi senior di PT Geo Dipa Energi (Persero). Label yang digunakan terdiri atas tiga kelas: Ringan untuk kegiatan *Preventive* dan *Predictive* dengan kondisi peralatan normal, Sedang untuk kegiatan *Corrective*, *Modification*, atau *Refurbishment* yang mengindikasikan adanya perbaikan aktif, dan Berat untuk kondisi *emergency*, *shutdown*, *blackout*, atau kegagalan total peralatan. Distribusi label pada dataset hasil *cleaning* dan pelabelan ditampilkan pada tabel berikut.

Kelas	Jumlah Data	Persentase
Ringan	4.272	96,91%
Sedang	133	3,02%
Berat	3	0,07%
Total	4.408	100%

Tabel 3. 1 Distribusi Label Data Daily Report Instrumentation and Control (I&C)

Kolom Uraian Pekerjaan yang telah diproses melalui *case folding*, tokenisasi, *stopword removal*, dan *stemming* selanjutnya dipisahkan menjadi data latih dan data uji dengan perbandingan 80% dan 20% sebelum dilakukan ekstraksi fitur, guna menghindari data *leakage*. Dari total 4.408 data, sebanyak 3.526 data digunakan untuk pelatihan model dan 882 data digunakan untuk pengujian. Ekstraksi fitur dilakukan menggunakan TF-IDF *Vectorizer* dengan jumlah maksimal 500 fitur dan konfigurasi *ngram_range*=(1,2) untuk menangkap kombinasi unigram dan bigram, sehingga frasa seperti “tidak normal”, “gagal total”, dan “kebocoran aktif” dapat direpresentasikan sebagai satu unit fitur yang bermakna.

#	A	B	C	D	E	F	G	H	I
1	Hari	Tanggal	Uraian Pekerjaan	Jenis Pekerjaan	Equipme Status	Keterangan Label	Label_Nama		
2	KAMIS	2022-06-02 00:00:00	Pemeriksaan status, kondisi dan pembersihan serta pengecatan jika diperlukan pressure tran	Preventive	Turbin	Complete Peralatan	0 Ringan		
3	KAMIS	2022-06-02 00:00:00	Pemeriksaan status, kondisi dan pembersihan serta pengecatan jika diperlukan pressure tran	Preventive	Generato	Complete Peralatan	0 Ringan		
4	KAMIS	2022-06-02 00:00:00	Pemeriksaan status, kondisi dan pembersihan serta pengecatan jika diperlukan pressure tran	Preventive	UPS	Complete Peralatan	0 Ringan		
5	KAMIS	2022-06-02 00:00:00	Pemeriksaan visual panel, membersihkan Air filter	Preventive	DCS Cabin	Complete Peralatan	0 Ringan		
6	KAMIS	2022-06-02 00:00:00	Pemeriksaan visual panel, membersihkan Air filter	Preventive	TCS Cabin	Complete Peralatan	0 Ringan		
7	KAMIS	2022-06-02 00:00:00	Pemeriksaan visual panel, membersihkan Air filter	Preventive	Excitasi C	Complete Peralatan	0 Ringan		
8	KAMIS	2022-06-02 00:00:00	Pemeriksaan status, kondisi dan pembersihan serta pengecatan jika diperlukan pressure tran	Preventive	FWDP	Complete Peralatan	0 Ringan		
9	KAMIS	2022-06-02 00:00:00	Pemeriksaan status, kondisi dan pembersihan serta pengecatan jika diperlukan pressure tran	Preventive	Turbin	Complete Peralatan	0 Ringan		
10	KAMIS	2022-06-02 00:00:00	Pemeriksaan status, kondisi dan pembersihan serta pengecatan jika diperlukan pressure tran	Preventive	Generato	Complete Peralatan	0 Ringan		
11	JUMAT	2022-06-03 00:00:00	Pemeriksaan status, kondisi dan pembersihan serta pengecatan jika diperlukan pressure tran	Preventive	V-101	Complete Peralatan	0 Ringan		
12	JUMAT	2022-06-03 00:00:00	Pemeriksaan status, kondisi dan pembersihan serta pengecatan jika diperlukan pressure tran	Preventive	MSV-102	Complete Peralatan	0 Ringan		
13	JUMAT	2022-06-03 00:00:00	Pemeriksaan status, kondisi dan pembersihan serta pengecatan jika diperlukan pressure tran	Preventive	K-102A	Complete Peralatan	0 Ringan		
14	JUMAT	2022-06-03 00:00:00	Pemeriksaan status, kondisi dan pembersihan serta pengecatan jika diperlukan pressure tran	Preventive	K-102B	Complete Peralatan	0 Ringan		
15	JUMAT	2022-06-03 00:00:00	Pemeriksaan status, kondisi dan pembersihan serta pengecatan jika diperlukan pressure tran	Preventive	Main stea	Complete Peralatan	0 Ringan		
16	SENIN	2022-06-06 00:00:00	Pemeriksaan status, kondisi dan pembersihan serta pengecatan jika diperlukan level transm	Preventive	Cooling Ti	Complete Peralatan	0 Ringan		
17	SENIN	2022-06-06 00:00:00	Pemeriksaan status, kondisi dan pembersihan serta pengecatan jika diperlukan level transm	Preventive	Cooling Ti	Complete Peralatan	0 Ringan		
18	SENIN	2022-06-06 00:00:00	Kalibrasi temperatur gauge TI-4716	Preventive	Temperat	Complete Peralatan	1 Sedang		
19	SENIN	2022-06-06 00:00:00	Kalibrasi temperatur gauge TI-3014	Preventive	Temperat	Complete Peralatan	1 Sedang		
20	SENIN	2022-06-06 00:00:00	Membanu team IKPT Kalibrasi Vacuum Gauge	Corrective	Vacum Ga	Complete Peralatan	1 Sedang		
21	SELASA	2022-06-07 00:00:00	Pemeriksaan status, kondisi dan pembersihan serta pengecatan jika diperlukan pressure gau	Preventive	Blowdown	Complete Peralatan	0 Ringan		
22	SELASA	2022-06-07 00:00:00	Pemeriksaan status, kondisi dan pembersihan serta pengecatan jika diperlukan control valve,	Preventive	After Co	Complete Peralatan	0 Ringan		
23	SELASA	2022-06-07 00:00:00	Pemeriksaan status, kondisi dan pembersihan serta pengecatan jika diperlukan control valve,	Preventive	Inter Con	Complete Peralatan	0 Ringan		
24	SELASA	2022-06-07 00:00:00	Pemeriksaan status, kondisi dan pembersihan serta pengecatan jika diperlukan, Pressure trar	Preventive	Control Vi	Complete Peralatan	0 Ringan		

Gambar 3. 1 Contoh Hasil Preprocessing Dataset

b. Implementasi Algoritma *Naive Bayes*

Implementasi algoritma *Naive Bayes* dilakukan sebagai tahap inti dalam proses pembangunan sistem klasifikasi laporan harian pada bagian *Instrumentation and Control (I&C)*. Algoritma yang digunakan adalah *Multinomial Naive Bayes*, karena secara teoritis dan empiris paling sesuai untuk menangani permasalahan klasifikasi data berbasis teks yang direpresentasikan dalam bentuk bobot TF-IDF. Proses implementasi dimulai dengan inisialisasi model, pelatihan menggunakan data latih yang telah diseimbangkan melalui *Random Oversampling*, kemudian dilanjutkan dengan prediksi terhadap data uji sehingga kinerja model dapat dianalisis dan dievaluasi.

c. Hasil Evaluasi

Proses evaluasi dilakukan pada data uji sebesar 20% dari total dataset, yaitu sebanyak 882 entri yang terdiri atas 866 data kelas Ringan, 15 data kelas Sedang, dan 1 data kelas Berat. Hasil pengujian menunjukkan nilai akurasi sebesar 96,49%, yang berarti sebagian besar data uji berhasil diprediksi dengan benar oleh model. Namun demikian, angka akurasi yang tinggi ini perlu diinterpretasikan secara hati-hati mengingat distribusi kelas yang sangat tidak seimbang dalam dataset, dengan kelas Ringan mendominasi 96,91% dari seluruh data. Rincian performa model pada setiap kelas ditampilkan pada tabel berikut.

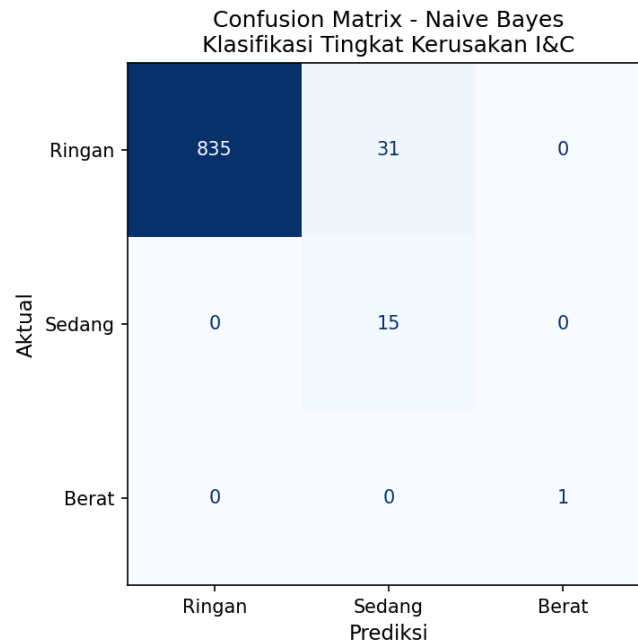
Kelas	Precision	Recall	F1-Score	Jumlah Data Uji
Ringan	1,00	0,96	0,98	866
Sedang	0,33	1,00	0,49	15
Berat	1,00	1,00	1,00	1
Macro avg	-	-	0,82	882
Weighted avg	-	-	0,97	882

Tabel 3. 2 Hasil Evaluasi Model *Naive Bayes*

	precision	recall	f1-score	support
Ringan (0)	1.00	0.96	0.98	866
Sedang (1)	0.33	1.00	0.49	15
Berat (2)	1.00	1.00	1.00	1
accuracy			0.96	882
macro avg	0.78	0.99	0.82	882
weighted avg	0.99	0.96	0.97	882

Gambar 3. 2 Classification Report Hasil Evaluasi

Nilai macro average f1-score sebesar 0,82 dihitung sebagai rata-rata aritmetika F1-Score dari ketiga kelas secara setara, yaitu $(0,98 + 0,49 + 1,00) / 3 = 0,82$. Nilai ini telah melampaui target Macro F1-Score minimal 70% yang ditetapkan dalam tujuan penelitian, dan merupakan indikator yang jauh lebih jujur dibandingkan akurasi global karena secara otomatis menghukum performa buruk pada kelas Sedang ($F1 = 0,49$). Namun demikian, nilai *precision* kelas Sedang sebesar 0,33 masih berada jauh di bawah ambang 65% yang ditargetkan, sehingga secara keseluruhan sistem belum sepenuhnya memenuhi kriteria keandalan pada kelas minoritas tersebut. Model juga belum dapat dievaluasi secara andal pada kelas Berat karena keterbatasan data uji yang hanya berjumlah 1 sampel, sehingga nilai sempurna 1,00 pada kelas tersebut secara metodologis tidak memiliki kekuatan statistik dan tidak dapat digeneralisasi.

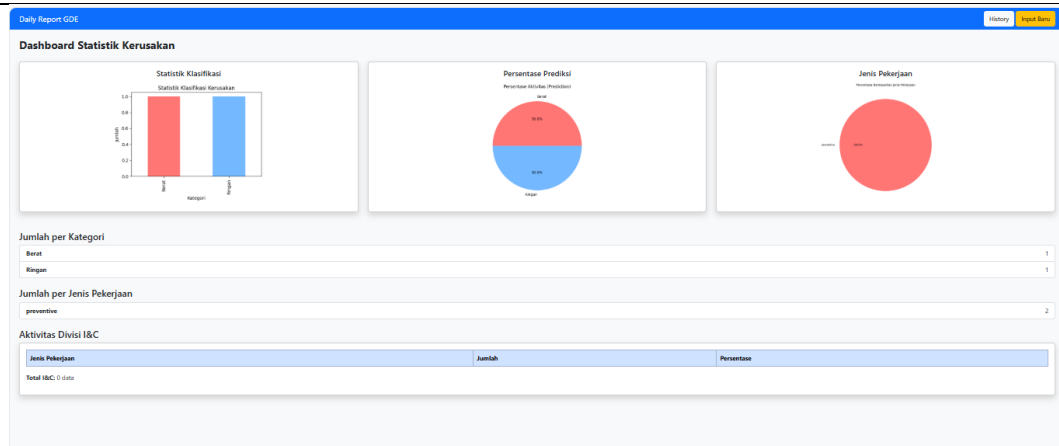


Gambar 3. 3 Confusion Matrix Hasil Pengujian

Confusion matrix menunjukkan bahwa kelas Ringan memiliki 835 prediksi benar dan 31 prediksi salah yang diklasifikasikan ke kelas Sedang, sedangkan kelas Sedang memiliki 15 prediksi benar tanpa kesalahan dan kelas Berat memiliki 1 prediksi benar. Tidak ditemukan kesalahan klasifikasi kelas Ringan ke kelas Berat, yang menunjukkan bahwa perbedaan karakteristik teks antara kondisi ringan dan berat cukup kontras dan mudah dibedakan oleh model. Nilai *precision* kelas Sedang yang rendah bukan semata-mata akibat tumpang tindih istilah teknis antar kelas, melainkan juga merupakan konsekuensi langsung dari penerapan *Random Oversampling* berbasis replikasi identik pada data latih, yang menggeser *decision boundary* model ke arah kelas Sedang sehingga model menjadi terlalu agresif dalam mengklasifikasikan sampel Ringan ke kelas Sedang. Dari perspektif operasional, kondisi ini berarti sistem terlalu sering memberikan peringatan kerusakan tingkat Sedang pada laporan yang sebenarnya hanya bersifat Ringan, yang dapat menyebabkan mobilisasi tim perbaikan yang tidak perlu. Untuk penelitian mendatang, disarankan penggunaan teknik oversampling yang lebih canggih seperti SMOTE (*Synthetic Minority Oversampling Technique*) yang menghasilkan sampel sintesis baru, serta mempertimbangkan algoritma yang lebih kebal terhadap ketidakseimbangan kelas seperti *Random Forest* atau *Support Vector Machine* (SVM) (Saepudin dkk., 2024; Fitriani dkk., 2021).

d. Implementasi Website

1) Halaman Dashboard



Gambar 3. 4 Halaman Dashboard

Halaman *Dashboard* menampilkan rangkuman statistik hasil klasifikasi kerusakan berdasarkan data *Daily Report*. Tampilan halaman memuat grafik batang untuk perbandingan jumlah kategori kerusakan, grafik pie untuk persentase tingkat kerusakan, serta grafik pie untuk distribusi jenis perbaikan. Bagian bawah halaman menyediakan tabel yang menampilkan informasi jumlah aktivitas berdasarkan kategori dan jenis pekerjaan sebagai referensi cepat bagi pengguna.

2) Halaman History

The 'History' page, titled 'Riwayat Aktivitas', includes a search bar and filters for 'Jenis Pekerjaan' and 'Filter Equipment'. The table below lists activities:

Mari	Tanggal	Uraian	Jenis	Equipment	Status	Keterangan	Prediction	Aksi
kamis	2025-12-04	turbine macet	preventive	Turbine	Completed	turbine mengalami macet tidak bergerak	Benar	Detail
Selasa	2025-12-02	Pemeriksaan rutin bulanan dan pembersihan sensor	preventive	Pressure Transmitter PT 101	In Progress	Sensor dibersihkan, dilakukan pengecekan wiring, tidak ditemukan anomali.	Benar	Detail

Gambar 3. 5 Halaman History

Halaman *History* menampilkan daftar lengkap aktivitas harian yang telah melalui proses klasifikasi. Setiap baris memuat tanggal laporan, uraian pekerjaan, jenis pekerjaan, nama peralatan, prediksi tingkat kerusakan, hasil prediksi tindakan korektif, serta status validasi. Terdapat fitur pencarian dan filter yang dapat digunakan untuk mempersempit tampilan data berdasarkan kategori tertentu, serta tombol aksi untuk melihat detail, mengedit, dan menghapus data.

3) Halaman Input

Gambar 3. 6 Halaman Input

Halaman *Input* digunakan untuk memasukkan data aktivitas baru ke dalam sistem. Formulir yang tersedia memuat kolom untuk tanggal, uraian pekerjaan, jenis pekerjaan, nama peralatan, status kondisi, keterangan teknis, serta catatan tambahan. Setiap data yang disimpan akan diproses menggunakan model *Naive Bayes* untuk menghasilkan klasifikasi otomatis tingkat kerusakan, dan hasilnya akan ditampilkan pada halaman *History*.

4. Kesimpulan

Penelitian ini menunjukkan bahwa algoritma *Multinomial Naive Bayes* berbasis teks berhasil diimplementasikan pada sistem *Daily Report Instrumentation and Control* (I&C) PT Geo Dipa Energi (Persero) untuk melakukan klasifikasi otomatis tingkat kerusakan peralatan ke dalam tiga kelas, yaitu Ringan, Sedang, dan Berat. Melalui tahapan *preprocessing* teks yang meliputi data *cleaning*, *case folding*, tokenisasi, *stopword removal*, *stemming*, serta ekstraksi fitur menggunakan TF-IDF Vectorizer, sistem mampu memproses laporan harian dan menghasilkan prediksi kategori kerusakan secara efisien sebagai pendukung pengambilan keputusan operasional. Hasil evaluasi menunjukkan nilai akurasi sebesar 96,49% dan Macro F1-Score sebesar 0,82, yang telah melampaui target minimal 70% yang ditetapkan. Meskipun demikian, penerapan *Random Oversampling* berbasis replikasi identik menyebabkan pergeseran *decision boundary* yang menurunkan *precision* kelas Sedang menjadi 0,33, sedangkan performa kelas Berat belum dapat dievaluasi secara representatif karena hanya didasarkan pada satu sampel uji. Kondisi ini menegaskan bahwa sistem saat ini lebih tepat diposisikan sebagai alat bantu pendukung keputusan yang memerlukan validasi manual dari teknisi, bukan sebagai sistem klasifikasi otomatis penuh. Secara keseluruhan, sistem *Daily Report* berbasis website yang dikembangkan mampu meningkatkan efektivitas pengelolaan laporan pemeliharaan serta mendukung keandalan operasional pembangkit listrik di PT Geo Dipa Energi (Persero).

Reference

- Aininnisa, F., & Nuryana, I. K. D. (2023). Rancang Bangun Helpdesk Web Application Dengan Klasifikasi Jenis Pelaporan Menggunakan Algoritma Naive Bayes (Studi Kasus: PT Pelindo Daya Sejahtera). *JINACS (Journal Of Informatics And Computer Science)*, 5(2), 85–94.
- Fitriani, R. D., Yasin, H., & Tarno, T. (2021). Penanganan Klasifikasi Kelas Data Tidak Seimbang Dengan Random Oversampling Pada Naive Bayes (Studi Kasus: Status Peserta KB IUD Di Kabupaten Kendal). *Jurnal Gaussian*, 10(1), 11–20.
- Handayani, S., & Susanto, R. (2023). Klasifikasi Laporan Kerja Berbasis Teks Menggunakan Algoritma Naive Bayes. *Jurnal Sistem Informasi*, 19(1), 45–54.
- Kadir, A. (2014). *Pengenalan Sistem Informasi* (Edisi Revisi). Andi Offset.
- Kadir, A. (2015). *Pemrograman Python Untuk Pemula*. Andi Offset.
- Pardede, J., & Darmawan, D. (2025). Perbandingan Algoritma Stemming Porter, Sastrawi, Idris, Dan Arifin & Setiono Pada Dokumen Teks Bahasa Indonesia. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, 12(1), 69–76.
- Pressman, R. S. (2015). *Software Engineering: A Practitioner's Approach* (8th Ed.). Mcgraw-Hill Education.
- Putra, R. A., Sartika, D., & Wijaya, H. (2021). Sistem Prediksi Risiko Kerusakan Mesin Industri Menggunakan Naive Bayes Berbasis Web. *Jurnal Teknologi Dan Sistem Informasi*, 7(3), 210–218.
- Rosa, A. S., & Shalahuddin, M. (2018). Rekayasa Perangkat Lunak Terstruktur Dan Berorientasi Objek. *Informatika*.
- Saepudin, A., Faqih, A., & Dwilestari, G. (2024). Perbandingan Algoritma Klasifikasi Support Vector Machine, Random Forest Dan Logistic Regression Pada Ulasan Shopee. *Jurnal Tekno Kompak*, 18(1), 178–192.
- Septiani, D., & Isabela, I. (2022). Analisis Term Frequency Inverse Document Frequency (TF-IDF) Dalam Temu Kembali Informasi Pada Dokumen Teks. *SINTESIA: Jurnal Sistem Dan Teknologi Informasi Indonesia*, 1(2), 81–88.
- Setiawan, M. D., & Eliyani. (2022). Pengembangan Aplikasi Helpdesk Ticketing System Dengan Algoritma Multinomial Naive Bayes Classifier. *Jurnal Ilmiah FIFO*, 14(2), 123–131.