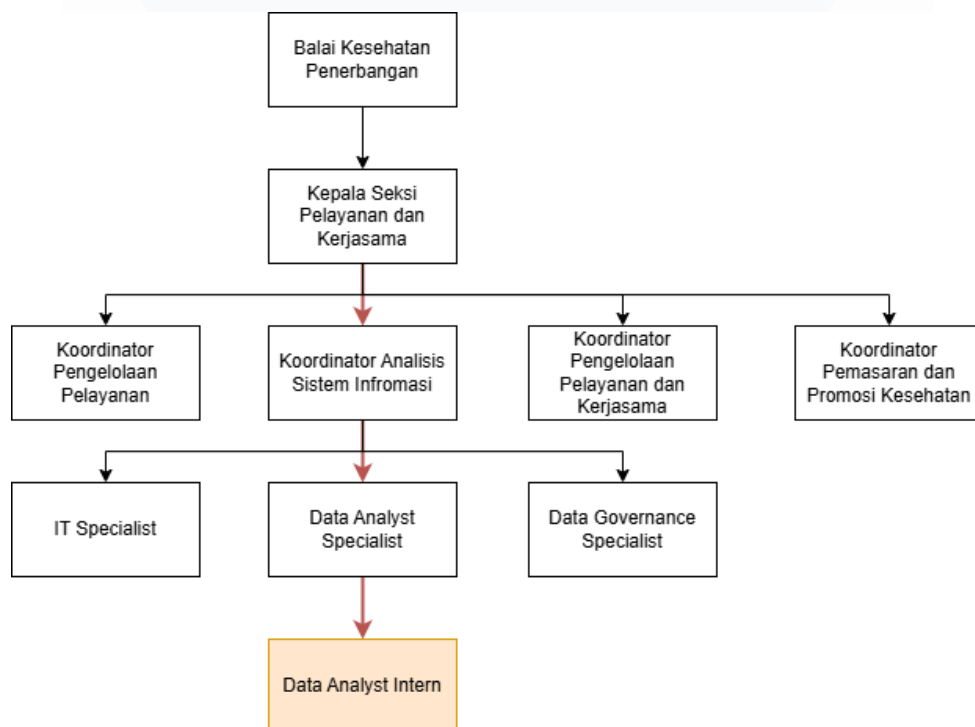


BAB III

PELAKSANAAN KERJA MAGANG

3.1 Kedudukan dan Koordinasi

Selama program magang di BLU Balai Kesehatan Penerbangan, posisi sebagai Data Analyst ditempatkan di Seksi Pelayanan dan Kerjasama, tepatnya di bagian Informasi Teknologi (IT). Dalam struktur organisasi, atasan langsung dalam unit ini adalah tim IT (Informasi Teknologi), sementara atasan tidak langsung adalah Koordinator Analis Sistem Informasi, dengan lingkup kerja yang berkoordinasi langsung dengan Kepala Seksi Pelayanan dan Kerjasama. Kedudukan ini memungkinkan adanya integrasi antara analisis data dan kebutuhan layanan kesehatan penerbangan, terutama dalam mendukung pengelolaan data survei kepuasan pelanggan serta optimalisasi proses digitalisasi layanan kesehatan.



Gambar 2.2 Struktur Organisasi Sub-divisi Data Analyst

Dalam menjalankan tugasnya, posisi ini memiliki alur koordinasi yang jelas yang dapat dilihat pada Gambar 2.2, dimulai dari penerimaan tugas yang diberikan oleh Koordinator Analis Sistem Informasi atau tim IT terkait kebutuhan analisis data tertentu. Setelah menerima tugas, proses selanjutnya adalah pengumpulan dan

pengolahan data yang bersumber dari survei kepuasan pelanggan dan berbagai sistem yang digunakan oleh Balai Kesehatan Penerbangan, seperti web app MEDEX yang menjadi platform utama dalam pencatatan hasil survei dari pengunjung. Data yang dikumpulkan dianalisis menggunakan dashboard interaktif dengan pemanfaatan *tool* Power BI untuk mengevaluasi sentimen serta mengidentifikasi isu-isu terkini berdasarkan data yang ada, bukan sekadar berdasarkan frekuensi kata.

Koordinasi dalam pekerjaan ini dilakukan secara berjenjang. Hasil analisis yang telah disusun dalam bentuk laporan atau dashboard akan dikomunikasikan terlebih dahulu dengan tim IT untuk dilakukan validasi dan evaluasi. Jika sudah sesuai, laporan tersebut akan diteruskan kepada Koordinator Analis Sistem Informasi sebelum akhirnya disampaikan ke Kepala Seksi Pelayanan dan Kerjasama untuk mendapatkan persetujuan dan pertimbangan lebih lanjut. Selain itu, koordinasi juga melibatkan komunikasi dengan berbagai tim di dalam organisasi, terutama yang terkait dengan pengambilan keputusan berbasis data di sektor layanan kesehatan penerbangan.

Selain tugas utama dalam analisis data survei kepuasan pelanggan, posisi ini juga berkontribusi dalam mendukung upaya peningkatan efisiensi sistem informasi di Balai Kesehatan Penerbangan, termasuk memastikan bahwa data yang digunakan dalam analisis memiliki kualitas yang baik serta membantu mengembangkan metode yang lebih optimal dalam penyajian hasil analisis. Dengan demikian, kedudukan dan koordinasi yang terjalin selama program magang tidak hanya memberikan pengalaman dalam analisis data, tetapi juga membantu dalam memahami bagaimana proses pengambilan keputusan berbasis data dilakukan di lingkungan pelayanan kesehatan penerbangan.

3.2 Tugas dan Uraian Kerja Magang

Sebagai Data Analyst di BLU Balai Kesehatan Penerbangan, peran yang dijalankan berfokus pada pengolahan dan analisis data untuk mendukung pengambilan keputusan berbasis informasi, terutama dalam meningkatkan kualitas layanan kesehatan penerbangan. Tugas utama mencakup pengumpulan,

pembersihan, analisis, serta visualisasi data survei kepuasan pelanggan yang dikumpulkan melalui *web app* MEDEX. Data ini kemudian diolah menjadi wawasan yang dapat membantu organisasi dalam memahami preferensi, keluhan, serta tingkat kepuasan masyarakat terhadap layanan yang diberikan. Dengan menyajikan data dalam bentuk laporan maupun *dashboard* interaktif, hasil analisis ini berperan dalam mendukung pengambilan keputusan berbasis bukti, terutama dalam mengidentifikasi permasalahan utama yang dihadapi pelanggan serta merancang strategi perbaikan layanan yang lebih efektif. Dalam praktiknya, peran ini juga mendukung upaya digitalisasi sistem informasi di BLU Balai Kesehatan Penerbangan dengan mengembangkan alat bantu yang mempermudah pemantauan data serta meningkatkan efisiensi dalam proses evaluasi layanan. Secara keseluruhan, kontribusi yang diberikan tidak hanya berdampak pada optimalisasi kinerja internal dalam pengelolaan data dan analisis informasi, tetapi juga berkontribusi secara langsung terhadap peningkatan pengalaman pelanggan dengan memastikan layanan kesehatan yang lebih responsif, transparan, dan berbasis kebutuhan nyata pengguna.



Tabel 3.1 Timeline Realisasi Agenda Data Analyst Intern

No.	Aktivitas	Minggu ke-	Tanggal Mulai Aktivitas	Tanggal Akhir Aktivitas	Urian Aktivitas
1.	Adaptasi & Pengumpulan Data Awal	3 – 5	20 Januari 2025	9 Februari 2025	
1.1	Pengenalan lingkungan kerja dan struktur organisasi	3	20 Januari 2025	26 Januari 2025	Mengikuti briefing awal, mengenal struktur organisasi, divisi, serta alur komunikasi internal di BLU Balai Kesehatan Penerbangan.
1.2	Pengenalan aplikasi MEDEX dan alur pelayanan	3	20 Januari 2025	26 Januari 2025	Mempelajari fungsi aplikasi MEDEX dalam proses pendaftaran dan pelayanan Medical Check-Up serta alur kerja digitalisasi layanan kesehatan.
1.3	Pengenalan alur data pendaftaran dan survei	4	27 Januari 2025	2 Februari 2025	Mengidentifikasi bagaimana data pasien direkam dari awal pendaftaran hingga pengisian survei, termasuk jenis data yang terkumpul.

No.	Aktivitas	Minggu ke-	Tanggal Mulai Aktivitas	Tanggal Akhir Aktivitas	Urian Aktivitas
1.4	Koordinasi akses data dan file master	4	27 Januari 2025	2 Februari 2025	Melakukan koordinasi dengan tim IT dan Kepala Analisis Sistem Informasi untuk mendapatkan akses ke file master data dan folder penyimpanan survei.
1.5	Pengumpulan data pendaftaran & survei, observasi struktur data	5	3 Februari 2025	9 Februari 2025	Mengumpulkan data mentah dari sistem dan memetakan struktur dataset (kolom, format isian, dan tipe data).
2.	Eksplorasi data & analisis deskriptif	6 – 7	10 Februari 2025	23 Februari 2025	
2.1	Eksplorasi data awal	6	10 Februari 2025	16 Februari 2025	Mengecek jumlah responden survei, distribusi skor kepuasan, serta mendokumentasikan jenis jawaban terbuka yang masuk dengan menggunakan Python.
2.2	Perhitungan rata-rata skor & NPS	6	10 Februari 2025	16 Februari 2025	Menghitung rata-rata skor survei per indikator dan menentukan nilai Net Promoter Score sebagai metrik awal kepuasan pelanggan.

No.	Aktivitas	Minggu ke-	Tanggal Mulai Aktivitas	Tanggal Akhir Aktivitas	Urian Aktivitas
2.3	Visualisasi deskriptif awal (Power BI)	7	17 Februari 2025	23 Februari 2025	Membuat grafik batang, pie chart, dan tabel interaktif di Power BI untuk menampilkan distribusi skor dan jumlah responden.
2.4	Pembangunan Model Machine Learning	7	17 Februari 2025	23 Februari 2025	Mengelompokkan dan menghitung keluhan yang paling sering muncul dari data jawaban terbuka, berdasarkan tema atau kata kunci.
3.	Text Preprocessing & Labeling Data	8 – 10	24 Februari 2025	16 Maret 2025	
3.1	Preprocessing teks (basic)	8	24 Februari 2025	28 Februari 2025	Melakukan pembersihan teks jawaban terbuka: menghilangkan stopwords, lowercase, serta lemmatization untuk standarisasi kata.
3.2	Preprocessing lanjutan dan labeling data	9 – 10	1 Maret 2025	16 Maret 2025	Membersihkan simbol, angka, dan tanda baca. Memberi label kepuasan (puas/netral/tidak puas) berdasarkan skor kuantitatif.

No.	Aktivitas	Minggu ke-	Tanggal Mulai Aktivitas	Tanggal Akhir Aktivitas	Urian Aktivitas
4.	Feature Extraction & Handling Imbalance	11 – 13	17 Maret 2025	6 April 2025	
4.1	TF-IDF untuk ekstraksi fitur teks	11	17 Maret 2025	23 Maret 2025	Mengonversi data teks ke bentuk numerik menggunakan teknik TF-IDF sebagai input model machine learning.
4.2	Cek sebaran kelas sentimen	11	17 Maret 2025	23 Maret 2025	Mengamati proporsi data per label sentimen untuk mendeteksi ketidakseimbangan antar kelas (imbalanced data).
4.3	SMOTE untuk balancing kelas	11 – 13	17 Maret 2025	6 April 2025	Menerapkan teknik SMOTE untuk menyeimbangkan data minoritas (tidak puas), sehingga model tidak bias terhadap kelas mayoritas.
5.	Pemodelan <i>Machine Learning</i>	13 – 14	31 Maret 2025	13 April 2025	

No.	Aktivitas	Minggu ke-	Tanggal Mulai Aktivitas	Tanggal Akhir Aktivitas	Urian Aktivitas
5.1	Training & evaluasi Logistic Regression	13	31 Maret 2025	6 April 2025	Melatih model klasifikasi Logistic Regression dan mengevaluasi performanya dengan F1-Score dan akurasi.
5.2	Training model LDA	13 – 14	31 Maret 2025	April 2025	Menerapkan Latent Dirichlet Allocation untuk mengelompokkan jawaban terbuka ke dalam topik-topik utama secara otomatis.
6.	Pembuatan Visualisasi Interaktif	11 - 15	17 Maret	9 Mei 2025	Pembuatan dashboard dengan pemanfaatan tool Excel untuk membersihkan data, seperti membenarkan format. Selain itu, menggunakan Power BI sebagai tool utama dalam melakukan dashboard Interaktif.
7.	Monitoring data dan handling data pelanggan	5 - 20	3 Februari 2025	16 Mei 2025	Menata dan memantau data pasien yang akan melakukan medical check-up tahunan melalui MEDEX, serta melakukan pencatatan harian.

Dalam menjalankan tugas selama program magang di BLU Balai Kesehatan Penerbangan, berbagai tools digunakan untuk mendukung analisis data serta pengembangan sistem yang berkaitan dengan survei kepuasan pelanggan dan digitalisasi layanan kesehatan.

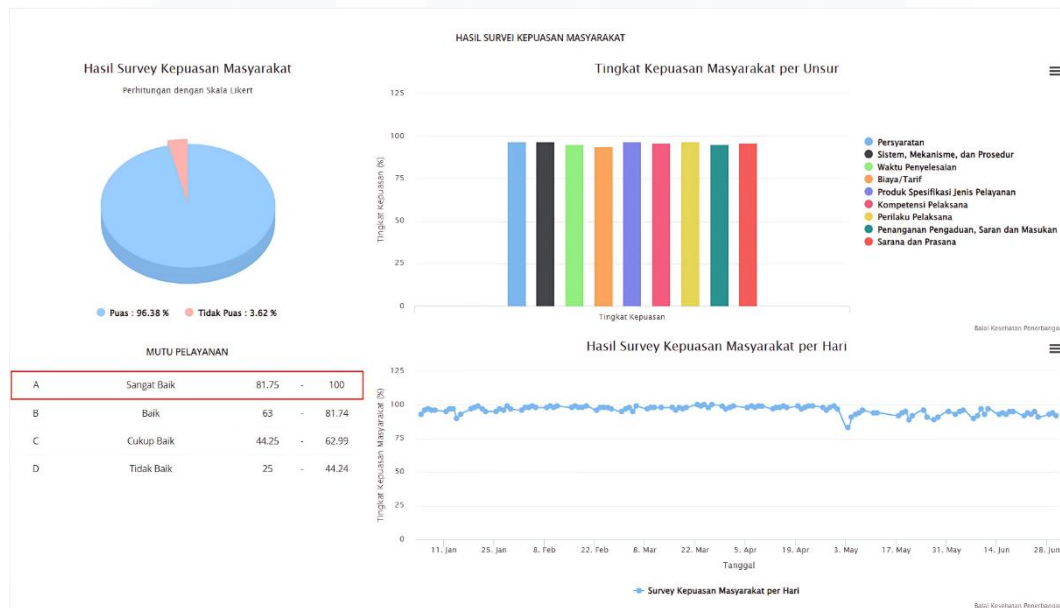
1. Microsoft Excel digunakan sebagai alat utama dalam pengelolaan data awal, terutama untuk mengambil dan merapikan hasil survei kepuasan pelanggan. Dengan Excel, data yang diperoleh dari *web app* MEDEX dapat diorganisir, difilter, serta disiapkan sebelum masuk ke tahap analisis lebih lanjut.
2. Python dimanfaatkan dalam proses pembersihan data dan pengolahan lebih lanjut, terutama untuk membangun model machine learning yang mampu menganalisis sentimen pelanggan. Melalui pemodelan *Natural Language Processing* (NLP), data survei diolah untuk mengidentifikasi pola sentimen serta mengurutkan permasalahan utama berdasarkan tingkat urgensinya.
3. Power BI digunakan untuk membuat dashboard interaktif yang membantu divisi keuangan dalam memahami pola transaksi serta tren yang berkaitan dengan layanan kesehatan penerbangan. Dengan Power BI, visualisasi data yang informatif dapat disajikan untuk mendukung pengambilan keputusan strategis.

3.2.1 Pengembangan dan Pemanfaatan dashboard interaktif

3.2.1.1 Dashboard interaktif pada Survei Kepuasan Masyarakat

Dalam rangka meningkatkan kualitas pelayanan dan memahami kebutuhan masyarakat secara lebih akurat, BLU Balai Kesehatan Penerbangan memanfaatkan dashboard interaktif berbasis web sebagai sarana visualisasi hasil survei kepuasan masyarakat. Dashboard ini dirancang untuk menyajikan data secara real-time dan informatif, sehingga mempermudah pengambilan keputusan oleh manajemen. Pemanfaatan dashboard menjadi sangat krusial mengingat tingginya volume data dari responden yang masuk setiap hari. Dengan tampilan visual yang ringkas namun informatif, pengguna internal dapat dengan mudah membaca tren, menemukan titik lemah layanan, serta memprioritaskan perbaikan. Dashboard ini juga menjembatani

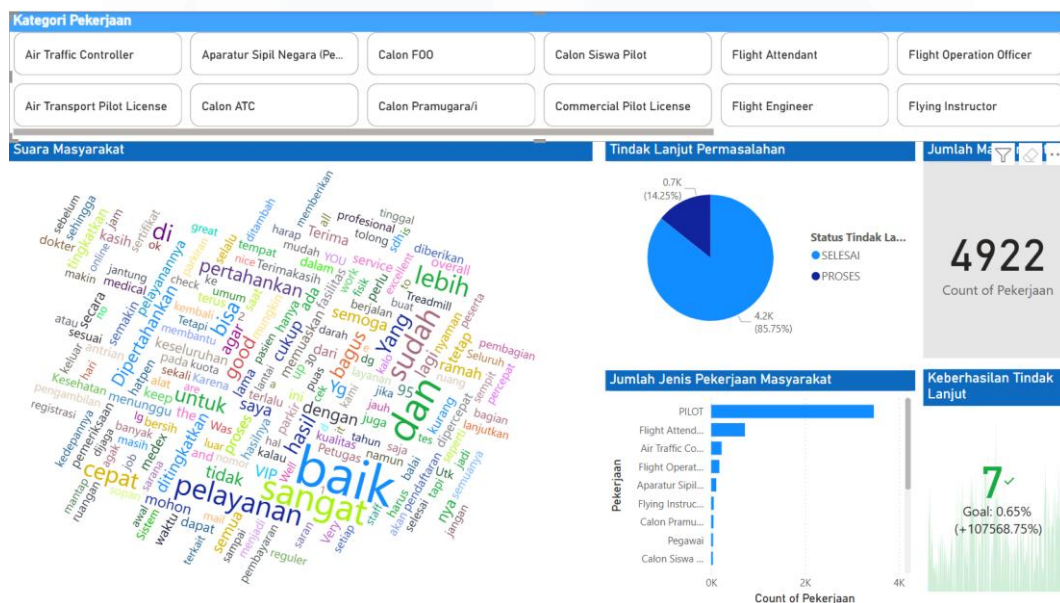
antara proses analitik dan eksekusi kebijakan, karena dapat diakses langsung oleh para penanggung jawab layanan. Salah satu wujud konkret dari upaya tersebut adalah dashboard hasil survei kepuasan masyarakat ini, yang secara langsung menjadi jawaban atas tantangan dalam memantau dan mengevaluasi performa pelayanan secara dinamis.



Gambar 3.1 Dashboard Interaktif Survei Kepuasan Masyarakat

Dashboard pada Gambar 3.1 menyajikan beberapa visualisasi penting yang saling melengkapi. Di sisi kiri atas, terdapat *pie chart* yang menunjukkan proporsi kepuasan masyarakat berdasarkan skala Likert yang telah diolah dengan hasil mayoritas responden menyatakan puas (96.38%) dan hanya 3.62% yang menyatakan tidak puas. Data tersebut diambil pada 21 Juli 2025 yang dapat diakses melalui laman Survey Hatpen, nilai tersebut dinilai berdasarkan rerata ulasan yang diisi oleh pasien setelah mereka melakukan seluruh pemeriksaan Kesehatan pada *website* MEDEX Hatpen. Di bawahnya, terdapat klasifikasi mutu pelayanan berdasarkan rentang nilai kepuasan, dibagi menjadi empat kategori: Sangat Baik, Baik, Cukup Baik, dan Tidak Baik, lengkap dengan skor intervalnya. Di sisi kanan atas, *bar chart* menunjukkan tingkat kepuasan masyarakat berdasarkan unsur pelayanan, seperti Persyaratan, Waktu Penyelesaian, Biaya Tarif, hingga Penanganan Pengaduan dan Sarana Prasarana sehingga tiap aspek layanan dapat

dievaluasi secara individual. Sementara itu, grafik di bagian bawah *dashboard* menampilkan tren hasil survei dari waktu ke waktu, memungkinkan pengguna untuk memantau fluktuasi harian yang mungkin disebabkan oleh perubahan operasional, kendala layanan, atau faktor musiman lainnya.



Gambar 3.2 *Dashboard* monitoring survei masyarakat

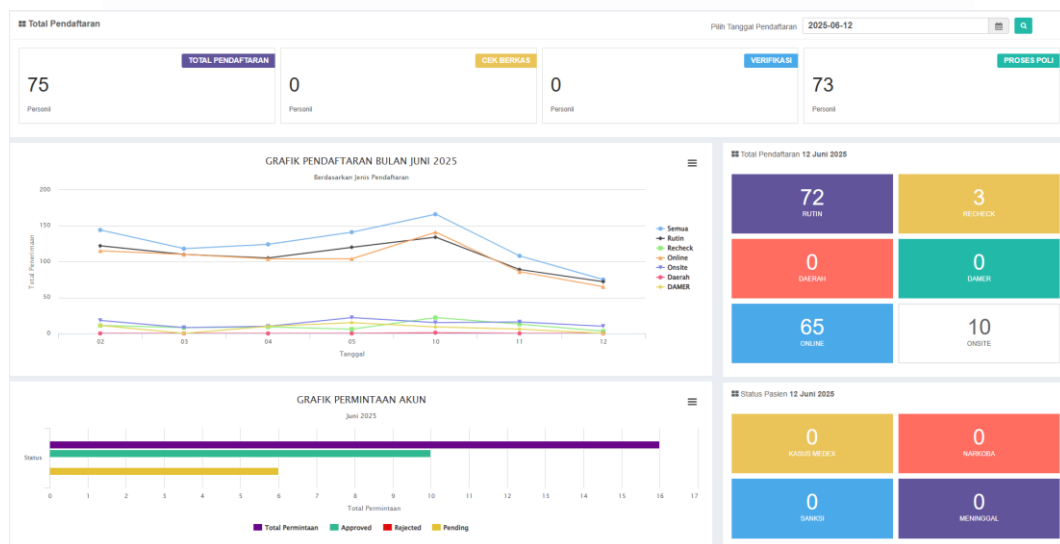
Dashboard pada Gambar 3.2 menunjukkan penggunaan *Word Cloud* untuk melihat secara lebih mendalam mengenai suara masyarakat berdasarkan jumlah kata terbanyak. Selain itu, *dashboard* ini juga dapat melihat proses dari tindak lanjut permasalahan yang diselesaikan. Dengan adanya visualisasi tersebut, memudahkan dalam jenis-jenis permasalahan yang belum diselesaikan.

Secara keseluruhan, *dashboard* ini tidak hanya berfungsi sebagai media pelaporan, tetapi juga sebagai alat monitoring strategis yang mendukung pendekatan berbasis data dalam meningkatkan kualitas pelayanan publik. Penyajian visual yang komprehensif dan mudah diinterpretasikan mempermudah tim internal dalam mengidentifikasi isu secara cepat dan objektif. Selain itu, *dashboard* ini memberikan peluang untuk evaluasi berkala yang lebih transparan dan terukur, sehingga BLU Balai Kesehatan Penerbangan dapat bertransformasi menjadi lembaga yang adaptif terhadap umpan balik masyarakat. Dengan adanya sistem ini, proses pengambilan keputusan dapat dilakukan lebih responsif dan berbasis bukti

nyata, serta menjadi bagian penting dalam penguatan sistem pelayanan yang berorientasi pada kepuasan pengguna jasa.

3.2.1.2 Pemanfaatan *dashboard* dalam memonitoring pengunjung

Sebagai bagian dari transformasi digital dan peningkatan efisiensi pelayanan, BLU Balai Kesehatan Penerbangan mengembangkan berbagai *dashboard* interaktif untuk mendukung proses monitoring dan pengambilan keputusan. *Dashboard* ini digunakan untuk menampilkan data pendaftaran pasien secara *real-time* yang meliputi berbagai tahapan seperti verifikasi, cek berkas, hingga proses ke poli. Tujuan dari penggunaan *dashboard* interaktif ini adalah untuk mengurangi proses manual dalam pelaporan dan membantu staf operasional maupun manajemen dalam melihat kondisi harian secara cepat dan akurat. Dalam praktiknya, *dashboard* ini menjadi salah satu solusi atas permasalahan pengolahan data yang sebelumnya bersifat statis dan tidak terintegrasi. Dengan adanya perubahan struktur data dan dinamika operasional harian, *dashboard* ini memberikan fleksibilitas serta kemudahan untuk beradaptasi terhadap perubahan sistem secara langsung.

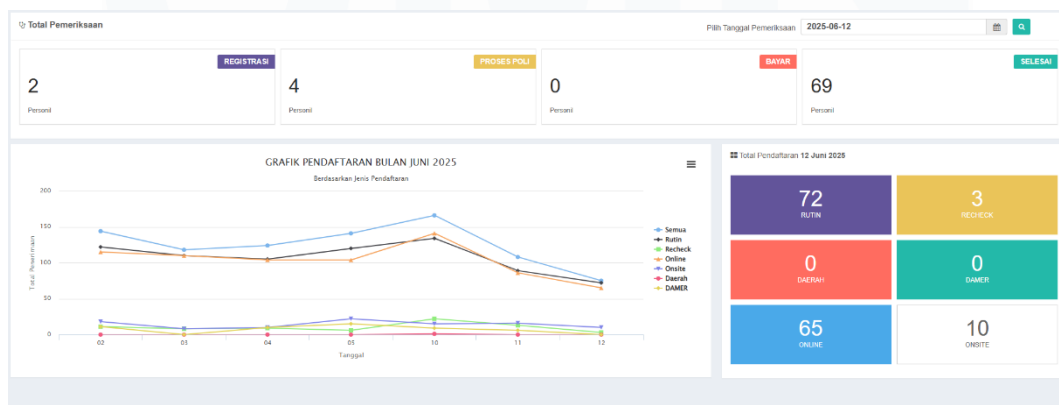


Gambar 3.3 *Dashboard* interaktif monitoring total pemeriksaan bagian Registrasi

Dashboard pada Gambar 3.3 menyajikan sejumlah visualisasi penting yang saling melengkapi dalam menampilkan aktivitas pendaftaran pasien. Di bagian atas,

pengguna dapat melihat rekap jumlah total pendaftar beserta status terkini seperti jumlah yang sudah diverifikasi, masih dalam cek berkas, atau yang sudah diproses ke poli. Tersedia pula fitur pemilihan tanggal untuk memantau data harian secara spesifik. Grafik di bagian tengah menampilkan tren pendaftaran selama bulan Juni 2025 (dapat diatur sesuai tanggal yang diinginkan) berdasarkan jenis layanan seperti rutin, recheck, onsite, dan online, memungkinkan analisis terhadap fluktuasi jumlah pasien. Di sisi kanan, panel jumlah pendaftaran pada tanggal tertentu dibedakan berdasarkan kategori layanan serta lokasi pemeriksaan (*onsite/online*), memberikan informasi yang lebih granular. Di bagian bawah, terdapat grafik permintaan akun serta status pasien pada hari yang sama yang berfungsi untuk memantau kebutuhan akun baru dan mendeteksi kategori khusus seperti kasus medex, narkoba, sanksi, atau yang meninggal.

Secara keseluruhan, *dashboard* ini tidak hanya memudahkan pelaporan operasional harian, tetapi juga memperkuat sistem monitoring pelayanan kesehatan berbasis data yang presisi. Integrasi visualisasi ini memungkinkan pihak internal untuk melakukan evaluasi secara cepat, mengidentifikasi lonjakan atau penurunan layanan, serta menyusun langkah tindak lanjut berdasarkan informasi yang terverifikasi. Dengan penyajian data yang terstruktur dan mudah dipahami, *dashboard* ini menjadi tulang punggung dalam mendukung pelayanan yang responsif, terukur, dan transparan. Pemanfaatannya juga memberikan kontribusi penting dalam menciptakan lingkungan kerja yang lebih adaptif terhadap data dan mendukung transformasi digital di BLU Balai Kesehatan Penerbangan.



Gambar 3.4 *Dashboard* interaktif untuk bagian Kasir

Dashboard pemeriksaan pada Gambar 3.4 terdiri dari beberapa elemen visual utama yang masing-masing memiliki fungsi spesifik dalam mendukung pengambilan keputusan. Di bagian atas, ringkasan jumlah pasien per tahapan proses, mulai dari registrasi, proses poli, pembayaran, hingga selesai yang memberikan gambaran singkat mengenai status pemeriksaan harian. Pemilihan tanggal memungkinkan pengguna untuk memantau pergerakan jumlah pasien berdasarkan waktu tertentu, meningkatkan fleksibilitas dalam analisis historis. Selanjutnya, grafik pendaftaran pasien berdasarkan jenis layanan selama bulan Juni 2025 memudahkan tim operasional untuk mengidentifikasi tren jumlah kunjungan dan lonjakan aktivitas pada kategori tertentu, seperti rutin, recheck, online, atau onsite. Visualisasi tambahan di sisi kanan menunjukkan distribusi jumlah pasien per jenis layanan pada tanggal terpilih, membantu dalam pengaturan sumber daya dan penjadwalan tenaga medis secara lebih efisien.

Secara keseluruhan, *dashboard* ini menawarkan pendekatan yang adaptif terhadap pemantauan proses pemeriksaan pasien secara menyeluruh. Tidak hanya memberikan transparansi terhadap kinerja layanan, khususnya bagian kasir, tetapi juga meningkatkan koordinasi antar unit dengan informasi yang tersedia secara *real-time*. Dengan visualisasi yang sederhana namun informatif, *dashboard* ini mampu menjembatani antara data teknis dan kebutuhan pengambilan keputusan yang cepat. Selain itu, pengembangan *dashboard* ini memperkuat budaya kerja berbasis data di lingkungan BLU Balai Kesehatan Penerbangan, yang pada akhirnya mendorong kualitas pelayanan kesehatan yang lebih responsif dan terukur.

3.2.1.3 Monitoring kasir dengan pemanfaatan *dashboard* interaktif

Sebagai bagian dari transformasi digital yang diterapkan di BLU Balai Kesehatan Penerbangan, pemanfaatan *dashboard* interaktif menjadi solusi utama dalam mengelola data pelayanan secara efisien dan transparan. Sistem pelaporan berbasis *dashboard* tidak hanya memudahkan dalam menyajikan informasi, tetapi juga mempercepat pengambilan keputusan berbasis data nyata. Salah satu contoh pemanfaatannya adalah dalam konteks pemantauan transaksi pembayaran, yang sebelumnya dilakukan secara manual atau terpisah antarunit. Keberadaan

dashboard ini menjawab tantangan penting, seperti kebutuhan akan sistem yang mampu merangkum informasi keuangan dan status pembayaran pasien secara terpusat dan dapat diakses secara real-time.



Gambar 3.5 *Dashboard* interaktif monitoring keuangan

Visualisasi pada *dashboard* Gambar 3.5 dirancang dengan berbagai komponen yang saling melengkapi. Di bagian utama terdapat tabel data yang menyajikan rincian informasi transaksi pasien, termasuk nama personil, nomor rekam medis, nomor kuitansi, dan total pembayaran yang telah dilakukan. Informasi ini sangat berguna dalam menelusuri detail pembayaran individu secara cepat. Di bagian bawah, terdapat ringkasan total pembayaran yang ditampilkan dalam bentuk nilai agregat penjumlahan (Rp248 juta) serta pembagian antara biaya awal dan potensi denda. Selain itu, *pie chart* di sisi kanan menampilkan proporsi status pembayaran antara “Hutang” dan “Tunai”, yang memudahkan tim keuangan dalam mengevaluasi tingkat kepatuhan pasien terhadap pembayaran yang berlaku.

Dengan adanya filter dinamis seperti pemilihan tanggal, nama personil, nomor rekam medis, dan maskapai, pengguna dapat menyaring data sesuai kebutuhan analisis. Fitur ini sangat penting dalam operasional harian, misalnya saat ingin menilai performa keuangan dari maskapai tertentu atau mengaudit riwayat pembayaran per individu. Interaktivitas ini memungkinkan *dashboard* menjadi alat

monitoring sekaligus audit yang efektif, mengurangi beban administrasi yang sebelumnya dilakukan secara terpisah antardivisi.

Secara keseluruhan, *dashboard* ini menjadi representasi nyata dari integrasi data pelayanan dan sistem keuangan dalam satu platform visual yang komprehensif. Kemampuan untuk menampilkan informasi secara terstruktur, cepat, dan dapat disesuaikan dengan kebutuhan analisis menjadikannya alat penting dalam menunjang transparansi dan akuntabilitas di lingkungan BLU Balai Kesehatan Penerbangan. Pendekatan ini tidak hanya meningkatkan efisiensi internal, tetapi juga membentuk budaya kerja berbasis data yang lebih adaptif dan responsif terhadap tantangan pelayanan publik.

3.2.2 Perancangan sistem klasifikasi sentiment berbasis *machine learning*

Kualitas layanan di sektor kesehatan penerbangan merupakan aspek vital yang memiliki implikasi langsung terhadap keselamatan dan kenyamanan operasional industri aviasi. Dalam lingkup pelayanan *Medical Check-Up* (MCU) bagi awak penerbangan, evaluasi kepuasan pelanggan menjadi indikator utama dalam menjamin mutu serta keberlangsungan layanan yang diberikan. Setiap pelanggan yang telah menjalani pemeriksaan diwajibkan mengisi survei digital sebagai bagian dari umpan balik pelayanan. Meskipun data tersebut telah terkumpul dalam jumlah besar dan bersifat rutin, proses analisis masih terbatas pada pengolahan deskriptif dasar seperti rata-rata skor survei tanpa eksplorasi lebih lanjut terhadap opini terbuka yang disampaikan pelanggan. Kondisi ini menimbulkan potensi kehilangan informasi penting yang terkandung dalam data teks, khususnya yang berkaitan dengan keluhan tersembunyi dan isu layanan yang bersifat kritis namun belum teridentifikasi secara sistematis.

Sebagai solusi terhadap keterbatasan tersebut, dikembangkan pendekatan analitik berbasis *machine learning* yang diarahkan untuk mengoptimalkan pemrosesan data survei pelanggan. Tujuan dari pengembangan sistem ini adalah melakukan klasifikasi otomatis terhadap sentimen pelanggan berdasarkan jawaban terbuka yang dikumpulkan dari survei digital. Proyek ini memanfaatkan pendekatan *Natural Language Processing* (NLP) dengan *pipeline* yang mencakup tahapan

preprocessing teks seperti pembersihan karakter non-alfabet, normalisasi huruf, penghapusan kata umum (*stopwords*), serta proses lemmatization untuk mendapatkan bentuk dasar kata. Setelah melalui tahapan *preprocessing*, teks yang telah dibersihkan diekstraksi fiturnya menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF), lalu dilanjutkan dengan tahap pelabelan data berdasarkan kategori kepuasan seperti puas, tidak puas, dan netral yang diperoleh dari skor survei. Untuk mengatasi ketidakseimbangan distribusi label dalam data, digunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) agar model pembelajaran dapat mengenali kelas minoritas secara lebih optimal.

Model klasifikasi utama yang digunakan dalam proyek ini adalah *Logistic Regression* karena efisiensinya dalam menangani data teks berdimensi tinggi hasil ekstraksi TF-IDF. Proses pelatihan model dilakukan dengan mengukur metrik evaluasi seperti akurasi dan F1-Score untuk menilai performa klasifikasi terhadap label sentimen yang ditetapkan. Dengan demikian model ini tidak hanya berfungsi sebagai alat evaluasi tetapi juga sebagai landasan pengambilan keputusan strategis dalam peningkatan kualitas pelayanan MCU. Contoh datasetnya adalah sebagai berikut:

#	Pengirim	Pekerjaan	Tgl. Pengaduan Masuk	Tgl. Tindak Lanjut	Status Tindak Lanjut	Keterangan
1	RADITYO PRIMANDANU KOESOEMO Indonesia Air Asia	PILOT	23 February 2024	23 February 2024	SELESAI	Baik
2	DERU BESTRIAN NURAJA Citilink Airline	PILOT	02 January 2024	-	PROSES	Dokter jantung sering terlambat sehingga membuang waktu menunggu
3	RYAN GHIFARI ASYUDI Indonesia Air Asia	PILOT	05 March 2024	-	SELESAI	
4	JONATHAN MARTIN Garuda Indonesia	PILOT	05 January 2024	05 January 2024	SELESAI	Pelayanan yang baik.
5	MERELLE ANASTASIA SURYAHASILAGA Wings Air	Flight Attendant	05 January 2024	05 January 2024	SELESAI	Excellent
6	EXELLINO KEVIN WBOWO API BANYUWANGI	PILOT	08 January 2024	-	SELESAI	
7	FATASYA NURMADIANTY BBN AIRLINES INDONESIA	Flight Attendant	02 January 2024	-	SELESAI	
8	NOVAN ZANZAHRI ZARFIE Citilink Airline	PILOT	04 January 2024	-	PROSES	Untuk tensi coba menggunakan manual. Karena akan lebih akurat

Gambar 3.6 Dataset Survey Pelanggan BLU Balai Kesehatan Penerbangan

Dataset yang digunakan dalam Gambar 3.6 terdiri atas data survei pengaduan dan kepuasan dari pelanggan layanan kesehatan penerbangan. Setiap entri merepresentasikan satu pengaduan atau umpan balik dari individu yang sebagian besar berasal dari profesi pilot yang bekerja di berbagai maskapai penerbangan seperti Indonesia Air Asia, Citilink Airline, dan Garuda Indonesia. Atribut yang dicatat dalam dataset ini mencakup nama pengirim beserta afiliasi

maskapainya, jenis pekerjaan, tanggal pengaduan masuk, tanggal tindak lanjut yang dilakukan oleh pihak penyedia layanan, status tindak lanjut (misalnya "SELESAI" atau "PROSES"), serta kolom keterangan yang berisi pernyataan langsung atau tanggapan dari pelanggan terkait kualitas pelayanan yang diterima.

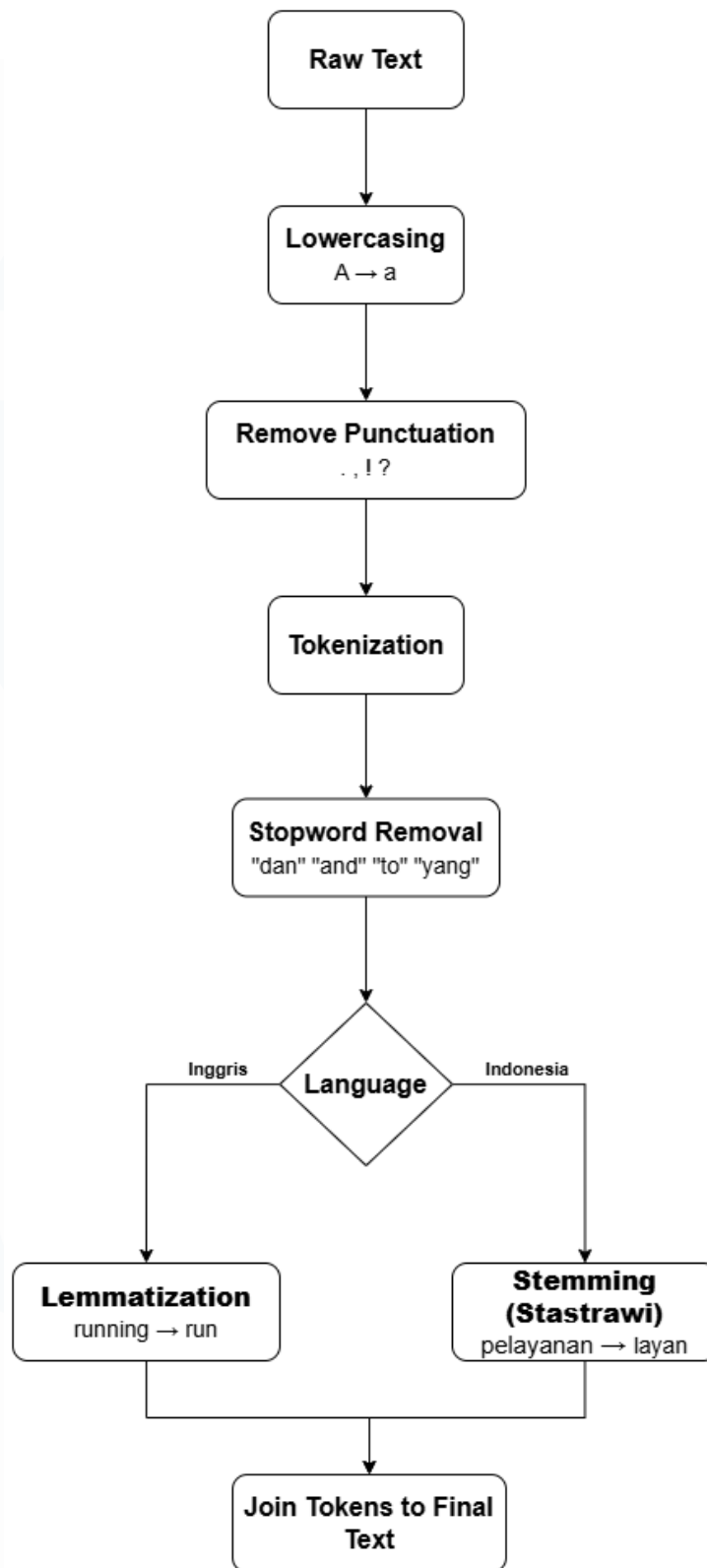
Kolom “Keterangan” memegang peran penting sebagai basis analisis dalam proyek ini karena berisi data tidak terstruktur berupa teks bebas yang mencerminkan persepsi pelanggan terhadap kualitas layanan yang diberikan. Beberapa entri mencerminkan kepuasan melalui komentar positif seperti “pelayanan yang baik” sementara yang lain mengandung keluhan spesifik seperti keterlambatan dokter yang menyebabkan waktu tunggu yang lama. Informasi ini menjadi sumber utama untuk ekstraksi sentimen dan pembelajaran model machine learning dalam mengkategorikan tingkat kepuasan pelanggan. Pendekatan ini memungkinkan sistem untuk tidak hanya mengidentifikasi status pengaduan secara administratif tetapi juga memahami dimensi kualitas layanan berdasarkan persepsi aktual pengguna.

Tahapan *praproses (Preprocessing)* data merupakan fondasi penting dalam analisis data teks khususnya dalam pengembangan model *machine learning* yang bertujuan untuk memahami sentimen dari masukan pelanggan. Data mentah yang diperoleh dari survei atau formulir pengaduan biasanya masih mengandung banyak elemen yang tidak relevan seperti simbol tanda baca angka acak atau bahkan teks yang tidak konsisten dalam hal kapitalisasi dan struktur kalimat. Keberadaan elemen-elemen tersebut dapat mengganggu proses ekstraksi informasi dan menurunkan performa model analitik yang digunakan. Oleh karena itu, sebelum data digunakan untuk pelatihan model diperlukan serangkaian tahapan sistematis guna memastikan bahwa data tersebut bersih terstruktur dan merepresentasikan informasi yang benar-benar diperlukan dalam proses analisis.

Langkah pertama yang dilakukan pada Gambar 3.7 yaitu praproses adalah pembersihan teks dari karakter-karakter yang tidak diperlukan, seperti simbol khusus angka yang tidak memiliki makna tertentu serta tanda baca yang tidak berkontribusi terhadap pemahaman konteks. Pembersihan ini dilakukan untuk

mengurangi kebisingan (*noise*) dalam data dan memperjelas isi teks agar proses selanjutnya dapat berjalan lebih efisien. Setelah proses pembersihan awal dilanjutkan dengan normalisasi teks di mana seluruh huruf diubah menjadi huruf kecil agar tidak terjadi duplikasi makna akibat perbedaan kapitalisasi. Misalnya kata “Dokter” dan “dokter” akan dianggap sebagai entitas yang berbeda jika normalisasi tidak dilakukan padahal keduanya merujuk pada hal yang sama. Selain itu, normalisasi juga mencakup penghapusan spasi ganda dan karakter tak terlihat yang sering kali muncul akibat kesalahan input atau hasil salinan data dari sistem lain.





Gambar 3.7 Pipeline Proses Preprocessing

Langkah lanjutan yang sangat penting adalah penghapusan *stopwords* yaitu kata-kata umum dalam bahasa, seperti “yang”, “dengan”, atau “adalah”, yang tidak membawa makna signifikan dalam konteks klasifikasi sentimen. Dengan menghapus kata-kata ini maka fokus analisis dapat diarahkan pada kata-kata kunci yang mengandung opini atau penilaian yang eksplisit terhadap pelayanan. Setelah itu, dilakukan proses *lemmatization* yaitu perubahan kata ke bentuk dasar atau lemma dengan mempertimbangkan konteks sintaksis dan semantik dari kata tersebut. Berbeda dengan *stemming* yang hanya memotong akhiran kata, *lemmatization* mempertahankan makna aslinya sehingga lebih sesuai untuk keperluan analisis linguistik yang sensitif terhadap makna.

3.2.2.1 Praproses: Menyiapkan *libraries* yang dibutuhkan

Langkah awal dalam pembangunan model *machine learning* untuk analisis kepuasan pelanggan adalah menyiapkan pustaka (*libraries*) yang mendukung seluruh proses mulai dari pengolahan data hingga evaluasi model. Penggunaan pustaka yang tepat sangat penting untuk memastikan efisiensi, keakuratan, dan skalabilitas proses analisis. Dalam proyek ini, berbagai pustaka digunakan dan dikelompokkan berdasarkan fungsinya, dimulai dari manipulasi data, preprocessing teks, hingga pemodelan machine learning. Pendekatan modular ini memudahkan proses debugging serta pengembangan lebih lanjut jika dibutuhkan.

Untuk manipulasi data yang terdapat pada Gambar 3.8, digunakan *pandas* dan *numpy*. *pandas* berperan dalam membaca, membersihkan, dan memanipulasi data survei pelanggan dalam format *DataFrame*. Proses seperti *filtering*, penggabungan data, dan agregasi dilakukan secara efisien dengan pustaka ini. *Numpy*, di sisi lain, digunakan dalam operasi numerik yang mendasari komputasi statistik dan transformasi data, termasuk pengolahan array dua dimensi yang sering muncul dalam representasi vektor. Kedua pustaka ini menjadi fondasi penting sebelum data diolah lebih lanjut oleh model *machine learning*.

```

import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
import numpy as np
from textblob import TextBlob
import string
import re
from nltk.tokenize import word_tokenize
from nltk.corpus import stopwords
from nltk.stem import WordNetLemmatizer
from sklearn.decomposition import LatentDirichletAllocation
import pyLDAvis
import pyLDAvis.lda_model
import spacy
from sklearn.pipeline import Pipeline
import joblib
import nltk
from langdetect import detect
import plotly.express as px
nltk.download('punkt')
nltk.download('stopwords')
nltk.download('wordnet')
from imblearn.over_sampling import SMOTE
from sklearn.naive_bayes import MultinomialNB
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix

```

Gambar 3.8 Daftar libraries yang dibutuhkan

Dalam tahap preprocessing teks, digunakan nltk, spaCy, Sastrawi, dan langdetect. langdetect digunakan untuk mendeteksi apakah respon teks ditulis dalam Bahasa Indonesia atau Inggris. Berdasarkan hasil deteksi bahasa ini, proses pembersihan teks dilakukan dengan pendekatan yang sesuai: untuk Bahasa Inggris digunakan nltk dan spaCy (untuk tokenisasi, stopwords removal, dan lemmatization), sedangkan untuk Bahasa Indonesia digunakan Sastrawi (untuk stemming). Proses ini bertujuan mengubah input teks menjadi bentuk yang lebih bersih dan seragam sebelum dikonversi menjadi fitur numerik.

Setelah teks dibersihkan, tahap transformasi dilakukan menggunakan TfidfVectorizer dari sklearn, yang mengubah teks menjadi representasi numerik berdasarkan frekuensi kata. Teknik TF-IDF ini membantu model membedakan kata-kata penting yang mencerminkan kepuasan atau keluhan pengguna. Untuk menangani data yang tidak seimbang antar kelas, digunakan teknik *oversampling* SMOTE dari imblearn. Selanjutnya, model dikembangkan menggunakan

LogisticRegression dari `sklearn.linear_model`, serta evaluasi performa dilakukan dengan `classification_report` dan `confusion_matrix`.

3.2.2.2 Praproses: Pengaturan format penulisan pada kolom Keterangan

Dalam proses analisis data berbasis teks, langkah awal yang sangat penting adalah pembersihan dan penyesuaian struktur data agar sesuai dengan kebutuhan model serta *pipeline* analisis. Pada data survei yang digunakan dalam proyek ini, terdapat beberapa kolom penting yang perlu diatur ulang format dan isinya, khususnya kolom Pengirim, Keterangan, dan Tgl. Pengaduan Masuk. Proses ini tidak hanya bertujuan meningkatkan keterbacaan, tetapi juga memastikan konsistensi data sebelum masuk ke tahap transformasi lanjutan seperti tokenisasi, vektorisasi, dan klasifikasi.

```
[ ] # Memisahkan nama pengirim dan maskapai (dengan batas 1 pemisahan)
df[['Nama Pengirim', 'Maskapai']] = df['Pengirim'].str.split('\n', n=1, expand=True)

# Mengisi NaN dengan string kosong jika tidak ada maskapai
df['Maskapai'] = df['Maskapai'].fillna('')

# Menghapus kolom Pengirim lama
df.drop(columns=['Pengirim'], inplace=True)

# Menyusun ulang kolom agar Nama Pengirim dan Maskapai ada di depan
df = df[['#', 'Nama Pengirim', 'Maskapai'] + [col for col in df.columns if col not in ['#', 'Nama Pengirim', 'Maskapai']]]
df.head()
```

Show hidden output

```
[ ] df["Keterangan"] = df["Keterangan"].str.lower()
df.head()
```

Show hidden output

```
▶ # Menghitung mode (nilai yang paling sering muncul) di kolom "Keterangan"
mode_keterangan = df["Keterangan"].mode()[0] # Ambil nilai pertama jika ada beberapa mode

# Mengganti NaN dengan nilai mode
df["Keterangan"].fillna(mode_keterangan, inplace=True)
df.head()
```

Show hidden output

Mengubah nilai NaN pada kolom "Keterangan" menjadi values paling sering muncul pada kolom "Komentar" dengan asumsi pengadu termasuk bagian dari rerata tersebut yang tidak ingin memberikan komentar.

```
[ ] # Contoh data (pastikan kolom tanggal ada di dataframe)
df["Tgl. Pengaduan Masuk"] = pd.to_datetime(df["Tgl. Pengaduan Masuk"]) # Konversi ke datetime
```

Gambar 3.9 Langkah dalam membenaran format

Langkah pertama pada Gambar 3.9 dilakukan pada kolom Pengirim, yang secara default berisi dua informasi dalam satu sel, yakni nama pengirim dan nama maskapai, yang dipisahkan oleh karakter baris baru atau *newline* (`\n`). Agar informasi ini dapat dianalisis lebih lanjut secara terpisah, data dipisah menjadi dua

kolom baru: Nama Pengirim, dan Maskapai, menggunakan metode `str.split()` dengan batas satu pemisahan. Namun, tidak semua entri memiliki informasi maskapai, sehingga kolom Maskapai akan berisi nilai kosong pada beberapa baris. Untuk mencegah error di proses selanjutnya, nilai kosong ini diganti dengan string kosong (") menggunakan `fillna()`. Setelah itu, kolom Pengirim lama dihapus untuk menghindari redundansi, dan susunan kolom disesuaikan kembali agar kolom hasil ekstraksi diletakkan di bagian depan. Ini memudahkan pemantauan dan validasi data secara manual oleh tim analis.

Setelah menangani struktur pengirim, fokus selanjutnya diarahkan ke kolom Keterangan, yang merupakan kolom inti dalam proses analisis sentimen karena berisi isi komentar atau pendapat dari responden. Untuk menjaga keseragaman dalam pemrosesan teks, seluruh isi kolom ini dikonversi ke huruf kecil menggunakan fungsi `str.lower()`. Hal ini mencegah duplikasi nilai akibat perbedaan kapitalisasi, yang bisa menyebabkan ketidakefisienan dan ketidakakuratan saat dilakukan tokenisasi atau pencocokan kata. Selain itu, beberapa data menunjukkan adanya nilai *NaN* atau kosong dalam kolom ini. Untuk menjaga kelengkapan data, nilai kosong ini diisi dengan nilai mode atau nilai yang paling sering muncul dalam kolom tersebut. Pendekatan ini dipilih agar pengisian tetap merepresentasikan konteks umum dari data yang ada, sekaligus meminimalkan kehilangan informasi penting.

Sebagai bagian akhir dari tahap praproses ini, kolom Tgl. Pengaduan Masuk juga mendapatkan perhatian khusus. Kolom ini semula bertipe *object* atau string, sehingga tidak dapat langsung digunakan untuk analisis berbasis waktu seperti agregasi per minggu atau bulan. Oleh karena itu, kolom ini dikonversi ke format *datetime* menggunakan fungsi `pd.to_datetime()`. Dengan format yang sesuai, kolom ini dapat digunakan untuk analisis tren waktu, melihat lonjakan keluhan berdasarkan periode, serta sebagai dasar segmentasi data pada laporan mingguan dan *dashboard* visualisasi.

Secara keseluruhan, pengaturan format ini merupakan bagian penting dari tahapan praproses yang sangat krusial untuk menjamin bahwa data dalam kondisi

bersih, konsisten, dan siap diproses lebih lanjut oleh *pipeline* analitik. Tanpa langkah ini, risiko error, kehilangan data penting, atau interpretasi yang salah bisa meningkat secara signifikan. Oleh karena itu, meskipun terlihat sederhana, tahap ini memberikan fondasi yang kokoh untuk keseluruhan proses analisis yang akan dilakukan berikutnya.

3.2.2.3 Praproses: Membangun pipeline untuk mendeteksi bahasa

Dalam analisis teks yang melibatkan responden dari berbagai latar belakang, termasuk dari sektor internasional di dunia aviasi, sangat mungkin ditemukan masukan atau komentar yang menggunakan lebih dari satu bahasa. Untuk itu, salah satu langkah krusial dalam proses praproses adalah mendeteksi bahasa dari setiap masukan pelanggan agar dapat dilakukan pemrosesan teks yang sesuai berdasarkan karakteristik bahasanya. Misalnya, teks berbahasa Indonesia perlu diproses dengan teknik dan pustaka yang berbeda dibandingkan teks berbahasa Inggris. Oleh karena itu, pipeline pendeteksian bahasa dibangun menggunakan library *langdetect*, yang mampu mengidentifikasi bahasa dari suatu string teks secara otomatis.

```
[ ] def detect_language(text):
    try:
        return detect(text)
    except:
        return 'unknown'

df['lang'] = df['Keterangan'].apply(detect_language)

[ ] df_id = df[df['lang'] == 'id']
df_en = df[df['lang'] == 'en']

from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
import re

# Inisialisasi tools
nlp_en = spacy.load("en_core_web_sm")
factory = StemmerFactory()
stemmer = factory.create_stemmer()
stop_words_en = set(stopwords.words("english"))
stop_words_id = set([
    "yang", "dan", "di", "ke", "dari", "pada", "adalah", "untuk", "dengan", "itu", "ini", "sudah",
    "sangat", "agar", "jika", "karena", "dalam", "sehingga", "tidak", "bisa", "lebih", "kurang",
    "terlalu", "banyak", "terima", "kasih", "semua", "dapat"
])
```

Gambar 3.10 Proses membangun pipeline bahasa

Fungsi `detect_language()` yang teradapat pada Gambar 3. 10 dibuat untuk membungkus proses deteksi bahasa. Fungsi ini mencoba mendeteksi bahasa dari

setiap entri teks yang ada di kolom Keterangan. Jika proses deteksi berjalan lancar, maka akan mengembalikan kode bahasa seperti 'id' untuk Bahasa Indonesia atau 'en' untuk Bahasa Inggris. Namun, dalam kasus tertentu di mana teks kosong atau terlalu singkat sehingga tidak dapat dianalisis, fungsi ini akan menangkap error dan mengembalikan nilai 'unknown'. Hasil deteksi bahasa disimpan ke dalam kolom baru bernama 'lang'. Dari hasil ini, data kemudian dipisahkan ke dalam dua subset: `df_id` untuk teks berbahasa Indonesia dan `df_en` untuk teks berbahasa Inggris. Pemisahan ini sangat penting untuk memastikan setiap bahasa mendapat perlakuan yang sesuai pada tahapan praproses berikutnya.

Untuk teks berbahasa Indonesia, digunakan library *Sastrawi* yang terkenal dalam dunia NLP Bahasa Indonesia, khususnya dalam proses *stemming* dan *stopword removal*. Inisialisasi dilakukan terhadap *StemmerFactory* dari *Sastrawi* untuk membentuk objek stemmer yang bertugas mengubah kata ke bentuk dasarnya. Selain itu, *stopword* juga didefinisikan secara manual menggunakan daftar kata-kata umum dalam Bahasa Indonesia yang tidak memiliki makna penting dalam konteks sentimen, seperti "yang", "sudah", "agar", "banyak", dan lain-lain. Daftar ini dirancang agar lebih kontekstual terhadap data survei yang dianalisis, karena banyak mengandung ungkapan netral atau repetitif dari responden.

Sementara itu, untuk teks berbahasa Inggris, digunakan pustaka *spaCy* dengan model `en_core_web_sm` sebagai NLP pipeline utama, yang mencakup berbagai proses seperti tokenisasi, *part-of-speech tagging*, dan *lemmatization*. Sebagai pelengkap, juga digunakan *stopword list* dari *NLTK* untuk menyaring kata-kata umum dalam Bahasa Inggris seperti "the", "and", "is", yang seringkali tidak memberikan nilai informatif dalam klasifikasi sentimen. Dengan pendekatan bilingual ini, pipeline praproses dapat menangani beragam input teks secara akurat dan efisien, sehingga hasil analisis menjadi lebih tajam dan relevan dengan kondisi nyata di lapangan.

PIPELINE BAHASA INGGRIS

```
[ ] def preprocess_english(text):
    text = re.sub(r"[^\w\s]", "", text) # hapus tanda baca
    doc = nlp_en(text.lower())
    tokens = [token.lemma_ for token in doc if token.text.lower() not in stop_words_en and not token.is_punct]
    return " ".join(tokens)
```

PIPELINE INDONESIA

```
[ ] def preprocess_indonesian(text):
    text = re.sub(r"[^\w\s]", "", text)
    tokens = text.lower().split()
    tokens = [t for t in tokens if t not in stop_words_id]
    tokens = [stemmer.stem(t) for t in tokens]
    return " ".join(tokens)
```

PENERAPAN

```
[ ] def detect_lang(text):
    try:
        return detect(text)
    except:
        return "unknown"

# Tambah kolom bahasa
df["detected_lang"] = df["Keterangan"].apply(detect_lang)

# Terapkan pipeline sesuai bahasa
def preprocess_mixed(text, lang):
    if lang == "id":
        return preprocess_indonesian(text)
    elif lang == "en":
        return preprocess_english(text)
    else:
        return text # return original if language unknown

df["processed_text"] = df.apply(lambda row: preprocess_mixed(row["Keterangan"], row["detected_lang"]), axis=1)
```

Gambar 3.11 Penerapan *pipeline* kedalam dataframe

Langkah pertama yang tertera pada Gambar 3.11 adalah membangun fungsi pendeteksi bahasa bernama `detect_lang()`, yang menggunakan pustaka `langdetect` untuk mengklasifikasikan setiap baris teks ke dalam kategori “id” (Bahasa Indonesia), “en” (Bahasa Inggris), atau “unknown” apabila pendeteksian tidak berhasil. Hasil deteksi ini kemudian disimpan dalam kolom baru bernama `detected_lang`, yang menjadi referensi untuk menentukan proses lanjutan yang harus dilakukan. Setelah itu, dua buah fungsi khusus dikembangkan untuk masing-masing bahasa: `preprocess_indonesian()` dan `preprocess_english()`. Masing-masing fungsi tersebut dirancang untuk menangani struktur linguistik khas dari tiap bahasa secara efektif, agar informasi yang relevan tetap terjaga meskipun dilakukan penyederhanaan kata.

Untuk teks berbahasa Indonesia, pipeline menerapkan beberapa tahapan kunci yaitu pembersihan tanda baca, konversi huruf ke lowercase, penghapusan *stopword* dari daftar yang dikembangkan secara khusus untuk konteks survei

publik, dan terakhir proses *stemming* menggunakan Sastrawi. Tujuan dari *stemming* ini adalah mengembalikan kata ke bentuk dasarnya agar makna inti tetap terwakili meskipun dalam berbagai bentuk turunan. Sedangkan untuk teks berbahasa Inggris, pipeline menerapkan lemmatization menggunakan pustaka *spaCy*. Proses ini dilakukan setelah pembersihan tanda baca dan penghapusan *stopword* dari pustaka *nlTK*, kemudian token yang masih relevan diubah ke bentuk lema yang paling mendekati akar katanya tanpa kehilangan konteks gramatikal. Ini penting untuk menjaga agar fitur teks yang dihasilkan tetap informatif dan dapat membedakan sentimen maupun urgensi komentar pengguna secara lebih akurat.

Setelah kedua pipeline bahasa tersebut selesai dibuat, diterapkan fungsi `preprocess_mixed()` yang bertugas menyesuaikan metode preprocessing berdasarkan bahasa yang telah terdeteksi. Fungsi ini kemudian dijalankan menggunakan metode `.apply()` pada setiap baris data, sehingga menghasilkan kolom baru `processed_text` yang berisi teks bersih dan siap digunakan untuk tahap *vectorization* selanjutnya. Dengan adanya pipeline preprocessing multibahasa ini, proses analisis sentimen maupun klasifikasi tingkat kepuasan dapat berjalan dengan lebih presisi. Hal ini juga menunjukkan pentingnya fleksibilitas teknis dalam menangani data real-world, di mana kebijakan satu pendekatan saja sering kali tidak cukup. Kemampuan untuk beradaptasi terhadap konten data menjadi krusial dalam memastikan keberhasilan sistem analitik berbasis machine learning yang dikembangkan di lingkungan BLU Balai Kesehatan Penerbangan.

Keterangan	lang	detected_lang	processed_text
baik	id	id	baik
dokter jantung sering terlambat sehingga membu...	id	id	dokter jantung sering lambat buang waktu tunggu
good	so	so	good
pelayanan yang baik	id	id	layan baik
excellent	ca	ca	excellent
good	so	so	good
good	so	so	good
untuk tensi coba menggunakan manual. karena ak...	id	id	tensi coba guna manual akan akurat
pertahankan dan tingkatkan	id	id	tahan tingkat
semua baik..	id	id	baik

Gambar 3.12 Hasil akhir data dari proses *Preprocessing*

Hasil akhir dari seluruh tahapan ini pada Gambar 3.12 adalah data teks yang telah melalui transformasi menjadi representasi yang lebih bersih konsisten dan siap untuk diproses lebih lanjut pada tahap ekstraksi fitur dan pelatihan model *machine learning* guna mengidentifikasi sentimen secara lebih efektif dan akurat.

3.2.2.4 Penggunaan model Logistic Regression untuk menganalisis sentimen

Model *Logistic Regression* digunakan sebagai algoritma klasifikasi utama dalam menganalisis sentimen yang terkandung dalam data pengaduan pelanggan. Model ini dipilih karena memiliki performa yang baik dalam menyelesaikan permasalahan klasifikasi kategori ganda seperti sentimen negatif, netral, dan positif. Sebelum model dilatih, dilakukan proses pelabelan awal secara otomatis berdasarkan keberadaan kata-kata kunci tertentu yang mencerminkan arah sentimen. Kata-kata yang merepresentasikan sentimen negatif dan positif dipetakan untuk mengklasifikasikan setiap baris data ke dalam salah satu dari tiga kelas

sentimen utama. Setelah data diberi label, fitur teks diekstraksi menggunakan teknik TF-IDF *Vectorization* dengan mempertimbangkan unigram dan bigram, serta membatasi jumlah fitur maksimal sebanyak 5000 agar proses pelatihan tetap efisien.

Model dilatih menggunakan skema train-test split dengan proporsi 80% data pelatihan dan 20% data pengujian. Hasil evaluasi model pada Gambar 3.13 menunjukkan tingkat akurasi yang sangat tinggi yaitu sebesar 98,37%, yang menunjukkan bahwa model dapat mengklasifikasikan sentimen dengan presisi yang sangat baik. Berdasarkan metrik evaluasi seperti *precision*, *recall*, dan *F1-score*, model menunjukkan performa sangat optimal khususnya dalam mendeteksi sentimen positif dan netral. Namun, performa pada kelas negatif masih rendah karena jumlah data pada kelas tersebut sangat terbatas, sehingga berdampak pada ketidakseimbangan distribusi kelas yang cukup signifikan. Ketidakseimbangan ini menyebabkan model kesulitan dalam mengenali pola dari sentimen negatif secara akurat. Meskipun demikian secara umum model mampu memberikan klasifikasi yang stabil dan efektif untuk mayoritas data yang dianalisis.

Accuracy: 0.983756345177665

Classification Report:

	precision	recall	f1-score	support
Negatif	0.00	0.00	0.00	5
Netral	0.97	0.97	0.97	216
Positif	0.99	0.99	0.99	764
accuracy			0.98	985
macro avg	0.65	0.65	0.65	985
weighted avg	0.98	0.98	0.98	985

Confusion Matrix:

```
[[ 0  3  2]
 [ 0 209  7]
 [ 0  4 760]]
```

Gambar 3.13 Evaluasi Model Logistic Regression

Penerapan model *Logistic Regression* ini memberikan manfaat langsung dalam proses identifikasi opini pelanggan terhadap layanan yang diberikan. Dengan klasifikasi sentimen yang dilakukan secara otomatis, proses analisis data dapat dilakukan secara lebih cepat dan efisien tanpa harus membaca setiap komentar secara manual. Hasil klasifikasi ini dapat digunakan untuk mendeteksi tren opini masyarakat terhadap pelayanan kesehatan, memantau kualitas pelayanan dari waktu ke waktu, serta mengidentifikasi potensi isu atau ketidakpuasan yang perlu segera ditindaklanjuti. Model ini juga dapat digunakan sebagai fondasi awal untuk pengembangan sistem pelaporan berbasis data yang lebih adaptif dan responsif terhadap masukan pelanggan.

3.2.2.5 Penggunaan model SMOTE untuk menganalisis sentimen

Untuk mengatasi permasalahan ketidakseimbangan kelas pada data sentimen pelanggan yang digunakan, dilakukan pendekatan *oversampling* menggunakan metode *RandomOverSampler* dari pustaka *imbalanced-learn*. Ketidakseimbangan ini tampak jelas pada distribusi label, di mana jumlah data berlabel positif dan netral jauh lebih dominan dibandingkan dengan data yang merepresentasikan sentimen negatif. Ketimpangan distribusi seperti ini dapat menyebabkan bias pada model klasifikasi, sehingga model cenderung terlalu sering memprediksi kelas mayoritas dan gagal mendeteksi kelas minoritas secara akurat. Hal ini tentu sangat merugikan dalam konteks pemantauan pelayanan publik, karena sentimen negatif justru merupakan indikator penting yang dapat memberikan sinyal perbaikan layanan. Oleh karena itu, dilakukan strategi penyeimbangan kelas pada data pelatihan agar model yang dihasilkan mampu mengenali pola dari semua jenis sentimen dengan lebih adil dan akurat.

Langkah awal yang dilakukan sebelum penerapan *oversampling* adalah transformasi data teks hasil preprocessing ke dalam bentuk representasi numerik menggunakan teknik *Term Frequency-Inverse Document Frequency* (TF-IDF). Teknik ini digunakan untuk menangkap informasi penting dari frekuensi kata dalam dokumen relatif terhadap seluruh korpus, sehingga model dapat membedakan kata-kata umum dari kata-kata yang lebih relevan terhadap konteks tertentu. Setelah

diperoleh matriks fitur dari teks, data kemudian dibagi menjadi data pelatihan dan data pengujian menggunakan metode stratified train-test split guna mempertahankan proporsi distribusi kelas dalam subset data yang lebih kecil. Selanjutnya, teknik *RandomOverSampler* diterapkan pada data pelatihan untuk memperbanyak sampel dari kelas yang memiliki jumlah lebih sedikit, yakni kelas negatif, dengan cara menduplikasi secara acak baris-baris data yang termasuk dalam kelas tersebut. Proses ini bertujuan untuk mengurangi bias terhadap kelas mayoritas dan memberikan kesempatan yang lebih besar bagi model untuk belajar dari karakteristik kelas minoritas.

Setelah proses *oversampling* selesai, dilakukan pelatihan ulang terhadap model Logistic Regression menggunakan data pelatihan yang sudah seimbang. Evaluasi terhadap model pada Gambar 3.14 dilakukan dengan menguji performanya pada data uji yang tidak diubah untuk menjaga objektivitas penilaian. Hasil evaluasi menunjukkan bahwa akurasi model secara keseluruhan tetap tinggi yaitu sebesar 97%, sementara distribusi nilai precision, recall, dan f1-score menunjukkan perbaikan pada kelas negatif dibandingkan dengan pemodelan sebelumnya. F1-score untuk kelas negatif meningkat menjadi 0.50 dari sebelumnya 0.00, meskipun data uji untuk kelas ini masih sangat terbatas. Selain itu, nilai f1-score untuk kelas positif dan netral juga tetap tinggi yang menunjukkan bahwa penyeimbangan data pelatihan tidak menurunkan kualitas prediksi untuk kelas mayoritas. Temuan ini menunjukkan bahwa penggunaan teknik oversampling secara signifikan dapat meningkatkan keandalan model klasifikasi dalam mengidentifikasi keluhan pelanggan tanpa mengorbankan akurasi prediksi untuk sentimen lain.

Accuracy: 0.9685279187817258

Classification Report:

	precision	recall	f1-score	support
Negatif	1.00	0.33	0.50	3
Netral	0.88	1.00	0.93	217
Positif	1.00	0.96	0.98	765
accuracy			0.97	985
macro avg	0.96	0.77	0.80	985
weighted avg	0.97	0.97	0.97	985

Confusion Matrix:

```
[[ 1  2  0]
 [ 0 217  0]
 [ 0  29 736]]
```

Gambar 3.14 Evaluasi Model SMOTE

Penerapan teknik ini memberikan manfaat strategis dalam pengembangan sistem analisis opini publik di sektor pelayanan kesehatan khususnya dalam konteks pemantauan pengalaman pengguna terhadap layanan yang diberikan. Dengan model yang lebih sensitif terhadap sentimen negatif, organisasi dapat lebih cepat dan tepat dalam mengenali isu atau keluhan yang muncul serta mengambil langkah korektif berdasarkan data yang valid. Hal ini tidak hanya berkontribusi terhadap peningkatan kualitas layanan tetapi juga membangun kepercayaan pelanggan terhadap sistem umpan balik yang diterapkan. Keberhasilan pendekatan ini juga membuka peluang untuk eksplorasi metode penyeimbangan data lainnya serta integrasi dengan algoritma klasifikasi lanjutan demi menghasilkan model analitik yang lebih robust dan adaptif dalam menghadapi variasi opini pelanggan yang kompleks.

3.2.3 Penyesuaian Data Pasien terkait Check atau Recheck

Sebagai seorang data analyst di BLU Balai Kesehatan Penerbangan, pemanfaatan sistem digital berbasis *dashboard* seperti yang ditampilkan pada serangkaian gambar ini memiliki peran sentral dalam mendukung efisiensi proses pelayanan dan analisis data pasien. Salah satu alur penting yang dianalisis adalah

alur pemeriksaan kesehatan personil penerbangan, yang terintegrasi dari input data personal, verifikasi status pemeriksaan, rekam medis historis, hingga rincian biaya layanan yang dijalani. *Dashboard* ini tidak hanya memfasilitasi petugas layanan dalam menyusun proses pemeriksaan yang tepat, namun juga mendukung pengambilan keputusan berbasis data bagi manajemen untuk memastikan layanan berjalan sesuai regulasi.

Profil Personil Penerbangan

No. Pemeriksaan
25051771

No. RM
00-

Nama Lengkap
L. RTO

Jenis Kelamin
Laki-laki

Tgl. Lahir / Usia
1994 / 31 th

Nomor Lisensi

Pekerjaan
PILOT

Maskapai
SUPER AIR JET

Kelas
FIRST

Foto Profil

Detail Pemeriksaan

Maskapai
Super Air Jet

Tgl. Pemeriksaan
2025-05-20

Jenis Pemeriksaan
Regular

Tempat Pemeriksaan
Hatpen

* maskapai daftar pemeriksaan

Status Pasien
Standart

Paket Pemeriksaan
Standart (Basic)

Gambar 3.15 *Dashboard* data diri pasien

Langkah awal dalam alur ini dimulai dari tampilan profil personil (Gambar 3.15), yang mencatat identitas lengkap seperti nama, maskapai, nomor lisensi, serta jenis pemeriksaan dan lokasi pelaksanaan. Bagian ini menjadi sumber penting dalam pemetaan demografis pasien, frekuensi kunjungan berdasarkan maskapai, serta segmentasi kebutuhan layanan (misalnya berdasarkan jenis pemeriksaan atau kelas penerbangan). Dengan data tersebut, analisis lanjutan dapat dilakukan untuk mengetahui apakah terdapat korelasi antara jenis maskapai atau kelas penerbangan tertentu dengan kebutuhan paket pemeriksaan tertentu. Selain itu, data pada tahap ini juga berfungsi untuk mengukur distribusi beban kerja fasilitas kesehatan berdasarkan lokasi pemeriksaan. Dengan visualisasi dan pengolahan yang tepat, informasi ini bisa digunakan untuk menyusun proyeksi kebutuhan sumber daya,

perencanaan jadwal pemeriksaan, hingga optimalisasi kapasitas pelayanan kesehatan berbasis lokasi dan waktu.

Status Pemeriksaan | Status Paket Pemeriksaan | Log Status Paket Pemeriksaan

Lab Paket Pemeriksaan

Audiometri | Fisik | Gigi | Laboratorium | Mata | Result

Daftar Paket Pemeriksaan

No	Nama Paket	Poliklinik	Re-check	Catatan
1	Pengisian Pendaftaran	Audiometri	Tidak	
2	Konsultasi Dokter Umum	Fisik	Tidak	
3	Paket Pengisian Fisik Lengkap	Fisik	Tidak	
4	Pengisian pemeriksaan Gigi	Gigi	Tidak	
5	Pemeriksaan Hunkulu Urine	Laboratorium	Tidak	
6	Pemeriksaan Paket Darah Lengkap	Laboratorium	Tidak	
7	Pemeriksaan Paket Kimia Darah	Laboratorium	Tidak	
8	Pemeriksaan Urine Lengkap	Laboratorium	Tidak	
9	Pemeriksaan Paket Mata Pili	Mata	Tidak	

Riwayat Hasil Pemeriksaan

Head Neck	Detail	Abdomen	Detail
Mouth Throat	Detail	Anus	Detail
Nose	Detail	Urinary	Detail
Ears	Detail	Endocrine	Detail
Eyes (Oculi)	Detail	Joint	Detail
Eyes (Papili)	Detail	Spleen	Detail
Eyes (Ocular)	Detail	Neurologic	Detail
Chest	Detail	Pulicudi	Detail
Heart	Detail	Skin	Detail
Vascular	Detail	General	Detail

Status Kelulusan :

Ekstensi Luka (Tanda lahir/tattoo) :

Limitasi :

Catatan : ** Status yang dalam proses dengan poliklinik berdasarkan paket pemeriksaan yang dilakukan oleh personel pemeriksaan sesuai dengan apa yang tertera di atas.

Gambar 3.16 Database Rekam Medis Pasien

Tahapan selanjutnya adalah pengecekan riwayat dan status pemeriksaan sebelumnya (gambar 3.16 dan 3.17), yang sangat krusial dalam memastikan bahwa layanan tidak dilakukan secara redundan. Beberapa jenis tes seperti *treadmill*, Elektrokardiogram (EKG/ECG), dan Elektroensefalogram (*EEG*) memiliki batasan frekuensi pemeriksaan dalam satu tahun, kecuali untuk keperluan *recheck*. Dari sisi analitik, data historis pemeriksaan ini juga dapat diolah untuk menghasilkan insights mengenai tingkat kesehatan umum personil penerbangan dalam jangka waktu tertentu. Misalnya, dengan mengamati tren abnormalitas yang muncul dari pemeriksaan tertentu, dapat dilakukan tindakan preventif yang lebih tepat sasaran, seperti penyesuaian pola jadwal kerja atau intervensi medis dini bagi profesi penerbangan dengan risiko tinggi. Selain itu, dengan membandingkan data historis antar individu atau kelompok maskapai, bisa dihasilkan metrik performa kesehatan kolektif sebagai bahan evaluasi pihak regulator.

Tanggal Periksa : 2024-06-14

FISIK	☒
LAB	☒
RONTGEN	☒
ECG	☒
TREADMILL	☐
MATA	☒
AUDIO	☒
GIGI	☒
EEG	☐

Tanggal Periksa : 2024-11-26

FISIK	☒
LAB	☒
RONTGEN	☐
ECG	☐
TREADMILL	☐
MATA	☒
AUDIO	☒
GIGI	☒
EEG	☐

Gambar 3.17 Pemeriksaan pasien beberapa bulan sebelumnya

Bagian selanjutnya dalam alur sistem adalah daftar pemeriksaan yang akan dilakukan di hari tersebut (Gambar 3.18). Di sinilah sistem menggabungkan data dari hasil pemilihan paket dan kebijakan layanan pemeriksaan untuk menyusun daftar tindakan medis dan estimasi biaya yang akan dikeluarkan. Dari sudut pandang data analyst, daftar pemeriksaan ini bisa dijadikan dasar dalam membuat analisis beban layanan berdasarkan waktu, misalnya peak-time pemeriksaan fisik atau radiologi, serta identifikasi potensi bottleneck dalam alur pelayanan. Lebih lanjut, analisis frekuensi pemilihan layanan juga dapat dimanfaatkan untuk merancang paket pemeriksaan yang lebih disesuaikan dengan kebutuhan pengguna dan efisiensi operasional. Kombinasi data harga layanan, jumlah pasien, dan waktu tunggu juga bisa digunakan untuk menghitung efektivitas biaya serta menentukan strategi promosi atau penyesuaian harga layanan di masa mendatang.

Pemeriksaan Kesehatan				Paket MCU	Paket Pemeriksaan	Jumlah Pengisian
No.	Nama Pengujian	Poliklinik	Harga			
FISIK						
1	— Paket Pengujian Fisik Lengkap	Fisik	120.000			2
2	— Konsultasi Dokter Umum	Fisik	50.000			2
MATA						
3	— Pemeriksaan Paket Mata Paket	Mata	100.000			2
AUDIOMETRI						
4	— Pengujian Pendengaran	Audiometri	75.000			2
GIGI						
5	— Pengujian pemeriksaan Gigi	Gigi	87.500			2
RADIOLOGI						
6	— Foto Thorax PA/AP	Radiologi	250.000			2
LABORATORIUM						
7	— Pemeriksaan Paket Darah Lengkap	Laboratorium	150.000			2
8	— Pemeriksaan Urine Lengkap	Laboratorium	120.000			2
9	— Pemeriksaan Paket Kimia Darah	Laboratorium	450.000			2
JANTUNG						
10	— Kurva Listrik Jantung (ECG)	Jantung	300.000			2
KASIR						
11	— Administrasi	Kasir	17.500			2
12	— Resume	Kasir	30.000			2
13	— Serifikasi Kesehatan	Kasir	60.000			2
14	— Berkas Rekam Medis	Kasir	25.000			2
			Total			1.825.000

Gambar 3.18 Pemeriksaan yang sudah dicocokkan dengan database

Terakhir, data yang sudah benar akan disimpan sama seperti pada Gambar 3.19. Akhir dari rangkaian ini memberikan gambaran lengkap mengenai total estimasi biaya yang ditanggung oleh pasien untuk seluruh rangkaian pemeriksaan hari itu. Informasi ini dapat dijadikan bahan utama dalam membangun laporan keuangan harian, menyusun prediksi pendapatan berdasarkan periode waktu, dan menilai tingkat profitabilitas masing-masing jenis layanan. Kombinasi seluruh data ini, bila dikelola dengan sistem yang terintegrasi dan diperkaya analisis statistik serta prediktif, mampu mendukung pengambilan keputusan yang lebih strategis bagi BLU Balai Kesehatan Penerbangan, baik dari sisi pelayanan medis maupun aspek bisnis operasionalnya.

Catatan
Isikan catatan dokter dari pemeriksaan terakhir (bila ada)

File Catatan
Upload Catatan

Pilih File

Tidak ada file yang dipilih

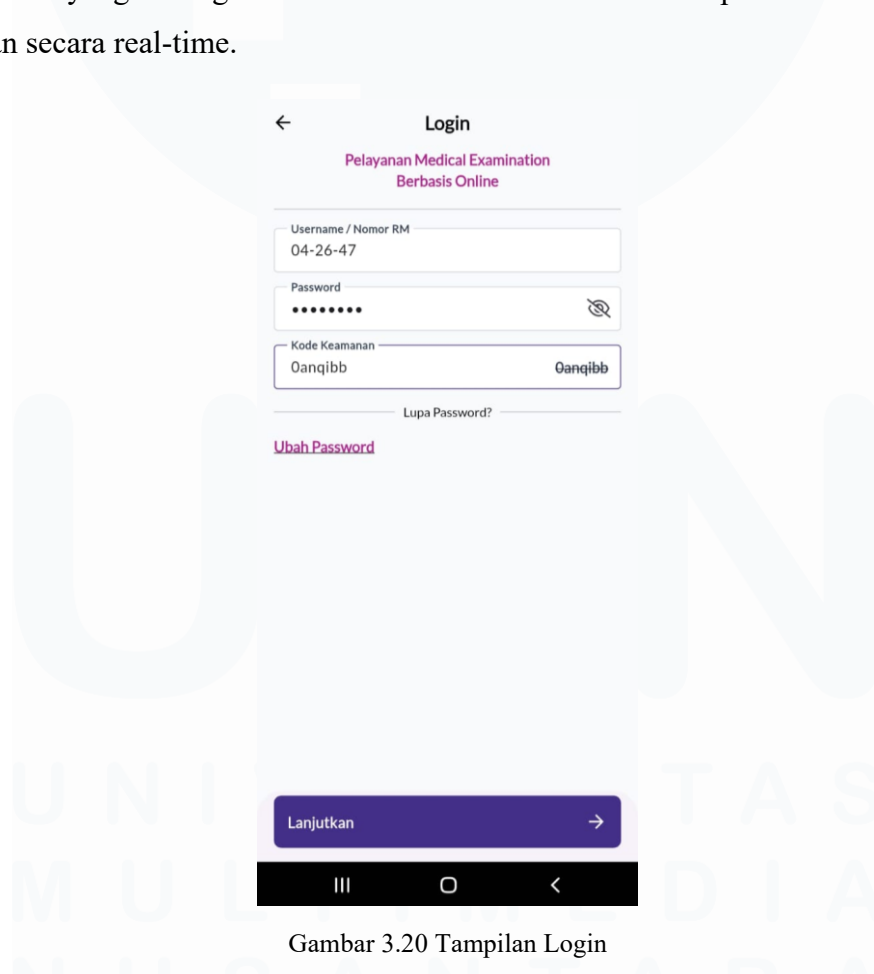
Simpan Foto

SIMPAN & UPDATE STATUS CEK BERKAS

Gambar 3. 19 Menyimpan data yang sudah bersih

3.2.4 Testing Aplikasi HATPEN Mobile

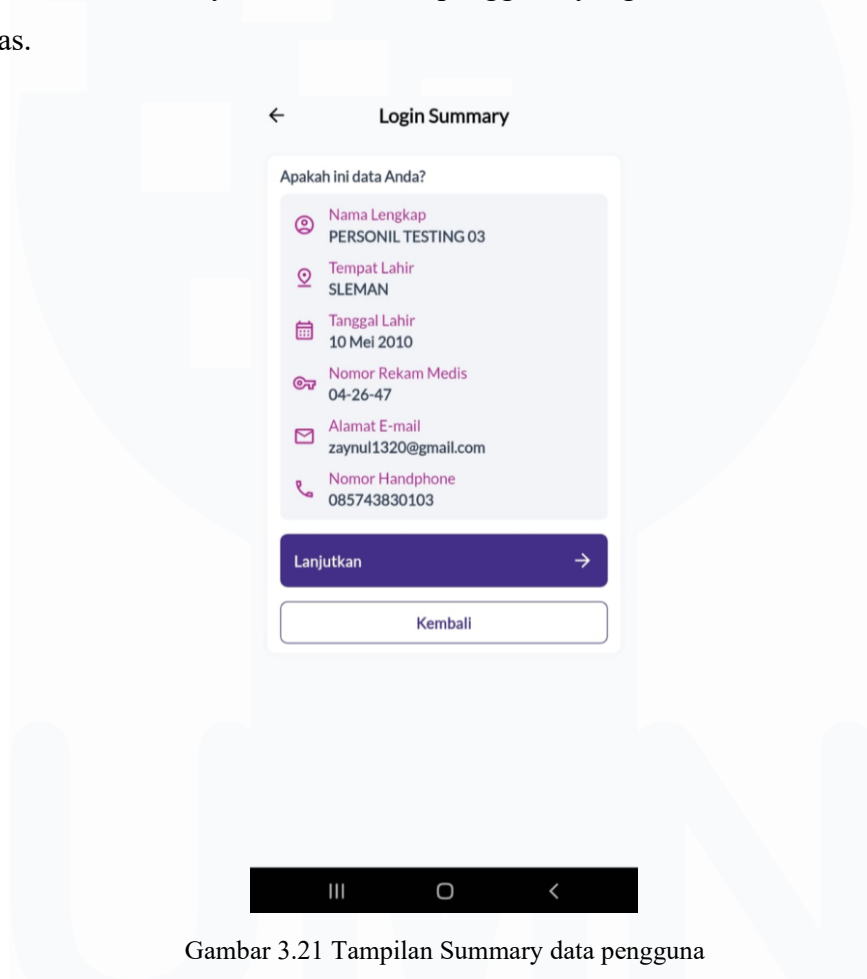
Dalam rangka mendukung transformasi layanan kesehatan berbasis digital, BLU Balai Kesehatan Penerbangan mengembangkan aplikasi *Pelayanan Medical Examination Berbasis Online* yang dapat diakses oleh personel penerbangan secara mandiri melalui perangkat seluler. Aplikasi ini dirancang sebagai solusi efisien untuk mempercepat proses administrasi dan monitoring pemeriksaan kesehatan secara daring tanpa perlu antri fisik di lokasi pelayanan. Melalui sistem ini, pengguna dapat melakukan login, verifikasi data, mengecek riwayat pemeriksaan, serta melakukan pemesanan jadwal secara langsung dalam satu platform yang terintegrasi. Adanya sistem ini menjadi bagian dari inovasi pelayanan publik yang berorientasi pada kemudahan akses dan efisiensi waktu, serta mendukung sistem *dashboard* yang dibangun oleh analis data untuk kebutuhan pemantauan kinerja layanan secara real-time.



Gambar 3.20 Tampilan Login

Pada Gambar 3.20, aplikasi mengharuskan pengguna mengisi tiga kolom utama: *Username/Nomor RM*, *Password*, dan *Kode Keamanan* (captcha). Kolom

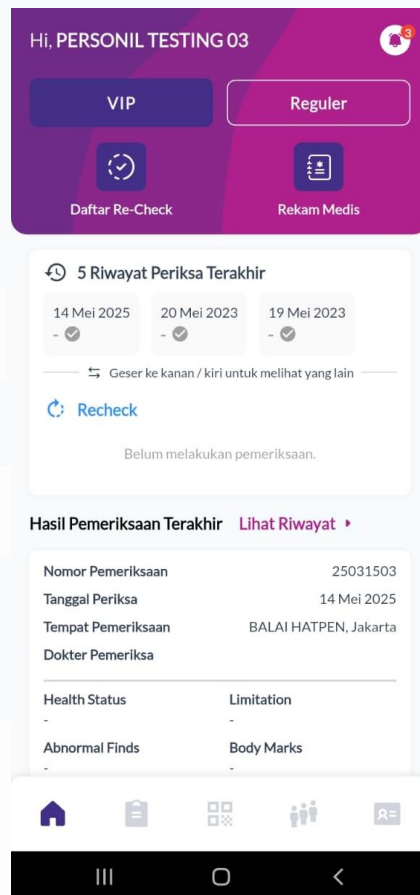
Username biasanya diisi dengan nomor rekam medis pengguna. Kode keamanan merupakan fitur validasi untuk memastikan input berasal dari pengguna nyata, bukan sistem otomatis. Setelah semua kolom diisi dengan benar, pengguna dapat menekan tombol “Lanjutkan” yang terletak di bagian bawah. Proses ini menjadi gerbang awal masuk ke dalam sistem informasi pelayanan medis secara daring, dan memastikan akses hanya dimiliki oleh pengguna yang telah terverifikasi secara identitas.



Gambar 3.21 Tampilan Summary data pengguna

Setelah berhasil login (seperti pada Gambar 3.21), pengguna akan disajikan dengan data ringkasan identitas seperti nama lengkap, tempat dan tanggal lahir, nomor rekam medis, alamat email, serta nomor handphone. Fungsi tampilan ini adalah untuk memberikan konfirmasi visual kepada pengguna bahwa data akun yang digunakan telah sesuai dengan identitas aslinya. Jika semua informasi telah benar, pengguna dapat menekan “Lanjutkan” untuk masuk ke halaman utama

aplikasi. Fitur ini penting untuk mencegah kesalahan identitas yang bisa berdampak pada validitas rekam medis dan informasi pendaftaran.



Gambar 3.22 Tampilan menu utama aplikasi

Halaman pada Gambar 3.22 memperlihatkan dua kategori layanan yang tersedia, yaitu *VIP* dan *Reguler*, yang dapat dipilih sesuai kebutuhan pengguna. Di bagian tengah, pengguna dapat mengakses menu *Daftar Re-Check* dan *Rekam Medis*. Fitur unggulan lainnya adalah daftar riwayat lima pemeriksaan terakhir lengkap dengan tanggal pelaksanaan dan statusnya. Pada bagian bawah, terdapat ringkasan hasil pemeriksaan terakhir termasuk nomor pemeriksaan, tanggal, tempat pelaksanaan, dan dokter pemeriksa. Semua informasi ditampilkan dengan antarmuka yang responsif dan sederhana untuk memudahkan navigasi pengguna.

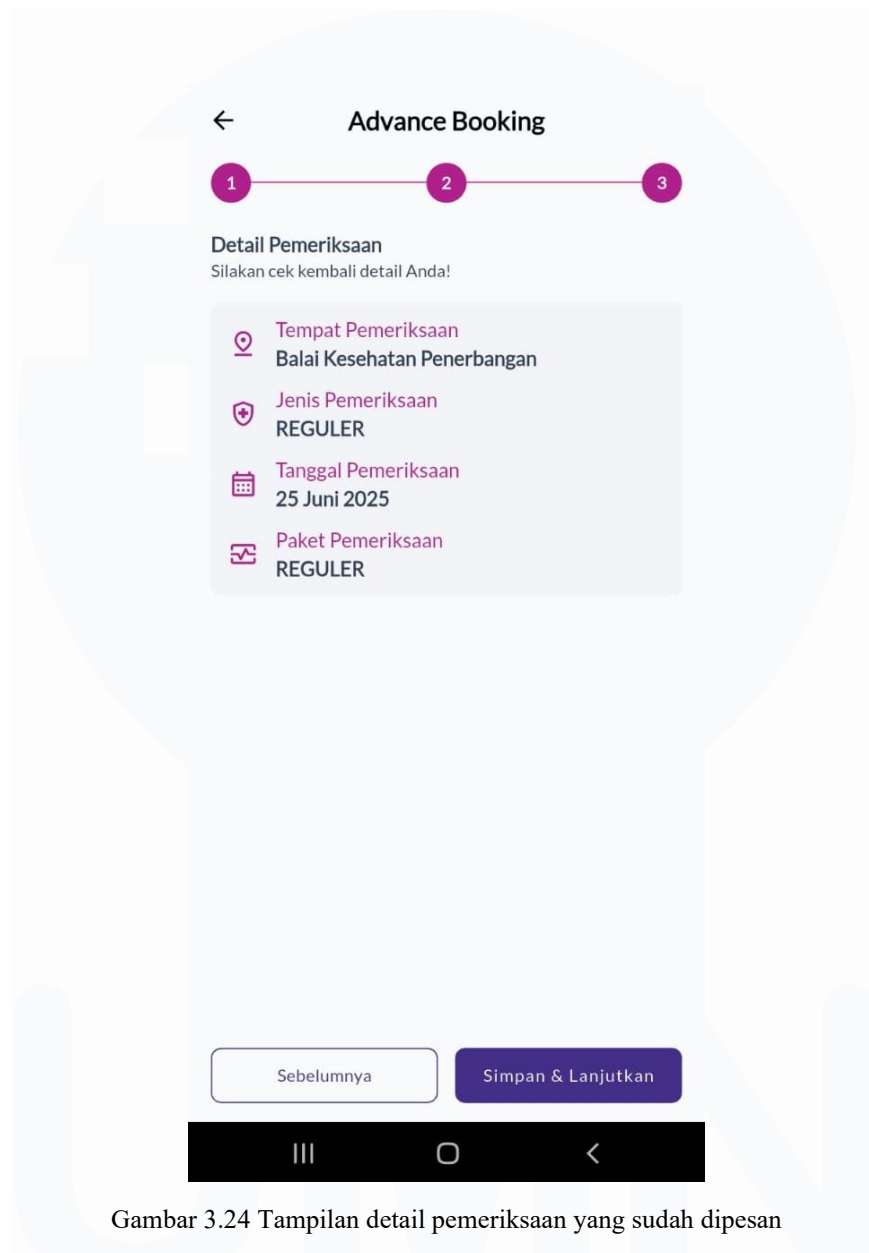
The screenshot shows a mobile application interface for 'Advance Booking'. At the top, there is a progress bar with three steps: 1 (selected), 2, and 3. Below the progress bar, the title 'Advance Booking' is displayed. The main section is titled 'Tanggal Pemeriksaan' (Examination Date) with the instruction 'Pilih tanggal pemeriksaan' (Select examination date). A calendar for June 2025 is shown, with the date 24 selected. Below the calendar, there is a section titled 'Paket Pemeriksaan' (Examination Package) with three options: 'Reguler' (selected), 'FAA', and 'Reguler & FAA'. At the bottom, there are two buttons: 'Sebelumnya' (Previous) and 'Simpan & Lanjutkan' (Save & Continue).

Gambar 3.23 Tampilan pada saat pemesanan jadwal pemeriksaan

Di tahap pada Gambar 3.23, pengguna dapat memilih tanggal pemeriksaan menggunakan kalender interaktif. Sistem memungkinkan pengguna untuk melihat tanggal-tanggal yang tersedia dan menentukan hari pemeriksaan berdasarkan preferensi pribadi atau kebutuhan maskapai. Fitur ini dirancang agar pengguna dapat merencanakan jadwal pemeriksaan jauh-jauh hari secara mandiri tanpa perlu datang langsung ke tempat layanan.

Setelah tanggal dipilih, pengguna diarahkan untuk memilih jenis paket pemeriksaan. Terdapat beberapa opsi seperti *Reguler*, *FAA*, dan kombinasi keduanya. Pilihan paket ini disesuaikan dengan kebutuhan medis atau regulasi dari maskapai penerbangan masing-masing. Antarmuka yang bersih dan intuitif

memudahkan pengguna dalam membuat keputusan dan menghindari kesalahan input.



Gambar 3.24 Tampilan detail pemeriksaan yang sudah dipesan

Semua data yang telah diinput mulai dari tempat pemeriksaan, jenis pemeriksaan, tanggal, hingga paket layanan akan ditampilkan secara ringkas dan terstruktur seperti pada Gambar 3.24. Pengguna diminta memeriksa ulang data tersebut sebelum melanjutkan proses dengan menekan tombol “Simpan & Lanjutkan.” Langkah ini krusial untuk memastikan bahwa data yang tersimpan di

sistem telah akurat dan dapat langsung ditindaklanjuti oleh tim administrasi dan medis BLU Balai Kesehatan Penerbangan.

Tabel 3. 2 Penilaian Aplikasi Hatpen Mobile

No.	Aspek yang Dinilai	Deskripsi Penilaian	Skor (1-5)	Catatan Tambahan
1.	Kemudahan Login dan Autentikasi	Proses login cukup intuitif dengan captcha sebagai pengaman tambahan.	4	Diperlukan pengingat password agar tidak mudah lupa.
2.	Verifikasi Data Pengguna	Ringkasan data setelah login membantu memastikan identitas sudah benar.	5	Fitur konfirmasi sangat berguna untuk mencegah salah akun.
3.	Navigasi Halaman Beranda	Antarmuka jelas dan pilihan menu mudah dipahami.	4	Posisi ikon notifikasi dapat sedikit mengganggu tampilan.
4.	Fitur Riwayat Pemeriksaan	Menyediakan riwayat lima pemeriksaan terakhir secara ringkas dan informatif.	5	Ideal untuk tracking data medis personal.
5.	Fitur Advance Booking Pemeriksaan	Proses pemesanan mudah dipahami dengan alur yang jelas dan terstruktur.	5	Tanggal pemeriksaan dan jenis layanan ditampilkan dengan jelas.
6.	Visual dan Desain Antarmuka	Kombinasi warna dan ikon menarik serta konsisten.	4	Warna tombol bisa diperkuat untuk aksesibilitas pengguna lansia.
7.	Kesesuaian Fungsi dan Respons Sistem	Responsif saat digunakan dan tidak ditemukan bug selama pengujian.	5	Performa sistem sudah optimal di perangkat mobile.

Berdasarkan hasil pengujian fungsional pada Tabel 3.2, aplikasi *Pelayanan Medical Examination Berbasis Online* menunjukkan performa yang optimal dalam mendukung layanan digitalisasi pemeriksaan kesehatan bagi personel penerbangan. Evaluasi dilakukan melalui sejumlah aspek, mulai dari kemudahan akses, keandalan sistem, hingga kenyamanan tampilan visual, dengan pendekatan berbasis pengalaman langsung sebagai bentuk simulasi penggunaan sistem.

Dari sisi keamanan dan aksesibilitas, fitur login dilengkapi dengan verifikasi ganda melalui kode keamanan sehingga meningkatkan proteksi data pengguna. Setelah berhasil masuk, sistem menampilkan informasi personal secara ringkas dan akurat melalui tampilan *Login Summary*. Proses verifikasi identitas ini meminimalkan risiko kesalahan data dan mendukung efisiensi proses administrasi pemeriksaan.

Kemudian, pada halaman utama aplikasi, navigasi diarahkan secara intuitif melalui fitur seperti *Rekam Medis*, *Daftar Re-check*, dan *Advance Booking*. Penyajian informasi riwayat pemeriksaan tampil ringkas namun cukup informatif, serta dilengkapi dengan opsi pendaftaran ulang jika pemeriksaan re-check diperlukan. Proses pemesanan melalui *Advance Booking* juga sudah terstruktur dalam beberapa tahap mulai dari pemilihan tanggal, jenis pemeriksaan, hingga konfirmasi jadwal secara sistematis.

Secara visual, desain antarmuka menggunakan palet warna yang konsisten serta ikon yang mudah dikenali, mendukung keterbacaan dan kenyamanan. Dalam uji coba, tidak ditemukan hambatan teknis signifikan dan semua elemen antarmuka berfungsi dengan stabil. Dengan total skor 32 dari 35 dan rata-rata evaluasi sebesar 4,57, aplikasi ini memenuhi standar kualitas sebagai sistem layanan kesehatan digital yang adaptif, akurat, dan efisien.

Testing aplikasi:

<https://drive.google.com/drive/folders/15u0SyDZXsZI9inQQsVFDOeJCFCDtuFe1?usp=sharing>

3.3 Kendala yang Ditemukan

Selama pelaksanaan program magang yang berfokus pada pengembangan dan pemanfaatan *dashboard* interaktif berbasis data survei kepuasan masyarakat, berbagai kendala ditemukan baik dari aspek teknis maupun alur integrasi data. Kendala-kendala ini menjadi pembelajaran penting dalam memahami kompleksitas pengolahan data berbasis pelayanan publik serta penerapan machine learning untuk mendukung proses pengambilan keputusan.

1) Implementasi awal *dashboard* sebagai media interaktif evaluasi layanan

Penggunaan *dashboard* sebagai media utama dalam mengevaluasi hasil survei merupakan pendekatan baru yang sebelumnya belum diterapkan secara sistematis di BLU Balai Kesehatan Penerbangan. Transformasi ini menuntut kesiapan data, kejelasan indikator visual, serta pemahaman pengguna internal terhadap interpretasi grafik dan metrik analisis yang ditampilkan. Kurangnya acuan standar visualisasi sebelumnya juga mempersulit tahap validasi efektivitas *dashboard* dalam pengambilan keputusan awal.

2) Perubahan struktur data selama proses pengembangan

Selama integrasi data ke dalam sistem *dashboard*, ditemukan perubahan struktur pada data survei, seperti penambahan kolom baru, format tanggal yang inkonsisten, serta nilai kosong pada entri penting. Hal ini memerlukan penyesuaian ulang terhadap tahap preprocessing dan transformasi data, khususnya untuk kolom seperti "Pengirim" yang harus dipecah menjadi dua atribut terpisah: nama dan maskapai. Perubahan tersebut turut memengaruhi susunan pipeline dan konfigurasi visualisasi.

3) Ketidakseimbangan kelas dalam data sentimen pelanggan

Pada proses klasifikasi sentimen menggunakan model machine learning, ditemui ketidakseimbangan signifikan antara jumlah data berlabel "Positif", "Netral", dan "Negatif". Dominasi label positif menyebabkan model cenderung bias dan kurang sensitif dalam mendeteksi isu atau keluhan yang bersifat krusial. Hal ini sempat menurunkan performa

klasifikasi dan berdampak pada ketepatan penyajian isu di dalam *dashboard*.

4) Akses terbatas terhadap data rekam medis atau feedback lanjutan

Dashboard interaktif hanya mencakup data dari formulir survei daring, tanpa terhubung langsung dengan sistem rekam medis atau umpan balik lanjutan dari pasien pasca-layanan. Keterbatasan akses ini menyulitkan proses triangulasi atau validasi terhadap keluhan yang dinyatakan dalam survei dengan kenyataan lapangan, sehingga insight yang ditampilkan masih bersifat satu sisi.

5) Keterbatasan waktu dalam proses penyempurnaan visualisasi lanjutan

Meskipun fitur visualisasi awal dan pendeteksi isu krusial telah diterapkan, proses penyempurnaan fitur lanjutan seperti pelacakan tren mingguan dan filtering dinamis masih belum sepenuhnya optimal akibat keterbatasan waktu. Hal ini menyebabkan beberapa bagian *dashboard* masih bersifat statis dan belum sepenuhnya otomatis dalam memperbarui informasi berdasarkan data terbaru yang masuk.

3.4 Solusi atas Kendala yang Ditemukan

Untuk mengatasi berbagai tantangan yang muncul selama proses pengembangan dan pemanfaatan *dashboard* interaktif, telah dilakukan sejumlah langkah strategis guna memastikan keberhasilan implementasi sistem berbasis data survei ini. Setiap solusi ditujukan untuk mengatasi hambatan teknis dan metodologis agar *dashboard* dapat memberikan informasi yang akurat, relevan, dan bermanfaat bagi pengambilan keputusan di lingkungan pelayanan kesehatan penerbangan.

1) Penguatan pemahaman internal terhadap penggunaan *dashboard* sebagai alat monitoring

Sebagai bentuk inovasi, *dashboard* diperkenalkan secara bertahap dengan pendekatan edukatif. Penyusunan visualisasi disesuaikan dengan kebutuhan pengguna internal, termasuk pemilihan indikator yang intuitif dan label yang jelas. Dokumentasi teknis dan panduan penggunaan juga

disiapkan agar seluruh tim memiliki pemahaman yang seragam dalam membaca informasi dan menindaklanjuti hasil visualisasi.

2) Penyesuaian pipeline preprocessing untuk adaptasi struktur data baru

Sebagai respon terhadap dinamika format data, pipeline preprocessing dibangun secara modular dan fleksibel. Hal ini memungkinkan penanganan perubahan seperti pemisahan kolom, pengisian nilai kosong, dan konversi tipe data dilakukan secara otomatis tanpa memengaruhi proses utama. Pembersihan teks juga ditingkatkan dengan memperkuat deteksi bahasa serta integrasi preprocessing untuk Bahasa Indonesia dan Inggris dalam satu alur terintegrasi.

3) Penerapan teknik SMOTE untuk menangani ketidakseimbangan data sentiment

Untuk mengurangi bias klasifikasi akibat dominasi sentimen positif, digunakan metode *Synthetic Minority Over-sampling Technique* (SMOTE) pada data pelatihan model. Pendekatan ini terbukti efektif meningkatkan akurasi dan sensitivitas model terhadap kategori minoritas, seperti “Negatif”. Dengan distribusi kelas yang lebih seimbang, sistem klasifikasi dapat memberikan hasil yang lebih adil dan mendukung visualisasi isu secara lebih akurat di *dashboard*.

4) Integrasi data relevan dari sumber pendukung sebagai acuan tambahan

Meskipun data rekam medis tidak dapat diakses langsung, upaya triangulasi dilakukan melalui analisis metadata seperti waktu pengaduan, maskapai, dan jenis pelayanan. Data tambahan ini kemudian diolah sebagai filter dalam *dashboard* untuk memberikan konteks terhadap keluhan pelanggan. Selain itu, dilakukan koordinasi untuk potensi pengembangan sistem yang terhubung langsung ke catatan pelayanan atau histori pemeriksaan pelanggan.

5) Penyederhanaan prioritas fitur dan pemodelan iteratif visualisasi

Agar keterbatasan waktu tidak menghambat pengembangan, dilakukan segmentasi tahapan fitur berdasarkan prioritas. Fitur esensial seperti klasifikasi sentimen, ringkasan topik krusial, dan tampilan tren

utama diselesaikan terlebih dahulu. Fitur pelengkap seperti filtering lanjutan dan pembaruan otomatis dijadwalkan dalam iterasi berikutnya. Dengan pendekatan iteratif ini, *dashboard* tetap dapat digunakan secara fungsional selama proses pengembangan berlanjut.

