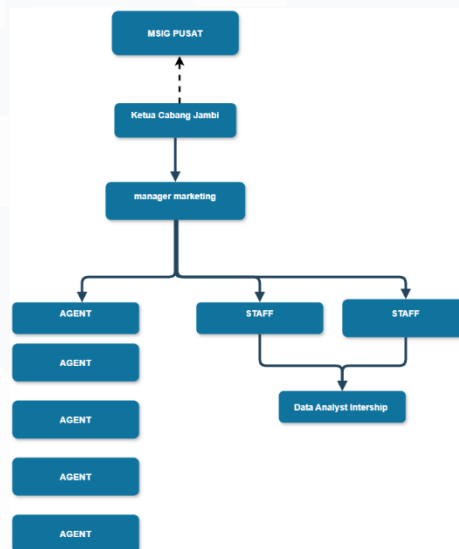


BAB III

PELAKSANAAN KERJA MAGANG

3.1 Kedudukan dan Koordinasi

Selama melaksanakan Magang Di MSIG Life. Posisi data analyst intern ditempatkan Bersama bagian staff dalam struktur organisasi. Yang berkordinasi langsung dengan manager marketing untuk melakukan menganalisis data polis nasabah dan pengklaiman nasabah pada Perusahaan tersebut.



Gambar 3.1 Stuktur organisasi Perusahaan Cabang jambi

Dalam struktur organisasi MSIG Life Cabang Jambi, Manager Marketing memiliki peran sentral dalam menjalankan fungsi pemasaran dan operasional di tingkat cabang. Posisi ini berada langsung di bawah koordinasi Ketua Cabang Jambi dan bertanggung jawab dalam mengelola tim yang terdiri dari agent serta staff administratif. Di bawah Manager Marketing, terdapat dua jalur koordinasi utama: jalur operasional yang terdiri dari beberapa agent, serta jalur administratif yang terdiri dari dua orang staff. Kedua staff tersebut memiliki tanggung jawab mendukung kelancaran aktivitas marketing dari sisi administratif dan dokumentasi. Salah satu staff bertugas membimbing serta mengoordinasikan kegiatan Data Analyst Internship, yang merupakan bagian dari program magang. Posisi intern ini berfokus pada pengolahan dan analisis data operasional serta pemasaran, guna memberikan insight yang bermanfaat dalam mendukung pengambilan keputusan. Koordinasi antara intern dengan staff dilakukan secara langsung, sementara pelaporan hasil kerja tetap berada di bawah pengawasan Manager Marketing secara

tidak langsung. Susunan ini menunjukkan hubungan kerja yang sistematis dan kolaboratif dalam pelaksanaan tugas di lingkungan cabang.

3.2 Tugas dan Uraian Kerja Magang

Selama masa magang berlangsung, sebagai Data Analysis di lingkungan kerja yang terfokus pada proses digitalisasi dan pengelolaan data operasional perusahaan. bertugas untuk mengelola data internal yang bersumber dari sistem informasi perusahaan (ORION), khususnya data yang berkaitan dengan proses pengajuan polis (SPAJ) dan klaim asuransi nasabah. Data ini meliputi informasi terkait tanggal pengajuan, jenis produk asuransi, metode pembayaran, status SPAJ, serta berbagai informasi terkait proses klaim seperti penyebab klaim, metode klaim, dan status klaim. bertanggung jawab melakukan proses pembersihan dan validasi data, termasuk menghapus duplikasi, mengisi data yang hilang, serta memastikan kesesuaian format data agar siap dianalisis lebih lanjut. Setelah proses persiapan data selesai, kemudian membangun beberapa dashboard interaktif menggunakan Power BI untuk menampilkan tren, pola, dan insight penting dari data tersebut. Visualisasi ini meliputi grafik tren SPAJ, status klaim berdasarkan waktu, serta analisis performa berdasarkan metode klaim atau agen. Serta melakukan eksplorasi awal terhadap algoritma machine learning untuk membangun model prediksi sederhana, dengan tujuan mengidentifikasi pola keterlambatan atau anomali dalam proses SPAJ dan klaim. Hasil visualisasi dan analisis ini disusun dalam bentuk dokumentasi teknis dan laporan magang yang menjadi salah satu keluaran akhir dari kegiatan ini.

Tabel 3.1 Uraian Kerja Magang

No	Pekerjaan	Minggu	Tanggal mulai	Tanggal selesai
Data cleaning				

1	Mengumpulkan data perusahaan (ORION)	1	8 April 2025	15 April 2025
2	Membersihkan data	2	16 April 2025	23 April 2025
3	Validasi Konsistensi data	3	23 April 2025	1 April 2025
Dashboard interaktif				
4	Dashboard Ringkasan Klaim Asuransi	4	1 April 2025	12 April 2025
5	Dashboard Monitoring SPAJ	5	13 April 2025	22 April 2025
6	Dashboard Analisis Premi & Produk SPAJ	6	23 April 2025	1 Juni 2025
7	Dashboard Kinerja Agen & Produksi SPAJ	7	2 Juni 2025	9 Juni 2025
Machine learning				
8	Praproses data: Label Encoding dan pembagian data training/testing	8	10 Juni 2025	16 Juni 2025
9	Pembuatan model klasifikasi dengan Decision Tree & evaluasi performa	9	17 Juni 2025	23 Juni 2025
10	Pembuatan dan evaluasi model Random Forest	10	24 Juni 2025	30 Juni 2025

3.2.1 Mengumpulkan data perusahaan (ORION)

Langkah awal yang dilakukan dalam proses kerja magang adalah mengumpulkan data dari sistem internal perusahaan, yaitu ORION, yang merupakan sistem operasional utama di MSIG Life Indonesia Tbk. ORION menyimpan berbagai informasi penting terkait aktivitas operasional, seperti data SPAJ (Surat Permintaan Asuransi Jiwa), data klaim asuransi nasabah, serta histori transaksi dan proses administratif lainnya. Mahasiswa melakukan

koordinasi langsung dengan bagian terkait di perusahaan untuk memperoleh akses data yang dibutuhkan, baik dalam bentuk file Excel maupun melalui pengambilan data secara langsung dari sistem. Proses pengumpulan data ini menjadi fondasi utama dalam pekerjaan magang, karena seluruh proses analisis, pembuatan dashboard, dan pengembangan model machine learning sangat bergantung pada kualitas dan kelengkapan data yang diperoleh dari sistem ORION tersebut. Selain itu, pemahaman terhadap struktur dan konteks data juga diperoleh selama proses pengumpulan, yang kemudian sangat membantu dalam tahap data cleaning dan pengolahan lanjutan.

3.2.2 Membersihkan data

Setelah data diperoleh dari sistem ORION, langkah selanjutnya yang dilakukan adalah proses pembersihan data atau *data cleaning*. Tahapan ini sangat krusial untuk memastikan bahwa data yang akan dianalisis benar-benar akurat, konsisten, dan siap digunakan dalam pengolahan lebih lanjut. Pada proses ini, mahasiswa mengidentifikasi dan menghapus duplikasi data yang sering terjadi pada pencatatan SPAJ dan klaim asuransi. Selain itu, dilakukan pengecekan terhadap nilai-nilai kosong (*missing values*) dan kesalahan format, terutama pada kolom tanggal, angka, serta teks. Data dengan format tidak seragam, seperti tanggal yang tidak konsisten atau nilai numerik yang tercampur dengan simbol, diubah ke dalam bentuk yang standar agar dapat dibaca oleh perangkat lunak analitik seperti Power BI dan Python.

3.2.3 Validasi Konsistensi data

Setelah proses pembersihan data selesai, langkah penting berikutnya adalah melakukan validasi konsistensi data untuk memastikan bahwa seluruh informasi yang akan digunakan dalam analisis bersifat logis, seragam, dan tidak menimbulkan bias dalam interpretasi. Validasi ini dilakukan dengan memeriksa keterkaitan antar kolom, seperti memastikan bahwa tanggal pengajuan SPAJ tidak lebih lambat dari tanggal pengiriman dokumen ke JNT, atau bahwa status klaim telah sesuai dengan logika proses bisnis (misalnya, klaim yang belum diproses tidak memiliki nilai pembayaran). Selain itu,

dilakukan juga pemeriksaan terhadap nilai-nilai yang ekstrem atau outlier yang berpotensi menimbulkan kesalahan dalam analisis, seperti durasi proses yang bernilai negatif atau tidak wajar

3.2.4 Dashboard Ringkasan Klaim Asuransi

Pembuatan Dashboard Ringkasan Klaim Asuransi dilakukan untuk menjawab kebutuhan perusahaan dalam menyajikan data klaim secara lebih efisien, informatif, dan mudah dipahami oleh berbagai pihak internal seperti tim operasional, supervisor, maupun manajemen. Dalam lingkungan perusahaan asuransi, pengelolaan klaim merupakan salah satu proses yang kompleks dan melibatkan berbagai tahapan, mulai dari pengajuan oleh nasabah, verifikasi data, hingga pencairan dana klaim. Tanpa alat bantu visual yang terintegrasi, pemantauan terhadap performa klaim seringkali harus dilakukan secara manual melalui laporan tabel atau file spreadsheet yang kurang efektif dalam menggambarkan kondisi secara keseluruhan.

3.2.5 Dashboard Monitoring SPAJ

Dashboard Monitoring SPAJ (Surat Permintaan Asuransi Jiwa) dikembangkan untuk membantu perusahaan dalam memantau proses pengajuan polis asuransi dari awal hingga status akhir secara lebih terstruktur dan real-time. Dalam proses bisnis asuransi jiwa, SPAJ merupakan dokumen penting yang diajukan oleh calon nasabah dan menjadi titik awal dalam proses akuisisi polis. Tanpa pemantauan yang sistematis, proses ini rentan terhadap keterlambatan, kehilangan data, atau ketidaksesuaian antara dokumen dan data sistem.

3.2.6 Dashboard Analisis Premi & Produk SPAJ

Dashboard Analisis Premi & Produk SPAJ dikembangkan sebagai sarana untuk memahami tren pengajuan polis dan distribusi premi berdasarkan jenis produk asuransi jiwa yang ditawarkan oleh MSIG Life Indonesia Tbk. Dalam industri asuransi, analisis terhadap variasi produk dan besaran premi sangat penting untuk mengetahui preferensi pasar, kinerja agen, serta efektivitas strategi pemasaran yang diterapkan.

3.2.7 Dashboard Kinerja Agen & Produksi SPAJ

Dashboard Kinerja Agen & Produksi SPAJ dirancang untuk memberikan gambaran menyeluruh mengenai produktivitas dan kontribusi masing-masing agen asuransi dalam proses pemasaran dan pengajuan polis SPAJ (Surat Permintaan Asuransi Jiwa). Dashboard ini menjadi alat bantu yang strategis bagi manajemen dalam mengevaluasi performa individu maupun kelompok agen secara kuantitatif dan visual.

3.2.8 Praproses data: Label Encoding dan pembagian data training/testing

mempersiapkan data sebelum dimasukkan ke dalam model machine learning yang akan digunakan untuk klasifikasi prediksi keterlambatan klaim asuransi. Salah satu proses utama yang dilakukan adalah Label Encoding, yaitu teknik untuk mengubah data kategorikal (seperti jenis klaim, penyebab klaim, dan metode klaim) menjadi bentuk numerik agar dapat dikenali oleh algoritma machine learning. Proses ini dilakukan dengan menggunakan library Python seperti LabelEncoder dari `sklearn.preprocessing`, yang secara otomatis memberikan nilai numerik untuk setiap kategori unik dalam fitur tertentu. Setelah proses encoding selesai, data kemudian dibagi menjadi dua bagian: data training dan data testing. Pembagian ini dilakukan dengan menggunakan fungsi `train_test_split` dari `sklearn.model_selection`, dengan proporsi umum seperti 70% untuk pelatihan dan 30% untuk pengujian.

3.2.9 Pembuatan model klasifikasi dengan Decision Tree & evaluasi performa

Setelah data melalui tahap praproses yang mencakup label encoding dan pembagian data training/testing, langkah selanjutnya adalah membangun model klasifikasi menggunakan algoritma Decision Tree. Algoritma ini dipilih karena kemampuannya dalam menangani data kategorikal dan numerik serta interpretasinya yang mudah dipahami. Model ini bertujuan untuk mengklasifikasikan apakah sebuah klaim asuransi akan diproses tepat waktu atau mengalami keterlambatan, berdasarkan fitur-fitur input seperti jenis klaim, penyebab klaim, metode klaim, dan durasi proses. Hasil evaluasi dari model Decision Tree menjadi indikator awal terhadap efektivitas pendekatan

klasifikasi yang diterapkan. Jika ditemukan ketidakseimbangan atau performa yang belum optimal, maka model dapat ditingkatkan dengan tuning parameter atau menggunakan pendekatan lain seperti Random Forest. Dengan membangun model ini, perusahaan memiliki alat bantu untuk memprediksi risiko keterlambatan klaim, sehingga dapat mengambil langkah antisipatif lebih awal dan meningkatkan kualitas pelayanan kepada nasabah.

3.2.10 Pembuatan dan evaluasi model Random Forest

analisis data klaim asuransi di MSIG Life adalah mengembangkan model klasifikasi menggunakan algoritma Random Forest. Random Forest dipilih karena kemampuannya yang unggul dalam meningkatkan akurasi prediksi dan mengurangi risiko overfitting yang umum terjadi pada model Decision Tree tunggal. Algoritma ini bekerja dengan membangun banyak pohon keputusan (decision trees) secara acak dan menggabungkan hasilnya untuk menghasilkan prediksi akhir yang lebih stabil dan akurat. Hasil evaluasi menunjukkan bahwa Random Forest memberikan performa yang lebih baik dibandingkan dengan Decision Tree, terutama dalam menangani variasi data dan meningkatkan sensitivitas terhadap klaim yang tergolong terlambat. Dengan penerapan model ini, perusahaan mendapatkan insight prediktif yang lebih akurat terkait potensi keterlambatan klaim, yang selanjutnya dapat digunakan sebagai dasar dalam pengambilan keputusan strategis dan peningkatan efisiensi layanan.

Dalam proses pelaksanaan magang di MSIG Life Indonesia Tbk Cabang Jambi, menggunakan beberapa perangkat lunak dan tools analisis data yang membantu dalam mengolah, menganalisis, dan menyajikan data secara visual berkaitan data SPAJ dan Data klaim nasabah :

1. Microsoft Excel

pada tahap awal sebagai alat bantu eksplorasi dan pembersihan data mentah. Excel memungkinkan mahasiswa untuk membuka file dengan cepat, melakukan peninjauan awal terhadap struktur dataset, serta membersihkan data dari duplikasi, nilai kosong, dan kesalahan

format sederhana. Selain itu, Excel juga sangat membantu dalam perhitungan dasar seperti menghitung jumlah kasus, merangkum data berdasarkan kategori tertentu, dan membuat pivot table sederhana sebagai langkah awal sebelum masuk ke tahap visualisasi atau analisis lanjutan.

2. Python

Tabel 3.2 Python dan VS code

Bahasa pemrograman	Kelebihan	kekurangan
Python	• Sintaks sederhana dan mudah dibaca • Cocok untuk AI, ML, dan analisis data • Komunitas besar	• Kecepatan eksekusi lambat • Kurang efisien untuk aplikasi mobile dan real-time
VS code	• Editor ringan dan fleksibel • Mendukung banyak bahasa pemrograman,	• Perlu instalasi ekstensi tambahan untuk fungsional penuh

Selama pelaksanaan magang, saya lebih memilih Python karena Python merupakan inti dari pengolahan dan analisis data, sedangkan VS Code hanyalah alat bantu (editor) untuk menulis kode. Python memiliki kekuatan sebagai bahasa pemrograman yang kaya akan pustaka (library) seperti pandas untuk data cleaning, matplotlib untuk visualisasi data, serta scikit-learn untuk membangun dan mengevaluasi model machine learning seperti Decision Tree dan Random Forest. Semua proses penting dalam proyek—mulai dari pengolahan data klaim asuransi, pembuatan model klasifikasi, hingga evaluasi performa—dilakukan dengan Python.

3. Power BI

Tabel 3.3 Power BI dan Tablue

Perbandingan	Power BI	Tablue
Keunggulan	• Terintegrasi erat dengan Microsoft Excel & Office 365 • Biaya lisensi lebih murah	• Visualisasi lebih interaktif dan powerful • Cocok untuk data besar dan analisis kompleks
Kelemahan	• Visualisasi kurang fleksibel untuk kustomisasi tinggi • Kurang efisien untuk data besar	• Biaya lisensi lebih mahal • Kurva pembelajaran sedikit lebih tinggi dibanding Power BI

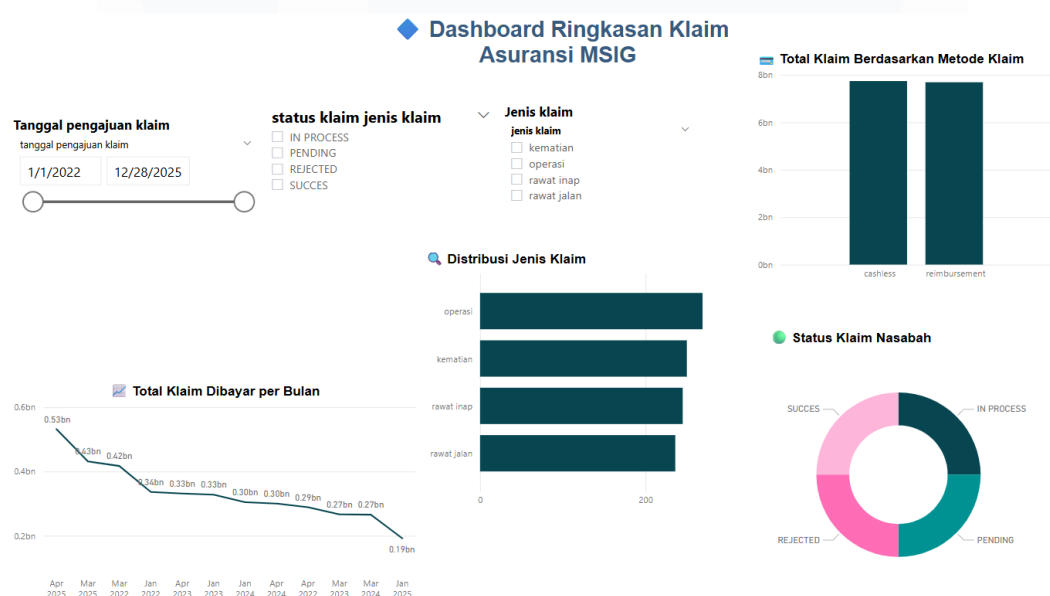
Selama pelaksanaan magang, saya lebih memilih tablue karena kebutuhan analisis data klaim dan SPAJ (Surat Permintaan Asuransi Jiwa) membutuhkan visualisasi yang interaktif, cepat, dan mendalam, terutama untuk mendeteksi keterlambatan klaim serta menganalisis sebaran wilayah atau produk. Tableau memiliki kemampuan drag-and-drop visualisasi yang kuat, serta mampu menangani data besar dengan performa tinggi. Selain itu, Tableau mempermudah saya dalam membuat dashboard yang dinamis, user-friendly, dan dapat dieksplorasi secara langsung oleh user.

3.2.9 Pembuatan dan pemanfaatan dashboard interaktif

3.2.9.1 Dashboard Ringkasan Klaim Asuransi MSIG

Dashboard Ringkasan Klaim Asuransi MSIG dibuat untuk memberikan gambaran menyeluruh mengenai proses dan status klaim nasabah di MSIG Life Indonesia. Dashboard ini dirancang agar pihak manajemen dapat dengan mudah memantau tren pengajuan klaim, mengetahui penyebab klaim yang paling sering

terjadi, serta mengevaluasi durasi proses klaim dari awal hingga selesai. Dengan menyajikan data dalam bentuk visual seperti grafik batang, pie chart, dan tren waktu, dashboard ini mempermudah proses identifikasi area layanan yang memerlukan perbaikan atau peningkatan efisiensi. Selain itu, dashboard ini juga membantu dalam mengetahui distribusi klaim berdasarkan jenis klaim, metode pengajuan, dan status klaim yang telah dibayarkan atau ditolak. Dengan informasi yang akurat dan dapat diakses secara interaktif, manajemen dapat mengambil keputusan berbasis data secara lebih cepat dan tepat sasaran. Dashboard ini menjadi alat bantu strategis untuk meningkatkan transparansi, efisiensi, dan kualitas layanan klaim kepada nasabah, serta mendukung langkah-langkah perbaikan berkelanjutan di lingkungan perusahaan.



Gambar 3.2 Dashboard ringkasan Klaim asuransi MSIG

Dashboard Ringkasan Klaim Asuransi MSIG dirancang untuk memberikan wawasan komprehensif mengenai aktivitas klaim asuransi nasabah selama periode tertentu. Dashboard ini menyajikan informasi kunci yang membantu tim manajemen dalam memantau dan menganalisis data klaim secara efektif, mulai dari jenis klaim, metode klaim yang digunakan, status penyelesaian klaim, hingga total nominal klaim yang dibayarkan setiap bulannya. Dengan adanya filter interaktif seperti rentang Tanggal Pengajuan Klaim, Status Klaim, dan Jenis Klaim, pengguna

dapat menyaring data sesuai kebutuhan untuk keperluan evaluasi yang lebih spesifik.

Visualisasi pertama yang disorot adalah grafik Total Klaim Berdasarkan Metode Klaim, yang menampilkan perbandingan jumlah klaim antara metode cashless dan reimbursement. Grafik ini membantu mengidentifikasi preferensi metode klaim nasabah serta efektivitas penanganan berdasarkan metode tersebut. Visualisasi kedua, yaitu diagram pie Status Klaim Nasabah, menunjukkan proporsi klaim dalam berbagai status seperti success, in process, pending, dan rejected. Ini memberikan gambaran umum terhadap kecepatan dan keberhasilan proses klaim yang sedang berjalan maupun yang sudah selesai.

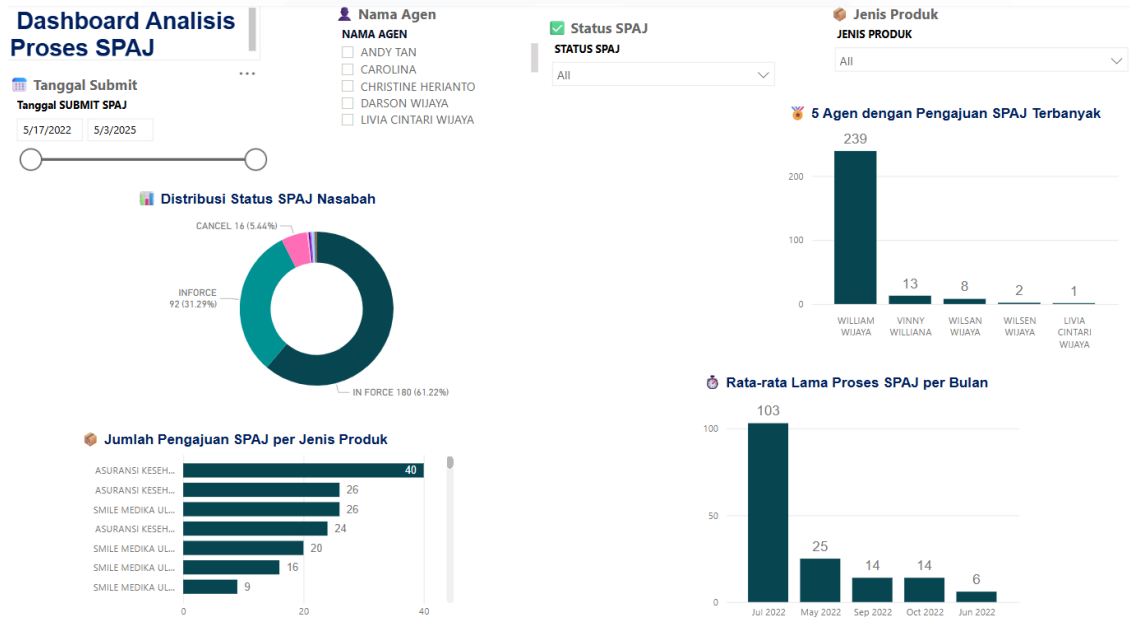
Selanjutnya, Visualisasi Distribusi Jenis Klaim menyajikan jumlah klaim berdasarkan kategori seperti operasi, kematian, rawat inap, dan rawat jalan. Dengan informasi ini, manajemen dapat mengidentifikasi jenis klaim yang paling sering diajukan oleh nasabah, sehingga bisa mempersiapkan strategi layanan yang lebih sesuai. Selain itu, visualisasi Total Klaim Dibayar per Bulan dalam bentuk grafik garis menyajikan tren nominal pembayaran klaim dari waktu ke waktu. Grafik ini berguna untuk mengetahui pola pengeluaran perusahaan dan mendeteksi bulan-bulan tertentu dengan beban klaim tinggi.

3.2.9.2 Dashbaord Analisis Proses SPAJ

Pembuatan dashboard analisis proses SPAJ (Surat Permintaan Asuransi Jiwa) dilakukan sebagai upaya untuk mempermudah pemantauan dan evaluasi alur proses SPAJ secara real-time dan efisien. Dalam proses bisnis asuransi jiwa, SPAJ merupakan dokumen penting yang menjadi awal dari proses akseptasi polis oleh perusahaan. Namun, dalam praktiknya, proses ini bisa mengalami kendala seperti keterlambatan input data, kekurangan dokumen pendukung, atau lamanya waktu persetujuan.

Melalui dashboard interaktif, seluruh tahapan proses SPAJ—mulai dari pengajuan, verifikasi, underwriting, hingga polis terbit—dapat divisualisasikan secara ringkas dan mudah dipahami. Manajemen dapat melihat secara langsung berapa banyak SPAJ yang tertunda, yang sudah selesai, atau yang sedang dalam proses, serta rata-rata durasi penyelesaian di tiap tahapan. Dengan begitu, dashboard ini tidak hanya menjadi alat pemantauan, tetapi juga sebagai dasar pengambilan keputusan untuk

memperbaiki efisiensi operasional, meningkatkan pelayanan kepada calon nasabah, dan meminimalisasi risiko keterlambatan dalam penerbitan polis.



Gambar 3.3 Dashboard Analisis Proses SPAJ

Pada gambar 3.3 dapat dilihat dashboard dirancang sebagai alat bantu visual yang bertujuan untuk memantau performa pengajuan surat permintaan asuransi jiwa (SPAJ). Dashboard ini memanfaatkan beberapa filter interaktif seperti rentang Tanggal Submit SPAJ, Nama Agen, Status SPAJ, dan Jenis Produk, yang memungkinkan pengguna menyesuaikan tampilan data sesuai kebutuhan analisis. manajer cabang atau staf operasional dapat mengevaluasi kinerja dan efisiensi proses pengajuan SPAJ secara lebih akurat dan fleksibel.

visualisasi pertama adalah grafik Distribusi Status SPAJ Nasabah dalam bentuk diagram pie, yang memperlihatkan proporsi pengajuan SPAJ berdasarkan status akhir, seperti “INFORCE” (aktif) atau “CANCEL” (dibatalkan). Dari grafik tersebut terlihat bahwa sebagian besar SPAJ telah berstatus aktif, menunjukkan tingkat keberhasilan yang tinggi dalam proses pengajuan. Visualisasi kedua yaitu Jumlah Pengajuan SPAJ per Jenis Produk, yang disajikan dalam bentuk grafik batang horizontal, menampilkan informasi mengenai frekuensi pengajuan berdasarkan produk asuransi yang dipilih nasabah. Melalui grafik ini, perusahaan dapat dengan mudah mengidentifikasi produk yang paling diminati atau mendeteksi produk yang kurang populer untuk evaluasi lebih lanjut.

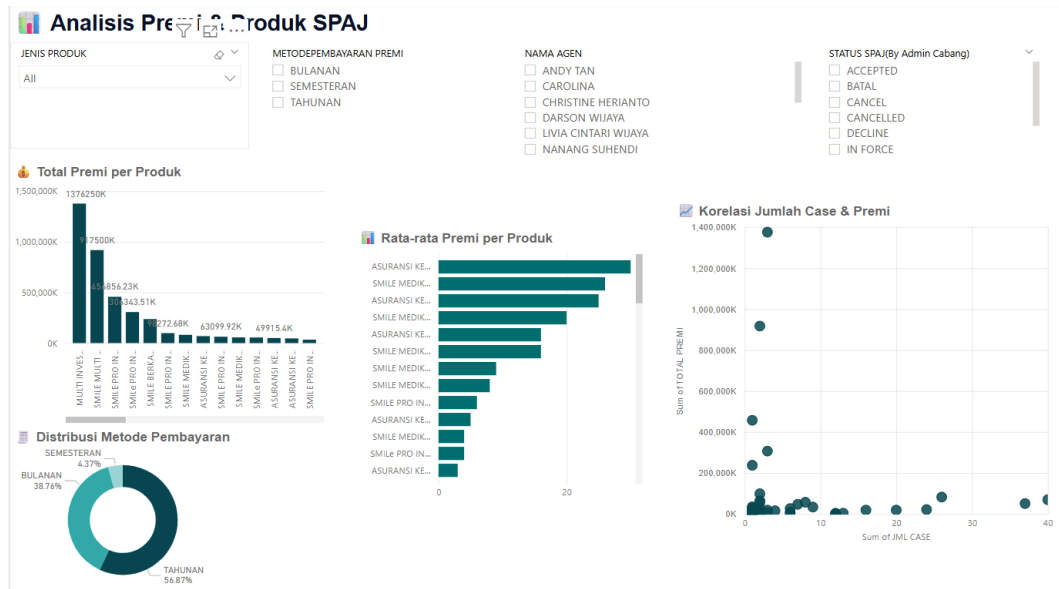
Selain itu, terdapat visualisasi ketiga adalah 5 Agen dengan Pengajuan SPAJ Terbanyak dalam bentuk grafik batang vertikal, yang menunjukkan tingkat produktivitas masing-masing agen. Agen yang paling menonjol dalam pengajuan SPAJ dapat dikenali dengan cepat, sehingga dashboard ini juga mendukung strategi insentif atau pelatihan berbasis kinerja individu. Terakhir, grafik Rata-rata Lama Proses SPAJ per Bulan memberikan informasi mengenai seberapa cepat proses SPAJ diselesaikan dari waktu submit hingga pengiriman dokumen. Data ini sangat berguna untuk mengukur efisiensi proses operasional dari waktu ke waktu serta menemukan bulan-bulan tertentu yang menunjukkan keterlambatan signifikan.

3.2.9.3 Dashboard Analisis Premi & Produk SPAJ

Dashboard Analisis Premi & Produk SPAJ memiliki peran penting dalam membantu perusahaan asuransi, khususnya MSIG Life Indonesia, untuk memahami dinamika produk dan kontribusi premi yang diperoleh dari setiap pengajuan SPAJ (Surat Permintaan Asuransi Jiwa). Dashboard ini menyajikan visualisasi interaktif terkait jumlah dan jenis produk asuransi yang paling banyak dipilih nasabah, tren pembayaran premi dari waktu ke waktu, serta sebaran premi berdasarkan jenis pembayaran seperti tahunan, semesteran, atau bulanan.

Melalui dashboard ini, manajemen dapat mengetahui produk asuransi mana yang paling diminati, dan bagaimana performa penjualannya dalam kurun waktu tertentu. Hal ini sangat berguna untuk merancang strategi pemasaran yang lebih tepat sasaran, mengevaluasi efektivitas kampanye produk tertentu, dan memperkuat produk-produk yang terbukti unggul di pasar. Selain itu, informasi seputar APE (Annual Premium Equivalent) yang ditampilkan juga membantu dalam mengukur kontribusi premi dari setiap agen atau channel distribusi.

dashboard ini juga memberikan gambaran mengenai pola pembayaran premi nasabah, sehingga perusahaan dapat mengidentifikasi kecenderungan nasabah dalam memilih metode pembayaran, serta memperkirakan cash flow masuk dari premi secara lebih akurat. perusahaan dapat meningkatkan efisiensi dalam pengelolaan produk dan perencanaan pendapatan, serta memperkuat posisi kompetitif di industri asuransi jiwa.



Gambar 3.4 Dashboard Analisis Premi & Produk SPAJ

Dashboard Analisis Premi & Produk SPAJ ini dirancang untuk memberikan gambaran menyeluruh mengenai kinerja penjualan produk asuransi di MSIG Life, terutama dari sisi jumlah premi yang terkumpul, jenis produk yang diminati metode pembayaran yang dipilih nasabah, serta korelasi antara jumlah pengajuan SPAJ (JML CASE) dengan total premi. Dashboard ini juga dilengkapi dengan fitur interaktif seperti slicer Jenis Produk, Metode Pembayaran, Nama Agen, dan Status SPAJ untuk memudahkan pengguna melakukan eksplorasi data secara lebih mendalam.

Visualisasi pertama, Total Premi per Produk, menunjukkan akumulasi premi dari masing-masing produk asuransi. Dari grafik ini terlihat bahwa produk "MULTI INVEST" dan "SMILE MULTI" menjadi kontributor terbesar dalam pendapatan premi, sehingga bisa menjadi fokus utama untuk strategi pemasaran berikutnya.

Visualisasi kedua, Rata-rata Premi per Produk, menampilkan nilai rata-rata premi yang dibayarkan per kasus untuk setiap produk. Ini membantu manajemen melihat produk mana yang bernilai tinggi per unit SPAJ dan dapat menjadi prioritas untuk segmen nasabah tertentu.

Visualisasi ketiga, Distribusi Metode Pembayaran, menunjukkan proporsi antara metode pembayaran tahunan, semesteran, dan bulanan. Dari grafik donat terlihat bahwa sebagian besar nasabah memilih metode tahunan (56.87%), diikuti

bulanan (38.76%), dan semesteran hanya 4.37%. Informasi ini penting untuk pengelolaan cash flow dan perencanaan keuangan perusahaan.

Visualisasi keempat, Korelasi Jumlah Case & Premi, menggunakan grafik scatter plot untuk memperlihatkan hubungan antara jumlah SPAJ (JML CASE) dengan total premi. Titik-titik pada grafik mewakili agen atau produk tertentu; dari sini dapat dianalisis apakah ada agen yang menghasilkan banyak case namun dengan premi kecil, atau sebaliknya. Visualisasi ini berguna untuk mengukur efektivitas agen dalam menghasilkan nilai ekonomi.

Secara keseluruhan, dashboard ini tidak hanya mendukung monitoring kinerja penjualan produk, tetapi juga menjadi alat strategis dalam menganalisis peluang, mengevaluasi performa, serta menyusun kebijakan pemasaran dan penjualan berbasis data di lingkungan MSIG Life.

3.2.9.4 Dashboard Kinerja Agen & Produksi SPAJ

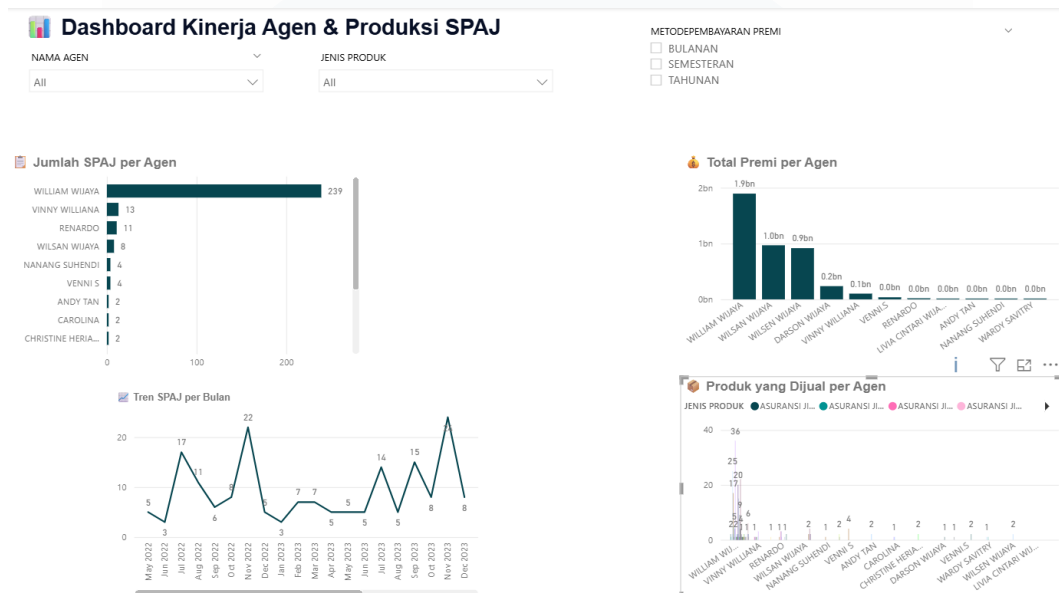
Dashboard dirancang untuk membantu manajemen Perusahaan, khususnya tim pemasaran dan manajer cabang, dalam memantau dan mengevaluasi kontribusi masing-masing agen terhadap kinerja penjualan polis (SPAJ). Dalam industri asuransi, Agen merupakan ujung tombak dalam menjangkau nasabah dan mendorong penjualan produk. Oleh karena itu, memahami seberapa besar peran dan efektivitas setiap agen menjadi hal yang sangat penting.

Dashboard ini menyajikan informasi seperti jumlah SPAJ yang diajukan oleh setiap agen, total premi yang berhasil dikumpulkan, serta performa berdasarkan waktu, status SPAJ, dan jenis produk yang dijual. Dengan dibuatnya visualisasi ini, Perusahaan dengan mudah mengidentifikasi agen-agen terbaik yang konsisten menghasilkan volume penjualan tertinggi, maupun agen yang perlu dipelatihan atau dukungan lebih.

Selain itu, dashboard ini berguna dalam merancang strategi pemberian insentif, bonus kerja, atau promosi internal. Manajemen mampu menggunakan data historis dan tren kinerja untuk membuat Keputusan dalam pengelolaan sumber daya manusia.

Secara strategis, dashboard ini mendukung peningkatan efisiensi tim penjualan, mendorong semangat kompetitif antar agen. Melalui pemantauan secara

real-time dan transparan, perusahaan dapat mengambil tindakan tepat waktu untuk meningkatkan produktivitas dan mencapai tujuan bisnis yang lebih optimal.



Gambar 3.5 Dashboard Kinerja Agen & produksi SPAJ

Dashboard Kinerja Agen & Produksi SPAJ yang ditampilkan di atas dirancang untuk memberikan gambaran menyeluruh mengenai performa para agen dalam proses pengajuan polis (SPAJ) di lingkungan MSIG. Dashboard ini terdiri dari beberapa visualisasi yang saling terhubung dan memungkinkan pengguna untuk melakukan filter berdasarkan nama agen, jenis produk, serta metode pembayaran premi (bulanan, semesteran, tahunan), sehingga memberikan fleksibilitas dalam analisis data.

Visualisasi pertama, yaitu Jumlah SPAJ per Agen, menggunakan diagram batang horizontal untuk menunjukkan seberapa banyak SPAJ yang berhasil diajukan oleh masing-masing agen. Terlihat jelas bahwa William Wijaya memiliki kontribusi paling besar dalam hal jumlah SPAJ yang diajukan, diikuti oleh beberapa agen lainnya dalam jumlah yang jauh lebih sedikit.

Visualisasi kedua, Total Premi per Agen, menggambarkan besarnya nilai premi yang berhasil dikumpulkan oleh tiap agen. Grafik ini memberikan wawasan

tambahan bahwa volume SPAJ tidak selalu berbanding lurus dengan nilai premi yang diperoleh, karena bisa jadi satu SPAJ mewakili premi yang besar.

Kemudian, visualisasi ketiga Tren SPAJ per Bulan disajikan dalam bentuk grafik garis yang menunjukkan dinamika atau fluktuasi jumlah SPAJ yang diajukan setiap bulan. Hal ini membantu manajemen memahami kapan terjadi peningkatan atau penurunan aktivitas pengajuan SPAJ, yang bisa berkaitan dengan musim promosi, kebijakan baru, atau faktor eksternal lainnya.

Terakhir, Visualisasi keempat Produk yang Dijual per Agen divisualisasikan dalam bentuk clustered bar chart (batang kelompok) berdasarkan agen dan jenis produk. Visualisasi ini sangat berguna untuk melihat preferensi produk masing-masing agen serta mendeteksi apakah terdapat agen yang hanya fokus pada satu jenis produk tertentu.

Secara keseluruhan, dashboard ini sangat penting bagi perusahaan karena dapat digunakan untuk mengevaluasi kinerja individu agen, menyesuaikan strategi pelatihan dan insentif, serta memetakan kontribusi terhadap target perusahaan berdasarkan data yang akurat dan real-time.

3.2.10 Perancangan Klasifikasi data berbasis machine learning

Proses pengajuan klaim asuransi merupakan bagian penting dalam layanan pascapenjualan yang berdampak langsung terhadap kepuasan nasabah. Di industri asuransi jiwa, seperti di MSIG Life, ketepatan waktu dalam memproses klaim menjadi salah satu tolak ukur kualitas layanan perusahaan. Nasabah mengharapkan proses klaim berjalan cepat dan transparan, namun pada kenyataannya, masih terdapat kasus keterlambatan yang menyebabkan ketidakpuasan dan berpotensi menurunkan kepercayaan pelanggan terhadap perusahaan. Saat ini, proses evaluasi performa klaim masih bersifat manual dan tidak terintegrasi dengan sistem prediktif yang mampu memperkirakan klaim mana yang berpotensi terlambat. Hal ini menjadi celah kritis yang perlu ditangani melalui pendekatan berbasis data dan kecerdasan buatan. Sebagai solusi terhadap permasalahan tersebut, dirancanglah model klasifikasi berbasis

machine learning menggunakan decision tree dan random forest yang bertujuan untuk memprediksi dan mengelompokkan klaim menjadi dua kategori, yaitu "Tepat Waktu" dan "Terlambat", berdasarkan pola historis data klaim yang diperoleh dari sistem informasi perusahaan (Orion). Dengan adanya klasifikasi ini, diharapkan manajemen dapat melakukan pengawasan lebih proaktif serta mengambil langkah antisipatif terhadap klaim yang cenderung berisiko lambat dalam prosesnya. Menggunakan dataset sebagai berikut:

A	B	C	D	E	F	G	H	I	J	K
nama	jenis kelamin	usia	daerah	tanggal pengajuan klaim	jenis klaim	penyebab klaim	jumlah klaim yang dibayar	durasi klaim	metode klaim	status klaim
Putra Nugroho	Wanita	33	Alam Barajo	15/04/2024	kematian	Kecelakaan	Rp10.009.502	15 hari	cashless	SUCCESS
Hadi Santoso	Wanita	31	Alam Barajo	18/12/2025	operasi	Konsultasi mata	Rp8.240.690	15 hari	cashless	SUCCESS
Nina Yuliana	Pria	57	Jambi Selatan	30/01/2025	rawat inap	Penyakit Dalam	Rp18.615.868	2 hari	cashless	SUCCESS
Rina Putri	Pria	61	Alam Barajo	26/04/2025	rawat jalan	Konsultasi mata	Rp11.250.921	2 hari	cashless	SUCCESS
Nina Setiawan	Wanita	47	Kotabaru	22/12/2023	rawat jalan	Kecelakaan	Rp15.477.095	2 hari	cashless	SUCCESS
Fajar Hidayat	Wanita	41	Jelutong	16/01/2024	kematian	Kecelakaan	Rp29.988.699	2 hari	reimbursement	SUCCESS
Joko Putri	Pria	61	Kotabaru	21/03/2023	rawat inap	Konsultasi mata	Rp15.164.185	7 hari	reimbursement	SUCCESS
Sari Yuliana	Pria	65	Alam Barajo	22/02/2022	kematian	Penyakit Dalam	Rp17.572.783	5 hari	reimbursement	SUCCESS
Joko Hidayat	Pria	35	Telanaipura	31/07/2024	rawat jalan	Penyakit Dalam	Rp13.239.872	5 hari	cashless	SUCCESS
Kiki Pratama	Pria	54	Danau Teluk	11/12/2022	kematian	Kelahiran	Rp25.353.287	2 hari	reimbursement	SUCCESS
Joko Kurniawan	Pria	22	Danau Sipin	26/04/2024	operasi	Penyakit Dalam	Rp6.833.825	7 hari	reimbursement	SUCCESS
Umi Hidayat	Pria	31	Jambi Selatan	19/05/2025	operasi	Konsultasi mata	Rp9.989.548	2 hari	cashless	SUCCESS
Oka Lestari	Pria	39	Danau Sipin	31/12/2022	rawat inap	Konsultasi mata	Rp29.287.805	2 hari	cashless	SUCCESS
Lina Siregar	Pria	49	Telanaipura	24/06/2022	kematian	Konsultasi mata	Rp20.861.882	5 hari	cashless	SUCCESS
Nina Nugroho	Wanita	38	Alam Barajo	24/04/2024	kematian	Kelahiran	Rp14.238.472	10 hari	reimbursement	SUCCESS
Gita Putri	Pria	22	Danau Teluk	07/11/2024	kematian	Penyakit Dalam	Rp19.906.173	10 hari	cashless	SUCCESS
Andi Lestari	Wanita	42	Jelutong	14/08/2025	kematian	Penyakit Dalam	Rp22.502.541	15 hari	cashless	SUCCESS
Dewi Santoso	Pria	26	Danau Sipin	10/04/2023	rawat jalan	Kecelakaan	Rp7.624.396	2 hari	cashless	SUCCESS
Andi Lestari	Pria	45	Jelutong	14/04/2025	rawat inap	Konsultasi mata	Rp18.243.091	10 hari	cashless	SUCCESS
Indra Kurniawan	Pria	22	Telanaipura	10/01/2025	operasi	Kecelakaan	Rp22.005.713	15 hari	cashless	SUCCESS
Dewi Saputra	Wanita	40	Pasar Jambi	15/12/2022	rawat inap	Kelahiran	Rp15.568.970	15 hari	cashless	SUCCESS
Rina Siregar	Wanita	51	Danau Teluk	03/04/2022	rawat inap	Penyakit Dalam	Rp13.056.823	7 hari	reimbursement	SUCCESS
Indra Santoso	Pria	53	Jelutong	13/08/2024	rawat inap	Kelahiran	Rp24.153.132	2 hari	reimbursement	SUCCESS
Budi Yuliana	Wanita	33	Telanaipura	21/09/2022	rawat jalan	Penyakit Dalam	Rp21.845.198	15 hari	cashless	SUCCESS
Kiki Purnama	Wanita	29	Telanaipura	20/06/2025	rawat inap	Konsultasi mata	Rp24.649.415	7 hari	reimbursement	SUCCESS
Mega Saputra	Pria	20	Jambi Timur	25/01/2025	operasi	Kelahiran	Rp7.044.503	2 hari	reimbursement	SUCCESS
Nina Pratama	Wanita	56	Jelutong	11/09/2023	rawat jalan	Konsultasi mata	Rp29.009.016	5 hari	reimbursement	SUCCESS
Eka Hidayat	Wanita	42	Jambi Timur	27/06/2025	rawat jalan	Konsultasi mata	Rp6.652.346	5 hari	cashless	SUCCESS
Mega Putri	Wanita	26	Pasar Jambi	10/04/2025	rawat jalan	Penyakit Dalam	Rp11.584.417	2 hari	cashless	SUCCESS

Gambar 3.6 Dataset Klaim nasabah MSIG

Dataset yang digunakan dalam kegiatan magang ini merupakan data historis pengajuan klaim nasabah PT MSIG Life Insurance Indonesia Tbk yang diambil dari sistem internal perusahaan. Dataset ini mencakup berbagai informasi penting terkait proses klaim yang diajukan oleh nasabah, mulai dari identitas demografis, rincian klaim, hingga status penyelesaiannya. Beberapa kolom utama yang terdapat dalam data ini antara lain nama nasabah, jenis kelamin, usia, dan daerah tempat tinggal nasabah. Selain itu, terdapat informasi mengenai tanggal pengajuan klaim, jenis klaim (seperti rawat inap, rawat jalan, operasi, maupun klaim kematian), serta penyebab klaim (misalnya karena kecelakaan, penyakit dalam, konsultasi mata, atau kelahiran).

Salah satu elemen penting dalam dataset ini adalah durasi klaim, yaitu waktu yang dibutuhkan sejak pengajuan hingga klaim selesai dibayarkan. Nilai durasi ini dinyatakan dalam satuan hari dan menjadi salah satu fokus utama analisis karena berkaitan langsung dengan ketepatan waktu layanan yang diberikan oleh perusahaan. Kolom lain seperti jumlah klaim yang dibayarkan, metode klaim

(cashless atau reimbursement), serta status klaim (misalnya SUCCES) juga disertakan sebagai bagian dari fitur analisis. Dataset ini digunakan untuk mendukung visualisasi dashboard interaktif serta menjadi bahan utama dalam pengembangan model klasifikasi berbasis machine learning, dengan tujuan memprediksi apakah klaim akan diproses tepat waktu atau mengalami keterlambatan. Melalui analisis terhadap data ini, perusahaan dapat mengidentifikasi pola-pola yang memengaruhi efisiensi layanan dan merancang strategi perbaikan yang lebih akurat.

3.2.10.1 Praproses : libraries yang dibutuhkan

In [75]:

import pandas as pd
import numpy as np
from sklearn.preprocessing import LabelEncoder

In [106]:

df = pd.read_excel('Data Klaim Nasabah MSIG.xlsx')
df

Out[106]:

	nama	jenis kelamin	usia	daerah	tanggal pengajuan klaim	jenis klaim	penyebab klaim	jumlah klaim yang dibayar	durasi klaim	metode klaim	status klaim
0	Putra Nugroho	Wanita	33	Alam Barajo	2024-04-15	kematian	Kecelakaan	10009502	15 hari	cashless	SUCCES
1	Hadi Santoso	Wanita	31	Alam Barajo	2025-12-18	operasi	Konsultasi mata	8240690	15 hari	cashless	SUCCES
2	Nina Yuliana	Pria	57	Jambi Selatan	2025-01-30	rawat inap	Penyakit Dalam	18615868	2 hari	cashless	SUCCES
3	Rina Putri	Pria	61	Alam Barajo	2025-04-26	rawat jalan	Konsultasi mata	11250921	2 hari	cashless	SUCCES
4	Nina Seliawan	Wanita	47	Kotabaru	2023-12-22	rawat jalan	Kecelakaan	15477095	2 hari	cashless	SUCCES
...	...	...	...	...	...	...	...	...	...	...	...
995	Citra Siregar	Pria	60	Kotabaru	2025-03-06	rawat jalan	Penyakit Dalam	27027016	7 hari	cashless	IN PROCESS
996	Mega Seliawan	Wanita	62	Kotabaru	2023-05-10	operasi	Penyakit Dalam	12010626	7 hari	reimbursement	IN PROCESS
997	Oka Putri	Pria	29	Pasar Jambi	2024-03-29	rawat jalan	Konsultasi mata	19410998	7 hari	reimbursement	IN PROCESS
998	Eka Saputra	Pria	41	Jambi Selatan	2024-06-25	rawat jalan	Kelahiran	9303507	10 hari	cashless	IN PROCESS
999	Sari Lestari	Pria	60	Danau Teluk	2023-08-01	kematian	Penyakit Dalam	28736905	10 hari	cashless	IN PROCESS

1000 rows × 11 columns

Gambar 3.7 Praproses : libraries yang dibutuhkan

Dalam tahap pra-pemrosesan data pada proyek klasifikasi keterlambatan klaim asuransi, beberapa library penting digunakan untuk menunjang proses manipulasi dan transformasi data. Pertama, library pandas digunakan untuk membaca dan mengelola data dalam bentuk tabel (dataframe). Dengan menggunakan fungsi read_excel(), data klaim nasabah dari file Excel yang bernama "Data Klaim Nasabah MSIG.xlsx" berhasil dimuat dan ditampilkan untuk dianalisis lebih lanjut. Library ini memudahkan dalam melakukan filter, seleksi kolom, hingga transformasi nilai pada dataset. Selain itu, library numpy juga diimpor sebagai pendukung dalam berbagai operasi numerik dan komputasi array, khususnya dalam hal konversi tipe data dan pengolahan data numerik yang mungkin diperlukan pada tahapan lanjutan, seperti normalisasi atau encoding. Selanjutnya, digunakan pula modul LabelEncoder dari sklearn.preprocessing, yang berfungsi untuk mengubah data kategorikal (seperti jenis klaim, metode klaim, status klaim, dan penyebab klaim) menjadi format numerik. Proses ini sangat

penting karena sebagian besar algoritma machine learning hanya dapat memproses input dalam bentuk angka, bukan teks. Dengan ketiga library utama ini, proses pra-pemrosesan data menjadi lebih efisien dan siap digunakan untuk pelatihan model klasifikasi yang akan dilakukan pada tahap berikutnya.

3.2.10.2 Praproses : Label encoding

```
In [109]: #c. Ambil angka dari durasi klaim
df['durasi klaim'] = df['durasi klaim'].astype(str).str.extract('(\\d+)').astype(float)

In [110]: # d. Buat kolom target 'keterlambatan': Cepat (<= 7 hari), Lambat (> 7 hari)
df['keterlambatan'] = df['durasi klaim'].apply(lambda x: 'Cepat' if x <= 7 else 'Lambat')

In [111]: # e. Label encoding untuk fitur kategorikal
label_cols = ['jenis kelamin', 'daerah', 'jenis klaim', 'penyebab klaim', 'metode klaim']
le = LabelEncoder()
for col in label_cols:
    if df[col].dtype == 'object':
        df[col] = le.fit_transform(df[col].astype(str))

In [112]: # f. Encode label target
df['keterlambatan'] = df['keterlambatan'].map({'Cepat': 0, 'Lambat': 1})

In [113]: # g. Hapus kolom yang tidak dibutuhkan
df_baru = df.drop(columns=['nama', 'daerah']) # hapus kolom ini jika tidak diperlukan

In [114]: # 4. Tampilkan hasil akhir yang siap untuk modeling
df_baru.head()

Out[114]:
```

	jenis kelamin	usia	tanggal pengajuan klaim	jenis klaim	penyebab klaim	jumlah klaim yang dibayar	durasi klaim	metode klaim	status klaim	keterlambatan
0	1	33	2024-04-15	0	0	10009502	15.0	0	SUCCE	1
1	1	31	2025-12-18	1	2	8240690	15.0	0	SUCCE	1
2	0	57	2025-01-30	2	3	18615868	2.0	0	SUCCE	0
3	0	61	2025-04-26	3	2	11250921	2.0	0	SUCCE	0
4	1	47	2023-12-22	3	0	15477095	2.0	0	SUCCE	0

Gambar 3.8 Label Encoding

Tahapan berikutnya adalah melakukan data dimulai dengan mengubah kolom durasi klaim, yang awalnya berbentuk string seperti "15 hari", menjadi nilai numerik. Proses ini dilakukan dengan mengekstrak angka menggunakan ekspresi reguler, lalu mengonversinya ke tipe data float agar dapat digunakan dalam analisis dan klasifikasi. Setelah itu, dibuat kolom baru bernama keterlambatan untuk mengkategorikan durasi klaim menjadi dua kelas, yaitu Cepat jika klaim diselesaikan dalam waktu kurang dari atau sama dengan 7 hari, dan Lambat jika lebih dari 7 hari. Kolom ini kemudian menjadi label atau target yang akan diprediksi dalam model klasifikasi.

Selanjutnya, dilakukan proses label encoding untuk mengubah data kategorikal menjadi numerik agar dapat diproses oleh algoritma machine learning. Kolom-kolom seperti jenis kelamin, daerah, jenis klaim, penyebab klaim, dan metode klaim dikodekan menjadi angka menggunakan LabelEncoder dari pustaka Scikit-learn. Kemudian, kolom target keterlambatan juga diubah dari teks menjadi angka, yaitu 0 untuk Cepat dan 1 untuk Lambat, agar bisa digunakan dalam model klasifikasi biner.

Setelah proses encoding selesai, kolom-kolom yang dianggap tidak relevan atau bersifat identitas unik, seperti nama dan daerah, dihapus dari dataset agar tidak memengaruhi proses pelatihan model secara negatif. Dataset akhir yang telah dibersihkan dan dikodekan ini kemudian siap digunakan dalam proses pelatihan dan pengujian model machine learning untuk memprediksi keterlambatan klaim berdasarkan atribut-atribut lainnya. Data yang sudah siap kemudian digunakan untuk pelatihan model dengan algoritma Random Forest Classifier dan decision tree, yang dikenal efektif dalam menangani klasifikasi dan memberikan interpretasi penting terhadap fitur yang paling berpengaruh.

3.2.10.3 Penggunaan Decision tree

Pemilihan algoritma Decision Tree sebagai model awal dalam proyek klasifikasi keterlambatan klaim asuransi dilakukan karena kemampuannya yang unggul dalam menangani berbagai jenis data serta memberikan interpretasi yang mudah. Decision Tree dapat memproses baik fitur numerik maupun kategorikal secara langsung tanpa memerlukan transformasi data yang kompleks, sehingga sangat sesuai dengan karakteristik dataset klaim asuransi yang terdiri dari kolom seperti usia, jenis klaim, metode klaim, hingga durasi proses. Selain itu, algoritma ini juga tidak memerlukan asumsi statistik tertentu seperti linearitas atau normalitas data, yang menjadikannya fleksibel untuk digunakan dalam kondisi data operasional yang cenderung beragam. Salah satu keunggulan utama Decision Tree adalah kemampuannya untuk menyajikan hasil dalam bentuk visual pohon keputusan yang sangat intuitif, memudahkan tim non-teknis seperti manajemen dalam memahami proses klasifikasi. Proses pelatihannya pun cepat, sehingga cocok digunakan sebagai langkah awal eksplorasi model machine learning sebelum beralih ke algoritma yang lebih kompleks. Dengan pertimbangan tersebut, Decision Tree menjadi pilihan strategis untuk mengidentifikasi faktor-faktor yang memengaruhi keterlambatan klaim secara efektif.

proses pelatihan dan evaluasi model Decision Tree Classifier dalam memprediksi keterlambatan klaim asuransi berdasarkan data yang telah diproses sebelumnya. Pertama-tama, data dibagi menjadi dua bagian menggunakan fungsi `train_test_split`, yaitu 70% untuk data pelatihan dan 30% untuk data pengujian, dengan agar hasil pembagian tetap konsisten dan dapat direproduksi.

```
In [123]: X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
dt_model.fit(X_train, y_train)

Out[123]: DecisionTreeClassifier(random_state=42)
In a Jupyter environment, please rerun this cell to show the HTML representation or trust the notebook.
On GitHub, the HTML representation is unable to render, please try loading this page with nbviewer.org.

In [124]: y_pred_dt = dt_model.predict(X_test)

In [125]: from sklearn.metrics import classification_report, confusion_matrix, accuracy_score
print("Akurasi:", accuracy_score(y_test, y_pred_dt))
print("Confusion Matrix:\n", confusion_matrix(y_test, y_pred_dt))
print("Classification Report:\n", classification_report(y_test, y_pred_dt))

Akurasi: 1.0
Confusion Matrix:
[[185  0]
 [ 0 115]]
Classification Report:
              precision    recall  f1-score   support

     0       1.00      1.00      1.00        185
     1       1.00      1.00      1.00        115

 accuracy          1.00      1.00      1.00        300
 macro avg          1.00      1.00      1.00        300
 weighted avg          1.00      1.00      1.00        300

In [126]: import seaborn as sns
```

Gambar 3.9 Hasil Desicion tree

Hasil evaluasi model klasifikasi menggunakan algoritma Decision Tree menunjukkan performa yang sangat tinggi dan optimal. Berdasarkan metrik evaluasi yang ditampilkan, model ini mampu mencapai akurasi sebesar 100%, yang berarti seluruh data dalam subset pengujian berhasil diprediksi dengan benar tanpa adanya kesalahan klasifikasi. Hal ini juga diperkuat oleh confusion matrix, di mana jumlah prediksi benar untuk masing-masing kelas (yaitu “Cepat” dan “Lambat”) adalah 185 dan 115, serta tidak terdapat nilai false positive maupun false negative (semuanya nol). Selain itu, hasil dari classification report menunjukkan bahwa setiap kelas memiliki nilai precision, recall, dan f1-score sebesar 1.00, menandakan tidak hanya ketepatan prediksi yang sempurna, tetapi juga cakupan dan keseimbangan yang sangat baik antara kelas minoritas dan mayoritas. Dengan nilai macro average dan weighted average yang juga mencapai 1.00, model terbukti mampu mempertahankan kinerja yang konsisten di seluruh kategori data. Hasil ini menunjukkan bahwa model Decision Tree sangat efektif dalam mengidentifikasi klaim yang mengalami keterlambatan dan yang diproses secara cepat, serta berpotensi besar untuk digunakan dalam mendukung pengambilan keputusan di perusahaan asuransi secara otomatis dan akurat. Namun, hasil akurasi yang terlalu sempurna ini juga perlu ditinjau ulang untuk memastikan bahwa model tidak mengalami overfitting terhadap data latih.

3.2.10.4 Penggunaan Random Forest

Algoritma Random Forest dipilih dalam proyek klasifikasi keterlambatan klaim asuransi karena mampu memberikan hasil prediksi yang lebih stabil, akurat, dan andal dibandingkan model pohon keputusan tunggal. Random Forest bekerja dengan membangun sejumlah pohon keputusan secara acak dari subset data dan kemudian menggabungkan hasilnya melalui proses voting untuk menghasilkan keputusan akhir. Pendekatan ini membuat model lebih tahan terhadap overfitting

dan lebih mampu melakukan generalisasi pada data baru yang belum pernah dilihat sebelumnya. Hal ini sangat penting dalam konteks industri asuransi, di mana pola data sangat bervariasi dan seringkali tidak beraturan. Selain itu, Random Forest juga dapat menangani baik data numerik maupun kategorikal tanpa transformasi kompleks, menjadikannya sangat sesuai untuk dataset klaim yang berisi informasi seperti usia, jenis klaim, metode klaim, dan durasi klaim. Keunggulan lain dari algoritma ini adalah kemampuannya dalam menampilkan tingkat kepentingan dari masing-masing fitur, sehingga perusahaan dapat mengetahui variabel mana yang paling memengaruhi keterlambatan klaim. Dengan demikian, Random Forest tidak hanya berfungsi sebagai alat prediksi, tetapi juga sebagai alat bantu strategis untuk pengambilan keputusan berbasis data di perusahaan asuransi seperti MSIG.

Random forest

```
In [127]: from sklearn.ensemble import RandomForestClassifier
rf_model = RandomForestClassifier(random_state=42)
rf_model.fit(X_train, y_train)
y_pred_rf = rf_model.predict(X_test)

print("Akurasi RF:", accuracy_score(y_test, y_pred_rf))
print(confusion_matrix(y_test, y_pred_rf))
print(classification_report(y_test, y_pred_rf))
```

Akurasi RF: 1.0
[[185 0]
[0 115]]

	precision	recall	f1-score	support
0	1.00	1.00	1.00	185
1	1.00	1.00	1.00	115
accuracy			1.00	300
macro avg	1.00	1.00	1.00	300
weighted avg	1.00	1.00	1.00	300

Gambar 3.10 Hasil Random forest

hasil evaluasi model Random Forest Classifier pada klasifikasi keterlambatan klaim asuransi, diperoleh tingkat akurasi yang sangat tinggi, yaitu 100%. Hal ini ditunjukkan oleh skor akurasi sebesar 1.00, serta hasil confusion matrix yang menunjukkan bahwa seluruh data uji (sebanyak 300 observasi) berhasil diklasifikasikan dengan benar, tanpa adanya kesalahan prediksi (false positive maupun false negative). Metrik evaluasi lainnya seperti precision, recall, dan f1-score untuk kedua kelas (0 untuk "cepat" dan 1 untuk "lambat") juga mencapai nilai sempurna, yaitu 1.00. Artinya, model mampu mengidentifikasi seluruh klaim tepat waktu maupun klaim lambat dengan presisi dan sensitivitas maksimal. Kinerja ini mengindikasikan bahwa Random Forest bekerja sangat efektif dalam mengenali pola-pola pada data klaim asuransi yang memengaruhi keterlambatan, serta menunjukkan kemampuan generalisasi yang sangat baik pada data uji. Meskipun hasilnya ideal, evaluasi lanjutan seperti validasi silang dan pengujian pada data real-world tambahan tetap disarankan untuk memastikan konsistensi performa model di luar data pelatihan.

3.3 Kendala yang Ditemukan

Dalam pelaksanaan program magang yang berfokus pada pengembangan klasifikasi keterlambatan klaim asuransi menggunakan metode machine learning dan visualisasi data melalui dashboard interaktif, ditemukan beberapa kendala teknis maupun non-teknis yang menjadi pembelajaran penting dalam proses implementasi teknologi data analytics di sektor asuransi. Kendala-kendala ini berkaitan dengan struktur data, model analitik, serta integrasi fitur visualisasi.

1) Variasi format data pada kolom tanggal dan durasi klaim

Selama proses pembersihan data (data cleaning), ditemukan inkonsistensi dalam format penulisan tanggal serta nilai durasi klaim. Beberapa entri menggunakan format teks seperti "15 hari", sehingga perlu dilakukan ekstraksi angka secara eksplisit. Hal ini menyebabkan penambahan tahapan preprocessing tambahan untuk mengonversi data menjadi format numerik yang bisa digunakan dalam proses modeling.

2) Kebutuhan label baru untuk target klasifikasi keterlambatan

Data asli tidak secara eksplisit menyertakan label “lambat” atau “cepat” pada klaim. Oleh karena itu, perlu dibuat fitur baru (keterlambatan) berdasarkan logika aturan tertentu, misalnya klaim dengan durasi lebih dari 7 hari dikategorikan sebagai “lambat”. Proses ini membutuhkan validasi manual dan eksplorasi untuk memastikan bahwa threshold yang digunakan sesuai dengan kebijakan klaim aktual perusahaan.

3) Ketidakseimbangan fitur kategorikal yang terlalu dominan di satu kelas

Beberapa fitur kategorikal seperti metode klaim dan status klaim didominasi oleh satu nilai (misalnya mayoritas klaim bersifat *cashless* dan berstatus *success*), sehingga dikhawatirkan dapat menyebabkan bias pada model klasifikasi. Penyesuaian perlu

dilakukan dengan mempertimbangkan kembali bobot atau pemilihan fitur yang benar-benar relevan terhadap target prediksi.

4) Overfitting pada model awal

Saat menggunakan algoritma Decision Tree sebagai baseline model, hasil akurasi yang sangat tinggi (100%) menunjukkan kemungkinan besar terjadinya overfitting pada data pelatihan. Meskipun hasil prediksi sempurna, validasi silang atau uji coba dengan data baru sangat penting dilakukan untuk menghindari bias evaluasi.

5) Keterbatasan waktu untuk eksplorasi model yang lebih kompleks

Meskipun telah diterapkan model Random Forest yang memberikan hasil evaluasi sangat baik, keterbatasan waktu selama magang membatasi eksplorasi terhadap model lanjutan seperti XGBoost atau hyperparameter tuning yang mungkin dapat meningkatkan performa lebih jauh. Selain itu, dashboard yang menampilkan hasil klasifikasi juga belum sepenuhnya terintegrasi dengan pipeline model secara otomatis.

3.4 Solusi atas Kendala yang Ditemukan

1) Normalisasi Format Tanggal dan Durasi

Kendala variasi format data pada kolom tanggal dan durasi klaim diatasi dengan menerapkan fungsi parsing dan ekstraksi angka menggunakan metode regular expression. Seluruh entri durasi dikonversi menjadi tipe numerik (float) agar dapat dianalisis lebih lanjut, serta format tanggal distandarkan ke dalam bentuk datetime agar kompatibel dengan analisis temporal atau visualisasi tren.

2) Pembuatan Label Target Keterlambatan

Untuk menangani tidak tersedianya label eksplisit mengenai keterlambatan, dibentuklah kolom baru bernama keterlambatan yang didefinisikan berdasarkan durasi klaim. Klaim dengan durasi lebih dari

7 hari diklasifikasikan sebagai “Lambat”, sementara lainnya sebagai “Cepat”. Aturan ini disesuaikan dengan kebijakan waktu pemrosesan standar perusahaan asuransi.

3) Encoding Fitur Kategorikal untuk Menghindari Bias

Seluruh fitur kategorikal seperti jenis kelamin, daerah, jenis klaim, metode klaim, dan penyebab klaim dilakukan encoding menggunakan LabelEncoder untuk memastikan dapat diproses oleh model machine learning. Selain itu, dilakukan analisis eksplorasi awal guna mengetahui distribusi nilai tiap kategori sehingga bisa diantisipasi jika terjadi dominasi satu kelas.

4) Pemilihan Model Random Forest untuk Mencegah Overfitting

Model Decision Tree awal menunjukkan kecenderungan overfitting, sehingga solusi yang diambil adalah mengganti dengan algoritma Random Forest. Random Forest memiliki kemampuan generalisasi lebih baik karena membangun banyak pohon keputusan (ensemble) dan mengurangi varians dari hasil prediksi. Model ini terbukti memberikan hasil evaluasi yang lebih stabil dan akurat.

5) Optimalisasi Visualisasi dan Automasi Dashboard

Untuk mengatasi keterbatasan visualisasi, dilakukan pengembangan dashboard dengan filter dinamis dan segmentasi data yang lebih fleksibel. Visualisasi berbasis Power BI disusun agar mampu membaca pembaruan data terbaru secara otomatis, meskipun penyempurnaan lebih lanjut disarankan jika terdapat integrasi langsung dengan sistem internal klaim.