

LAPORAN

PREMIUM OPTIMIZATION
PT AXA INSURANCE INDONESIA



OLEH:

Kelompok 11 (SD-A2)

RIZAL HIDAYATULLAH	164221031
FAUZIAH HAMIDAH AL HANIEF	164221055
ARKAN SYAFIQ ATTAQY	164221062
ULVI AZZAHRA	164221109
AHMAD ZULFIKAR AL GHIFFARI	164221116

PROGRAM STUDI S1
TEKNOLOGI SAINS DATA
DEPARTEMEN TEKNIK
FAKULTAS TEKNOLOGI MAJU DAN MULTIDISIPLIN
UNIVERSITAS AIRLANGGA
SURABAYA
2025

DAFTAR ISI

DAFTAR ISI	ii
DAFTAR TABEL	iv
DAFTAR GAMBAR	v
DAFTAR LAMPIRAN	vi
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	2
1.3 Tujuan.....	2
1.4 Manfaat.....	3
1.4.1 Manfaat Bagi PT AXA Insurance Indonesia.....	3
1.4.2 Manfaat Bagi Pelanggan PT AXA Insurance Indonesia	3
1.4.3. Manfaat Bagi Mahasiswa	3
1.5 Batasan Masalah	3
BAB II TINJAUAN PUSTAKA	4
2.1 PT AXA Insurance Indonesia.....	4
2.2 Optimasi Premi dalam Asuransi	4
2.3 Data Profil Pelanggan dan Aspek Risiko dalam Asuransi	5
2.4 Exploratory Data Analysis (EDA).....	5
2.5 Data Preprocessing	5
2.6 Analisis Regresi.....	6
2.7 Model Machine Learning	6
2.8 Model Deep Learning.....	7
2.9 Evaluasi Model	7
BAB III METODE PELAKSANAAN	9
3.1 Alur Kerja.....	9
3.2 Sumber Data	9
3.3 Exploratory Data Analysis (EDA).....	11
3.3.1 Histogram	12
3.3.2 Boxplot	12
3.3.3 Heatmap Correlation.....	12
3.3.4 Bar Plot.....	12
3.4 Data Preprocessing	12
3.4.1 Handling Missing Value	12
3.4.2 Feature Engineering.....	12

3.4.3	Encoding Data	13
3.5	Model Training	13
3.6	Model Evaluation	13
BAB IV HASIL DAN PEMBAHASAN		14
4.1	Exploratory Data Analysis (EDA).....	14
4.1.1	Histogram	14
4.1.2	Box Plot	15
4.1.3	Heatmap Correlation.....	16
4.1.4	Bar Plot.....	16
4.2	Data Preprocessing	17
4.2.1	Handling Missing Value.....	17
4.2.2	Feature Engineering.....	17
4.2.3	Encoding Data	18
4.2.4	Feature Selection	19
4.3	Model Training	20
4.4	Model Evaluasi	21
BAB V PENUTUP		22
5.1	Kesimpulan.....	22
5.2	Saran	23
KONTRAK		24
DAFTAR PUSTAKA		28
LAMPIRAN		30

DAFTAR TABEL

Tabel 2.1. Variabel penelitian	10
Tabel 4.1. Variable missing value	17
Tabel 4.2. Variabel feature engineering	18
Tabel 4.3. Variabel encoding data	19
Tabel 4.4. Feature selection	20
Tabel 4.5. Model training	20
Tabel 4.6. Metrik evaluasi	21

DAFTAR GAMBAR

Gambar 3.1. Alur Kerja.....	9
Gambar 4.1. Histogram ‘Customer Service’	14
Gambar 4.2. Histogram ‘Age’	14
Gambar 4.3. Histogram ‘Claims History’	14
Gambar 4.4. Histogram ‘Coverage Amount’	14
Gambar 4.5. Histogram ‘Credit Score’	14
Gambar 4.6. Histogram ‘Deductible’	14
Gambar 4.7. Histogram ‘Premium Amount’	15
Gambar 4.8. Boxplot ‘Customer Service’	15
Gambar 4.9. Boxplot ‘Claims History’	15
Gambar 4.10. Boxplot ‘Age’	15
Gambar 4.11. Boxplot ‘Coverage Amount’	15
Gambar 4.12. Boxplot ‘Credit Score’	15
Gambar 4.13. Boxplot ‘Deductible’	15
Gambar 4.14. Heatmap correlation	16
Gambar 4.15. Barplot ‘Gender’	16
Gambar 4.16. Barplot ‘Geographic Information’	16
Gambar 4.17. Barplot ‘Products Owned’	17
Gambar 4.18. Barplot ‘Life Events’	17
Gambar 4.19. Barplot ‘Occupation’	17
Gambar 4.20. Barplot ‘Customer Preferences’	17

DAFTAR LAMPIRAN

Lampiran 1 - Link Drive	30
Lampiran 2 - Link GitHub	30
Lampiran 3 - Link Dashboard	30

BAB I PENDAHULUAN

1.1 Latar Belakang

Industri asuransi memiliki peran strategis dalam perekonomian global dengan menyediakan perlindungan terhadap berbagai risiko yang dihadapi oleh individu maupun entitas bisnis. Dalam menghadapi dinamika pasar yang semakin kompetitif, yang dipacu oleh digitalisasi cepat dan perubahan ekspektasi pelanggan, perusahaan asuransi harus terus berinovasi dalam pengembangan produk yang relevan sekaligus menjaga efisiensi operasional. Penetapan premi yang optimal merupakan aspek kunci untuk memastikan keberlanjutan, profitabilitas, dan daya saing perusahaan. Optimasi premi berarti menentukan harga premi yang mencerminkan risiko pelanggan secara akurat berdasarkan analisis data historis dan profil risiko, sehingga perusahaan dapat meningkatkan keuntungan sekaligus menjaga kepuasan dan loyalitas pelanggan.

Penetapan premi yang tidak sesuai dengan tingkat risiko aktual dapat memicu dua masalah utama, yaitu kerugian akibat klaim yang tinggi dan kehilangan pangsa pasar akibat premi yang tidak kompetitif. *Moral hazard*, sebagaimana dijelaskan (Akerlof, 2020), terjadi ketika nasabah cenderung mengajukan klaim berlebihan karena merasa terlindungi oleh polis asuransi. Sementara itu, *adverse selection* muncul ketika nasabah dengan risiko tinggi lebih berminat membeli asuransi, sedangkan nasabah dengan risiko rendah enggan melanjutkan pembelian polis. Kedua fenomena ini dapat merugikan perusahaan jika premi tidak dirancang berdasarkan profil risiko yang tepat, sehingga memperburuk posisi perusahaan di pasar yang kompetitif.

Untuk mengatasi tantangan tersebut, penerapan teknologi modern seperti *machine learning* dan *big data* menjadi sangat relevan. Teknologi ini memungkinkan perusahaan asuransi untuk menganalisis data pelanggan secara mendalam, mengidentifikasi pola risiko dengan lebih akurat, dan menetapkan premi yang sesuai dengan karakteristik individu. Sebagai contoh, PT AXA Insurance Indonesia telah memanfaatkan model prediktif berbasis data untuk meminimalkan kesalahan dalam penetapan premi, yang berkontribusi pada pengurangan risiko *moral hazard* dan *adverse selection*, sebagaimana diuraikan (Buhl et al., 2021). Dengan memanfaatkan data seperti usia, status pernikahan, pendapatan, dan riwayat klaim, perusahaan dapat membangun model yang lebih presisi dalam memperkirakan risiko, sehingga premi yang ditetapkan lebih adil dan kompetitif.

Sebagai pemain utama di industri asuransi jiwa di Indonesia, PT AXA Insurance Indonesia terus berkomitmen untuk meningkatkan akurasi penetapan premi dalam menghadapi dinamika pasar yang kompleks. Laporan tahunan 2024 menunjukkan bahwa penerapan model prediktif berbasis data telah mendukung pertumbuhan premi bruto sebesar Rp997 miliar, naik 10% dibandingkan tahun sebelumnya, serta peningkatan pendapatan *underwriting* sebesar 21% menjadi Rp191 miliar (PT AXA Insurance Indonesia, 2024). Implementasi teknologi ini membantu mengurangi kesalahan *underwriting* secara signifikan dan meningkatkan kepuasan pelanggan, dengan premi yang lebih sesuai dengan profil risiko dan kemampuan finansial nasabah. Selain itu, model ini memungkinkan penyesuaian premi secara dinamis untuk merespons perubahan kondisi eksternal, termasuk fluktuasi ekonomi, perubahan regulasi, dan dampak pandemi COVID-19, yang berpengaruh pada daya beli nasabah dan biaya klaim. Penyesuaian responsif tersebut, sebagaimana diuraikan dalam laporan (Löffler et al., 2020), membantu

perusahaan mempertahankan relevansi dan daya saing di pasar asuransi yang terus berkembang.

Penelitian ini bertujuan untuk menganalisis data historis dan profil pelanggan guna mengembangkan pendekatan optimasi premi yang lebih efektif bagi PT AXA Insurance Indonesia, khususnya dalam operasional di Jakarta. Tujuan utama adalah menghasilkan model penetapan premi yang lebih adil dan akurat, yang mencerminkan risiko nasabah secara tepat. Pendekatan ini diharapkan dapat meningkatkan efisiensi proses *underwriting*, mengurangi potensi *moral hazard* dan *adverse selection*, serta memperkuat daya saing produk asuransi di pasar. Dengan model optimasi premi yang lebih cerdas, PT AXA Insurance Indonesia dapat menawarkan harga yang kompetitif sambil menjaga keberlanjutan finansial, sehingga mampu bertahan dan berkembang di tengah persaingan pasar yang semakin kompleks. Hasil penelitian ini diharapkan memberikan kontribusi langsung bagi PT AXA Insurance Indonesia dengan memperkenalkan pendekatan berbasis data yang lebih tepat dalam penetapan premi. Selain mengurangi risiko kerugian akibat klaim berlebihan, pendekatan ini juga akan mendukung strategi perusahaan untuk mempertahankan daya tarik produk di mata nasabah, sekaligus memastikan stabilitas finansial dalam menghadapi tantangan pasar yang dinamis.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Faktor-faktor apa saja dari data profil pelanggan, detail polis, dan riwayat interaksi pada operasional PT AXA Insurance Indonesia di Jakarta yang paling signifikan mempengaruhi tingkat risiko dan menjadi dasar penentuan besaran premi asuransi?
2. Bagaimana rancangan model optimasi premi yang efektif dapat dikembangkan untuk operasional PT AXA Insurance Indonesia di Jakarta dengan memanfaatkan data historis dan profil pelanggan, guna menghasilkan penetapan premi yang lebih akurat, adil, dan selaras dengan risiko serta nilai polis?
3. Strategi implementasi apa yang dapat direkomendasikan kepada PT AXA Insurance Indonesia untuk operasionalnya di Jakarta berdasarkan temuan model optimasi premi, agar dapat meningkatkan akurasi penetapan harga dan daya saing produk?

1.3 Tujuan

Berdasarkan rumusan masalah yang telah diuraikan sebelumnya, maka tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Mengidentifikasi dan menganalisis faktor-faktor kunci dari data pelanggan dan polis dalam lingkup operasional PT AXA Insurance Indonesia di Jakarta yang paling berpengaruh terhadap risiko dan penentuan besaran premi.
2. Mengembangkan dan menguji sebuah model optimasi premi yang efektif untuk operasional PT AXA Insurance Indonesia di Jakarta, yang mampu memberikan rekomendasi penetapan premi yang lebih akurat, adil, dan sesuai dengan profil risiko nasabah serta nilai polis.
3. Merumuskan rekomendasi strategi bagi PT AXA Insurance Indonesia untuk mengimplementasikan model optimasi premi dalam proses bisnisnya pada

operasional di Jakarta, guna meningkatkan akurasi penetapan harga dan daya saing.

1.4 Manfaat

Penelitian ini diharapkan dapat memberikan manfaat bagi berbagai pihak, antara lain:

1.4.1 Manfaat Bagi PT AXA Insurance Indonesia

1. Meningkatkan akurasi dan efisiensi dalam penetapan premi asuransi dengan memanfaatkan model prediktif berbasis data, sehingga mengurangi risiko kerugian finansial.
2. Mendukung otomatisasi proses *underwriting* dan penyesuaian premi yang lebih cepat, yang memungkinkan tim fokus pada strategi bisnis dan pengembangan produk.
3. Memperkuat daya saing produk asuransi dengan menawarkan premi yang lebih adil dan kompetitif sesuai profil risiko nasabah di Jakarta.
4. Mengoptimalkan biaya operasional melalui pengurangan risiko *moral hazard* dan *adverse selection* dengan model premi yang lebih tepat sasaran.

1.4.2 Manfaat Bagi Pelanggan PT AXA Insurance Indonesia

1. Mendapatkan penetapan premi yang transparan dan sesuai dengan risiko masing-masing, meningkatkan keadilan dan kepuasan pelanggan.
2. Akses ke produk asuransi yang lebih sesuai dengan kebutuhan dan preferensi pribadi, sehingga meningkatkan loyalitas dan retensi pelanggan.
3. Meningkatkan kepercayaan pelanggan terhadap perusahaan karena penggunaan teknologi *data-driven* dalam penetapan harga.

1.4.3. Manfaat Bagi Mahasiswa

1. Mendapatkan pengalaman praktis dalam menerapkan metode data science dan machine learning pada permasalahan bisnis nyata di industri asuransi.
2. Mengasah keterampilan analisis data, pemodelan risiko, dan optimasi harga melalui studi kasus yang relevan dan aplikatif.
3. Menyelesaikan tantangan akademik dengan menghasilkan solusi yang berdampak nyata bagi perusahaan dan masyarakat.

1.5 Batasan Masalah

Untuk menjaga fokus dan kedalaman analisis, penelitian ini memiliki batasan-batasan sebagai berikut:

1. Data yang digunakan terbatas pada dataset sekunder PT AXA Insurance Indonesia untuk operasional di wilayah Jakarta, tanpa menggunakan data eksternal tambahan. Namun, variabel baru boleh dibuat melalui *feature engineering* dari data yang tersedia.
2. Penelitian ini fokus pada optimasi premi untuk produk asuransi yang tercakup dalam dataset, tidak mencakup produk lain yang mungkin dimiliki PT AXA Insurance Indonesia.
3. Model yang dikembangkan berbasis analisis kuantitatif dan data historis; tidak mencakup metode optimasi kompleks atau pengembangan model AI tingkat lanjut.
4. Faktor eksternal dinamis seperti regulasi, kondisi ekonomi makro, dan aspek teknis integrasi sistem tidak dibahas dalam penelitian ini.

BAB II TINJAUAN PUSTAKA

2.1 PT AXA Insurance Indonesia

PT AXA Insurance Indonesia, sebelumnya dikenal sebagai PT Mandiri AXA General Insurance (MAGI), merupakan bagian dari AXA Group, perusahaan asuransi dan manajemen aset terbesar di dunia yang berbasis di Paris. AXA Group beroperasi di 51 negara dan melayani lebih dari 93 juta nasabah dengan lebih dari 145.000 karyawan (PT AXA Insurance Indonesia, 2025).

Perusahaan ini didirikan pada tahun 1961 dengan nama awal PT Asuransi Dharma Bangsa. Pada tahun 2011, perusahaan ini berubah menjadi PT Mandiri AXA General Insurance melalui kemitraan strategis antara PT Bank Mandiri Tbk dan AXA Group. Pada tahun 2019, MAGI melakukan merger dengan PT Asuransi AXA Indonesia sebagai bagian dari kepatuhan terhadap regulasi Otoritas Jasa Keuangan (OJK). Kemudian, pada semester kedua tahun 2023, MAGI resmi berganti nama menjadi PT AXA Insurance Indonesia.

Sebagai perusahaan asuransi umum, PT AXA Insurance Indonesia menawarkan berbagai produk asuransi umum dan kesehatan untuk individu maupun korporasi melalui saluran distribusi *omni-channel* (PT AXA Insurance Indonesia, 2025). Perusahaan ini dikenal sebagai salah satu penyedia asuransi kesehatan karyawan, asuransi perjalanan, dan pengangkutan terkemuka di pasar Indonesia.

PT AXA Insurance Indonesia berkomitmen untuk memberikan layanan terbaik kepada nasabah dan terus berinovasi dalam menghadirkan produk serta layanan yang sesuai dengan kebutuhan masyarakat Indonesia (PT AXA Insurance Indonesia, 2024).

2.2 Optimasi Premi dalam Asuransi

Penetapan premi yang optimal merupakan hal penting dalam industri asuransi untuk memastikan keberlanjutan dan profitabilitas perusahaan. Dengan berkembangnya teknologi seperti *machine learning* dan *big data*, perusahaan asuransi kini dapat lebih akurat dalam mengidentifikasi dan menganalisis pola risiko nasabah. Ejjami (2024) mengungkapkan bahwa algoritma *machine learning*, seperti *random forest* dan *gradient boosting*, mampu menghasilkan prediksi yang lebih tepat terkait premi asuransi dibandingkan metode tradisional yang lebih sederhana, seperti *generalized linear models* (GLM). Teknologi ini juga memungkinkan perusahaan untuk menyesuaikan premi secara dinamis berdasarkan data *real-time*, yang sangat penting untuk merespon perubahan kondisi pasar dan perilaku pelanggan (Calvo, 2025). Penerapan teknologi tersebut tidak hanya meningkatkan akurasi dalam penilaian risiko tetapi juga mempercepat proses *underwriting* dan mengurangi kemungkinan terjadinya kesalahan dalam penetapan premi (Pike, Roggenbaum, & Garsson, 2023).

Namun, penerapan *machine learning* dan *big data* dalam industri asuransi juga memiliki tantangan. Menurut Pike et al. (2023), penggunaan algoritma dalam *underwriting* dapat menyebabkan diskriminasi jika data yang digunakan tidak dikelola dengan baik. Oleh karena itu, penting bagi perusahaan untuk menjaga kualitas dan keterandalan data agar model yang dihasilkan tetap adil dan tidak bias. Di Indonesia, perusahaan seperti PT AXA Insurance Indonesia telah mulai mengimplementasikan model prediktif berbasis data untuk mengoptimalkan penetapan premi, mengurangi risiko *moral hazard* dan *adverse selection*, serta meningkatkan daya saing produk mereka di pasar yang semakin kompetitif (PT

AXA Insurance Indonesia, 2024). Penggunaan model prediktif ini juga memungkinkan perusahaan untuk melakukan penyesuaian premi yang lebih fleksibel dalam menghadapi perubahan ekonomi dan regulasi, yang mendukung daya tahan perusahaan dalam situasi yang dinamis.

2.3 Data Profil Pelanggan dan Aspek Risiko dalam Asuransi

Pemanfaatan data pelanggan dalam industri asuransi sangat penting untuk penetapan premi yang akurat. Dengan menggunakan teknologi seperti *machine learning*, perusahaan asuransi dapat menganalisis data perilaku dan profil risiko nasabah untuk menetapkan premi yang lebih tepat. Ini juga mencakup analisis data dari media sosial dan perangkat IoT untuk meningkatkan pemahaman profil risiko pelanggan (NAIC, 2025). Namun, tantangan muncul terkait privasi dan diskriminasi, di mana penggunaan data seperti skor kredit dan lokasi dapat mempengaruhi keadilan dalam penetapan premi (Investopedia, 2022).

Penetapan premi juga dipengaruhi oleh aspek risiko, seperti *moral hazard* dan *adverse selection*. *Moral hazard* terjadi ketika nasabah merasa aman dan mengajukan klaim berlebihan, sementara *adverse selection* terjadi ketika nasabah dengan risiko tinggi lebih cenderung membeli asuransi. Penggunaan teknologi berbasis AI dapat membantu dalam mengurangi kesalahan penetapan premi dengan memperhitungkan perubahan kondisi pasar dan perilaku konsumen (Löffler et al., 2020).

Walaupun teknologi dapat meningkatkan akurasi dalam penetapan premi, terdapat tantangan dalam hal integrasi data dan perlindungan privasi. Perusahaan asuransi perlu mengembangkan model yang transparan dan sesuai regulasi untuk menghindari bias dan ketidakadilan dalam penggunaan data (Swiss Re, 2022).

2.4 Exploratory Data Analysis (EDA)

EDA merupakan langkah penting dalam setiap analisis penelitian. Tujuan utama analisis eksplorasi adalah untuk memeriksa data untuk distribusi, *outlier*, dan anomali untuk mengarahkan pengujian spesifik terhadap hipotesis (Komorowski, dkk., 2016). EDA digunakan untuk mengidentifikasi informasi penting dalam data yang kita miliki. Langkah ini diperlukan untuk menentukan tahapan analisis yang akan dilakukan dan memahami jenis informasi yang dapat diolah dari data tersebut. EDA membantu kita untuk mengungkap pola, anomali, dan hubungan dalam data, sehingga dapat mengarahkan proses analisis selanjutnya secara lebih efektif.

Ide sentral dalam korpus EDA adalah tujuan mengembangkan model mental yang rinci dan akurat yang memberikan rasa keintiman dengan nuansa data (Behrens, dkk., 2003). Ini menjadi penting karena model yang sesuai dengan data akan memberikan prediksi atau hasil yang lebih akurat dan relevan. Dengan demikian, kita dapat membuat keputusan yang lebih tepat berdasarkan informasi yang diperoleh dari model tersebut.

2.5 Data Preprocessing

Data preprocessing adalah proses mengubah data mentah menjadi format yang lebih berguna dan efektif untuk analisis. Tahap ini mencakup berbagai teknik seperti pembersihan data, transformasi data, normalisasi, dan pengurangan dimensi. Pembersihan data melibatkan penghapusan atau pengisian nilai yang hilang, mengoreksi kesalahan data, dan mengidentifikasi serta menghapus duplikasi. Transformasi data, di sisi lain, mungkin melibatkan konversi data dari satu format ke format lain atau menerapkan fungsi matematika untuk mengubah skala data. Normalisasi adalah proses untuk memastikan bahwa data berada dalam skala yang

konsisten, sering kali dengan merubah nilai data ke dalam rentang tertentu seperti 0 hingga 1. Pengurangan dimensi bertujuan untuk mengurangi jumlah fitur dalam dataset tanpa mengorbankan informasi penting, sering dilakukan melalui teknik seperti Principal Component Analysis (PCA). Pemrosesan awal data penting karena mengadaptasi data untuk memenuhi permintaan input algoritma penambangan data, mengurangi kompleksitas dan menghilangkan elemen yang tidak relevan dan mengganggu (García, dkk., 2014).

Tahapan ini sangat krusial karena kualitas data mentah sering kali tidak memadai untuk analisis yang tepat. Data yang belum diproses bisa mengandung *noise*, kesalahan, dan variasi yang bisa mengaburkan hasil analisis atau model *machine learning*. Oleh karena itu, *data preprocessing* tidak hanya meningkatkan kualitas data tetapi juga efisiensi dan akurasi analisis selanjutnya. Dengan data yang telah diproses dengan baik, algoritma *machine learning* dapat bekerja lebih efektif dan memberikan hasil yang lebih akurat dan andal. Proses ini adalah fondasi penting dalam *pipeline data science*, memungkinkan peneliti dan analis untuk membuat keputusan yang lebih baik dan mendapatkan wawasan yang lebih mendalam dari data mereka. Pemrosesan awal data mengurangi kompleksitas kumpulan data dunia nyata, memungkinkan proses pembelajaran lebih cepat dan tepat serta struktur data mentah yang lebih mudah dipahami (Gallego, dkk., 2017).

2.6 Analisis Regresi

Analisis regresi merupakan teknik yang sangat penting dalam optimasi premi di industri asuransi, khususnya untuk mengevaluasi hubungan antara variabel risiko dan penetapan harga premi. Dalam konteks asuransi, regresi digunakan untuk memodelkan berbagai faktor yang mempengaruhi tingkat risiko nasabah dan, pada gilirannya, harga premi yang harus dibayar. Sebagai contoh, model regresi linier dapat digunakan untuk memprediksi premi berdasarkan faktor-faktor seperti usia, pendapatan, riwayat kesehatan, dan jenis kelamin, yang semuanya berpengaruh terhadap kemungkinan nasabah mengajukan klaim.

Dalam beberapa tahun terakhir, model regresi distribusional mulai mendapatkan perhatian dalam industri asuransi, berkat kemampuannya untuk memodelkan seluruh distribusi variabel dependen, seperti jumlah klaim, berdasarkan faktor-faktor independen (Klein, 2024). Pendekatan ini memberikan gambaran yang lebih lengkap daripada hanya menggunakan rata-rata, sehingga lebih akurat dalam mengelola risiko di sektor asuransi. Selain itu, regresi logistik banyak digunakan untuk memprediksi probabilitas kejadian, seperti klaim, yang relevan dalam menentukan premi berdasarkan tingkat risiko yang teridentifikasi.

Selain regresi linier dan logistik, model regresi *robust* nonparametrik juga semakin diterapkan untuk mengatasi masalah heteroskedastisitas dan *outlier* dalam data asuransi. Data risiko dalam asuransi sering menunjukkan variabilitas tinggi dan tidak selalu mengikuti distribusi normal (Salibian-Barrera, 2022). Penggunaan model ini memungkinkan perusahaan asuransi seperti PT AXA Insurance Indonesia untuk menetapkan premi yang lebih akurat dan lebih tahan terhadap gangguan atau kesalahan data yang dapat mempengaruhi kinerja model tradisional.

2.7 Model Machine Learning

Model *machine learning* adalah algoritma atau sistem yang dirancang untuk mengenali pola dan membuat keputusan atau prediksi berdasarkan data. Model ini dikembangkan melalui proses pelatihan, di mana *raw* data yang telah diproses diberikan kepada algoritma untuk mempelajari hubungan antara variabel *input* dan

output. Sistem perangkat lunak yang belajar dari data semakin banyak digunakan dalam skenario aplikasi industri (Schelter, dkk., 2015). Terdapat berbagai jenis model *machine learning*, termasuk regresi, klasifikasi, *clustering*, dan jaringan saraf. Setiap jenis model ini digunakan untuk berbagai tugas tertentu, misalnya, regresi digunakan untuk memprediksi nilai numerik, klasifikasi untuk mengkategorikan data, dan *clustering* untuk mengelompokkan data berdasarkan kesamaan. Pemilihan model bergantung pada jenis data dan masalah yang ingin dipecahkan, serta ketersediaan sumber daya komputasi dan kebutuhan akurasi.

Setelah model *machine learning* dilatih, model tersebut dievaluasi menggunakan data yang belum pernah dilihat selama pelatihan untuk mengukur kinerjanya. Evaluasi ini penting untuk memastikan bahwa model tidak hanya menghafal data pelatihan tetapi juga dapat melakukan generalisasi dengan baik pada data baru. Berbagai metrik seperti akurasi, presisi, recall, dan F1-score digunakan untuk menilai performa model. Peningkatan model dapat dilakukan melalui *tuning hyperparameter* dan penggunaan metode *ensemble*. Banyak keputusan untuk merancang sistem yang mengelola model ML memerlukan pemahaman mendalam tentang algoritma ML dan konsekuensinya terhadap sistem terkait (Schelter, dkk., 2015).

2.8 Model Deep Learning

Model *deep learning* (DL) telah mengalami kemajuan signifikan dalam beberapa tahun terakhir, terutama dalam penerapan jaringan syaraf tiruan (neural networks) untuk berbagai aplikasi. Jaringan syaraf tiruan, yang terinspirasi oleh struktur dan fungsi otak manusia, terdiri dari lapisan-lapisan neuron buatan yang saling terhubung. Arsitektur seperti *Convolutional Neural Networks* (CNNs) dan *Recurrent Neural Networks* (RNNs) telah menjadi dasar dalam pengolahan citra dan data sekuensial. Namun, perkembangan terbaru menunjukkan munculnya arsitektur baru seperti *Transformer*, *Generative Adversarial Networks* (GANs), dan *Graph Neural Networks* (GNNs), yang menawarkan kemampuan lebih dalam menangani data kompleks dan tidak terstruktur (Mienye & Swart, 2024).

Selain itu, teknik pelatihan baru seperti *self-supervised learning*, *federated learning*, dan *deep reinforcement learning* telah diperkenalkan untuk meningkatkan efisiensi dan generalisasi model DL. *Self-supervised learning* memungkinkan model untuk belajar dari data tidak berlabel, sementara *federated learning* memungkinkan pelatihan model secara terdistribusi tanpa memindahkan data sensitif. *Deep reinforcement learning*, di sisi lain, menggabungkan pembelajaran berbasis penghargaan dengan teknik DL untuk pengambilan keputusan yang lebih baik dalam lingkungan dinamis (Mienye & Swart, 2024).

Namun, meskipun kemajuan ini, tantangan seperti *overfitting*, kebutuhan komputasi yang tinggi, dan interpretabilitas model masih menjadi perhatian utama dalam penelitian DL. Oleh karena itu, pengembangan model yang lebih efisien dan dapat dijelaskan menjadi fokus penting untuk memastikan penerapan DL yang efektif dan etis di berbagai bidang (Mienye & Swart, 2024).

2.9 Evaluasi Model

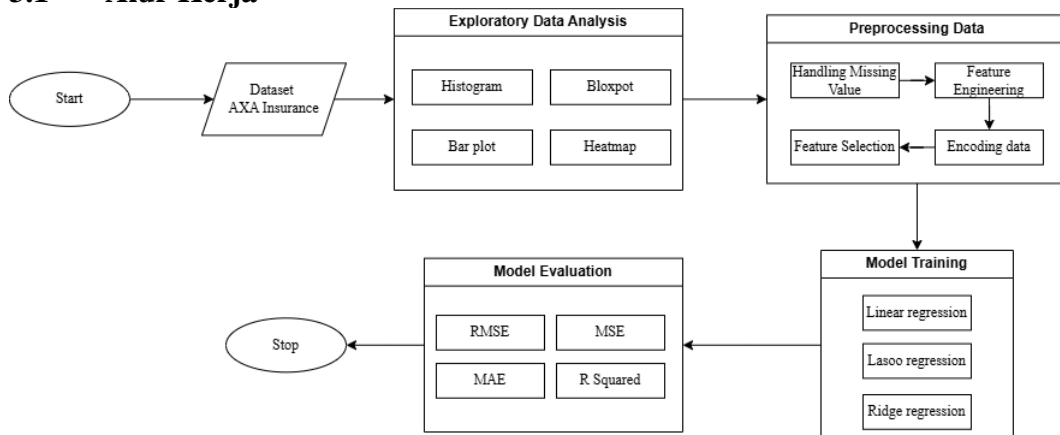
Evaluasi model merupakan tahap yang sangat penting dalam pengembangan model *machine learning* dan *deep learning*, terutama dalam tugas-tugas regresi. Beberapa metrik yang umum digunakan untuk mengevaluasi kinerja model adalah *Mean Squared Error* (MSE), *Root Mean Squared Error* (RMSE), *Mean Absolute Error* (MAE), dan *R-squared* (R^2). MSE mengukur rata-rata kuadrat selisih antara

nilai prediksi dan nilai aktual, memberikan penalti yang lebih besar pada kesalahan yang lebih besar, karena kesalahan dihitung dalam kuadrat. Hal ini membuat MSE sangat sensitif terhadap *outlier*, menjadikannya pilihan yang umum dalam pelatihan model deep learning, karena sifatnya yang mudah didiferensiasi, memudahkan dalam proses optimasi menggunakan algoritma seperti *gradient descent*. RMSE adalah akar kuadrat dari MSE, yang memberikan ukuran kesalahan dalam satuan yang sama dengan data asli, sehingga lebih mudah diinterpretasikan oleh pengguna. Metrik ini juga memberikan penalti yang lebih besar terhadap kesalahan besar dibandingkan MAE, membuatnya sensitif terhadap *outlier*.

Di sisi lain, MAE menghitung rata-rata nilai absolut selisih antara prediksi dan nilai aktual, yang tidak sensitif terhadap *outlier* dan lebih stabil dalam kasus kesalahan besar yang jarang terjadi. MAE menjadi pilihan yang baik ketika kesalahan besar tidak diinginkan dalam evaluasi model. *R-squared* (R^2), sebagai metrik tambahan, mengukur proporsi variasi dalam data yang dapat dijelaskan oleh model. Nilai R^2 berkisar antara 0 dan 1, di mana nilai yang lebih tinggi menunjukkan bahwa model mampu menjelaskan variasi data dengan lebih baik.

BAB III METODE PELAKSANAAN

3.1 Alur Kerja



Gambar 3.1. Alur Kerja

Berikut adalah alur kerja pelaksanaan penelitian yang dilakukan untuk memastikan bahwa hasil yang didapatkan akurat:

1. **Import data:** Data yang digunakan adalah data sekunder, oleh karena itu peneliti dapat langsung *import dataset* yang akan digunakan. Dalam hal ini, peneliti menggunakan data historis pelanggan asuransi PT AXA Insurance Indonesia.
2. **Exploratory Data Analysis (EDA):** Langkah selanjutnya adalah melakukan *Exploratory Data Analysis* (EDA) untuk memahami bentuk data yang kita miliki sebelum nantinya akan diolah. Langkah ini dilakukan dengan visualisasi data menggunakan *histogram*, *boxplot*, *heatmap*, dan *bar plot*.
3. **Preprocessing Data:** Setelah memahami data, selanjutnya adalah membersihkan data. Tahap ini mencakup proses mengisi atau menghapus data yang hilang, menambah dan memilih variabel yang digunakan, mengubah data yang belum sesuai. Hal ini penting dilakukan agar data menjadi bersih dan terstruktur sehingga proses pengolahannya akan lebih mudah.
4. **Model Training:** Lakukan pelatihan model pada data yang telah dibersihkan. Model yang digunakan adalah Linear regression, Lasso regression, dan Ridge Regression.
5. **Model Evaluation:** Setelah model dilatih, tahap selanjutnya adalah mengevaluasinya untuk melihat seberapa baik model dapat memprediksi data baru. Metrik evaluasi yang digunakan adalah MSE, RMSE, MAE, dan R-Squared.

3.2 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang didapat dari klien, yaitu PT AXA Insurance Indonesia. Data tersebut berupa kumpulan data pelanggan asuransi, meliputi berbagai aspek penting untuk analisis risiko dan optimasi premi. Data mencakup informasi demografis seperti usia, jenis kelamin, status pernikahan, pendidikan, pekerjaan, serta informasi geografis termasuk lokasi dan wilayah tempat tinggal pelanggan. Selain itu, data juga berisi riwayat klaim, perilaku pelanggan, riwayat pembelian produk asuransi, jenis produk asuransi yang dimiliki, serta rincian polis seperti tanggal mulai polis, tanggal perpanjangan,

jumlah premi, jumlah pertanggungan, dan *deductible*. Tabel 3.1 menunjukkan variabel penelitian yang digunakan.

Tabel 2.1. Variabel penelitian

Jenis variabel	Variabel	Tipe data	Keterangan
Variabel dependen	Premium Amount	Float	Besaran premi yang dibayarkan pelanggan
Variabel waktu	Policy Start Date	Date	Tanggal mulai berlakunya polis
	Policy Renewal Date	Date	Tanggal perpanjangan polis
Variabel independen	Age	Integer	Usia pelanggan
	Gender	String	Jenis kelamin pelanggan
	Marital Status	String	Status pernikahan pelanggan
	Occupation	String	Pekerjaan pelanggan
	Income level	String	Kategori atau rentang penghasilan pelanggan
	Education level	String	Tingkat pendidikan tertinggi pelanggan
	Geographic Information	String	Wilayah tempat tinggal pelanggan
	Location	String	Kota atau daerah spesifik tempat tinggal pelanggan
	Purchase history	String	Riwayat pembelian produk asuransi
	Claim history	String	Jumlah atau catatan klaim yang pernah diajukan pelanggan
	Interactions with Customer Service	Integer	Frekuensi atau catatan interaksi pelanggan dengan layanan pelanggan
	Insurance products owned	String	Daftar produk asuransi yang dimiliki pelanggan
	Coverage amount	Integer	Besaran nilai pertanggungan polis
	Deductible	Float	Besaran biaya yang harus dibayar pelanggan sebelum klaim asuransi berlaku
	Policy type	String	Jenis polis asuransi (Life, Auto, dll.)

	Customer preferences	String	Preferensi pelanggan terkait produk dan layanan
	Preferred Communication Channel	String	Cara komunikasi yang disukai pelanggan (Email, Phone, dll.)
	Preferred Contact Time	String	Waktu yang disukai pelanggan untuk dihubungi
	Preferred language	String	Bahasa komunikasi yang disukai pelanggan
	Risk profile	String	Penilaian tingkat risiko pelanggan berdasarkan berbagai faktor
	Previous claims history	Integer	Jumlah riwayat klaim sebelumnya
	Credit score	Integer	Skor kredit pelanggan
	Driving record	String	Catatan riwayat mengemudi pelanggan (jika ada)
	Life events	String	Peristiwa penting dalam hidup pelanggan yang dapat mempengaruhi kebutuhan asuransi
	Segmentation Group	String	Klasifikasi pelanggan berdasarkan segmentasi demografis atau perilaku

Lebih jauh, terdapat data terkait interaksi pelanggan dengan layanan pelanggan, preferensi komunikasi, profil risiko, nilai kredit, dan catatan pengemudi (bagi produk asuransi kendaraan). Informasi ini memberikan gambaran menyeluruh mengenai pelanggan, yang sangat berguna dalam melakukan analisis segmentasi dan optimasi premi.

Tujuan utama dari penggunaan data ini adalah untuk mengembangkan model optimasi premi yang dapat menentukan besaran premi yang optimal berdasarkan profil risiko dan nilai polis pelanggan. Dengan data historis yang lengkap dan variabel yang beragam, model diharapkan dapat membantu PT AXA Insurance Indonesia dalam mengatur premi secara lebih efektif dan sesuai dengan risiko yang dihadapi tiap pelanggan.

Secara teknis, data ini bersifat historis dan dikumpulkan secara terstruktur, memungkinkan untuk analisis prediktif dan pemodelan risiko. Hal ini sangat relevan untuk proyek premium optimization yang menuntut pemahaman mendalam terhadap karakteristik pelanggan dan pola klaim mereka.

3.3 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) merupakan tahap awal dalam proses penelitian yang bertujuan untuk memahami karakteristik dan pola yang terdapat dalam data. Pada tahap ini, berbagai teknik visualisasi dan statistik digunakan untuk mengidentifikasi distribusi data, *outlier*, korelasi antar variabel, serta ketidaksesuaian data. Dengan melakukan EDA, peneliti dapat memperoleh insight

penting yang membantu dalam menentukan langkah selanjutnya dalam pengolahan data dan pemilihan model yang sesuai.

3.3.1 Histogram

Histogram digunakan untuk melihat distribusi frekuensi dari variabel numerik dalam dataset. Melalui histogram, dapat diketahui apakah data memiliki distribusi normal, *skewed*, atau multimodal. Informasi ini sangat berguna untuk menentukan metode statistik yang tepat dan juga untuk mengidentifikasi adanya *outlier* atau nilai ekstrim yang mungkin perlu ditangani lebih lanjut.

3.3.2 Boxplot

Boxplot digunakan untuk mendeteksi *outlier* dan memahami penyebaran data dari setiap variabel. Dengan melihat *boxplot*, peneliti dapat dengan cepat mengidentifikasi median, kuartil, serta nilai minimum dan maksimum yang ada dalam data. Analisis ini membantu dalam menilai apakah terdapat data yang tidak wajar yang bisa mempengaruhi hasil pemodelan.

3.3.3 Heatmap Correlation

Heatmap Correlation digunakan untuk memvisualisasikan korelasi antar variabel numerik dalam dataset. Dengan melihat peta warna yang merepresentasikan tingkat korelasi, peneliti dapat memahami hubungan antar fitur, apakah terdapat fitur yang sangat berkorelasi positif atau negatif. Hal ini penting untuk menghindari multikolinearitas dan memilih fitur yang paling relevan untuk pemodelan.

3.3.4 Bar Plot

Bar plot merupakan metode visualisasi data yang digunakan untuk menampilkan perbandingan antar kategori melalui batang-batang yang tingginya merepresentasikan nilai dari masing-masing kategori. Dengan menggunakan bar plot, peneliti dapat dengan mudah mengidentifikasi kategori mana yang memiliki nilai tertinggi atau terendah serta melakukan perbandingan antar kategori secara visual. Visualisasi ini sangat efektif untuk menunjukkan distribusi data kategori, membandingkan frekuensi kejadian, dan menampilkan hasil ringkasan data seperti rata-rata atau jumlah dari suatu kategori tertentu. Penggunaan bar plot membantu dalam pengambilan keputusan berbasis data dengan cepat dan jelas.

3.4 Data Preprocessing

Dalam proses pengolahan data, data *preprocessing* menjadi tahapan yang sangat penting untuk memastikan data yang akan diolah adalah data yang akurat. Pada proses ini, langkah yang dilakukan adalah *handling missing value*, *feature engineering*, *feature selection*, *encoding data*, dan normalisasi data.

3.4.1 Handling Missing Value

Handling missing value dilakukan untuk mengatasi data yang hilang atau tidak lengkap dalam dataset. Teknik yang umum digunakan meliputi penghapusan baris atau kolom yang memiliki *missing value*, pengisian nilai yang hilang dengan mean, median, modus, atau metode imputasi yang lebih kompleks. Pemilihan teknik ini bergantung pada jenis data dan jumlah *missing value* yang ada, dengan tujuan meminimalkan bias dan kehilangan informasi penting.

3.4.2 Feature Engineering

Feature engineering adalah proses pembuatan atau transformasi fitur baru dari data yang sudah ada untuk meningkatkan performa model. Teknik ini dapat berupa ekstraksi fitur dari data waktu, penggabungan beberapa fitur menjadi satu,

atau transformasi matematis pada fitur. Dengan *feature engineering*, model dapat menangkap pola yang lebih kompleks dan meningkatkan kemampuan prediksi.

3.4.3 Encoding Data

Encoding data digunakan untuk mengubah variabel kategorikal menjadi format numerik yang dapat diproses oleh algoritma *machine learning*. Beberapa teknik encoding yang umum digunakan antara lain *one-hot encoding*, *label encoding*, dan *ordinal encoding*. Pemilihan metode *encoding* harus mempertimbangkan sifat dari data kategori dan algoritma yang digunakan agar tidak menimbulkan bias.

3.4.4 Feature Selection

Feature selection bertujuan untuk memilih *subset* fitur yang paling relevan dan memberikan kontribusi signifikan terhadap variabel target. Teknik ini membantu mengurangi kompleksitas model, mempercepat proses *training*, serta menghindari *overfitting*. Metode yang digunakan dapat berupa teknik statistik, algoritma berbasis model, atau teknik *wrapper* dan *embedded*.

3.5 Model Training

Model training adalah tahap di mana algoritma *machine learning* belajar dari data yang sudah diproses untuk menemukan pola dan hubungan yang dapat digunakan untuk prediksi. Pada tahap ini, data *training* digunakan untuk mengoptimalkan parameter model sehingga mampu meminimalkan error. Proses ini juga dapat melibatkan teknik validasi silang untuk memastikan model tidak *overfit* dan memiliki generalisasi yang baik.

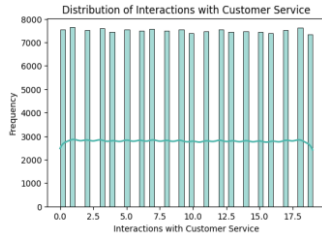
3.6 Model Evaluation

Model evaluation dilakukan untuk menilai performa model yang telah dilatih menggunakan data *testing* atau data validasi. Berbagai matrik evaluasi digunakan sesuai dengan jenis masalah, seperti RMSE, MSE, MAE, dan R Squared. Evaluasi ini penting untuk mengetahui apakah model sudah cukup baik digunakan dalam aplikasi nyata atau perlu dilakukan penyempurnaan lebih lanjut.

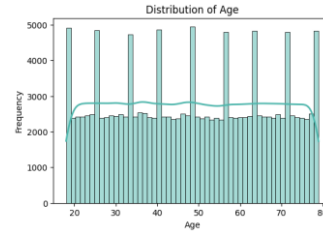
BAB IV HASIL DAN PEMBAHASAN

4.1 Exploratory Data Analysis (EDA)

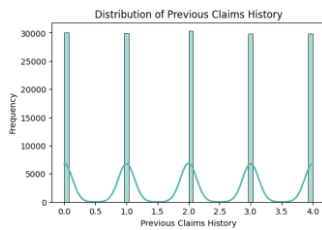
4.1.1 Histogram



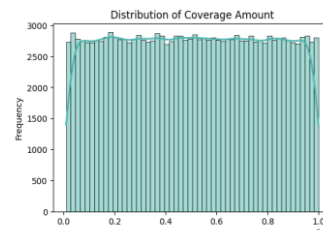
Gambar 4.1. Histogram 'Customer Service'



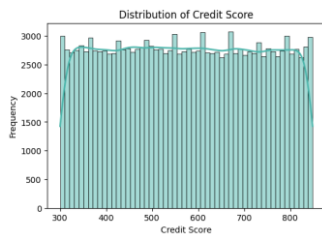
Gambar 4.2. Histogram 'Age'



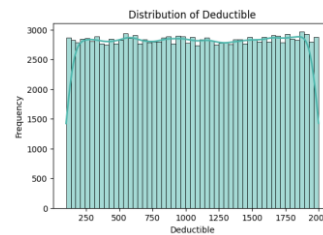
Gambar 4.3. Histogram 'Claims History'



Gambar 4.4. Histogram 'Coverage Amount'

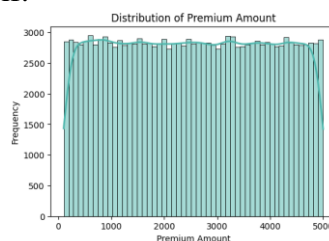


Gambar 4.5. Histogram 'Credit Score'



Gambar 4.6. Histogram 'Deductible'

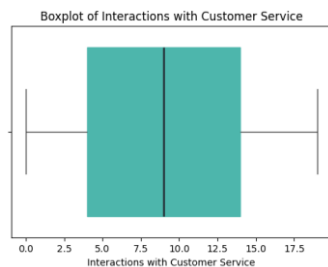
Distribusi umur pelanggan cukup merata dari muda hingga senior, menunjukkan bahwa dataset mencakup berbagai kelompok usia yang beragam. Nilai *Coverage Amount* tersebar luas dari yang kecil hingga sangat besar, mencerminkan variasi produk asuransi yang dimiliki oleh nasabah. Begitu juga dengan *Credit Score*, yang berkisar dari rendah hingga tinggi, mengindikasikan profil risiko yang berbeda-beda. Nilai *Deductible* juga beragam, memperlihatkan berbagai tingkat risiko yang dipilih pelanggan. Interaksi dengan layanan pelanggan cukup beragam, dari yang minim hingga sangat aktif, sedangkan riwayat klaim sebelumnya didominasi oleh pelanggan dengan sedikit klaim, tetapi ada juga yang memiliki beberapa klaim. Keseluruhan distribusi ini menunjukkan bahwa data kita memiliki variasi yang sehat dan representatif untuk melakukan analisis premi asuransi secara komprehensif.



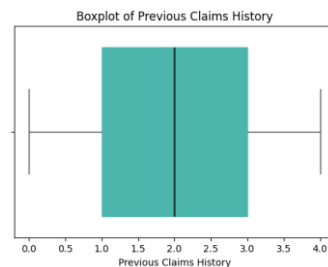
Gambar 4.7. Histogram 'Premium Amount'

Distribusi premi asuransi sebagai variabel target dalam data ini menunjukkan penyebaran nilai yang relatif merata di seluruh rentang premi, dari yang rendah hingga tinggi. Frekuensi yang konsisten di berbagai level premi mencerminkan adanya variasi profil risiko dan kebutuhan asuransi yang beragam di antara para nasabah. Meskipun terdapat sedikit penurunan frekuensi pada nilai premi yang sangat rendah dan sangat tinggi, distribusi ini memberikan gambaran yang baik untuk membangun model prediksi premi yang mampu menangkap beragam kondisi pasar dan karakteristik pelanggan.

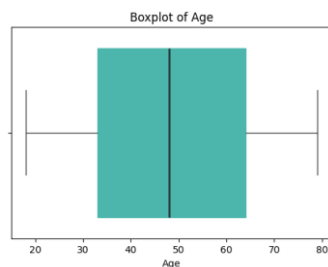
4.1.2 Box Plot



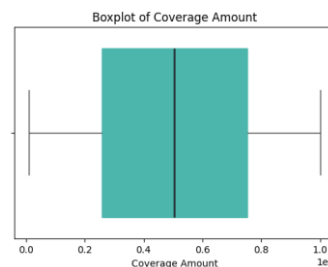
Gambar 4.8. Boxplot 'Customer Service'



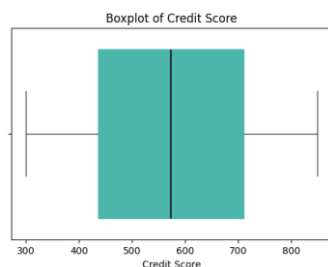
Gambar 4.9. Boxplot 'Claims History'



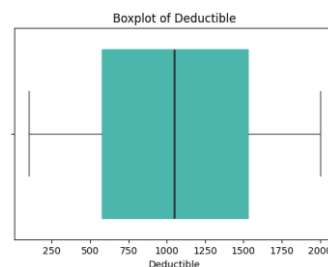
Gambar 4.10. Boxplot 'Age'



Gambar 4.11. Boxplot 'Coverage Amount'



Gambar 4.12. Boxplot 'Credit Score'

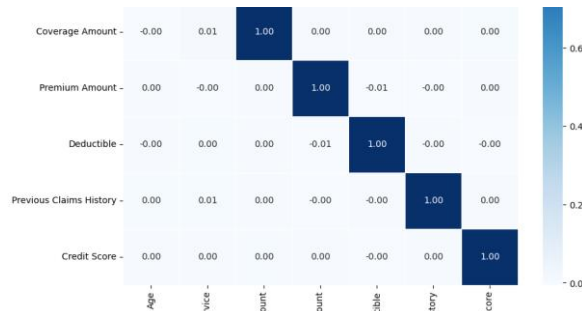


Gambar 4.13. Boxplot 'Deductible'

Dari visualisasi *boxplot* yang ditampilkan untuk variabel-variabel numerik utama seperti *Age*, *Coverage Amount*, *Credit Score*, *Deductible*, *Interactions with Customer Service*, dan *Previous Claims History*, terlihat bahwa distribusi setiap variabel cukup simetris tanpa adanya *outlier* ekstrem. Ini menandakan bahwa data relatif bersih dan sebarannya cukup merata di rentang nilai yang wajar. Misalnya, usia berada di kisaran 18 sampai 79 tahun dengan median sekitar 48 tahun, sedangkan *Coverage Amount* dan *Deductible* juga memiliki rentang nilai yang luas namun tidak ada nilai ekstrim yang mencolok. Kondisi ini memudahkan pemodelan

karena model tidak perlu terlalu khawatir terhadap efek *outlier* yang bisa mempengaruhi kestabilan prediksi. Secara umum, data numerik ini tampak sehat dan representatif untuk digunakan dalam pengembangan model prediksi premi asuransi.

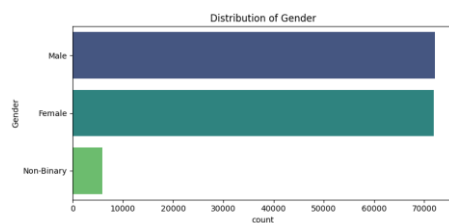
4.1.3 Heatmap Correlation



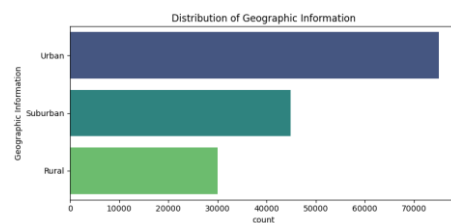
Gambar 4.14. Heatmap correlation

Dari matriks korelasi yang ditampilkan, dapat diamati bahwa variabel-variabel numerik utama dalam dataset memiliki korelasi yang sangat rendah satu sama lain, termasuk dengan variabel target, yaitu *Premium Amount*. Nilai korelasi antar fitur seperti *Age*, *Interactions with Customer Service*, *Coverage Amount*, *Deductible*, *Previous Claims History*, dan *Credit Score* cenderung mendekati nol, menunjukkan tidak adanya hubungan linier yang kuat antar fitur-fitur tersebut. Begitu pula dengan korelasi antara fitur-fitur ini terhadap *Premium Amount* yang juga sangat kecil, hampir nol. Kondisi ini menandakan bahwa variabel-variabel tersebut relatif independen dan tidak saling mempengaruhi secara langsung, serta bahwa hubungan linier sederhana mungkin kurang memadai untuk menjelaskan variabilitas premi. Oleh karena itu, pendekatan pemodelan yang menggabungkan interaksi atau metode non-linear mungkin diperlukan untuk meningkatkan kualitas prediksi.

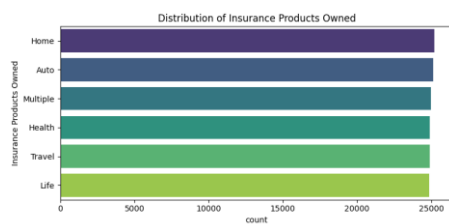
4.1.4 Bar Plot



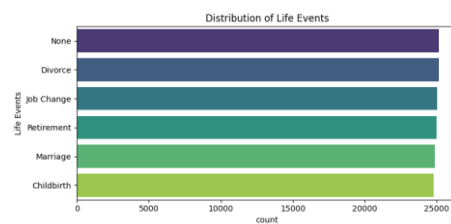
Gambar 4.15. Barplot 'Gender'



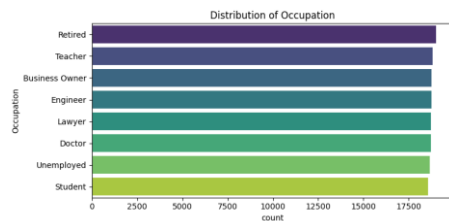
Gambar 4.16. Barplot 'Geographic Information'



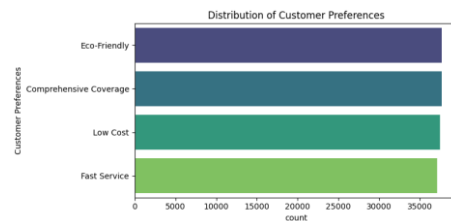
Gambar 4.17. Barplot 'Products Owned'



Gambar 4.18. Barplot 'Life Events'



Gambar 4.19. Barplot ‘Occupation’



Gambar 4.20. Barplot ‘Customer Preferences’

Data kategori utama pada pelanggan asuransi menunjukkan distribusi yang cukup seimbang. Gender terbagi merata antara laki-laki dan perempuan, lokasi mayoritas urban, dan produk asuransi yang dimiliki beragam. Peristiwa hidup seperti pernikahan dan pensiun tersebar merata, serta latar belakang pekerjaan beragam dari pelajar hingga profesional. Preferensi pelanggan juga bervariasi, mencakup layanan ramah lingkungan hingga biaya rendah. Data ini mencerminkan variasi yang baik untuk mendukung model prediksi premi.

4.2 Data Preprocessing

4.2.1 Handling Missing Value

Setelah melakukan pengecekan *missing value*, terdapat lima variabel yang memiliki *missing value* yang dijelaskan pada Tabel 4.1.

Tabel 4.1. Variable *missing value*

Variabel	Jumlah missing value
Income Level	7500
Education Level	7500
Behavioral Data	7500
Claim History	7500
Risk Profile	7500

Kelima variabel “Income Level”, “Education Level”, “Behavioral Data”, “Claim History”, dan “Risk Profile” yang memiliki *missing value*, daripada membuang baris-baris tersebut dan mengurangi ukuran sampel, setiap nilai kosong digantikan dengan nilai modus dari masing-masing variabel. Dengan imputasi modus ini kita mempertahankan keseluruhan data, menjaga distribusi asli agar tidak terdistorsi oleh kategori baru, dan memastikan tidak ada lagi nilai kosong yang dapat mengganggu proses analisis atau pemodelan.

4.2.2 Feature Engineering

Feature engineering dilakukan untuk membuat variabel tambahan yang didapatkan dari variabel yang telah ada. Hal ini bertujuan untuk memudahkan dalam mengolah data. Tabel 4.2 menunjukkan sepuluh variabel yang terbentuk pada tahap ini.

Tabel 4.2. Variabel *feature engineering*

Variabel	Tipe Data	Keterangan
Policy Duration Days	Integer	Selisih hari antara Policy Renewal Date dan Policy Start Date

Policy Duration Months	String	Durasi polis dalam bulan (hari dibagi 30, lalu di-cast ke integer)
Age Group	String	Binning usia menjadi “AYA” (18–39), “Adult” (40–64), “Senior Adult” (>64), atau “Unknown”
Age Coverage Interaction	Integer	Interaksi antara usia (Age) dan jumlah pertanggungan (Coverage Amount)
Coverage Deductible Interaction	Float	Interaksi antara jumlah pertanggungan (Coverage Amount) dan deductible (Deductible)
Credit Risk Interaction	Integer	Interaksi antara skor kredit (Credit Score) dan profil risiko (Risk Profile)
Income Marital Interaction	String	Interaksi antara tingkat pendapatan (Income Level) dan status pernikahan (Marital Status)
Has Claims	String	Penanda biner: 1 jika ada riwayat klaim, 0 jika “No Claims”
Multiple Claims	String	Penanda biner: 1 jika Previous Claims History > 1 (lebih dari satu klaim), 0 jika tidak
Customer Service Interaction Level	String	Level interaksi CS yang di binning (<5=Low, 5–14=Medium, ≥15=High) kemudian di-encode ke 0/1/2

Feature engineering dilakukan untuk membuat variabel tambahan yang didapatkan dari variabel yang telah ada. Hal ini bertujuan untuk memudahkan dalam mengolah data. Tabel 4.2 menunjukkan sepuluh variabel yang terbentuk pada tahap ini.

4.2.3 Encoding Data

Encoding data merupakan tahap penting dalam pra-pemrosesan data yang bertujuan mengubah variabel kategorikal menjadi bentuk numerik sehingga dapat diproses oleh algoritma machine learning. Proses ini memudahkan komputer dalam memahami dan mengolah data yang sebelumnya berbentuk kategori atau teks. Pada sub bab ini, dilakukan *encoding* terhadap berbagai variabel kategorikal yang ada pada dataset sesuai dengan skala dan kategorinya masing-masing. Tabel 4.3 menjelaskan skema *encoding* yang diterapkan pada setiap variabel, yang meliputi pengkodean numerik untuk kategori seperti tingkat pendidikan, jenis kelamin, jenis pekerjaan, hingga preferensi pelanggan.

Tabel 4.3. Variabel *encoding data*

Variabel	Keterangan
Education Level	0=High School, 1=Master, 2=Bachelor, 3=PhD
Gender	0=Female, 1=Male, 2=Non-Binary
Occupation	0=Retired, 1=Student, 2=Doctor, 3=Unemployed, 4=Lawyer, 5=Teacher, 6=Business Owner, 7=Engineer

Geographic Information	0=Suburban, 1=Rural, 2=Urban
Behavioral Data	0=Inactive, 1=New, 2=Active, 3=Occasional
Purchase History	0=Frequent, 1=Infrequent, 2=One-time
Claim History	0=1-2 Claims, 1=No Claims, 2=3+ Claims
Insurance Products Owned	0=Health, 1=Auto, 2=Multiple, 3=Life, 4=Home, 5=Travel
Policy Type	0=Annual, 1=Monthly, 2=Quarterly, 3=Semi-Annual
Customer Preferences	0=Eco-Friendly, 1=Low Cost, 2=Comprehensive Coverage, 3=Fast Service
Preferred Communication Channel	0=SMS, 1=Phone, 2=Email, 3=App Notification
Preferred Contact Time	0=Evening, 1=Afternoon, 2=Morning
Preferred Language	0=French, 1=English, 2=Spanish, 3=Mandarin, 4=Indonesian
Life Events	0=Marriage, 1=Divorce, 2=None, 3=Retirement, 4=Childbirth, 5=Job Change
Segmentation Group	0=C, 1=B, 2=A, 3=D, 4=E
Age Group	0=AYA (18–39), 1=Adult (40–64), 2=Senior Adult (>64), 3=Unknown
Customer Service Interaction Level	0=Low (<5 interaksi), 1=Medium (5–14 interaksi), 2=High (≥15 interaksi)
Marital Status	0=Widowed, 1=Divorced, 2=Married, 3=Single
Income Level	0=High, 1=Very High, 2=Medium, 3=Low
Risk Profile	0=Medium, 1=Low, 2=High

Dengan melakukan *encoding data* seperti yang ditunjukkan dalam Tabel 4.3, variabel kategorikal dapat direpresentasikan dalam format numerik yang konsisten dan terstruktur. Hal ini memungkinkan model *machine learning* untuk mengakses informasi yang terkandung dalam variabel tersebut secara efektif, serta meningkatkan kualitas analisis dan prediksi. Proses *encoding* ini juga menjadi dasar penting untuk tahap *feature engineering* dan seleksi fitur selanjutnya, yang secara keseluruhan bertujuan mengoptimalkan performa model prediksi.

4.2.4 Feature Selection

Dalam proses pemilihan fitur, digunakan tiga metrik utama untuk memastikan fitur yang dipilih relevan dan bermakna. Korelasi absolut mengukur hubungan linear antara fitur dan target, membantu mengeliminasi fitur yang kurang terkait secara langsung. *Mutual information* menangkap hubungan non-linear yang tidak terdeteksi korelasi, sehingga fitur dengan informasi penting tetap terpilih. Koefisien *Lasso* menunjukkan kontribusi fitur dalam model regresi dengan

regularisasi, memfilter fitur yang kurang berpengaruh. Kombinasi ketiga metrik ini membuat proses seleksi fitur lebih akurat dan meningkatkan performa model.

Tabel 4.4. Feature selection

Variabel	Correlation (abs)	Mutual Information	Lasso Coefficient
Education Level	0.00287	0	3.77364
Interactions with Customer Service	0.00280	0.00188	-0.68473
Insurance Products Owned	0.00625	0.00305	5.17302
Deductible	0.00252	0	-0.01357
Customer Preferences	0.00252	0	-3.20297

Dari tabel tersebut terlihat bahwa meskipun nilai korelasi absolut dan mutual information beberapa fitur tergolong kecil, koefisien *Lasso* memberikan gambaran kontribusi yang lebih jelas dalam model regresi. Fitur seperti *Education Level* dan *Insurance Products Owned* memiliki koefisien positif yang signifikan, menandakan pengaruh kuat terhadap prediksi, sedangkan fitur seperti *Interactions with Customer Service* dan *Customer Preferences* memiliki koefisien negatif, menunjukkan hubungan terbalik dengan target. Proses seleksi fitur ini memastikan hanya variabel yang relevan dan berkontribusi signifikan yang digunakan dalam model, sehingga dapat meningkatkan akurasi dan kestabilan prediksi.

4.3 Model Training

Setelah tahap *preprocessing data* selesai, langkah berikutnya adalah melakukan pelatihan model regresi untuk memprediksi nilai target, yaitu premium asuransi. Dalam studi ini, tiga model regresi digunakan yaitu Linear Regression, Lasso Regression, dan Ridge Regression. Ketiga model ini dipilih untuk membandingkan kemampuan prediksi dan kestabilan model, terutama dalam menghadapi kemungkinan multikolinearitas dan overfitting pada dataset dengan banyak fitur. Evaluasi performa setiap model dilakukan menggunakan metrik MSE, RMSE, MAE, dan R-Squared, yang akan dijelaskan pada tabel berikut.

Tabel 4.5. Model training

Model	MSE	RMSE	MAE	R-Squared
Linear Regression	1.992.060,13662	1.411.403,60	1.221.934,82	0,00005
Lasso Regression	1.992.059,53837	1.411.403,39	1.221.935,16	0,00005
Ridge Regression	1992060,13324	1.411.403,60	1.221.934,83	0,00005

Berdasarkan hasil evaluasi pada Tabel 4.5, ketiga model menunjukkan performa yang relatif mirip dengan nilai metrik kesalahan yang hampir sama. Namun, Ridge Regression dipilih sebagai model terbaik karena kemampuannya dalam menangani multikolinearitas dan menjaga kestabilan koefisien tanpa menghilangkan fitur. Model ini lebih andal saat diaplikasikan pada data baru dan efektif dalam mengurangi risiko overfitting melalui regularisasi L2, sehingga memberikan keseimbangan optimal antara kompleksitas model dan performa prediksi.

4.4 Model Evaluasi

Setelah proses pelatihan model selesai, tahap selanjutnya adalah melakukan evaluasi untuk mengukur kinerja model dalam memprediksi nilai target. Evaluasi ini penting untuk mengetahui seberapa baik model dapat merepresentasikan pola data dan menghasilkan prediksi yang akurat. Berbagai metrik evaluasi regresi digunakan dalam penelitian ini, yaitu Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), dan koefisien determinasi (R Square). Tabel 4.6 menyajikan hasil evaluasi model berdasarkan metrik-metrik tersebut.

Tabel 4.6. Metrik evaluasi

Metrik Evaluasi	Score
MSE	1.992.060,13324
RMSE	1.411.403,60
MAE	1.221.934,83
R Square	0,0005

Hasil evaluasi pada Tabel 4.6 menunjukkan bahwa meskipun nilai kesalahan prediksi seperti MSE sebesar 1.992.060,13, RMSE sebesar 1.411,40, dan MAE sebesar 1.221,93 memberikan gambaran tingkat akurasi model, nilai R Square yang sangat rendah yaitu 0,0005 mengindikasikan keterbatasan model dalam menjelaskan variasi data target. Angka R Square tersebut menunjukkan bahwa model hanya mampu menjelaskan sekitar 0,05% variasi target, sehingga performa model dalam menangkap pola data masih sangat terbatas. Kondisi ini menunjukkan perlunya pengembangan model lebih lanjut, baik dengan memperkaya fitur, menerapkan teknik seleksi fitur yang lebih optimal, ataupun mencoba metode model lain. Evaluasi ini menjadi acuan penting untuk peningkatan kualitas model pada tahap berikutnya agar prediksi premi asuransi dapat lebih akurat dan terpercaya.

BAB V PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian ini, kesimpulan yang dapat diambil adalah sebagai berikut:

1. Faktor-Faktor yang Mempengaruhi Tingkat Risiko dan Penetapan Premi Asuransi

Berdasarkan analisis data yang dilakukan, faktor-faktor yang paling signifikan dalam mempengaruhi tingkat risiko dan menjadi dasar penetapan premi asuransi di PT AXA Insurance Indonesia, khususnya di Jakarta, meliputi data profil pelanggan, detail polis, dan riwayat interaksi nasabah. Variabel-variabel seperti tingkat pendapatan (Income Level), tingkat pendidikan (Education Level), perilaku pelanggan (Behavioral Data), riwayat klaim (Claim History), dan profil risiko (Risk Profile) memiliki peran penting dalam menentukan estimasi risiko serta besar premi yang harus dibayar oleh nasabah. Dalam hal ini, langkah imputasi data pada variabel-variabel dengan missing value sangat berperan dalam menjaga keakuratan model, memastikan data yang digunakan dalam proses analisis dan prediksi tetap utuh, serta menghindari distorsi data. Dengan demikian, proses ini memastikan bahwa hasil analisis menjadi lebih representatif dan dapat diandalkan dalam pengambilan keputusan terkait penetapan premi asuransi.

2. Rancangan Model Optimasi Premi yang Efektif

Pengembangan model optimasi premi yang efektif untuk PT AXA Insurance Indonesia dilakukan melalui beberapa tahap penting, yang meliputi data preprocessing, feature engineering, encoding data, dan seleksi fitur. Pada tahap ini, tiga model regresi digunakan untuk memprediksi nilai premi, yaitu Linear Regression, Lasso Regression, dan Ridge Regression. Berdasarkan hasil evaluasi yang dilakukan, model Ridge Regression dipilih sebagai model terbaik karena kemampuannya dalam menangani multikolinearitas serta menjaga kestabilan koefisien fitur tanpa menghilangkan fitur yang relevan. Meskipun hasil evaluasi menunjukkan nilai R-Square yang sangat rendah (0.0005), yang mengindikasikan keterbatasan model dalam menjelaskan variasi data, model ini tetap lebih stabil dan dapat digunakan sebagai dasar untuk pengembangan lebih lanjut. Hal ini menunjukkan bahwa meskipun model prediksi premi asuransi yang digunakan sudah cukup efektif, masih ada ruang untuk pengembangan guna meningkatkan akurasi dan ketepatan dalam penetapan premi di masa depan.

3. Strategi Implementasi yang Direkomendasikan

Berdasarkan temuan dari model optimasi premi, terdapat beberapa rekomendasi strategi implementasi yang dapat diterapkan oleh PT AXA Insurance Indonesia untuk meningkatkan akurasi penetapan premi dan daya saing produk. Pertama, penerapan model yang lebih canggih disarankan untuk mengatasi rendahnya nilai R-Square yang menunjukkan keterbatasan model dalam menjelaskan variasi data. Dengan memperkaya fitur yang relevan dan mencoba teknik model lainnya, seperti ensemble methods atau algoritma non-linear, model dapat lebih efektif dalam menangani data yang lebih kompleks.

Kedua, penyesuaian premi yang responsif dapat dilakukan dengan menggunakan model prediktif berbasis data untuk menyesuaikan premi secara dinamis sesuai dengan perubahan kondisi pasar, fluktuasi ekonomi, dan

perubahan regulasi. Hal ini akan memastikan bahwa premi yang ditetapkan lebih sesuai dengan kondisi terkini dan profil risiko nasabah. Terakhir, peningkatan kualitas layanan dan kepuasan pelanggan sangat penting. Dengan menggunakan hasil analisis ini untuk lebih memahami profil nasabah, perusahaan dapat menawarkan premi yang lebih adil dan sesuai dengan kebutuhan mereka, yang pada gilirannya dapat meningkatkan kepuasan dan loyalitas pelanggan terhadap produk asuransi yang ditawarkan.

5.2 Saran

Untuk meningkatkan efektivitas model optimasi premi di PT AXA Insurance Indonesia, disarankan untuk memperkaya fitur data dengan menambahkan variabel eksternal seperti tren pasar dan faktor ekonomi makro, yang dapat membantu meningkatkan akurasi prediksi. Selain itu, eksperimen dengan model machine learning lain seperti random forest atau neural networks bisa memperbaiki prediksi pada data yang lebih kompleks. Penggunaan teknik seleksi fitur yang lebih canggih, seperti feature importance dan PCA, juga dapat membantu dalam memilih fitur yang paling relevan, sehingga model menjadi lebih efisien dan akurat. Selain itu, penting untuk melakukan evaluasi model secara berkala untuk memastikan relevansi dan akurasi prediksi seiring perubahan kondisi pasar. Pengembangan sistem penyesuaian premi secara dinamis yang responsif terhadap perubahan ekonomi dan regulasi juga dapat menjadi strategi yang efektif. Terakhir, fokus pada pengalaman pelanggan dengan menawarkan premi yang lebih personal dan sesuai dengan profil risiko nasabah akan meningkatkan loyalitas pelanggan dan memperkuat posisi PT AXA Insurance Indonesia di pasar yang semakin kompetitif.

KONTRAK

A. Judul dan Deskripsi Proyek

Judul Proyek:

PREMO (Premium Optimization through Regression Modeling)

Deskripsi:

Proyek Premium Optimization PT AXA Insurance Indonesia menggunakan metode regresi dan teknik machine learning bertujuan untuk meningkatkan akurasi prediksi premi asuransi. Model ini diharapkan dapat membantu PT AXA Insurance Indonesia dalam menetapkan premi yang optimal berdasarkan karakteristik pelanggan dan data historis, sehingga mendukung keputusan bisnis yang lebih tepat dan efisien.

B. Pihak yang Terlibat

Klien:

Nama Perusahaan : PT AXA Insurance Indonesia

Alamat : AXA Tower Lantai 11, Jalan Prof. Dr. Satrio Kav. 18, Karet Kuningan, Setiabudi, RT.14/RW.4, Kuningan, Karet Kuningan, Kecamatan Setiabudi, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12940

Email : risna.yodanti@axa.co.id

Perwakilan : Yodanti Risna Widya

Konsultan:

Nama Perusahaan : PT Vonetis

Alamat : Jl. Dr. Ir. H. Soekarno, Mulyorejo, Kecamatan Mulyorejo, Kota Surabaya, Jawa Timur 60115

Email : arkanattaqy09@gmail.com

Perwakilan : Arkan Syafiq At'taqy

C. Timeline Proyek

Tugas	Week - 1	Week - 2	Week - 3	Week - 4
Menyusun Rencana				
Eksplorasi Data				
Preprocessing Data				

Modelling dan Train Data				
Evaluasi Model				
Deployment & Dokumentasi				

D. Biaya dan Pembayaran

Biaya Proyek:

- Uang Starter: **Rp26.380.000**
 - Biaya ini mencakup manajemen proyek, pengolahan dan pra-pemrosesan data, pengembangan model, evaluasi, pengujian, deployment website prediksi, serta keahlian tenaga kerja yang diperlukan.
- Harga maintenance: **Rp5.276.000/tahun**
 - Biaya ini meliputi pembaruan data, pemeliharaan model, perbaikan bug, optimasi performa, penyesuaian fitur, dan dukungan teknis setelah proyek selesai.

Skema Pembayaran:

- Uang Muka (DP)
 - Jumlah: 50% dari uang starter (Rp26.380.000)
 - Waktu Pembayaran: Dibayarkan di muka sebelum proyek dimulai
 - Nilai: Rp13.190.000
- Pembayaran Sisa Uang Muka
 - Jumlah: 50% dari uang starter (Rp26.380.000)
 - Waktu Pembayaran: Dibayarkan setelah penyelesaian dan pengujian model, serta serah terima hasil proyek
 - Nilai: Rp13.190.000
- Pembayaran Maintenance
 - Jumlah: Rp440.000/bulan
 - Waktu Pembayaran: Dibayarkan secara berkala sesuai kesepakatan setelah proyek selesai

Rincian Biaya:

Posisi/Layanan	Estimasi Biaya (Rp)	Keterangan
IT Project Manager	4.000.000	Manajemen, Koordinasi, Monitoring
Data Scientist	7.000.000	Data Processing, Modelling, Evaluasi
Data Engineer	4.000.000	Preprocessing, Pipeline Setup
Front End Developer	3.000.000	UI/UX Website
Back End Developer	3.000.000	API & Backend Integration
Google Colab Pro	180.000	Langganan Selama Proyek
Infrastruktur & Tools	1.000.000	Cloud Storage, Komunikasi, Repositori
Manajemen Proyek	2.000.000	Reporting, Dokumentasi, Meeting

& Dokumentasi		
Kontinjensi (10%)	2.200.000	Cadangan Risiko dan Kebutuhan Tambahan
Total	26.380.000	

E. Hak dan Kewajiban

Hak Klien

- Klien berhak menerima hasil pekerjaan sesuai dengan ruang lingkup dan spesifikasi yang telah disepakati.
- Klien berhak mendapatkan laporan lengkap mengenai proses, hasil pemodelan, serta dokumentasi teknis dan manual pengguna.
- Klien berhak melakukan review dan meminta klarifikasi atas deliverable proyek

Kewajiban Klien

- Klien wajib menyediakan data yang lengkap, akurat, dan relevan untuk keperluan pengolahan dan pengembangan model.
- Klien wajib memberikan akses serta keterangan data yang diperlukan kepada konsultan secara tepat waktu.
- Klien wajib melakukan pembayaran sesuai dengan skema dan jadwal yang telah disepakati dalam kontrak.

Hak Konsultan

- Konsultan berhak mendapatkan kompensasi sesuai dengan kesepakatan yang tertulis dalam kontrak.
- Konsultan berhak mendapatkan dukungan dan informasi yang diperlukan dari klien untuk kelancaran proyek.
- Konsultan berhak melakukan revisi sesuai dengan ruang lingkup perubahan yang telah disetujui.

Kewajiban Konsultan

- Konsultan wajib menyelesaikan proyek sesuai dengan spesifikasi yang telah ditentukan.
- Konsultan wajib memberikan laporan hasil akhir proyek kepada klien.

F. Perubahan dan Revisi

Prosedur Perubahan

- Setiap permintaan perubahan dari klien harus disampaikan secara tertulis dan disetujui oleh kedua belah pihak.
- Perubahan yang mempengaruhi ruang lingkup, biaya, atau jadwal proyek harus dinegosiasikan dan dituangkan dalam addendum kontrak.
- Konsultan akan mengevaluasi permintaan perubahan dan memberikan estimasi dampak terhadap biaya dan waktu pengerjaan.

Permintaan Perubahan

- Revisi atas hasil pekerjaan dapat diajukan oleh klien dalam batas waktu yang disepakati setelah deliverable diserahkan.
- Revisi yang tidak termasuk dalam ruang lingkup awal dapat dikenakan biaya tambahan sesuai kesepakatan.

G. Penghentian Kontrak

Syarat Penghentian

- Kontrak dapat dihentikan oleh salah satu pihak dengan pemberitahuan tertulis minimal 7 (tujuh) hari sebelum tanggal penghentian.
- Penghentian kontrak dapat dilakukan jika salah satu pihak tidak memenuhi kewajiban secara material setelah diberikan kesempatan perbaikan.

Dampak Penghentian

- Jika kontrak dihentikan sebelum proyek selesai, klien wajib membayar biaya sesuai dengan progres pengerjaan yang telah dilakukan sampai saat penghentian.
- Konsultan wajib menyerahkan semua hasil kerja dan dokumen terkait yang telah diselesaikan sampai tanggal penghentian.
- Kerahasiaan dan hak kekayaan intelektual atas hasil kerja yang telah diserahkan tetap berlaku meskipun kontrak dihentikan.

DAFTAR PUSTAKA

- Akerlof, G. A. (2020). The market for "lemons": Quality uncertainty and the market mechanism. *Quarterly Journal of Economics*, 84(3), 488-500. <https://doi.org/10.2307/1885326>
- Anchen, J., & Dave, A. (2019). Advanced analytics: Unlocking new frontiers P&C insurance. Swiss Re Institute. <https://www.swissre.com/institute/research/sigma-research/sigma-4-2019.html>
- Buhl, A., Dreiling, J., & Müller, M. (2021). Machine learning in insurance: A review of predictive modeling in premium optimization. *Journal of Risk and Insurance*, 88(4), 1049-1067. <https://doi.org/10.1111/jris.12435>
- Calvo, R., & Jin, D. (2025). Reimagining insurance premium optimization with AI and pricing tech. NTT DATA. <https://www.nttdata.com/global/en/insights/blog/reimagining-insurance-premium-optimization-with-ai-and-pricing-tech>
- Ejjami, Rachid. (2024). Machine Learning Approaches for Insurance Pricing: A Case Study of Public Liability Coverage in Morocco. *International Journal For Multidisciplinary Research*. Volume 6. 10.36948/ijfmr.2024.v06i03.20511.
- Festa, E. (2022). Actuarial group takes steps to identify racial bias in insurance rates. Investopedia. <https://www.investopedia.com/news/actuarial-group-takes-steps-identify-racial-bias-insurance-rates/>
- Hu, Y., Le, Y., Hang, T., & Feng, J. (2020). Stream-flow forecasting of small rivers based on LSTM. *arXiv preprint*. <https://doi.org/10.48550/arXiv.2001.05681>
- Klein, N. (2024). Distributional regression for data analysis. *Annual Review of Statistics and Its Application*, 11, 321–346. <https://doi.org/10.1146/annurev-statistics-040722-053607>
- Li, X., & Zhang, X. (2023). A comparative study of statistical and machine learning models on carbon dioxide emissions prediction of China. *Research Square*. <https://doi.org/10.21203/rs.3.rs-3070359/v1>
- Löffler, G., Mühlbacher, P., & Müller, M. (2020). Dynamic pricing in insurance: A survey. *Journal of Risk and Insurance*, 87(4), 1011-1045. <https://doi.org/10.1111/jori.12215>
- Löffler, M., Klemperer, P., & Aydin, S. (2020). Data-driven pricing and its effects on market efficiency: The case of insurance. *Journal of Risk and Insurance*, 87(5), 1151-1173. <https://doi.org/10.1111/jris.12409>
- Mienye, I. D., & Swart, T. G. (2024). A comprehensive review of deep learning: Architectures, recent advances, and applications. *Information*, 15(12), 755. <https://doi.org/10.3390/info15120755>
- National Association of Insurance Commissioners (NAIC). (2025). Big data. <https://content.naic.org/insurance-topics/big-data>
- Pike, K. P., Roggenbaum, F., & Garsson, F. (2023). Big data and algorithms are revolutionizing the insurance industry. *Pennsylvania Association of Mutual Insurance Companies*. <https://pamic.org/big-data-and-algorithms-are-revolutionizing-the-insurance-industry/>
- PT AXA Insurance Indonesia. (2024). Laporan Tahunan 2024. Diakses dari <https://www.axa.co.id/annual-report-2024>

- PT AXA Insurance Indonesia. (2025). Mengenal AXA Insurance Indonesia. Diakses dari <https://www.axa.co.id/tentang-axa/mengenal-axa-insurance-indonesia>
- Salibian-Barrera, M. (2022). Robust nonparametric regression: Review and practical considerations. arXiv preprint arXiv:2211.08376. <https://doi.org/10.48550/arXiv.2211.08376>

LAMPIRAN

Lampiran 1 - Link Drive

<https://drive.google.com/drive/folders/1wEPrrZ4n98pMzRvToKtyinNlxVg70T-b>

Lampiran 2 - Link GitHub

<https://github.com/arknsa/Premium-Optimization/tree/main>

Lampiran 3 - Link Dashboard

<https://premo-vonetis.streamlit.app/>