PREMIUM OPTIMIZATION

PT AXA INSURANCE INDONESIA

Group 11 - VONETIS

OUR BEST TEAM MEMBERS

Rizal Hidayatullah
164221031

Fauziah Hamidah Al Hanief
164221055

Arkan Syafiq At'taqy
164221062

Ulvi Azzahra
164221109

Ahmad Zulfikar Al Ghiffar
164221116

01

PERMASALAHAN

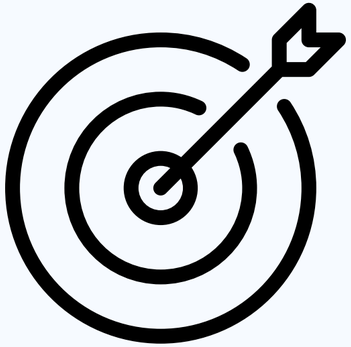
PT AXA Insurance Indonesia menghadapi **tantangan dalam menetapkan premi** yang mencerminkan risiko pelanggan secara tepat. Penetapan premi yang tidak sesuai dapat menyebabkan kerugian akibat klaim berlebihan (moral hazard) dan menarik lebih banyak nasabah berisiko tinggi (adverse selection). Dengan pesatnya digitalisasi dan persaingan yang ketat, perusahaan asuransi harus mengadopsi teknologi modern seperti machine learning untuk menganalisis data historis dan profil pelanggan secara mendalam.

Penggunaan model optimasi premi yang cerdas memungkinkan PT AXA Insurance Indonesia **untuk meningkatkan profitabilitas sekaligus menjaga kepuasan dan loyalitas nasabah**. Namun, tantangan utama adalah mengolah data kompleks dan variatif agar model dapat menghasilkan premi yang akurat dan responsif terhadap perubahan pasar.

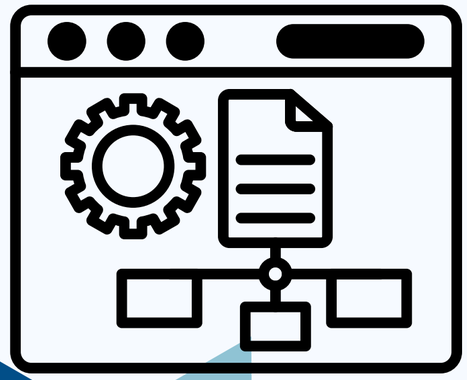
02

BATASAN

Data **internal** wilayah Jakarta



Keterbatasan cakupan produk asuransi



Model **sederhana**

TUJUAN

Identifikasi Faktor

Rekomendasi **Strategis**

Pengembangan Model

strategi & cara OPTIMASI premi

EDA

- Statistik Deskriptif
- Distribusi Data
- Korelasi antar Variabel

Preprocessing

- Handling Missing Value
- Feature Engineering
- Encoding
- Feature Selection

Modeling

- **Lasso Regression**
  - Regularisasi L1 untuk seleksi fitur dan model yang lebih sederhana.
- **Linear Regression**
  - Model dasar yang mudah dipahami sebagai baseline.
- **Ridge Regression**
  - Regularisasi L2 untuk mengatasi multikolinearitas dan mencegah overfitting.

Evaluasi

- **MSE**
  - Mengukur rata-rata kuadrat kesalahan prediksi, menunjukkan seberapa besar perbedaan prediksi dengan nilai aktual.
- **RMSE**
  - Akar dari MSE, memberikan interpretasi kesalahan dalam satuan asli target.
- **MAE**
  - Rata-rata kesalahan absolut, mengukur rata-rata besarnya kesalahan tanpa memperhatikan arah.
- **R-Squared**
  - Menunjukkan proporsi variasi target yang dapat dijelaskan oleh model.

04

EXPLORATORY DATA ANALYSIS

DIMENSI

Kolom: 30
Baris: 150.000

TIPE

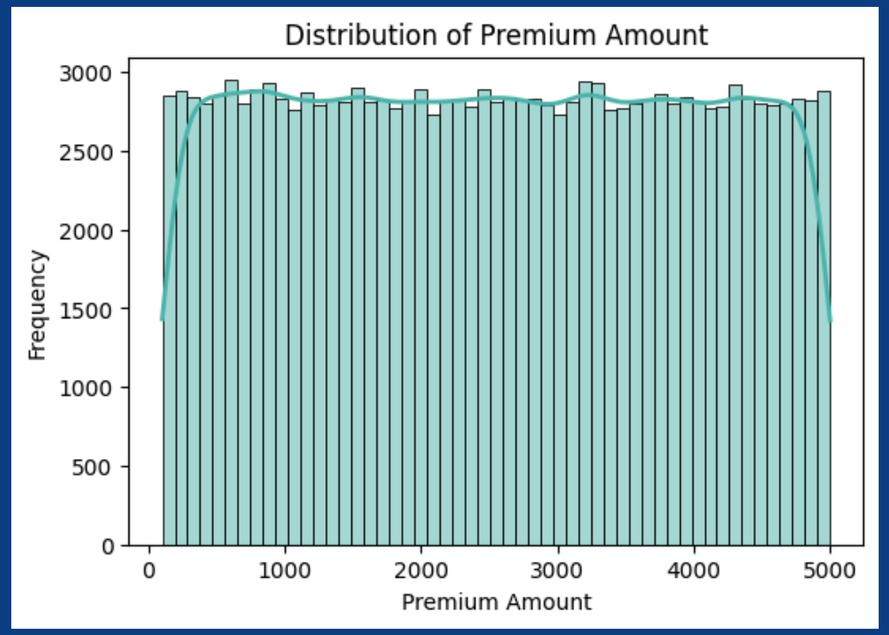
Kategorik: 20
Numerik: 6
Tanggal: 2

Dataset yang digunakan dalam proyek ini terdiri dari 150.000 entri dengan 30 kolom yang mencakup berbagai fitur terkait karakteristik pelanggan, termasuk 20 variabel kategorik, 6 variabel numerik, dan 2 variabel tanggal. Target variabel dalam dataset ini adalah kolom Premium Amount, yang menunjukkan nilai premi yang dibayarkan oleh pelanggan. Distribusi premi tersebut ditampilkan pada gambar di atas, yang memperlihatkan frekuensi nilai premi yang relatif merata di seluruh rentang nilai, tanpa konsentrasi pada nilai tertentu.

EXPLORATORY DATA ANALYSIS

05

Matriks menunjukkan korelasi antar variabel sangat rendah, hampir mendekati nol. Ini berarti tidak ada hubungan linear kuat antar fitur dan dengan premi. Model perlu menangkap pola non-linear agar prediksi lebih akurat.



Distribusi target (Jumlah Premi) cenderung merata di seluruh rentang nilai, tanpa adanya konsentrasi di nilai tertentu. Hal ini menunjukkan variasi premi yang luas dan seimbang di antara pelanggan.

06

DATA PREPROCESSING

Handling Missing Values

Mengatasi nilai yang hilang dengan menghapus kolom yang memiliki banyak missing dan mengimputasi nilai modus untuk kolom yang tersisa agar data lengkap dan konsisten.

Encoding Data

Mengubah variabel kategorik menjadi bentuk numerik menggunakan teknik label encoding agar dapat diproses oleh algoritma machine learning.

Feature Engineering

Membuat variabel baru dari data asli untuk meningkatkan kualitas model, seperti mengubah tanggal menjadi durasi polis dan membuat interaksi antar fitur penting.

Feature Selection

Memilih fitur-fitur paling relevan dan berkontribusi signifikan terhadap target menggunakan metode korelasi, mutual information, dan regularisasi Lasso untuk meningkatkan akurasi dan efisiensi model.

MODELING

07

Model	MSE	RMSE	MAE	R-Squared
Linear Regression	1992060.13662	1411.40360	1221.93482	0.00005
Lasso Regression	1992059.53837	1411.40339	1221.93516	0.00005
Ridge Regression	1992060.13324	1411.40360	1221.93483	0.00005

Tiga model regresi digunakan untuk memprediksi premi, yaitu Linear Regression, Lasso Regression, dan Ridge Regression. Ketiga model menunjukkan hasil yang sangat mirip dengan nilai MSE sekitar 1.992.060 dan R-Squared sangat kecil (0,00005), menandakan model saat ini memiliki keterbatasan dalam menjelaskan variasi data. Meski demikian, penggunaan Lasso dan Ridge membantu mengatasi potensi overfitting dan multikolinearitas, menjaga kestabilan model

EVALUASI & HASIL

Metrik Evaluasi	Skor
MSE	1991743.8124164576
RMSE	1411.2915405459134
MAE	1221.8415552208162
R Square	0.00020612850556966933

Kemampuan model untuk menjelaskan variabel target hanya

0,05%

Hal ini menandakan performa model masih lemah dalam menangkap pola data, sehingga perlu pengembangan lanjutan melalui peningkatan fitur atau eksplorasi metode lain untuk menghasilkan prediksi premi yang lebih akurat.

TEMUAN MASALAH PADA DATA

09

Distribusi Variabel Target dan Fitur yang Hampir Merata

Banyak variabel seperti usia, premi, deductible, credit score, dan jumlah interaksi dengan customer service menunjukkan distribusi yang sangat merata atau hampir uniform. Hal ini menyulitkan model dalam menemukan pola yang signifikan karena tidak ada konsentrasi atau variasi yang menonjol.

Korelasi Antar Fitur dan dengan Target Sangat Rendah

Matriks korelasi memperlihatkan nilai korelasi antar fitur maupun dengan premi sangat rendah, bahkan hampir nol. Ini menandakan fitur tidak memiliki hubungan linear yang kuat dengan target, sehingga prediksi menggunakan model linear menjadi sangat menantang.

Variabel Kategorikal yang Seimbang Namun Tidak Memiliki Pola Jelas

Distribusi kategori pada fitur seperti occupation dan purchase history tersebar cukup merata tanpa adanya dominasi kategori tertentu, yang juga bisa membuat model sulit membedakan pengaruh tiap kategori.

Keterbatasan Data dalam Memodelkan Pola Premi

Dengan kondisi ini, model prediksi premi perlu strategi lanjutan untuk menangkap hubungan kompleks dan non-linear agar performa prediksi dapat meningkat.

PREMO (Premium Optimization through Regression Modeling)

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.

TIMELINE PROYEK

FIRST WEEK

- Menyusun rencana
- Eksplorasi data

SECOND WEEK

Preprocessing data

THIRD WEEK

- Modeling
- Train data

FOURTH WEEK

- Evaluasi model
- Deployment dan dokumentasi

OUR ADVANTAGE

Fitur mencakup

a. Pengolahan data & pemodelan
b. Evaluasi + dokumentasi
c. Deployment website prediksi premi

AFFORDABLE PRICES

BIAYA UTAMA PROYEK

Hanya **Rp26.380.000**

MAINTENANCE TAHUNAN

Hanya **Rp5.276.000**

SKEMA PEMBAYARAN

UANG MUKA (DP)

Rp13.190.000 (50%)

Dibayar sebelum proyek dimulai

PELUNASAN

Rp13.190.000 (50%)

Dibayar setelah serah terima

MAINTENANCE

Rp440.000/bulan setelah proyek selesai

Previous Next

RINCIAN BIAYA PROYEK

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

Posisi/Layanan	Estimasi Biaya (Rp)	Keterangan
Google Colab Pro	180.000	Langganan Selama Proyek
Infrastruktur & Tools	1.000.000	Cloud Storage, Komunikasi, Repositori
Manajemen Proyek & Dokumentasi	2.000.000	Reporting, Dokumentasi, Meeting
Kontinjensi (10%)	2.200.000	Cadangan Risiko dan Kebutuhan Tambahan
Total	26.380.000	

SARAN

Perluas dan Perkaya Data Pelanggan

Mengumpulkan dan mengintegrasikan data tambahan, seperti data perilaku nasabah, kondisi keuangan yang lebih detail, dan informasi eksternal lain, untuk memberikan gambaran risiko yang lebih lengkap.

Lakukan Evaluasi dan Penyesuaian Premi Secara Berkala

Dengan pasar dan perilaku pelanggan yang dinamis, penyesuaian premi secara rutin penting agar tetap relevan dan kompetitif.

Fokus pada Pemahaman Pelanggan Secara Holistik

Melihat profil pelanggan dari berbagai aspek, tidak hanya demografi dan klaim, tapi juga preferensi, interaksi layanan, dan faktor lain yang mungkin memengaruhi risiko dan loyalitas.

Kolaborasi Antara Tim Bisnis dan Data

Pastikan hasil analisis dan model mudah dipahami oleh tim bisnis agar dapat diimplementasikan secara efektif dalam strategi penetapan premi.

Thank You