



Sains Data Consulting SD-A2



FAKULTAS TEKNOLOGI MAJU DAN MULTIDISIPLIN
fimmair

SISTEM PENDUKUNG KEPUTUSAN CROSS-SELLING PRODUK ASURANSI MENGGUNAKAN K-NEAREST NEIGHBOR DAN PREDIKSI BESARAN PREMI (STUDI KASUS AXA)

Bela (004), Jalu (009), Aura (053), Jovita (094), Kania (111)



01



PROBLEM STATEMENT

- Sebagian besar pelanggan asuransi tidak melakukan pembelian silang (cross-buying), bahkan setelah lima tahun memiliki produk.
- Pendekatan rule-based dan segmentasi statis yang digunakan perusahaan kurang adaptif terhadap dinamika perilaku pelanggan.
- Studi terdahulu lebih fokus pada klasifikasi peluang pembelian, tanpa mempertimbangkan personalisasi rekomendasi produk secara mendalam.
- Belum banyak pendekatan yang memadukan kesamaan profil pelanggan (KNN) dengan prediksi estimasi premi (regresi).



RUANG LINGKUP

Sistem rekomendasi produk asuransi tambahan menggunakan algoritma K-Nearest Neighbors (KNN), serta prediksi besaran premi dengan model regresi Random Forest.

02

DATA DAN PREPROCESSING

Jumlah Data

125025, khusus customer yang Insurance Product Owned-nya bukan Multiple →
target cross-selling

Variabel Prediktor KNN

27 variabel dari data → Age, Education Level, Income Level, dsb
5 variabel hasil Feature Engineering → (Tenure Days, Days Since Renewal, Interaction Rate, Deductible Coverage Ratio, Premium Coverage Ratio)
1 variabel hasil mapping (Location → Location Type)

Variabel Filtering Untuk KNN

Insurance Product Owned

Variabel Target Random Forest Regressor

Premium Amount

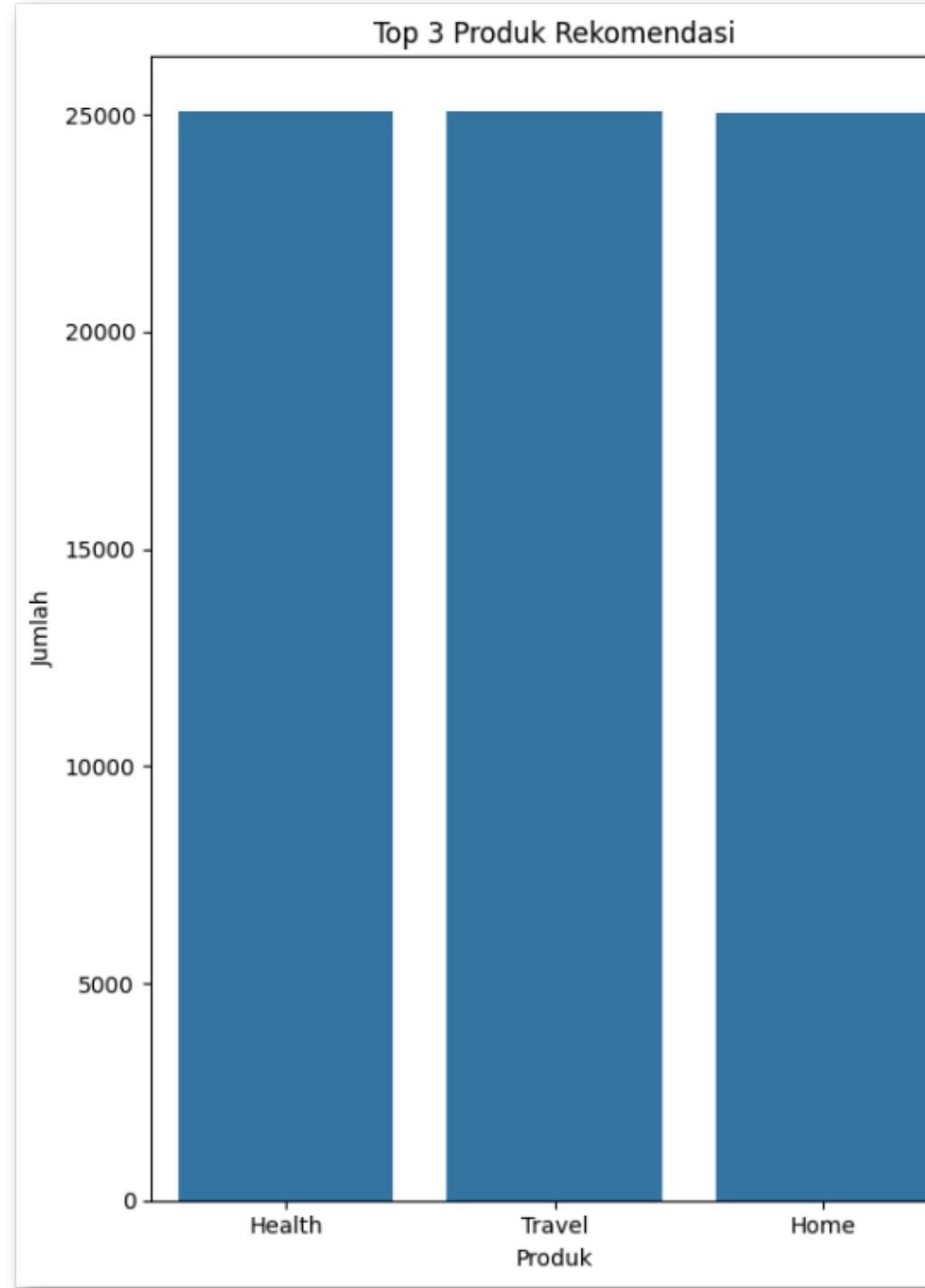
Preprocessing

- Imputasi Missing Values
- Melakukan Feature Engineering
- Encoding dan normalisasi variabel
 - Variabel numerik → Normalisasi dengan StandarScaler
 - Variabel kategorik ordinal (Income Level, Education Level, dsb) → OrdinalEncoder
 - Variabel kategorik nominal (Gender, Geographic Location, dsb) → OneHot Encoding
 - Variabel kategorik nominal dengan banyak kategori (Occupation, Location Type) → OrdinalEncoder untuk efisiensi memori

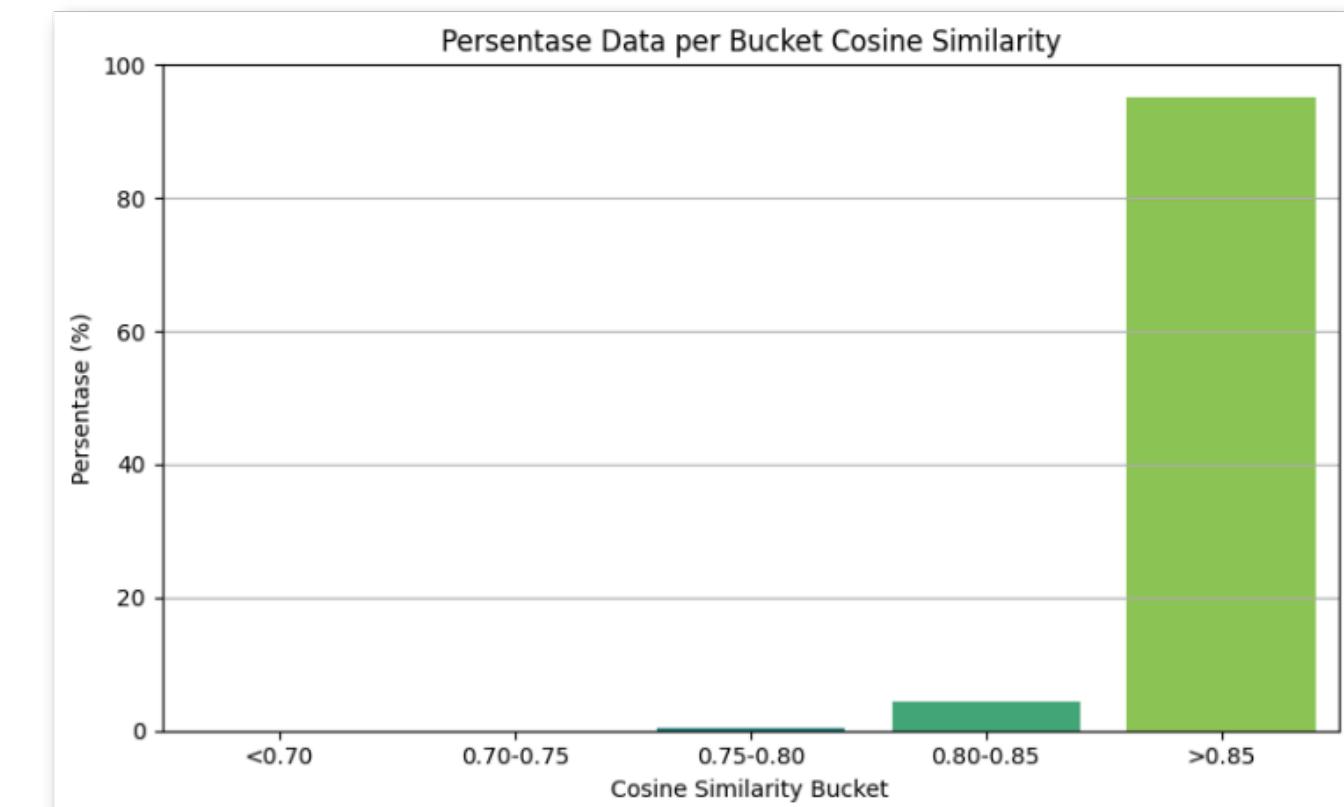
03



REKOMENDASI PRODUK DENGAN KNN



Produk Rekomendasi	Persentase (< 0,75)
Auto	3.62%
Health	2.79%
Home	2.39%
Life	5.63%
Travel	5.98%



Apa itu Cosine Similarity?

Ukuran tingkat kemiripan dua objek berbasis vektor. Nilainya berkisar dari minus satu hingga satu; semakin mendekati satu, semakin mirip kedua objek tersebut."

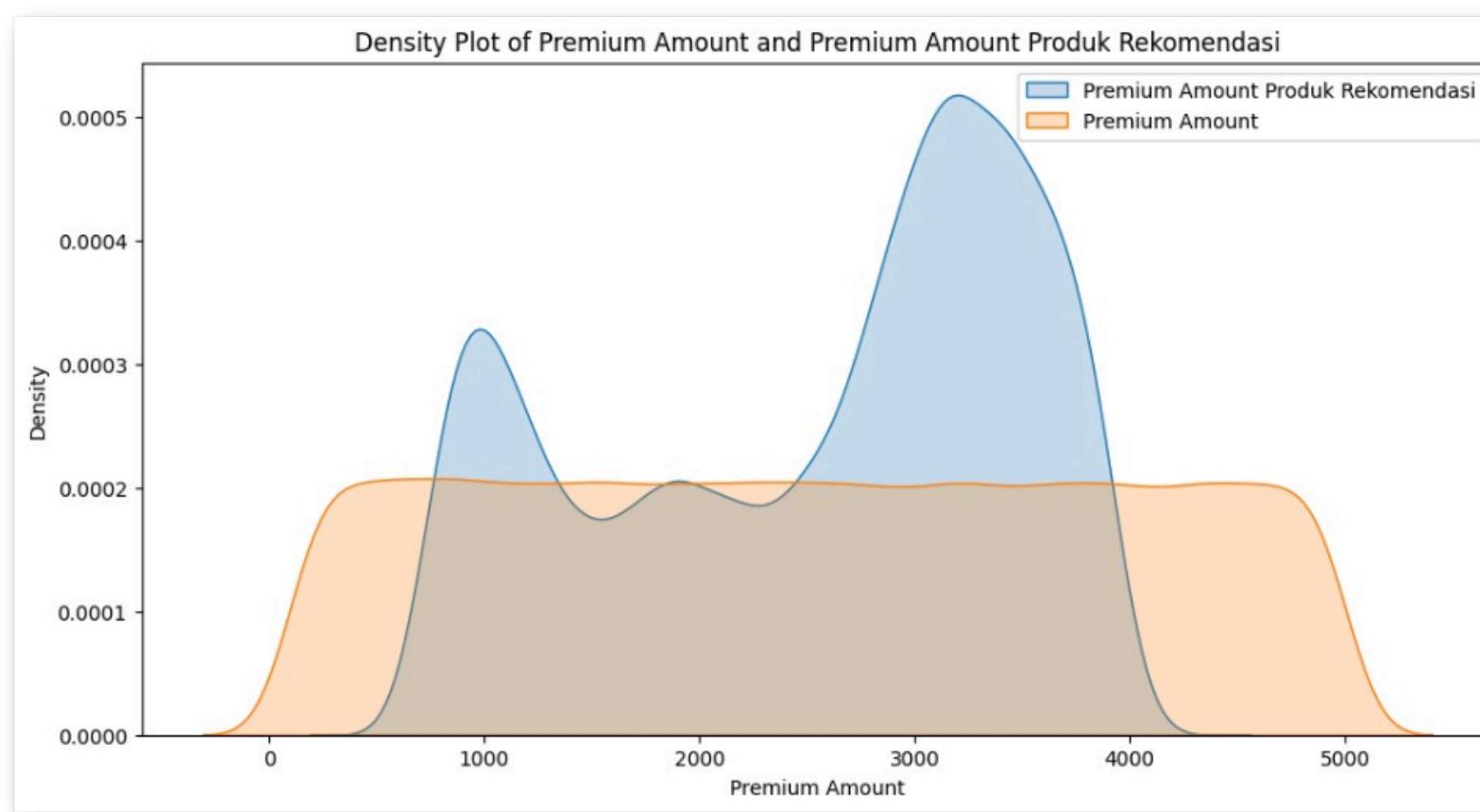
4,07 persen

hasil rekomendasi punya cosine similarity di bawah 0.75

PREDIKSI BESARAN PREMI

MODEL YANG DIGUNAKAN
RANDOM FOREST REGRESSOR

Produk	Mean Absolut Eror
Travel	495.77
Home	500.9
Health	504.01
Auto	505.56
Life	509.9

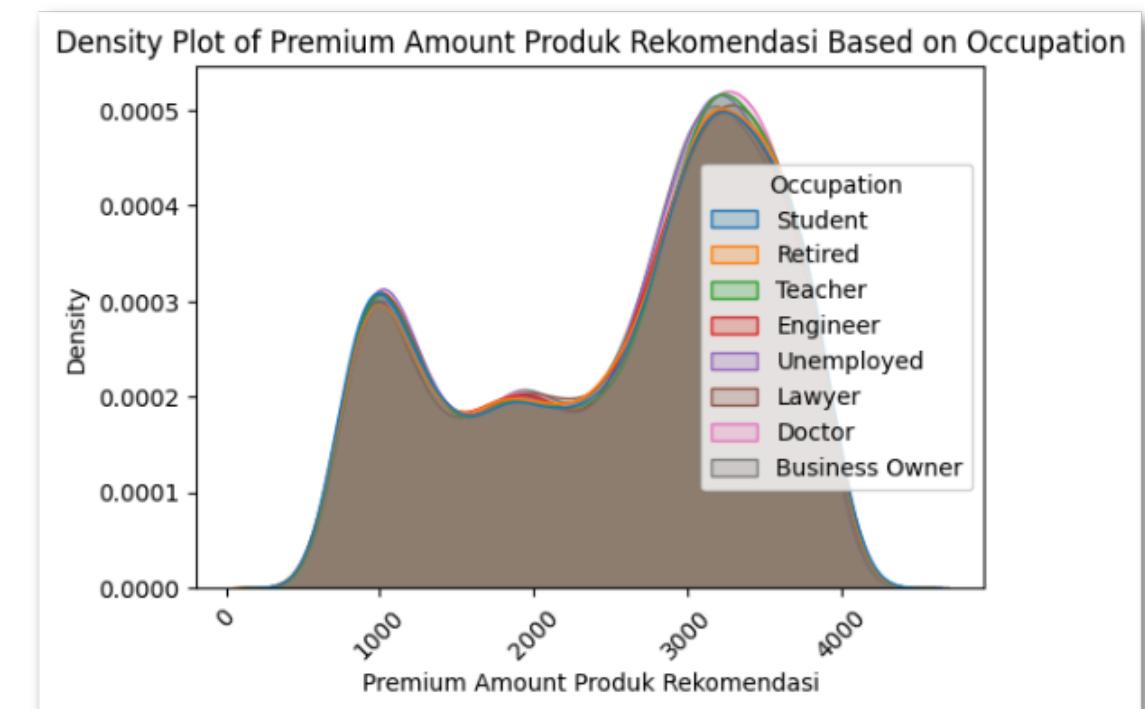
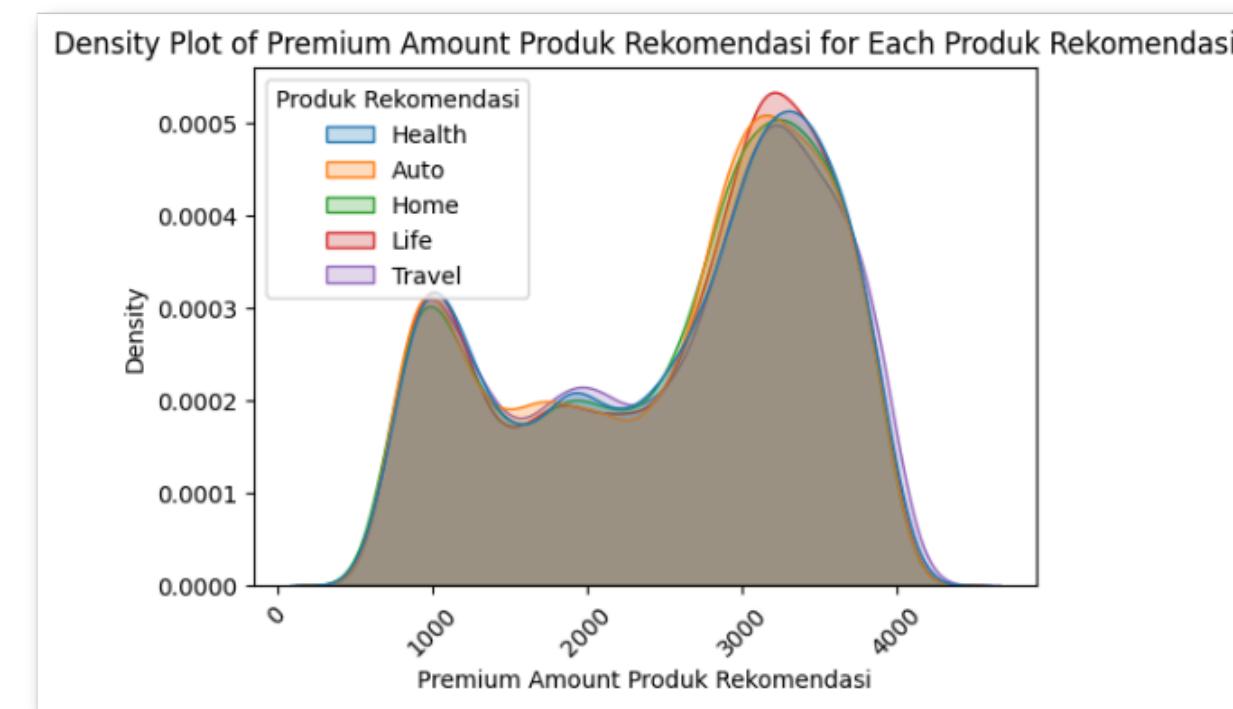
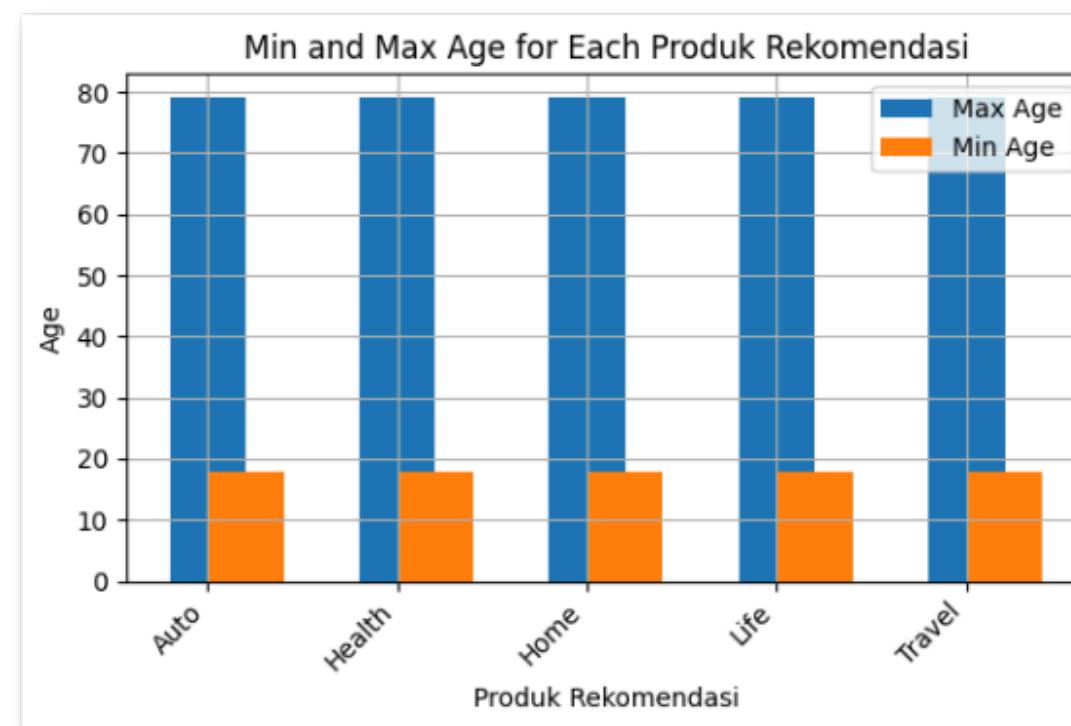
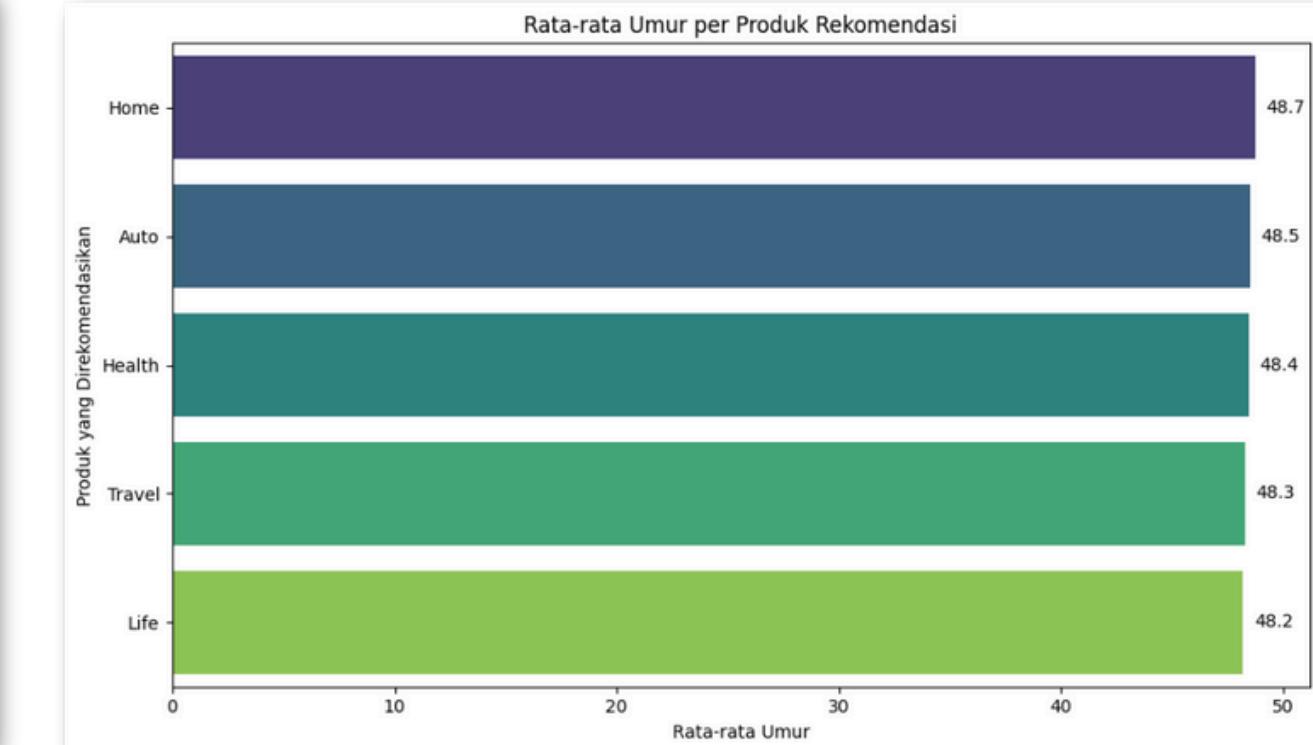
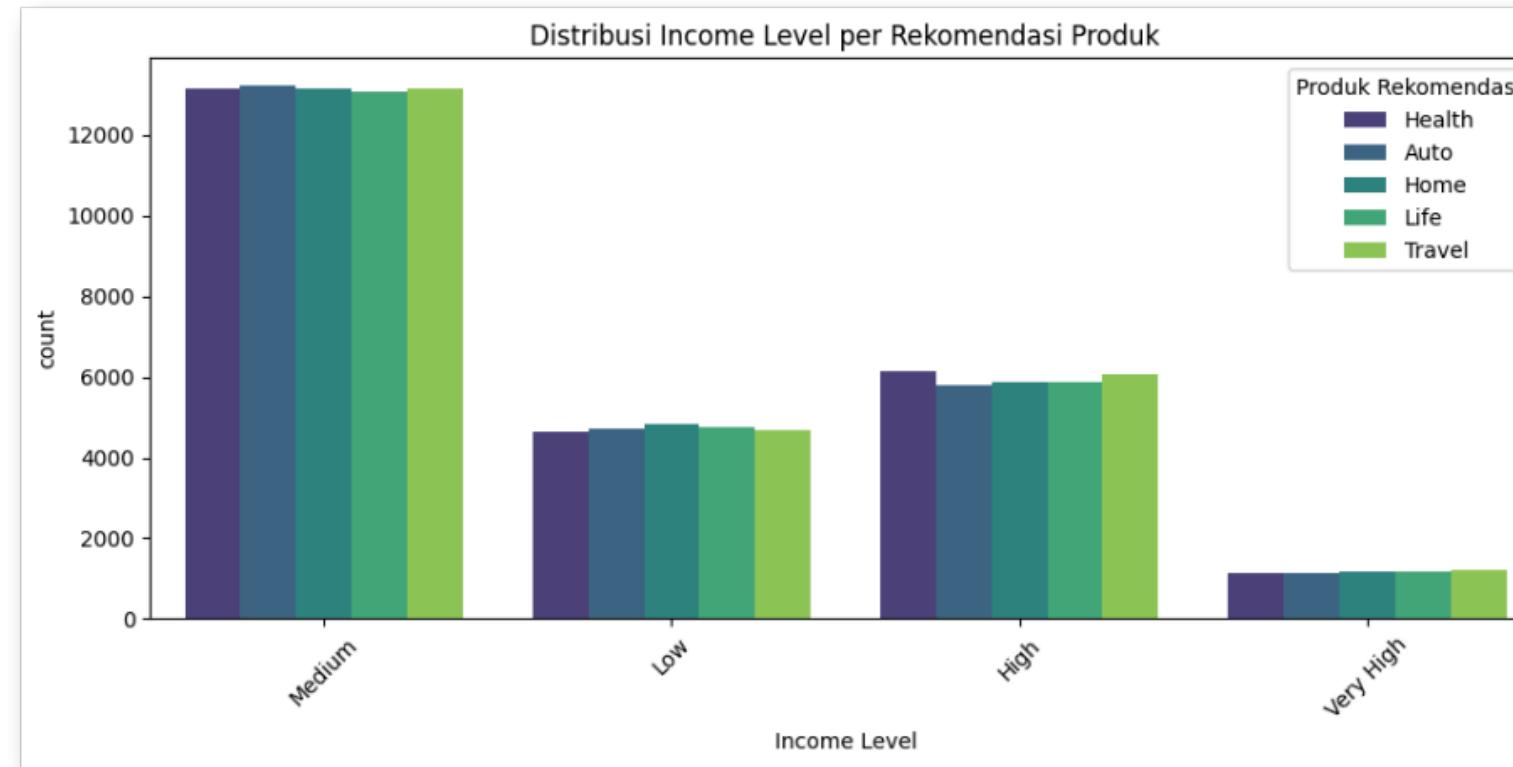


Produk Asuransi	Fitur	Nilai
Health	Premium_Coverage_Ratio	0.5761
	Coverage Amount	0.0839
	Deductible_Coverage_Ratio	0.0492
	Deductible	0.0259
	Tenure_Days	0.0211
Auto	Premium_Coverage_Ratio	0.6147
	Coverage Amount	0.0827
	Deductible_Coverage_Ratio	0.0451
	Deductible	0.0239
	Credit Score	0.0193
Home	Premium_Coverage_Ratio	0.5758
	Coverage Amount	0.0851
	Deductible_Coverage_Ratio	0.0503
	Deductible	0.0256
	Days_Since_Renewal	0.0214
Life	Premium_Coverage_Ratio	0.6098
	Coverage Amount	0.0847
	Deductible_Coverage_Ratio	0.0487
	Deductible	0.0231
	Tenure_Days	0.0188
Travel	Premium_Coverage_Ratio	0.5915
	Coverage Amount	0.0828
	Deductible_Coverage_Ratio	0.046
	Deductible	0.0243
	Tenure_Days	0.0205

HASIL REKOMENDASI DAN PREDIKSI

Customer ID	Produk Dimiliki	Produk Rekomendasi	Premium Amount Produk Rekomendasi	Cosine Similarity	Tetangga ID	Produk Tetangga	Premium Amount Tetangga
CUST086471	Travel	Auto	1576.18	0.92815	CUST140191	Auto	584.22
CUST124670	Home	Auto	2853.22	0.924487	CUST016587	Auto	3825.07
CUST091300	Auto	Health	2405.32	0.906572	CUST144921	Health	2319.83
CUST121697	Life	Health	3783.23	0.902447	CUST072743	Health	4608.27
CUST045183	Health	Home	2018.07	0.92455	CUST148597	Home	513.8
CUST093953	Auto	Home	3527.5	0.869616	CUST080360	Home	2957.18
CUST032952	Auto	Life	3242.4	0.930461	CUST108890	Life	2937.56
CUST083643	Travel	Life	3595.45	0.921393	CUST139990	Life	4903.83
CUST043894	Home	Travel	3258.69	0.908112	CUST147373	Travel	1844.75
CUST063970	Auto	Travel	3318.41	0.911759	CUST029772	Travel	4721.29

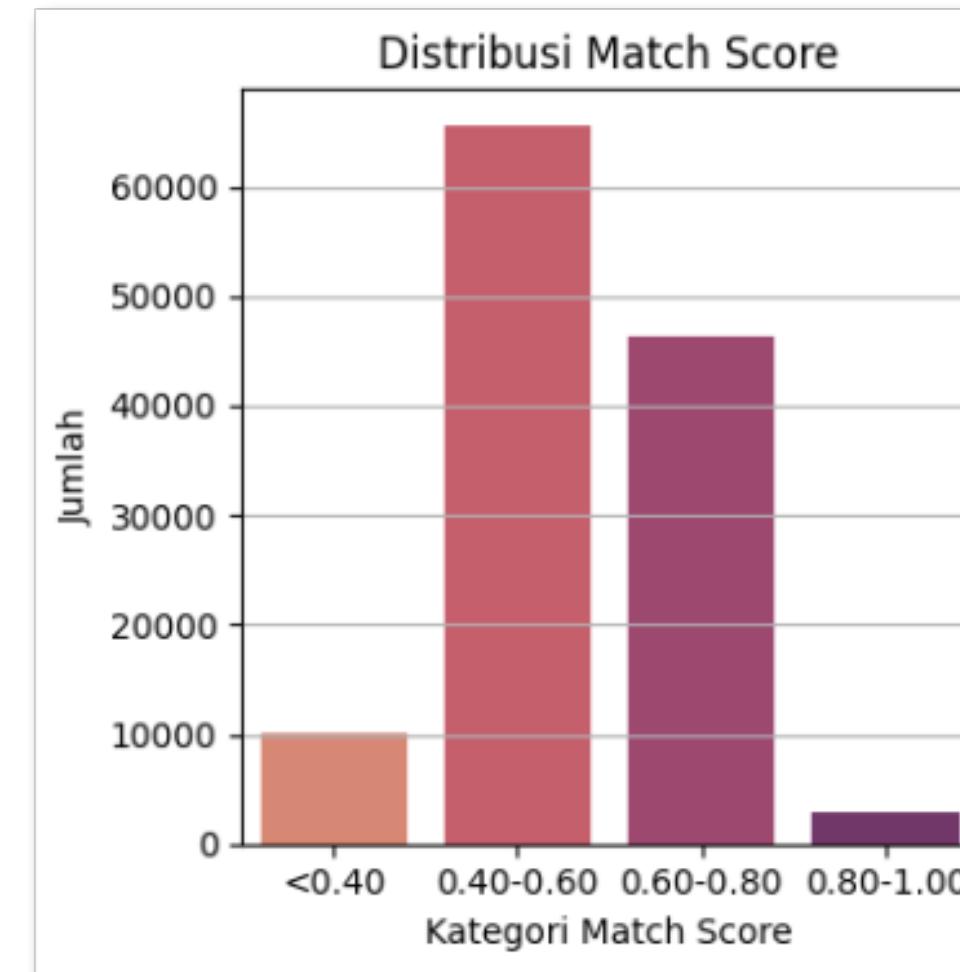
EXPLORATORY DATA ANALYSIS



INSIGHT DAN DISKUSI

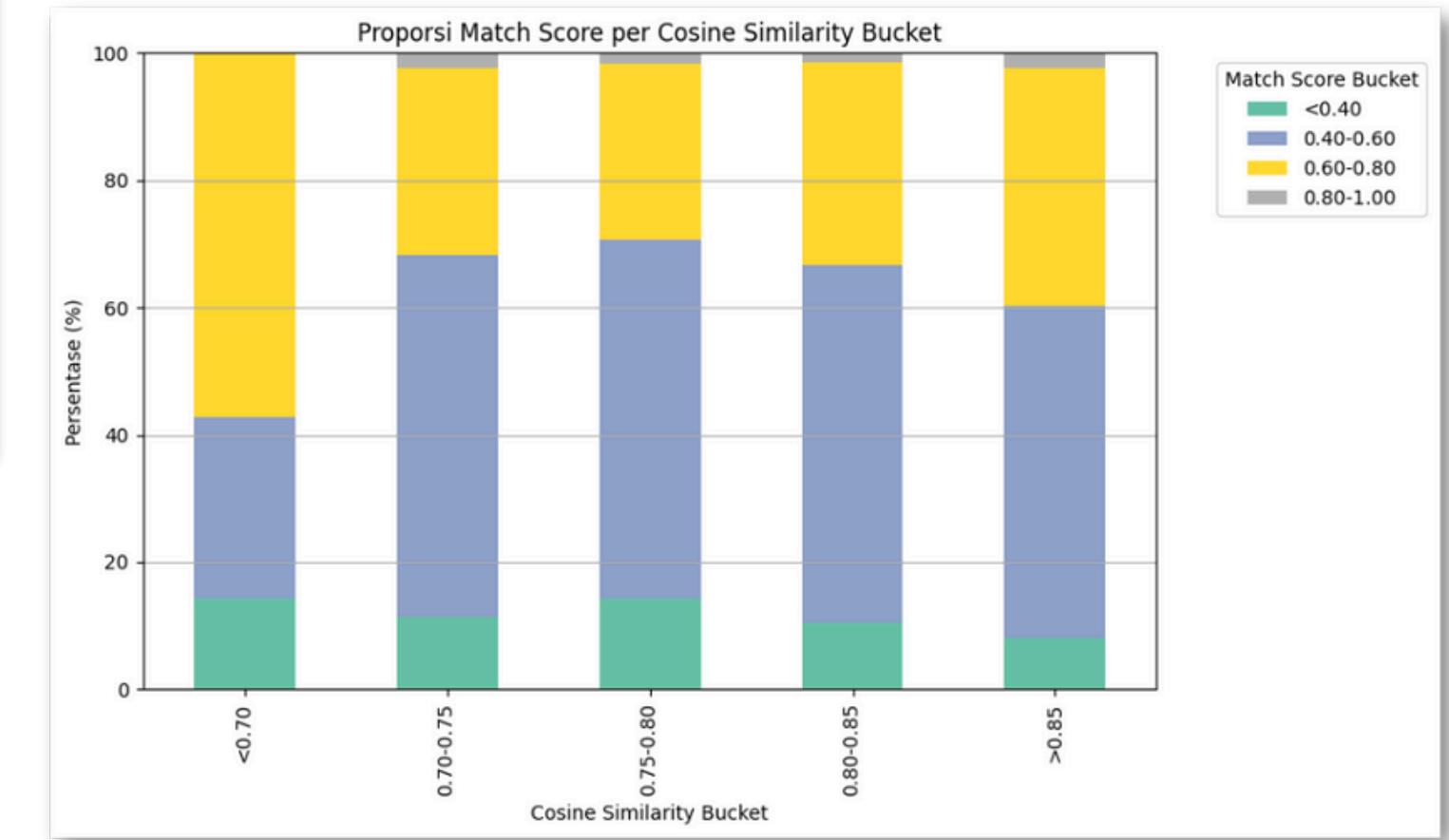
KRITERIA MATCH SCORE

Variabel	Kriteria
Age	Selisih 5 tahun
Coverage Amount	Selisih 20.000
Premium Amount	Selisih 10.000
Income Level	Harus sama persis
Risk Profile	Harus sama persis
Claim History	Harus sama persis
Customer Preferences	Harus sama persis
Segmentation Group	Harus sama persis
Premium Coverage Ratio	Selisih 0.05 (5%)
Deductible Coverage Ratio	Selisih 0.05 (5%)



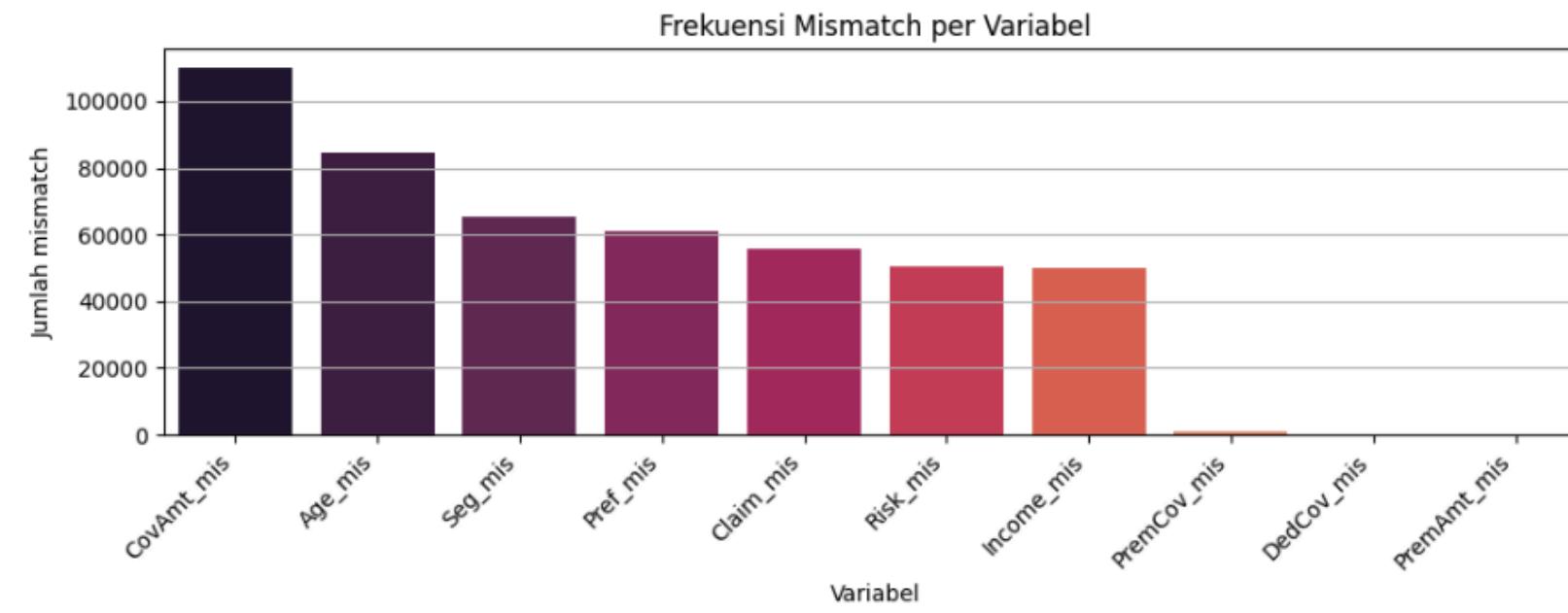
Sebagian besar rekomendasi berada pada match score 0,4–0,6, menunjukkan hasil yang cukup sesuai namun masih menyisakan mismatch yang perlu diperbaiki.

Meskipun nilai cosine similarity yang tinggi menunjukkan kemiripan profil, hal tersebut tidak selalu berbanding lurus dengan ketepatan rekomendasi, sehingga dibutuhkan validasi tambahan untuk meningkatkan relevansi hasil.

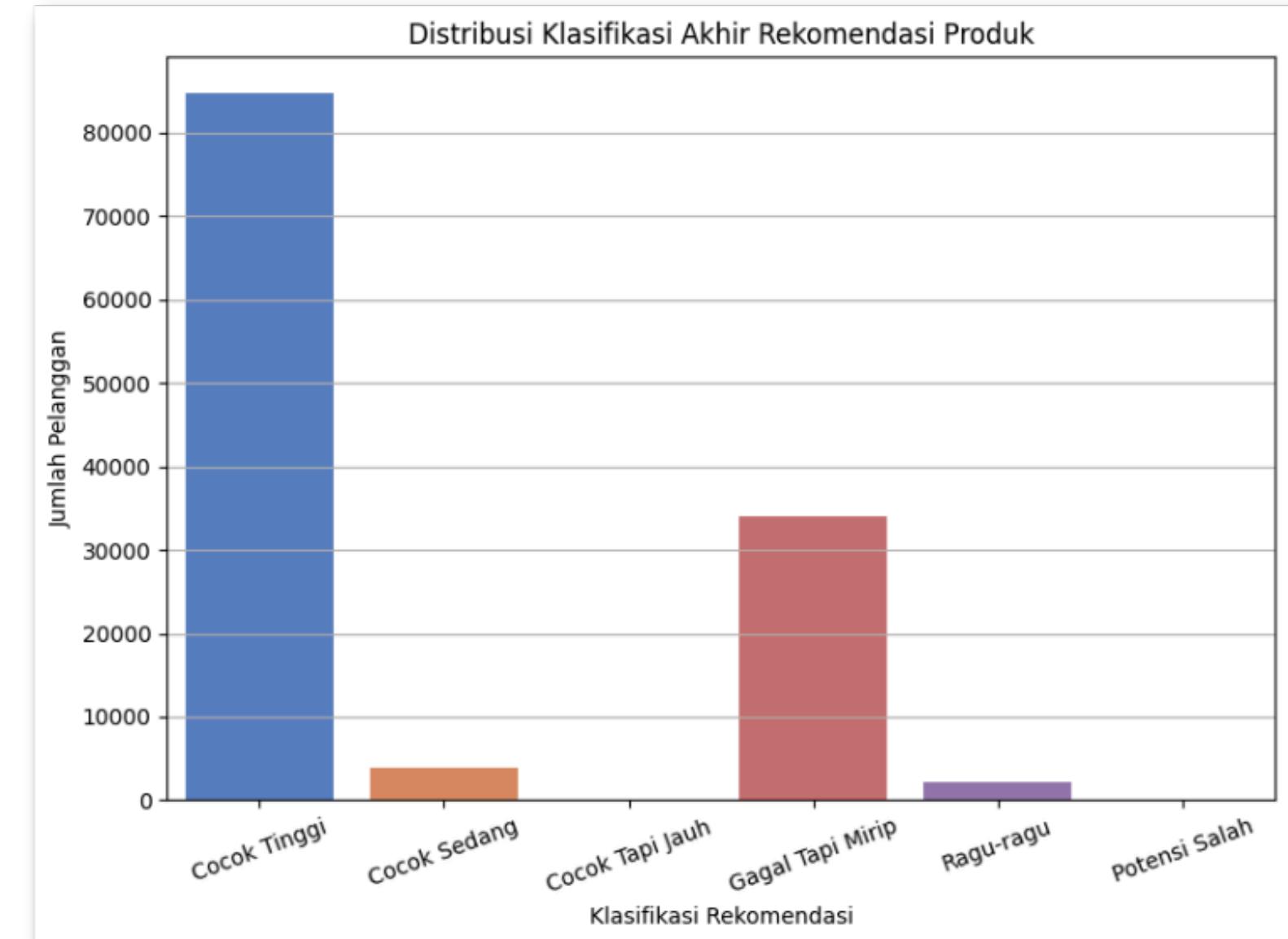




INSIGHT DAN DISKUSI



Cosine Similarity	Match Score	Klasifikasi
≥ 0.85	≥ 0.60	Cocok Tinggi
0.75 – 0.85	≥ 0.60	Cocok Sedang
< 0.75	< 0.60	Potensi Salah
≥ 0.85	< 0.60	Gagal Tapi Mirip
< 0.75	≥ 0.60	Cocok Tapi Jauh
0.75 – 0.85	< 0.60	Ragu-ragu

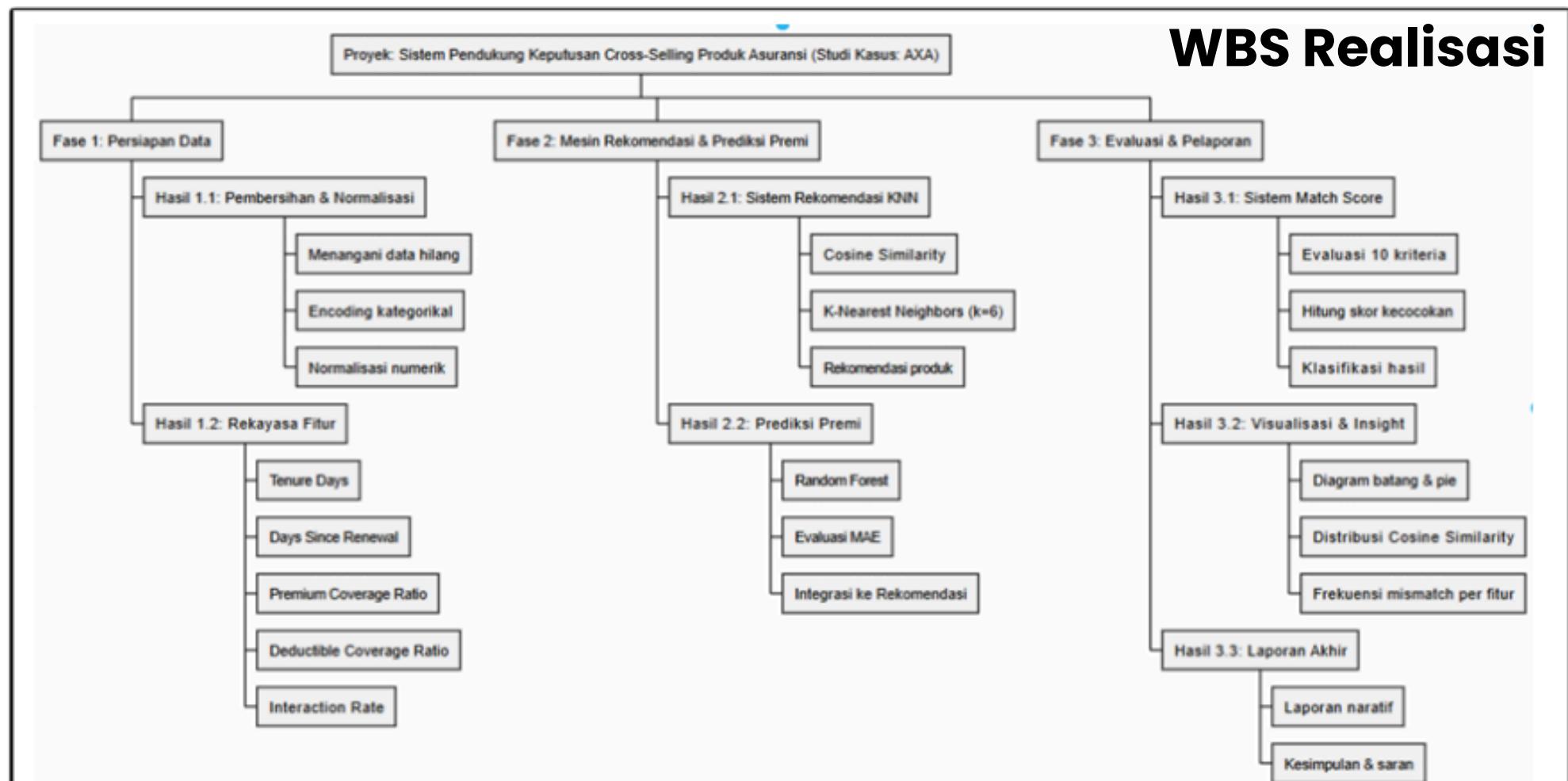
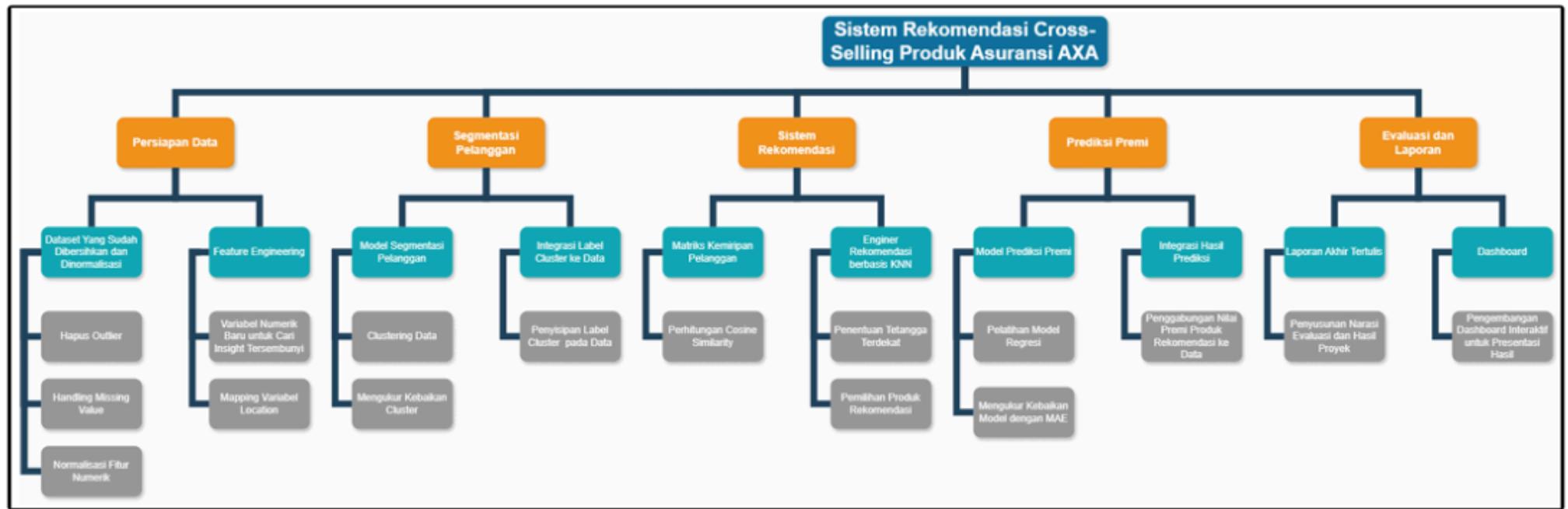


Mayoritas hasil rekomendasi berada pada kategori "Cocok Tinggi", menandakan sistem telah bekerja cukup baik meskipun masih perlu perbaikan pada kasus "Gagal Tapi Mirip".



PROJECT SCOPE MANAGEMENT

WBS Perencanaan



Realisasi

- Tanpa clustering, langsung menggunakan KNN berdasarkan kesamaan profil.
- Dashboard ditiadakan, diganti dengan visualisasi statis & laporan naratif.
- Ditambahkan analisis Match Score untuk interpretasi kecocokan rekomendasi.
- Deliverables akhir: Dataset + fitur rekayasa, Model KNN & RF, Evaluasi berbasis Match Score, Visualisasi hasil (bar chart, stacked bar), Laporan naratif lengkap sebagai output utama

Catatan: Proyek dibatasi hanya sampai pengembangan model offline, tanpa implementasi real-time atau integrasi sistem AXA.

HUMAN RESOURCE & COMMUNICATION MANAGEMENT

Aktivitas / Tahapan Proyek	PM	DS1	DS2	DE	BA
Penyusunan jadwal dan manajemen proyek	A	C	C	I	I
Pemahaman kebutuhan bisnis (business understanding)	C	I	I	I	R
Persiapan dan pembersihan data	I	I	I	R	C
Feature engineering dan normalisasi	I	R	R	C	C
Pengembangan model KNN	I	R	I	I	C
Pengembangan model prediksi premi (RF)	I	I	R	I	C
Evaluasi model dan metrik	I	R	R	C	C
Interpretasi hasil untuk rekomendasi bisnis	C	C	C	I	R
Penyusunan laporan akhir proyek	A	C	C	I	R
Komunikasi dengan stakeholder eksternal (AXA)	R	I	I	I	C

Keterangan

- R (**Responsible**): Pelaksana utama tugas
- A (**Accountable**): Penanggung jawab akhir
- C (**Consulted**): Dilibatkan untuk memberi masukan
- I (**Informed**): Diberi informasi terkait progres

Struktur Peran

- Project Manager (PM)
- Data Scientist (DS)
- Data Engineer (DE)
- Business Analyst (BA)
- Technical Writer (TW)

Aktivitas / Tahapan Proyek	PM	DS	DE	BA	TW
Penyusunan jadwal dan manajemen proyek	A	C	I	I	I
Pemahaman kebutuhan bisnis	C	I	I	R	I
Persiapan dan pembersihan data	I	I	R	C	I
Feature engineering dan normalisasi	I	R	C	C	I
Pengembangan model KNN	I	R	I	C	I
Pengembangan model prediksi premi (RF)	I	R	I	C	I
Evaluasi model dan metrik	I	R	C	C	I
Interpretasi hasil untuk rekomendasi bisnis	C	C	I	R	I
Penyusunan laporan akhir proyek	C	C	I	C	R
Komunikasi dengan stakeholder eksternal (AXA)	R	I	I	C	I

Dalam realisasinya, hanya ada satu data scientist dan ada role baru, yaitu Technical Writer

PROJECT TIME MANAGEMENT

Kegiatan	Waktu Perencanaan	Waktu Realisasi	Keterangan
Business & Data Understanding	1 – 16 Februari 2025	1 – 16 Februari 2025	Sesuai rencana
Data Preparation & Feature Engineering	17 Februari – 15 Maret 2025	17 Februari – 15 Maret 2025	Sesuai rencana
Modeling KNN & Random Forest	16 Maret – 20 April 2025	16 Maret – 25 April 2025	Durasi sedikit lebih panjang
Evaluasi Model & Validasi Rekomendasi	21 April – 5 Mei 2025	25 April – 13 Mei 2025	Dimulai sedikit mundur dan durasi penggerjaan sedikit lebih lama
Pembuatan Dashboard Interaktif	6 – 15 Mei 2025	-	Tidak direalisasikan
Penulisan Laporan & Finalisasi	11 – 26 Mei 2025	14 – 26 Mei 2025	Dimulai sedikit mundur tapi tetap selesai tepat waktu

Realisasi time management proyek berjalan cukup baik dan sesuai rencana secara keseluruhan. Meskipun terdapat penyesuaian scope seperti peniadahan tahap clustering dan dashboard interaktif, waktu penggerjaan tetap efisien karena tahapan tersebut digantikan dengan aktivitas yang lebih relevan seperti evaluasi Match Score dan visualisasi statis. Tidak ada keterlambatan signifikan karena risiko sudah diantisipasi melalui buffer waktu.



PROJECT COST MANAGEMENT

Komponen Biaya	Rincian	Estimasi	Realisasi
Tenaga Kerja	5 orang × 80 jam × Rp 60.000	Rp 24.000.000	Rp 24.000.000
Dashboard Interaktif	1 produk	Rp 1.000.000	-
Biaya Internet & Listrik	Rp 150.000/orang/bulan × 4 bulan × 5 orang	Rp 3.000.000	Rp 3.000.000
Total Biaya Proyek		Rp 28.000.000	Rp 27.000.000

Proyek direncanakan dengan anggaran Rp 28.000.000, mencakup tenaga kerja, operasional, dan pembuatan dashboard. Realisasi hanya Rp 27.000.000 karena dashboard interaktif ditiadakan. Efisiensi ini tidak memengaruhi kualitas hasil, karena visualisasi tetap disajikan secara informatif di laporan akhir.

PROJECT QUALITY MANAGEMENT

Perencanaan

Metode	Variabel yang Dievaluasi	Metrik Evaluasi	Batasan Kualitas
Clustering	Silhouette Score	Nilai Silhouette	$\geq 0,5$
	Davies Bouldin Index (DBI)	Nilai DBI	< 1
KNN (Rekomendasi Produk)	Cosine Similarity	Nilai Cosine Similarity	$\geq 0,75$
Prediksi Premi (Random Forest)	Mean Absolute Error (MAE)	Nilai MAE	≤ 500 dolar

Realisasi

Metode	Variabel yang Dievaluasi	Metrik Evaluasi	Batasan Kualitas
KNN	Cosine Similarity	Nilai Cosine Similarity	$\geq 0,75$
	Tingkat Kesesuaian Atribut	Match Score	$\geq 0,6$
Prediksi Premi (Random Forest)	Mean Absolute Error (MAE)	Nilai MAE	≤ 500 dolar

Namun, dalam pelaksanaannya, metode clustering ditiadakan dan digantikan dengan evaluasi kecocokan atribut menggunakan Match Score. Sementara itu, model prediksi premi tetap mempertahankan target $MAE \leq 500$ dolar, yang tercapai dengan hasil aktual sekitar 495–510 dolar.

PROJECT RISK MANAGEMENT

Perencanaan

ID	Risiko	Probabilitas (%)	Dampak (minggu)	EMV (minggu)
R1	Data nasabah tidak lengkap, banyak missing, atau terlalu seragam	70%	2 minggu	1,4 minggu
R2	Segmentasi clustering tidak menghasilkan klaster yang bermakna	50%	1,5 minggu	0,75 minggu
R3	Rekomendasi produk dari KNN tidak relevan secara logis/bisnis	65%	2 minggu	1,3 minggu
R4	Model prediksi premi memiliki tingkat kesalahan (MAE) tinggi	60%	1 minggu	0,6 minggu
R5	Interpretasi hasil sulit dipahami oleh pengguna non-teknis atau stakeholder	30%	1 minggu	0,3 minggu

ID	Resiko	Probabilitas	Dampak (minggu)	EMV (minggu)
R1	Data tidak lengkap, missing, atau terlalu seragam	70%	2 minggu	1,4 minggu
R2	Match score (cosine similarity) tidak valid dalam mencerminkan kemiripan	50%	1,5 minggu	0,75 minggu
R3	Rekomendasi KNN tidak logis atau tidak relevan secara bisnis	65%	2 minggu	1,35 minggu
R4	Error analysis menunjukkan bias atau pola anomali	40%	1 minggu	0,4 minggu
R5	Business insight sulit dipahami oleh pengguna akhir	25%	1 minggu	0,25 minggu

Perubahan metode dari clustering + KNN menjadi KNN langsung berhasil menekan risiko validasi segmentasi, meski menimbulkan tantangan baru terkait validitas kemiripan dan relevansi rekomendasi. Namun, proyek tetap selesai tepat waktu dengan buffer time sedikit berkurang, dari 4,35 menjadi 4,1 minggu.

KESIMPULAN DAN SARAN

- Proyek berhasil membangun sistem rekomendasi produk asuransi tambahan berbasis algoritma K-Nearest Neighbors (KNN) tanpa segmentasi clustering.
- Sistem mampu memberikan rekomendasi berdasarkan kemiripan profil pelanggan dengan menggunakan cosine similarity.
- Prediksi besaran premi produk rekomendasi menggunakan Random Forest Regressor menunjukkan performa yang cukup baik, dengan nilai MAE di bawah 510 dolar untuk semua kategori produk.
- Evaluasi sistem dilakukan dengan Match Score yang mencakup 10 kriteria kesesuaian, menghasilkan klasifikasi rekomendasi
- Proyek selesai tepat waktu dengan risiko yang berhasil diminimalkan, dan tidak ada keterlambatan signifikan dari rencana awal.

15
●
Proyek masih dapat dikembangkan dengan mengeksplorasi model sistem rekomendasi lainnya, misalnya dengan deep learning





TERIMA KASIH ATAS PERHATIANNYA

SELESAI