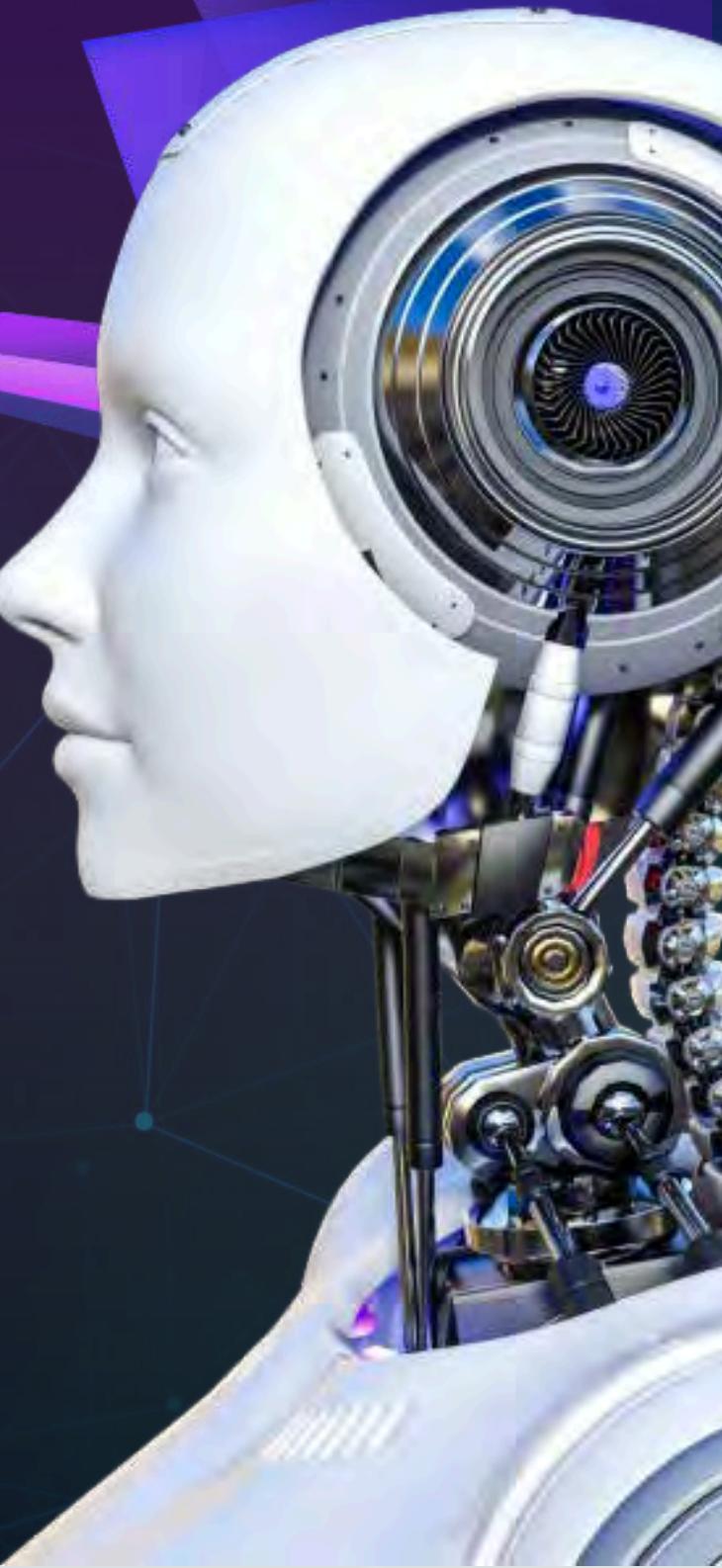




Team
Dawnbringer



Mikael Alexander Adikara Purnama | Raphael Angelo Adikara Purnama | Timotius Christopher Patrick Sihombing

“Setiap keluhan pelanggan adalah alarm dini bernilai jutaan dolar, data merupakan sumber suaranya”

66%

konsumen meninggalkan brand karena layanan buruk

Artificial Intelligence dapat menghemat waktu hingga

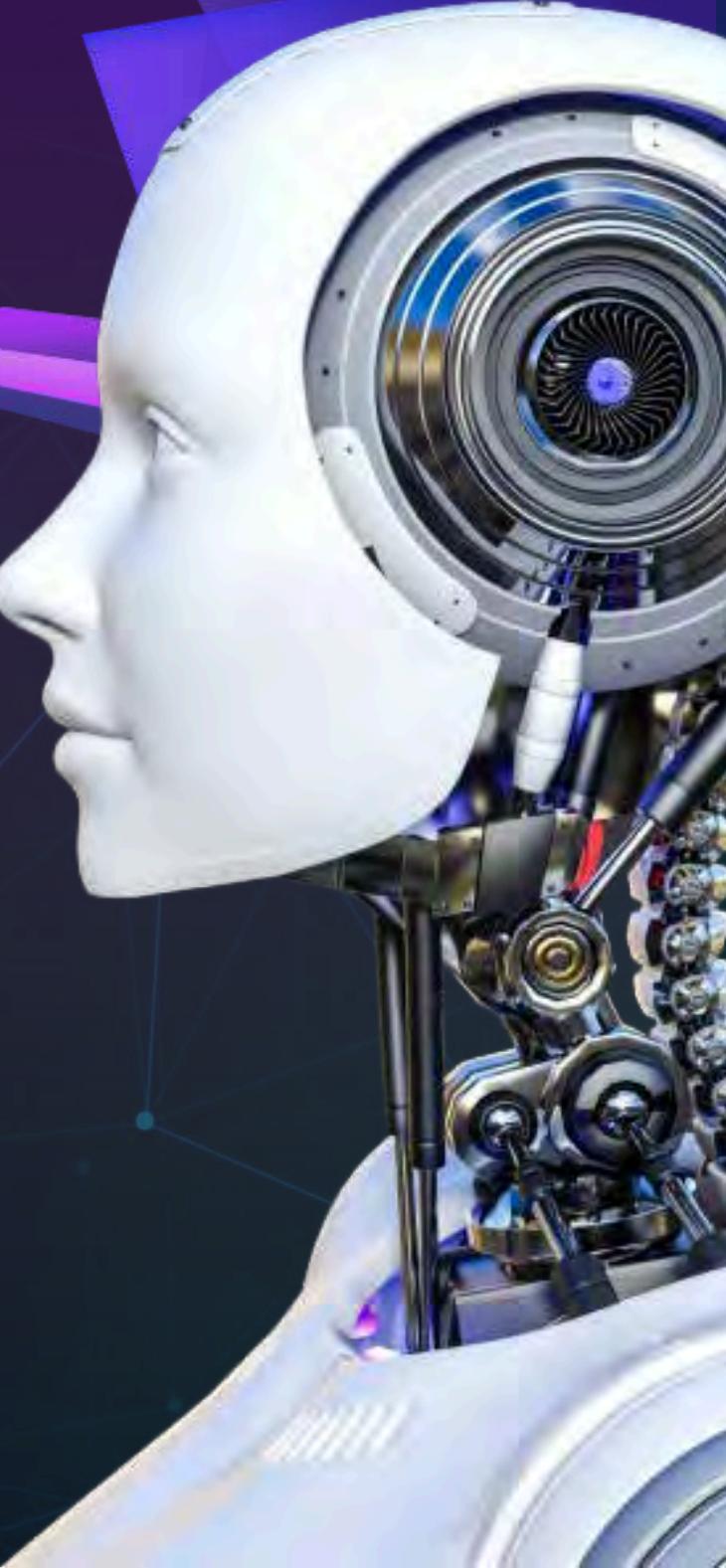
50%



Team
Dawnbringer

PROJECT RED FLAG

**"Keluhan Hari Ini, Krisis Besok
Kecuali Kita Bertindak"**



WHY NOW?

1 BIAYA TINGGI

Penanganan keluhan oleh agen = **\$6-\$12** per interaksi vs otomatis hanya beberapa sen.



2 PENGHEMATAN BESAR

AI percakapan dapat menghemat **\$80 miliar** biaya tenaga kerja global.



3 SIKLUS MERUGIKAN

Respons lambat → biaya operasional membengkak



Stres agen meningkat → turnover **30-45%** per tahun



Biaya penggantian agen: **\$10.000-\$15.000** per orang



Tidak mengadopsi sistem cerdas = menyerahkan keunggulan kompetitif + membayar “biaya ganda” (interaksi mahal + rekrutmen berulang).

Status quo bukan sekadar inefisiensi operasional, ini adalah kerentanan bisnis kritis.

14%

churn
langsung

66%

stop hubungan
bisnis

LEMBAGA KEUANGAN RENTAN

Retensi + **5%**

profit naik **25% - 95%**

McKinsey
& Company

churn bank ritel naik hingga 30%
di segmen tertentu

01

Tujuan & Value Proposition

1. Membangun model prediksi risiko keterlambatan respons keluhan.
2. Menghasilkan dashboard prioritas untuk membantu agen/customer support.
3. Menggali tema tersembunyi dalam keluhan untuk insight strategis.

02

Dampak Finansial

Jika rata-rata kerugian per churn Rp 2 juta dan 5% pelanggan churn karena keterlambatan → kerugian potensi Rp 20 miliar per 200 ribu pelanggan.

Konsumen → Respon lebih cepat & adil.
Perusahaan → Efisiensi reputasi meningkat.
Regulator → Transparansi penanganan isu.

03

Tantangan:

- Volume keluhan sangat besar → sulit diprioritaskan manual.
- Tema keluhan tidak selalu jelas (teks bebas, panjang, bervariasi).



DATASET:

Consumer Complaint Database – Consumer Financial Protection Bureau (CFPB), USA

- ± **331.793** keluhan konsumen terkait produk/layanan keuangan.

60.02% Dengan Narasi
39.98% Tanpa Narasi



FITUR

Company response to consumer
Consumer complaint narrative
Company public response
Consumer consent provided?
Submitted via
Date sent to company
Consumer disputed?
Complaint ID

Date received
Product
Sub-product
Issue
Sub-issue
Company
State
ZIP code
Tags



TARGET:

Timely Response = NO
(is_delayed)
apakah telat ditangani
sejak awal masuk?

Ekosistem Berbasis AI untuk Triase Proaktif

REAKTIF ➤ PROAKTIF



PENDEKATAN TERPADU

"Ekosistem AI dengan dua komponen utama"

1. **Intelligent Triage System (ITS)**
→ Fokus operasional (efisiensi respon cepat).
2. **Strategic Insights Engine (SIE)** → Fokus strategis (wawasan & pencegahan jangka panjang).



STRATEGIC INSIGHTS ENGINE

- Dasbor interaktif untuk wawasan pelanggan.
- Fitur utama:
 - **Topic Modeling** → Identifikasi tema berulang (penagihan, KPR, mobile app).
 - **Sentiment Analysis** → Skor emosional tiap keluhan.
 - **Root Cause Analysis** → Dari isu besar → sub-topik → kata kunci spesifik.

INTELLIGENT TRIAGE SYSTEM

- Mengubah paradigma:
First-In, First-Out → **Risk-First**.
- Output model:
 - **Risk Probability %**
 - **Severity Score**
- Keluhan langsung diberi:
 - Priority Level (**Rendah / Sedang / Tinggi**)
 - Recommended Action (misal: **eskalasi ke agen senior < 24 jam**).

Arsitektur Teknis & Validasi: Intelligent Triage System



1 PEMILIHAN MODEL

Algoritma	Kelebihan	Keterbatasan
CatBoost	- Native handling kategori - Prevent overfitting dengan ordered boosting - High performa tanpa tuning berlebih	- Training bisa lebih lambat dibanding LightGBM pada data numerik kosong
LightGBM	Sangat cepat & efisien memori (leaf-wise growth, histogram binning)	Butuh encoding untuk kategorial, tidak seterbuka CatBoost
XGBoost	Efisien, fleksibel, regulasi kuat (L1/L2), sangat populer	Memerlukan encoding kategorikal; tuning hyperparameter lebih kompleks

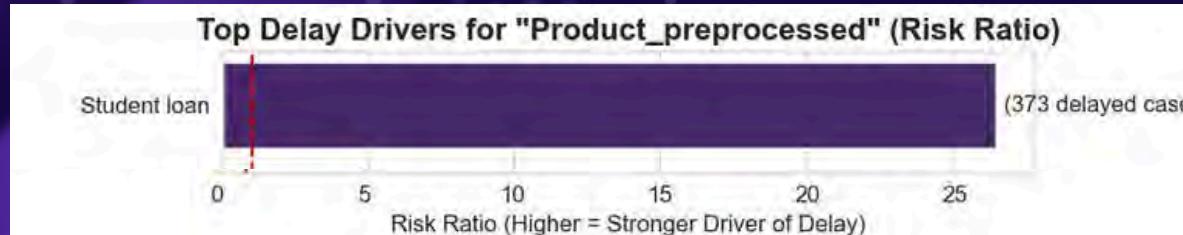
CatBoost dipilih karena kemampuannya yang ***superior*** dalam menangani **fitur kategorikal**

- Menangani fitur kategorikal secara native, tanpa preprocessing seperti encoding
- Menggunakan ***ordered boosting*** yang mencegah ***target leakage*** dan ***overfitting*** melalui mekanisme statistik berbasis target dan permutasi acak.
- Membentuk kombinasi fitur kategorikal secara otomatis untuk greedy feature combinations.

3

EXPLORATORY DATA ANALYSIS

Pinjaman mahasiswa (**Student loan**) muncul sebagai pendorong utama keterlambatan dengan rasio risiko jauh di atas baseline.



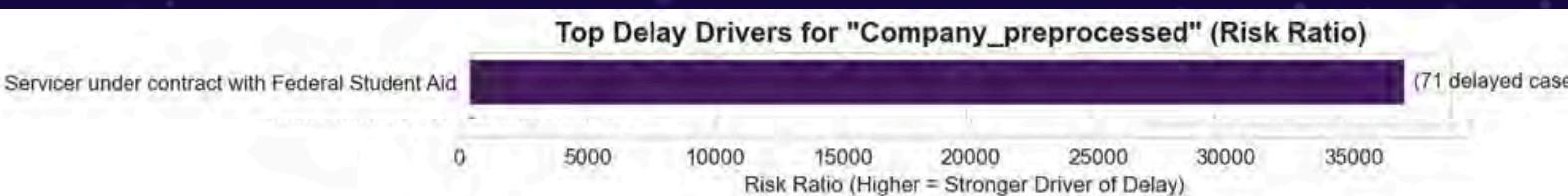
Perusahaan dengan kontrak **Federal Student Aid** menjadi pendorong keterlambatan paling dominan dengan rasio risiko sangat tinggi, jauh melampaui baseline.

Visualisasi Sementik Keluhan Menggunakan UMAP

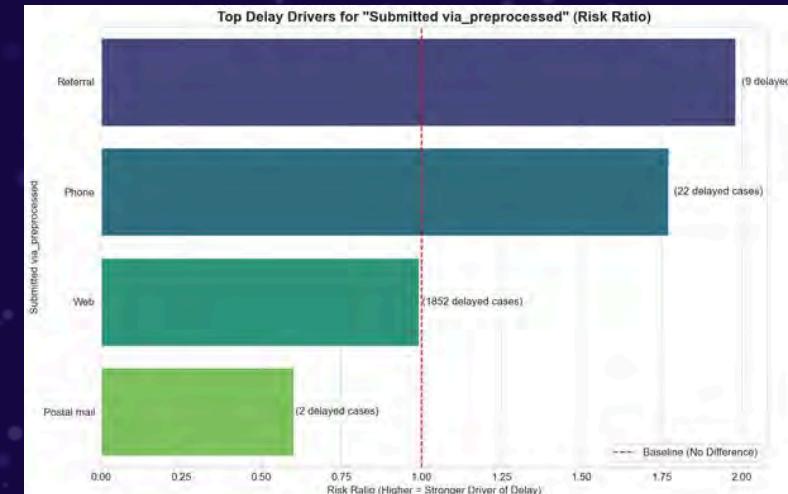
Titik yang berdekatan menunjukkan keluhan dengan makna atau topik yang serupa



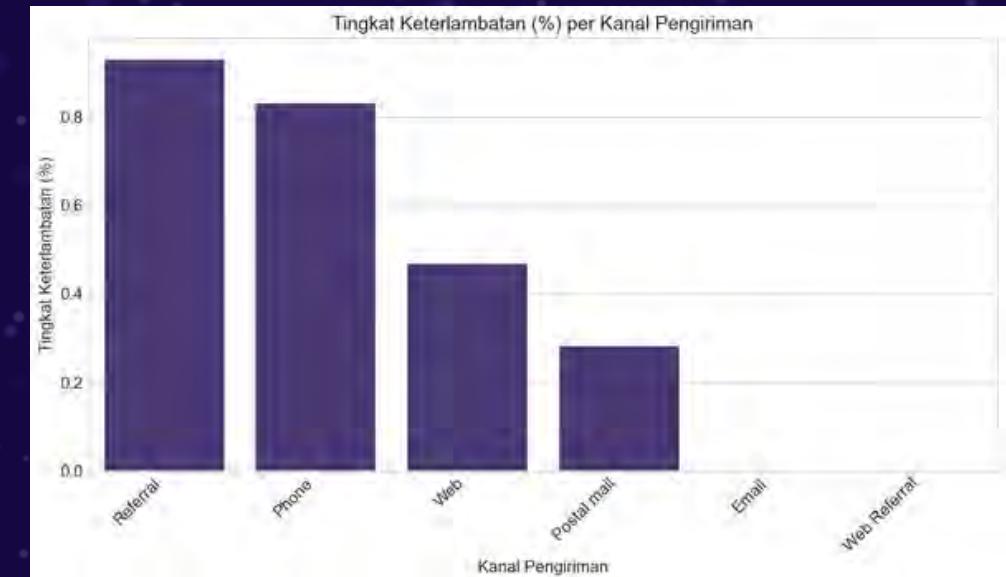
Scatter text ini memvisualisasikan distribusi sementik keluhan berdasarkan dimensi teks Terlihat bahwa titik oranye jumlahnya relatif sedikit dan tersebar di area yang sama dengan titik biru, menandakan bahwa secara sementik, keluhan tertunda tidak membentuk pola atau cluster khusus, melainkan bercampur dengan keluhan lain.



Saluran Referral dan Phone menunjukkan rasio risiko tertinggi sebagai pendorong keterlambatan, meskipun jumlah kasus tertunda relatif kecil. Sementara itu,



Tingkat keterlambatan tertinggi terdapat pada kanal Referral (~0,9%) dan Phone (~0,8%), menunjukkan potensi hambatan signifikan pada jalur ini. Kanal Web memiliki tingkat keterlambatan moderat (~0,47%), sementara Postal mail relatif rendah (~0,27%).



3

DATA CLEANING & PREPROCESSING



quality data cleaning meningkatkan akurasi model prediktif hingga 30% dalam domain data besar.

NORMALISASI ENTITAS PERUSAHAAN

Normalisasi Nama Perusahaan dengan Fuzzy Matching (Levenshtein)

Variasi kecil nama perusahaan → diperlakukan sebagai entitas berbeda oleh model.



Sinyal prediktif lemah → model tidak bisa belajar pola perusahaan dengan benar



Fuzzy String Matching (Levenshtein Distance) → mengukur jumlah minimum operasi edit (hapus, tambah, ganti karakter) untuk menyamakan dua string.



1

Hapus sufiks umum (INC, LLC, Bank, Corp)

2

Ubah ke huruf besar & trimming.

3

Hitung skor kemiripan (token_sort_ratio).

4

Jika skor > 85 → dipetakan ke nama kanonikal.

sebelum normalisasi:
2346

Convergent Resources, Inc.

setelah normalisasi
2344



Convergent Resources

4

FEATURE ENGINEERING

01

Variabel Target

is_delayed

- 1 = Respons Terlambat
- 0 = Respons Tepat Waktu

Mendefinisikan problem
supervised learning.



- Distribusi target tidak seimbang (<1% Terlambat) → pentingnya FE untuk memperkuat sinyal.
- FE berbasis teks + waktu sering meningkatkan recall model dalam mendekripsi kasus minoritas.

<i>Narrative length</i>	<i>has_narrative</i>	<i>Day</i>	<i>Month</i>	<i>is_delayed</i>
245	1	0	7	0

02

Fitur Teks Sederhana

- *narrative_length* → Panjang narasi keluhan.
 - Hipotesis: keluhan panjang → lebih kompleks → lebih berisiko terlambat.
- *has_narrative* → Ada / tidaknya narasi.
 - Narasi sering memuat detail penting untuk prediksi.

03

Testing and Deployment

- *day_of_week_received* → Hari keluhan diterima.
 - *month_received* → Bulan keluhan diterima.
- Menangkap pola temporal (misal weekend lebih lambat).



04

Fitur Gabungan Teks

combined_text = *product* + *sub_product* + *issue* + *sub_issue* + *narrative*.

Memberikan representasi teks yang lebih kaya untuk pemodelan (input TF-IDF).



5

MODELLING

Pipeline Otomatis: Konsistensi, Reproducibility, Anti-Leakage

Pipeline = Tulang Punggung Workflow ML

01

Konsistensi preprocessing
antara train & test.

02

Mencegah data leakage
(info test data bocor ke train)

03

Alur kerja dapat direproduksi
& siap untuk produksi.

TF-IDF Vectorizer
(Text Features)

01

- Ubah narasi teks → representasi numerik.
- Fokus pada kata informasi kunci (TF = frekuensi, IDF = bobot kata langka).

StandardScaler
(Numerical Features)

02

- Skala numerik (contoh: narrative_length) → mean=0, std=1.
- Penting untuk model alternatif (Regresi, SVM).
- Passthrough categorical feature tetap dipertahankan → ditangani CatBoost secara native.

Penanganan
Ketidakseimbangan
Kelas

03

- Data "Terlambat" jauh lebih sedikit → model cenderung hanya prediksi "Tepat Waktu".
- Solusi (auto_class_weights='Balanced')

CatBoostClassifier
(Final Model)

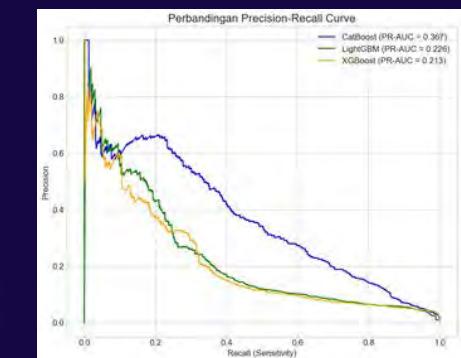
04

- Menangani fitur kategorikal secara langsung.
- Parameter: iterations=500, learning_rate=0.05, auto_class_weights='Balanced'.
- GPU enabled → training lebih cepat.



6

EVALUASI DAN PERBANDINGAN MODEL

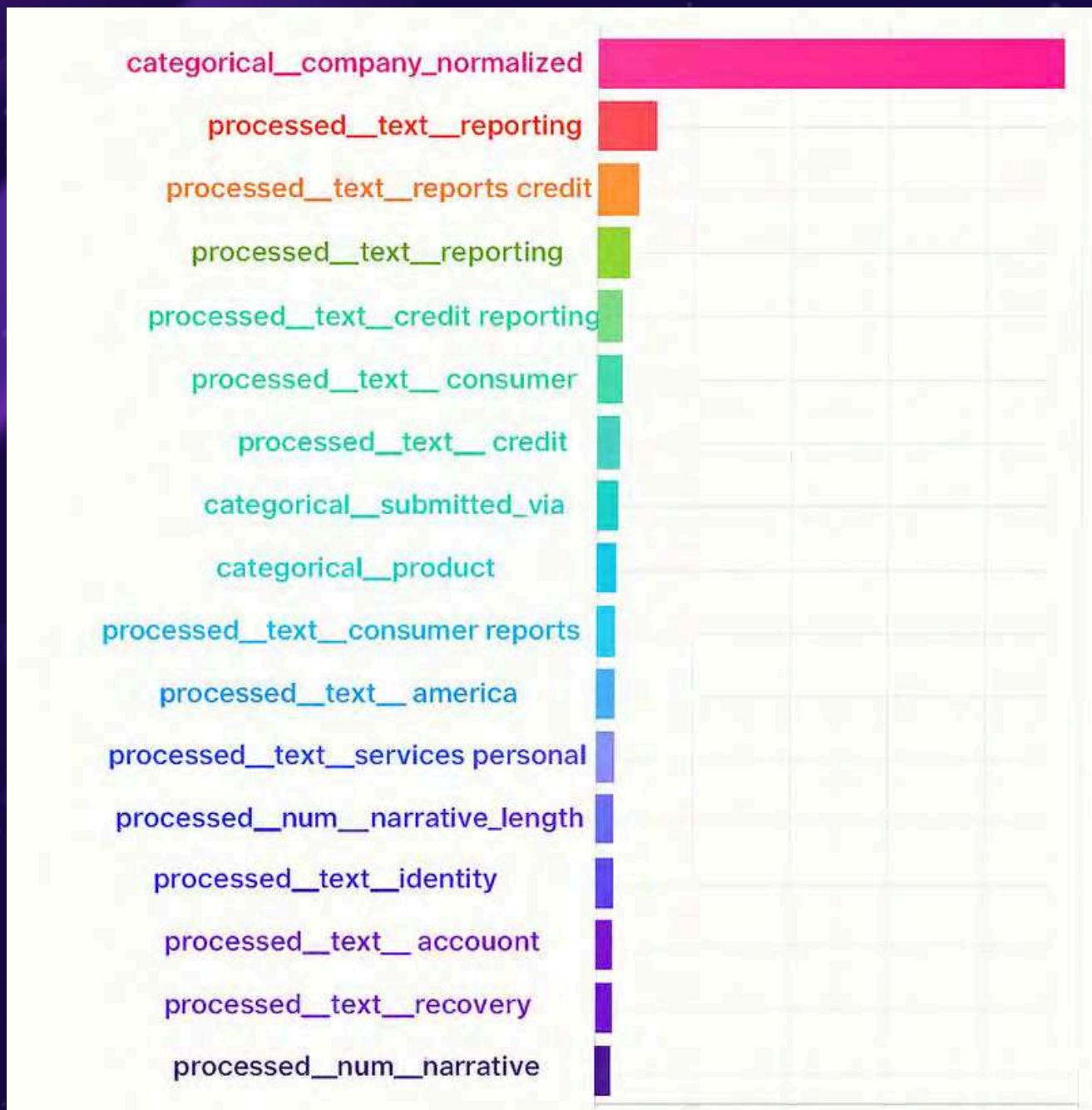


CatBoost
XGBoost
LightGBM

Metrik	CatBoost	LightGBM	XGBoost	Penjelasan
Recall (Terlambat)	0.85	0.9	0.9	CatBoost sedikit lebih rendah, namun masih sangat baik dalam menemukan sebagian besar keluhan yang terlambat.
Precision (Terlambat)	0.12	0.05	0.05	CatBoost jauh lebih unggul. Artinya, dari semua yang diprediksi terlambat, 12% di antaranya benar-benar terlambat, dibandingkan hanya 5% pada model lain. Ini mengurangi jumlah prediksi positif palsu secara signifikan.
F1-Score (Terlambat)	0.21	0.1	0.1	CatBoost memiliki F1-score dua kali lebih tinggi , menunjukkan keseimbangan terbaik antara Precision dan Recall.
PR-AUC	351	91	213	Nilai PR-AUC CatBoost adalah yang tertinggi , yang secara visual menunjukkan kinerja superior dalam menangani kelas minoritas (terlambat).

7

INTERPRETASI SHAP



SHAP (SHapley Additive exPlanations) →
metode interpretasi berbasis teori permainan.

1

Perusahaan Penanggung Jawab

- Perusahaan yang menangani keluhan adalah prediktor paling dominan. Ini menunjukkan adanya perbedaan kinerja yang signifikan antar perusahaan.

2

Kata Kunci "Pelaporan Kredit"

- Jenis keluhan ini secara konsisten menjadi pendorong keterlambatan, kemungkinan karena tingkat kompleksitasnya yang lebih tinggi.

3

Hari Pengiriman Keluhan

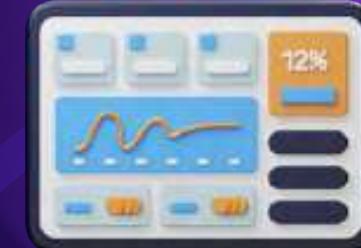
- Hari saat keluhan diterima ikut memengaruhi prediksi, mengindikasikan adanya potensi masalah alur kerja mingguan atau penumpukan pekerjaan.



Berdasarkan analisis SHAP, faktor paling krusial yang mendorong keterlambatan respons adalah **perusahaan yang menangani keluhan tersebut**, karena dampaknya jauh lebih signifikan dibandingkan faktor lainnya. Oleh karena itu, upaya perbaikan akan memberikan hasil paling efektif jika difokuskan pada evaluasi kinerja perusahaan-perusahaan yang berisiko tinggi.



Membuka Suara Pelanggan: Dasbor Strategic Insights Engine



BERTOPIC UNTUK ANALISIS TOPIK & SENTIMEN

"Dari Kebisingan Menjadi Sinyal: BERTopic & Sentiment Analysis"

- **BERTopic** menemukan tema-tema tersembunyi yang tidak terdeteksi oleh kategori produk standar.
- Lebih unggul dibanding **LDA** pada teks pendek, tidak terstruktur, & bahasa kompleks (contoh: Bahasa Indonesia).

Klasifikasi tiap keluhan & topik:
Positif / Netral / Negatif

“Sangat Membantu”

“Aplikasi Mobile Lambat”

“KPR Tidak Jelas”

01

Embedding
(BERT/IndoBERT)

teks keluhan

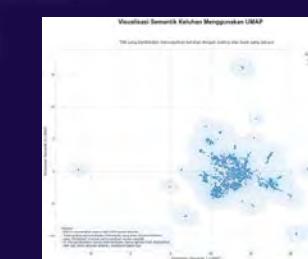


vektor semantik

02

UMAP

reduksi dimensi untuk mempermudah visualisasi & clustering.



03

HDBSCAN

klasterisasi keluhan berdasarkan makna, bukan sekadar kata kunci.

04

c-TF-IDF

ekstraksi kata representatif tiap topik.

O PUSAT KOMANDO: DASBOR INTERAKTIF

Menyediakan single source of truth bagi seluruh stakeholder untuk pemantauan risiko, tren, dan insight operasional.



Tinjauan Eksekutif

- KPI: Total Keluhan, Skor Risiko Rata-rata, SLA Compliance, Top 5 Isu Terbesar.
- Memberi gambaran cepat untuk C-Level dan regulator.



Screenshot of a data visualization tool interface showing a table titled "Tabel Antrian Prioritas".

The table displays the following data:

Complaint ID	Product	Company	Recommended Action	Priority Level
12782397	Credit reporting or other person...	INDEPENDENT BANK GROUP, INC.	Eskalasikan ke agen senior; target respons < 24 jam.	0.9282 Rendah
12998243	Student loan	EdFinancial Services	Eskalasikan ke agen senior; target respons < 24 jam.	0.9945 Sedang
12453082	Student loan	Servicer under contract with Federal Student Aid	Eskalasikan ke agen senior; target respons < 24 jam.	0.9943 Sedang
12248799	Student loan	Servicer under contract with Federal Student Aid	Eskalasikan ke agen senior; target respons < 24 jam.	0.9900 Tinggi
12463303	Credit reporting or other person...	Servicer under contract with Federal Student Aid	Eskalasikan ke agen senior; target respons < 24 jam.	0.9994 Tinggi
7784262	Credit reporting or other person...	ADIR INTERNATIONAL LLC	Eskalasikan ke agen senior; target respons < 24 jam.	0.9985 Rendah
13071941	Credit reporting or other person...	EdFinancial Services	Eskalasikan ke agen senior; target respons < 24 jam.	0.9986 Rendah
13363641	Credit reporting or other person...	EdFinancial Services	Eskalasikan ke agen senior; target respons < 24 jam.	0.9986 Rendah
13025731	Credit reporting or other person...	Servicer under contract with Federal Student Aid	Eskalasikan ke agen senior; target respons < 24 jam.	0.9984 Rendah
8107659	Credit reporting or other person...	Mobilloans, LLC	Eskalasikan ke agen senior; target respons < 24 jam.	0.9979 Rendah
7153766	Credit reporting, cred...	Porania, LLC	Eskalasikan ke agen senior; target respons < 24 jam.	0.9989 Rendah
6876424	Credit reporting, cred...	Palm Beach Credit Adjustors, Inc	Eskalasikan ke agen senior; target respons < 24 jam.	0.9986 Rendah

O PUSAT KOMANDO: DASBOR INTERAKTIF

Menyediakan single source of truth bagi seluruh stakeholder untuk pemantauan risiko, tren, dan insight operasional.

01

Explorer Topik & Sentimen

- Interaktif
- Memberi konteks di balik keluhan

02

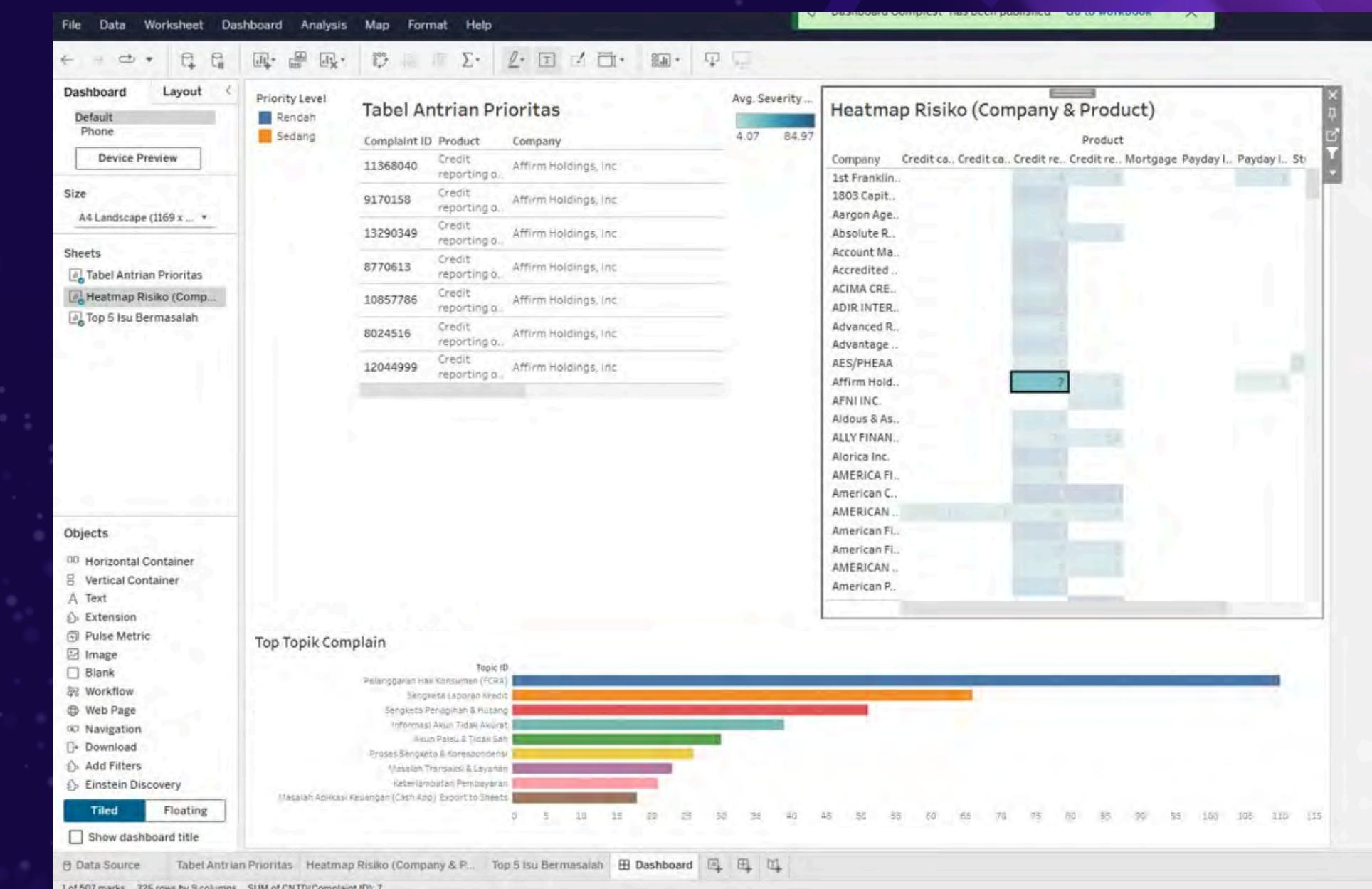
Matriks Risiko (Heatmap)

- Menghubungkan Produk x Topik.

03

Tabel Antrian Prioritas

- *escalation path* ke agen senior (<24 jam).



MLOps, Integrasi, dan Peningkatan Berkelanjutan

01

Pipeline MLOps Otomatis

CI/CD/CT Pipeline ➔
otomatis validasi,
retraining, evaluasi

Versioning ➔ data, kode,
parameter, & model
terdokumentasi
➔ mudah diaudit &
direproduksi.

02

Pemantauan Drift Proaktif

Data Drift ➔

uji statistik
(Kolmogorov-Smirnov)

Concept Drift ➔

monitor performa (AUC, akurasi)
pada data terbaru.

Prediction Drift ➔

deteksi perubahan distribusi skor
risiko. Model tetap relevan

03

Champion Challenger Framework

Champion
model produksi
saat ini.



Challenger

model baru → diuji dalam
shadow mode.



Model dipromosikan hanya jika
performa signifikan lebih baik →
aman & berbasis data.

Analisis Bisnis & Proyeksi ROI

EFISIENSI OPERASIONAL

- ⌚ Pengurangan AHT: -20%
→ hemat \$1.28M/tahun.
- ⟳ Pengurangan Eskalasi
→ routing pintar, beban agen turun.
- 👤 Perputaran Agen ↓ 25%
→ hemat \$375K/tahun.
- 🤖 Defleksi Panggilan 20% →
hemat \$800K/tahun.

DAMPAK PENDAPATAN

- 🔒 Churn ↓ 10%
→ retensi tambahan 3.750 pelanggan/tahun.
💰 Peningkatan CLV
→ tambahan revenue \$5.6M/tahun.

Biaya **\$6–\$12** per interaksi mencerminkan biaya penuh (**fully loaded cost**) yang mencakup gaji agen, overhead operasional, manajemen & dukungan, serta biaya perekrutan dan pelatihan akibat tingginya turnover.

Proyeksi Dampak Finansial & ROI (Outlook 3 Tahun)

Metrik	Tahun 1	Tahun 2	Tahun 3	Total 3 Tahun	Catatan & Asumsi
PENGHEMATAN BIAYA					Berdasarkan 500.000 keluhan/tahun, biaya interaksi agen \$8.
Pengurangan AHT (20%)	\$1.280.000	\$1.280.000	\$1.280.000	\$3.840.000	Penghematan 2 menit/interaksi pada 25% keluhan kompleks.
Pengurangan Perputaran Agen (25%)	\$375.00	\$375.00	\$375.00	\$1.125.000	Menghindari biaya penggantian 25 agen @\$15.000/agen.
Defleksi Panggilan (20%)	\$800.00	\$800.00	\$800.00	\$2.400.000	100.000 interaksi dialihkan ke swalayan (biaya \$0).
Total Penghematan Biaya	\$2.455.000	\$2.455.000	\$2.455.000	\$7.365.000	
PENINGKATAN PENDAPATAN					Berdasarkan tingkat churn 15% & CLV \$1.500.
Pengurangan Churn (10%)	\$5.625.000	\$5.625.000	\$5.625.000	\$16.875.000	Mempertahankan 3.750 pelanggan tambahan per tahun.
Total Peningkatan Pendapatan	\$5.625.000	\$5.625.000	\$5.625.000	\$16.875.000	
MANFAAT FINANSIAL					
Total Manfaat (Penghematan + Pendapatan)	\$8.080.000	\$8.080.000	\$8.080.000	\$24.240.000	
INVESTASI					
Biaya Implementasi (Platform, SDM)	\$2.500.000	\$0	\$0	\$2.500.000	Biaya satu kali untuk pengembangan & lisensi awal.
Biaya Operasional (Lisensi, Infrastruktur)	\$1.200.000	\$1.200.000	\$1.200.000	\$3.600.000	Biaya tahunan untuk pemeliharaan & lisensi.
Total Investasi	\$3.700.000	\$1.200.000	\$1.200.000	\$6.100.000	
ANALISIS ROI					
Arus Kas Bersih	\$4.380.000	\$6.880.000	\$6.880.000	\$18.140.000	
Nilai Sekarang Bersih (NPV) @ 10%				\$14.538.918	
Tingkat Pengembalian Investasi (ROI)				212%	
Periode Pengembalian (Payback Period)				< 6 Bulan	

Kesimpulan dan Rekomendasi

TRANSFORMASI STRATEGIS

1 Proyek ini mengubah manajemen keluhan dari pusat biaya yang reaktif menjadi mesin proaktif yang cerdas untuk mitigasi risiko, pengurangan biaya, dan peningkatan retensi pelanggan.

REKOMENDASI

2 Rekomendasikan untuk memulai dengan **program percontohan** selama 3 bulan pada lini produk kartu kredit untuk memvalidasi ROI yang diproyeksikan.

LANGKAH BERIKUTNYA

3 Proyek ini akan mengimplementasikan ekosistem **AI Hybrid** yang terdiri dari Intelligent Triage System (**IndoBERT + CatBoost**) untuk prediksi risiko real-time yang dapat diinterpretasikan dengan **SHAP**, dan Strategic Insights Engine (**BERTopic**) untuk analisis akar masalah.



TEAM DAWNBRINGER

THANK YOU!

We appreciate your time and attention.

Mikael Alexander Adikara Purnama

Raphael Angelo Adikara Purnama

Timotius Christopher Patrick Sihombing