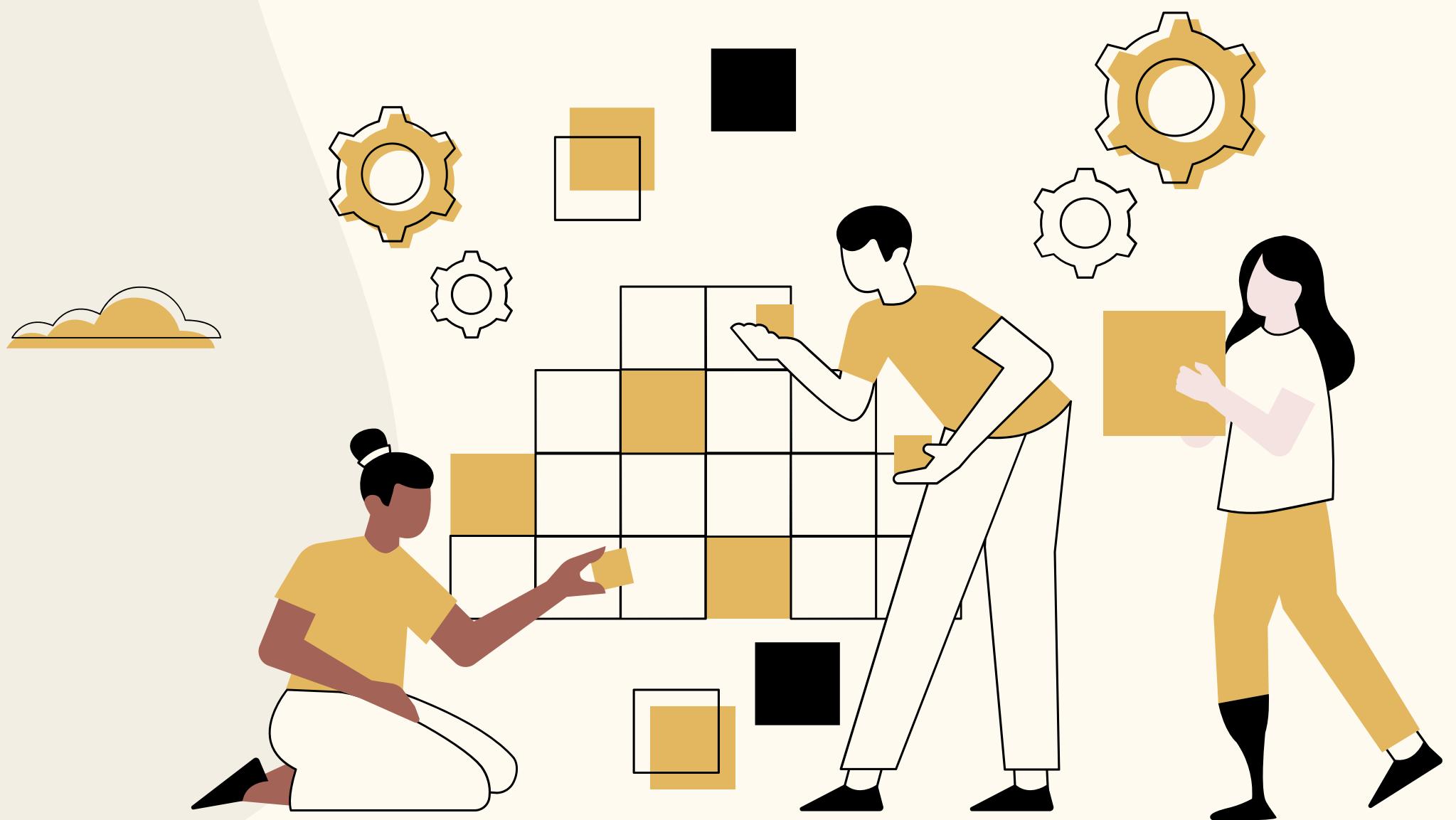


PROBLEM STATEMENT: 3

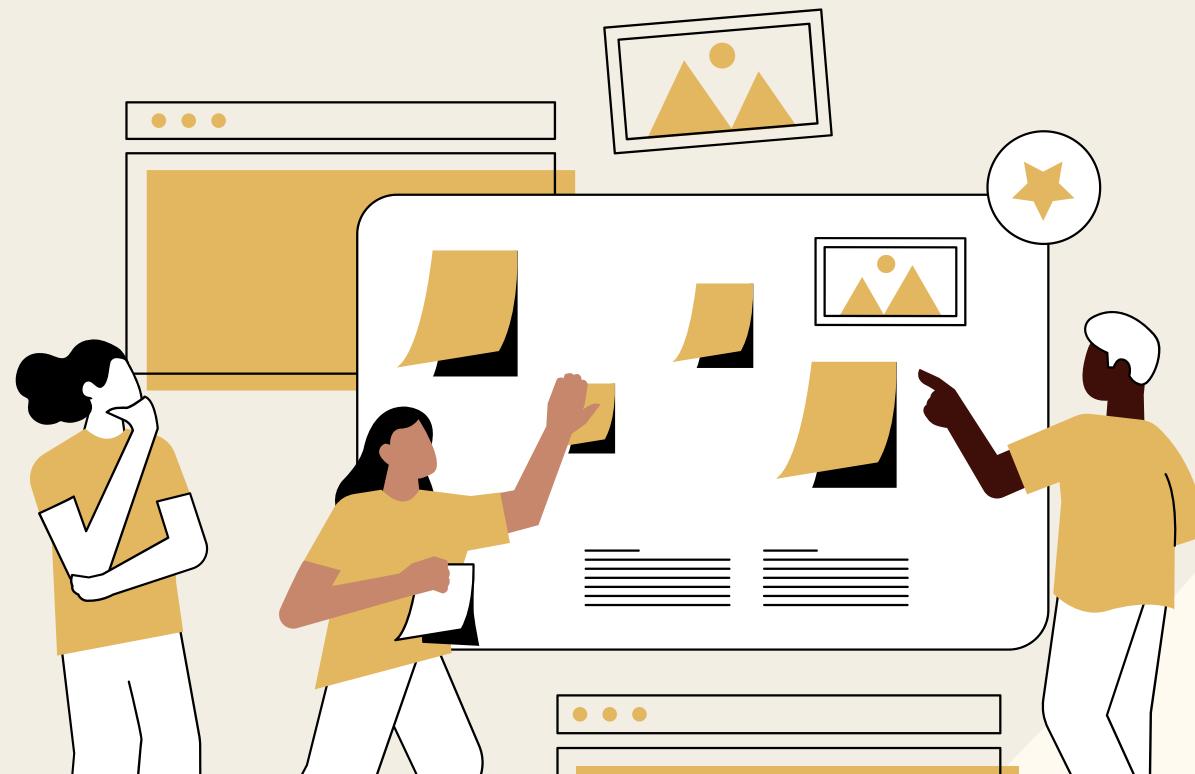
# CREDSENTINEL

Created by  
CasualCraft



# The Problem

Build a proactive system to identify customers at risk of delinquency before missed payments occur, enabling early intervention to reduce credit losses, lower collection costs, and protect long-term customer relationships.



## Late Detection:

Financial stress starts weeks before default but banks detect it too late.

## High Collections Costs:

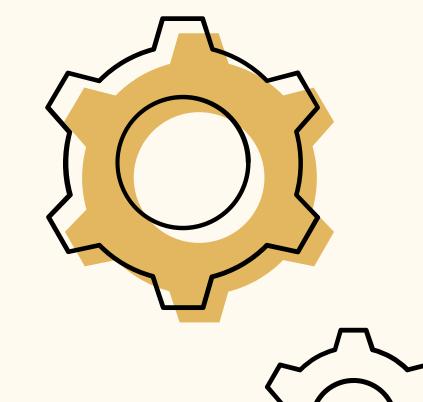
Aggressive late interventions damage trust and long term lifetime values.

## Customer Harm:

Collections consume 15-20% of recovered amounts and erode margins.

## Scattered Signals:

Early warning signals exist across channels but are siloed and under-used.



# The Solution

We built a Pre-Delinquency Intervention Engine that monitors behaviour in real time, produces explainable risk scores and recommends precise, timely interventions so banks can support customers before missed payments occur.



## Real-time behavioural monitoring

Continuously ingest transaction, account and engagement events to detect anomalies and trends as they emerge.

## Early intervention recommendations

Actionable treatments (SMS, offers, hardship routing) prioritised by expected cure probability and customer impact.

## Explainable risk scoring

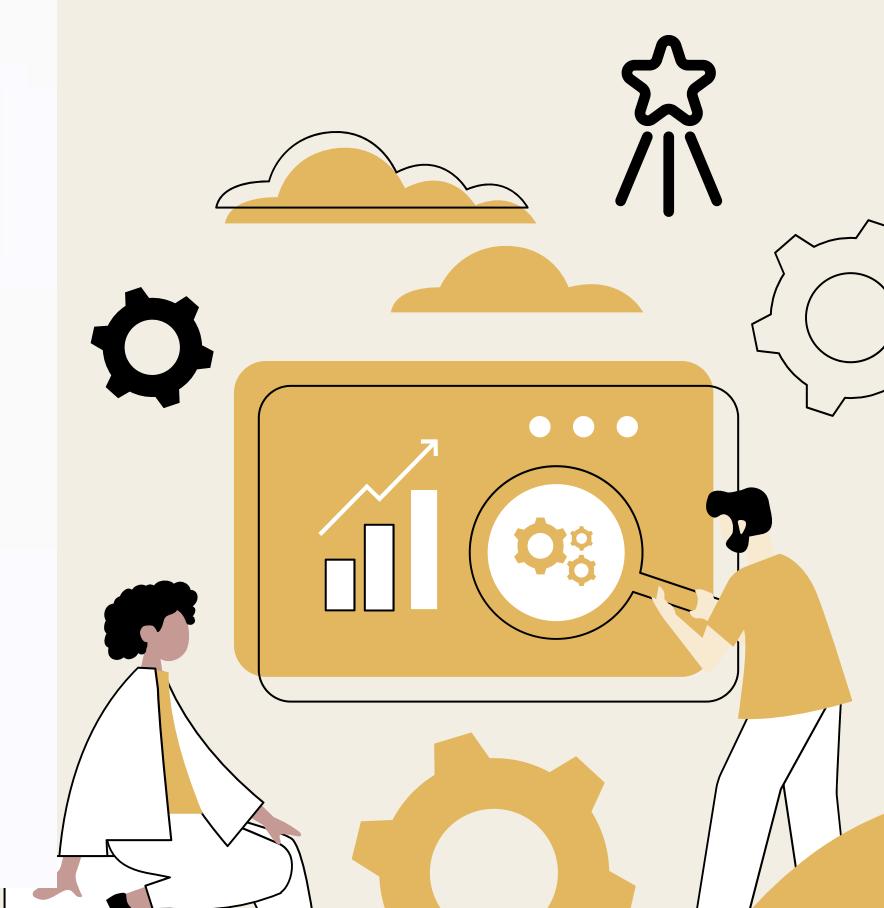
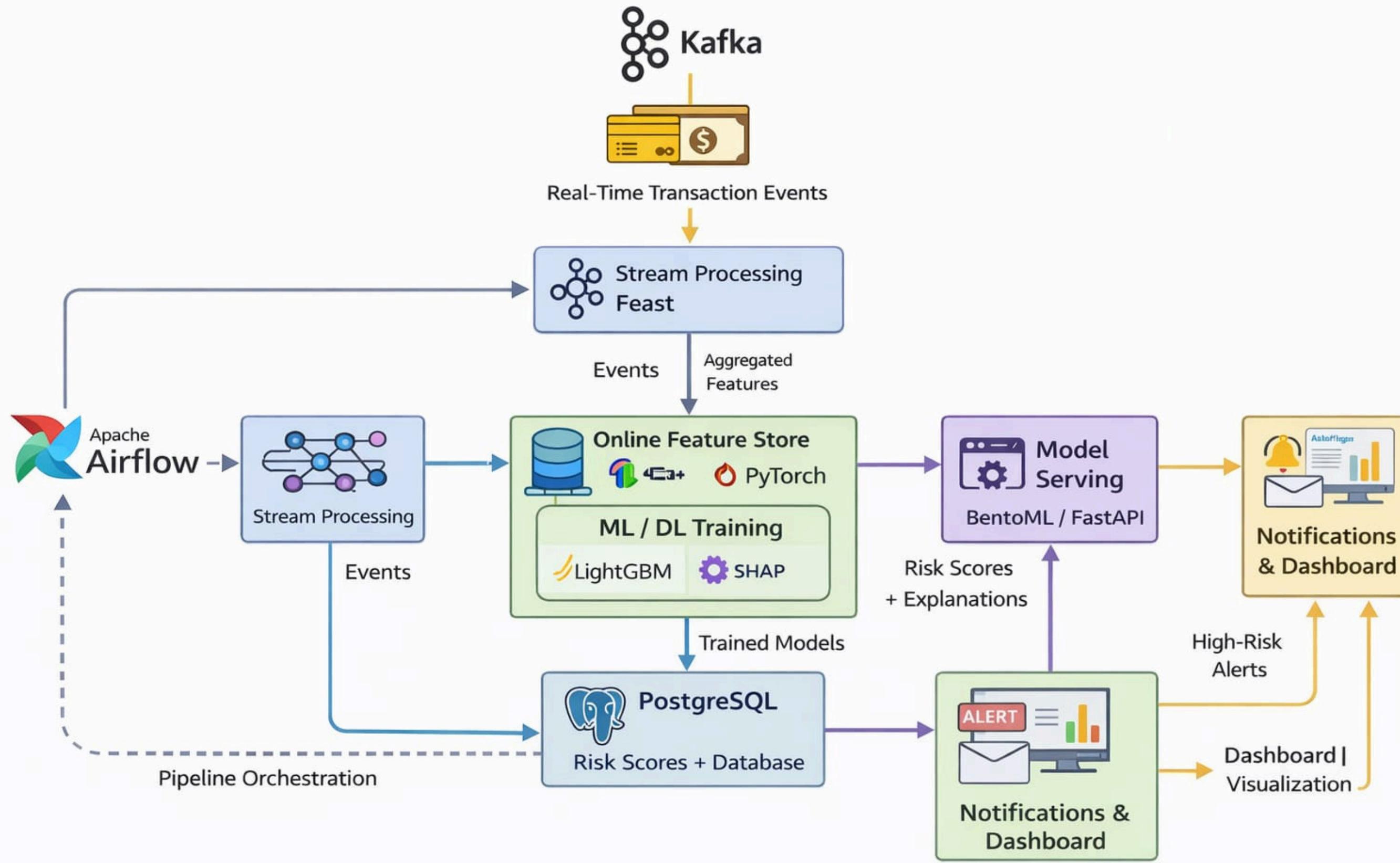
Interpretable scores with SHAP attribution so analysts and regulators can understand why a customer is flagged.

# ABSTRACT

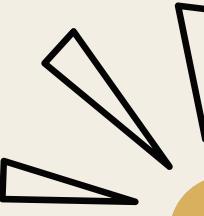
Banks often detect customer delinquency only after payments are missed, leading to higher financial losses and strained customer relationships. This project proposes a proactive, AI-driven system that predicts delinquency risk by analyzing real-time behavioral and transaction data. Using Apache Kafka for event streaming and Feast as a centralized feature store, the system continuously generates behavioral features that capture early signs of financial stress. Machine learning and deep learning models, including XGBoost, LightGBM, and PyTorch, evaluate these signals to produce predictive risk scores, while SHAP ensures explainable and transparent decision-making.

The trained models are deployed through BentoML and FastAPI to enable scalable real-time inference, with risk insights stored in PostgreSQL for visualization and automated alerts. Apache Airflow orchestrates the end-to-end pipeline, ensuring efficient data flow and workflow management. By shifting from reactive collections to proactive engagement, the solution helps financial institutions reduce credit losses, improve operational efficiency, and maintain stronger long-term customer relationships.

# System Architecture



# SYSTEM ARCHITECTURE OVERVIEW



We propose a real-time, privacy-preserving AI system that identifies customers at risk of delinquency before missed payments occur, enabling early and responsible intervention.

**How It Works :**

Kafka ingests real-time transaction and repayment events. Stream Processing + Feast generate behavioral features (EMI stress, spend volatility, utilization trends). ML (LightGBM) + Deep Learning (LSTM/GRU) detect early financial stress patterns.

Differential Privacy (DP-SGD) ensures customer-level data protection during model training.

SHAP Explainability provides transparent risk drivers for every prediction.

FastAPI/BentoML serves real-time risk scores.

PostgreSQL stores risk history for monitoring and audit.

Notification & Intervention Engine delivers early alerts and personalized EMI options.

Dash Dashboard enables risk monitoring and operational visibility.

Airflow automates retraining and pipeline orchestration.

# Tech Stack

Layer	Tech
Streaming	Apache Kafka
Features	Feast
ML	XGBoost, LightGBM
DL	PyTorch (LSTM/GRU)
Privacy	Opacus (DP-SGD)
Explainability	SHAP
Serving	BentoML + FastAPI
Database	PostgreSQL
Notifications	Email / SMS Stub
Dashboard	Plotly + Dash
Orchestration	Apache Airflow

## Apache Kafka

Real-time ingestion of transaction and behavioral events.

## Feast

Centralized feature management for consistent training and inference

## XGBoost, LightGBM

Tabular models for customer risk prediction.

## PyTorch (LSTM / GRU)

Captures temporal financial behavior patterns.

## Opacus (DP-SGD)

Privacy-preserving model training.

## SHAP

Provides interpretable risk explanations.

## BentoML + FastAPI

Deploys models as scalable real-time APIs.

## PostgreSQL

Stores risk scores, predictions, and historical data.

## Plotly + Dash

Visualization of risk insights and analytics.

## Apache Airflow

Workflow scheduling and pipeline automation.

# Feature Engineering

\$15000

## Reduced discretionary spending

Decline in discretionary categories as customers prioritise essential payments.

## Late utility payments

Shifts in bill payment dates and partial payments indicate cash flow strain.

## Savings balance drop

Rapid depletion of buffers flagged against seasonal baselines and replacement rates.

## ATM withdrawal spikes

High cash outflows or atypical withdrawal patterns suggest liquidity stress.

## Salary delay detection

Detect missed or irregular salary credits relative to employer patterns and payroll cycles.

# NOVELTY

## Proactive Instead of Reactive

Unlike traditional systems that act after missed payments, our solution predicts financial stress weeks in advance, enabling early and supportive intervention.

## Behavioral Intelligence over Static Credit Scoring

The system analyzes dynamic behavioral patterns such as salary timing, spending trends, and savings changes

## Real-Time Streaming Architecture

Integration of Kafka with a Feature Store enables continuous monitoring and live risk assessment instead of batch-based analysis.

## Explainable AI for Trust & Compliance

SHAP-based explanations provide transparent reasoning behind risk predictions, ensuring regulatory alignment and human-understandable decisions

## Privacy-Preserving Model Design

Use of differential privacy concepts ensures sensitive financial data remains protected while maintaining predictive accuracy.

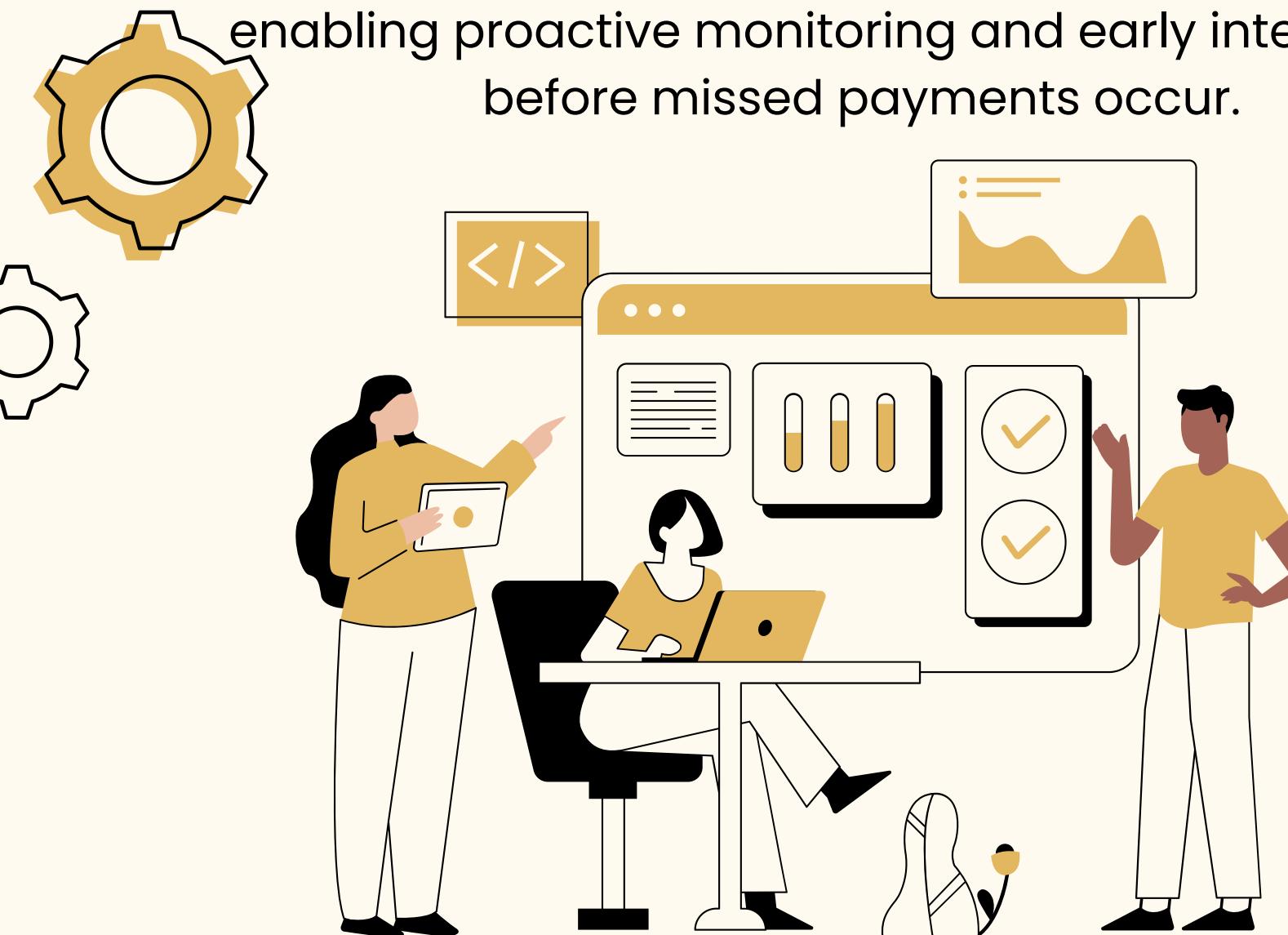
## Production-Ready Serving Layer

Deployment using BentoML and FastAPI bridges machine learning models with real-world banking workflows, enabling scalable and actionable decision-making.

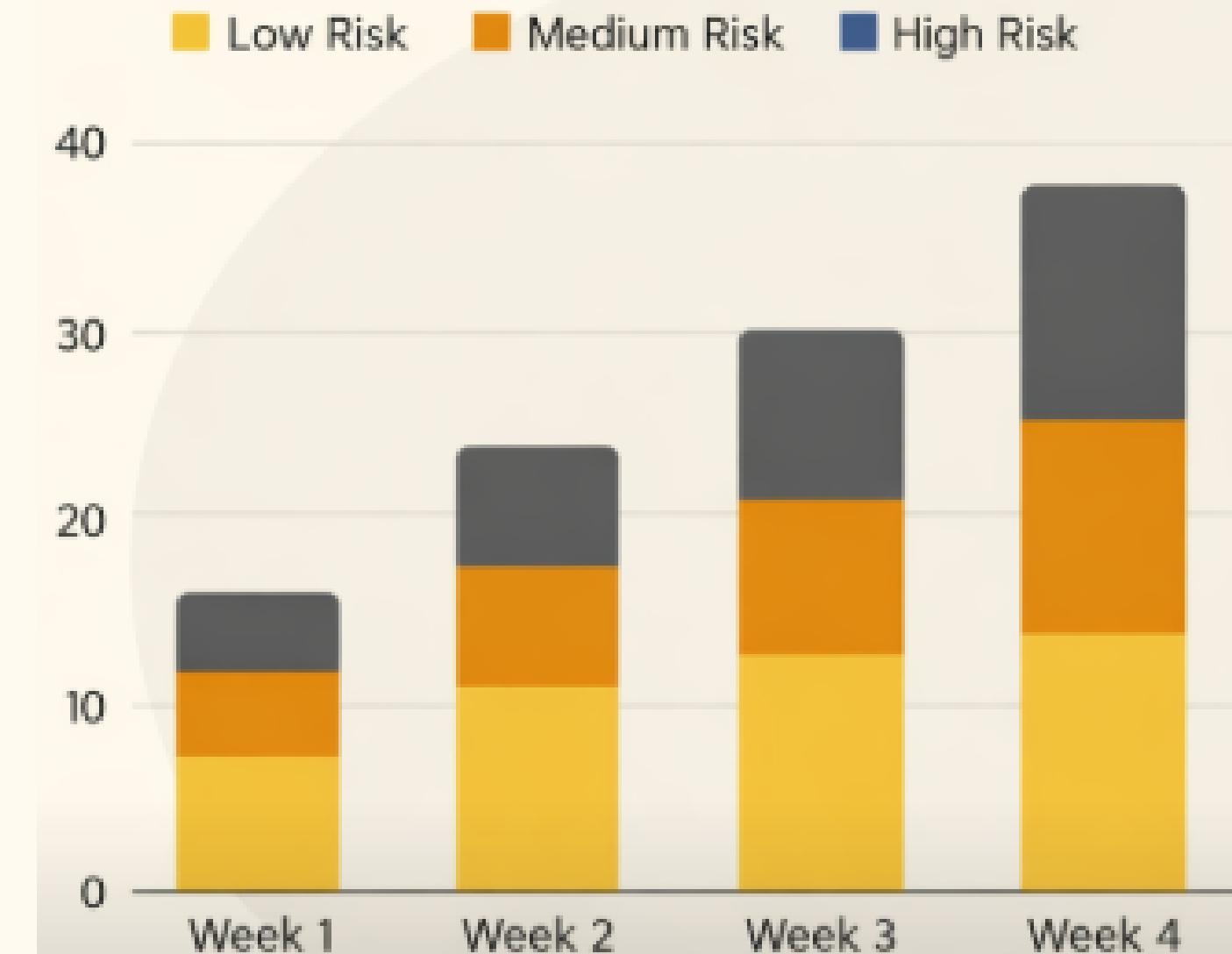


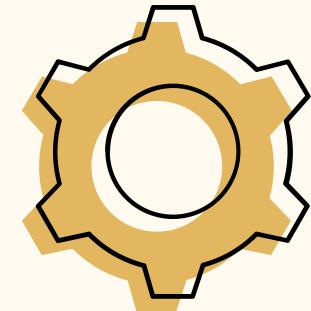
# AI-Driven Customer Risk Monitoring

This graph illustrates predicted customer delinquency risk levels over four weeks, showing a gradual increase in medium and high-risk cases, enabling proactive monitoring and early intervention before missed payments occur.

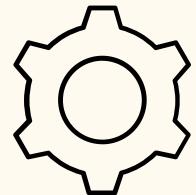


The graph below shows the predicted delinquency risk levels over the past four weeks.





# Model Serving Layer



## 01 API endpoints (FastAPI)

Expose low-latency prediction and explanation endpoints for downstream services and UI

## 02 Model packaging (BentoML)

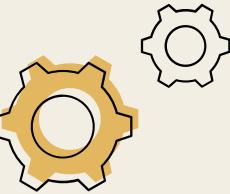
Containerised models with versioning, reproducible environments and easy rollbacks.

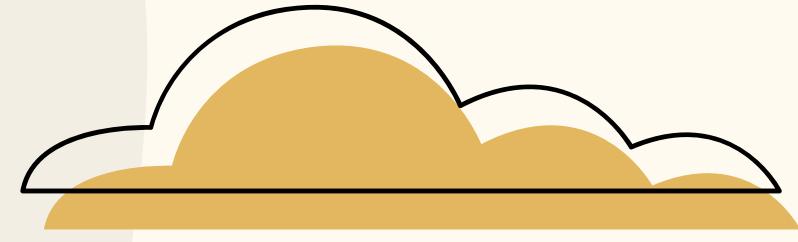
## 03 Feature fetch (Feast)

Real-time feature retrieval ensures consistent inputs between training and serving.

## 04 Inference + Explanation

Return risk score with SHAP breakdown; persist results to PostgreSQL and emit alerts to dashboard and orchestrator.





# Benefits to Banks

Our real-time credit-risk solution delivers significant advantages across three critical dimensions: financial performance, customer relationships, and operational efficiency. By proactively addressing risk, banks can achieve stronger outcomes and foster greater trust.

## 1. Financial Impact

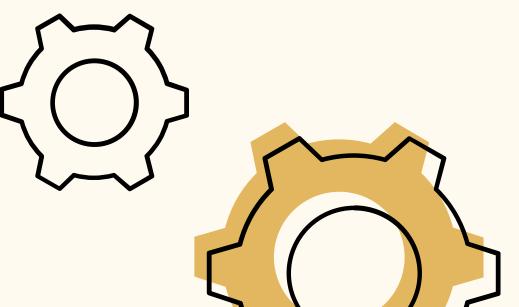
- Significantly reduced credit losses
- Lower collection and recovery costs

## 2. Customer Relationships

- Proactive support for customers before crises
- Cultivation of stronger, trust-based relationships

## 3. Operational Excellence

- Automated, data-driven decision support
- Reduced analyst fatigue and improved efficiency



# Innovation Highlights

## Real-time Streaming Architecture

Process vast volumes of data instantly, enabling immediate risk assessments and interventions.

## Privacy-Preserving Training

Leverage advanced techniques to train robust models while strictly adhering to data privacy regulations.

## Explainable AI (XAI)

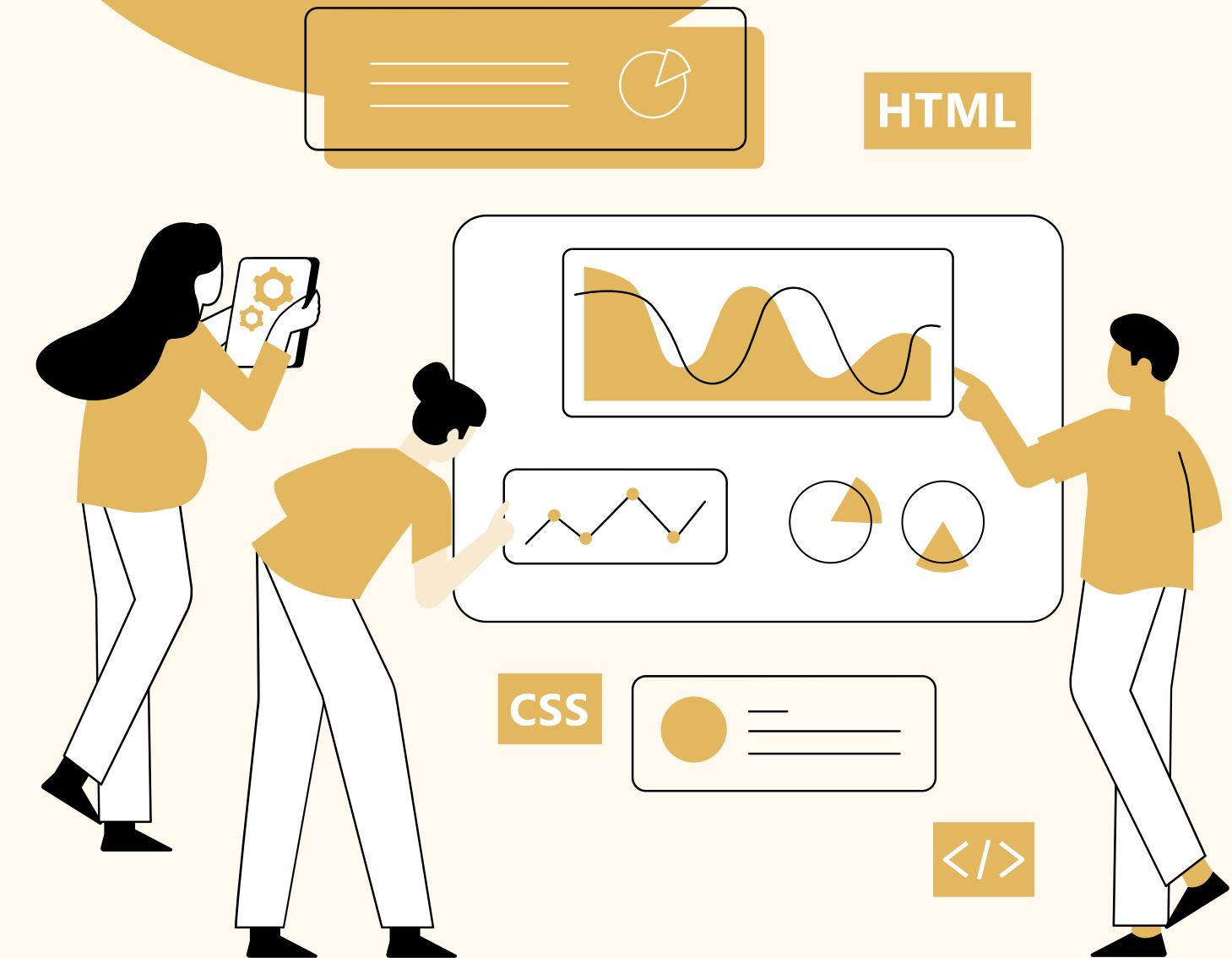
Gain transparent insights into model decisions, ensuring compliance and building trust with stakeholders.

## Feature Store Design

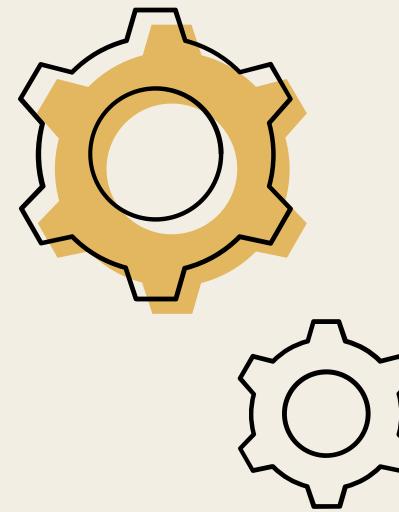
Centralise and manage features for consistent model training accelerating development cycles.

## Production-Ready Model Serving

Deploy and scale models efficiently with robust infrastructure designed for high-performance, real-world banking environments.



# Future Scope



## Multi-product Risk Correlation:

Understand interdependent risks across a customer's entire product portfolio.



## Reinforcement Learning Interventions:

Develop adaptive strategies for optimal customer engagement and risk mitigation.

## Real Bank Integration:

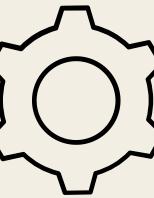
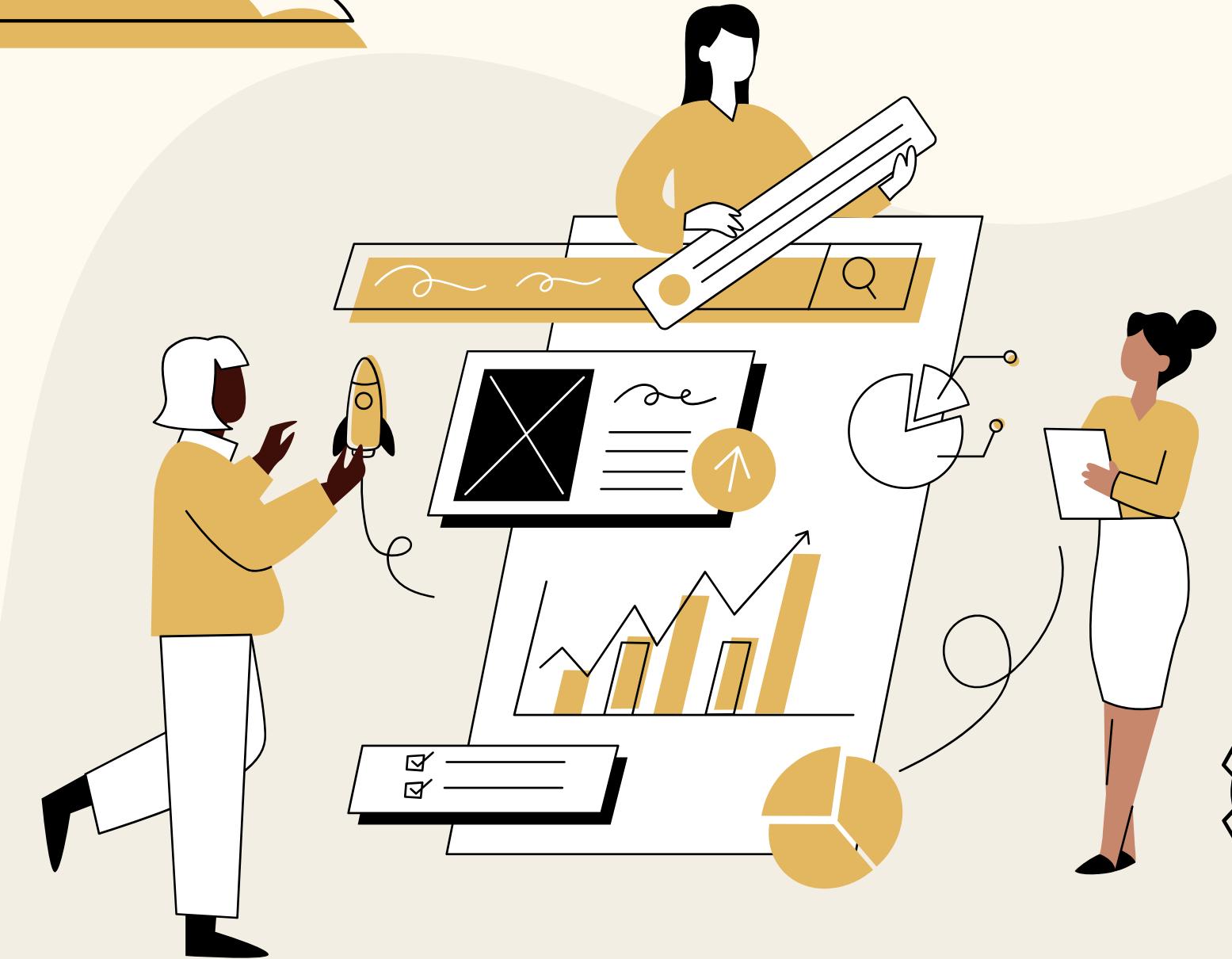
Seamlessly integrate with existing core banking systems for a holistic risk management framework.

## Graph-based Behavioral Modeling:

Uncover intricate patterns and relationships in customer behavior for enhanced predictive power.

# Thank You

Instead of reacting to defaults,  
our system helps banks act  
early, protecting customers  
and reducing risk through  
explainable, real-time  
intelligence.



- Team CasualCraft