Created by CasualCraft

SYSTEM ARCHITECTURE

CrediSentinel: Pre-Delinquency Risk Prediction Platform

SYSTEM ARCHITECTURE

High level overview

Flow:

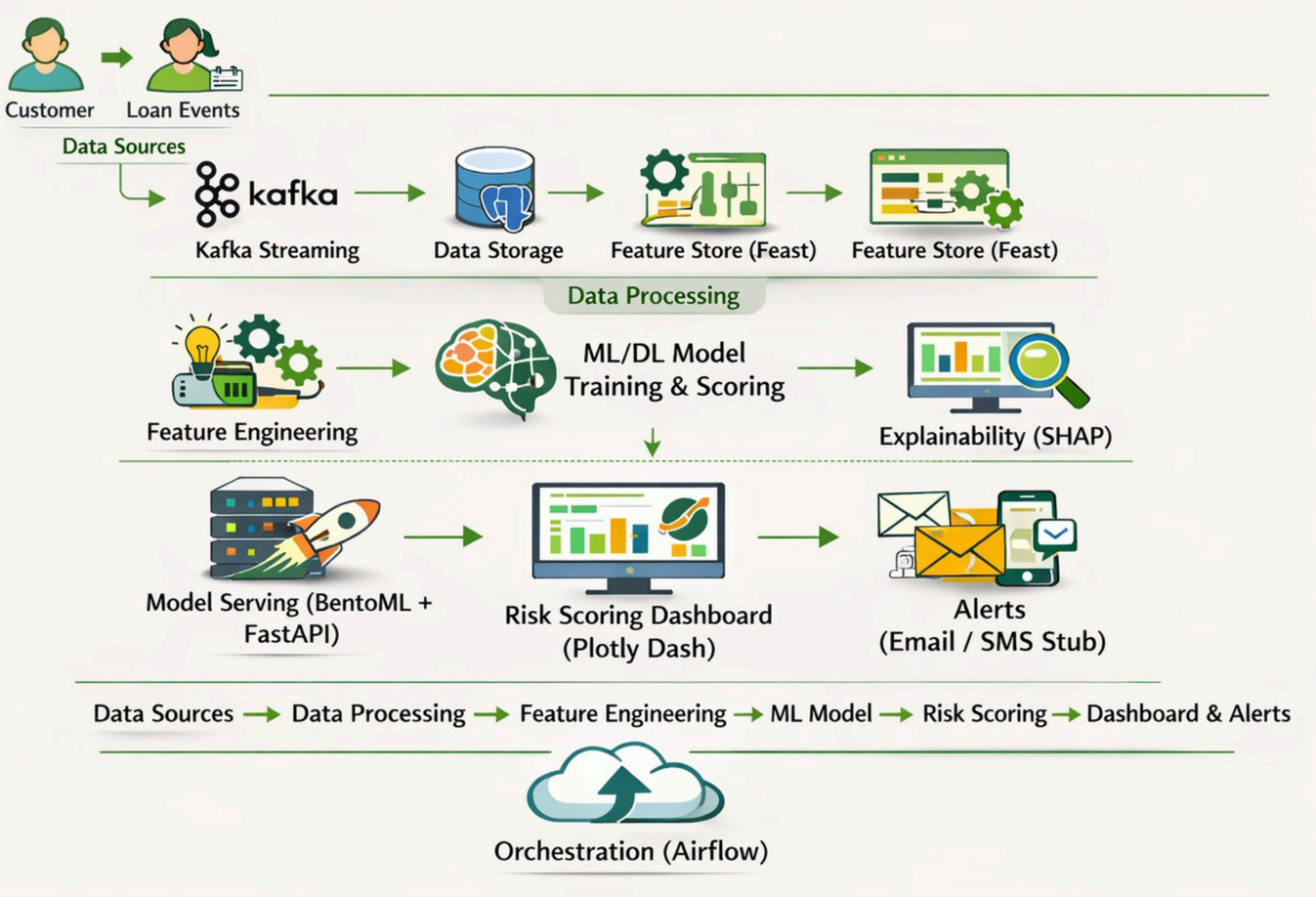
Data Sources → Data Processing → Feature Engineering → ML Model → Risk Scoring → Dashboard + Alerts

SYSTEM ARCHITECTURE

High level overview

Pipeline Flow:

- Customer + Loan Events (repayments, utilization, missed EMI)
- Kafka Streaming Layer
- Data Storage (PostgreSQL)
- Feature Store (Feast)
- ML/DL Model Training & Scoring
- Explainability (SHAP)
- Model Serving (BentoML + FastAPI)
- Dashboard (Plotly Dash)
- Alerts (Email/SMS Stub)
- Orchestration (Airflow)



STREAMING AND DATA LAYER

Streaming

- Apache Kafka
- Captures:
- Payment events
- Loan updates
- Transaction patterns

Database

PostgreSQL

Stores:

Raw customer/loan data

Historical repayment logs

Risk scoring outputs

Real-Time Data Processing & Storage

FEATURE ENGINEERING LAYER

The Feature Engineering Layer ensures consistent and reliable data preparation using a centralized feature store (Feast). Financial attributes such as income patterns, creditworthiness, spending behavior, loan exposure, and macro-risk factors are organized into structured feature groups. These engineered features are then used by both Machine Learning models (XGBoost, LightGBM) for structured data and Deep Learning models (LSTM/GRU) for sequential risk analysis. The processed data ultimately generates a probability score that classifies customers into Low, Medium, or High risk categories, supporting accurate and scalable risk prediction.

FEATURE STORE

- Feast
- Eature Groups Used
- Income & Employment
- Income_monthly, employment_type
- Income_volatility_6m
- Creditworthiness & Bureau
- Credit_score
- Bureau_inquiries_6m
- Utilization & Spending Behaviour
- Credit_card_utilization
- Spend_growth_6m, expense_volatility_6m
- Balance_slope_6m
- Loan Exposure & EMI Burden
- personal_loan_outstanding,
- home_loan_outstanding
- total_outstanding, total_emi
- emi_to_income_ratio
- Macro & Risk Context
- macro_stress, delinquency_stage

ML AND DL INTELLIGENCE LAYER

Risk Prediction Models

ML Models

XGBoost

LightGBM

Used for:

Structured/tabular banking data

High accuracy & fast inference

DL Models

PyTorch (LSTM/GRU)

Used for:

Sequential repayment patterns

Time-series risk behavior

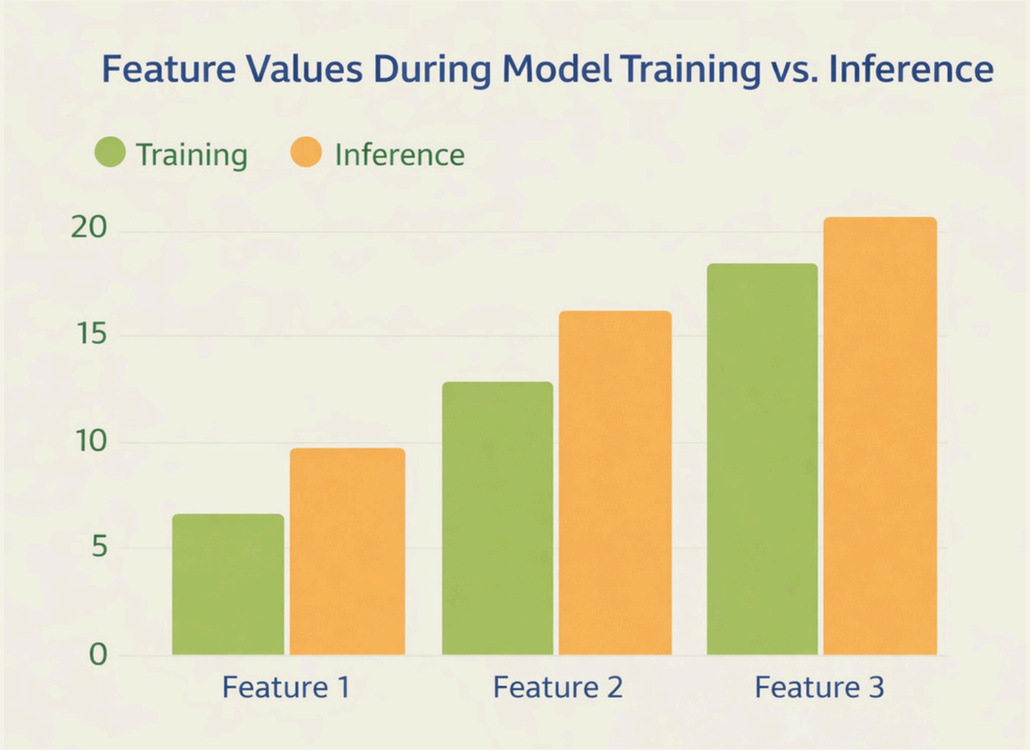
Output

Probability score (0–1)

Risk bucket: Low / Medium / High

FEATURE DISTRIBUTION

This bar chart compares key feature values used during model training and real-time inference. It helps validate that the feature distribution remains consistent across both stages, reducing the risk of data drift and ensuring the model performs reliably in production.



EXPLAINABILITY AND PRIVACY LAYER

- In banking, accuracy alone is not enough — predictions must be transparent, auditable, and privacy-safe
- SHAP (Explainable AI) is used to explain every risk score
- SHAP highlights the top drivers behind pre-delinquency predictions, such as:
- emi_to_income_ratio
- credit_card_utilization
- balance_slope_6m
- bureau_inquiries_6m
- This helps risk teams and collections teams trust model outputs and take informed action
- Opacus (Differential Privacy in PyTorch) is applied during training to:
- Prevent customer data memorization
- Reduce the risk of sensitive information leakage
- This layer makes CrediSentinel a Responsible AI system:
- Accurate
- Explainable
- Privacy-compliant / deployment-ready

Trust, Compliance, and Responsible AI

MODEL SERVING AND API LAYER

Serving

BentoML + FastAPI

API Outputs

Risk Score

Risk Band

SHAP Explanation

Suggested action label

This is enterprise-grade because :

- Scalable
- Fast inference
- Easy integration with bank systems

Production-Ready Model Deployment

DASHBOARD AND ALERTS LAYER

Dashboard

Plotly + Dash

Displays:

Portfolio risk summary

High-risk customers list

Risk trends over time

Customer drill-down

Notifications

Email / SMS Stub

Triggers:

High-risk alerts

Early warning alerts

Escalation alerts

The Dashboard and Alerts Layer visualizes portfolio risk insights using Plotly Dash and triggers email or SMS notifications for high-risk cases, early warnings, and escalation alerts to support proactive decision-making.

ORCHESTRATION LAYER

Automated ML Workflow Management

Orchestration

Apache Airflow

Scheduled Pipelines

Daily ingestion

Feature computation

Batch scoring

Weekly/monthly retraining

Monitoring & logging

Automates ML workflows using Airflow, managing pipelines, scoring, retraining, monitoring, and scheduled data processing.

THANK YOU

By Team - CasualCraft