LOANLENS

A MULTI-STAGE MACHINE LEARNING FRAMEWORK FOR CREDIT RISK ASSESSMENT AND PERSONALIZED FINANCIAL RECOMMENDATION

UCE2023502 – ADVIKA KHORGADE

UCE2023515 – ARPITA DEODIKAR

UCE2023532 – JUI KULKARNI

THE FINANCIAL CHALLENGE
LIMITATIONS OF CONVENTIONAL CREDIT SCORING

Lack of Unified Systems

- The Problem: Prior research focuses on isolated components (prediction or segmentation), lacking a single, unified decision workflow.
- LoanLens Solution: We use a sequential architecture where models feed directly into each other (e.g., K-Means label informs XGBoost), building a cohesive, end-to-end FinTech system.

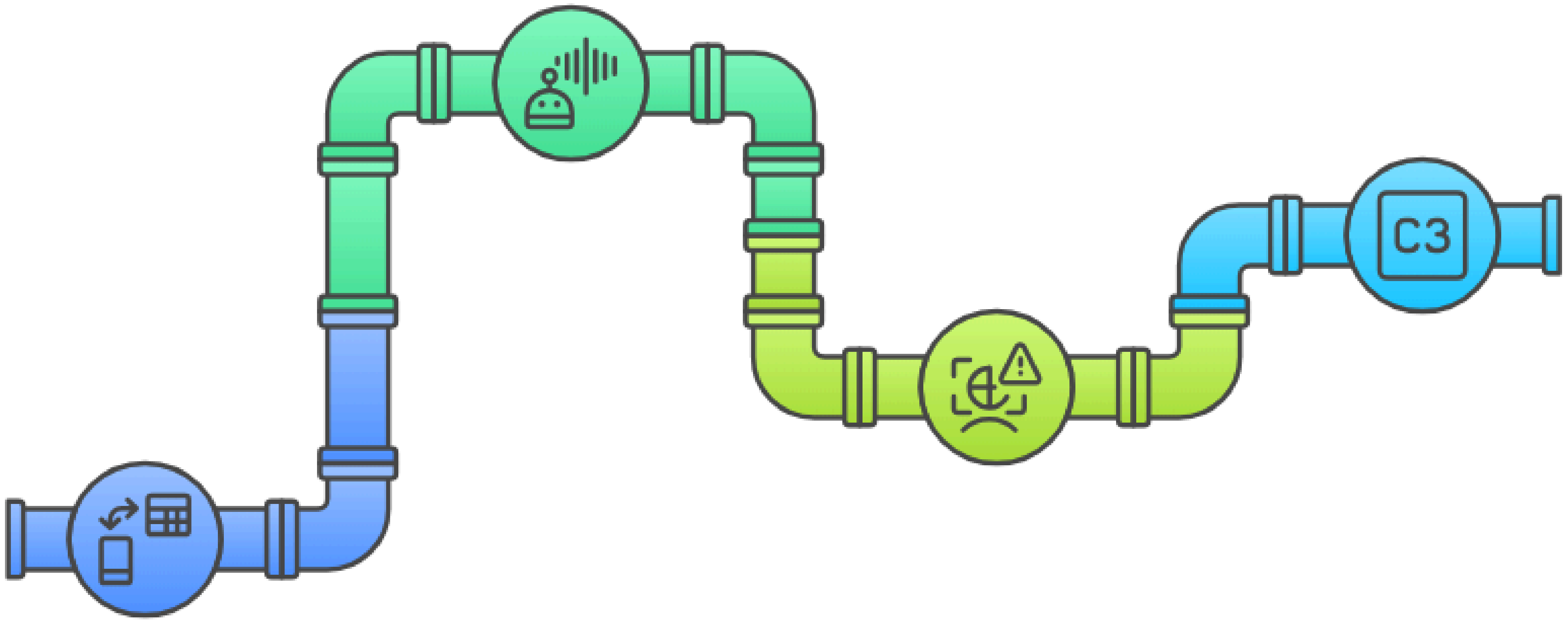
Limited Interpretability

- The Problem: Models often rely on opaque feature sets, lacking the domain-specific justification required by financial regulators and users.
- LoanLens Solution: We engineered a custom, shared RiskScore (from CreditScore, LoanAmount, Income) and ratio metrics, ensuring transparency across all three ML stages.

Absence of Deployable Systems

- The Problem: Academic models are typically restricted to offline training, failing to meet the real-time, low-latency demands of digital lending.
- LoanLens Solution: The multi-stage system is operationalized as a Flask API microservice, optimized for production readiness and millisecond-latency inference.

MULTISTAGE



01

Data Preparation

Initial data processing and cleaning

02

Stage 1 - Prediction

Random Forest Classifier estimates default probability

03

Stage 2 - Segmentation

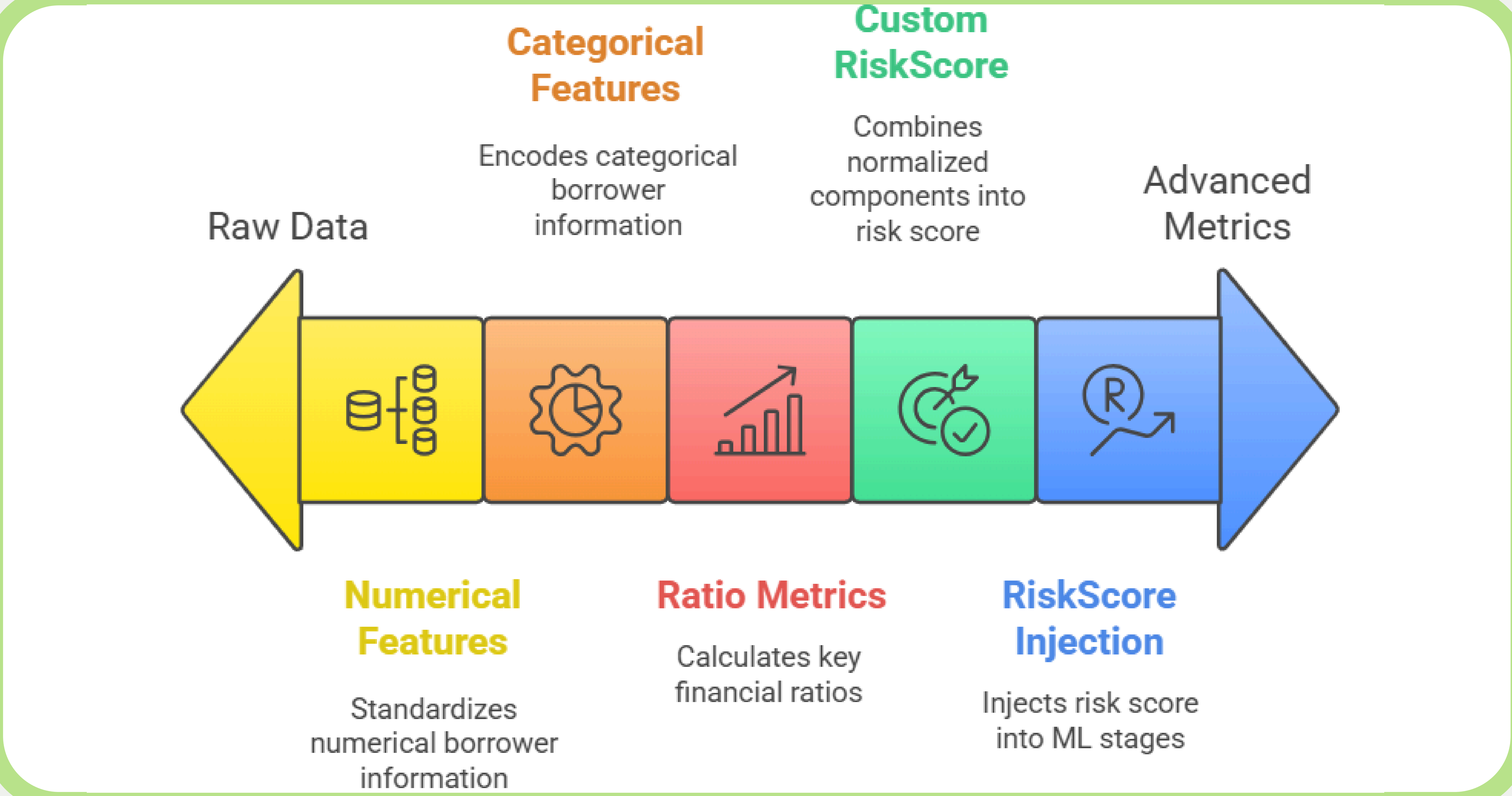
Risk-Weighted K-Means Clustering segments users

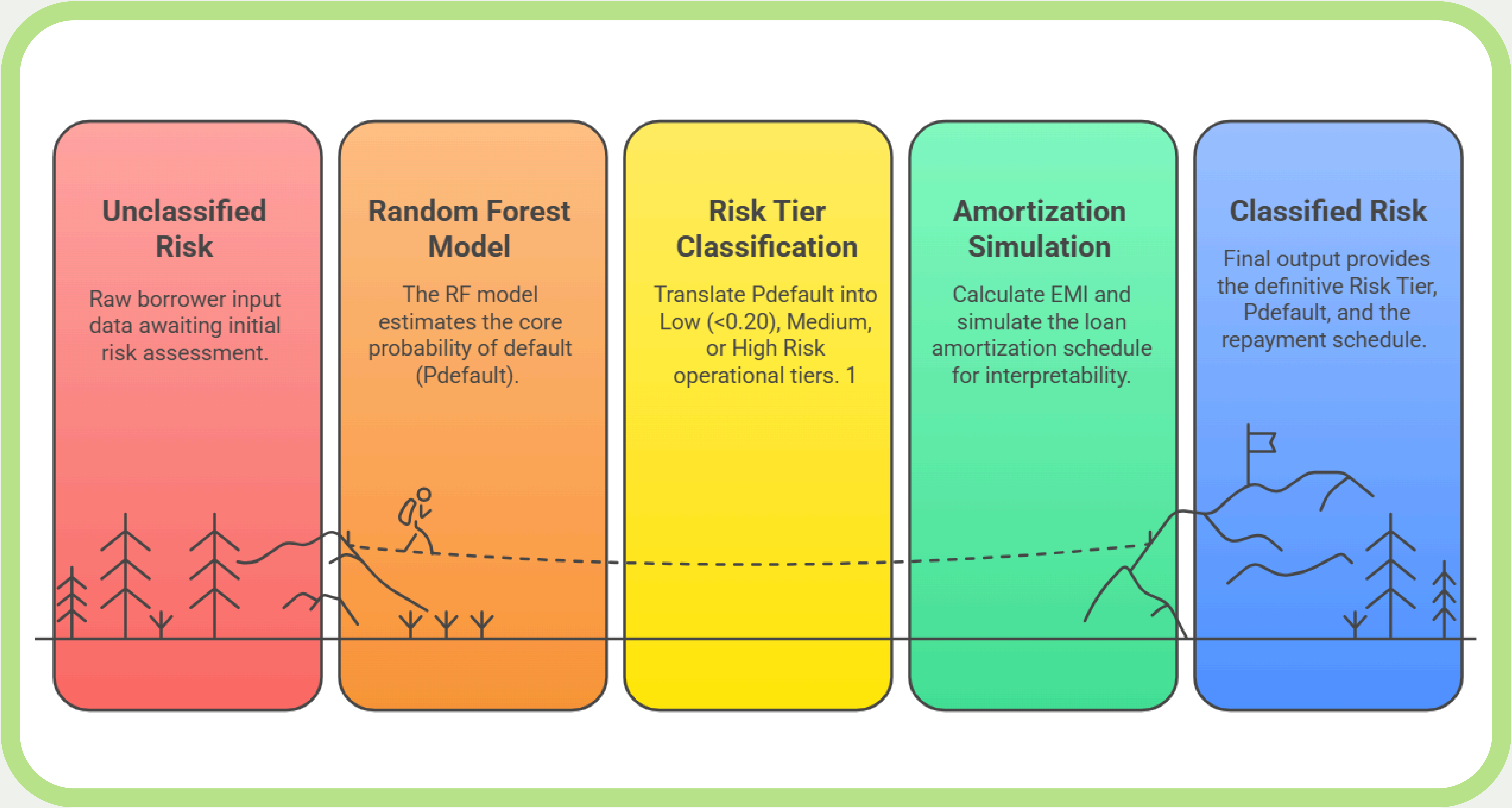
04

Stage 3 - Recommendation

XGBoost Multiclass Classifier matches users with products

DATA ENGINEERING AND CUSTOM RISK SCORING




Unclassified Risk
Raw borrower input data awaiting initial risk assessment.
Random Forest Model
The RF model estimates the core probability of default (Pdefault).
Risk Tier Classification
Translate Pdefault into Low (<0.20), Medium, or High Risk operational tiers. 1
Amortization Simulation
Calculate EMI and simulate the loan amortization schedule for interpretability.
Classified Risk
Final output provides the definitive Risk Tier, Pdefault, and the repayment schedule.

STRATEGIC BORROWER SEGMENTATION (RISK-WEIGHTED K-MEANS)

Model Goal & Input

- The system utilizes Risk-Weighted K-Means clustering (K=5) to efficiently partition borrowers into homogeneous clusters .
- Cluster corresponds to distinct operational risk groups. This is achieved by clustering on a feature RiskScore

Strategic Benefits

Segmentation is critical for strategic portfolio management, as it supports highly specific policy applications like optimized interest rate setting, differentiated underwriting strategies, and credit policy personalization, all of which contribute to improved portfolio stability

Interpretability & Actionability

- The resulting clusters are highly interpretable and directly actionable .
- For example, Cluster 1 represents "Very Low Risk" borrowers eligible for premium offers, while Cluster 5 captures "High Risk" individuals who require restrictive terms or higher oversight.

XGBOOST PERSONALIZED PRODUCT MATCHING

USE THE XGBOOST MULTICLASS CLASSIFIER TO PREDICT THE MOST SUITABLE BANK PRODUCT

Feature Synergy (Critical Input)

The model benefits from preceding stages by incorporating:

- All Engineered Financial Ratios
- Custom RiskScore
- Clustering Labels (from K-Means segmentation)

Safety Logic: Hierarchical Fallback

Ensures safe and compliant recommendations when the top prediction is "No_Match":

1. Choose the next-highest probability bank.
2. Apply hard, rule-based constraints (e.g., minimum income/credit score).
3. If all fail, return "No suitable bank found."

Decision Output (Final Recommendation Layer)

The system generates a transparent, explainable financial decision using:

- Top predicted bank product with confidence score
- Ranked probability distribution across all banks
- Explanation summary using feature importance + cluster behavior

MLOPS: FLASK-BASED REAL-TIME INFERENCE SERVICE

• Deployment Architecture

- Packaged as a Flask microservice optimized for low-latency, real-time inference.
- Supports seamless communication between frontend dashboard and backend ML pipeline.
- Designed for stateless request handling, making it cloud-ready and container-friendly.

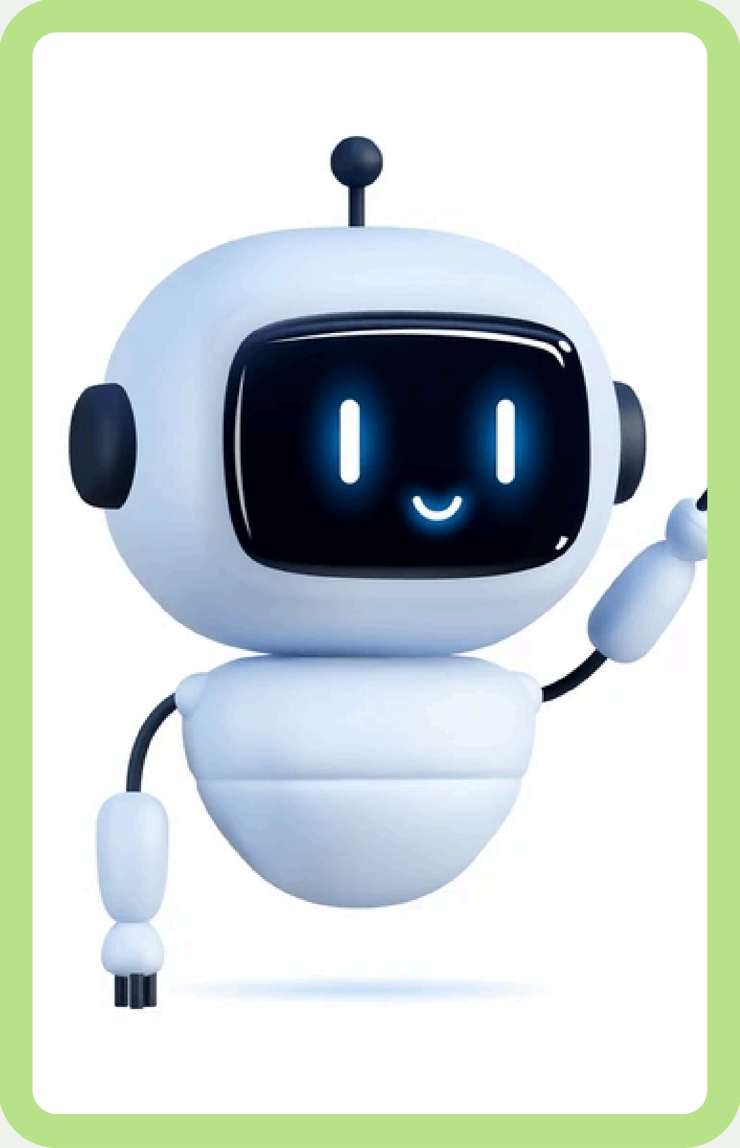
• Model Operationalization

- All trained models (RF, K-Means, XGBoost) and preprocessing assets are serialized with joblib.
- Loaded once into backend memory during server startup to avoid repeated disk I/O.
- Ensures consistent predictions, reproducibility, and quick warm-start performance.

• API Endpoints (Modularity)

- Provides clean, modular REST APIs for different ML tasks:
- /predict_default – Computes default probability and assigns a risk tier.
- /segment_user – Returns the user's behavioral/financial cluster from K-Means.
- /recommend_bank – Generates bank product recommendation using XGBoost + fallback rules.
- /chat – Chatbot endpoint using Gemini for financial guidance.
- Modular design allows endpoints to be scaled, monitored, and versioned independently.

CORE INNOVATION:
THE GEMINI API FINTECH CHATBOT



Financial Guidance

Answers queries on credit scores, loans, EMIs, and interest rates.

Model Explanations

Translates complex machine learning model outputs into simple language.

Contextual Help

Provides contextual support for interpreting bank or lender recommendations.

Health Advice

Offers actionable advice to improve financial health.

EXPERIMENTAL VALIDATION AND PERFORMANCE METRICS

Random Forest Default Prediction (Stage 1)

Metric	Value
Accuracy	88.4%
Precision (default)	79.2%
Recall (default)	75.6%
AUC-ROC	0.92

- Strong discriminative capability

Feature Importance

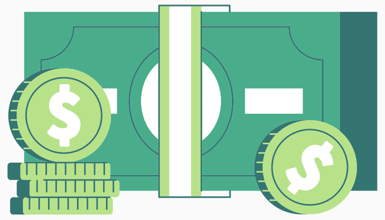
CreditScore   Debt-to-Inome Ratio   Income

Most influnential predictors

XGBoost Recomendation (Stage 3)

- High precision from feature synergagy.
- 95.2% Top 3 Accuracy.
- Overall: Accuracy 90.7%, AUC-ROC 0.93.
- Example: High-credit users matched with prime lenders

Conclusion:
LoanLens Unified System Impact



Contribution 1: Unified Architecture

Integrated Random Forest prediction, K-Means segmentation, and XGBoost recommendation into a single FinTech pipeline.

Contribution 2: Interpretability & Actionability

Introduced a custom RiskScore and engineered ratios that improve explainability and provide actionable borrower insights.

Contribution 3: Operational Readiness

Implemented real-time inference through a Flask microservice for seamless integration with financial workflows.

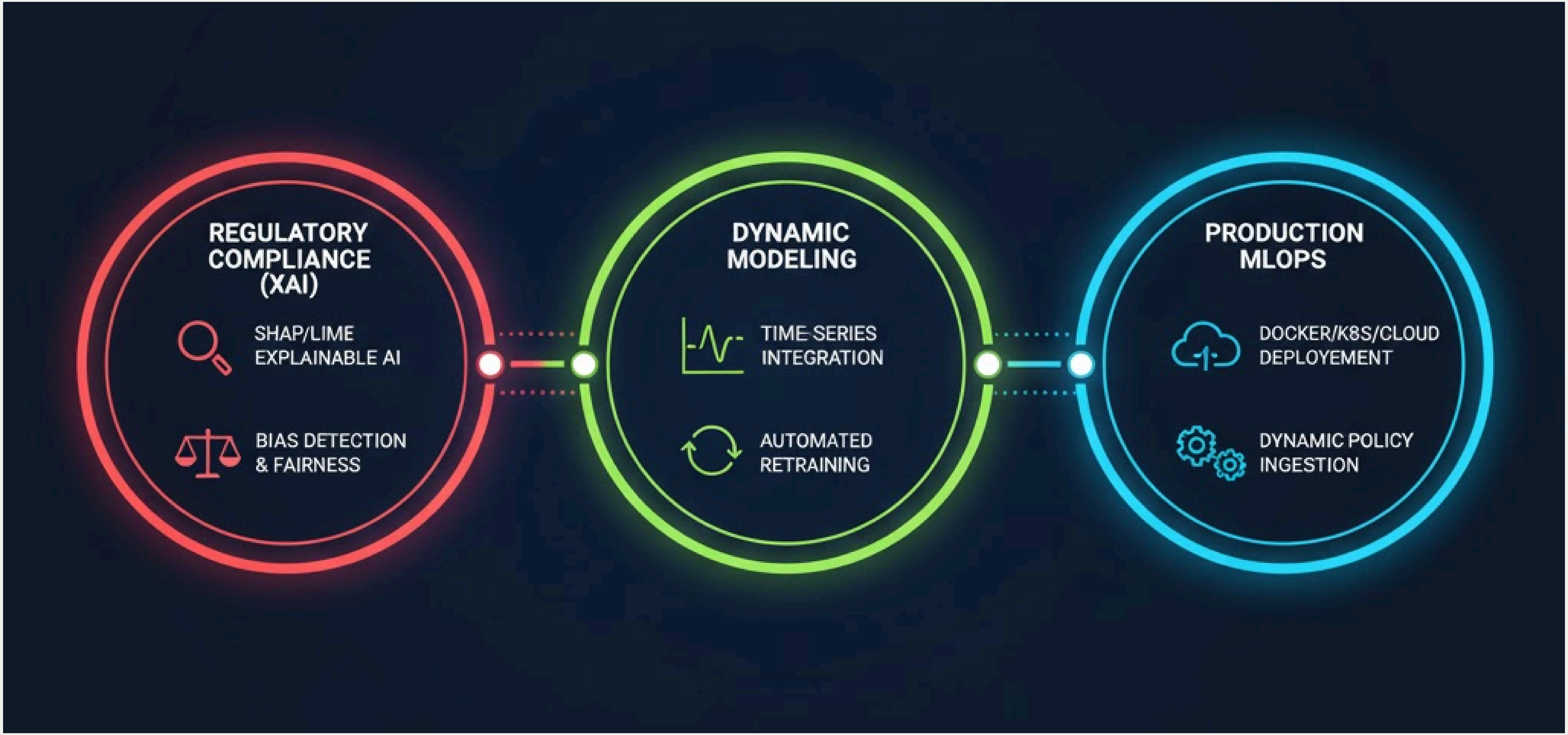
Contribution 4: Enhanced Transparency

Added a Gemini API–powered chatbot offering conversational explanations, boosting user trust and system accessibility.

Overall Finding

A hybrid of supervised and unsupervised models delivers more comprehensive and reliable risk intelligence than isolated approaches.

FUTURE SCOPE



THANK YOU