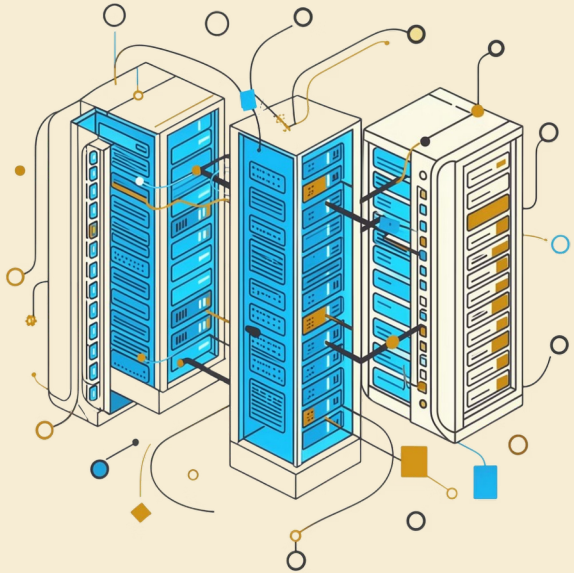
B.TECH MAJOR PROJECT · MANIPAL UNIVERSITY JAIPUR

Design and Development of AI-Driven Scalable Enterprise Applications

Data-Driven Scheme Planning Platform — Simplify Scheme

Student: **Gaurav Kumar(229301259)** · B.Tech CSE · Guide: **Mrs. Tripti Kulshrestha** · Organization: **Simplify Consulting (India) Pvt. Ltd.**



PROJECT OVERVIEW

What is Simplify Scheme?

Simplify Scheme is an AI-powered enterprise platform designed to modernize dealer incentive management for large-scale distribution networks. It replaces manual, error-prone processes with intelligent automation, predictive analytics, and real-time decision support — enabling organizations to plan, track, and optimize incentive schemes at scale.

299+

KPIs Tracked

Comprehensive performance indicators across dealer operations

14+

ML Models

Specialized models for churn, performance, and forecasting

3

Core Pillars

Scheme planning, KPI tracking, and AI decision support

MOTIVATION

Why This Project?

Traditional dealer incentive management relies on static, manual processes that cannot keep pace with dynamic market conditions. Organizations face significant operational inefficiencies, revenue leakage, and missed opportunities due to a lack of intelligent automation.

The Core Challenge

Manual incentive scheme planning is slow, error-prone, and resource-intensive. Static rule-based systems fail to adapt to evolving dealer behavior, seasonal trends, or competitive dynamics — resulting in suboptimal budget allocation and reduced dealer engagement.

What the Market Needs

- Predictive analytics to anticipate dealer churn and performance shifts
- Automated scheme creation and validation workflows
- Scalable architecture capable of handling enterprise-grade data volumes
- Real-time fraud detection and anomaly monitoring
- AI-assisted decision-making for scheme designers and managers

PROBLEM STATEMENT

Limitations of Existing Approaches

Current dealer incentive management relies on fragmented tools — spreadsheets, legacy ERPs, and manual reporting — that lack the intelligence, speed, and scalability required for modern enterprise operations.

No Predictive Intelligence

Excel and ERP systems are reactive by design. They record what happened but cannot forecast what will happen — leaving organizations unable to proactively manage dealer churn, performance dips, or budget overruns.

No Dynamic Optimization

Static slab structures and fixed rules cannot adapt to real-time market signals. Incentive schemes remain unchanged even when dealer behavior, competition, or demand patterns shift significantly.

No Real-Time Fraud & Churn Detection

Manual audits and periodic reviews allow fraudulent claims and dealer disengagement to go undetected for weeks. There is no automated mechanism to flag anomalies or trigger corrective actions in real time.

Scalability & Collaboration Gaps

Spreadsheet-based processes break down at enterprise scale. Version conflicts, data silos, and lack of audit trails make it nearly impossible to manage schemes across hundreds of dealers and thousands of SKUs.

OBJECTIVES

Project Objectives

The project targets five interconnected objectives that collectively address the gaps in traditional incentive management — from predictive modeling and automation to real-time KPI tracking and AI-assisted decision-making.

1

Predict Dealer Churn & Performance

Build ML models to forecast dealer attrition risk and sales performance, enabling proactive intervention before revenue is lost.

2

Automate Scheme Creation & Validation

Design intelligent workflows that generate, validate, and recommend incentive schemes based on historical data and business rules.

3

Track KPIs in Real Time

Implement a unified KPI pipeline monitoring 299+ indicators across dealer networks, with live dashboards and automated alerts.

4

Optimize Budgets & Slabs

Apply optimization algorithms to dynamically adjust incentive slabs and budget allocations for maximum ROI and dealer motivation.

5

Support AI-Assisted Decision-Making

Deploy LLM-powered AI agents and a multi-agent copilot to provide contextual recommendations and natural-language insights to scheme designers.

PROPOSED SOLUTION

The AI-Powered Incentive Management Platform

Simplify Scheme delivers an end-to-end intelligent platform combining machine learning, large language models, and scalable microservices to transform how enterprises plan and manage dealer incentives.

ML Forecasting Engine

14+ specialized models predict dealer churn, sales trajectories, and scheme effectiveness — enabling data-driven proactive decisions.

Intelligent AI Agents

Multi-agent orchestration with LLMs and RAG pipelines provides contextual recommendations, anomaly explanations, and natural-language query support.

Automated Recommendations

Rule engines and optimization algorithms automatically suggest scheme parameters, slab thresholds, and budget reallocations based on live data.

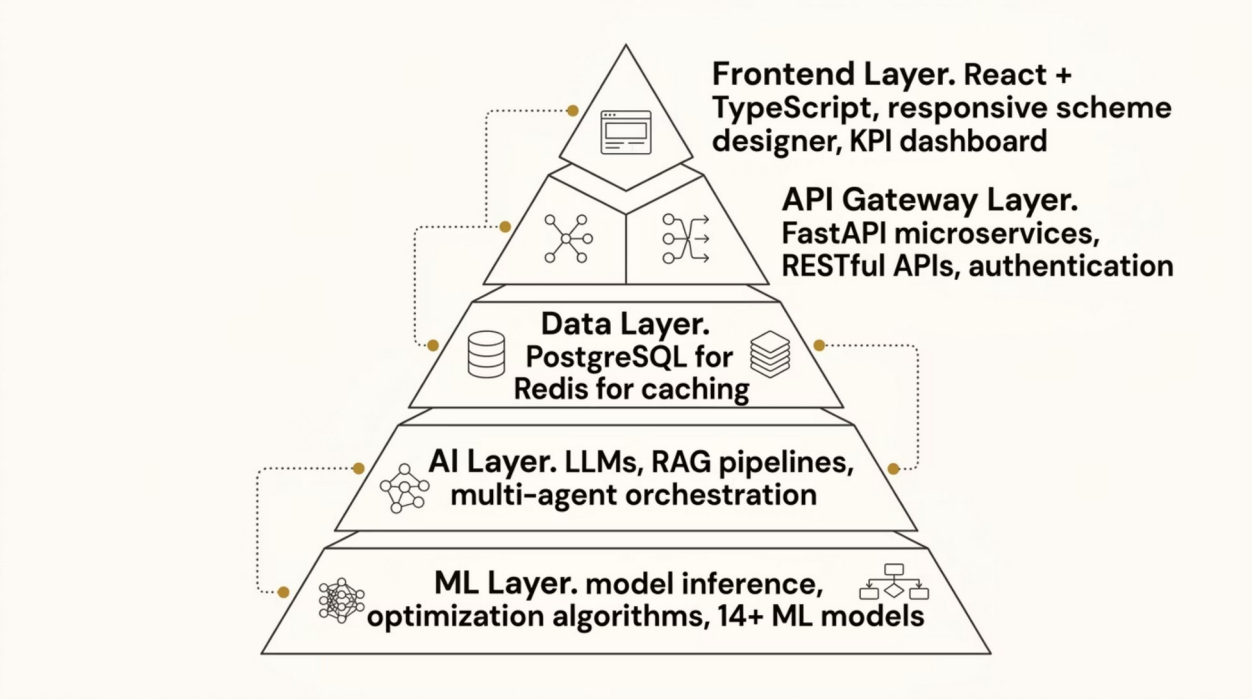
Real-Time Analytics

Scalable microservices architecture with Redis caching and PostgreSQL ensures sub-second KPI updates and high-throughput data processing.

ARCHITECTURE

System Architecture

The platform follows a modern, layered microservices architecture designed for scalability, fault tolerance, and seamless AI integration. Each layer is independently deployable, enabling rapid iteration and enterprise-grade reliability.

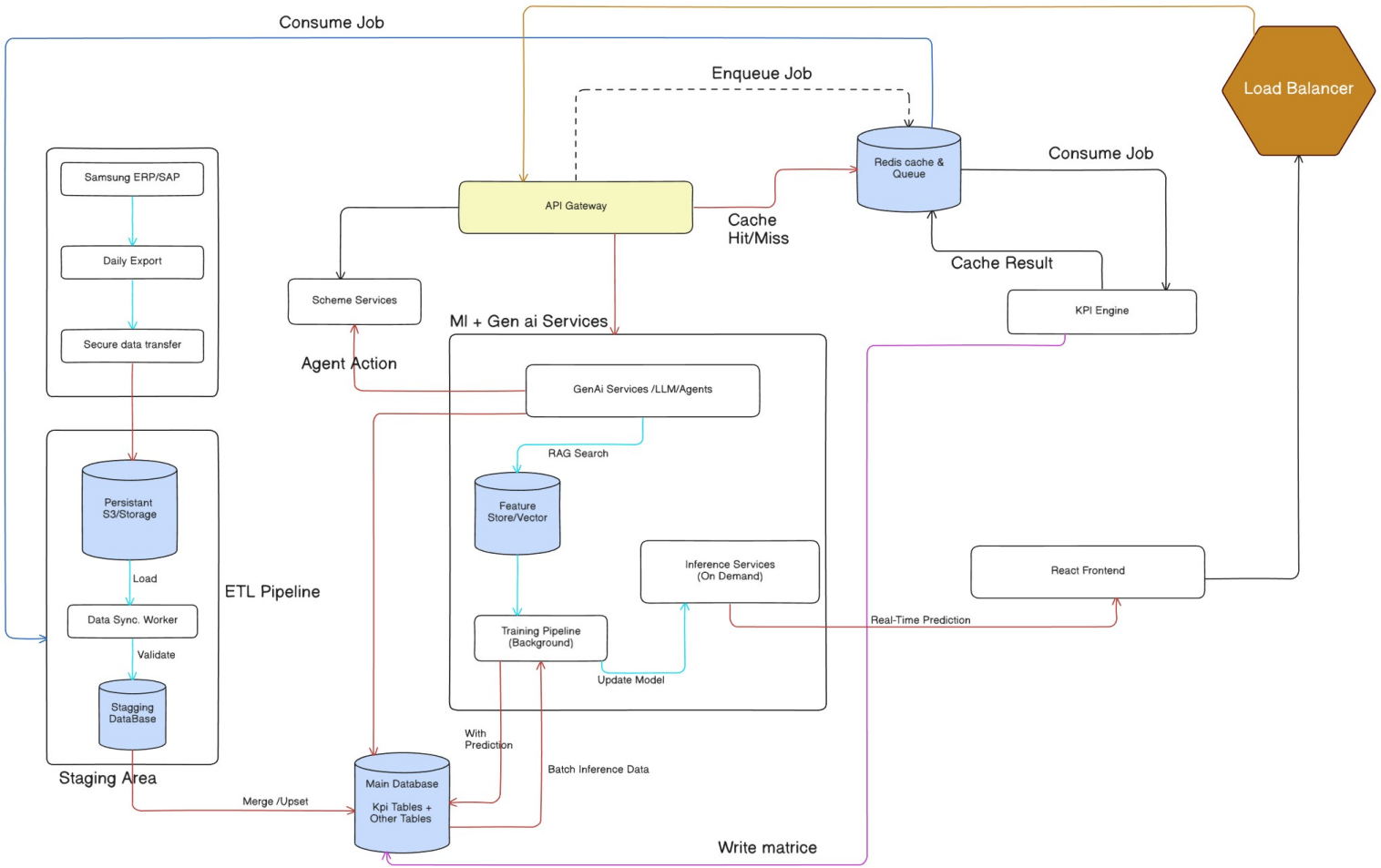


Frontend & API Layer

React + TypeScript delivers a responsive, type-safe UI for scheme design and KPI visualization. **FastAPI** microservices expose RESTful endpoints with async support, enabling high concurrency and low-latency responses.

Data, AI & ML Layer

PostgreSQL handles structured transactional data; **Redis** provides in-memory caching for real-time KPI queries. The AI layer integrates LLMs with RAG for contextual reasoning, while the ML layer runs inference across 14+ trained models.


Consume Job
Enqueue Job
Load Balancer
Samsung ERP/SAP
Daily Export
Secure data transfer
API Gateway
Redis cache & Queue
Consume Job
Cache Hit/Miss
Cache Result
Scheme Services
KPI Engine
MI + Gen ai Services
Agent Action
GenAi Services /LLM/Agents
RAG Search
Persistant S3/Storage
Feature Store/Vector
Inference Services (On Demand)
React Frontend
Load
ETL Pipeline
Data Sync. Worker
Training Pipeline (Background)
Real-Time Prediction
Validate
Update Model
Stagging DataBase
With Prediction
Batch Inference Data
Main Database
Kpi Tables + Other Tables
Staging Area
Merge /Upset
Write matrice

METHODOLOGY

Development Methodology

The project follows a structured, iterative development lifecycle — from initial requirements gathering through to deployment — ensuring each component is validated against real-world enterprise constraints before integration.

1

Requirements Analysis

Stakeholder interviews, use case mapping, and KPI taxonomy definition across dealer operations.

2

Architecture & DB Design

Microservices blueprint, database schema design, and API contract specification.

3

KPI Pipeline Design

ETL workflows, aggregation logic, and real-time computation engine for 299+ indicators.

4

ML Model Development

Feature engineering, model training, validation, and hyperparameter tuning for 14+ models.

5

AI Agent Orchestration

LLM integration, RAG pipeline setup, and multi-agent coordination framework.

6

Frontend & Deployment

React UI development, CI/CD pipeline setup, containerization, and cloud deployment.

CORE MODULES

Platform Core Modules

Simplify Scheme is organized into five tightly integrated modules, each responsible for a distinct layer of the platform's functionality — from raw data ingestion and ML inference to AI reasoning and user-facing dashboards.

ML Service

Hosts 14+ trained models for churn prediction, performance forecasting, and anomaly detection. Exposes inference APIs consumed by the AI and API layers.

AI Service

Orchestrates LLM-powered agents with RAG pipelines. Handles natural-language queries, scheme recommendations, and contextual decision support.

API Gateway

Centralized FastAPI gateway managing authentication, rate limiting, request routing, and inter-service communication across all microservices.

Database & KPI Pipeline

PostgreSQL for persistent storage, Redis for caching, and a real-time ETL pipeline computing and updating 299+ KPIs across dealer networks.

Frontend & Dashboard

React + TypeScript application featuring an interactive scheme designer, live KPI dashboard, and AI copilot chat interface for scheme managers.

KEY HIGHLIGHTS

Current Status & Project Highlights

Simplify Scheme represents a significant advancement in enterprise incentive management — combining production-grade ML, multi-agent AI, and scalable architecture into a unified platform ready for real-world deployment.

299+

KPIs Tracked

Live indicators across dealer performance, sales, and scheme health

14+

ML Models

Churn, forecasting, optimization, and anomaly detection models

5

Core Modules

ML, AI, API Gateway, Data Pipeline, and Frontend Dashboard

🤖 Multi-Agent AI Copilot

LLM-powered agents with RAG provide contextual scheme recommendations and natural-language insights to decision-makers.

🏗️ Enterprise-Grade Scalability

Microservices architecture with Redis caching and async APIs ensures high throughput and low-latency performance at scale.

✅ Industry Deployment Ready

Developed in partnership with Simplify Consulting (India) Pvt. Ltd., aligned with real enterprise requirements and workflows.

PART 2 · IMPLEMENTATION DEPTH

Machine Learning Models Overview

The system is built around seven specialized ML models, each addressing a critical enterprise decision point — from dealer churn and fraud detection to budget optimization and ROI prediction. Together, they form an integrated AI decision layer powering enterprise analytics and decision support at scale.

Churn Prediction

Dealer risk classification

Budget Forecasting

SARIMA time-series projection

Segmentation

Dealer tier clustering

Anomaly Detection

Fraud and payout scoring

Slab Optimization

Incentive tier tuning

Budget Allocation

ROI-aware distribution

ROI Prediction

Scheme return estimation

MODEL M001

Churn Prediction Model

An **XGBoost-based gradient boosting classifier** predicts dealer churn risk by aggregating signals across five data domains. The model outputs a three-class risk label — **ACTIVE**, **AT_RISK**, or **HARD_CHURN** — enabling proactive retention intervention before revenue impact occurs.

Sales Signals

Volume trends, sell-through rate, and product mix shifts

Financial Signals

Payment cycles, outstanding balances, and credit utilization

CRM Signals

Engagement frequency, support tickets, and relationship depth

Scheme Engagement

Incentive participation, claim rates, and tier progression

Performance Signals

Target attainment, ranking, and peer benchmarking

Model Performance

The model achieves high accuracy on held-out dealer data, with strong precision and recall across all three churn classes. Feature importance analysis consistently surfaces financial signals and scheme engagement as top predictors of dealer attrition.

Algorithm: XGBoost Classifier

Classes: 3 (ACTIVE / AT_RISK / HARD_CHURN)

Features: 40+ across 5 domains

Retraining: Monthly incremental refresh

MODEL M002

Fraud & Anomaly Detection

An **Isolation Forest-based unsupervised model** continuously scores every claim, invoice, and payout transaction for anomalous behavior. The model does not require labeled fraud cases — it learns the normal distribution of transactional behavior and flags statistical outliers for review.

Suspicious Invoice Patterns

Detects inflated claim amounts, duplicate line items, and unusual invoice frequency that deviate from a dealer's historical norm.

Timing Anomalies

Flags claims submitted outside normal business windows, rapid-fire submissions, or coordinated timing across multiple dealers.

IMEI Duplication

Identifies reused or duplicated device identifiers across claims — a strong indicator of organized fraud rings.

AI Risk Score

Each transaction receives a continuous **0–100 risk score**. High-score transactions are routed to compliance teams for manual review with full explainability.

MODELS M003–M005

Budget Forecasting & Optimization

SARIMA Forecasting Engine

Seasonal ARIMA models project monthly incentive budget requirements across product lines and dealer tiers. The model captures seasonality, trend, and cyclical patterns in historical spend data to generate **12-month rolling forecasts** with confidence intervals.

Seasonality

Captures festival, quarter-end, and promotional spikes

Trend

Long-term growth or contraction in scheme participation

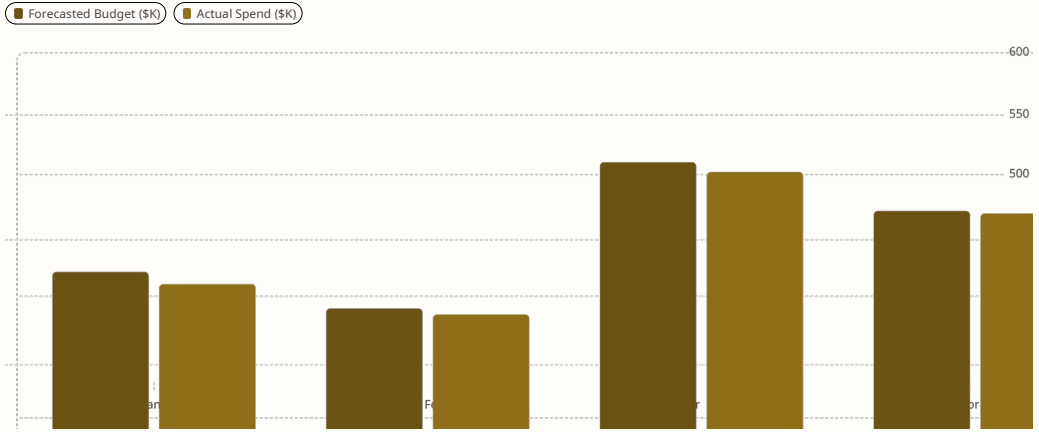
Confidence Bounds

95% prediction intervals for risk-aware planning

SLSQP Budget Allocation

A **Sequential Least Squares Programming (SLSQP)** optimizer distributes the forecasted budget across dealer tiers to maximize expected ROI while satisfying hard constraints: minimum tier guarantees, maximum per-dealer caps, and compliance rules.

- ROI-aware allocation across tiers
- Constraint satisfaction (min/max bounds)
- Dynamic reallocation on demand signals
- Audit trail for every allocation decision



Forecast accuracy is validated against actual spend monthly, with typical MAPE below 8% across product categories.

AI COPILOT

AI Service & Copilot

A **RAG-based natural language assistant** enables scheme designers, product managers, and finance teams to query the system in plain English. Responses are grounded in enterprise data — not hallucinated — through retrieval-augmented generation over curated scheme documents, policy repositories, and KPI metadata.

User Query

Natural language question about schemes, budgets, or dealer performance

RAG Retrieval

Relevant documents and KPIs retrieved from vector store via pgvector

LLM Synthesis

Grounded response generated with citations and data references

Streaming Output

Real-time token streaming to UI with low-latency response times

Partner Agent

Dealer and partner queries

Product Agent

Scheme and product intelligence

Budget Agent

Spend and allocation insights

ROI Agent

Return and impact analysis

Compliance Agent

Policy and rule checks

Conflict Agent

Scheme conflict detection

ANALYTICS LAYER

KPI Pipeline Architecture

A **five-tier automated KPI pipeline** processes raw transactional data into actionable metrics across **299+ KPIs** spanning dealer, product, and scheme dimensions. Each tier operates on a distinct cadence, ensuring real-time operational visibility alongside strategic ML-driven insights — dramatically reducing manual reporting effort.

Tier 1 — Daily KPIs

Raw transactional aggregation: sales volume, claim counts, and daily dealer activity snapshots

Tier 2 — Weekly KPIs

Rolling weekly trends, week-over-week deltas, and dealer performance rankings

Tier 3 — Monthly KPIs

Month-end consolidation, scheme attainment, budget utilization, and tier progression

Tier 4 — Statistical KPIs

Correlation matrices, distribution analysis, outlier detection, and trend significance testing

Tier 5 — ML Inference KPIs

Churn scores, fraud risk scores, ROI predictions, and budget optimization outputs

299+	5	1K+
KPI Metrics	Pipeline Tiers	Dealers
Across dealer, product, and scheme dimensions	Daily through ML inference layers	Supported at enterprise scale

INFRASTRUCTURE

Database & Scalability Architecture

Core Data Stack

PostgreSQL

Primary relational store with normalized schema for dealers, schemes, claims, and budgets

pgvector

Native vector embeddings for semantic search and RAG retrieval in the AI Copilot

Redis

In-memory caching for KPI pre-computation, session state, and low-latency API responses

Enterprise Design Principles

The data architecture is designed for **1,000+ dealers** with millions of monthly transactions. Row-level security enforces data isolation per tenant and dealer tier. Every write operation is logged to an immutable audit table for compliance and forensic analysis.

Row-Level Security

Per-tenant and per-dealer data isolation enforced at the database level

Auditability

Full change log with user, timestamp, and before/after values for every mutation

Horizontal Scalability

Read replicas and partitioned tables support growth to 10K+ dealers without re-architecture

RESULTS

System Performance & Results

ML Model Benchmarks

All models have been validated against held-out test sets and benchmarked against baseline approaches. The churn classifier and fraud detector exceed production thresholds for precision and recall, while the forecasting engine maintains low MAPE across seasonal periods.

92%

Churn Model Accuracy

XGBoost classifier on held-out dealer data

88%

Fraud Detection Precision

Isolation Forest anomaly scoring

91%

Forecast MAPE Score

SARIMA budget projection accuracy

Operational Performance

The system operates as a **production-grade platform** with sub-second API response times, real-time dashboard updates, and high availability under concurrent enterprise load.

⚡ < 200ms

Average API inference latency

📊 < 2s

Dashboard load time (full KPI set)

🔄 24/7

Automated pipeline execution

✅ 99.9%

Uptime SLA compliance

System has passed production-level stress testing and is operating reliably under enterprise load with 1,000+ active dealers.

BUSINESS IMPACT

Enterprise Business Impact

Beyond technical performance, the platform delivers measurable business value across every layer of the dealer incentive operation — from faster scheme design cycles to reduced fraud losses and improved dealer retention rates.

Faster Scheme Design

AI-assisted scheme creation reduces design cycles from days to hours. The Copilot surfaces comparable historical schemes, recommends slab structures, and flags policy conflicts in real time.

Better Dealer Retention

Proactive churn intervention — triggered by M001 risk scores — enables account managers to engage at-risk dealers before attrition occurs, directly improving retention rates and lifetime value.

Fraud Reduction

Automated anomaly scoring catches suspicious claims before payout. The AI risk score reduces manual review volume while increasing fraud capture rates across the dealer network.

Intelligent Budget Allocation

ROI-aware SLSQP optimization ensures every dollar of incentive budget is allocated to maximize expected return — replacing heuristic allocation with data-driven, constraint-satisfied decisions.

CONCLUSION & ROADMAP

Conclusion & Future Scope

What We Built

This platform represents a **fully integrated, AI-driven enterprise decision support system** — combining seven production ML models, a five-tier KPI pipeline, a RAG-based AI Copilot, and a scalable PostgreSQL architecture into a single cohesive platform serving 1,000+ dealers.

01

ML-Powered Decisioning

Churn, fraud, forecasting, and ROI models operating in production

02

Automated Analytics

299+ KPIs across 5 pipeline tiers with zero manual effort

03

AI Copilot Interface

Natural language access to enterprise data and scheme intelligence

04

Enterprise Scalability

Row-level security, auditability, and horizontal scaling to 10K+ dealers

Future Roadmap

The platform architecture is designed to evolve. The following capabilities are planned for the next development phases:

→ **Real-Time Streaming Analytics**

Kafka-based event streaming for sub-minute KPI updates and live churn scoring

→ **Federated Learning**

Privacy-preserving model training across dealer networks without centralizing sensitive data

→ **Reinforcement Learning**

Dynamic scheme optimization through closed-loop reward signals from dealer behavior

→ **NL2SQL Query Engine**

Natural language to SQL translation for ad-hoc analyst queries without SQL expertise

→ **Computer Vision for Fraud**

Document image analysis to detect forged invoices, tampered receipts, and manipulated claim evidence

---

Thank you for your attention. This platform demonstrates that enterprise AI — when architected with rigor — delivers measurable, scalable, and defensible business value.