FinFlow: Automated Credit Risk Management & ROI Optimization

Project Lead: Yue Dong (Lead Machine Learning Engineer)

Date: 2026-02-05

---

1. Business Objectives

In the traditional lending landscape, manual credit reviews are slow, subjective, and prone to human error. FinFlow was designed as a data-driven solution to achieve three primary strategic goals:

- **Operational Efficiency:** Automate over **70%** of routine loan applications, allowing credit officers to focus on high-complexity cases.
- **Net Profit Maximization:** Use mathematical modeling to identify the "Sweet Spot" between interest income and default loss.
- **Systemic Stability:** Establish a production-grade monitoring framework to detect market shifts (Data Drift) in real-time.

---

2. Variable Definitions & Feature Engineering

The model utilizes both raw financial data and engineered "ratio features" to capture the nuanced repayment capacity of small businesses.

Core Feature Set

Category	Variable	Business Rationale
Solvency	loan_to_income	Ratio of total debt to annual income (The #1 risk predictor).
Leverage	rev_to_loan_ratio	Measures how much revenue is generated for every $1 borrowed.
Credit History	credit_tier	Binned categorical rankings (Elite, Average, Subprime).
Efficiency	rev_per_employee	Revenue per head; reflects management quality and overhead risk.
Market Segment	industry	Weights risk based on sector-specific volatility (e.g., Retail vs. Tech).

---

3. Methodology

The project follows the standard **Machine Learning Life Cycle (MLOps)**, divided into five distinct phases:

1. **Data Quality Audit (EDA):** Cleaned 5,000+ historical records and addressed a 15% class imbalance.
2. **End-to-End Pipeline:** Built a robust feature pipeline using SimpleImputer for missing values and OneHotEncoder for categorical alignment.
3. **Model Selection:** Evaluated Logistic Regression vs. Random Forest vs. XGBoost; **Random Forest** was selected for its superior performance in capturing non-linear risk patterns.
4. **Strategy Optimization:** Utilized **SHAP** for model explainability and plotted **Profit-Maximization Curves**.
5. **Engineering Deployment:** Modularized experimental code into production-ready Python scripts with built-in validation.

---

4. Key Findings & Business Strategy

Through rigorous testing and simulation, the project yielded the following outcomes:

- **Performance:** The champion model (Random Forest) achieved an **81% Precision** rate, effectively flagging the vast majority of potential defaults.
- **The 0.29 Pivot Point:** Our research proved that the standard 0.5 probability threshold was sub-optimal. The **0.29 threshold** was identified as the point of maximum net recovery:
    - **Prob < 0.20:** 🟢 **Auto-Approve** (High velocity, low risk).
    - **0.20 <=Prob < 0.29:** 🟡 **Manual Review** (Edge cases for expert audit).
    - **Prob >=0.29:** 🔴 **Auto-Reject** (Immediate capital protection).

---

5. Production Practice & MLOps

To ensure stability in a live environment, FinFlow includes a comprehensive automation and monitoring suite:

- **Weekly Pipeline (weekly_run.py):** Automatically processes new batches of applications every week.
- **Stability Monitoring (PSI):** Real-time tracking of the Population Stability Index to prevent model decay.

$$PSI = \sum (Actual\% - Expected\%) \times \ln\left(\frac{Actual\%}{Expected\%}\right)$$

    - **Latest Result: 0.0003** (Indicates near-perfect alignment with training data).
- **Automated Auditing:** Generates ROI reports for every run, quantifying "Potential Loss Avoided."

---

6. Future Roadmap

While the system is currently live and stable, the following steps are planned to maintain a competitive edge:

A/B Testing Strategy

- **Experiment Design:** Allocate 10% of traffic to an "Aggressive Group" (Threshold set at 0.35).
- **Goal:** Observe if the increased interest income from higher approval volume offsets the marginal increase in default costs.

System Evolution

- **Real-Time API:** Transition from weekly batch processing to an API-based "Instant Decision" architecture.
- **Alternative Data:** Integrate external datasets (e.g., utility payments or social sentiment) to further refine the credit profiles of small businesses.