Final Report / Blog Post – Credit Risk Probability Model for Alternative Data

1. Executive Summary

Bati Bank seeks to responsibly expand credit to customers without traditional financial histories. This project leverages alternative behavioral data to predict loan default probability.

Key Achievements:

- Built a **robust, modular credit risk model** using alternative data.
- Implemented **unit/integration tests** and **CI/CD** pipelines for reliability.
- Developed an **interactive Streamlit dashboard** with **SHAP explainability**.
- Delivered a **finance-ready report and presentation** demonstrating business impact.

Business Impact:

- Reduced lending risk via accurate default prediction.
- Enabled responsible market expansion.
- Increased transparency for non-technical stakeholders and regulators.

---

2. Problem Statement

Traditional credit scoring excludes customers without prior financial history. This creates:

- Missed business opportunities.
- Increased portfolio risk due to uninformed lending decisions.

Solution: Predict loan default probability using alternative behavioral data to guide responsible lending decisions.

---

3. Approach & Methodology

1. **Data Collection & Preprocessing**
    - Sources: Mobile usage, transactional behavior, alternative financial indicators.
    - Cleaning: Handling missing values, normalization, feature engineering.
    - Example Screenshot: *[Insert preprocessed dataset snapshot]*
2. **Model Development**
    - Algorithms: Logistic Regression, Random Forest, Gradient Boosting.
    - Evaluation: Accuracy, AUC, Precision-Recall, F1-score.

- Example Screenshot: *[Insert evaluation metrics chart]*

3. **Explainability**
    - SHAP summary and force plots show **feature importance per prediction**.
    - Helps explain why a loan is approved or declined.
    - Example Screenshot: *[Insert SHAP plots]*
4. **Deployment & Dashboard**
    - **FastAPI** backend for prediction endpoints.
    - **Streamlit Dashboard**: Input data → see prediction & SHAP explanation.
    - Example Screenshot: *[Insert dashboard screenshot]*
5. **Engineering Reliability**
    - **Code Refactor:** Modular scripts for easy maintenance.
    - **Testing:** Unit & integration tests validate pipeline and endpoints.
    - **CI/CD:** Automated testing ensures stability on every update.
    - Example Screenshot: *[Insert CI/CD workflow screenshot]*

---

4. Results & Business Value

Metric	Value / Observation
Model Accuracy	[Insert value]
AUC Score	[Insert value]
Top Features (SHAP)	[List top 3–5 features]
Dashboard Usage	Non-technical users can interpret results
Risk Reduction Potential	Estimated default risk reduced by X%

Key Insights:

- Certain behavioral indicators strongly predict default risk.
- The dashboard empowers decision-makers with transparent and actionable insights.
- CI/CD + tests ensure trustworthiness for production deployment.

---

5. Future Improvements

- Incorporate additional alternative data sources for more coverage.
- Add real-time monitoring and alerts for model drift.
- Extend the dashboard for portfolio-level risk simulation.
- Consider ensemble or deep learning models for higher predictive accuracy.

---

6. Conclusion

This project demonstrates **how robust engineering, explainable AI, and clear business storytelling** can transform a data science solution into a finance-ready tool. By combining risk-aware modeling, reliable deployment, and intuitive dashboards, Bati Bank can confidently expand credit to underserved customers while minimizing default risk.

---

7. Supporting Materials

- **GitHub Repository:** [Insert link]
- **Screenshots:**
    - Preprocessed Data → *[Insert]*
    - Model Evaluation Metrics → *[Insert]*
    - SHAP Visualizations → *[Insert]*
    - Streamlit Dashboard → *[Insert]*
    - CI/CD Workflow → *[Insert]*