Credit Risk Modeling Proposal

Introduction

Machine learning (ML) can transform credit risk assessment by:

1. Enhancing prediction accuracy.  
2. Reducing manual effort and bias.  
3. Providing actionable insights for loan officers.  
4. Analyzing large datasets efficiently.

Data Requirements

1. Customer Information:  
Demographics (e.g., age, location).  
Employment and income details.

2. Credit History:  
Repayment behavior.  
Credit limits and utilization.  
Records of defaults or bankruptcies.

3. Transaction Data:  
Spending patterns.  
Recurring payments.

4. Additional Features:  
Economic indicators (e.g., interest rates).  
Behavioral data (e.g., loan inquiries).

Data Outputs

1. Risk Score: Numerical score representing default risk.  
2. Probability of Default (PD): Statistical likelihood of non-repayment.  
3. Segment Analysis: Classification into risk tiers (low, medium, high).  
4. Feature Insights: Key factors driving the risk score.  
5. Recommendations: Suggested actions (e.g., approve, decline, modify terms).

Architecture

Recommended model types:

1. Logistic Regression:  
A simple and interpretable baseline.  
Useful for binary classification (default vs. non-default).

2. Tree-Based Models:  
High performance with non-linear relationships.  
Offers feature importance metrics.

3. Neural Networks:  
Handles large-scale, complex data.  
Effective but less interpretable.

4. Hybrid Approaches:  
Combines tree-based models with explainability tools.

Risks and Challenges

Key challenges include:

1. Data Quality:  
Incomplete or biased data can affect accuracy.  
Needs robust preprocessing and engineering.

2. Explainability:  
Regulatory requirements demand transparency.  
Use tools for model interpretability.

3. Compliance:  
Avoid biases and adhere to regulations (e.g., GDPR, Fair Lending).

4. Overfitting:  
Ensure models generalize well with regularization and validation.

5. Dynamic Changes:  
Economic shifts require regular model updates.