***Credit Risk Modeling Proposal***  
A machine learning-based credit risk modeling system will enhance Citi's loan management system by automating creditworthiness evaluations, reducing manual effort, and improving prediction accuracy. This system can assist loan officers in making data-driven decisions, mitigate the risk of loan defaults, and streamline the approval process.

***Data Requirements***  
 **Demographics**: Age, employment status, income, education level.

 **Credit History**: Credit score, previous loans, repayment behavior, defaults.

 **Financial Metrics**: Debt-to-income ratio, current liabilities, net worth.

 **Loan Details**: Loan type, amount, tenure, interest rate.

 **Behavioral Data**: Payment patterns, savings trends, transaction history.

 **Macroeconomic Indicators**: Interest rate trends, inflation rates, unemployment rates.

***Data Outputs***  
 **Credit Risk Score**: A numerical score representing the applicant's likelihood of default.

 **Approval Recommendation**: Classification of applications (e.g., Approve, Reject, Refer for Manual Review).

 **Default Probability**: A predicted probability of default over the loan tenure.

 **Loan Limits and Terms**: Recommended loan amount, interest rate, and repayment tenure based on risk assessment.

***Architecture***  
 **Logistic Regression**: A baseline model for binary classification (low vs. high risk).

 **Decision Trees and Random Forests**: For capturing non-linear relationships and feature importance.

 **Gradient Boosting Models**: Ideal for handling tabular data with high accuracy.

 **Neural Networks**: For large datasets and complex patterns, though interpretability may be challenging.

**Architecture**: Gradient Boosting Models, such as LightGBM, due to their high performance with tabular data, interpretability through feature importance, and scalability.

***Risks and Challenges***  
 **Data Quality**: Inconsistent, incomplete, or biased data may degrade model performance.

 **Regulatory Compliance**: Adherence to financial regulations like the Fair Lending Act and GDPR is critical.

 **Interpretability**: Complex models may lack transparency, making it hard to justify decisions to stakeholders.

 **Economic Changes**: Sudden macroeconomic shifts (e.g., recession) may affect model accuracy.

 **Bias and Fairness**: Ensuring the model does not inadvertently discriminate against any demographic group.