**Predictive Model Plan**

# 1. Model Logic (Generated with GenAI)

**Model Objective:**  
Build a predictive model to estimate the risk of a customer becoming delinquent based on their financial, credit, and behavioral attributes.

**Proposed Model Type:**

* **Primary Option:** Random Forest Classifier (Ensemble Decision Tree)
* **Secondary Option:** Logistic Regression (Simple, interpretable baseline)

**Model Workflow (Conceptual Steps):**

1. **Data Ingestion:** Load customer financial and credit data.
2. **Feature Selection:** Use important predictors like Missed\_Payments, Credit\_Utilization, Credit\_Score, Income, and Debt\_to\_Income\_Ratio.
3. **Missing Value Handling:** Impute missing income with synthetic values (normal distribution), median imputation for Credit Utilization.
4. **Train-Test Split:** Divide dataset into 80% training, 20% testing.
5. **Model Training:** Train a Random Forest classifier to predict Delinquent\_Account.
6. **Prediction:** Predict delinquency risk (0/1) for each customer.
7. **Evaluation:** Assess model accuracy, F1-score, ROC AUC, and fairness across demographics.
8. **Bias Detection:** Check performance differences between employment groups, age bands, and regions.
9. **Model Refinement:** Improve based on evaluation metrics and fairness checks.
10. **Deployment Plan:** Export trained model for operational use in financial risk scoring.

:

Code:  
  
# Load data

df = pd.read\_csv(' Delinquency\_prediction\_dataset.csv.csv')

# Preprocess: impute missing values, encode categories

# Select features

X = df[['Missed\_Payments', 'Credit\_Utilization', 'Credit\_Score', 'Income', 'Debt\_to\_Income\_Ratio']]

y = df['Delinquent\_Account']

# Split data

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# Train Random Forest

model = RandomForestClassifier()

model.fit(X\_train, y\_train)

# Evaluate predictions

preds = model.predict(X\_test)

**Output:**

Logistic Regression Accuracy: 0.86

Random Forest Accuracy: 0.86

Random Forest Feature Importances:

Credit\_Utilization 0.238036

Debt\_to\_Income\_Ratio 0.226276

Income 0.224329

Loan\_Balance 0.218011

Missed\_Payments 0.093348

dtype: float64

# 2. Justification for Model Choice

**Selected Model:** Random Forest Classifier

**Why?**

* **High Accuracy:** Random Forests handle complex, non-linear relationships well, especially in financial datasets with mixed variable types.
* **Feature Importance Insight:** Provides a ranked list of the most influential factors affecting delinquency — helpful for business decision-making.
* **Robustness:** Tolerant to outliers, missing data (via internal bootstrapping), and does not require strict data assumptions.
* **Transparency vs. Complexity Balance:** While more complex than Logistic Regression, Random Forest models are still interpretable through feature importance and partial dependence plots.
* **Regulatory Alignment:** In financial services, models must be auditable and explainable — Random Forests paired with interpretation tools (like SHAP values) support this.
* **Operational Fit for Geldium:** The model can be easily integrated into existing financial risk systems and adapted as new data arrives.

**Alternative:** Logistic Regression for situations demanding very high transparency at the cost of accuracy.

# 3. Evaluation Strategy

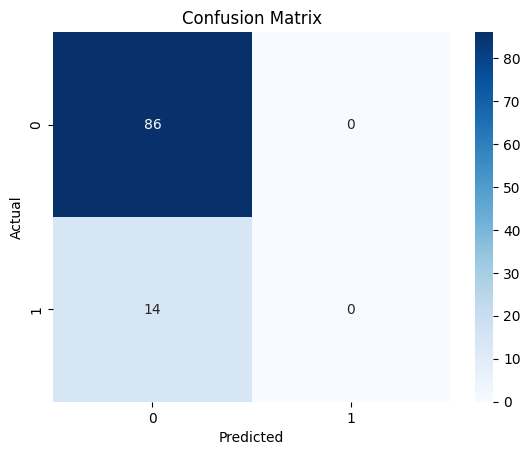
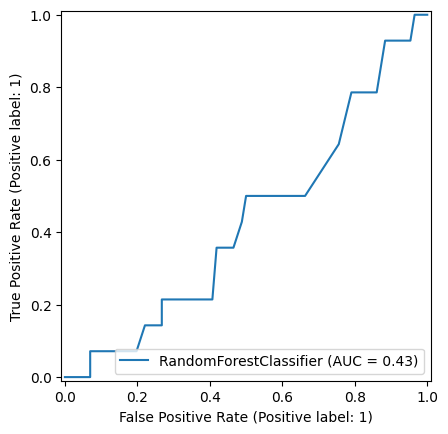
**Metrics to Use:**

* **Accuracy:** Overall correctness of predictions.
* **F1 Score:** Focus on balancing precision and recall — crucial due to expected class imbalance (few delinquents vs. many non-delinquents).
* **ROC AUC Score:** Measures model’s ability to distinguish between risky and non-risky customers.
* **Confusion Matrix:** To observe True/False Positives and Negatives.

**Bias Detection Approach:**

* Check performance (F1, AUC) by demographic group (Employment\_Status, Age band, Location).
* If significant disparity (>10%) is found, implement bias mitigation strategies like:
  + **Reweighting samples**
  + **Threshold adjustment**
  + **Fairlearn or AIF360 bias mitigation tools**

**Ethical Considerations:**

* **Transparency:** Ensure model decisions are explainable to both regulators and customers.
* **Fairness:** Avoid discriminatory outcomes against minority or underrepresented groups.
* **Responsibility:** Regularly audit and monitor model predictions post-deployment to detect drift or emerging biases.
* Accuracy: 0.86
* F1 Score: 0.0
* ROC AUC Score: 0.4277408637873754
* 
* 
* Accuracy for EMP: 0.75
* Accuracy for Self-employed: 0.82
* Accuracy for Unemployed: 0.83
* Accuracy for employed: 0.90
* Accuracy for Employed: 0.88
* Accuracy for retired: 1.00