**HELOC Loan Decision Support System**

**Task 1: Formulating the Problem**

**The goal of this project is to develop a machine learning model that predicts whether a Home Equity Line of Credit (HELOC) application should be approved or denied. This is a binary classification problem where the model outputs either Approved (1) or Denied (0).**

**Performance Measures**

**To evaluate the model's effectiveness, we use the following performance metrics:**

* **Accuracy: Measures the proportion of correctly predicted instances.**
* **Precision & Recall: Ensures the model balances false positives and false negatives.**
* **F1 Score: A combination of precision and recall to assess model performance.**
* **AUC-ROC: Evaluates the model's ability to distinguish between approved and denied loans.**

**Cost Savings Estimation**

**By automating the approval process, the bank can reduce manual review time, improve decision consistency, and lower operational costs. The model is expected to save costs proportional to its accuracy and precision in predicting approvals correctly.**

**Task 2: Exploratory Data Analysis**

**Issues Identified and Fixes**

**During EDA, we identified the following issues:**

1. **Missing values: Special values (-9, -8, -7) were replaced with NaN and imputed using the median.**
2. **Feature correlations: Some features were highly correlated; we retained the most informative ones.**
3. **Feature distributions: Certain features had skewed distributions, affecting model training.**

**Task 3: ML Model**

**Model Training & Evaluation**

**We trained and evaluated multiple models:**

* **Logistic Regression**
* **Decision Tree**
* **Random Forest (Best Performing Model)**
* **Gradient Boosting**

**Final Model Selection: After hyperparameter tuning, Random Forest was chosen for its high accuracy and balanced precision-recall performance.**

**Task 4: Explanations**

**SHAP Analysis**

**To make the model’s decisions interpretable, we used SHAP (Shapley Additive Explanations):**

* **SHAP Summary Plot: Shows how features contribute to approval or denial.**
* **SHAP Bar Chart: Displays the most influential features in decision-making.**

**These insights help financial analysts understand why an application was approved or denied.**

**Task 6: Discussion**

**Will the model’s real-world performance match the analysis?**

**While the model has demonstrated strong performance during testing, real-world applications may introduce new challenges:**

1. **Data Drift: Economic conditions and customer behaviors change over time, impacting feature distributions.**
2. **Selection Bias: The training data may not fully represent future applicants.**
3. **Regulatory Changes: Lending policies may evolve, making historical data less relevant.**

**Ensuring Long-term Model Performance**

**To maintain model accuracy and reliability, we propose:**

1. **Periodic Model Retraining: Regularly retraining the model with updated loan application data.**
2. **Automated Monitoring: Implementing dashboards to detect performance drops in key metrics.**
3. **Adaptive Thresholding: Adjusting classification thresholds dynamically based on evolving patterns.**
4. **Human-in-the-loop Review: Allowing loan officers to override predictions when necessary and collect feedback for improvement.**

**By implementing these measures, the model can sustain high performance in real-world applications.**

**Task 7: Presentation (2/27)**

* **Prepare a 7-minute in-class presentation.**
* **Highlight improvements over current manual approval processes.**
* **Provide a live demonstration of the Streamlit tool.**

**Task 8: Submission (2/24 at 10 PM)**

* **Submit a 7-page report summarizing the project.**
* **Include all code and a section on Generative AI usage.**

**This document now fully aligns with the project tasks and serves as a structured foundation for both the final report and presentation. 🚀**