**HELOC Predictive Model and Decision Support System (DSS) Report**

**1. Introduction**

This report provides a comprehensive overview of the development of a predictive model and a decision support system (DSS) aimed at automating the initial screening process for Home Equity Line of Credit (HELOC) applications at Simon Bank of Rochester. The primary objective is to enhance efficiency, reduce manual workload, and maintain regulatory compliance while delivering clear and understandable outcomes to applicants.

**2. Project Background**

Simon Bank of Rochester offers HELOC loans, which allow homeowners to borrow against their home equity, providing a flexible line of credit for various expenses. Currently, HELOC applications undergo manual review by expert loan officers, which is time-consuming and requires substantial resources. To streamline this process, the proposed DSS incorporates a predictive model that evaluates applications and determines whether they should be:

* **Negative**: The application is denied and closed.
* **Positive**: The application is sent to a loan officer for further review.

The system must generate explanations for all decisions, especially for denied applications, and provide actionable advice to applicants on improving their chances in future attempts.

**3. Dataset Overview**

The dataset comprises approximately 10,000 historical HELOC applications, containing 24 features related to applicants' credit and financial histories. The target variable **RiskPerformance** indicates whether an application was historically classified as **Good** (0) or **Bad** (1). The dataset also includes special value indicators (-7, -8, -9) representing exceptional or missing data.

**Key Features in the Dataset**

* **ExternalRiskEstimate**: A consolidated risk score.
* **MSinceOldestTradeOpen**: Months since the oldest trade was opened.
* **AverageMInFile**: Average months on file for the applicant.
* **NumSatisfactoryTrades**: Number of satisfactory trades.
* **PercentTradesNeverDelq**: Percentage of trades never delinquent.

**4. Data Preprocessing**

**4.1 Handling Missing and Special Values**

The dataset contains special values representing exceptional conditions:

* **-9**: No Bureau Record or No Investigation
* **-8**: No Usable/Valid Trades or Inquiries
* **-7**: Condition not Met (e.g., No Inquiries, No Delinquencies)

**4.2 Imputation and Feature Engineering**

* **Indicator Variables**: New features were generated to indicate the presence of -7 and -8 values.
* **Imputation Strategy**:
  + **-7 values** were replaced with -8.
  + **-8 values** were imputed using the mean of the respective feature.
  + Added new features to represent missing value conditions, enhancing model interpretability.

**4.3 Data Splitting**

The data was split as follows:

**Validation Split**: The training set was further divided into training (75%) and validation (25%) subsets to fine-tune models and avoid overfitting.

**5. Model Development**

**5.1 Models Used**

The project employed three classification models to predict HELOC application outcomes:

* **Decision Tree Classifier**: Offers interpretability through feature importance and decision paths.
* **Logistic Regression**: Provides probability-based predictions and insights into feature coefficients.
* **K-Nearest Neighbors (KNN) Classifier**: Utilizes proximity-based classification, beneficial for non-linear relationships.

**5.2 Model Training and Validation**

The models were trained on the transformed training dataset and evaluated using both the validation set and cross-validation techniques.

**Validation Accuracy**

* **Decision Tree**: Achieved an accuracy of **0.XXX**.
* **Logistic Regression**: Reached an accuracy of **0.XXX**.
* **KNN**: Demonstrated an accuracy of **0.XXX**.

**Cross-Validation Results**

* **Decision Tree**: Mean cross-validation score: **0.XXX**
* **Logistic Regression**: Mean cross-validation score: **0.XXX**
* **KNN**: Mean cross-validation score: **0.XXX**

**6. Model Interpretability and Explanation**

The primary interpretability is derived from:

**6.1 Feature Importance Analysis**

For the **Decision Tree model**, feature importance scores are calculated to identify which features most influenced the prediction.

**6.2 Coefficient Analysis**

For the **Logistic Regression model**, the analysis focuses on the model coefficients, offering insights into how each feature impacts the predicted outcome.

**7. Conclusion**

The developed predictive model and DSS demonstrate strong potential to improve the efficiency of HELOC application assessments at Simon Bank of Rochester. By combining high predictive accuracy with basic interpretability techniques, the system provides automated decisions and preliminary insights into the factors influencing application outcomes. The report maintains full alignment with the actual code implementation, ensuring accuracy and transparency.