Explanation of the Machine Learning Task

The script SK.py builds a machine learning model for predicting loan defaults using the **Gradient Boosting Classifier**. The script follows a structured approach to the machine learning pipeline, which includes:

1. **Data Understanding and Preprocessing**:
   * The dataset is loaded, and basic data inspection is performed using df.info(), df.describe(), and df.isnull().sum() to check for missing data and the overall structure of the dataset.
   * Missing data is handled by imputing missing values for numeric columns using the mean and for categorical columns using the most frequent value.
   * Numeric features are scaled using StandardScaler, and categorical features are encoded using OneHotEncoder.
2. **Exploratory Data Analysis (EDA)**:
   * Visualizations are used to understand the relationships between features. A correlation heatmap is plotted to visualize correlations between numerical features.
   * Histograms are generated for each numerical feature to understand their distributions.
3. **Model Building**:
   * The data is split into training and testing sets (80-20 split).
   * A pipeline is built that combines preprocessing steps (imputing missing values, scaling, and encoding) with a **Gradient Boosting Classifier**.
   * Hyperparameter tuning is performed using **GridSearchCV** to optimize the number of estimators, learning rate, and maximum depth of the classifier.
4. **Model Evaluation**:
   * The model is evaluated on the test set using performance metrics such as accuracy, precision, recall, F1 score, and ROC-AUC.
   * The best hyperparameters found using GridSearchCV are displayed, and predictions are made on the test set.
5. **Feature Importance**:
   * Feature importance is calculated from the Gradient Boosting model, and a bar plot is generated to visualize which features contribute most to the predictions.
6. **Documentation**:
   * The script emphasizes the need for detailed documentation in the form of a README.md file, explaining the model and the process.