Homework 2 Overview

You are tasked with creating two classifiers, **SVM (Support Vector Machine)** and **KNN (K-Nearest Neighbours)**, using the dataset `loan_approval.csv`. ***(Please note that this file requires EDA (Exploratory Data Analysis) and data cleansing before it can be used.)***

After building the models, evaluate their performance using the following metrics:

1. Accuracy
2. Precision
3. Recall
4. F1-Score
5. Confusion Matrix

Step-by-Step Instructions

1. **Set Up Your Google Colab Notebook**
   - Open Google Colab at https://colab.research.google.com/.
   - Create a new notebook and rename it to `Loan_Approval_Classification`.
2. **Load the Dataset**
   - Download the provided `loan_approval.csv` file to your local machine.
   - Upload the file to Colab using the following steps:
     - Click on the folder icon in Colab (left sidebar).
     - Click on the upload icon and select your file.
   - Use the following code to load the dataset into a pandas DataFrame:

```
import pandas as pd
df = pd.read_csv('loan_approval.csv')
```

3. **Explore and Preprocess the Data**
   - View the dataset structure using:

```
df.head()  # View the first 5 rows
df.info()  # Check for missing values and data types
```

   - Handle missing values:
     - Decide how to fill or drop missing values. For example:

```
df.fillna(df.mean(), inplace=True)  # Example for numerical columns
```

   - Encode categorical variables:
     - Use `pd.get_dummies()` or `LabelEncoder` for categorical features.

```
from sklearn.preprocessing import LabelEncoder
le = LabelEncoder()
df['Category_Column'] =
le.fit_transform(df['Category_Column'])
```

   - Split the dataset into features and labels:

```
X = df.drop('Target_Column', axis=1)
y = df['Target_Column']
```

   - Split into training and testing sets:

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)
```

4. **Build Classifiers**
    - **SVM Classifier**:
        - Import and train the model:

```
from sklearn.svm import SVC
svm_model = SVC()
svm_model.fit(X_train, y_train)
```

        - Make predictions:

```
y_pred_svm = svm_model.predict(X_test)
```

    - **KNN Classifier**:
        - Import and train the model:

```
from sklearn.neighbors import KNeighborsClassifier
knn_model = KNeighborsClassifier()
knn_model.fit(X_train, y_train)
```

        - Make predictions:

```
y_pred_knn = knn_model.predict(X_test)
```

5. **Evaluate the Models**
    - Import the required metrics:

```
from sklearn.metrics import accuracy_score, precision_score,
recall_score, f1_score, confusion_matrix
```

    - Calculate and print the metrics for each model:

```
# Example for SVM
print("SVM Metrics:")
print("Accuracy:", accuracy_score(y_test, y_pred_svm))
print("Precision:",      precision_score(y_test,      y_pred_svm,
average='binary'))
print("Recall:",         recall_score(y_test,         y_pred_svm,
average='binary'))
print("F1-Score:",          f1_score(y_test,          y_pred_svm,
average='binary'))
print("Confusion       Matrix:\n",       confusion_matrix(y_test,
y_pred_svm))
```

6. **Submission Requirements**
    - **Code File**: Download your `.ipynb` file:
        - Go to **File > Download > Download .ipynb** in Colab.
    - **PDF Screenshot**: Take a full-page screenshot of your notebook using the **GoFullPage** Chrome extension:
        - Install GoFullPage from the Chrome Web Store.
        - Use it to capture the full notebook view and save it as a PDF.
7. **File Naming**
    - Name your files as follows:
        - `[YourName]_Loan_Approval_SVM_KNN.ipynb`
        - `[YourName]_Loan_Approval_SVM_KNN.pdf`
8. **Submission Platform**
   Upload the `.ipynb` and PDF files to the MSTeam Assignment by the deadline.

Deadline: Submit by **19/01/2025 – 23.59 (11.59 PM)**.