**Predicting Banking Crises Using Economic Indicators**

**Project Overview**

In this project, we aimed to build a machine learning model to predict banking crises using a dataset that includes various economic indicators for different countries across several years. Our goal was to use logistic regression to predict the binary target variable banking\_crisis (whether a banking crisis occurred or not) based on features such as inflation rates, exchange rates, debt defaults, and other macroeconomic indicators.

**Step-by-Step Process**

**Step 1: Importing Necessary Libraries**

We started by importing essential libraries for data analysis, visualization, and machine learning, such as:

python

Copy code

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LogisticRegression

from sklearn.metrics import classification\_report, confusion\_matrix, accuracy\_score

**Step 2: Data Preprocessing**

The dataset was loaded into a Pandas DataFrame, and we performed initial data inspection:

* **Missing Data**: We confirmed there were no missing values in the dataset.
* **Feature Selection**: We identified the features most relevant to predicting banking crises: currency\_crises, inflation\_crises, systemic\_crisis, exch\_usd, inflation\_annual\_cpi, and domestic\_debt\_in\_default.
* **Target Encoding**: The target variable banking\_crisis, which contained values crisis and no\_crisis, was mapped to binary values 1 and 0 respectively.

python

Copy code

data['banking\_crisis'] = data['banking\_crisis'].map({'crisis': 1, 'no\_crisis': 0})

**Step 3: Feature and Target Variables**

We selected the features and target variable for the model:

python

Copy code

features = ['currency\_crises', 'inflation\_crises', 'systemic\_crisis', 'exch\_usd', 'inflation\_annual\_cpi', 'domestic\_debt\_in\_default']

target = 'banking\_crisis'

X = data[features]

y = data[target]

**Step 4: Train-Test Split**

We split the data into training and test sets using an 80-20 ratio. The model was trained on the training set and evaluated on the test set.

python

Copy code

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

**Step 5: Model Training**

We used **Logistic Regression** for the initial model. Logistic regression is a simple yet powerful algorithm for binary classification tasks.

python

Copy code

model = LogisticRegression()

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

**Step 6: Model Evaluation**

After training the model, we evaluated its performance using accuracy, classification report, and confusion matrix. The confusion matrix showed how well the model predicted the banking crises:

plaintext

Copy code

[[192 1]

[ 6 13]]

* **Accuracy**: 96.7%
* **Precision for Crisis (1)**: 0.93
* **Recall for Crisis (1)**: 0.68
* **F1-Score for Crisis (1)**: 0.79

**Step 7: Key Insights**

From the confusion matrix, we observed the following:

* The model is performing very well at predicting "no crisis" (class 0), with only 1 false positive.
* However, the model is less effective at identifying banking crises (class 1), with 6 false negatives.
* This discrepancy in performance suggests the model might benefit from additional tuning or different techniques, especially since the class distribution is imbalanced.

**Conclusions and Next Steps**

The model performed quite well with a high accuracy of 96.7%, but there’s room for improvement, especially in detecting banking crises. To improve the model, we can consider the following steps:

1. **Handling Class Imbalance**: The dataset may have an imbalanced distribution of classes (crisis vs no\_crisis). Techniques such as SMOTE (Synthetic Minority Over-sampling Technique) can be used to balance the dataset and improve recall for the minority class.
2. **Trying More Advanced Models**: Logistic regression worked well but may not be the best model for this task. Exploring more advanced models like **Random Forests** or **XGBoost** could improve the prediction performance.
3. **Hyperparameter Tuning**: Tuning the hyperparameters of the logistic regression model (like regularization strength) or other models could help increase performance, particularly in terms of recall for the minority class.

**Final Thoughts**

This project demonstrates how logistic regression can be applied to predict banking crises using economic indicators. Although the results were promising, there are several areas for improvement that could lead to better performance and more accurate predictions.