**FINANCIAL CRISES, EXCHANGE RATE VOLATILITY, AND SOCIOECONOMIC IMPACT**

Rohith Reddy Jonnalagadda, Allu Sagarika

Valluru Anjula Chakravarthy

University of North Texas

**Acknowledgement**

I would like to express my sincere gratitude to everyone who supported me throughout the course of this research project. I extend my heartfelt thanks to Dr. Sam Shamroukh, whose guidance, expertise, and constructive feedback were invaluable at every stage of this study.

Special thanks to Raahul Raj Akula and Bhavana Raj Rayapudi for providing necessary environment to carry out this research. I am also grateful to my peers and colleagues for their insightful discussions and constant encouragement.

**ABSTRACT**

Financial crises have had profound and recurring impacts on global economies, yet predicting their occurrence remains a complex challenge due to the non-linear nature of macroeconomic indicators and the rarity of crisis events. This study aims to develop a data-driven early warning system capable of forecasting multiple types of financial crises including banking, currency, inflation, and systemic crises using machine learning techniques.

To achieve this, historical crisis data from the Reinhart-Rogoff Global Financial Crisis dataset was merged with macroeconomic indicators retrieved from the International Monetary Fund (IMF), resulting in a comprehensive dataset covering 70 countries from 1980 to 2016. Exploratory Data Analysis (EDA) and statistical assumption testing revealed significant non-normality and weak linearity in most features, guiding the selection of non-parametric models for classification. Multi-output classification models were employed, including Random Forest, K-Nearest Neighbors (KNN), and XGBoost.

Among the models tested, XGBoost emerged as the best-performing classifier, achieving high accuracy and F1-scores across all crisis types, particularly excelling in predicting inflation and systemic crises. The findings underscore the potential of ensemble-based machine learning models to detect early signs of economic distress and offer actionable insights for policymakers and financial institutions.

This research contributes to the integration of macroeconomic theory with modern AI tools, providing a foundation for scalable and real-time financial crisis monitoring systems. Future work may explore deep learning models and incorporate additional temporal or sentiment-based data sources to enhance predictive performance

Keywords : Financial Crises, Crisis Prediction, Machine Learning, Multi-Output Classification, Macroeconomic Indicators, XGBoost, Random Forest, IMF Data, Early Warning Systems

**Table of Contents**

[1. Introduction 6](#_Toc195457338)

[1.1 Problem Statement 6](#_Toc195457339)

[1.2 Description of the Global Financial Crisis Dataset (1800 - Present) 6](#_Toc195457340)

[1.2.1 Crisis Classification & Economic Events. 6](#_Toc195457341)

[1.2.2 Temporal Information. 6](#_Toc195457342)

[1.2.3 Geographical Information. 7](#_Toc195457343)

[1.2.4 Macroeconomic Indicators. 7](#_Toc195457344)

[1.2.5 Sovereign Debt & Government Finance. 7](#_Toc195457345)

[1.2.6 Additional Economic Data. 7](#_Toc195457346)

[1.3 Data Source Link: 7](#_Toc195457347)

[1.4 Why that specific source of data? 7](#_Toc195457348)

[1.5 Project Goals and Tools 8](#_Toc195457349)

[1.6 Tools 8](#_Toc195457350)

[1.7 Research Question: 9](#_Toc195457351)

[2. Literature Review 10](#_Toc195457352)

[2.1 Financial Crises: Patterns, Indicators, and Historical Trends 10](#_Toc195457353)

[2.2 The Dollar Exchange Rate as a Global Risk Factor: Evidence from Investment 10](#_Toc195457354)

[2.3 Foreign Currency Debt, Financial Crises, and Economic Growth: A Long-Run Perspective 11](#_Toc195457355)

[2.4 Are Leading Indicators of Financial Crises Useful for Assessing Country Vulnerability? Evidence from the 2008-09 Global Crisis 12](#_Toc195457356)

[2.5 Global Risk and the Dollar 12](#_Toc195457357)

[2.6 Research Gaps and what our study adds 13](#_Toc195457358)

[3. Methodology 14](#_Toc195457359)

[3.1 Data Source and Sample 14](#_Toc195457360)

[3.2 Variable Description 15](#_Toc195457361)

[3.3 Assumption Testing – Normality 17](#_Toc195457362)

[4. Exploratory Data Analysis 19](#_Toc195457363)

[4.1 Key findings from the summary statistics: 19](#_Toc195457364)

[4.2 Correlation Analysis 20](#_Toc195457365)

[4.3 Banking Crisis By Countries 22](#_Toc195457366)

[4.4 Trend Analysis Over Time 23](#_Toc195457367)

[4.5 Scatter Plot 24](#_Toc195457368)

[5. RESULTS 25](#_Toc195457369)

[5.1 XGBoost Classifier Performance 26](#_Toc195457370)

[5.2 Random Forest Classifier Performance 27](#_Toc195457371)

[5.3 K-Nearest Neighbors (KNN) Classifier Performance 27](#_Toc195457372)

[6. DISCUSSION 28](#_Toc195457373)

[7. Conclusion 30](#_Toc195457374)

[References 33](#_Toc195457375)

**Table of Figures**

[Figure 1: Data Description 15](#_Toc195457927)

[Figure 2: Normal Distribution Test (Normality Check) 18](#_Toc195457928)

[Figure 3: Summary Statistics 20](#_Toc195457929)

[Figure 4: Correlation Matrix 22](#_Toc195457930)

[Figure 5: Bar Graph 23](#_Toc195457931)

[Figure 6: Time-Series 24](#_Toc195457932)

[Figure 7: Scatter Plot 25](#_Toc195457933)

[Figure 8: Model Results 28](#_Toc195457934)

# 1. Introduction

## 1.1 Problem Statement

This research investigates historical financial crises across different countries, analysing patterns in sovereign debt defaults, banking crises, currency collapses, and inflation spikes. A key focus is on USD exchange rate fluctuations against global currencies and how these changes impact financial stability worldwide. The study evaluates long-term economic trends using the Reinhart-Rogoff Global Crisis Dataset, aiming to determine whether macroeconomic indicators, such as GDP growth, inflation, and exchange rate volatility can serve as early warning signals for financial instability. Additionally, this research explores strategies used by the United States to maintain its currency standard and global economic influence.

## 1.2 Description of the Global Financial Crisis Dataset (1800 - Present)

The Reinhart-Rogoff Financial Crisis Dataset contains 15,191 records covering more than 70 countries from 1800 to the present. The dataset tracks various types of financial crises, including banking crises, currency crises, sovereign debt defaults, and inflation crises. It provides a broad historical perspective on macroeconomic fluctuations and financial stability.

Key Features of the Dataset

### 1.2.1 Crisis Classification & Economic Events.

* Banking Crisis – Indicator for systemic banking failures.
* Currency Crisis – Instances of significant currency depreciation.
* Sovereign Debt Default – Government default or restructuring events.
* Inflation Crisis – High inflation or hyperinflation periods.

### 1.2.2 Temporal Information.

* Year – The year of the recorded economic event.
* Gold Standard – Indicator of whether a country was on the gold standard at that time.

### 1.2.3 Geographical Information.

* Country – The country experiencing the crisis.
* CC3 Code – Three-letter country code.

### 1.2.4 Macroeconomic Indicators.

* Exchange Rate (exch\_usd) – Country’s currency value against the USD.
* GDP-Weighted Default Rate – Default rates adjusted for economic size.
* Inflation Rates – Annual consumer price inflation percentages.

### 1.2.5 Sovereign Debt & Government Finance.

* Domestic Debt in Default – Status of domestic debt default.
* External Debt Defaults – Frequency and restructuring events of external debt.

### 1.2.6 Additional Economic Data.

* Independence – Indicator of whether the country was independent during the recorded period.
* Conversion Notes – Annotations related to exchange rate transformations.

1.3 Data Source Link:  
[Reinhart-Rogoff Historical Financial Crisis Dataset](https://www.reinhartandrogoff.com/)

1.4 Why that specific source of data?

The Reinhart-Rogoff Financial Crisis Dataset (1800-present) is a comprehensive resource that tracks various types of financial crises, including banking crises, currency crises, sovereign debt defaults, and inflation crises across over 70 countries. With more than 15,000 records, it offers a unique longitudinal perspective spanning over two centuries, making it invaluable for studying long-term trends in financial instability. The dataset includes macroeconomic indicators such as exchange rates (exch\_usd), GDP-weighted default rates, and inflation rates, providing critical insights into how crises affect national and global economies. It also offers data on sovereign debt defaults, including domestic and external debt, and includes temporal and geographical information to facilitate detailed country-level analysis. The inclusion of variables like the gold standard and country independence allows researchers to explore the historical evolution of financial systems. Given its credibility and the expertise of Carmen Reinhart and Kenneth Rogoff, this dataset is widely used for analysing financial crises and their long-term economic impacts, making it a vital tool for both academic research and policy analysis.

## **1.5 Project Goals and Tools**

The main purpose of this research is to explore historical financial crisis events alongside their relevant economic indicators and warning signs.

The research examines how different countries with different periods experience debt defaults by their governments and banking breakdowns as well as currency devaluation and volatile inflation rates.

Examine the role that changes in the USD exchange rate play regarding worldwide economic stability.

The research examines if economic indicators including GDP growth together with inflation rates and exchange rate volatility behaviour can alert to financial instability before it happens.

The United States implements various approaches to sustain its status as the chief monetary standard along with its domination in global economics.

## **1.6 Tools**

A complete financial crisis dataset from Reinhart-Rogoff Global Crisis contains data about banking crises along with currency crises and sovereign debt defaults and inflation crises. The above-mentioned dataset serves as the primary information base for the present research investigation. The data cleaning operations together with statistical analysis rely on Pandas Python library combined with NumPy and SciPy packages for data preprocessing. The analysis of data visualization will incorporate both Matplotlib with Seaborn packages. Both statistical calculations and data handling functions within R operate through the combination of tidyverse and dplyr packages. The SQL tool manages, and queries structured datasets located in databases.

The cleaning operations for exploratory evaluations will utilize Excel as a data processing tool.

The chosen tools deliver general functionality plus friendly interfaces along with strong processing features. Through the use of these tools the research can achieve a complete financial crisis analysis and develop early economic instability warning systems.

## 1.7 Research Question:

What are the key economic indicators that signal an impending financial crisis across different countries and time periods, and how does the USD exchange rate impact global economic stability? (Financial Crisis Patterns, USD Exchange Rates, and Macroeconomic Indicators)

Sub-variables:

* How do inflation rates, GDP, and exchange rates fluctuate before a banking crisis?
* Do sovereign debt defaults follow a predictable pattern based on GDP growth?
* How does the USD perform against other global currencies during financial crises?
* Are currency crises more frequent in developing economies compared to developed ones?

# 2. Literature Review

## 2.1 Financial Crises: Patterns, Indicators, and Historical Trends

This is a very broad analysis based on the paper This Time is Different: A Panoramic View of Eight Centuries of Financial Crises written by Reinhart and Rogoff (2008). Unlike machine learning models, the study does not use but rather analyses historical data using approaches and statistical methodologies to assess economic downturns. The paper describes financial crises spanning more than 800 years and recognizes recurring patterns in sovereign default on debt, banking crises, peaks in inflation, and currency crises. It emphasizes how financial instability is cyclical in nature, more often than not followed by large capital mobility, exogenous shocks, and policy errors (Reinhart & Rogoff, 2008). The research utilizes historical data analysis and macroeconomic trend evaluation but not predictive model techniques like machine learning or real-time early warning systems. Though the research gives useful information about financial crisis tendencies, future research can utilize AI-based prediction models like Random Forest and LSTMs to further improve the predictability and create automated financial crisis monitoring systems. This synthesis would close the gap between predictive analytics and past economic studies and provide forward-facing solutions to reduce financial instability.

## 2.2 The Dollar Exchange Rate as a Global Risk Factor: Evidence from Investment

The article by Avdjiev et al. (2018) discusses how US dollar movements affect investment flows around the world. The article looks at the "financial channel" of exchange rates, and we learn that a stronger dollar reduces dollar cross-border bank lending and decreases real investment in emerging market economies (EMEs). Using both macroeconomic (country-level) and microeconomic (firm-level) specifications, the authors use Structural Panel Vector Autoregressions (SPVAR) and firm-level panel regressions to estimate these relations. The results indicate that dollar appreciation squeezes global financial conditions, affecting investment more than conventional trade effects (Avdjiev et al., 2018). The paper employs models like the "risk-taking channel" (Bruno & Shin, 2015), capital flows models, and financial frictions models without employing AI or machine learning-based predictive modeling. The paper identifies the prevalence of the financial channel over the trade channel in international investment flows, noting gaps in real-time crisis prediction and the application of machine learning approaches in predicting financial risk.

## 2.3 Foreign Currency Debt, Financial Crises, and Economic Growth: A Long-Run Perspective

The article by Bordo et al. (2010) discusses the contribution of foreign currency debt to financial crises and economic growth in two separate eras: 1880-1913 and 1973-2003. The study concludes that a larger proportion of foreign currency debt relative to total debt exposure raises the vulnerability to currency as well as debt crises, especially in nations that have fragile financial systems and low levels of reserves (Bordo et al., 2010). It sees that financial crises triggered by foreign currency exposure will result in substantial long-term losses in output. Empirical probability models, trend analysis of macroeconomic variables, and historical data comparisons are the tools used by the research work to reach such conclusions. It also develops theoretical models like IS-LM models with capital market frictions and balance sheet-based transmission mechanisms for describing how currency mismatch amplifies financial instability. That being said, it does not employ machine learning and predictive modelling as a means of crisis forecasting. Its findings emphasize that though lower foreign currency exposure can minimize crisis risk, also crucial is the presence of strong financial institutions together with policy credibility for the buffer against financial instability.

## 2.4 Are Leading Indicators of Financial Crises Useful for Assessing Country Vulnerability? Evidence from the 2008-09 Global Crisis

The article by Frankel and Saravelos (2012) examines whether leading economic indicators can forecast the severity of financial crises, and more particularly the 2008-09 crisis. The paper thoroughly reviews more than eighty previous studies on early warning indicators and concludes that central bank reserves and real exchange rate appreciation are two of the most robust predictors of exposure to crisis among nations (Frankel & Saravelos, 2012). Employing empirical evidence, the authors evaluate six indicators of crisis, such as decline in GDP, currency depreciation, stock market performance, loss of reserves, and IMF support, to check how the indicators performed in 2008-09. The study uses a range of econometric specifications, such as regression-based methods, exchange market pressure indices, and probit models to determine key predictors of crisis occurrence. The study emphasizes the role of international reserves in reducing crisis risk and concludes that conventional early warning indicators are still useful despite changing financial conditions. However, the study acknowledges gaps in forecasting crises in real time and calls for further research in integrating AI-driven predictive models for more effective crisis warning.

## 2.5 Global Risk and the Dollar

The article by Georgiadis et al. (2021) analyses spillovers of global risk shocks to the world economy with a special focus on the U.S. dollar's role in the international financial system. The paper uses a Bayesian Proxy Structural Vector Autoregression (BPSVAR) model in which high-frequency gold price surprises are employed to identify global risk shocks. The results show that global risk shocks result in a synchronized slowdown of global economic activity while, at the same time, triggering the appreciation of the U.S. dollar (Georgiadis et al., 2021). The appreciation exacerbates economic recessions in the rest of the world via a financial channel by tightening global financial conditions and lowering cross-border bank credit. The authors perform counterfactual simulations to show that, absent dollar appreciation, the adverse impacts of global risk shocks on economic activity are much less pronounced. The paper further emphasizes the offsetting roles of the trade channel and the financial channel with the latter playing the predominant role. Although the study offers strong empirical proof of risk transmission in the international economy, it does not employ machine learning or artificial intelligence-based predictive models and thus offers an opportunity for future research to scrutinize real-time mechanisms of financial crisis forecasting.

## 2.6 Research Gaps and what our study adds

While the reviewed literature provides a robust understanding of the historical dynamics, macroeconomic patterns, and theoretical frameworks related to financial crises, a notable gap persists in the application of real-time, data-driven prediction methods using machine learning. Most prior studies rely heavily on traditional econometric models and historical trend analysis, without incorporating modern predictive techniques that can handle large-scale, high-dimensional datasets. This research addresses that gap by integrating historical crisis data from the Reinhart-Rogoff dataset with macroeconomic indicators sourced from the IMF, enabling a more comprehensive and enriched analytical framework. By applying multi-output Random Forest models, we move beyond static, one-crisis-at-a-time analyses and demonstrate the feasibility of predicting multiple types of crises such as banking, inflation, and currency crises simultaneously. Additionally, our approach includes rigorous preprocessing, assumption testing, and uses panel data structures to capture time-dependent patterns. This synthesis of traditional economic theory and contemporary AI-driven methods contributes a novel early-warning framework for financial instability detection, bridging the methodological gap between historical macroeconomic research and applied machine learning.

# 3. Methodology

## 3.1 Data Source and Sample

This study utilizes a uniquely constructed dataset that merges the Reinhart-Rogoff Global Financial Crisis Dataset(1800–present) with complementary macroeconomic indicators retrieved from the International Monetary Fund (IMF). The merged dataset spans the years 1980 to 2023 and encompasses economic and crisis-related data for 70 countries, providing both temporal and geographical diversity for cross-sectional and longitudinal analysis. The final dataset consists of 2,520 observations across 22 variables, including both crisis indicators and macroeconomic fundamentals. Key dependent variables include binary classifications for Banking Crisis, Currency Crises, Systemic Crisis, and Inflation Crises, as well as sovereign debt default metrics. Independent variables consist of indicators such as GDP per capita, exchange rate against the USD (exch\_usd), inflation rates, unemployment rate, current account balance, trade volume metrics, and government fiscal positions. These variables enable a comprehensive analysis of the economic environment preceding financial crises. The dataset was cleaned and standardized to ensure consistency in variable naming, formatting, and scaling. Missing data were addressed using Multiple Imputation by Chained Equations (MICE) to preserve sample integrity while minimizing bias. The resulting dataset offers a rich basis for both exploratory analysis and predictive modeling, supporting the research objectives of identifying leading indicators of financial crises and assessing the systemic role of the USD in global economic stability.

A screenshot of a computer program

AI-generated content may be incorrect.

Figure : Data Description

## 3.2 Variable Description

IMF Macroeconomic Indicators (Manually Added)

Country – The name of the country for the given observation.

Year – The corresponding year of the data point.

Current account balance – Net exports and financial flows as a percentage of GDP.

General government net lending/borrowing – Fiscal surplus or deficit position of the government.

Gross domestic product per capita, current prices – GDP per person, measured in current US dollars.

Gross domestic product, current prices – Total economic output of a country, in current US dollars.

Inflation, average consumer prices – Annual inflation rate based on consumer price index (CPI).

Population – Total midyear population estimate for the country.

Unemployment rate – Percentage of the labor force that is unemployed.

Volume of exports of goods and services – Index measuring the quantity of exports (volume-based).

Volume of imports of goods and services – Index measuring the quantity of imports (volume-based).

*Crisis and Risk Indicators from Reinhart-Rogoff Dataset*

Banking Crisis – Binary indicator (0/1) denoting whether a banking crisis occurred.

Systemic Crisis – Binary flag indicating systemic financial crisis events.

exch\_usd – Exchange rate of local currency to US dollars.

Domestic\_Debt\_In\_Default – Indicator of whether domestic debt was in default.

Sovereign Debt 1 – Binary variable for default on external sovereign debt.

Sovereign Debt 2 – Binary variable for default on private sector sovereign debt.

GDP\_Weighted\_default – Weighted index of sovereign default events by GDP.

Inflation, Annual percentages of average consumer prices – Percentage change in CPI from the previous year.

Independence – Indicator of whether the country was politically independent in the given year.

Currency Crises – Binary (or categorical) indicator representing currency crisis events.

Inflation Crises – Binary flag for occurrences of extreme inflation episodes

Assumptions – Shapiro-Wilk Test

Prior to implementing classification models, a series of assumption tests were conducted to assess the suitability of the macroeconomic variables for predictive modeling. These tests focused on normality, multicollinearity, and linearity, which are commonly considered when selecting appropriate modeling strategies.

To begin, the Shapiro–Wilk test was applied to 12 continuous macroeconomic variables, including inflation rates, exchange rates, trade volumes, GDP indicators, and fiscal metrics. The results revealed that all variables had p-values below the 0.05 threshold, strongly rejecting the null hypothesis of normality. This indicates that the data does not follow a Gaussian distribution—a finding consistent with the characteristics of real-world economic data. Even variables with high W-statistics, such as *General government net lending/borrowing* (W = 0.9676), showed statistically significant non-normality. These findings support the use of non-parametric, distribution-free classification algorithms like Random Forest, which do not assume normality of the features.

Additionally, Variance Inflation Factor (VIF) scores were computed to test for multicollinearity among the independent variables. All VIF values were well below the standard cutoff of 5, indicating that there is no significant collinearity that could distort model interpretation or accuracy.

Lastly, scatter plots and regression line visualizations were used to assess linearity between features and target variables. The plots showed no strong linear relationships, further justifying the choice of non-linear, tree-based models for classification tasks. These results guided the research toward using multi-output Random Forest classifiers, which are robust to these assumption violations and effective in capturing non-linear, complex patterns in multidimensional datasets.

## 3.3 Assumption Testing – Normality

Normality testing was conducted on continuous predictor variables to assess their suitability for modeling, particularly to determine whether parametric methods like logistic regression would be appropriate. The Shapiro–Wilk test, a widely accepted formal test for normality, was applied to 12 macroeconomic variables including GDP per capita, inflation, exports, and fiscal indicators.

The results revealed that nearly all variables significantly deviated from a normal distribution, with p-values less than 0.05, and in most cases approaching zero. This statistical evidence was further supported by distribution plots, which showed strong skewness, kurtosis, and outliers in many variables. For instance, Inflation (Annual Percentages) and GDP per Capita showed severe right skew, while others like Government Net Lending/Borrowing only approximated a bell shape but failed the formal normality test.

Given the widespread violation of the normality assumption, it was determined that parametric models would not be appropriate for this dataset. As a result, the study proceeded with non-parametric, tree-based classification models such as Random Forest, which do not rely on distributional assumptions and are more robust to skewed and non-linear data.

A graph of a graph

AI-generated content may be incorrect.A graph of blue bars

AI-generated content may be incorrect.A white rectangular object with black text

AI-generated content may be incorrect.A graph of a number of people

AI-generated content may be incorrect.

Figure : Normal Distribution Test (Normality Check)

# 4. Exploratory Data Analysis

A descriptive statistical analysis was conducted on all continuous variables to understand the central tendencies, variability, and distributional properties of the dataset. The dataset consists of 2,359 observations across 22 variables, including macroeconomic indicators and financial crisis flags across 70 countries from 1980 to 2016.

## 4.1 Key findings from the summary statistics:

Current Account Balance ranges from –816.65 to 420.57, with a mean near zero (–0.48) but a high standard deviation of 57.67, indicating wide dispersion and possible outliers.

General Government Net Lending/Borrowing also shows a left-skewed distribution, with a minimum of –32.12 and a mean of –2.08, reflecting common fiscal deficits in many countries.

Gross Domestic Product (GDP) values vary widely:

Per capita GDP ranges from 238.50 to 103,215 USD, with a highly skewed distribution (mean = 12,746; median = 4,540).

Total GDP ranges from near-zero to 18,804 billion USD, again showing right skewness and high variance.

Inflation (average consumer prices) has a maximum value exceeding 13,000%, suggesting the presence of hyperinflation in specific countries/years. The large standard deviation (474.84) further confirms high variability.

Unemployment Rate ranges from 0.04% to over 40%, indicating strong variation across time and regions.

Trade Volume Indicators (exports and imports) also show substantial variance, with some countries experiencing dramatic shifts in trade activity (e.g., export volumes ranging from –60.40 to 649.15).

Exchange Rate (exch\_usd) includes values as high as 1.92 million, reflecting historical devaluations or hyperinflation episodes in some economies.

A computer screen with numbers and numbers

AI-generated content may be incorrect.A screenshot of a computer

AI-generated content may be incorrect.A screenshot of a computer screen

AI-generated content may be incorrect.A computer screen shot of numbers and symbols

AI-generated content may be incorrect.A computer screen shot of numbers

AI-generated content may be incorrect.A screen shot of a black screen

AI-generated content may be incorrect.

Figure : Summary Statistics

## 4.2 Correlation Analysis

To better understand the relationships between numerical features and potential multicollinearity within the dataset, a Pearson correlation heatmap was generated. The heatmap reveals that most variables exhibit weak to moderate linear correlations, with few strong relationships observed.

There is a moderately positive correlation (0.70) between Banking Crisis and Systemic Crisis, which aligns with economic intuition as systemic events often coincide with banking collapses.

Currency Crises and Inflation Crises show a moderate correlation (0.47), suggesting that periods of extreme inflation may often coincide with currency instability.

Variables such as Gross Domestic Product per capita and Exchange Rate (exch\_usd) also demonstrate modest positive relationships, indicating potential economic scale effects.

Most economic indicators like Inflation, Unemployment, and Trade Volumes have low pairwise correlations, indicating that these features are relatively independent and unlikely to cause multicollinearity issues in model training.

Interestingly, the target variables (such as Banking Crisis, Currency Crises, and Inflation Crises) appear to have low correlations with most numerical features, reinforcing the necessity of using non-linear models capable of capturing complex, non-obvious patterns in the data.

This correlation structure supports the choice of tree-based classification models, which are not sensitive to multicollinearity and can capture non-linear interactions between weakly correlated predictors.

A screenshot of a graph

AI-generated content may be incorrect.

Figure : Correlation Matrix

## 4.3 Banking Crisis By Countries

The bar chart illustrates the total number of banking crises experienced by each country in the dataset. **Zimbabwe, Central African Republic**, and **Republic of Congo** lead with the highest number of crises, each exceeding 13 events, indicating severe financial instability over time. Several developed economies like Sweden, Italy, and UnitedStates also report frequent banking crises, highlighting that such events are not exclusive to developing nations. This distribution reveals the widespread nature of banking vulnerabilities across both emerging and advanced economies.

A graph of a number of banking crisis

AI-generated content may be incorrect.

Figure : Bar Graph

## 4.4 ****Trend Analysis Over Time****

The time series plots highlight the frequency trends of four types of financial crises between 1980 and 2016. Banking Crises and Systemic Crises show a noticeable rise in the 1990s, peaking around the 1997–1998 Asian Financial Crisis, followed by a sharp decline in the early 2000s. Inflation Crises were more common in the 1980s and early 1990s but dropped significantly after 2000, reflecting global stabilization policies and better inflation control. Currency Crises display a highly volatile trend with multiple peaks, notably around 1985, 1990, and 2008. Post-2005, most crisis types exhibit a declining trend, likely due to improved fiscal and monetary regulations in many countries. A small resurgence in crises can be observed during the 2008 Global Financial Crisis. Overall, the patterns suggest that while financial crises remain cyclical, their intensity and frequency have moderated over time. These visualizations support the temporal context for building predictive models.

A graph showing the growth of a financial crisis

AI-generated content may be incorrect.A graph showing the time of the year

AI-generated content may be incorrect.A graph showing the growth of the economy

AI-generated content may be incorrect.A graph showing the currency crisis

AI-generated content may be incorrect.

Figure : Time-Series

## 4.5 Scatter Plot

The scatterplot highlights important trends in the interplay between a country’s economic output and labor market conditions in the context of financial crises. A dense clusterof observations is seen at the lowerendofGDP, where unemployment rates show significant variability ranging from near 0% to over 40%. This suggests that low-GDP economies exhibit greater volatility inemployment, potentially due to structural weaknesses, lack of social safety nets, or dependency on unstable sectors.

As GDP increases, unemploymentratesgenerallystabilize and converge around 5–10%, reflecting stronger and more resilient economies. Interestingly, bankingcrises **(**markedinorange) are more prevalent in countries with low to mid-rangeGDP and moderate to high unemployment, which may indicate vulnerabilities due to credit constraints, under-regulation, or weak governance.

At higher GDP levels, banking crises are less frequent, though not absent — implying that even advanced economies are not immune to systemic shocks but may possess better mechanisms to contain their fallout. Overall, the scatterplot suggests a non-linear andasymmetricrelationship between macroeconomic health and banking system stability, reinforcing the need for non-parametric modeling techniques in crisis prediction.

A graph with blue and orange dots

AI-generated content may be incorrect.

Figure : Scatter Plot

# 5. RESULTS

The primary objective of this study was to assess the capability of machine learning models in predicting multiple types of financial crises, specifically Banking Crises, Currency Crises, and Systemic Crises. To accomplish this, we employed a multi-output classification approach using three models: XGBoost, Random Forest, and K-Nearest Neighbors (KNN). These models were trained on a dataset enriched by merging the Reinhart-Rogoff Global Financial Crisis dataset with macroeconomic indicators from the International Monetary Fund (IMF). The dataset comprised 2,359 observations from 70 countries covering the period from 1980 to 2016, with 18 predictive features and three binary crisis indicators as target variables. Standard evaluation metrics such as accuracy, precision, recall, and F1-score were used to gauge performance.

Among the three models, XGBoost delivered the best overall results. It performed especially well in predicting Systemic Crises, achieving an accuracy of 91 percent, and demonstrated solid performance for Currency Crises and Banking Crises, with accuracy scores of 86 percent each. The model’s gradient boosting mechanism allowed it to effectively focus on hard-to-classify instances, making it particularly suitable for the detection of rare and complex events in economic data. Its robustness to non-linearity, class imbalance, and high-dimensional features reinforced its effectiveness in crisis prediction.

The Random Forest classifier also yielded strong results, with performance comparable to XGBoost in several areas. For Currency Crises, the model achieved 87 percent accuracy and a similarly high F1-score, nearly matching XGBoost’s effectiveness. In the cases of Banking Crises and Systemic Crises, the model displayed slightly lower recall, indicating that it missed a few crisis instances that XGBoost successfully identified. This reduction in sensitivity is likely due to Random Forest’s averaging mechanism across decision trees, which can dilute the model’s responsiveness to less frequent events. Despite this, Random Forest remains a reliable and interpretable model, making it a practical choice for understanding variable importance in economic forecasting.

In contrast, the K-Nearest Neighbors (KNN) model lagged behind the ensemble models across all evaluated crisis types. While it managed a respectable 86 percent accuracy in predicting Systemic Crises, its performance dropped for Banking Crises, with an accuracy of just 78 percent. More notably, KNN struggled with recall, failing to correctly identify a significant portion of actual crisis years. This limitation highlights the model’s vulnerability to issues like class imbalance, high dimensionality, and sensitivity to feature scaling—all common in macroeconomic datasets. Given these challenges, KNN proved less suitable for this application, especially where early detection of rare crisis events is critical.

Overall, these results confirm that ensemble methods—particularly XGBoost—are the most effective in capturing the complex, non-linear relationships that underlie financial crisis prediction. Their ability to handle noisy data and emphasize rare but important patterns makes them powerful tools for building early warning systems in economic contexts.

A screenshot of a screen

AI-generated content may be incorrect.

Figure : Model Results

# 6. DISCUSSION

This study set out to answer a fundamental question in global economics: What are the key economic indicators that signal an impending financial crisis across different countries and time periods, and how does the USD exchange rate impact global economic stability? Using machine learning models applied to macroeconomic and historical crisis data, we identified key indicators—such as exchange rate volatility, GDP per capita, sovereign debt exposure, and inflation trends—as critical signals of forthcoming financial crises. These indicators showed varying degrees of influence across different crisis types including banking, currency, and systemic crises.

An important observation emerged during model evaluation: the prediction of Inflation Crises using the XGBoost model achieved 100% accuracy, precision, and recall. While this might initially appear as a success, such a perfect score raised concerns regarding overfitting or potential data leakage—both of which are serious red flags in real-world predictive modeling. In practical terms, no model should perfectly predict rare economic events without some margin of error. Consequently, to maintain the integrity and realism of our study, we made the decision to exclude Inflation Crises from further interpretation and discussion. This allowed us to focus on the more plausible and generalizable findings from the remaining crisis categories.

Through our analysis, we confirmed that fluctuations in the USD exchange rate have a significant impact on global economic stability. Countries experiencing sharp depreciations against the dollar—especially those with high levels of foreign-denominated debt—were more susceptible to currency and systemic crises. This aligns with existing economic theories on the financial channel of exchange rate risk and highlights the dollar’s central role in amplifying global financial vulnerabilities.

Our multi-output classification models, particularly Random Forest and XGBoost, demonstrated strong capability in capturing non-linear relationships between economic indicators and crisis events. The ability to predict multiple crisis types simultaneously underscores the interconnected nature of financial instability and the shared macroeconomic patterns that often precede it.

In conclusion, this discussion confirms that certain macroeconomic indicators can indeed function as early warning signals of financial crises, and that USD exchange rate dynamics are central to understanding global financial health. By integrating historical insights with machine learning, this research contributes a scalable and evidence-based foundation for future crisis prediction systems aimed at supporting both economic policy and financial risk management.

# 

# 7. Conclusion

This research aimed to build a predictive framework for identifying multiple types of financial crises using macroeconomic indicators and machine learning techniques. By combining historical crisis data from the Reinhart-Rogoff Global Financial Crisis dataset with macroeconomic variables sourced from the IMF, we developed a multi-dimensional dataset that allowed for in-depth analysis across 70 countries and over three decades. The use of ensemble-based models like XGBoost and Random Forest enabled effective identification of patterns leading up to banking, currency, and systemic crises, thereby affirming the potential of data-driven early warning systems in economic policymaking.

Despite these strengths, several limitations in our study merit attention and offer directions for future work. First, the dataset used was annual in frequency, which limits the model’s responsiveness to short-term economic shocks or rapidly evolving crises. Incorporating higher-frequency data—such as quarterly or monthly economic indicators—could significantly enhance the temporal resolution and sensitivity of predictive models. Second, our models did not account for temporal dependencies or lag structures, which are often critical in understanding how economic crises evolve. Future efforts could integrate time series models such as LSTM or Transformer architectures to capture sequential dynamics and intertemporal risk buildup.

Third, although the dataset included a broad selection of countries, it treated all observations with equal weight. This overlooks structural differences between economies—such as size, development stage, or policy frameworks—which may influence how crises manifest. Future researchers could improve model robustness by stratifying or clustering countries based on economic characteristics or regional groupings. Additionally, behavioral and market sentiment data—such as news analytics, investor confidence indexes, or commodity price trends—were not included in this study. Incorporating such qualitative data could provide a more holistic view of crisis formation, especially in today's interconnected and media-driven financial landscape.

Lastly, the removal of Inflation Crises from the final model output—due to an unrealistic 100% accuracy score—highlights the importance of continuously validating model performance against real-world expectations. Future researchers using similar data should apply rigorous cross-validation techniques and guard against data leakage or overfitting when dealing with rare events.

In sum, while this study provides a strong foundation for predicting financial crises using historical and macroeconomic data, it also opens avenues for richer, more nuanced research. By addressing these limitations, future work can advance the development of intelligent, reliable, and context-aware early warning systems for global economic stability.

# 

# References

Avdjiev, S., Du, W., Koch, C., & Shin, H. S. (2018). The dollar exchange rate as a global risk factor: Evidence from investment. Brookings Papers on Economic Activity, 2018(2), 313–365.

Barroga, E., Matanguihan, G. J., Furuta, A., Arima, M., Tsuchiya, S., Kawahara, C., Takamiya, Y., & Izumi, M. (2023). Conducting and writing quantitative and qualitative research. Journal of Korean Medical Science, 38(37), Article e291.

Bordo, M. D., Meissner, C. M., & Stuckler, D. (2010). Foreign currency debt, financial crises, and economic growth: A long-run view. Journal of International Money and Finance, 29(4), 642–665.

Frankel, J. A., & Saravelos, G. (2012). Are leading indicators of financial crises useful for assessing country vulnerability? Evidence from the 2008–09 global crisis. Journal of International Economics, 87(2), 216–231.

Georgiadis, G., Müller, G. J., & Schumann, B. (2021). Global risk and the dollar (European Central Bank Working Paper Series No. 2628). European Central Bank.

Ioan, B., Malar Kumaran, R., Larissa, B., Anca, N., Lucian, G., Gheorghe, F., Horia, T., Ioan, B., & Mircea-Iosif, R. (2020). A panel data analysis on sustainable economic growth in India, Brazil, and Romania. Journal of Risk and Financial Management, 13(8), Article 170.

Lau, F., & Holbrook, A. (2017). Methods for comparative studies. In F. Lau & C. Kuziemsky (Eds.), Handbook of eHealth evaluation: An evidence-based approach (pp. 193–214). University of Victoria.

Li, P., Stuart, E. A., & Allison, D. B. (2015). Multiple imputation: A flexible tool for handling missing data. JAMA, 314(18), 1966–1967.

Li, Q. (2020). Overview of data visualization. In Embodying data (pp. 17–47). Springer.

Pearl, J. (2010). An introduction to causal inference. The International Journal of Biostatistics, 6(2), Article 7.

Reinhart, C. M., & Rogoff, K. S. (2008). This time is different: A panoramic view of eight centuries of financial crises (NBER Working Paper No. 13882). National Bureau of Economic Research.

Setia, M. S. (2016). Methodology series module 3: Cross-sectional studies. Indian Journal of Dermatology, 61(3), 261–264.

Sivakumar, G. (2025). HMM-LSTM fusion model for economic forecasting. arXiv.

Zapf, A., Wiessner, C., & König, I. R. (2024). Regression analyses and their particularities in observational studies: Part 32 of a series on evaluation of scientific publications. Deutsches Ärzteblatt International, 121(4), 128–134.