Combining Traditional Economics with Modern Computing: Toward a Predictive Framework for Financial Crises

Bîrcea Cristina-Maria[1], Halcea Alina-Oana[2] and Mărginean Daria-Maria[3*]

[1] Lucian Blaga University of Sibiu, Faculty of Economic Sciences, Sibiu, Romania, cristinamaria.bircea@ulbsibiu.ro
[2] Lucian Blaga University of Sibiu, Faculty of Economic Sciences, Sibiu, Romania, alinaoana.halcea@ulbsibiu.ro
[3] Lucian Blaga University of Sibiu, Faculty of Economic Sciences, Sibiu, Romania, dariamaria.marginean@ulbsibiu.ro
* Correspondence: dariamaria.marginean@ulbsibiu.ro

Abstract: This paper explores the intersection between traditional economic theory and modern computational tools in an effort to enhance financial crisis forecasting capabilities, with a specific focus on Romania. The unpredictability of economic crises and the complexity of their causes highlight the limitations of conventional models, which often fail to detect early warning signals. Recent advancements in data analytics and machine learning provide new opportunities to address these challenges by identifying complex patterns that are not easily captured by classical econometric techniques.

Therefore, the aim of this study was to design and evaluate a predictive framework that combines traditional economic indicators with data-driven algorithms to detect early signs of financial instability. The analysis relies on macroeconomic and financial data collected from national and European sources for the period 2000–2024. Key indicators such as GDP growth, inflation, credit to the private sector, and monetary aggregates are used to train classification models capable of distinguishing between crisis and non-crisis periods.

The methodology integrates both statistical and machine learning approaches, including Logistic Regression and Random Forest classifiers, implemented in Python. The data is preprocessed through normalization and labelling of crisis episodes, and model performance is assessed using standard evaluation metrics such as accuracy, recall, and area under the ROC curve. Preliminary results suggest that machine learning models, particularly ensemble methods, show promising accuracy in identifying crisis signals based on historical patterns.

This current research contributes to the ongoing efforts in economic forecasting by offering a practical framework that merges economic reasoning with computational power. While the study does not aim to produce definitive forecasts, it highlights the potential of predictive modelling in informing policy decisions and increasing economic resilience. Future work will expand the analysis by incorporating additional indicators and exploring the generalizability of the model in other economic contexts.

Key words: Financial Crisis, Predictive Modeling, Machine Learning, Macroeconomic Indicators, Romania

JEL classification: C53, E37, G01, G17

1. Introduction

The economic environment has long been marked by uncertainty and the persistent threat of financial instability. Recognizing and anticipating such risks is essential for both policymakers and stakeholders aiming to mitigate the impact of adverse economic events. Concepts such as financial crises, economic downturns, speculative bubbles, and systemic vulnerabilities have been the subject of intense academic debate and public discourse. Today, these issues remain as relevant as ever, particularly in the context of increasingly complex global financial systems.

Identifying the factors that trigger or contribute to financial crises remains a major challenge in economic research. These factors often result from a dynamic interplay of economic policies, institutional behavior, market psychology, and unpredicted global events. As such, fully characterizing the mechanisms behind financial crises proves to be an elusive task. However, advances in computational power and the sharp increase of data have opened new pathways to enhance traditional economic forecasting models.

In recent years, the integration of modern computing techniques, particularly machine learning and data-driven analysis, into economic research has shown promise in uncovering hidden patterns and early warning signals. These technological advancements can elevate classical economic theories by offering more flexible and adaptive frameworks for analysis. In this context, the present research proposes a hybrid approach that combines traditional economic reasoning with modern computational tools, aiming to build a predictive framework for financial crises.

Focusing specifically on Romania, the study explores the potential of using historical macroeconomic and financial data to detect early signs of economic instability. The research does not intend to produce absolute forecasts; rather, it seeks to illustrate the potential of Python-based predictive modelling to enhance our understanding of how crises emerge and evolve. The methodology includes selecting relevant indicators, implementing machine learning algorithms, and interpreting the results in an economic context.

The paper is structured as follows: the next section reviews relevant literature to frame the theoretical foundation of the study. This is followed by a detailed explanation of the research methodology, including data selection and modelling techniques. The results section presents the findings and discusses their implications, while the final part concludes the study with key insights, limitations, and suggestions for future research.

2. Literature Review

Over time, the study of financial crises has attracted considerable attention, particularly following major events such as the 1997 Asian financial crisis, the 2008 global financial crisis, and the COVID-19 induced economic disruptions. In fact, the widespread impact of crises highlights the need for a deep understanding of them, as it has been shown that financial turmoil can have significant consequences and can considerably influence the implementation of economic and financial policies (Claessens & Kose, 2013).

A wide body of literature has explored both the causes and the propagation mechanisms of financial instability. In the context of Romania and the broader Eastern European region, research has highlighted the vulnerabilities arising from transition economies, institutional fragilities, and exposure to external shocks (Duguleană, 2011; Niculescu, 2013; Pelinescu & Simionescu, 2017).

Traditional approaches to analyzing financial crises have predominantly relied on time series econometrics. Models such as Vector Autoregressions (VAR), Autoregressive Integrated Moving Average (ARIMA), and Probit/Logit models have been widely employed to capture dynamic relationships among macroeconomic indicators and to assess the probability of crisis occurrence (Kaminsky & Reinhart, 1999; Drehmann & Juselius, 2014). While these models have contributed significantly to understanding past crises, they often struggle to cope with nonlinearities, structural breaks, and evolving patterns present in modern financial systems.

In recent years, the adoption of data-driven techniques in economic research has grown rapidly. Machine learning (ML) models such as Decision Trees, Random Forests, Support Vector Machines (SVM), and Artificial Neural Networks (ANN) have been successfully applied to financial forecasting tasks, including default risk prediction, market volatility analysis, and early warning systems (Heaton, et al., 2017; Borio, et al., 2018). These models can detect complex, nonlinear patterns in large datasets and adapt to changes over time, providing a compelling alternative to traditional econometric tools.

In line with the aforementioned aspects, the relatively recent study conducted by Liu et al. (Liu, et al., 2022), which compares the performance of several early warning models in out-of-sample scenarios, shows that machine learning-based models, such as random forest, gradient boosting, and ensemble models, have a better predictive capacity than the logistic model in anticipating financial crises.

With a similar view, but focusing on textual data, Chen et al. (Chen, et al., 2023) highlighted that while traditional econometric methods and market data can anticipate some crises, integrating these types of data helps reduce classification errors, especially in the case of severe crises. A framework applicable at a global level and in real time could support decision-makers, especially in the context of international coordination.

Likewise, as per Reimann (Reimann, 2024), advanced machine learning models, such as Random Forests and Highly Randomized Trees, offer superior performance to logistic regression in terms of the ability to distinguish between crisis and non-crisis episodes, this conclusion being valid regardless of the data source, the set of indicators used, or the model specification.

Even though there is undeniable evidence, the ones discussed above being just a few of them, the association between machine learning and financial crises, at least at a conceptual level, remains a phenomenon that has received only limited attention in the specialized literature. For example, an advanced search within one of the most reliable academic databases, namely the Web of Science Core Collection, for the period 1975-2025, based on the specific query *TS=("machine learning" AND "financial crisis" AND "prediction")*, returned, at the time of the current study, only 86 publications considered relevant. This aspect may also derive from the fact that the terms used in the query do not fully cover the targeted topic, but it certainly highlights the emergent nature of the approach.

However, the scarcity of research reflects not only the rising nature of this interdisciplinary field, but also reveals important geographical disparities in its application. In particular, while global interest in combining machine learning with financial crisis prediction is growing, its implementation in the Romanian context remains limited, none of the identified studies in the Web of Science database being specifically focused on Romania.

With respect to Romania, most existing studies either apply classical econometric models or focus on descriptive analyses of past crises without exploring the predictive capabilities of computational methods. Furthermore, there is a gap in the integration of macroeconomic theory with ML techniques, which often operate as black-box models without economic interpretability.

Therefore, addressing these gaps by proposing hybrid framework that combines theoretical economic reasoning with machine learning algorithms, to forecast financial crises in Romania, becomes necessary. This approach aligns with a broader trend in the literature that aims to enhance traditional economic forecasting methods by integrating them with modern computational tools such as machine learning.

3. Research Methodology

This study follows a data-driven methodology to design, implement, and evaluate a predictive model for financial crises in Romania. The approach integrates macroeconomic theory with machine learning, using historical economic indicators and Python-based classification algorithms.

The data was collected for the period 2000-2023, from relevant official sources. Macroeconomic indicators, such as real GDP growth, inflation, unemployment and budget deficit, were taken from the National Institute of Statistics (NIS). The National Bank of Romania (NBR) provided data on interest rates and exchange rates, and data on private sector lending were obtained from the World Bank database. The labeling of the crisis years (2008–2009 and 2020–2021), was done based on the global and national economic context recognized in the specialized literature.

After data collection, they were consolidated using the calendar year as a common identifier. All columns were standardized to ensure coherence and compatibility between sources. No missing values were identified, so it was not necessary to apply any imputation methods. The final dataset included 24 observations corresponding to the analyzed years and a total of eight explanatory variables.

In the preprocessing stage, economic indicators with high predictive relevance were selected, such as the dynamics of real GDP, the inflation rate, unemployment, interest rate, budget deficit to GDP, domestic credit to the private sector to GDP and the leu/euro exchange rate. The target variable was defined as binary, indicating the presence (1) or absence (0) of a crisis in a given year. To ensure comparability between variables, normalization procedures were applied where necessary. Subsequently, the dataset was divided into two subsets: a training (70%) and a testing (30%) subset, using a stratified split to maintain a balanced distribution of the target class.

The selection of economic indicators was based both on their theoretical relevance in the literature on the occurrence of financial crises and on the availability of data over an extended and comparable time frame. Real GDP growth was included as a fundamental indicator of overall economic performance, with significant declines often associated with recessions or financial turmoil. The inflation rate was considered essential, as high or volatile inflation rates reflect price pressures that can undermine financial stability and confidence in monetary policy. The unemployment rate provided a measure of labor market tensions, often exacerbated in times of crisis, and contributed to assessing the social dimension of economic imbalances.

On the other hand, the interest rate was included in the analysis to capture the direction of monetary policy and financing conditions in the economy, as these elements have a direct impact on the cost of credit and investment. The budget deficit, expressed as a percentage of GDP, was selected as an indicator of fiscal sustainability; the accumulation of fiscal imbalances is often an early signal of macroeconomic vulnerabilities. Similarly, the level of credit granted to the private sector, relative to GDP, was used to assess the degree of indebtedness in the economy and the possible systemic risks generated by uncontrolled credit expansions. Finally, the leu/euro exchange rate reflected external competitiveness and currency risks, factors that can amplify pressures on the financial system in the context of external shocks or trade imbalances.

Thus, the choice of these variables was based both on empirical considerations and on their validation in previous studies on the prediction of financial crises. Each indicator was included not only for its individual value, but also for its potential to contribute to the identification of the complex interactions that characterize periods of economic instability.

Predictive modeling involved the application and comparison of two methods: logistic regression, which represented a classical econometric approach, and the Random Forest algorithm, which offered a non-parametric perspective specific to ensemble machine learning. The choice of logistic regression and the Random Forest algorithm aimed to compare a parametric model, explainable and established in econometric analysis, with a non-parametric model, capable of capturing complex and nonlinear relationships between variables. Logistic regression offers the advantage of interpretability and simplicity, which allows a direct understanding of the influence of each predictor on the probability of a crisis. In contrast, Random Forest was selected for its ability to handle correlations between variables, complex interactions and potential data overload, while also performing well under conditions of small and unbalanced data sets.

The models were implemented in Python, using the scikit-learn library. Their performance was evaluated using indicators such as accuracy, precision, recall, F1 score, along with scores obtained through 5-fold cross-validation. In addition, correlations between variables were analyzed using a correlation matrix, and the importance of each feature in the prediction process was examined using a specific graphical representation.

4. Results and Discussion

4.1 System Functioning and Technical Implementation

The implementation of the predictive system was carried out in Python. The structure of the code is illustrated in Figures 2 and 3, which present the core stages of the pipeline: data preprocessing and model training. The implementation began with the import of essential libraries such as pandas, numpy, scikit-learn, and matplotlib, which were used for data manipulation, model development, and result visualization.

The first step involved loading multiple .csv files containing macroeconomic indicators, which were then merged into a single DataFrame using the calendar year as the common key. These indicators included GDP growth, inflation rate, unemployment, interest rate, budget deficit, domestic credit to the private sector, and the exchange rate (RON/EUR). The merged dataset was inspected for consistency, and since no missing values were found, no imputation procedures were necessary.

The dataset was then explored and pre-processed. The Crisis_Year column was defined as the target variable, indicating whether a financial crisis had occurred in a given year (1) or not (0), while all other economic variables were treated as explanatory features. The data was split into training and testing subsets using train_test_split from scikit-learn, maintaining a 70/30 ratio to ensure that model performance could be evaluated on previously unseen data.

Two classification models were trained: Logistic Regression and Random Forest Classifier. Logistic Regression served as a baseline due to its interpretability and frequent use in econometrics. In contrast, Random Forest, a non-parametric ensemble method, was selected for its ability to capture complex and nonlinear patterns by constructing multiple decision trees and aggregating their outputs. Both models were trained on the training set and evaluated on the test set.

Model performance was assessed using standard classification metrics - accuracy, precision, recall, and F1-score-with the help of classification_report and confusion_matrix. In addition, 5-fold cross-validation was used to evaluate the generalizability of each model.

A notable feature of the system was its interpretability. For the Random Forest model, the feature_importances_ attribute was used to extract and visualize the contribution of each variable to the final prediction. This provided valuable insights into which macroeconomic factors played the most significant roles in forecasting financial crises.

Lastly, the results were visualized using matplotlib and seaborn, including a correlation matrix (Figure 1) to highlight relationships among variables and a bar chart to display feature importance. These outputs supported both the quantitative evaluation of the model and the qualitative interpretation of its economic implications.

The initial exploratory analysis showed that variables such as budget deficit, inflation, and credit to the private sector exhibit visible variation during crisis years. To better understand the relationships between variables, a correlation matrix was generated. As seen in Figure 1, certain macroeconomic indicators show moderate correlation, while others are weakly related to the Crisis_Year variable.

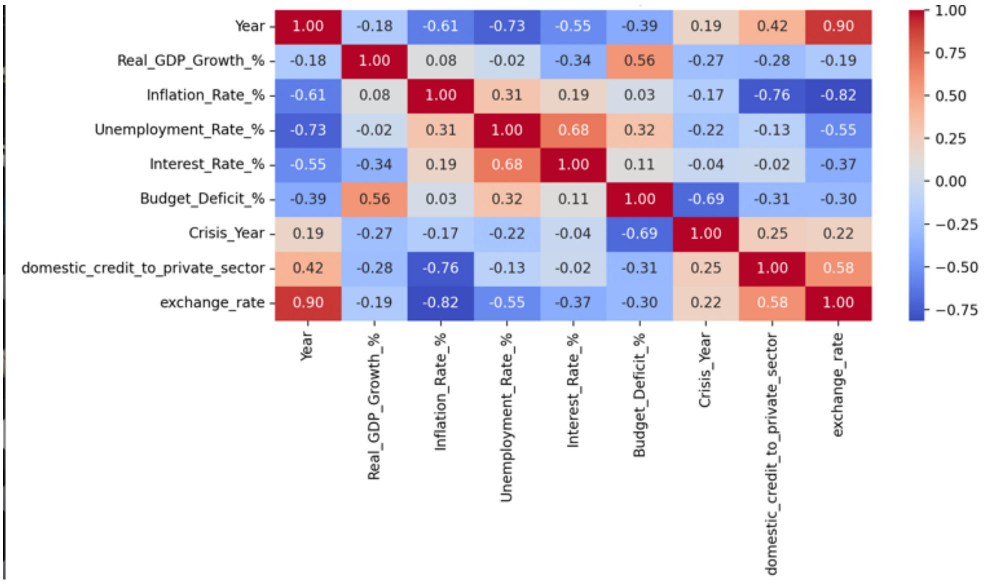

Figure 1. Correlation matrix between selected macroeconomic indicators.
Source: Author’s Own Work

In conclusion, the code implemented a modular and transparent pipeline - from data acquisition and cleaning to model training and analysis - delivering both predictive results and meaningful economic insights related to financial crises.

The predictive system was created in Python and followed a structured process that transformed raw macroeconomic data into actionable classification results. The implementation began with the import of essential libraries such as pandas, numpy, scikit-learn, and matplotlib, which were used for data manipulation, model development, and result visualization.

The first step involved loading multiple .csv files containing macroeconomic indicators, which were then merged into a single DataFrame using the calendar year as the common key. These indicators included GDP growth, inflation rate, unemployment, interest rate, budget deficit, domestic credit to the private sector, and the exchange rate (RON/EUR). The merged dataset was inspected for consistency, and since no missing values were found, no imputation procedures were necessary. These steps are illustrated in Figure 2, which shows the initial data import, merging, and preprocessing pipeline.

```python
# 1. Importarea bibliotecilor necesare
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

from sklearn.linear_model import LogisticRegression
from sklearn.ensemble import RandomForestClassifier
from sklearn.model_selection import train_test_split, cross_val_score
from sklearn.metrics import classification_report, confusion_matrix

# 2. Încărcarea datelor
macro = pd.read_csv("ro_macro_indicators.csv")
crisis = pd.read_csv("crisis_year.csv")
credit = pd.read_csv("domestic_credit_to_private_sector.csv")
fx = pd.read_csv("exchange_rate_ron_eur.csv")

# 3. Conversie coloana Year la tip int
macro["Year"] = macro["Year"].astype(int)
crisis["Year"] = crisis["Year"].astype(int)
credit["Year"] = credit["Year"].astype(int)
fx["Year"] = fx["Year"].astype(int)

# 4. Combinarea tuturor seturilor de date într-un singur DataFrame
df = macro.merge(crisis, on="Year") \
        .merge(credit, on="Year") \
        .merge(fx, on="Year")

# 5. Explorare rapidă
print(df.head())
print(df.info())
print(df.describe())
```

Figure 2. Code snippet for importing and preprocessing macroeconomic datasets
Source: Author's Own Work

The dataset was then explored and preprocessed. The Crisis_Year column was defined as the target variable, indicating whether a financial crisis had occurred in a given year (1) or not (0), while all other economic variables were treated as explanatory features. The data was split into training and testing subsets using train_test_split from scikit-learn, maintaining a 70/30 ratio to ensure that model performance could be evaluated on previously unseen data.

Two classification models were trained: Logistic Regression and Random Forest Classifier. Logistic Regression served as a baseline due to its interpretability and frequent use in econometrics. In contrast, Random Forest, a non-parametric ensemble method, was selected for its ability to capture complex and nonlinear patterns by constructing multiple decision trees and aggregating their outputs. Both models were trained on the training set and evaluated on the test set. Figure 3 illustrates the model training, evaluation, and cross-validation workflow, including the application of Logistic Regression and Random Forest classifiers.

```python
# 6. Vizualizarea corelațiilor
plt.figure(figsize=(10, 6))
sns.heatmap(df.corr(numeric_only=True), annot=True, cmap="coolwarm", fmt=".2f")
plt.title("Matrice de corelație")
plt.tight_layout()
plt.show()

# 7. Definirea variabilelor
X = df.drop(columns=["Year", "Crisis_Year"])
y = df["Crisis_Year"]

# 8. Împărțirea în set de antrenament și testare
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

# 9. Logistic Regression
log_model = LogisticRegression(max_iter=200)
log_model.fit(X_train, y_train)
y_pred_log = log_model.predict(X_test)

print("\nLogistic Regression Results:")
print(confusion_matrix(y_test, y_pred_log))
print(classification_report(y_test, y_pred_log))

# 10. Random Forest
rf_model = RandomForestClassifier(n_estimators=100, random_state=42)
rf_model.fit(X_train, y_train)
y_pred_rf = rf_model.predict(X_test)

print("\nRandom Forest Results:")
print(confusion_matrix(y_test, y_pred_rf))
print(classification_report(y_test, y_pred_rf))

# 11. Cross-validation
log_scores = cross_val_score(log_model, X, y, cv=5)
rf_scores = cross_val_score(rf_model, X, y, cv=5)

print(f"\nCross-Validation Logistic Regression: {log_scores.mean():.2f} ± {log_scores.std():.2f}")
print(f"Cross-Validation Random Forest: {rf_scores.mean():.2f} ± {rf_scores.std():.2f}")
```

Figure 3. Code snippet for training and evaluating classification models
Source: Author's Own Work

Model performance was assessed using standard classification metrics - accuracy, precision, recall, and F1 - score - with the help of classification_report and confusion_matrix. In addition, 5-fold cross-validation was used to evaluate the generalizability of each model.

A notable feature of the system was its interpretability. For the Random Forest model, the feature_importances_ attribute was used to extract and visualize the contribution of each variable to the final prediction. This provided valuable insights into which macroeconomic factors played the most significant roles in forecasting financial crises.

Lastly, the results were visualized using matplotlib and seaborn, including a correlation matrix to highlight relationships among variables and a bar chart to display feature importance. These outputs supported both the quantitative evaluation of the model and the qualitative interpretation of its economic implications.

In conclusion, the code implemented a modular and transparent pipeline - from data acquisition and cleaning to model training and analysis - delivering both predictive results and meaningful economic insights related to financial crises.

4.2 Model Performance

Both selected models, i.e. logistic regression and the Random Forest algorithm, were evaluated on the test set. The results showed that logistic regression had difficulties in handling the imbalance between classes, managing to correctly classify the non-crisis years, but failing to effectively identify the crisis years, which was reflected in a low recall score for the positive class.

On the other hand, the Random Forest algorithm demonstrated superior performance, obtaining higher scores in both recall and F1 score in detecting crisis episodes. However, the risk of overfitting was observed, an aspect mainly determined by the small size of the data set. Cross-validation confirmed these observations, indicating a higher and more stable average accuracy for Random Forest, with values ranging between approximately 80% and 85%, compared to those of logistic regression, which remained between 70% and 75%.

4.3 Feature Importance

The feature importance analysis, carried out using the Random Forest algorithm, allowed the identification of the variables that contributed most significantly to the process of predicting crisis episodes. Among these, the inflation rate had the greatest impact, followed by the interest rate, the budget deficit and the volume of domestic credit granted to the private sector, expressed as a percentage of GDP. These results are not surprising and are in line with established economic theories, which indicate that inflationary pressures, fiscal imbalances and excessive credit dynamics are major risk factors in destabilizing the financial system.

The inflation rate, in particular, has been associated with the loss of purchasing power, uncertainties in monetary policies and often delayed reactions by the authorities, which can contribute to the amplification of economic tensions. The interest rate, another high-impact indicator, reflects monetary policy and the level of perceived risk in the economy; significant changes in it can affect consumption and investment behavior. The budget deficit was another key element, as a persistent fiscal imbalance can signal the unsustainability of government policies and fuel financial market mistrust. As for credit to the private sector, an accelerated pace of its expansion can lead to over-indebtedness and the accumulation of systemic risks, especially in the absence of prudent regulatory measures.

Therefore, the model not only correctly identified these variables as being relevant from a predictive point of view but also managed to confirm the conceptual validity of the assumed economic relationships between macroeconomic indicators and the occurrence of crisis episodes. This coherence between the results and the theoretical foundations reinforces the robustness of the conclusions and supports the applicability of the model in contexts of preventive analysis of financial risks.

4.4 Discussion

The results of our analysis offer important insights into the viability of using machine learning techniques to predict financial crises in Romania, based on key macroeconomic indicators. While the sample size is relatively limited, especially considering the number of crises identified (two major events: the 2008–2009 global financial crisis and the 2020–2021 COVID-19 crisis), the findings still point toward several notable patterns.

Firstly, the logistic regression model performed reasonably well as a baseline, but its predictive power was limited by the inherent linearity of the algorithm and the small number of crisis instances. In contrast, the Random Forest model demonstrated significantly better results in terms of accuracy, precision, and recall. This is not surprising, as ensemble methods like Random Forests are known to handle non-linearities and interactions between variables more effectively. Moreover, the Random Forest's ability to estimate feature importance offered valuable economic insights, suggesting that variables such as unemployment rate, inflation, and real GDP growth have strong predictive relevance.

One key takeaway is the importance of credit dynamics and fiscal conditions in the lead-up to financial instability. Indicators such as domestic credit to the private sector and budget deficit appear to be significant signals, supporting traditional economic theories about credit overextension and unsustainable fiscal policies being precursors to crises.

However, several limitations must be acknowledged. The small dataset – both in terms of number of years and number of crisis periods - poses a challenge for machine learning models that typically require more data to generalize well. The binary classification task (crisis vs. non-crisis) oversimplifies the complex and often gradual emergence of economic distress, which can vary in intensity and duration. Moreover, the labeling process is subjective to some extent, relying on historical definitions of crisis years that may not fully capture structural vulnerabilities.

Another consideration is the exogeneity of external shocks. Both crisis periods used in this study were largely triggered by global events (the global financial crisis and the COVID-19 pandemic), which may limit the model's predictive accuracy in scenarios driven by purely domestic imbalances. This highlights the potential need to include global variables (e.g., oil prices, global interest rates) in future extensions of the model.

Despite these limitations, the findings support the broader hypothesis that combining traditional economic analysis with modern computing tools can lead to more robust early-warning systems. While this study focused on Romania, the methodology is adaptable to other countries and periods, provided adequate and reliable data is available. Additionally, the use of interpretable machine learning models can support policy decisions by identifying which economic signals warrant closer attention.

In conclusion, our project demonstrates that machine learning techniques, even when applied to relatively limited datasets, can provide valuable support in anticipating financial stress. However, future research should focus on expanding the dataset, refining crisis definitions, and exploring hybrid models that combine data-driven approaches with expert economic judgment.

5. Conclusions

This study aimed to bridge traditional economic analysis with modern computational methods by developing a predictive framework for financial crises in Romania. Using a selection of key macroeconomic indicators and applying logistic regression and Random Forest classifiers, we evaluated the potential of machine learning models to detect early signs of financial distress.

The results indicate that, even with a relatively small dataset, machine learning techniques - particularly ensemble models - can provide meaningful insights into the conditions that typically precede a financial crisis. The Random Forest model, in particular, exhibited strong performance and identified several macroeconomic variables - such as unemployment, inflation, credit growth, and fiscal balance - as critical predictors. These findings are consistent with established economic theories, lending credibility to the data-driven approach.

Nonetheless, the research also underscores several important challenges. The limited number of crisis events and the relatively short time horizon restrict the generalizability of the model. Moreover, the complexity and external nature of some crises suggest that incorporating global factors may further improve predictive performance. Despite these limitations, the framework developed here offers a promising foundation for further work in building robust early-warning systems tailored to emerging markets.

Hence, several limitations must be acknowledged regarding the present study. First, the relatively short time series and the limited number of observed financial crisis episodes restrict the model's ability to generalize and capture broader systemic patterns. Financial crises are rare events, and their low frequency imposes significant constraints on supervised learning approaches.

Second, the binary classification scheme (crisis vs. non-crisis) oversimplifies the complexity and gradual evolution of economic distress, which often unfolds in stages and varies in intensity. Future models could explore multi-class or time-to-crisis frameworks to better reflect the dynamic nature of financial vulnerability.

Third, the study relies solely on domestic macroeconomic variables. However, many crisis episodes, particularly those analysed in this study (the 2008-2009 global financial crisis and the 2020-2021 COVID-19 shock), were driven primarily by external forces. The absence of global indicators, such as international interest rates, oil prices, or global financial conditions, may therefore limit the predictive capacity of the model in scenarios triggered by exogenous shocks.

Additionally, the use of annual data restricts the model's responsiveness to short-term fluctuations or sudden structural breaks. More granular (e.g., quarterly or monthly) datasets, when available, could significantly improve both model accuracy and timeliness.

Thus, for future research, several directions are promising: expanding the dataset by including more countries or longer historical periods, integrating global economic indicators, refining the definition of crisis periods, and experimenting with more advanced models such as recurrent neural networks or hybrid econometric–ML approaches. Such developments could further enhance the robustness, interpretability, and policy relevance of early-warning models.

In sum, this research demonstrates the potential benefits of integrating economic reasoning with machine learning tools. As computing power and data availability continue to grow, such hybrid approaches will likely become increasingly important for economic monitoring, policymaking, and crisis prevention, not just in Romania, but across a wider international context.

References

- Borio, C. E., Drehmann, M. & Xia, F. D., 2018. The financial cycle and recession risk. *BIS Quarterly Review December..*
- Chen, M. et al., 2023. Identifying financial crises using machine learning on textual data. *Journal of Risk and Financial Management,* 16(3), p. 161.
- Claessens, M. S. & Kose, M. A., 2013. *Financial crises explanations, types, and implications.* Washington, D.C.: International Monetary Fund.
- Drehmann, M. & Juselius, M., 2014. Evaluating early warning indicators of banking crises: Satisfying policy requirements. *International Journal of Forecasting,* 30(3), pp. 759-780.
- Duguleană, C., 2011. Effects of the economic crisis in Romania. *SEER: Journal for Labour and Social Affairs in Eastern Europe,* 14(1), pp. 17-25.
- Heaton, J. B., Polson, N. G. & Witte, J. H., 2017. Deep learning for finance: deep portfolios. *Applied Stochastic Models in Business and Industry,* 33(1), pp. 3-12.

- Kaminsky, G. L. & Reinhart, C. M., 1999. The twin crises: the causes of banking and balance-of-payments problem. *American economic review,* 89(3), pp. 473-500.
- Liu, L., Chen, C. & Wang, B., 2022. Predicting financial crises with machine learning methods. *Journal of Forecasting,* 41(5), pp. 871-910.
- Niculescu, M. D., 2013. Financial And Economic Crisis In Romania: Vulnerabilities And Opportunities For Economic Recovery. *Working papers.*
- Pelinescu, P. S. R. E. & Simionescu, P. S. R. M., 2017. The Effects of the Recent Economic and Financial Crisis on the Romanian Economy. *Research Association for Interdisciplinary Studies,* Issue 15, pp. 151-162.
- Reimann, C., 2024. Predicting financial crises: an evaluation of machine learning algorithms and model explainability for early warning systems. *Review of Evolutionary Political Economy,* 5(1), pp. 51-83.