FMDB *Transactions on Sustainable Technoprise Letters*

Development of Novel Inflation-Aware AI Framework for Credit Risk Assessment under Economic Stress

Carla Popolizio[1,*], Rejwan Bin Sulaiman[2], Maruf Farhan[3], Naser Abosata[4]

[1,2,3]Department of Computer Science and Technology, Northumbria University, Newcastle upon Tyne, London, United Kingdom.
[4]Department of Science and Engineering, School of Computing, University of Derby, Derby, England, United Kingdom.
carla.popolizio@northumbria.ac.uk[1], rejwan.binsulaiman@gmail.com[2], maruf2.farhan@northumbria.ac.uk[3], n.abosata@derby.ac.uk[4]

Abstract: Traditional credit scoring models use historical financial data to estimate borrower risk, but they rarely consider how borrower affordability might change under inflation or economic instability. This limits lenders, as borrower default risk may rise even when underlying credit quality remains the same. To address this gap, this study created and tested an Inflation-Aware AI Credit Risk Decision Support Framework that uses supervised machine learning and inflation stress testing to help lenders make lending decisions during economic uncertainty. Using a Design Science Research methodology, researchers built and evaluated three supervised machine learning algorithms (Logistic Regression, Random Forest, and Gradient Boosting) on the Give Me Some Credit dataset, which contains 150,000 borrower records. Model performance was measured using accuracy, precision, recall, F1-score, and ROC-AUC, with stratified 5-fold cross-validation and McNemar's test used to validate performance differences. Gradient Boosting performed best and was later chosen for deployment in an interactive Streamlit application. The approach predicts borrower default probability, categorises debtors, and assesses how default risk changes under mild, moderate, and severe inflation scenarios by simulating income losses and increases in debt ratio and credit usage. To support decision-making, the application includes a simple dashboard that displays default probability, loan recommendations, borrower-level default risk, and inflation-scenario comparisons. The results show that applying inflation stress testing to credit risk assessments adds more value beyond traditional methods by identifying borrowers whose risk profile worsens under uncertain scenarios. The suggested approach therefore provides a useful tool that shows how AI and economic scenario analysis can further enhance evaluations in real financial contexts.

Keywords: Credit Risk Assessment; Decision Support Framework; Machine Learning; Random Forest; Borrower Default Risk; Gradient Boosting; AI Framework.

Received on: 23/06/2025, **Revised on:** 28/09/2025, **Accepted on:** 01/12/2025, **Published on:** 05/06/2026

Journal Homepage: https://www.fmdbpub.com/user/journals/details/FTSTPL

DOI: https://doi.org/10.69888/FTSTPL.2026.000813

Cite as: C. Popolizio, R. B. Sulaiman, M. Farhan, and N. Abosata, "Development of Novel Inflation-Aware AI Framework for Credit Risk Assessment under Economic Stress," *FMDB Transactions on Sustainable Technoprise Letter*, vol. 4, no. 2, pp. 104–126, 2026.

Copyright © 2026 C. Popolizio *et al.,* licensed to Fernando Martins De Bulhão (FMDB) Publishing Company. This is an open access article distributed under CC BY-NC-SA 4.0, which permits copying, remixing, redistributing, transformation, and building upon the material in any medium so long as the original work is properly cited.

1. Introduction

[*]Corresponding author.

To develop a novel inflation-aware AI framework that integrates machine learning and inflation stress testing to support credit risk assessment under changing economic conditions [1]. Credit risk assessment is one of the most important operations performed by financial institutions, as it directly affects lending decisions, financial stability and risk management [2]. Banks and other lending institutions must assess borrowers' capacity to meet their financial commitments while balancing profitability against potential losses from loan defaults [3]. This approach relies on statistical credit scoring models that include borrower attributes such as income, debt, repayment history and credit utilisation [4]. Although these tools have supported lending decisions, their ability to adapt to fast-changing economic conditions remains limited [8]. Recent economic shocks have underlined the need for more flexible methods for credit risk assessment [33]. For example, the increase in worldwide inflation caused by the COVID-19 pandemic, along with supply chain disruptions and geopolitical uncertainties, has severely affected family budgets [9]. This is because rising interest rates and declining buying power have put further financial strain on borrowers, increasing the risk of repayment issues [26]. Moreover, many existing credit scoring algorithms assume that borrower attributes stay constant throughout the loan process, whereas in actuality, economic circumstances change over time, and borrowers who might look financially stable at first could become much more vulnerable during times of inflation [30]. Therefore, incorporating stress testing into credit assessment could improve lending choices by providing a more dynamic picture of borrower risk [31]. This study creates a Novel Inflation-Aware AI Framework for assessing credit risk under economic stress [16]. The framework is implemented as an interactive Streamlit application that allows users to examine inflation scenarios and assess borrower risk [38]. This technique aims to provide a more comprehensive approach to credit risk assessment compared to traditional methods by merging predictive analytics with economic stress testing. [15]

1.1. Research Questions

To achieve the research aim, the study addresses the following research questions:

- How can an inflation-aware AI framework be designed and implemented to support credit risk assessment under changing economic conditions?
- How does integrating inflation stress testing into the proposed AI framework influence borrower credit risk assessment under simulated economic conditions?

1.2. Objectives

The objectives of this research are:

- To review the existing literature on AI, machine learning, credit risk assessment and the influence of macroeconomic conditions on lending decisions.
- To develop and compare machine learning models for predicting borrower default.
- To identify the most suitable predictive model for integration into the proposed framework.
- To develop an inflation stress-testing methodology that simulates different economic conditions.
- To develop and implement a novel inflation-aware AI framework for credit risk assessment.
- To investigate how simulated inflation scenarios influence borrower default probability and borrower risk classification within the proposed framework.

2. Literature Review

Artificial Intelligence (AI) is becoming increasingly important in credit risk assessment because it allows financial institutions to analyse large amounts of borrower data and support lending decision-making through predictive modelling. Machine Learning (ML) approaches are commonly used to identify patterns associated with borrower default risk, and they have been shown to increase the efficiency and consistency of credit rating procedures. However, concerns remain about these models' reliability when economic conditions change, especially during inflation or financial uncertainty. The review is organised in six key areas: AI in credit risk assessment, ML models for credit risk prediction, the impact of economic factors on borrower default risk, AI reliability under economic stress, trust and explainability of AI, and ethical considerations for AI-based financial decision-making. By analysing the existing literature, this paper identifies the research gaps that justify developing an Inflation-Aware AI Credit Risk Decision Support Tool and establishes the study's theoretical foundation. AI is becoming crucial for financial services, especially in areas such as credit scoring, fraud detection, lending and risk management. In credit risk assessment, machine learning (ML) algorithms evaluate the probability of borrower default or repayment issues. Traditional methods often rely on statistical methods or human review. However, financial institutions can now use more advanced ML models to support decision-making, thanks to improved data availability and processing capacity. Compared with traditional methods, growing research indicates that ML can improve the accuracy of credit risk assessment. According to Machado et al. [24], ML models are valuable because they can identify complex relationships between borrower behaviour and default

probability. Similarly, Noriega et al. [29] and Lu et al. [34] found that ML approaches can improve credit risk prediction by detecting nonlinear patterns that statistical models may miss. This matters because borrower risk can be influenced by several factors, including income, debt levels, repayment history, credit utilisation, and overall financial behaviour. Ayari et al. [3] further highlight that AI-based credit scoring enables faster, more reliable decision-making, making it a crucial area within financial technology. Recent research has also shifted toward privacy-preserving and distributed AI architectures for financial applications, particularly through federated learning, offering opportunities to improve financial intelligence while limiting the centralisation of sensitive financial information.

Despite these advantages, many existing studies evaluate model performance using historical data and pay less attention to how models perform during economic instability. Li [22] notes that credit risk model performance can degrade during severe economic scenarios, implying that models based on past data can become less reliable as borrower behaviour shifts due to inflation and interest rate changes. Similarly, Chang et al. [12] found that ML and deep learning models can accurately predict credit risk; however, borrower behaviour and default risk are not static. More specifically, risk patterns may change over time because of economic factors such as inflation and recession, which requires continuous monitoring and updating of predictive models to maintain accuracy and reliability. Their findings show that the models' ability to adjust to shifting macroeconomic scenarios is as important as the predictive performance of AI-based credit risk assessment. While the literature generally agrees that AI can enhance credit risk prediction, little is known about how these models perform during inflation or economic instability. Instead of evaluating whether models can still be trustworthy as borrower behaviour changes, the majority of the existing studies focus on predicted accuracy under typical circumstances, which highlights a significant gap in the literature and emphasises the need for a deeper investigation into inflation-aware AI systems that can provide more reliable credit risk predictions in unstable economic times.

ML is becoming more important in credit risk assessment because it can analyse vast financial information and identify patterns linked to borrower default. Unlike conventional statistical approaches, ML algorithms can accurately estimate credit risk by identifying complex, non-linear interactions between variables. As a result, several financial institutions have integrated ML methods into current credit scoring systems to support lending decisions and mitigate potential losses. Machine learning has increasingly been applied to financial risk and fraud-related decision-making, with algorithms such as Logistic Regression, Random Forest, Support Vector Machines and ensemble methods showing different performance depending on the characteristics and imbalance of financial datasets. Comparative evaluations of machine-learning models within digital payment environments further demonstrate the potential of data-driven approaches to identify complex financial risk patterns. Zhang [36] shows that Logistic Regression is a commonly used technique in credit risk modelling. Despite being one of the oldest predictive approaches, it remains widely used because of its simplicity, efficiency and interpretability. The model evaluates the likelihood of borrower default using a set of predictor variables, enabling lenders to understand how different factors influence risk assessments. However, as data availability and processing power have advanced, Ayari et al. [3] note that models like Random Forest, Gradient Boosting, and XGBoost can reveal complex connections between borrower characteristics that traditional approaches may miss. Machado et al. [24] similarly argue that ML techniques often outperform traditional methods in default prediction. Further, Chang et al. [12] reported that advanced ML and deep learning models performed well in prediction tasks, suggesting these methods can improve lending decision accuracy.

Despite these advantages, there is still ongoing debate about whether prediction accuracy should be prioritised over transparency. Hang [36] argues that simpler models like Logistic Regression remain relevant since their predictions are easier to understand and explain. From this perspective, a transparent, simpler model might be preferred over a complex one whose decision-making process is harder to explain. On the other hand, Machado et al. [24] and Chang et al. [12] contend that emphasising explainability can restrict financial institutions' capacity to take advantage of the higher predictive performance provided by advanced ML techniques, which creates a trade-off between achieving high predictive accuracy and maintaining transparency in lending decisions. Explainable AI (XAI) has been used to address this challenge. Explainability approaches, according to Do et al. [14], can improve understanding of predictions while maintaining the predictive advantages of more complex models. This implies that explainability and accuracy should be seen as complementary aspects of successful credit risk assessment systems rather than as competing objectives. Overall, while advanced ML models often provide better accuracy, traditional methods can offer more transparency and easier regulatory approval. Therefore, choosing a suitable model requires balancing explainability and accuracy to meet the demands of the lending environment. Further, it remains unclear how these models perform during inflation or economic instability, since most research assesses model performance using historical data collected under comparatively stable economic conditions. This emphasises the need for further investigation into evaluating credit risk under changing economic conditions.

Macroeconomic factors can significantly affect borrower trustworthiness and overall financial stability. Economic growth, unemployment, inflation and interest rates all affect an individual's finances, borrowing ability and repayment behaviour. According to Carvalho et al. [10], adding macroeconomic data can further enhance the precision of credit default prediction models, with economic growth being the key factor influencing default risk. In fact, their results show that borrower outcomes

are influenced not only by the individual financial qualities but also by the economic conditions. Similarly, Sharma et al. [32] argued that inflation and unemployment were among the strongest predictors of credit risk in the UK banking sector, and that deteriorating economic conditions can increase financial vulnerability, contributing to a higher likelihood of borrower default. In recent years, inflation has been a major concern due to rising living costs and increased financial pressures. In fact, according to the Bank of England [4], inflationary pressures still affect affordability and remain a source of economic concern, despite the decrease from the high levels seen in prior years. Household buying power decreases when prices rise, reducing the disposable income available for debt repayment and suggesting that prolonged inflation can make borrowers less resilient and repayments more challenging.

Further, interest rate fluctuations can amplify the effects of inflation. In fact, central banks often raise interest rates in response to inflationary pressures to reduce spending and stabilise prices. However, higher interest rates can make borrowing more expensive for both individuals and companies. To examine the factors that drive inflation, Molnar [28] created an ML framework to analyse the determinants of inflation. They found that supply shocks, labour market conditions, and inflation expectations interact to affect inflationary pressures. This suggests that economic environments are dynamic and continuously changing, so relationships observed in historical data may not remain stable over time. Overall, most ML research focuses on borrower-level factors and assesses model performance using historical datasets gathered under stable conditions, despite the general understanding that macroeconomic conditions affect borrower behaviour and default risk. Macroeconomic research, by contrast, emphasises that borrower risk can vary drastically as economic conditions change. As a result, little attention has been given to whether AI-based credit risk assessment systems can provide valid predictions when borrower affordability declines due to inflation, rising interest rates, or broader economic instability.

The increasing use of AI in financial services has raised significant concerns about how reliable ML models are under changing economic conditions. In fact, when economic conditions shift from those in the training data, the efficacy and accuracy of credit risk models may be questioned, even if many show strong predictive performance during development and testing. Thus, because ML models are usually trained on historical data, they assume similar patterns will appear in future observations, but the state of the economy keeps shifting. Bello [8] argues that economic uncertainty may affect ML systems' performance, as borrowers' financial behaviour can fluctuate over time. More specifically, during periods of inflation, rising living costs, and increased borrowing pressures might change repayment behaviour, making models trained on data gathered under normal conditions less reliable. One explanation for this challenge is distribution shift and concept drift. According to Hovakimyan and Bravo [17], concept drift happens when the link between predictor variables and outcomes changes over time; distribution shift happens when the features of incoming data differ from those seen and used during model training. Lu et al. [23] also argued that concept drift is an important issue for ML systems operating in dynamic contexts, since previously predicted patterns may lose significance when economic circumstances change. Similarly, Mannapur [25] argues that when models face shifting operational contexts, data drift and concept drift can lower forecast accuracy; for example, inflationary pressures may change borrower affordability, credit utilisation and repayment behaviour, increasing default risk.

Despite these concerns, the literature disagrees on how economic stress affects AI performance. Lu et al. [23] argued that concept drift indicates that when the economic context changes drastically, ML models may perform worse; therefore, strong performance on past datasets does not always ensure dependable future performance. Thus, Hovakimyan and Bravo [17] explain that both concept and distribution shift can affect model reliability in dynamic environments, suggesting continuous monitoring and adaptation to maintain predictive accuracy. However, Lan [21] contends that because more sophisticated ML models can detect complex, non-linear patterns in financial data, they can remain useful across a variety of economic contexts. Thus, when borrower characteristics and market circumstances shift, ML could be more flexible than traditional statistical methods. To address this, financial institutions are increasingly using stress-testing frameworks to assess model strength. Mannapur [25] argues that stress testing lets businesses evaluate how prediction models perform under economic instability scenarios. Thus, stress testing offers important insights into whether ML systems remain dependable when exposed to settings outside their original training environment by simulating situations like inflation shocks. Overall, the literature presents contrasting views on the reliability of AI under economic stress. While studies on more advanced ML models argue that they remain stable across contexts, research on concept drift and economic uncertainty shows that shifting circumstances can reduce model accuracy. Despite this controversy, little research has examined the reliability of AI-based credit risk assessment systems in inflationary environments, suggesting further research is needed to assess whether ML models can continue to provide accurate and trustworthy risk evaluations as borrower behaviour changes under economic stress.

As AI becomes increasingly integrated into financial decision-making, concerns about trust, explainability and ethical considerations have also become more important. In fact, while AI systems can improve efficiency and predictive performance, institutions must fully trust them before using them in sectors like banking. For this reason, when implementing AI-based credit risk assessment systems, institutions must consider transparency, accountability, and fairness. According to Ayari et al. [3], one of the main challenges of contemporary ML models remains low interpretability. For example, Gradient Boosting and other ensemble methods often function as "black box" systems, making them difficult to interpret. Further, Research by Černevičienė

and Kabasinskas [11] on Explainable AI (XAI) in finance suggests that improving model interpretability is essential for increasing trust in decision-making. In response, XAI has emerged, offering techniques that help people understand how ML models reach their predictions. According to Credit Connect [13], explainability enhances transparency and facilitates the use of AI in financial services.

Similarly, Mohsin and Nasim [27] noted that explainability has grown significantly as financial organisations seek to balance accountability and accuracy, and their review highlighted that techniques such as SHAP are widely used to improve understanding of AI outcomes. However, there is an ongoing discussion over the connection between explainability and predictive performance. Explainable models can increase users' trust and transparency, but they may not be as predictive as more sophisticated ML methods. For this reason, financial institutions must choose between maintaining transparency and optimising prediction accuracy, which makes it harder for organisations, as both performance and accountability are needed. Trust is another important factor in adopting AI systems. In fact, research by Do et al. [14] and Vuković et al. [35] suggests that users are more inclined to trust AI-generated suggestions when systems are transparent, dependable and fair. However, Ioannou et al. [18] highlight that trust must be treated with caution, as inadequate trust may lead people to reject potentially useful insights and may also lead to overreliance on automated suggestions. For this reason, successful adoption of AI in financial services depends not only on performance but also on ensuring that users understand and evaluate AI-generated outputs.

Ethical concerns further complicate the adoption of AI in credit risk assessment. In fact, ML algorithms learn from historical data and can inadvertently reproduce biases in training datasets. If those biases go undetected, AI systems can suggest unfair outcomes that disfavour specific groups of borrowers. Moreover, Kowald et al. [20] suggested that AI should be fair, transparent, responsible and resilient, as these characteristics matter in lending situations where credit choices can directly affect financial opportunity and overall economic wellbeing. Recent legislative developments have further emphasised the need for effective AI governance. Transparency, accountability, and human review are highlighted as crucial areas of responsible AI deployment. Similarly, policymakers increasingly acknowledge the need for governance measures that keep AI systems intelligible, auditable, and aligned with social norms. These changes reflect growing concerns about hazards associated with unclear decision-making in high-impact areas such as finance. The banking industry has also become more aware of the risks of extensive AI use. In fact, according to the Bank of England [5], businesses must address rising governance, operational, and systemic concerns connected with AI deployment, even when it offers benefits for boosting efficiency and decision-making. This highlights the need for strong governance structures, human supervision, and efficient monitoring when adopting AI-based financial systems. However, human monitoring does not fully resolve AI-related hazards, as it can over-trust automated suggestions or fail to spot algorithmic flaws.

Overall, when tested using historical credit datasets, ML models usually perform better than traditional methods. Similarly, Sharma et al. [32] and the Bank of England [6] note that borrowers' financial vulnerability and default risk may increase with rising interest rates, declining household affordability or overall economic uncertainty. Despite these developments, a substantial gap remains in how AI systems operate in credit risk assessment when economic scenarios change. A key limitation of the available research is the lack of agreement on the reliability of ML models during economic volatility. As mentioned by Lan [21], advanced ML model techniques are sufficient to adapt to changing environments, as they can capture complex and non-linear correlations within financial data, despite studies by Li [22] and Mannapur [25] on concept drift and economic uncertainty suggesting that changes in borrower behavior may reduce model effectiveness over time if not monitored. This disagreement draws attention to a basic uncertainty within the literature, as while predictive performance is often reported under normal scenarios, there is still little research on whether AI credit risk systems can continue to make accurate assessments when inflation changes the attributes that serve as the basis for their prediction.

A second key limitation is the division of macroeconomic and ML research, as while ML research mostly focuses on model performance measured with accuracy, precision, and ROC-AUC, studies looking at inflation usually concentrate on borrower affordability, debt repayment, and default behaviour. As a result, the relationship between shifting economic scenarios and AI credit risk assessment remains poorly understood. The reviewed literature provides limited evidence on how ML predictions change under inflationary stress scenarios, even though researchers recognise that macroeconomic conditions can affect borrowers' default risk [4]. As a result, it is unclear whether AI credit risk evaluations effectively reflect changes in borrower risk. Moreover, researchers evaluate ML models using static historical datasets. Although this method offers useful insights into predictive performance, it may not capture the dynamics of actual lending decisions. In fact, borrower characteristics and repayment patterns can change over time, especially during periods of economic stress. Therefore, assessing predictive performance alone may not adequately capture the model’s utility. Understanding how ML systems respond to economic situations different from those in their training data requires more attention. Practical implementation is another important limitation. In fact, research has mostly focused on creating or contrasting prediction models. Still, very little emphasis has been paid to converting these models into useful decision-support frameworks that can actually help with lending choices in those settings.

As a result, a gap remains between the creation of theoretical models and their implementation. Frameworks that not only support forecasting but also help make decisions in response to economic shifts are becoming increasingly necessary as financial institutions continue to implement and use AI technology. In addition, although explainability, trust, and responsible AI have drawn more attention in recent research, these concerns are often examined in terms of model robustness and economic stress. Consequently, the evidence remains limited [5]. Therefore, this research addresses these gaps by designing and assessing a framework that combines ML-based credit risk predictions with inflation stress-testing scenarios to evaluate how borrower risk varies under different economic conditions. This study integrates economic stress testing into the assessment process, allowing a more thorough examination of model behaviour under changing scenarios, in contrast with other studies that focus only on predictive performance. This study contributes to research on the practical application of AI in financial decision-making by combining predictive modelling, inflation-aware risk assessment, and decision-support functionality in a single framework. It also provides new insights into the reliability of AI-based credit risk assessments during periods of economic uncertainty.

3. Research Methodologies

This paper describes the approach used to develop the suggested Novel Inflation-Aware AI Credit Risk Decision Support Framework. The methodology addresses the research questions by integrating ML approaches with the development of an artefact that can evaluate borrower credit risk under economic shifts. The paper begins by outlining the research philosophy and research design adopted for the study. It then describes the dataset used, the preprocessing procedure and the feature engineering approaches used to prepare the data for ML. Afterward, it explains the inflation stress-testing framework incorporated into the application, and the development and evaluation of the predictive models before describing how the selected model was integrated into the Streamlit application through Python. Finally, it discusses the ethical considerations that are associated with AI lending decision support. For this study, a pragmatist approach was chosen because it allowed both quantitative model development and software creation, rather than restricting the work to statistical analysis or software development.

Although a quantitative approach was considered, it would have produced a prediction model rather than the decision-support tool that was the main outcome of the research. The quantitative component focuses on developing and evaluating predictive models using past data, whereas the practical components integrate the selected model into the application, which can simulate different inflation scenarios. This study uses a quantitative research design with a Design Science Research (DSR) approach. While DSR focuses on developing and assessing a useful software artefact to address the issue, the quantitative element supports the development and evaluation of ML models capable of predicting borrowers' default risk using historical data. Therefore, DSR is especially relevant to this study, as it goes beyond analysing existing ML models. Instead, the research aims to design, develop, and evaluate a framework capable of integrating predictive models with inflation stress-testing simulations. As shown in Figure 1, the study follows a systematic, multi-stage development process.

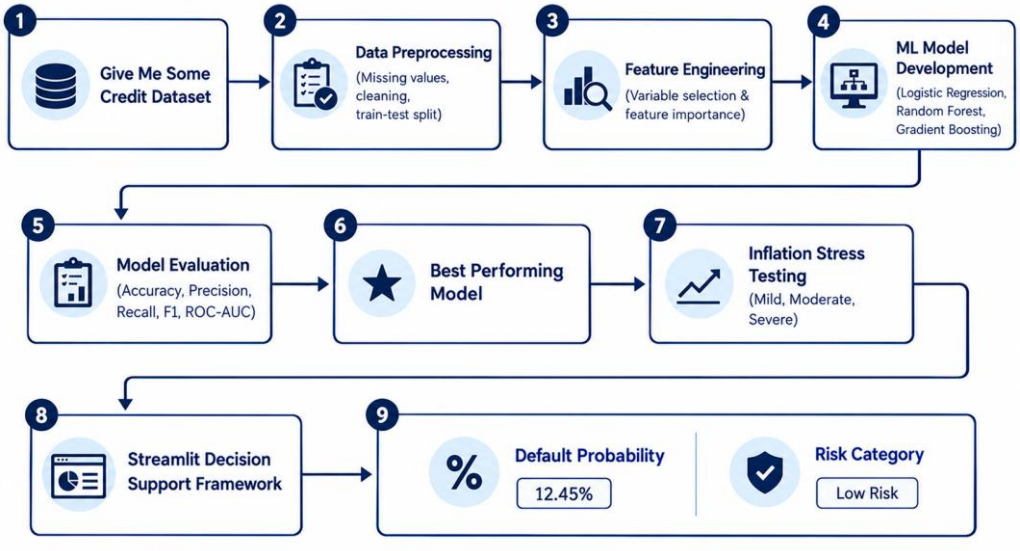


Figure 1: Research methodology overview

The study begins by selecting the "Give Me Some Credit" dataset from Kaggle [19]; before selecting the relevant variables through feature engineering, the dataset undergoes several preparation steps, including data cleaning, missing-value treatment,

and validation. Next, researchers build and evaluate several supervised ML algorithms to determine which model best balances predictive performance and robustness; the top-performing algorithm is then implemented in the web application using Python and Streamlit. Users can input borrowers' financial information to receive an estimated probability of default and an associated risk category. The framework also includes stress-testing, in contrast to traditional credit rating methods. The predefined inflation scenarios are simulated by changing a few borrowers' financial parameters, such as income, debt ratio, and credit utilisation, rather than retraining the algorithm on other datasets, which keeps the prediction model consistent while assessing borrower variation in risk under mild, moderate, and severe scenarios. Lastly, the framework is then evaluated by examining both the predictive performance of the selected ML model and the effect of inflation on borrower risk classification.

3.1. Dataset Description

The predictive model in this study was trained on the Give Me Some Credit dataset from Kaggle [19], which contains 150,000 borrower records and eleven variables, as shown in Table 1, describing demographic and financial features associated with credit risk.

Table 1: Dataset variables

Variable	Description
SeriousDlqin2yrs	Target variable indicating serious delinquency within two years (1 = Yes, 0 = No)
RevolvingUtilizationOfUnsecuredLines	Percentage of unsecured credit utilised
Age	Borrower's age (years)
DebtRatio	Monthly debt payments relative to income.
MonthlyIncome	Borrower's monthly income
NumberOfOpenCreditLinesAndLoans	Total number of active credit accounts.
NumberOfTime30-59DaysPastDueNotWorse	Number of late payments (30–59 days)
NumberOfTime60-89DaysPastDueNotWorse	Number of late payments (60–89 days)
NumberOfTimes90DaysLate	Number of late payments exceeding 90 days
NumberRealEstateLoansOrLines	Number of mortgage or real estate loans
NumberOfDependents	Number of financial dependents

The dependent variable, SeriousDlqin2yrs, shows whether a borrower has a major delinquency during the following two years. As a result, the issue is framed as a binary classification task, aiming to predict the probability of borrower default. The predictive variables include measures of repayment history, debt exposure, income, credit utilisation and borrowing behaviour.

3.2. Data Preprocessing

Before model development, the dataset underwent a structured preprocessing approach to improve data quality and ensure the ML algorithms were trained on accurate data. The first step was to check missing values across all variables, ensuring the dataset was complete, consistent and ready for further analysis. The second step was to examine the dataset for duplicates and unrealistic values; no duplicates were found, and unrealistic age values were corrected before confirming that no missing values remained.

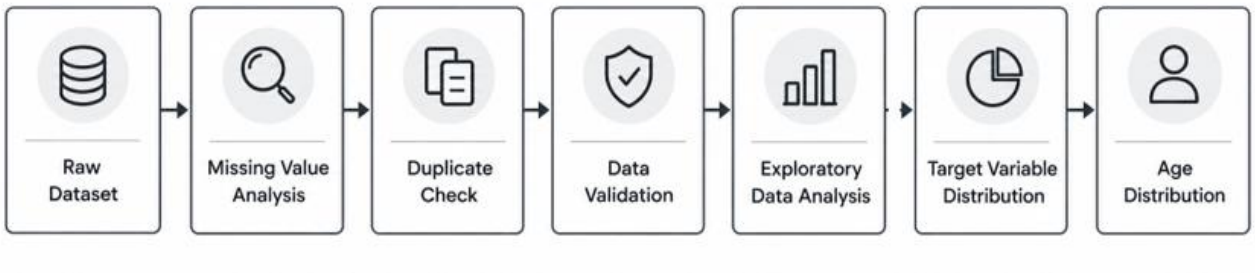


Figure 2: Data preprocessing pipeline

Before developing the model, Exploratory Data Analysis (EDA) examined the target variable distribution to assess the balance between defaulting and non-defaulting borrowers. Further, additional visualisation was added by creating histograms of borrower classes and borrower age. The overall data preprocessing procedure is summarised in Figure 2.

3.3. Feature Engineering

Researchers carried out feature engineering before the train/test split because they created the new variables using existing borrower information, which did not include the target variable or any information from the training or testing sets. Three new ratio features were added to represent borrowers' financial situation accurately: DebtPerDependent (DebtRatio divided by NumberOfDependents + 1), CreditLinesPerAge (number of open credit lines divided by age), and IncomeToDebt (MonthlyIncome divided by DebtRatio + 1). After feature engineering, a correlation heatmap was created to analyse correlations between the original and engineered predictor variables, providing an overview of linear relationships and allowing evaluation of the new variables before model development. Afterward, the engineered dataset was checked for outliers using the Z-score approach, and observations with an absolute Z-score of three or above were identified across the numerical variables to assess extreme values. These observations reflected real borrower features rather than data entry errors; therefore, they were kept for model training.

3.4. ML Model Development

The prediction model was built using supervised ML, as the study aims to categorise borrowers based on their probability of default. Before training, the dataset was split into training and testing sets using an 80:2 stratified train-test split (test_size=0.2, random_state=42).



Figure 3: ML model development and evaluation process

To determine the best prediction model, researchers compared three commonly used classification algorithms: Logistic Regression, Random Forest, and Gradient Boosting. Logistic Regression was chosen because of its ease of use, high interpretability, and computational efficiency, and remains widely used in financial institutions since its forecasts can be easily explained and justified, even though it assumes a linear connection between predictor variables and the chance of default. Because its coefficients are sensitive to feature scale, unlike the tree-based models below, it was developed in scikit-learn with class_weight="balanced", max_iter=1000, and random_state=42, and trained on standardised features (via StandardScaler, fitted only on the training set). Random Forest was chosen for its ability to reduce overfitting through ensemble learning. By combining the predictions of many decision trees, it can represent complex, non-linear relationships while maintaining stable predictive accuracy. It was configured with 200 trees, class_weight="balanced," and random_state=42 (n_estimators=200), and trained on the unscaled feature set, since tree-based splits are unaffected by feature magnitude. Gradient Boosting was added because it builds decision trees sequentially, with each tree correcting the previous one, resulting in strong predictive performance on difficult classification tasks. It was configured with 200 trees, a learning rate of 0.05, and a maximum depth of 3 (n_estimators=200, learning_rate=0.05, max_depth=3, random_state=42). Because scikit-learn's GradientBoostingClassifier has no class_weight parameter, class imbalance was addressed by computing per-instance weights with compute_sample_weight(class_weight="balanced", y=y_train) and passing these into fit () via the sample_weight argument, replicating the balancing effect applied to the other two models. Finally, a confusion matrix was generated to examine the distribution of true positives, true negatives, false positives and false negatives.

All three models were trained on the same 80/20 stratified split and evaluated on accuracy, precision, recall, F1-score, and ROC-AUC. Accuracy alone can give a misleading impression of the model's performance because the dataset has far fewer default instances than non-default scenarios. Further, failing to identify high-risk borrowers can have serious financial consequences; therefore, Recall and ROC-AUC were prioritized because they provide a more relevant evaluation for imbalanced credit-risk datasets. Researchers then re-evaluated all three models using StratifiedKFold 5-fold cross-validation (n_splits=5, shuffle=True, random_state=42) to assess model robustness. To ensure the validation fold was not used during weighting, researchers calculated sample weights for each training fold in Gradient Boosting using compute_sample_weight. In contrast, Logistic Regression and Random Forest were evaluated using scikit-learn's cross_validate(), with Logistic Regression performed inside a pipeline that included StandardScaler with the classifier. Researchers recorded the mean and standard deviation of accuracy, precision, recall, F1-score, and ROC-AUC across the five folds for each model. To assess if the performance difference between Random Forest and Gradient Boosting was significant, McNemar's test was also performed

on their predictions on the held-out testing set. The test compared the models' predictions using a 2x2 contingency table (Table 1, and the mcnemar() function from the statsmodels library was used with continuity correction (exact=False, correction=True) to calculate the test statistic and p-value. After comparative evaluation, the top-performing model was analysed using feature importance and SHAP, then selected for incorporation into the decision-support framework. Figure 3 summarises the complete model development and selection process.

3.5. Development of the Inflation Stress-Testing Framework

After selecting the best-performing model, researchers created an inflation stress-testing framework to assess how changes in economic conditions would affect borrower credit risk. Instead of including inflation as a predictor variable during model training, inflation was simulated before making new predictions, ensuring that any variation in estimated default risk reflected changing economic conditions rather than changes to the ML system. Three inflation scenarios were created to reflect the rising levels of financial stress: mild, moderate and severe inflation. Each scenario altered variables expected to be influenced by inflation, such as Monthly Income, Debt-to-Income Ratio, and Revolving Utilization of Unsecured Lines. Monthly Income was lowered to reflect the reduced buying power, while the debt ratio and revolving credit utilisation were raised to imitate the escalating financial pressure on borrowers. The size of these modifications was based on observable UK economic scenarios, as shown in Table 1: income was multiplied by 0.95, 0.90 and 0.85; debt ratio by 1.10, 1.20 and 1.30; and utilisation by 1.05, 1.15 and 1.25 for the mild, moderate and severe scenarios, respectively (Table 2).

Table 2: Inflation stress-testing scenario parameters

Scenario	Monthly Income	Debt Ratio	Revolving Utilization
Mild	×0.95	×1.10	×1.05
Moderate	×0.90	×1.20	×1.15
Severe	×0.85	×1.30	×1.25

The “mild” scenario is comparable to the real household disposable income decline of 4.3% in 2022-2023, which coincided with CPI inflation peaking at 11.1% and was the worst single-year decline since ONS records started in 1956-57. Model behaviour can be examined beyond the range of inflation through the “moderate” and “severe” scenarios, which represent larger hypothetical shocks at approximately two and three times the recorded 2022-2023 event. To ensure each inflation scenario produced a significant change in the data, researchers generated the Population Stability Index (PSI) for the three stressed variables using the original test set as a baseline, allowing us to quantify changes in the data distribution as well as changes in default probability. Therefore, stress testing can support the evaluation of financial risk under economic conditions rather than relying on historical model performance. Finally, the trained Gradient Boosting model was applied to each stressed dataset; for each case, new default probabilities were calculated and compared with the original predictions. Borrowers were then classified as (Low, Medium, High, and Very High Risk) to examine how the economy influences lending choices, allowing the system to provide a more dynamic rating while still using a consistent prediction model.

3.6. Design and Implementation of the Inflation-Aware Decision Support Framework

After completing the model creation and inflation stress-testing scenarios, the next step was to transform the Python analysis into the online application. Using the joblib library, we saved the trained Gradient Boosting model as credit_risk_model.pkl. The feature names used during model training were also stored separately as feature_names.pkl, allowing the application to reproduce the same input structure as the trained model. A new Python file, app.py, was then created to design the user interface using Streamlit. The program allows users to input borrower information such as age, monthly income, debt ratio, revolving credit usage, credit limits, payback history and number of dependents.

Once the user enters this information, the application recreates the borrower profile, produces the engineered variables the model needs, and sends the full record to the stored Gradient Boosting classifier. predict_proba() generates the prediction result, delivering the estimated probability of default risk and generating support suggestions. The inflation stress-testing logic from Section 3.8 was also included in app.py, allowing the borrower to be evaluated under Mild, Moderate, and Severe inflation scenarios. To make the results easier to understand, the interface was designed as a dashboard rather than only showing numerical model outputs. To provide more transparency around the prediction outcomes, we also included other analytical elements, such as borrower-level sensitivity analysis and model feature importance. Once the program was functioning, we produced the files required for deployment, such as app.py, credit_risk_model. pkl, feature_names.pkl, and a requirements.txt file listing the Python libraries needed to run the application.

A runtime.txt file was also used to manage the Python environment during deployment, ensuring compatibility with the stored ML model. Afterward, a GitHub repository was built to contain the application’s code, and the finalised application files were

submitted to the repository. The GitHub repository was then linked to the Streamlit Community Cloud, and app.py was chosen as the primary application file. Streamlit automatically installed the dependencies specified in requirements.txt, loaded the trained model, and created the online version of the framework. The final implementation transformed the ML and stress-testing code into an interactive, browser-accessible application, demonstrating how the trained credit risk model could be integrated into a practical lending decision-support tool.

3.7. Rejected Alternatives

Before deciding on the final methodology, researchers considered several approaches. SMOTE was not used to correct class imbalance since creating synthetic borrower records may modify the dataset's original features. Instead, balanced sample weighting was used to preserve all original observations while improving minority-class detection. The final model was also rejected since the dataset included about 93% non-default borrowers, which made accuracy potentially misleading.

As a result, precision, recall, F1-score, and ROC-AUC were used, as well as Stratified 5-Fold Cross-Validation and McNemar's test. Further, inflation was not included as a direct predictor variable; instead, researchers used stress-testing scenarios to simulate economic conditions. Also, more complex models, such as deep learning and XGBoost, were excluded to ensure a fair comparison of predicted performance, interpretability, and implementation complexity.

3.8. Ethical Considerations

Ethical questions remain essential because AI systems can directly affect lending decisions and, in turn, people's access to financial services. The research uses the publicly available Give Me Some Credit dataset, which was anonymised before release, ensuring privacy and confidentiality risks remained minimal. Nonetheless, historical financial statistics can reflect pre-existing social or economic bias, and ML algorithms trained on such data can recreate similar patterns without appropriate precautions. As a result, the suggested framework is designed as a decision support tool, aiming to help financial professionals by providing additional analytical insights while retaining the importance of human judgement in final loan decisions.

Transparency was also emphasised during the research, as multiple ML models were evaluated before selecting the final predictive algorithm, and feature importance analysis was conducted to improve understanding of the variables influencing the model predictions. This method increases user trust in the system while helping them to evaluate AI-generated results. Finally, the methodology includes inflation stress testing to better understand model behaviour under changing economic conditions, rather than assuming prior performance will always hold. The methodology highlights AI's potential benefit in credit risk assessment, although it was not designed to replace human judgement; instead, it adds another analytical tool for making better-informed loan decisions during times of economic instability.

4. Results and Discussions

This paper summarises the findings from developing and evaluating the proposed Inflation-Aware AI Credit Risk Decision Support Framework. Before developing the model, the paper begins with an EDA to examine the dataset features. The paper then determines the best-performing algorithm by comparing the predictive performance of the three ML models. After selecting the top-performing models, the paper conducts a feature importance analysis to examine which borrower attributes most affect default risk. Afterward, the paper describes how the proposed framework is implemented and assesses its performance in simulated inflation scenarios. Lastly, the results are examined in relation to the research objectives and the practical application of AI credit risk assessment.

4.1. Exploratory Data Analysis (EDA)

EDA was conducted to understand the dataset's characteristics and identify issues that could affect prediction performance before developing the ML models. The prediction task is a binary classification problem because the target variable, SeriousDlqin2yrs, indicates whether a borrower suffered financial delinquency within two years.

Data preprocessing was done before model training, and missing values in the Monthly Income and Number of Dependents variables were replaced using median imputation; duplicate records were removed, and unrealistic observations were corrected. Further, new variables were developed through feature engineering, such as (CreditLinesPerAge) and (IncomeToDebt), to provide more detail on borrowers' financial behaviour.

4.1.1. Class Distribution

The distribution of the target variable was analysed to assess if the dataset had balanced class frequencies.

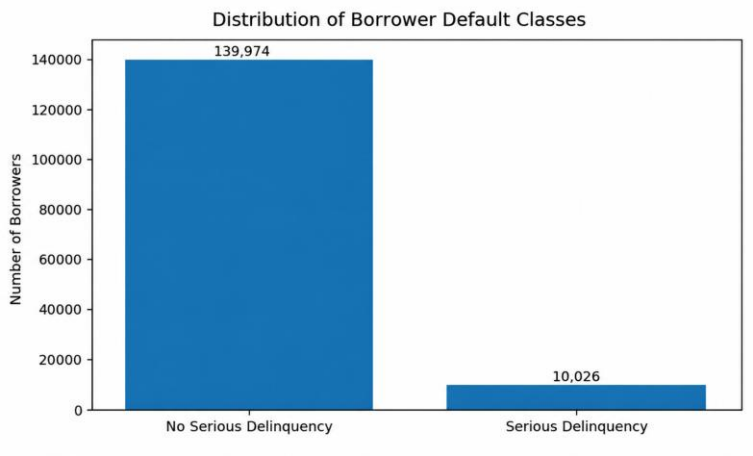


Figure 4: Dataset class distribution

Figure 4 and Table 3 show the count and percentage of borrowers who experienced serious delinquency compared with those who did not. The dataset shows significant class imbalance, with 93% of borrowers in the non-default category and just 7% in the default category. This imbalance, according to Zhao and Zou [37], is typical in credit risk datasets because defaults occur less frequently than successful loan repayments.

Table 3: Class distribution of the credit risk dataset

Class	Count	Percentage
No Serious Delinquency	139974	93.32
Serious Delinquency	10026	6.68

4.1.2. Distribution of Key Borrower Characteristics

To better understand the borrower population, researchers analysed the distributions of age (Figure 5) and monthly income (Table 4) after data cleaning.

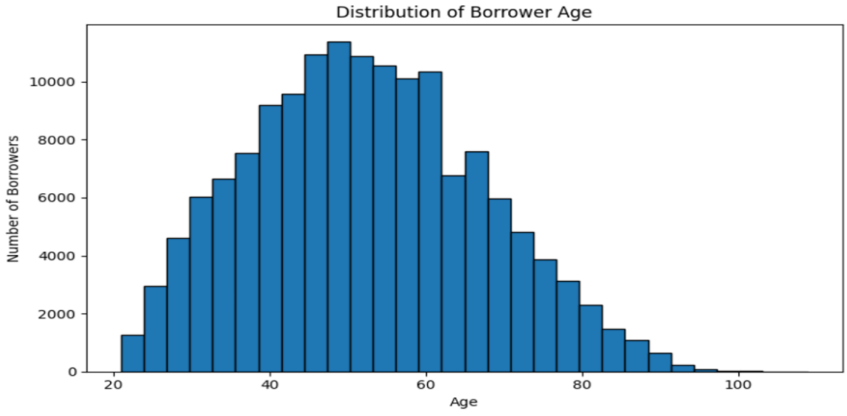


Figure 5: Distribution of borrower age

The age distribution shows that most borrowers are economically active adults, with fewer observations among younger and older individuals. After deleting unrealistic values during preprocessing, the distribution matches the predicted demographic profile of those seeking customer credit.

Table 4: Descriptive statistics of the engineered features

	Debt Per Dependent	Credit Lines Per Age	Income To Debt
count	150000.000000	150000.000000	1.500000e+05

mean	301.527526	0.169249	4.307583e+03
std	1782.173573	0.105246	1.279459e+04
min	0.000000	0.000000	0.000000e+00
25%	0.090326	0.093750	1.233142e+03
50%	0.228309	0.151515	3.385536e+03
75%	0.701663	0.225000	5.724012e+03
max	329664.000000	1.357143	3.004333e+06

Monthly income is positively skewed, with borrowers reporting modest salaries and few observations indicating higher values (Figure 6).

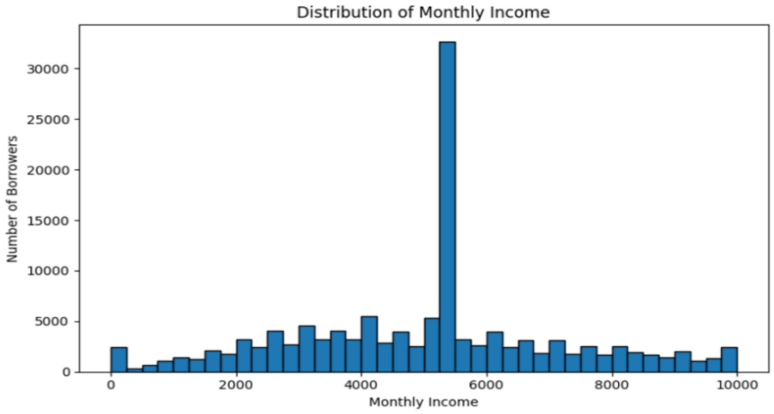


Figure 6: Distribution of monthly income

As a result, median imputation was preferred over mean imputation because the median is less affected by extreme values. The concentration of borrowers around £5,400 reflects the median imputation used to replace missing monthly income values during preprocessing (Figure 7).

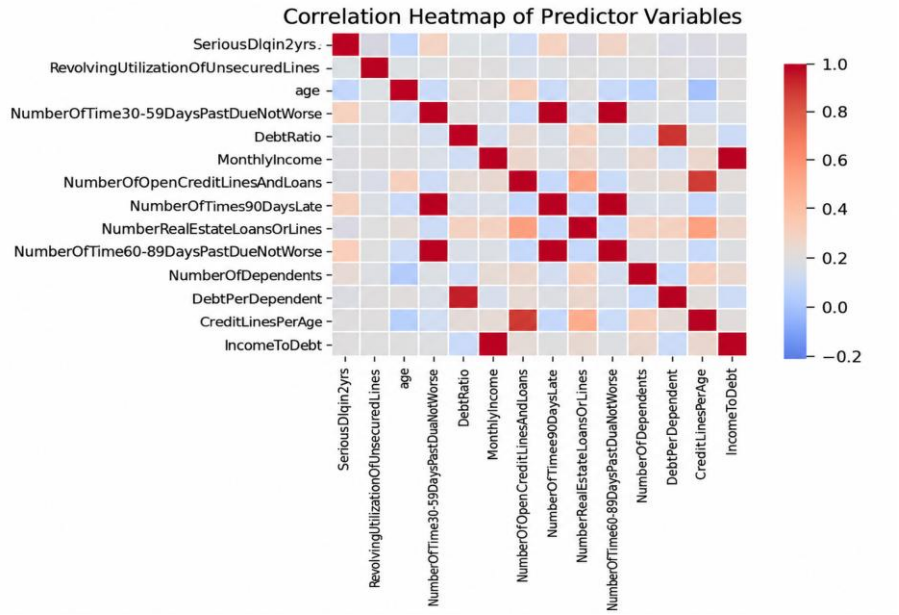


Figure 7: Correlation heatmap

The analysis of the predictor variables showed no evidence of significant multicollinearity that could impede the model's development. Although the variables related to previous late payment behaviour were naturally connected, these correlations were expected, as they capture different parts of borrowers' delinquency history. Overall, the exploratory study indicates that the dataset has enough diversity across borrower features while retaining enough data quality for predictive modelling.

4.2. ML Model Performance

To determine the most appropriate predictive model for the proposed framework, researchers trained and evaluated three supervised machine learning algorithms using the same pre-processed dataset: Logistic Regression, Random Forest and Gradient Boosting. On the held-out test set, the models achieved mean accuracies of 0.784, 0.935, and 0.792, with ROC-AUC values of 0.802, 0.843, and 0.869 (Table 5).

Table 5: Performance comparison of machine learning models

	Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
0	Logistic Regression	0.784200	0.186562	0.663342	0.291220	0.801939
1	Random Forest	0.935133	0.556622	0.144638	0.229612	0.842943
2	Gradient Boosting	0.792300	0.213686	0.786534	0.336068	0.868844

The results show that each model has different prediction qualities. Logistic Regression had a recall of 0.663 but poor precision (0.186), indicating that although it identified a large proportion of defaulting borrowers, it also produced many false positives (Table 6).

Table 6: Cross-validation performance comparison of machine learning models

	Model	Accuracy_ Mean	Accuracy_Std.	Precision_Mean	Precision_Std.	Recall_Mean	Recall_Std.	F1_Mean	F1_Std	Roc_Auc_Mean	Roc_Auc_Std.
0	Logistic Regression	0.775493	0.002331	0.177192	0.001827	0.647317	0.004091	0.278220	0.002408	0.790402	0.003892
1	Random Forest	0.934880	0.000538	0.548694	0.015582	0.145222	0.006606	0.229599	0.009027	0.842152	0.004349
2	Gradient Boosting	0.792993	0.003409	0.213257	0.002271	0.779674	0.013466	0.334887	0.003136	0.865262	0.003549

This could lead to unnecessary rejection of creditworthy applicants, limiting the model's applicability for lending decisions where balancing risk and client approval is crucial.

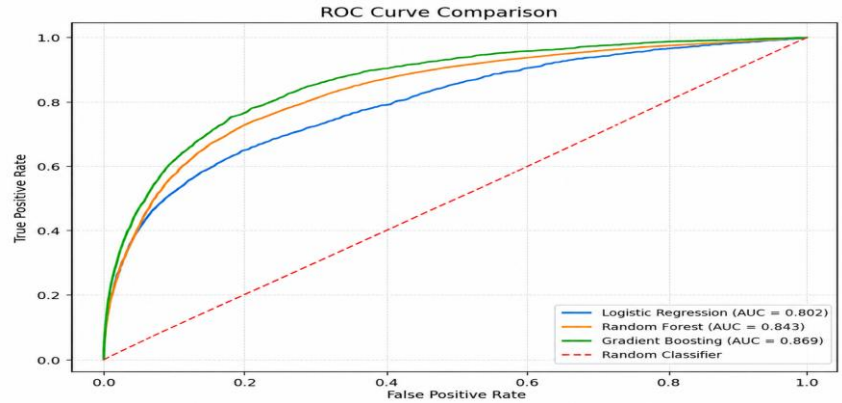


Figure 8: ROC curve comparison

On the other hand, Random Forest outperformed Logistic Regression with an accuracy of 0.935 and precision of 0.557, indicating a better ability to identify and differentiate low- and high-risk borrowers. However, its recall dropped significantly to 0.144, suggesting that the model missed many default cases. The model's limited ability to identify financially vulnerable borrowers reduces its effectiveness in credit risk assessment, as it may misclassify borrowers likely to default as low risk.

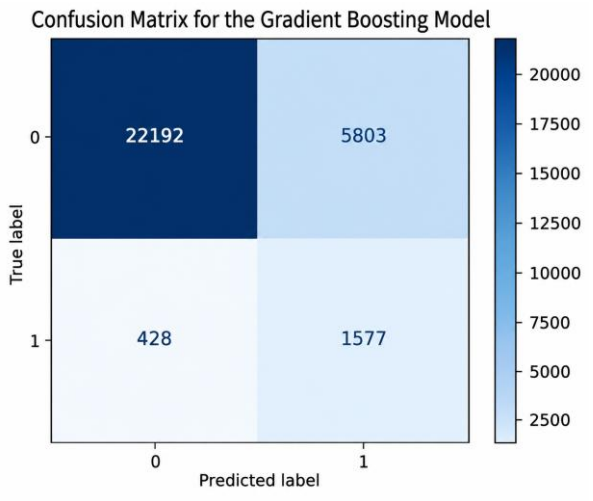


Figure 9: Confusion matrix for the gradient boosting model

Lastly, Gradient Boosting achieved an accuracy of 0.792, precision of 0.214, recall of 0.787, F1-score of 0.336, and ROC-AUC of 0.869. Although it has lower precision, the model identified most borrowers who defaulted while maintaining the highest ROC-AUC of the three models. Stratified five-fold cross-validation produced similar results. Gradient Boosting achieved a mean ROC-AUC of 0.865 (±0.004) and a mean recall of 0.780 (±0.013), further supporting the consistency of its performance. Random Forest maintained the highest mean accuracy (0.935) but considerably lower mean recall (0.145), while Logistic Regression achieved a mean ROC-AUC of 0.790 and recall of 0.647. McNemar's test was also conducted on the held-out test predictions of both Random Forest and Gradient Boosting, yielding a statistic of 2674 and $p < 0.001$, indicating a significant difference between the two models' prediction outcomes. The ROC curves show that Gradient Boosting performed better than the other two models across different classifications, with an ROC-AUC of 0.869, indicating a stronger ability to differentiate between borrowers likely or unlikely to default (Figure 9). The confusion matrix shows that Gradient Boosting identified a high proportion of actual default cases but also produced more false-positive predictions (Figure 8). Altogether, these findings show that ensemble learning techniques perform better in this dataset than traditional methods. More importantly, the results indicated that the best-performing model is not simply the one with the highest recall, but a combination of precision and predictive performance.

4.3. Selection of the Best Performing Model

Selecting the final prediction model required considering all assessment criteria rather than focusing only on classification accuracy. While Random Forest performed better than the other two models in accuracy and precision, it identified only a small percentage of real defaults, limiting its applicability when the main goal is detecting high-risk borrowers. Logistic Regression found a much higher percentage of defaulting borrowers, but it also produced many false-positive predictions that could lead to the unfair rejection of creditworthy applicants. To solve the dataset's class imbalance, Gradient Boosting was trained with balanced sample weighting. This resulted in a model that maintained a ROC-AUC of 0.869 while identifying 78.7% of real default cases (recall=0.787), the highest recall of the three models.

4.4. Feature Importance Analysis

After selecting Gradient Boosting as the final prediction model, feature importance analysis identified the factors that most affected credit risk prediction. Figure 10 shows that (Revolving Utilisation of Unsecured Lines) was the most significant predictor, with around 54.2% of the model's total feature importance, suggesting that credit utilisation is a key early indicator of default probability. The second most important variable was (NumberOfTimes90DaysLate), followed by (NumberOfTime30–59DaysPastDueNotWorse) and (NumberOfTime60–89DaysPastDueNotWorse). These variables highlighted the importance of borrowers' credit management and repayment history in predicting financial risks. In fact, borrowers with high credit utilisation and frequent late payments are more likely to face future repayment issues, which makes these factors useful predictors in the model.

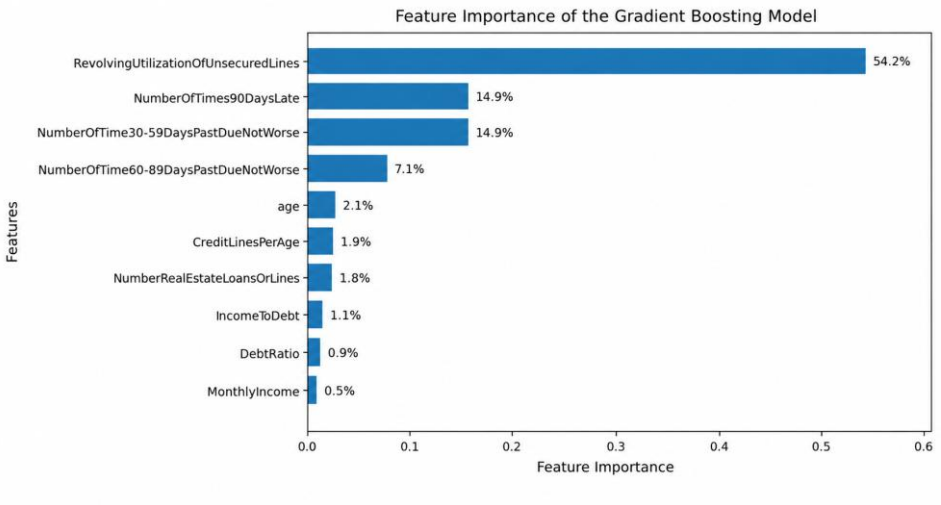


Figure 10: Feature importance of the gradient boosting model

Financial variables such as (Age, CreditLinesPerAge and NumberRealEstateLoansOrLines) also contributed to the prediction process, although with a smaller influence. Similarly, by capturing connections between borrower affordability and credit exposure, the designed variables (IncomeToDebt, DebtRatio and MonthlyIncome) provided more explanatory data, which together enhanced the model's ability to differentiate between borrowers with different levels of financial risk, despite their lower individual significance scores. Overall, the feature importance analysis showed that, to evaluate borrower creditworthiness, the Gradient Boosting model relies primarily on borrowers' available revolving credit currently in use, while also incorporating other factors. These results align with Chang et al. [12] and Machado et al. [24], which highlighted credit utilisation and delinquency history as important factors influencing default risk.

4.5. Local Explainability With SHAP

SHAP values were calculated using TreeExplainer on a sample of 1,000 test borrowers. (Figure 11) presents the summary plot, which confirms that Revolving Utilisation of Unsecured Lines remains the most important predictor, with high utilisation (red) consistently pushing predictors toward higher default risk and low utilisation (blue) pushing predictors lower, followed by the payment-delinquency variables.

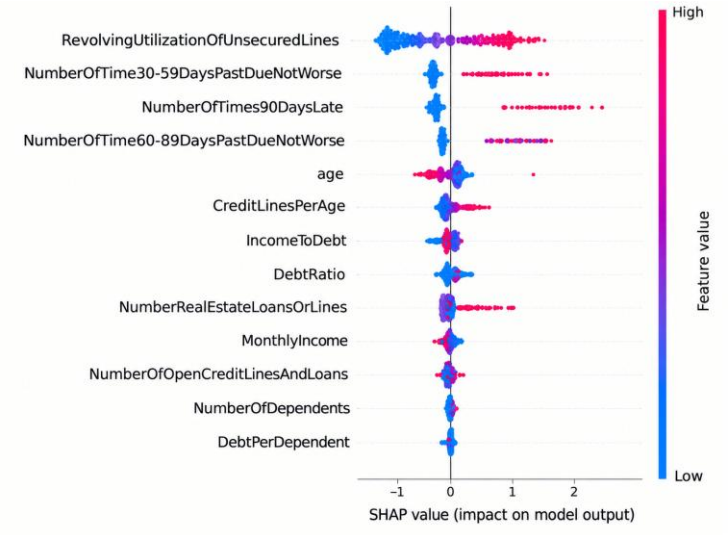


Figure 11: SHAP summary plot

Figure 12 shows this at the individual level for a borrower from the test set. This borrower's low credit utilisation (0.121) was the greatest single contribution to their projection, lowering it by 0.87, while a higher-than-average CreditLinesPerAge ratio (0.356) raised it by 0.34.

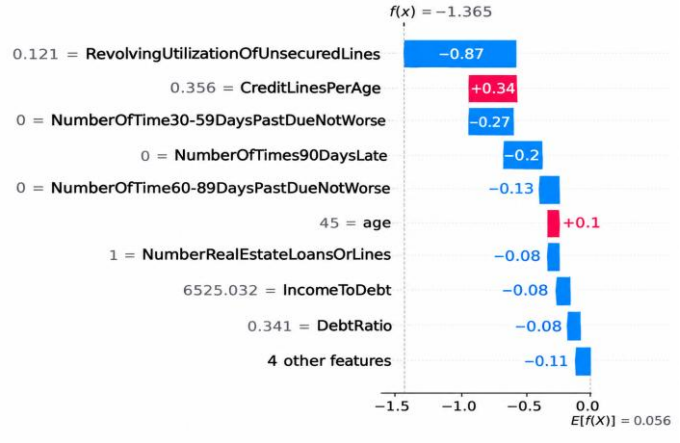


Figure 12: SHAP waterfall plot

The combined influence of all characteristics produced a final projected value of -1.365, equivalent to a default chance about 20.3% below the model's average baseline, highlighting how the framework gives lending professionals a clear, transparent basis for risk assessment.

4.6. Inflation Stress Testing Results

The final stage of this research examined how the framework behaved under different inflation scenarios; instead of relying only on fixed financial data, it allows users to understand how borrower risk can change if economic conditions worsen. Further, three inflation scenarios: mild, moderate and severe, were created by reducing monthly income while increasing debt ratios and revolving credit usage, resulting in progressively higher financial stress on borrowers.

Changes in Default Probability: Table 7 shows the average probability of default, which was derived by the Gradient Boosting model for each inflation scenario.

Table 7: Average predicted default probability

Scenario	Average Default Probability
0 — Normal	0.3264
1 — Mild Inflation	0.3410
2 — Moderate Inflation	0.3534
3 — Severe Inflation	0.3645

The results have shown an increase in expected default as inflation worsens. Under normal circumstances, the average probability of default was 32.64%, increasing to 34.10%, 35.34% and 36.45% under mild, moderate and severe scenarios, respectively. This supports Sharma et al. [32] and the Bank of England [7], who suggest that inflation and unstable economic conditions increase borrowers' financial vulnerability and increase default risk. Although the probability changes seem small, they remain consistent across the sample and show that even a small deterioration can change lending risk. The gradual progression across all three scenarios shows that the model produces consistent predictions rather than drastically changing borrower classification.

Changes in Borrower Risk Categories: To better improve the interpretability of the predictions, borrowers were grouped into four risk categories based on their predicted probability of default: Low Risk, Medium Risk, High Risk, and Very High Risk (Table 8).

Table 8: Borrower risk distribution under different inflation scenarios

Risk Category	Normal	Mild	Moderate	Severe
Low Risk	16012	15632	15114	14647
Medium Risk	6608	6359	6287	6176
High Risk	4596	4841	5133	5483
Very High Risk	2784	3168	3466	3694

Borrower risk changes in all of the inflation scenarios (Figure 13).

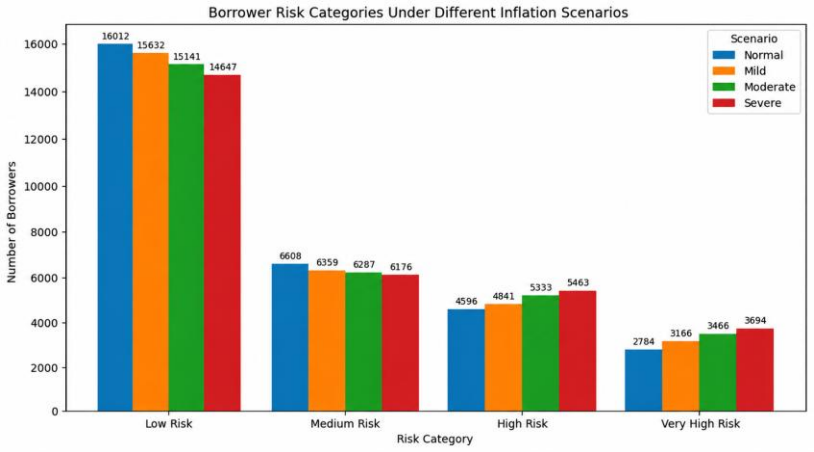


Figure 13: Borrower risk categories under different inflation scenarios

As conditions deteriorate, the percentage of borrowers in the low-risk group gradually decreases, falling from 53.4% under normal scenarios to 48.8% with high inflation. Meanwhile, the high-risk category increased from 15.3% to 18.3%, and the very-high category increased from 9.3% to 12.3%. The medium-risk category decreased slightly from 22.0% to 20.6%. Overall, the results showed a clear shift from lower to higher categories as economic stress increased.

4.7. Implementation of the Inflation-Aware AI Credit Risk Decision Support Framework

Following model evaluation, the Gradient Boosting model was included in the proposed framework, developed in Python and Streamlit. The application follows a simple workflow which includes four integrated modules: Borrower Assessment, What-if Simulator, Portfolio Batch Assessment and Model Governance & Monitoring. Borrower evaluation starts with an interactive input interface that contains the financial and behavioural variables required by the trained Gradient Boosting model. Users enter information such as financial and repayment details, and once submitted, the application automatically processes the data and generates an estimated probability of default (Figure 14).

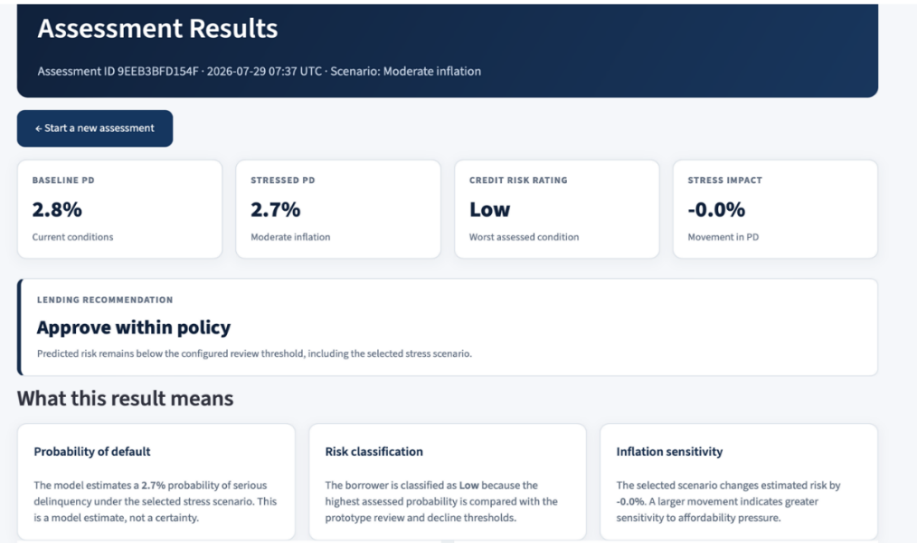


Figure 14: Assessment results 1

The assessment dashboard shows the baseline and stressed default probability, risk categorisation, predicted stress effect and lending recommendation (Figures 14 and 15). The additional visualisations explain the prediction using feature contribution analysis, borrower sensitivity to inflation, and stress financial variables adjusted based on the stress simulation. This improves

transparency and supports better decision-making. The framework also includes a What-if Simulator that lets users view different economic scenarios and examine how they influence borrower default probability.

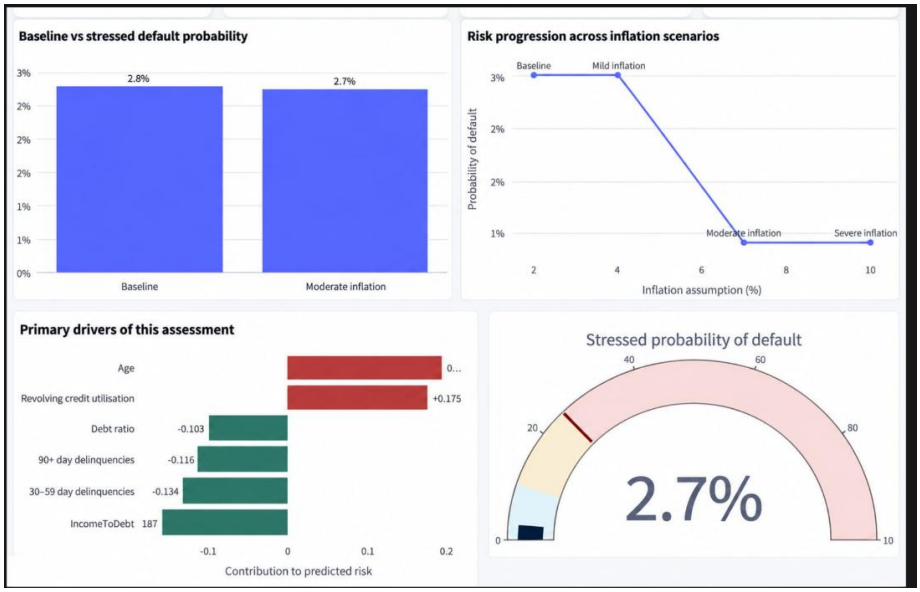


Figure 15: Assessment results 2

In addition, the Portfolio Batch Assessment module allows users to assess multiple borrowers' profiles at the same time by uploading a CSV file. In contrast, the Model Governance & Monitoring module shows assessment records, monitoring activities and governance controls that could further support the responsible use of AI in lending. The final trained model, along with its feature set and training environment, was saved to ensure results could be reproduced and the model could be properly governed. In line with responsible AI governance standards, this ensures that the framework's predictions can be linked to a particular model artefact. Gradient Boosting, trained with balanced sample weighting to address class imbalance, correctly detected 78.7% of default instances and achieved the best balance between recognising true defaulters and discriminative performance (ROC-AUC = 0.869). Although this resulted in lower accuracy than an unweighted configuration, this compromise aligns with this study's priority of flagging default while maintaining reasonable discrimination. Additionally, the feature importance analysis identified past repayment behaviour as the strongest predictor of borrower default. In fact, variables relating to previous late payments contributed much more to the prediction process than demographic data. However, this study's main contribution is incorporating stress testing into the decision-support framework, and the results showed that simulated inflation moved borrowers into higher-risk categories and increased predicted default percentages. This suggests that borrower creditworthiness should not be seen as static, especially in uncertain economic times when affordability may decline. Overall, the results showed that integrating ML with inflation stress testing provides a more dynamic approach to assessing credit risk than traditional models. As a result, the proposed approach benefits lending professionals and lays the foundation for future research on integrating economic factors into credit risk decision support systems.

5. Evaluation and Implications

Based on the results, this paper assesses the Novel Inflation-Aware AI framework created for credit risk assessment under economic stress. While the previous paper focused on the framework's implementation and results, this paper discusses the effectiveness of the developed artefact, its practical contribution to credit risk assessment, and its potential application in financial institutions. This paper also discusses the consequences of combining ML with inflation stress testing, evaluates the framework's weaknesses, and proposes areas for future research.

5.1. Evaluation of the Developed Framework

The results demonstrate that the proposed framework met the key goals of this study. The project integrated supervised machine learning, inflation stress testing, and interactive software development into a unified application to evaluate borrower credit risk under various simulated economic conditions. Rather than serving only as a forecast model, the framework was designed to help users make loan decisions by assessing how borrower risk can change as financial conditions worsen. One of the framework's main achievements is effectively integrating ML into an accessible application. In this study, a Gradient Boosting model was incorporated into the Streamlit application, allowing users to input borrower information, generate default forecasts,

run stress tests, and view recommendations through the graphical interface. The comparative study of the three ML algorithms also showed that choosing a predictive model should not be based on a single performance indicator. For example, Logistic Regression had much higher recall than both ensemble models, indicating a larger percentage of failing loans. However, this improved performance came with a significant drop in accuracy, resulting in more false-positive predictions. In a lending environment, many false positives can lead to the unjustified rejection of financially competent candidates, reducing loan chances and consumer satisfaction. On the other hand, Random Forest performed better in accuracy but detected fewer defaulting borrowers, limiting its practical use in situations where financial risk detection is the primary goal.

Lastly, Gradient Boosting achieved a strong balance between overall discriminative ability and detecting actual defaults, accepting reduced accuracy and precision as a compromise. Given the greater risk of failing to identify a genuine defaulter in financial environments, this behaviour aligns with the study's assessment that recall and ROC-AUC matter more than raw accuracy, making it the most suitable model to include in the proposed framework. Beyond predicted accuracy, the feature importance analysis provided valuable insights about the selected model's decision-making process. According to the results, the variable (RevolvingUtilizationOfUnsecuredLines) contributed the most to total model relevance, along with additional indicators related to past late payments. One of the framework's most important contributions is integrating inflation stress testing into the credit risk assessment process. In fact, traditional ML credit rating algorithms often assess borrowers' attributes in a stable historical context, whereas by lowering income while raising debt and credit usage, the suggested approach enables consideration of borrower affordability in more difficult economic situations. This reflects that borrowers' financial situations can change over time, especially during periods of inflation. Across all inflation scenarios, the stress testing results showed an increase in the default risk probability. The methodology produced higher default probabilities and an apparent transition in borrowers from lower- to higher-risk categories as financial pressure shifted from moderate to severe inflation.

However, instead of producing sudden changes in borrower risk, the framework responded gradually, suggesting stable sensitivity to increasing simulated stress; however, it does not develop predictive stability under real inflation conditions. From a practical perspective, these results show that integrating macroeconomic simulation into credit risk assessment can help lenders access data not available with traditional credit scoring techniques. Before approving a loan, the approach allows lenders to consider how future economic decline can affect repayment capability, rather than assessing borrower affordability based only on current financial conditions. This encourages more proactive risk management and aligns with research highlighting the growing importance of accounting for economic uncertainty when assessing credit risk. The framework also supports responsible AI by serving as a decision-support tool rather than replacing human decision-making. In fact, the goal was to give extra insights that could enhance current decision-making processes rather than replace professionals. In addition to contextual information that may not be included in the datasets, users still interpret the prediction, and human supervision remains a significant role in high-impact financial choices. Overall, the proposed methodology shows that integrating interactive software implementation, inflation stress testing and ML predictions can produce a useful tool that further enhances credit risk assessment under changing economic conditions, while giving financial institutions a more thorough understanding of borrower vulnerability.

5.2. Evaluation Against Research Aim and Objectives

To assess if the study achieved its intended purpose, the research findings are compared to the aim and six objectives outlined (Table 9).

Table 9: Summarises the achievements made in this study

Objective	Evidence of Achievement	Achievement
Literature review	• Comprehensive critical review of relevant literature • Research gap and justification for the study established.	✓ Achieved
O2Compare ML models	• Logistic Regression, Random Forest and Gradient Boosting developed and compared • 5-fold cross-validation and multiple metrics used.	✓ Achieved
Select suitable model	• Gradient Boosting selected as the most suitable model based on performance and trade-offs • McNemar's test identified a statistically significant difference between the models ($p < .001$).	✓ Achieved
Develop inflation methodology	• Three inflation stress scenarios (mild, moderate, severe) developed, anchored to UK economic data • Distributional changes validated using Population Stability Index (PSI).	✓ Achieved
Develop framework	• Gradient Boosting model and stress-testing methodology integrated into a Streamlit decision support application	✓ Achieved

	• Features include borrower assessment, what-if simulator, portfolio analysis and monitoring.	
Investigate inflation effects	• Predicted default probability increased across scenarios • Borrowers migrated to higher risk categories with increasing stress severity.	✓ Achieved within simulated conditions

Overall, the research aim was achieved by creating a functional framework. However, the results demonstrate the approach's feasibility rather than its potential effectiveness in real lending environments, as the inflation scenarios were simulated rather than based on real economic conditions.

5.3. Practical Implications

The results of this study have several practical implications for financial organisations that are seeking to improve credit risk assessment during times of economic instability. While many current credit scoring systems rely mostly on historical borrower data, the proposed approach shows how ML can integrate simulated macroeconomic events to provide a more dynamic evaluation of borrower creditworthiness. One of the framework's main advantages is its ability to support better-informed financing choices. In fact, traditional credit evaluation methods typically assess borrowers based on financial information gathered at the time of application. Although this approach remains useful in stable economic conditions, it can become less reliable during inflation, rising interest rates or declining household affordability. Borrowers who are financially stable at first can become more vulnerable as their income declines and their debt rises. Therefore, including inflation stress testing in the assessment process allows lenders to examine how these changes can affect default risk before deciding whether to approve a loan. This capability can also help enhance portfolio risk management. In fact, instead of examining individual borrowers separately, financial institutions can use comparable stress-testing approaches across larger loan portfolios to identify groups of clients who may become economically vulnerable under certain economic scenarios. Early identification of higher-risk borrowers could help lenders apply suitable risk-mitigation methods such as increased monitoring, altered lending regulations, or early client assistance. The Portfolio Batch Assessment module included with the Streamlit app further demonstrates this potential by enabling multiple borrower profiles at the same time rather than focusing on a single application.

Although the current implementation is only a research prototype, similar capability could be developed for banking systems to assess thousands of borrowers under various economic scenarios. The framework can also benefit financial technology (FinTech) companies and digital lending platforms. Automated lending systems increasingly rely on ML algorithms to quickly assess credit applications while maintaining consistency in decisions. Integrating inflation stress testing into these systems could help them adapt to changing conditions without redesigning the prediction model. Responsible AI is another key factor in financial services, especially in industries like banks where transparency, accountability and governance are critical. The methodology used in this study was intended to assist loan professionals by providing additional insights, including estimated default probability, borrower risk classification, and inflation stress-testing results. This strategy can encourage better, more informed decision-making and reduce the risk of overreliance on automated advice. Also, integrating feature importance analysis improves transparency by revealing which features most influence the model's predictions. The results also have implications for future AI research in credit risk assessment. Much research has focused on boosting predicted accuracy using a more powerful ML system. While predicted accuracy is crucial, this research shows that practical implementation, interpretability and flexibility to changing economic conditions are equally important. Models with slightly higher classification accuracy can offer little practical value if they are not successfully integrated into financial institutions' decision support systems. Although developed in an academic setting, the principles illustrated in this study have broader implications for companies seeking to improve lending decisions during economic instability. As financial institutions continue to use AI technology, frameworks that can integrate both predictive analysis and economic stress simulation can become more effective in promoting ethical and informed lending practices.

6. Conclusion

The increasing complexity of economic conditions has highlighted the limitations of traditional credit risk assessment methodologies, which rely mainly on past borrower data and fail to account for changes in the broader economic environment. Periods of sustained inflation reduce purchasing power, increase debt repayment pressures and weaken affordability, making it harder for financial institutions to assess credit risk. These developments have increased interest in AI techniques that support more adaptive, data-driven lending decisions. This study aimed to create a novel inflation-aware AI framework for credit risk assessment under economic stress by combining ML and inflation stress testing into an interactive application system. The findings show that this objective was accomplished, demonstrating that the framework can be designed, implemented and evaluated to support credit risk assessment under changing economic conditions. To do so, three supervised ML algorithms (Logistic Regression, Random Forest, and Gradient Boosting) were developed and tested on the Give Me Some Credit dataset. After a comparative evaluation, Gradient Boosting was chosen as the best model because of its strong predictive performance

and discriminative ability, which supported the study's goal of avoiding unnoticed default over classification accuracy. The findings also showed that integrating inflation stress testing into the framework increased predicted default probability and moved borrowers into higher-risk categories as simulated inflation scenarios worsened, indicating that borrower credit risk is not static and that considering changing economic conditions provides insights beyond traditional credit scoring methods.

In addition to predictive modelling, the study successfully converted the chosen ML model into a software application that is capable of supporting loan decisions. Through an easily accessible interface, users can enter borrower information, generate default probability estimates, run inflation stress tests, evaluate multiple borrowers at once, and visualise feature importance. The results indicate that combining ML with inflation stress testing can improve support in financial institutions. Rather than replacing human judgement, the framework provides additional insights to help professionals assess borrower affordability under changing economic conditions. This aligns with current trends in responsible AI implementation, where AI supports professional judgement while maintaining transparency and human oversight. Although the study makes an important contribution, it also has several limitations. The framework was developed using a historical dataset and evaluated with simulated inflation scenarios rather than real data. Therefore, although the prototype indicates the potential of inflation-aware credit evaluation, additional development is necessary before implementation in commercial financial institutions. Overall, this study met its stated goals by creating and testing a practical AI system that combines ML prediction with stress testing, and the framework provides a solid foundation for future research on AI systems that can adapt to changing economic conditions for credit risk assessment.

6.1. Limitations

Even if the framework met study goals, there are several constraints to consider when assessing the outcomes. These issues involve the ML model, inflation simulation, dataset, and framework implementation. The study's shortcomings should be considered when assessing the results for future research. The model's dataset is the first limitation. This dataset from Kaggle [19] has approximately 150,000 borrower records and historical financial data from stable conditions, which is utilized in credit risk research. The dataset may not correctly reflect borrower behavior or financial institution lending choices. Job stability, educational background, banking transaction history, and behavioral financial indicators used in real lending environments could improve prediction performance and borrower evaluation, but the dataset lacks them. Second, inflation simulations are limited. Instead of using real data to mimic inflation, the methodology modified monthly income, debt ratio, and revolving credit utilization. This strategy was chosen because it assessed borrower risk consistently without ML model retraining. The real economy is more complicated because inflation affects borrowers through interest rates, unemployment, wages, housing costs, and other factors. The simulated situations are simplistic approximations of economic stress, not actual economic behavior.

Despite Gradient Boosting's success, no ML model can reliably anticipate borrower behavior. Credit default is affected by several environmental, financial, and psychological elements, which organized numerical variables cannot always reflect. Although not reflected in earlier training data, unforeseen life events like illness, layoffs, or big economic shocks can greatly alter repayment behavior. Thus, the framework's projections are probability estimates, not loan outcomes. The methodology was tested using simulated economic circumstances and historical borrower data, not in a professional setting. The study didn't use loan decisions or banking professionals' feedback to test the framework's efficacy. Instead, testing the framework in an industrial setting would reveal its reliability, applicability, and compliance with lending operations. Even though Streamlit provides an efficient platform for building interactive ML applications, the framework as a research prototype lacks scalability, security, regulatory compliance, and integration with existing databases and corporate software. Thus, the framework must be refined before being employed in lending. Despite these drawbacks, the system shows that interactive decision assistance, inflation stress testing, and ML may operate together. Some limitations exist, but the study shows ways to improve the framework. Thus, the framework is a robust proof of concept that provides the groundwork for inflation AI credit risk assessment system enhancements.

6.2. Recommendations

The architecture indicates that ML and inflation stress testing can be connected, however it should be improved. Future advancement may improve regulatory compliance, explainability, prediction accuracy, and financial institution implementation. Borrower factors were carefully changed to simulate economic stress and inflation scenarios under the existing framework. Even if this technique was practicable for measuring borrower risk changes, the framework could contain genuine data from the Bank of England [7] or ONS. By automatically updating inflation, interest, unemployment, and wage growth, the system could improve economic judgments. This method would make the framework more adaptable and beneficial for financial institutions. Adding borrower data to the ML model is another major development area. The present architecture incorporates dataset features such monthly income, revolving credit utilization, debt ratio, and payback history. These criteria provide a good foundation for credit risk prediction, but lenders have more client data. Adding more financial and behavioral

data could improve prediction performance and give the framework a more full picture of borrower affordability. This application is a prototype, not a finished software solution. The framework may be put on cloud infrastructure for greater scalability and multi-user safety.

Integration with financial institutions' databases and loan management systems would increase the framework's practicality by automating borrower information handling rather than manual data entry. To meet organizational and regulatory standards, the system would need cybersecurity, user authentication, and audit logging before commercial use. Expanding the system to evaluate each borrower allows for more detailed examination. Financial institutions monitor several borrowers to uncover new financial vulnerability tendencies. Future portfolio dashboards could measure default risk under inflation, helping risk managers identify sectors or customer groups that need greater attention. A strategy that supports loan options and organizational risk management would benefit financial institutions more. Finally, hybrid ensemble methods and deep learning models could be studied in the future. Gradient Boosting fared best, but AI developments could enable new algorithms to capture more intricate financial data relationships. Compare these models while maintaining interpretability and regulatory compliance to see how credit risk assessment tools will evolve. This study approach provides a solid foundation for future progress. Future research could incorporate real financial data, increase borrower variables, improve explainability, strengthen security, and validate the framework in operational lending environments to make the prototype a complete decision-support system for modern financial institutions.

Acknowledgement: The authors gratefully acknowledge Northumbria University and the University of Derby for their valuable support and encouragement. They also appreciate the academic resources and research facilities provided, which contributed to the successful completion of this research.

Data Availability Statement: The data for this study are available upon request from the corresponding author.

Funding Statement: This manuscript and research paper were prepared without any financial support or funding.

Conflicts of Interest Statement: The authors have no conflicts of interest to declare.

Ethics and Consent Statement: This research adheres to ethical guidelines and obtains informed consent from all participants.

References

1. S. Gregor and O. Zwikael, "Design science research and the co-creation of project management knowledge," *International Journal of Project Management*, vol. 42, no. 3, pp. 1–5, 2024.
2. J. Alvi, I. Arif, and K. Nizam, "Advancing financial resilience: A systematic review of default prediction models and future directions in credit risk management," *Heliyon*, vol. 10, no. 21, pp. 1–20, 2024.
3. H. Ayari, P. R. Guetari, and P. N. Kraïem, "Machine learning powered financial credit scoring: A systematic literature review," *Artificial Intelligence Review*, vol. 59, no. 11, pp. 1–54, 2025.
4. Bank of England, "Financial Stability in Focus: Interest Rate Risk in the Economy and Financial System," *Bank of England,* London, United Kingdom, 2023.
5. Bank of England, "Monetary Policy Report—November 2024," *Bank of England,* London, United Kingdom, 2024.
6. Bank of England, "Financial Stability Report—July 2025," *Bank of England,* London, United Kingdom, 2025.
7. Bank of England, "Financial Stability in Focus: Artificial Intelligence in the Financial System," *Bank of England,* London, United Kingdom, 2025.
8. O. A. Bello, "Machine learning algorithms for credit risk assessment: An economic and financial analysis," *International Journal of Management Technology*, vol. 10, no. 1, pp. 109–133, 2023.
9. Bank of England, "What can we learn from microdata about household savings in recent years?" *Bank of England*, London, United Kingdom, 2024.
10. P. V. Carvalho, J. D. Curto, and R. Primor, "Macroeconomic determinants of credit risk: Evidence from the Eurozone," *International Journal of Finance & Economics*, vol. 27, no. 2, pp. 2054–2072, 2020.
11. J. Černevičienė and A. Kabasinskas, "Explainable artificial intelligence (XAI) in finance: A systematic literature review," *Artificial Intelligence Review*, vol. 57, no. 7, pp. 1–45, 2024.
12. V. Chang, S. Sivakulasingam, H. Wang, S. T. Wong, M. A. Ganatra, and J. Luo, "Credit risk prediction using machine learning and deep learning: A study on credit card customers," *Risks*, vol. 12, no. 11, pp. 1–33, 2024.
13. Credit Connect, "Household disposable incomes fell below pre-pandemic levels in 2023," *Credit Connect*, 2024. [Accessed by 07/01/2025].
14. T. T. Do, G. Babaei, and P. Pagnottoni, "Explainable machine learning for credit risk management when features are dependent," *Measurement: Interdisciplinary Research and Perspectives*, vol. 22, no. 4, pp. 315–340, 2024.

15. A. C. B. Garcia, M. G. P. Garcia, and R. Rigobon, "Algorithmic discrimination in the credit domain: What do we know about it?" *AI & Society*, vol. 39, no. 5, pp. 2059–2098, 2024.
16. T. Schneider, P. E. Strahan, and J. Yang, "Bank Stress Testing: Public Interest or Regulatory Capture?" *Review of Finance*, vol. 27, no. 2, pp. 423–467, 2023.
17. G. Hovakimyan and J. M. Bravo, "Evolving strategies in machine learning: A systematic review of concept drift detection," *Information*, vol. 15, no. 12, pp. 1–24, 2024.
18. A. Ioannou, M. M. Tavalaei, D. Vatn, and P. Mikalef, "The role of explainability in AI-driven financial decision making," *Information Systems Frontiers*, vol. 28, no. 4, pp. 1–17, 2026.
19. Kaggle, "Give Me Some Credit," *Kaggle*, 2011. [Accessed by 12/01/2025].
20. D. Kowald, S. Scher, V. Pammer-Schindler, P. Müllner, K. Waxnegger, L. Demelius, A. Fessl, M. Toller, I. G. Mendoza Estrada, I. Šimić, V. Sabol, A. Trügler, E. Veas, R. Kern, T. Nad, and S. Kopeinik, "Establishing and evaluating trustworthy AI: Overview and research challenges," *Frontiers in Big Data*, vol. 7, no. 11, pp. 1–21, 2024.
21. V. T. P. Lan, "Applying machine learning in credit risk pricing: Opportunities and challenges for the global financial system," *Edelweiss Applied Science and Technology*, vol. 10, no. 4, pp. 745–755, 2026.
22. S. Li, "Machine learning in credit risk forecasting—A survey on credit risk exposure," *Accounting and Finance Research*, vol. 13, no. 2, pp. 107–118, 2024.
23. J. Lu, A. Liu, F. Dong, F. Gu, J. Gama, and G. Zhang, "Learning under concept drift: A review," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 12, pp. 2346–2363, 2019.
24. M. R. Machado, D. T. Chen, and J. R. Osterrieder, "An analytical approach to credit risk assessment using machine learning models," *Decision Analytics Journal,* vol. 16, no. 9, p. 1–14, 2025.
25. S. B. Mannapur, "Understanding data drift and concept drift in machine learning systems," *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, vol. 11, no. 1, pp. 318–330, 2025.
26. A. M. F. Moffo, "A machine learning approach in stress testing US bank holding companies," *International Review of Financial Analysis,* vol. 95, no. 10, pp. 1–19, 2024.
27. M. T. Mohsin and N. B. Nasim, "Explaining the unexplainable: A systematic review of explainable AI in finance," *arXiv preprint arXiv:2503.05966*, 2025. [Accessed by 13/04/2026].
28. C. Molnar, "Interpretable Machine Learning: A Guide for Making Black Box Models Explainable," 2nd ed., *Independently published*, 2022. [Accessed by 15/03 /2024].
29. J. P. Noriega, L. A. Rivera, and J. A. Herrera, "Machine Learning for Credit Risk Prediction: A Systematic Literature Review," *Data*, vol. 8, no. 11, pp. 1–17, 2023.
30. O. E. Aro, "Predictive analytics in financial management: Enhancing decision-making and risk management," *International Journal of Research Publication and Reviews*, vol. 5, no. 10, pp. 2181–2194, 2024.
31. M. N. K. Saunders, P. Lewis, and A. Thornhill, "Research Methods for Business Students," 9th ed., *Pearson*, Harlow, United Kingdom, 2023.
32. H. Sharma, A. Andhalkar, O. Ajao, and B. Ogunleye, "Analysing the influence of macroeconomic factors on credit risk in the UK banking sector," *Analytics*, vol. 3, no. 1, pp. 63–83, 2024.
33. S. Shi, R. Tse, W. Luo, S. D'Addona, and G. Pau, "Machine learning-driven credit risk: A systemic review," *Neural Computing and Applications*, vol. 34, no. 7, pp. 14327–14339, 2022.
34. Z. Lu, D. Wang, and M. Yin, "Does More Advice Help? The Effects of Second Opinions in AI-Assisted Decision Making," *Proceedings of the ACM on Human-Computer Interaction*, vol. 8, no. 4, pp. 1–31, 2024.
35. D. B. Vuković, S. Dekpo-Adza, and S. Matović, "AI integration in financial services: A systematic review of trends and regulatory challenges," *Humanities and Social Sciences Communications*, vol. 12, no. 1, pp. 1–29, 2025.
36. Z. Zhang, "An application of logistic regression in bank lending prediction: A machine learning perspective," *Highlights in Business, Economics and Management*, vol. 45, no. 12, pp. 342–347, 2024.
37. S. Zhao and J. Zou, "Predicting loan defaults using logistic regression," *Journal of Student Research*, vol. 10, no. 1, pp. 1–14, 2021.
38. Z. Zhao, T. Cui, S. Ding, J. Li, and A. G. Bellotti, "Resampling techniques study on class imbalance problem in credit risk prediction," *Mathematics*, vol. 12, no. 5, pp. 1–27, 2024.

Publisher's Note: *The publisher remains impartial concerning jurisdictional claims in published maps and institutional affiliations. Responsibility for the content rests entirely with the authors and does not necessarily reflect the publisher's perspectives.*