Hybrid Banking Decision Support System

Harsh Chhatwani[1] Gaurav Amrutkar[2] Bushara Shaikh[3] Arya Khopey[4]
Mr. Yogeshchandra Puranik[5]
[1,2,3,4]Student [5]Assistant Professor
[1,2,3,4,5]Master of Computer Application
[1,2,3,4,5]PES Modern College of Engineering, Pune, Maharashtra, India

Abstract — Banking decision-support systems increasingly use machine learning to assess credit risk, loan eligibility, and repayment behaviour. However, systems trained on static historical datasets may not adequately reflect changes in borrower behaviour and financial conditions. In addition, predictions may communicate a risk label without explaining influential factors or indicating what applicants could review. This paper proposes a Hybrid Banking Decision Support System that combines historical banking data with sequentially replayed observations, concept-drift detection, adaptive model updating, continuous evaluation, and Explainable Artificial Intelligence (XAI). The system is designed to predict credit or loan risk and provide understandable explanations with advisory improvement guidance. Supporting facilities include EMI calculations, fixed- and floating-rate scenarios, moratorium handling, existing obligations, top-up assessment, and KYC document upload and detail extraction. The planned evaluation will use public datasets whose terms permit research use and compare static and adaptive approaches using Accuracy, Precision, Recall, F1-score, ROC-AUC, drift-detection effectiveness, and update efficiency. The proposed architecture uses Python, Django, MySQL, Scikit-learn, XGBoost, River, SHAP, and LIME. The evaluation will test whether adaptive learning maintains or improves robustness under changing data distributions; an improvement in accuracy is not assumed.

Keywords: Banking Decision Support System, Credit Risk, Loan Eligibility, Streaming Data, Online Learning, Concept Drift, Adaptive Machine Learning, Explainable AI, SHAP, LIME, EMI, KYC

I. Introduction

The banking sector generates large volumes of structured and continuously changing information through loan applications, customer profiles, credit histories, repayment behaviour, account activity, and financial transactions. Machine learning has become an important approach for analysing such information and supporting creditrisk assessment. Traditional statistical techniques such as Logistic Regression remain useful, while models such as Random Forest, Gradient Boosting, and XGBoost can capture nonlinear relationships among financial variables. Prior research has examined explainable bank creditworthiness prediction and machine-learning approaches to consumer credit-risk modelling [1], [2].

The project synopsis identifies a central limitation of conventional credit-risk systems: dependence on static historical data. A model trained once on a fixed dataset may become less representative when customer behaviour or financial conditions change. The proposed system therefore treats new records as an ongoing data stream or as sequentially replayed observations during experimentation. Concept-drift detection is used to identify meaningful changes, after which the model can be updated or retrained according to defined criteria. Work on online concept-drift detection and adaptive learning further motivates explicit monitoring and update strategies when data distributions change [3], [4].

A second contribution concerns decision transparency. The system is not intended to stop at a probability or classification. For an applicant, the platform should explain influential parameters, identify areas of weakness, and provide understandable improvement-oriented guidance. For example, factors such as a low credit score, high existing debt, high loan-toincome ratio, or poor repayment history can be surfaced as contributors to risk. This transforms the platform from a prediction-only model into a decision-support and financial-understanding interface. Related studies also examine explanation stability under class imbalance, SHAP/LIME-based credit-risk interpretation, and transparency and fairness in automated credit decisions [5]–[13].

The proposed platform further connects predictive assessment with practical loan planning. It can support EMI calculations under different assumptions, comparison of fixed and floating rates, moratorium periods, existing obligations, top-up loan considerations, and KYC-document submission. These facilities are treated as decision-support functions rather than as claims of automated bank approval.

II. Literature Survey

Recent literature shows strong interest in machine learning for credit-risk assessment, but also emphasizes interpretability, class imbalance, and adaptation to changing data. The following studies provide the principal research context for the proposed system.

Chang et al. [1]. Studied explainable machine learning for bank creditworthiness and showed how feature-level explanations can identify factors influencing credit decisions and support transparency.

Sowmiya et al. [7]. Applied SHAP and LIME to credit-risk models to explain both global feature importance and individual predictions.

Chen, Calabrese, and Martin-Barragán [5]. Examined interpretable models on imbalanced credit-scoring datasets, including the effects of class imbalance on predictive performance and the stability of SHAP/LIME explanations.

Li and Wu [6]. Compared machine-learning models for loan-default prediction and used SHAP to examine influential predictors.

Montevechi et al. [2]. Reviewed machinelearning approaches to consumer credit-risk modelling and highlighted the importance of transparent modelling procedures before practical adoption.

All rights reserved by www.ijsrd.com

Mulimani et al. [3]. Presented online detection and adaptation methods for concept drift in streaming classification, including incremental updates and sliding-window approaches.

Su, Guo, and Wang [4]. Proposed an online deep-learning approach for dynamic streams and discussed performance degradation caused by changes in data distributions.

Machado, Chen, and Osterrieder [8]. Investigated machine-learning models and SHAP-based interpretation for proactive creditrisk assessment and early-warning analysis in commercial banking.

Nwafor, Nwafor, and Brahma [9]. Studied a hybrid machine-learning approach for automated credit decisions, with attention to explainability and fairness.

Manjunath et al. [10]. Examined Logistic Regression and Decision Tree models with SHAP and LIME for explaining individual credit-risk predictions.

Bhuvaneswari and Muthukrishnan [11]. Examined SHAP and LIME as ways to make AIbased credit-card approval decisions more interpretable.

de Lange et al. [12]. Combined a machinelearning credit-default model with SHAP to explain predictors in a bank credit-assessment setting.

Osterrieder et al. [13]. Discussed the use of LIME and SHAP for local and global explanations of credit-scoring models and the practical challenges of explainability.

III. Research Gap

The project synopsis identifies five major gaps: static credit-risk models, changing data patterns, separation of adaptation and explainability, continuous model monitoring, and efficient retraining. The literature further indicates that interpretability itself can be affected by data characteristics such as class imbalance. These observations motivate a system in which data adaptation and decision explanation are treated as connected components rather than independent features.

- Static credit-risk models: Many systems are trained on historical datasets and evaluated as if the underlying distribution remains stable.
- Changing data patterns: Borrower behaviour and financial conditions can change over time, creating the possibility of concept drift.
- Prediction without actionable explanation: A probability or approval label does not necessarily tell an applicant why the outcome occurred or what can be improved.
- Adaptation and explainability are often separated: Adaptive learning may be studied independently from customer-facing explanations.
- Continuous monitoring and retraining: Realworld deployment requires monitoring predictive performance and determining when updating is justified.
- Practical loan planning is often outside the prediction model: EMI, interest-rate type, moratorium, obligations, top-up loans, and KYC are typically handled as separate processes.

IV. Research Objectives

The main objective is to develop a Hybrid Banking Decision Support System that uses machine learning and data analytics to predict credit risk while adapting to changing banking data and providing transparent, actionable decision support.

1) Study existing credit-risk prediction and banking decision-support systems.
2) Identify limitations and research gaps associated with static machine-learning models.
3) Develop and compare baseline models for credit-risk, loan-default, or loan-eligibility assessment.
4) Process newly arriving banking records sequentially to represent streaming data.
5) Detect concept drift and determine when incremental updating or retraining is required.
6) Use SHAP and LIME to explain both global feature importance and individual applicant predictions.
7) Convert explanation outputs into understandable improvement guidance for applicants.
8) Compare static and adaptive models using predictive and operational metrics.
9) Provide supporting loan-planning features such as EMI calculation, fixed/floating-rate scenarios, moratorium, obligations, and top-up assessment.
10) Provide a KYC document-upload workflow while keeping the predictive model separate from sensitive document storage and validation logic.

V. Proposed System

The proposed Hybrid Banking Decision Support System is a web-based decision-support platform. It combines a conventional batch-learning path with a dynamic path. Historical data is first cleaned, transformed, and used to establish baseline models. New records are subsequently introduced sequentially. The streaming layer evaluates predictions, monitors changes, and invokes an adaptive update or retraining process when predefined conditions are met.

The decision-support layer receives applicant attributes and model outputs and produces three levels of information: (1) a predicted risk/eligibility outcome, (2) the major factors contributing to the outcome, and (3) improvement-oriented guidance. The guidance is presented as decision support and not as a guarantee of future approval.

VI. System Architecture

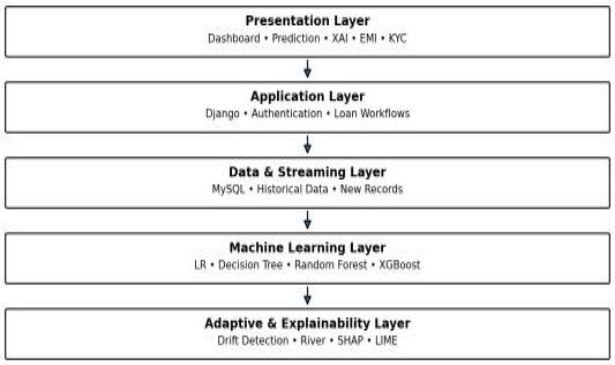


Fig. 1: Proposed Hybrid Banking Decision Support System Architecture

All rights reserved by www.ijsrd.com

The architecture follows a layered design that separates the user interface, application logic, persistent data, and model-specific functions. For this banking system, dedicated machine-learning, streaming/adaptation, explainability, and document/KYC responsibilities are separated to support traceable model updates and controlled document handling.

– Presentation Layer. Provides the applicant dashboard, prediction results, explanation views, EMI tools, loan-planning screens, and KYC upload interface using HTML5, CSS3, Bootstrap, and JavaScript.
– Application Layer. Coordinates authentication, input validation, loan workflows, recommendation logic, and API interactions using Python and Django.
– Data and Database Layer. Stores applicant and application records, engineered features, model metadata, and document metadata using MySQL and the Django ORM.
– Machine-Learning Layer. Trains and compares credit-risk models, including Logistic Regression, Decision Tree, Random Forest, and XGBoost.
– Streaming and Adaptation Layer. Processes sequential observations, tracks online evaluation, detects drift, and manages incremental updates or retraining using River or another suitable implementation.
– Explainability Layer. Produces global and local feature-attribution explanations using SHAP and LIME; counterfactual methods such as DiCE remain an optional extension.
– Document and KYC Layer. Validates uploaded files and manages controlled document handling and associated metadata separately from the prediction model.

VII. WORKING OF THE SYSTEM

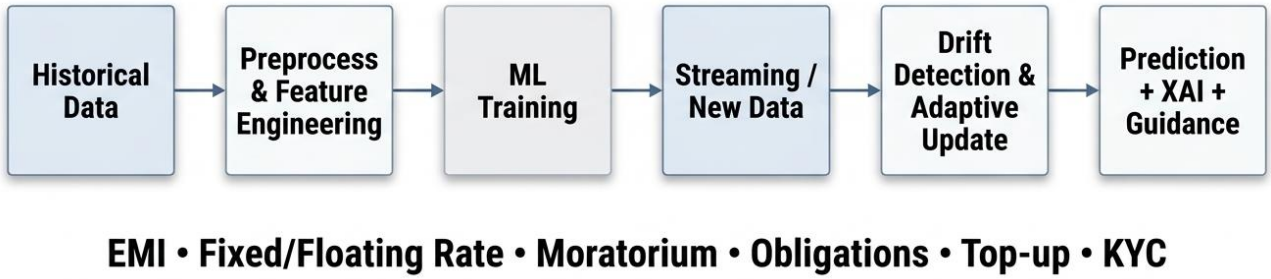
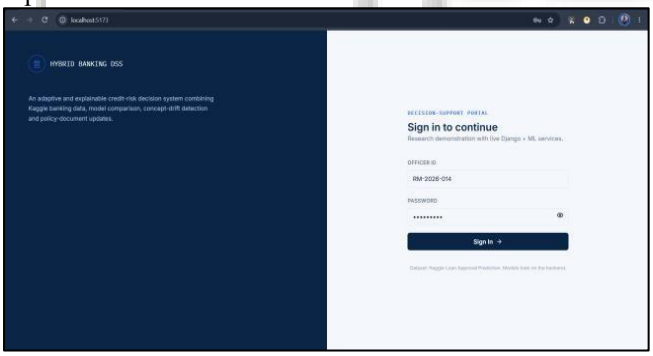


Fig. 2: Working Flow of the Proposed System

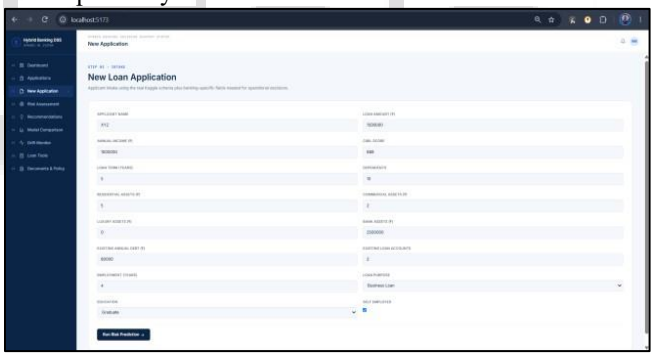
A. User Registration and Authentication

The user registers and authenticates before accessing protected banking decision-support functions. Role-based access can distinguish applicant, analyst, and administrator capabilities.



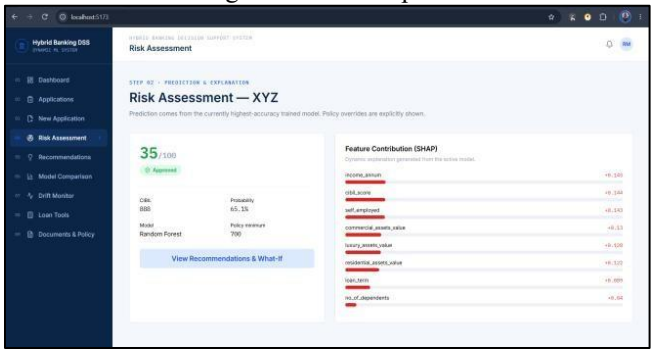
B. Loan Application and Applicant Data

The application captures relevant financial and credit attributes. Candidate features may include income, employment information, credit history, existing liabilities, loan amount, tenure, repayment behaviour, and derived ratios such as loan-to-income or debt-related measures, subject to dataset availability.



C. Baseline Prediction

Historical data is preprocessed and used to train multiple baseline models. Logistic Regression provides an interpretable statistical baseline, while Decision Tree, Random Forest, and XGBoost provide alternative nonlinear approaches. The models are evaluated using consistent train/test or chronological evaluation procedures.



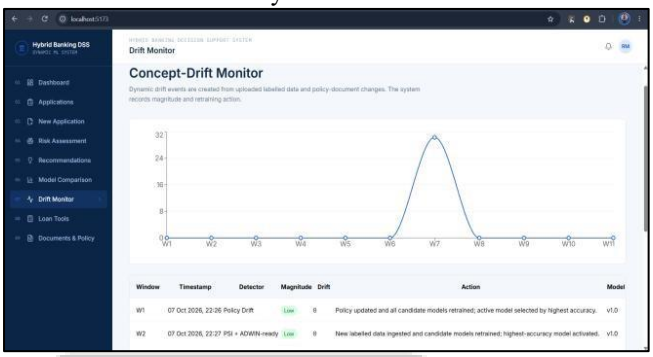
All rights reserved by *www.ijsrd.com*

D. Streaming Data Processing

After baseline training, new observations are processed sequentially. Public datasets can be replayed chronologically or divided into simulated time windows to represent continuously arriving records. This allows the research to compare a fixed model against an adaptive model under changing data.

E. Concept-Drift Detection

The system monitors prediction performance and/or relevant data-distribution changes. When a drift criterion is satisfied, the system records the event and can update the online model or trigger retraining. Model versions and update events are maintained for traceability.

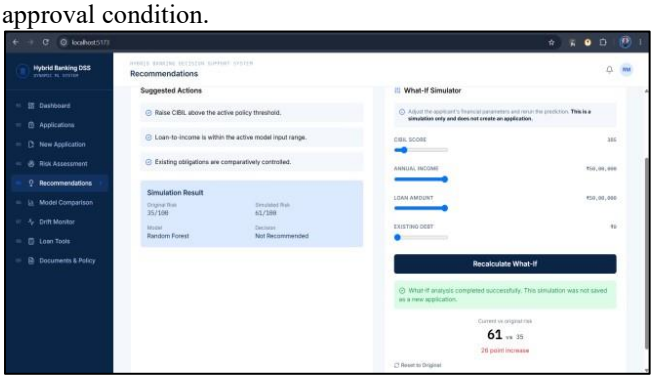


F. Prediction and Risk Interpretation

For an applicant, the system presents the predicted outcome or risk level together with relevant feature contributions. SHAP can provide global and local feature importance, while LIME can provide a local approximation for an individual case.

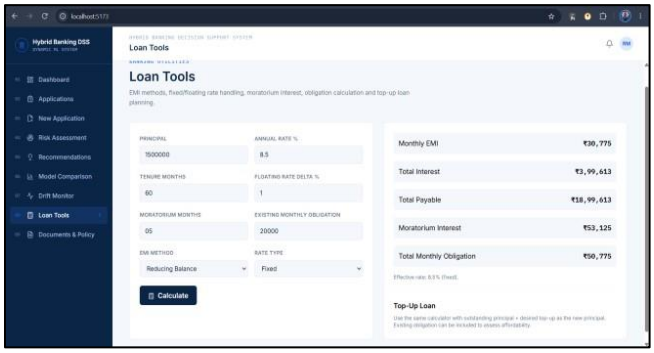
G. Actionable Improvement Guidance

The explanation layer is converted into useroriented guidance. For example, if an applicant has a relatively high existing debt burden, the system can identify that factor as a contributor and suggest reviewing obligations or reducing outstanding debt before a future application. Such guidance is advisory and should not be presented as a guaranteed approval condition.



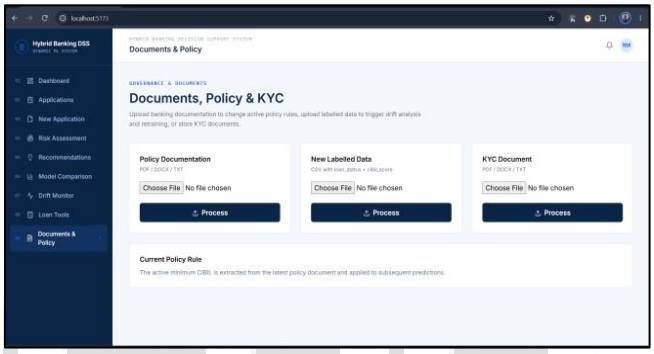
H. Loan Planning Facilities

The system provides calculators and supporting decision views for EMI, fixed versus floating rates, moratorium periods, existing obligations, and top-up loans. Different EMI formulas or calculation methods can be presented for comparison, with assumptions clearly displayed.



I. KYC Document Workflow

The user can upload supported KYC documents. The system stores document metadata and, where implemented, extracts relevant details for verification. KYC processing is treated as a controlled workflow and not as an unrestricted input to the ML model.



J. Monitoring and Reporting

Administrators or analysts can inspect model performance, drift events, retraining events, model versions, and relevant evaluation metrics. This supports continuous model governance.



VIII. MACHINE LEARNING METHODOLOGY

A. Data Preprocessing

Preprocessing includes missing-value handling, duplicate checking, categorical encoding, numerical scaling where required, outlier review, target construction, and feature engineering. Data leakage must be avoided, especially when time dependent information is used.

B. Baseline models

– Logistic Regression: baseline probabilistic classifier and interpretable benchmark.

All rights reserved by *www.ijsrd.com*

- Decision Tree: rule-oriented model useful for comparison and intuitive explanations.
- Random Forest: ensemble model capable of representing nonlinear relationships and interactions.
- XGBoost: gradient-boosting model for highperformance tabular classification.

C. Adaptive Learning

The adaptive path uses sequentially arriving observations. River or another suitable onlinelearning implementation can be used for incremental learning and online evaluation. The research will compare a static model that remains unchanged with an adaptive model that updates as new observations become available.

D. Concept Drift

Concept drift refers to changes in the relationship between input variables and the target or changes in the underlying data distribution that can degrade model performance. Drift detection is used as a trigger for investigation, adaptation, or retraining rather than assuming that every new record should immediately change the production model.

E. Explainable AI

SHAP provides feature contribution information that can be aggregated for global interpretation and examined for an individual prediction. LIME can provide a local surrogate explanation for a particular applicant. The research will compare the usefulness and consistency of these explanations, particularly under changing data.

F. Evaluation Metrics

- Accuracy
- Precision
- Recall
- F1-score
- ROC-AUC
- Confusion matrix
- Drift detection effectiveness
- Retraining/update frequency
- Computational or update efficiency
- Stability and usefulness of explanations where measurable

IX. Loan Decision-Support Features

- Loan eligibility/risk prediction. Estimates a risk or eligibility category using the selected model and a defined target variable.
- Decision explanation. Shows influential parameters and their contributions to an individual prediction.
- Improvement guidance. Translates influential risk factors into understandable areas the applicant may review.
- EMI calculation. Estimates periodic repayments using specified calculation methods and clearly stated assumptions.
- Fixed-rate scenario. Calculates repayments under a fixed interest-rate assumption.
- Floating-rate scenario. Compares repayment scenarios under changing or scenario-based interest-rate assumptions.
- Moratorium. Models a repayment holiday or deferral and displays its effect on the repayment schedule, based on configured assumptions.
- Existing obligations. Includes current liabilities or monthly obligations in affordability and risk analysis.
- Top-up loan. Assesses an additional borrowing request in the context of an existing loan and other obligations.
- KYC upload. Allows supported KYC documents to be submitted and associated with an application.
- Model monitoring. Shows performance, drift events, model versions, and update information to authorised users.

X. Proposed Algorithm / Workflow

1) Collect and validate historical banking/creditrisk data.
2) Preprocess the data and construct relevant features.
3) Split data using a suitable evaluation strategy, preferably respecting chronology when temporal information is available.
4) Train baseline models and record evaluation metrics.
5) Select one or more candidate models for adaptive evaluation.
6) Feed new observations sequentially to simulate streaming data.
7) Generate predictions and update online evaluation metrics.
8) Monitor for distributional or performance drift.
9) When the defined drift/update condition is satisfied, incrementally update or retrain the model.
10) Store the model version, update event, evaluation metrics, and relevant drift information.
11) Generate SHAP/LIME explanations for individual decisions.
12) Map major risk factors to non-guaranteed improvement guidance.
13) Combine the prediction with loan-planning facilities such as EMI, rate scenarios, moratorium, obligations, and top-up analysis.
14) Evaluate the final system against the static baseline.

XI. Research Experiment Design

Because the project is a proposed academic research system, the paper does not invent experimental results before implementation. The intended experiment compares at least two conditions: (A) a static model trained on an initial historical dataset and held fixed, and (B) an adaptive model that processes later observations sequentially and updates when appropriate. Controlled distribution changes may be introduced during simulation to test the response to concept drift.

- Baseline model comparison. Compare Logistic Regression, Decision Tree, Random Forest, and XGBoost using Accuracy, Precision, Recall, F1score, and ROC-AUC.

All rights reserved by www.ijsrd.com

- Static versus adaptive model. Compare a fixed model with an online/adaptive model using metric changes over time and recovery after drift.
- Drift response. Introduce a controlled distribution change and measure detection delay, performance degradation, and recovery.
- Explanation comparison. Compare SHAP and LIME for feature consistency, interpretability, and usefulness for individual predictions.
- Update/retraining efficiency. Compare update triggers using update count, processing time, and performance recovery.
- Class-imbalance handling. Compare suitable imbalance-handling strategies using Recall, F1score, ROC-AUC, and explanation stability where measurable.

XII. Expected Outcomes

1) A functional web-based Hybrid Banking Decision Support System.
2) A machine-learning module capable of predicting credit risk or a defined loandecision target.
3) A comparative evaluation of Logistic Regression, Decision Tree, Random Forest, and XGBoost or selected alternatives.
4) A mechanism for processing newly arriving banking data sequentially.
5) Concept-drift detection and adaptive model updating/retraining.
6) Transparent individual predictions using SHAP and LIME.
7) Actionable, understandable guidance based on influential risk factors.
8) Integrated loan-planning facilities covering EMI, fixed/floating rates, moratorium, obligations, and top-up loans.
9) A KYC document-upload workflow with controlled document handling.
10) A comparative research result showing whether adaptive learning improves robustness under simulated changing conditions rather than assuming an accuracy improvement.

XIII. Advantages

- Adaptive: the system can respond to newly arriving data rather than depending only on one fixed training snapshot.
- Explainable: predictions are accompanied by feature-level explanations.
- Actionable: applicants can see areas that contribute to their outcome and possible areas for improvement.
- Integrated: prediction, loan planning, obligations, top-up assessment, and KYC workflow are brought into one platform.
- Research-oriented: static and adaptive approaches can be compared experimentally.
- Modular: the architecture separates UI, application logic, data, ML, adaptation, and explainability.
- Extensible: additional datasets, models, data streams, document types, or financial products can be added later.

XIV. Limitations

- The quality of predictions depends on the representativeness, licensing, and quality of the selected public datasets.
- Streaming behaviour may initially be simulated by sequential replay of historical data rather than obtained from a live bank feed.
- Concept-drift detection does not guarantee that every detected change is economically meaningful.
- Explainability methods provide model-based explanations and should not be interpreted as complete causal explanations.
- KYC document extraction/verification depends on document quality, supported formats, OCR capability if used, and validation rules.
- Loan approval, interest rates, moratorium terms, and eligibility are ultimately subject to institutional policies and applicable regulations; the prototype is a decisionsupport system, not a banking approval engine.
- Personal financial data and KYC documents require appropriate security, access control, retention, and privacy safeguards.

XV. Applications

- Retail loan pre-assessment and customer selfservice.
- Credit-risk analyst decision support.
- Loan portfolio monitoring and early-warning analysis.
- Applicant-facing financial improvement guidance.
- Scenario-based EMI and affordability planning.
- Top-up loan and obligation analysis.
- Academic research on adaptive and explainable financial machine learning.

XVI. Future Scope

- Integration with authenticated real-time banking or open-banking data feeds where legally and technically permitted.
- More advanced online-learning ensembles and drift-aware model selection.
- Fairness analysis across relevant applicant groups and monitoring for unintended bias.
- Counterfactual recommendations using DiCE or related methods to estimate feasible changes to selected features.
- OCR-assisted KYC extraction with human verification and stronger document-security controls.
- Integration with credit-bureau or institutionapproved data sources.
- Mobile application support and multilingual explanations.
- Model registry, automated monitoring, audit logs, and deployment pipelines for productiongrade governance.

XVII. Conclusion

The proposed Hybrid Banking Decision Support System extends conventional credit-risk prediction by combining adaptive learning, streaming-data processing, concept-drift detection, Explainable AI, and practical loan decision support. The central research contribution is the integration

All rights reserved by www.ijsrd.com

of two needs that are often treated separately: the ability of a model to adapt as new data arrives and the ability of the system to explain an individual decision in a form useful to the applicant.

Instead of providing only an approval or risk prediction, the system is designed to identify influential parameters and translate them into understandable improvement-oriented guidance. The additional EMI, fixed/floating-rate, moratorium, obligations, top-up, and KYC modules broaden the platform from a pure predictive model into an integrated banking decision-support prototype.

The proposed evaluation deliberately avoids assuming that streaming data will automatically increase accuracy. The research will test whether adaptive learning can maintain or improve robustness when the data distribution changes, while measuring computational cost, drift response, and explanation quality. This combination of adaptation and transparency provides a practical research direction for more responsive and understandable machine-learning based banking decision support.

REFERENCES

[1] V. Chang, Q. A. Xu, S. H. Akinloye, V. Benson, and K. Hall, "Prediction of bank credit worthiness through credit risk analysis: An explainable machine learning study," Annals of Operations Research, vol. 354, pp. 247–271, 2025, doi: 10.1007/s10479-024-06134-x.

[2] A. A. Montevechi, R. de Carvalho Miranda, A. L. Medeiros, and J. A. B. Montevechi, "Advancing credit risk modelling with machine learning: A comprehensive review of the state-of-the-art," Engineering Applications of Artificial Intelligence, vol. 137, Art. no. 109082, 2024, doi: 10.1016/j.engappai.2024.109082.

[3] D. Mulimani, P. Patil, S. Totad, and R. Benni, "Online detection and adaptation of concept drift in streaming data classification," Procedia Computer Science, vol. 235, pp. 2803–2811, 2024, doi: 10.1016/j.procs.2024.04.265.

[4] R. Su, H. Guo, and W. Wang, "Elastic online deep learning for dynamic streaming data," Information Sciences, vol. 676, Art. no. 120799, 2024, doi: 10.1016/j.ins.2024.120799.

[5] Y. Chen, R. Calabrese, and B. MartinBarragán, "Interpretable machine learning for imbalanced credit scoring datasets," European Journal of Operational Research, vol. 312, no. 1, pp. 357–372, 2024, doi: 10.1016/j.ejor.2023.06.036.

[6] H. Li and W. Wu, "Loan default predictability with explainable machine learning," Finance Research Letters, vol. 60, Art. no. 104867, 2024, doi: 10.1016/j.frl.2023.104867.

[7] M. N. Sowmiya, S. Jaya Sri, S. Deepshika, and G. Hanushya Devi, "Credit risk analysis using explainable artificial intelligence," Journal of Soft Computing Paradigm, vol. 6, no. 3, pp. 272–283, 2024, doi: 10.36548/jscp.2024.3.004.

[8] 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, Art. no. 100605, 2025, doi: 10.1016/j.dajour.2025.100605.

[9] C. N. Nwafor, O. Nwafor, and S. Brahma, "Enhancing transparency and fairness in automated credit decisions: An explainable novel hybrid machine learning approach," Scientific Reports, vol. 14, Art. no. 25174, 2024, doi: 10.1038/s41598-024-75026-8.

[10] R. Manjunath et al., "A study on application of explainable AI for credit risk management of an individual," in Proc. 8th International Conference on Computational System and Information Technology for Sustainable Solutions (CSITSS), 2024, doi: 10.1109/CSITSS64042.2024.10816861.

[11] U. Bhuvaneswari and S. Muthukrishnan, "Enhancing transparency in credit decisions: Explainability for AI models," in Proc. International Conference on Advances in Computing, Communication and Materials (ICACCM), 2024, pp. 1–6, doi: 10.1109/ICACCM61117.2024.11059189.

[12] P. E. de Lange, B. Melsom, C. B. Vennerød, and S. Westgaard, "Explainable AI for credit assessment in banks," Journal of Risk and Financial Management, vol. 15, no. 12, Art. no. 556, 2022, doi: 10.3390/jrfm15120556.

[13] J. Osterrieder, B. H. Misheva, A. Hirsa, O. Kulkarni, and S. F. Lin, "Explainable AI in credit risk management," arXiv:2103.00949, 2021, doi: 10.48550/arXiv.2103.00949.

All rights reserved by www.ijsrd.com