Automated Personal Loan Underwriting Using Integrated Rule-Based and Behavioral Analytics

Name - Savani Kavindya Seneviratne
Supervisor Name - Alqa Husni
UOW ID - w1958946
IIT NO - 20221746

Problem statement

The personal loan underwriting process in Sri Lanka remains largely manual, leading to slow decision-making and inconsistent risk assessment across banks. Major financial institutions such as **Sampath Bank**, **Commercial Bank**, **DFCC Bank**, and **HNB** approximately takes **3-7 working days** to process and grant a personal loan due to labor-intensive verification procedures, manual creditworthiness checks, and fragmented data systems [1]-[4]. Although these banks already maintain substantial internal customer data like salary credits, account turnover, repayment history, and transaction patterns this information is not systematically leveraged to automate underwriting for existing customers. Consequently, applicants experience unnecessary delays despite having a complete financial history within the same bank.

The purpose of this project is to develop an automated underwriting solution that validates customer information through rule-based checks, analyzes behavioral and transactional data, generates risk personas, and delivers instant, consistent, and fair loan decisions for existing customers.

Project Aim

This project aims to develop an automated personal loan underwriting system that evaluates existing customers using rule-based validation, transactional analytics, and risk persona generation to support instant and consistent loan decisions.

Objectives

- To automatically check the customer entered data against the bank's internal records to reduce mistakes.
- To analyze salary patterns, transactions, and overall financial behavior to determine creditworthiness using existing account data.
- To automatically determine whether the requested amount can be granted, partially granted, or declined, reducing processing time from several days to minutes.
- To reduce staff workload, standardize assessments, and enable faster customer service for personal loan applications.

Tools and Software Used

What is it used for	Tools Used
Data Preprocessing & Mining	• Pandas – Data cleaning, wrangling, and tabular transformations • Scikit-learn – Preprocessing, feature engineering, and model preparation • Mlxtend – Generating association rules (Apriori / FP-Growth) • NumPy – Numerical operations and array handling
Databases	• **MS SQL Server** – Storing data • **SQL Query Language** – For validating active accounts, retrieving turnovers, checking salary deposits, etc.
Development environment for coding	• Jupyter Notebook
Data Visualization	• **Matplotlib** – Charts for data analysis • **Seaborn** – Statistical visualizations
Customer Interface	• **SSRS (SQL Server Reporting Services)** – Customer interface and report rendering

Skills

Data Preprocessing & Cleaning, Feature Engineering & Statistical Analysis, Machine Learning (Clustering & Rule Mining), Database Management & SQL, Report Development & Interface Design (SSRS)

Dataset Description and Sampling Framework

Branch	CIF	Legal id	Legal id type	Customer Name	Age	Occupation	Proposition	Sector
LK0010044	1072801	77187****V	NIC	CUSTOMER NAME 1	48	Entrepreneur / Business Owners	100	8000
LK0010034	1031617	88625****V	NIC	CUSTOMER NAME 2	37		100	8000
LK0010049	1073805	77724****V	NIC	CUSTOMER NAME 3	48	Entrepreneur / Business Owners	100	8000
LK0010021	1151118	49621****V	NIC	CUSTOMER NAME 4	76		100	8000
LK0010034	1172465	60808****V	NIC	CUSTOMER NAME 5	65	Private Sector Executives	160	8000
LK0010111	1370615	86596****V	NIC	CUSTOMER NAME 6	39	Entrepreneur / Business Owners	100	8000
LK0010036	1191411	92361****V	NIC	CUSTOMER NAME 7	32	Private Sector Executives	100	8000
LK0010011	277432	70262****V	NIC	CUSTOMER NAME 8	55	Shop Salespersons	100	8000
LK0010025	117892	90634****V	NIC	CUSTOMER NAME 9	35	Entrepreneur / Business Owners	100	8000
LK0010014	1057674	66360****V	NIC	CUSTOMER NAME 10	58	Entrepreneur / Business Owners	100	8000
LK0010024	1048636	52766****V	NIC	CUSTOMER NAME 11	73	Entrepreneur / Business Owners	100	8000
LK0010021	1188446	85279****V	NIC	CUSTOMER NAME 12	40	Private Sector Non-Executives	100	8000
LK0010075	1168290	88263****V	NIC	CUSTOMER NAME 13	37	Entrepreneur / Business Owners	100	8000
LK0010137	212942	92833****V	NIC	CUSTOMER NAME 14	32		120	8000
LK0010099	221101	19821660****	EIC	CUSTOMER NAME 15	43	Private Sector Non-Executives	100	8000
LK0010111	113049	99506****V	NIC	CUSTOMER NAME 16	26	Entrepreneur / Business Owners	100	8000

A representative sample was extracted from a population of approximately one million customer records from the DFCC Bank database, excluding individuals under 18 in compliance with Central Bank regulations. The sample was created by selecting 10% from each demographic category, such as age, gender, district, proposition, sector, and occupation, to ensure balanced representation.
From this, an additional 10% was selected to form the final customer base comprising approximately 750 customers, with the total sample size amounting to around 15,000 records. Corresponding transaction data, account data, and credit risk data were extracted for analysis. All data were masked, anonymized, and used solely for analytical and modeling purposes under strict confidentiality and governance standards.

Proposed Research Methodology

Pre-Processing
Removing duplicates and identifying outliers in the retrieved data set. After standardizing the data, key financial and behavioral features such as salary patterns, transaction stability, and engineered metrics like the EMI Ratio are extracted.

Applying Association Rule Mining
ARM algorithms (Apriori, FP-Growth, Eclat) are applied to discover hidden relationships between financial behaviors.

Clustering for Persona Generation
Behavioral and financial features (including ARM-derived insights) are used to cluster customers. Methods such as K-Means, Hierarchical Clustering, or DBSCAN are used to group customers into personas. Each cluster represents a unique risk persona for decision-making.

Risk Scoring
Based on the generated persona and financial indicators, the system assigns a risk score to the customer. Risk scoring considers salary consistency, transaction stability, EMI ratio, credit turnover, and persona attributes.

Creating the Interface
The customer-facing interface is built using SSRS, enabling structured data entry and direct integration with backend validation modules. Through this interface, customers provide NIC, salary details, and loan requirements, which are subsequently processed for verification and analysis.

Testing & Validation
The system undergoes validation to ensure the accuracy and reliability of the generated personas and risk-scoring mechanism. Clustering results are evaluated using measures such as silhouette score and consistency checks, while risk-scoring models are validated using historical data and performance metrics. Interface workflows are also tested for accuracy, usability, and system-level integration before deployment.

Deployment
After the model is fully tested and validated, it is deployed into the SSRS-based loan processing system. The clustering and risk-scoring algorithms are integrated with backend APIs that process customer inputs such as NIC number, salary information, and loan details in real time. The deployed system generates the customer's financial persona and corresponding risk score instantly

Related Research

Focus Area	Methods Used / Tested	Key Findings
Predicting consumer credit risk using behavioral and transactional data	Machine Learning models such as Tree-based methods and Support Vector Machines (SVMs) to predict consumer credit risk [5].	ML models outperform traditional statistical models and can reduce bank losses by 6%–25%.
Loan default prediction using real-world data	XGBoost, LightGBM, Gradient Boosting, Random Forest, SMOTE, ROC/AUC, SHAP analysis to predict loan defaults [6].	Gradient Boosting performs best; XGBoost shows the highest AUC. key predictors identified, such as improving risk assessment, but challenges remain regarding interpretability and regulatory acceptance
Automation of loan-processing systems	AI, RPA, OCR, and workflow automation solutions were analyzed to assess their effectiveness in automating key stages of the loan processing workflow [7].	Faster decisions, reduced operational costs, improved accuracy. Implementation takes 10 to 15 or more months. Potential ROI around 225%.

Project Implementation Timeline

Mile stones	Sep'25	Oct'25	Nov'25	Dec'25	Jan'26	Feb'26	Mar'26	Apr'26
Finalize project scope (Litriture Review)								
Define sample size using statistical methods								
Data Cleaning & Preprocessing								
Clustering & Persona Generation								
Association Rule Mining								
Combine Persona → Risk Score → Approval Flow								
Interface Development								
Testing & Validation								
Documentation & Final Submission								

RISK AND MITIGATION

RISK	MITIGATION
Data security and compliance issues	Apply data-masking, encryption, and maintain proper audit and access controls.
Insufficient or low-quality data	Improve data quality through cleaning, validation, and the use of anonymised or synthetic data if needed.
Incorrect or biased rule/score outputs	Validate rules with experts, test and recalibrate regularly.
System integration difficulties	Develop the system in separate parts and test each part early.

[1] Sampath Bank PLC, "Professional Personal Loan," 2025. [Online]. Available: https://www.sampath.lk/personal-banking/loan/personal-loans/Professional-Loan?category=personal_banking . [Accessed: 30-Oct-2025].
[2] Commercial Bank of Ceylon PLC, "Personal Loans," 2025. [Online]. Available: https://www.combank.lk/personal-banking/loans/personal-loans . [Accessed: 30-Oct-2025].
[3] DFCC Bank PLC, "Personal Loans," 2025. [Online]. Available: https://www.dfcc.lk/products-categories/personal-loans/ .[Accessed: 30-Oct-2025].
[4] Hatton National Bank PLC, "Personal Loans," 2025. [Online]. Available: https://hnb.lk/personal/loans/personal-loans . [Accessed: 30-Oct-2025].
[5] L. E. Breeden, "Machine learning methods for consumer credit-risk assessment," *Journal of Banking & Finance*, vol. 34, no. 11, pp. 2767–2787, 2010. [Online]. Available: https://www.sciencedirect.com/science/article/abs/pii/S0378426610002372?via%3Dihub . [Accessed: 31-OCT-2025].
[6] X. Zhang, T. Zhang, L. Hou, X. Liu, Z. Guo, Y. Tian and Y. Liu, "Data-driven loan default prediction: A machine learning approach for enhancing business process management," *Systems*, vol. 13, no. 7, art. no. 581, 2025. [Online]. Available: https://www.mdpi.com/2079-8954/13/7/581 . [Accessed: 01-Nov-2025].
[7] ScienceSoft Inc., "Loan automation: Features, benefits and implementation," 2024. [Online]. Available: https://www.scnsoft.com/lending/loan-automation. [Accessed: 01-Nov-2025].