Cash Flow Underwriting with Bank Transaction Data: Advancing MSME Financial Inclusion in Malaysia

Chun Chet Ng*
AI Lens
Kuala Lumpur, Malaysia
chunchet.ng@ailensgroup.com

Wei Zeng Low†*
AI Lens
Kuala Lumpur, Malaysia
alexlow@ailensgroup.com

Yin Yin Boon
AI Lens
Kuala Lumpur, Malaysia
yinyin.boon@ailensgroup.com

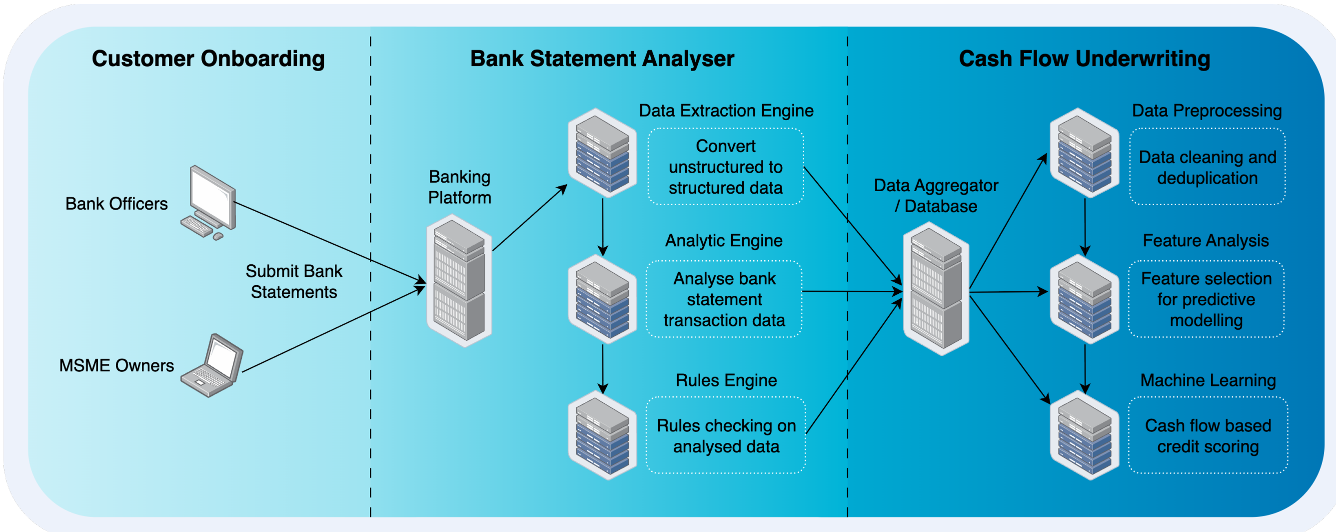


Figure 1: Proposed end-to-end cash flow underwriting workflow leveraging bank statement transaction data for credit scoring.

Abstract

Despite accounting for 96.1% of all businesses in Malaysia [15], access to financing remains one of the most persistent challenges faced by Micro, Small, and Medium Enterprises (MSMEs). Newly established or young businesses are often excluded from formal credit markets as traditional underwriting approaches rely heavily on credit bureau data. This study investigates the potential of bank statement data as an alternative data source for credit assessment to promote financial inclusion in emerging markets. Firstly, we propose a cash flow-based underwriting pipeline where we utilise bank statement data for end-to-end data extraction and machine learning credit scoring. Secondly, we introduce a novel dataset of 611 loan applicants from a Malaysian lending institution. Thirdly, we develop and evaluate credit scoring models based on application information and bank transaction-derived features. Empirical results show that the use of such data boosts the performance of all models on our dataset, which can improve credit scoring for new-to-lending MSMEs. Lastly, we intend to release the anonymised bank transaction dataset to facilitate further research on MSMEs financial inclusion within Malaysia's emerging economy.

Keywords

Bank Statement Data, Alternative Data, Credit Scoring

*Both authors contributed equally to this work.
†Corresponding author.

FinRem Workshop @ ACM ICAIF 2025, Singapore

1 Introduction

Financial inclusion remains a pressing challenge in emerging markets, where a significant portion of the population lacks sufficient credit history to access formal lending. According to the Securities Commission Malaysia [26], MSMEs are estimated to contribute around 60% of Malaysia's gross domestic product. Despite their significant contributions, MSMEs remain underserved by financial institutions in terms of access to financing, giving rise to an estimated MYR 90 billion funding gap [11].

One fundamental issue is that many MSMEs do not have lending history. Traditional credit assessment relies heavily on credit bureau data such as repayment history, outstanding obligations, and past delinquencies. While this model works well for established businesses, it has shortcomings for MSMEs with thin credit files:

i It is inherently backward-looking and focuses on past repayment behaviour rather than current or forward-looking capacity to repay.
ii It overlooks real-time financial signals or operational dynamics that more accurately reflect a firm's present financial health.
iii It omits alternative indicators of creditworthiness, such as cash flow consistency, receivables and payables patterns, digital transactions, and behavioural metrics from financial activity.

On the other hand, bank statements represent an up-to-date and verifiable source of financial behaviour which also capture income regularity, spending patterns, and cash flow stability. This study explores the use of Malaysian bank statement transactions as alternative data for credit scoring model development and evaluates its predictive value for MSMEs. The objectives are as follows:

- To propose a cash flow-based underwriting workflow capable of ingesting and analysing bank transaction data for credit decisioning to narrow the MSMEs financing gap.
- To introduce the first ever Malaysian bank statement dataset based on loan application of MSMEs.
- To evaluate the performance of machine learning–based credit scoring models trained on the proposed bank statement transaction dataset, and to examine the feasibility and predictive power of transaction-derived features in assessing MSME creditworthiness.

2 Related Works

With the recent technological advancements in machine learning algorithms, the research community begins to realize the limitations of existing traditional credit scoring methods [20]. Such studies are important to enhance financial inclusion of MSMEs in emerging markets where MSMEs are often the backbone of economic growth although they have limited access to financing [17]. As such, alternative data from non-traditional sources that are not included in standard credit bureau files are being used for credit scoring [21].

2.1 Limitations of Traditional Credit Scoring

Traditional credit scoring models used by financial institutions primarily rely on credit bureau data such as payment history, amount of debt, and other indicators [35]. Unfortunately, reliance on such data creates a high entry barrier for MSMEs in emerging economies where they usually lack audited financial statements or loan servicing history required by financial institutions [12, 17]. This leads MSMEs to be perceived as high-risk applicants and results in a perpetual cycle of financial exclusion that severely limits their growth potential [12]. Besides, traditional methods fail to capture recent cash flow status of MSMEs, which can be an accurate representation of financial health [17]. A recent survey indicates that 58% of lending institutions feel less confident to make decisions based on traditional credit data only [38]. The increasing dissatisfaction motivates the urgent need to use alternative data for a more adaptive and inclusive credit assessment [20].

2.2 Credit Scoring with Alternative Data

In order to overcome the limitations of traditional methods, credit underwriting using alternative data has emerged as a new approach [20, 28, 38]. For instance, mobile network data are used for credit scoring in Africa [19, 28] and bank account transactions are used for cash flow underwriting in consumer lending [38]. Recent studies suggest that incorporating transactional data can improve predictive accuracy and expand credit to underserved populations [16]. For instance, research has shown that retail transaction data can help construct alternative credit scores [24]. A study examining two Indian financial technology (FinTech) firms found that transactional data from platform activity worked as well as credit bureau data in predicting creditworthiness and improved the predictive accuracy when combined with bureau data [8]. However, the adoption of alternative data in MSMEs lending in Malaysia remains at an early stage, with limited research focus from the industry.

2.3 Machine Learning Models in Credit Scoring

Machine learning (ML) models have proven to be highly effective in processing financial data to produce accurate credit scores [6, 22, 37]. These include statistical models like Logistic Regression, Naive Bayes (NB), and ensemble methods like Random Forest, AdaBoost, and Gradient Boosting [6, 37]. For example, a Random Forest model achieved the best performance on Taiwanese credit card data [1]. A NB classifier performed well on a highly imbalanced dataset from a New Zealand lender, especially when undeclared features derived from bank statements were incorporated to supplement application form data [7]. Given the positive impact of ML models in credit scoring [2], we aim to improve financial inclusion of Malaysian MSMEs through our ML-based cash flow underwriting workflow.

3 Bank Statement Cash Flow Underwriting

The proposed end-to-end cash flow underwriting workflow enables financial institutions to integrate bank statement transaction data analysis into their credit decision-making process. Its primary objective is to enhance traditional credit underwriting by leveraging bank statement-derived features. As illustrated in Figure 1, the workflow consists of three main layers:
(i) Web Layer (Customer Onboarding): This layer acts as the entry point for all loan application-related data including bank statements. A web-based platform is built to allow MSME owners to submit loan applications and for bank officers to review them.
(ii) Application Layer (Bank Statement Analyser): This layer processes uploaded bank statements from the web interface. Multiple specialized engines automate the extraction and conversion of unstructured transactions into structured data, analyses cash flow-based indicators and bank statement related features, and applies rule-based checks to detect potential fraud and data abnormalities.
(iii) Data & Scoring Layer (Cash Flow Underwriting): This layer stores the analysed data and engineered features in a secure database to ensure traceability, regulatory compliance, and auditability. The stored data are then preprocessed, and feature selection methods are applied to retain the most informative variables. Using the selected features, a predictive model is trained to estimate the probability of default and classify credit risk.

The proposed workflow offers several advantages for financial institutions. It shortens turnaround time by automating manual tasks, thereby improving operational efficiency. It also expands credit access for thin-file MSMEs by leveraging transactional data instead of relying primarily on historical credit bureau records, allowing MSMEs with limited or no bureau data to obtain financing.

4 Bank Statement Transaction Dataset

Table 1: Statistics of MSME loan application datasets.

Split	Non-Event	Event	Total
Train (60%)	310	56	367
Validation (40%)	208	37	245
Overall	518	93	611

Transaction data derived from bank statements represents a valuable alternative resource for cash flow underwriting. To the best of our knowledge, no published studies have examined the use of bank statement data for credit assessment of MSMEs in Malaysia. In collaboration with a lending institution, we constructed the first Malaysian bank statements dataset to address this gap.

This study adopts the Cross-Industry Standard Process for Data Mining (CRISP–DM) framework [9] as the methodological basis. CRISP–DM is a widely recognized process model for systematic FinTech studies [10, 14, 29] and comprises six main phases i.e. business understanding, data understanding, data preparation, modeling, evaluation, and deployment. The first phase involves evaluating bank-statement transaction data as an alternative source for MSME credit risk assessment in Malaysia. The second phase focuses on constructing the proposed dataset of 611 MSME loan applicants and is split into training and validation sets with a 60:40 ratio as shown in Table 1. Among these, 518 applicants have a good credit record, while 93 have a history of default. Each applicant's record contains two main components: application form information submitted during the loan application process (e.g., demographic and business characteristics) and bank statement transaction data capturing detailed inflows and outflows over a six-month period. The third phase is about data preparation and includes several key tasks:

i Data cleaning and de-duplication to ensure consistency.
ii Fixing missing values, standardizing transaction categories and validating data integrity across all records.
iii Deriving features and variables from transaction data, such as determining cash flow stability, deposit regularity, and balance volatility. All personally identifiable information was anonymised prior to analysis. To preserve confidentiality, the features are grouped into application form information (7 features) and bank statement features (10 features). Although the specific feature calculations cannot be disclosed due to a non-disclosure agreement, all bank statement features are derived solely from bank statement data.

5 Transaction Data-based Credit Scoring

The fourth phase of CRISP–DM in this study employs Logistic Regression (LR) [13] on application form and bank transaction data derived from our dataset as the baseline credit scoring model. LR is widely adopted in credit risk modeling due to its interpretability, statistical robustness, and capacity to estimate the probability of default [1, 6, 7, 37]. To ensure that only informative predictors are retained, feature selection and transformation are guided by the Weight of Evidence (WOE) and Information Value (IV) framework, which are widely used in credit risk modeling for quantifying the predictive power of individual variables [28]. In practice, each feature is divided into discrete intervals based on thresholds derived from the data. These bins allow the calculation of WOE and IV at the group level, which makes them easier to interpret within a logistic regression model. The following notation is used to quantify the distribution of defaults and non-defaults across feature bins:

Let $y_i \in \{0, 1\}$ indicate default ($y_i = 1$) or non-default ($y_i = 0$) for applicant i. The total number of default and non-default applicants is defined as:

$$N_b = \sum_{i=1}^{n} \mathbb{I}(y_i = 1), \qquad N_g = \sum_{i=1}^{n} \mathbb{I}(y_i = 0), \tag{1}$$

Furthermore, let x_{ij} denote the value of feature j for applicant i. For a given feature j, suppose it is divided into K_j disjoint bins $\{B_{j1}, \ldots, B_{jK_j}\}$, then the corresponding counts within each bin is defined as:

$$\begin{aligned} n_{bjk} &= \sum_{i=1}^{n} \mathbb{I}(x_{ij} \in B_{jk})\, \mathbb{I}(y_i = 1), \\ n_{gjk} &= \sum_{i=1}^{n} \mathbb{I}(x_{ij} \in B_{jk})\, \mathbb{I}(y_i = 0). \end{aligned} \tag{2}$$

The distribution of defaults and non-defaults in each bin are:

$$\mathrm{Dist}^{(b)}_{jk} = \frac{n_{bjk}}{N_b}, \qquad \mathrm{Dist}^{(g)}_{jk} = \frac{n_{gjk}}{N_g}. \tag{3}$$

5.1 Weight of Evidence (WOE)

Firstly, we calculate the WOE for bin k of feature j using:

$$\mathrm{WOE}_{jk} = \log\left(\frac{\mathrm{Dist}^{(g)}_{jk}}{\mathrm{Dist}^{(b)}_{jk}}\right) = \log\left(\frac{n_{gjk}/N_g}{n_{bjk}/N_b}\right). \tag{4}$$

A positive WOE_{jk} indicates that bin B_{jk} is more common among non-default cases, suggesting lower risk, whereas a negative value indicates a higher likelihood of default. Besides, a small smoothing constant $\epsilon > 0$ is added to both numerators and denominators to handle bins with zero counts. Once WOE is computed for each bin, the predictive strength of feature j can be summarized by its IV.

5.2 Information Value (IV)

Secondly, we measure the IV of feature j by aggregating the WOE across all respective bins:

$$\mathrm{IV}_j = \sum_{k=1}^{K_j} \left(\mathrm{Dist}^{(g)}_{jk} - \mathrm{Dist}^{(b)}_{jk}\right) \mathrm{WOE}_{jk}. \tag{5}$$

A larger IV_j indicates stronger discriminatory power of feature j between default and non-default classes. In this study, IV serves as both a feature-ranking criterion and an interpretability measure, helping to identify the most influential transaction-level indicators for creditworthiness. Following industry practice [36], we adopt the commonly used thresholds: IV < 0.02 (not predictive), $0.02 \leq$ IV < 0.1 (weak), $0.1 \leq$ IV < 0.3 (medium), $0.3 \leq$ IV < 0.5 (strong), and IV ≥ 0.5 (suspiciously high; suggesting potential data leakage).

5.3 Binning of Bank Statement Features

Specifically, we use quantile or supervised monotonic binning to obtain $\{B_{jk}\}_{k=1}^{K_j}$ for continuous transaction-derived variables (e.g., cash-flow stability, balance volatility). Then, we group rare datapoints to ensure each bin has sufficient defaults and non-defaults for sparse categorical variables (e.g., business sector code). Computations of WOE and IV are performed on training folds only to avoid data leakage.

5.4 Logistic Regression Model

Given that WOE provides a log-odds transformation of each feature, this aligns naturally with the LR model. The WOE transformed feature value can be represented as $\text{WOE}_j(x_{ij})$ and the LR model can be expressed in a simplified form as:

$$\log \frac{P(y_i = 1 \mid \mathbf{x}_i)}{P(y_i = 0 \mid \mathbf{x}_i)} = \beta_0 + \sum_{j=1}^{d} \beta_j \, \text{WOE}_j(x_{ij}). \tag{6}$$

where β_0 is the intercept and β_j represents the coefficient for feature j. This allows direct interpretation of coefficients in terms of the credit risk associated with each feature. Positive β_j values indicate that higher WOE corresponds to lower default risk, and vice versa. Hence, WOE encoding stabilizes estimation and supports consistent, monotonic relationships between predictors and credit outcomes.

Formally, let $\mathbf{x}_i \in \mathbb{R}^d$ denote the feature vector for applicant i, where d is the number of features engineered from both application form data and bank statement transactions, and let $y_i \in \{0, 1\}$ be the binary output indicating default ($y_i = 1$) or non-default ($y_i = 0$). The LR model specifies the conditional probability of default as

$$P(y_i = 1 \mid \mathbf{x}_i; \boldsymbol{\beta}) = \sigma(\beta_0 + \mathbf{x}_i^\top \boldsymbol{\beta}), \tag{7}$$

where $\sigma(z) = \frac{1}{1+e^{-z}}$ is the sigmoid function, $\beta_0 \in \mathbb{R}$ is the intercept term, and $\boldsymbol{\beta} \in \mathbb{R}^d$ is the coefficient vector associated with the predictors. The parameters are estimated by maximizing the penalized log-likelihood function:

$$\mathcal{L}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left[y_i \log p_i + (1 - y_i) \log(1 - p_i) \right] - \lambda \, \|\boldsymbol{\beta}\|_2^2, \tag{8}$$

where $p_i = P(y_i = 1 \mid \mathbf{x}_i; \boldsymbol{\beta})$ and $\lambda \geq 0$ controls the strength of the ℓ_2 regularization [27]. This mitigates overfitting by shrinking coefficient magnitudes, which is particularly important when modeling high-dimensional, transaction-derived features.

6 Experiments and Deployment

Aligned with the fifth phase of CRISP–DM, we conduct a series of experiments to evaluate the effectiveness of the proposed cash flow underwriting workflow and the role of bank statement data in enhancing MSME credit scoring performance. The experiments include comparative analyses with established machine learning methods and ablation studies on transaction-derived features.

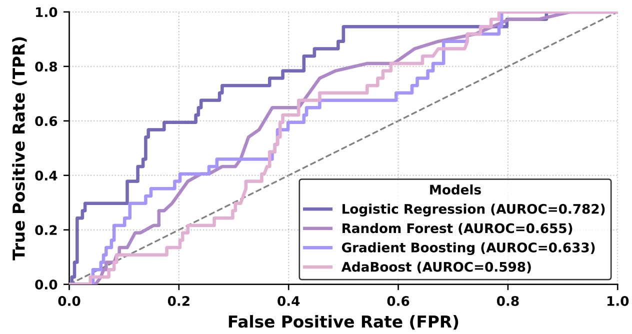


Figure 2: ROC curves of all models evaluated on the validation split of the proposed dataset.

6.1 Implementation Details

We benchmark the baseline LR model against several widely used machine learning methods, including Random Forest (RF) [4], Gradient Boosting (GB) [18], and AdaBoost (AB) [30]. All models are implemented in scikit-learn [31–34]. For the results in Section 6.2, models were trained with default hyperparameters. In Section 6.3, hyperparameters were tuned using a randomized grid search with 50 trials to ensure robustness. Model performance is evaluated using the Area Under the Receiver Operating Characteristic Curve (AUROC) [3], which measures the ability to discriminate between default and non-default cases across varying thresholds. An AUROC of 0.5 indicates no discriminative power (equivalent to random guessing), whereas a value of 1.0 indicates perfect discrimination. To ensure interpretability and systematic analysis, both application and bank statement features are grouped into three categories:

i Account Behaviour: logarithmic growth rate of average balances, average balance over the past six months, ratio between lowest balance in recent three months, percentage difference between the lowest balance, and highest average balance for the past three months.
ii Repayment Capacity: repayment capability calculated based on monthly installment.
iii Business Demographics: years in business, location, business sector code, number of directors, and director's minimum age.

This structured feature grouping facilitates clearer attribution of model performance to different behavioural and demographic dimensions of MSME credit profiles. It also supports experimental reproducibility, while respecting confidentiality constraints around proprietary feature derivations.

6.2 Quantitative Results

Quantitative results on the validation split of the proposed dataset are shown in Figure 2. The LR model achieves the highest AUROC of 0.782, representing a 30.77% improvement over AB, the lowest-performing model with a score of 0.598. RF follows with 0.655, and GB with 0.633. This ranking can be explained by the fact that LR is well suited to capturing the linear structure of the data, whereas more complex ensemble methods such as RF, GB, and AB are prone to overfitting when applied to relatively small

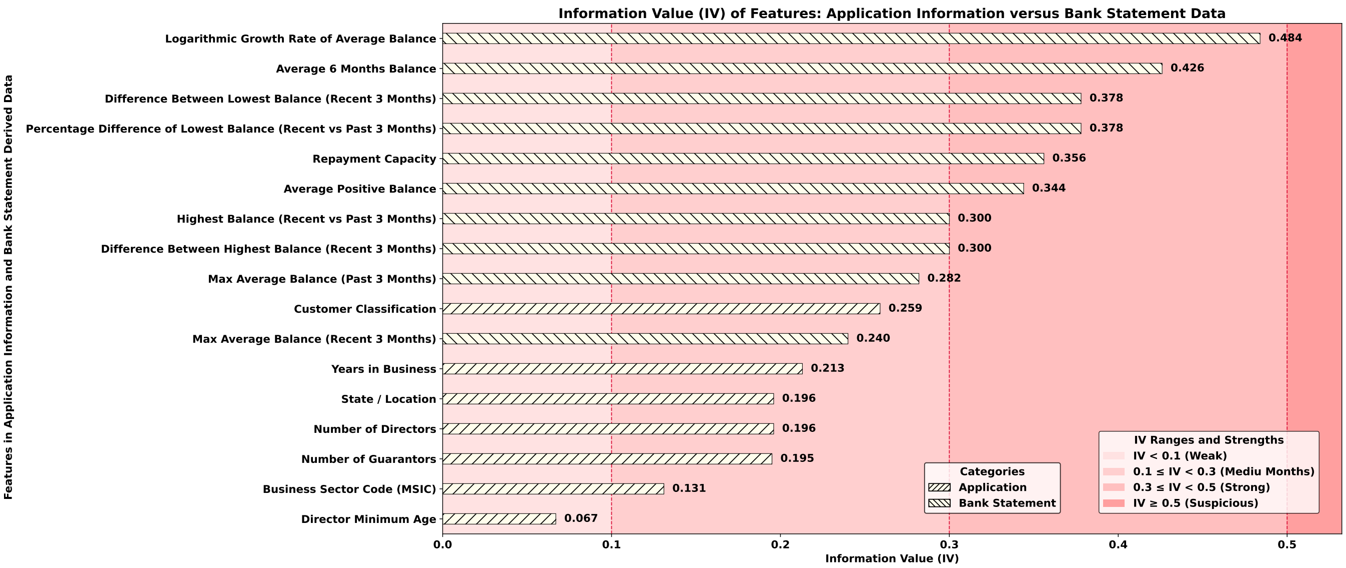


Figure 3: Information Value (IV) distribution of features derived from application and bank statement data. Features with higher IV indicate stronger predictive power for creditworthiness classification.

samples, particularly under class imbalance (310 non-event vs. 56 event cases in the training set). In such settings, models that provide well-calibrated probabilities and strong generalization tend to perform better, which is consistent with prior benchmarking studies showing that LR often matches or outperforms tree-based ensembles on small-to-medium credit datasets [5, 25].

6.3 Ablation Studies

Results in Figure 3 show that bank statement–derived features exhibit stronger predictive power than application form information. Nine of the ten transaction-based features rank above all application features, with the sole exception of customer classification, a pre-assigned label from the lending institution describing the business entity. Customer classification slightly surpasses the recent three-month maximum average balance with a relative difference of 7.92%, while also maintaining a substantial margin of 8.67% over the next highest-ranked application feature. These findings reinforce the superior discriminatory strength of bank statement features in distinguishing default from non-default outcomes.

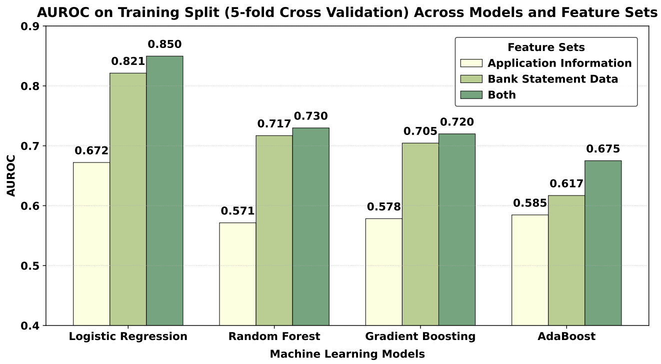
In addition, we conducted ablation experiments by progressively removing feature groups to study which features matter most for MSMEs credit scoring. We applied 5-fold cross-validation on the training split of the proposed dataset to reduce the risk that results are due to random chance and to increase confidence in model stability. From results illustrated in Figure 4a, we observe a substantial uplift when incorporating bank transaction data. The AUROC of LR improves from 0.672 (application information only) to 0.821 (bank transaction only), representing an uplift of 0.149 points (approximately 22% relative improvement). When both feature sets are combined, the AUROC of LR further increases to 0.8497, confirming that transactional data provides significant incremental predictive value. The validation results shown in Figure 4b are consistent with the cross-validation findings. The LR model achieves the highest AUROC overall, improving from 0.647 (application only) to 0.763 (bank statement only), and further to 0.806 when both feature sets are combined. Similar trends are observed for ensemble models such as RF, GB and AB, which record respective gains of 0.127, 0.145 and 0.181 points with the inclusion of bank transaction data. On both splits, all models that use bank transaction data outperform those based solely on application information. The combined features lead to the best result for all models. In general, these findings strongly support the hypothesis that bank statement transaction data offers significant predictive power in credit risk assessment for MSMEs in Malaysia. While application form information provides some discriminatory power, the inclusion of bank transaction-based features substantially enhances model performance, reinforcing the effectiveness of our proposed workflow.

7 Deployment and Operational Framework

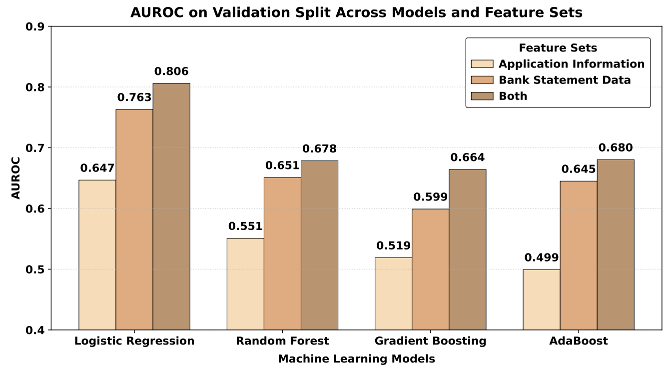
The sixth phase of CRISP–DM focuses on the deployment of the proposed cash flow underwriting workflow in a production environment. This workflow is designed to leverage verifiable transaction data to enable transparent and evidence-based credit decisions while establishing a continuous feedback cycle for data collection, model retraining, and performance monitoring within existing core banking systems. To support this, a robust machine learning operations framework is implemented to structure the end-to-end pipeline into five key stages. Lastly, an integrated credit scoring framework is introduced to provide comprehensive risk evaluation for both established and new-to-credit MSMEs.

7.1 Data Ingestion

The pipeline begins when bank statements are received. Extraction models parse and structure the raw transaction data using Optical Character Recognition (OCR) and layout analysis techniques. This



(a) AUROC performance on the training split with different features.



(b) AUROC performance on the validation split with different features.

Figure 4: Comparison of AUROC performance across models and feature sets on the training and validation splits.

process transforms unstructured documents into a standardized format suitable for downstream processing. The ingestion stage also performs integrity verification and data validation checks to ensure document authenticity.

7.2 Feature Engineering and Feature Store

Following ingestion, structured data are fed into the feature engineering module, which computes transaction-derived and behavioral features. These features are time-stamped, versioned, and stored in a repository. The feature store provides a single source of truth for both training and inference, ensuring the same logic used to generate historical features is consistently applied to new applicants in real time. This design guarantees reproducibility, prevents feature skew, and supports traceability across model iterations.

7.3 Continuous Integration (CI)

The CI pipeline automates model retraining and validation before integration into the production registry. It can be triggered by two key events: (1) on a predefined schedule (e.g., annually), when new labeled data become available from the core banking system, and (2) in response to alerts from the monitoring system indicating model or data drift. When triggered, the pipeline retrieves the latest feature set and ground truth labels, retrains the model, and evaluates it against the deployed model using predefined performance metrics.

7.4 Model Registry and Continuous Deployment (CD)

If the retrained model demonstrates superior performance, it is automatically versioned and stored in the model registry. The CD process then packages the validated model into a containerized microservice and deploys it as a secure REST API endpoint. To ensure reliability and minimize production risk, a canary deployment strategy is applied—initially routing a small fraction of live traffic to the new model for performance validation before full-scale rollout.

7.5 Continuous Monitoring and Feedback Loop

Once deployed, the model operates within a Champion–Challenger framework to enable continuous performance optimization [23]. The Champion model serves as the current production baseline, while a Challenger model is periodically retrained on the latest data and evaluated in parallel. Their performance is continuously monitored and compared across predefined metrics. When the Challenger consistently outperforms the Champion, the system automatically notifies bank officers to review the results and promote the new model to production. This approach ensures ongoing model improvement, robustness, and operational stability.

7.6 Integrated Credit Scoring Framework

In production, the cash flow underwriting system functions as an integrated decision engine embedded within the lending process. Upon submission of bank statements, AI models automatically extract and structure the raw data using OCR and natural language processing models. The processed data then undergoes fraud detection, cash flow analysis, and network analytics to generate transaction-derived features and credit scores for MSME applicants. The existing bureau-based scorecard, designed for borrowers with established credit histories, operates in parallel with the newly developed cash flow–based model. Each model independently produces a risk rating (Low, Medium, or High). A risk override mechanism ensures conservative decisioning; if either model indicates higher risk, the final classification adopts that rating. This framework provides an evidence-based and data-driven approach to MSME credit assessment using adaptive, explainable AI scoring. It also extends financial access to new-to-credit businesses through verifiable transaction data.

8 Conclusion

In conclusion, this study introduces bank statement transactions as alternative data for MSMEs loan underwriting in Malaysia. Our study demonstrates that transaction-derived features capture dynamic aspects of MSMEs' financial behaviour that are often overlooked in traditional credit models. These features also exhibit strong predictive power with high IV scores. Empirical results show that models built solely on application information perform modestly, whereas those based on bank transaction data achieve substantially better results. The combined features further improve the model's predictive power, confirming the complementary nature and practical value of alternative data.

Acknowledgments

We thank the Cradle Fund and the Microsoft for Startups program for their continuous support. This work was made possible through the contributions of the DocLab and DocSuite teams at AI Lens.

References

[1] Elaf Adel Abbas and Nisreen Abbas Hussein. 2024. Algorithm comparison for data mining classification: Assessing bank customer credit scoring default risk. *Jurnal Kejuruteraan* 36, 5 (2024), 1935–1944.
[2] Oluwabusayo Adijat Bello. 2024. The role of data analytics in enhancing financial inclusion in emerging economies. *International Journal of Developing and Emerging Economies* 11, 3 (2024), 90–112.
[3] Andrew P Bradley. 1997. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition* 30, 7 (1997), 1145–1159.
[4] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.
[5] Iain Brown and Christophe Mues. 2012. An experimental comparison of classification algorithms for imbalanced credit scoring data sets. *Expert systems with applications* 39, 3 (2012), 3446–3453.
[6] Michael Bücker, Gero Szepannek, Alicja Gosiewska, and Przemyslaw Biecek. 2022. Transparency, auditability, and explainability of machine learning models in credit scoring. *Journal of the Operational Research Society* 73, 1 (2022), 70–90.
[7] Rory P Bunker, Wenjun Zhang, and M Asif Naeem. 2016. Improving a credit scoring model by incorporating bank statement derived features. *arXiv preprint arXiv:1611.00252* (2016).
[8] Dean Caire and Maria Fernandez Vidal. 2024. *Leveraging Transactional Data for Micro and Small Enterprise (MSE) Lending.* Retrieved September 30, 2025 from https://www.cgap.org/sites/default/files/publications/Leveraging%20Transactional%20Data.pdf
[9] Pete Chapman, Julian Clinton, Randy Kerber, Colin Shearer Thomas Khabaza ands Thomas Reinartz, and Rüdiger Wirth. 2000. *CRISP-DM 1.0 Step-by-step data mining guide.* Retrieved September 30, 2025 from https://www.kde.cs.uni-kassel.de/wp-content/uploads/lehre/ws2012-13/kdd/files/CRISPWP-0800.pdf
[10] Aijun Cheng. 2023. Evaluating Fintech industry's risks: A preliminary analysis based on CRISP-DM framework. *Finance Research Letters* 55 (2023), 103966.
[11] International Finance Corporation. 2017. *MSME Finance Gap: Assessment of the Shortfalls and Opportunities in Financing Micro, Small, and Medium Enterprises in Emerging Markets.* World Bank, Washington, D.C., United States.
[12] MJ Courchane and AP Baines. 2020. *The Use of Cash-Flow Data in Underwriting Credit.* Retrieved September 30, 2025 from https://finreglab.org/wp-content/uploads/2023/12/FinRegLab_2020-03-03_Research-Report_The-Use-of-Cash-Flow-Data-in-Underwriting-Credit_Market-Context-and-Policy-Analysis.pdf
[13] David R Cox. 1958. The regression analysis of binary sequences. *Journal of the Royal Statistical Society Series B: Statistical Methodology* 20, 2 (1958), 215–232.
[14] B Carneiro da Rocha and R Timoteo de Sousa Junior. 2010. Identifying bank frauds using CRISP-DM and decision trees. *International Journal of Computer Science and Information Technology* 2, 5 (2010), 162–169.
[15] DOSM Department of Statistics Malaysia. 2025. *Profile of MSMEs in 2015-2024.* Retrieved September 30, 2025 from https://smecorp.gov.my/index.php/en/policies/2020-02-11-08-01-24/profile-and-importance-to-the-economy
[16] Viani B Djeundje, Jonathan Crook, Raffaella Calabrese, and Mona Hamid. 2021. Enhancing credit scoring with alternative data. *Expert Systems with Applications* 163 (2021), 113766.
[17] Okeoghene Elebe and Chikaome Chimara Imediegwu. 2021. A credit scoring system using transaction-level behavioral data for MSMEs. *Journal of Frontiers in Multidisciplinary Research* 2, 1 (2021), 312–322.
[18] Jerome H Friedman. 2001. Greedy function approximation: a gradient boosting machine. *Annals of statistics* (2001), 1189–1232.
[19] Anthony Gathu. 2020. *The Role of alternative data in accurately determining credit score for mobile lending on digital wallets in Kenya.* Ph. D. Dissertation. Strathmore University.
[20] Amol Gote and Vikas Mendhe. 2024. Building a Cash Flow Underwriting System: Insights from Implementation. *International Journal of Computer Trends and Technology* 72, 2 (2024), 70–74.
[21] World Bank Group. 2024. *The Use of Alternative Data in Credit Risk Assessment: Opportunities, Risks, and Challenges.* Retrieved September 30, 2025 from https://documents1.worldbank.org/curated/en/099031325132018527/pdf/P179614-3e01b947-cbae-41e4-85dd-2905b6187932.pdf
[22] Björn Rafn Gunnarsson, Seppe Vanden Broucke, Bart Baesens, María Óskarsdóttir, and Wilfried Lemahieu. 2021. Deep learning for credit scoring: Do or don't? *European Journal of Operational Research* 295, 1 (2021), 292–305.
[23] Eunji Kim, Jehyuk Lee, Hunsik Shin, Hoseong Yang, Sungzoon Cho, Seung-kwan Nam, Youngmi Song, Jeong-a Yoon, and Jong-il Kim. 2019. Champion-challenger analysis for credit card fraud detection: Hybrid ensemble and deep learning. *Expert Systems with Applications* 128 (2019), 214–224.
[24] Jung Youn Lee, Joonhyuk Yang, and Eric T. Anderson. 0. EXPRESS: Who Benefits from Alternative Data for Credit Scoring? Evidence from Peru. *Journal of Marketing Research* 0, ja (0), 00222437251360996. arXiv:https://doi.org/10.1177/00222437251360996 doi:10.1177/00222437251360996
[25] Stefan Lessmann, Bart Baesens, Hsin-Vonn Seow, and Lyn C Thomas. 2015. Benchmarking state-of-the-art classification algorithms for credit scoring: An update of research. *European journal of operational research* 247, 1 (2015), 124–136.
[26] Security Commissions Malaysia. 2025. *Enabling a More Relevant, Efficient and Diversified Market.* Retrieved September 30, 2025 from https://www.sc.com.my/api/documentms/download.ashx?id=61f72817-e06d-4e54-b25c-8be8abddcd28
[27] Andrew Y Ng. 2004. Feature selection, L 1 vs. L 2 regularization, and rotational invariance. In *Proceedings of the twenty-first international conference on Machine learning.* 78.
[28] Mfanasibili Ngwenya. 2024. Credit Scoring in Africa: Employing Logistic Regression on Alternative Data. In *2024 4th International Conference on Electrical, Computer, Communications and Mechatronics Engineering (ICECCME).* IEEE, Maldives, 1–6. doi:10.1109/ICECCME62383.2024.10796011
[29] Kuldeep Rawat. 2023. Applying CRISP-DM Methodology in Developing Machine Learning Model for Credit Risk Prediction. In *Science and Information Conference.* Springer, 522–538.
[30] Robert E Schapire. 2013. Explaining adaboost. In *Empirical inference: festschrift in honor of vladimir N. Vapnik.* Springer, 37–52.
[31] scikit learn. 2025. *AdaBoostClassifier.* Retrieved September 30, 2025 from https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.AdaBoostClassifier.html
[32] scikit learn. 2025. *GradientBoostingClassifier.* Retrieved September 30, 2025 from https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html
[33] scikit learn. 2025. *LogisticRegression.* Retrieved September 30, 2025 from https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html
[34] scikit learn. 2025. *RandomForestClassifier.* Retrieved September 30, 2025 from https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.RandomForestClassifier.html
[35] Tanmay Shivhare. 2024. *How to Utilize Bank Statements as a New Credit Scoring Method.* Ph. D. Dissertation. Dublin, National College of Ireland.
[36] Naeem Siddiqi. 2017. *Intelligent credit scoring: Building and implementing better credit risk scorecards.* John Wiley & Sons, North Carolina, USA.
[37] Shrawan Kumar Trivedi. 2020. A study on credit scoring modeling with different feature selection and machine learning approaches. *Technology in Society* 63 (2020), 101413.
[38] Stewart Watterson. 2024. *Cash Flow Underwriting: Reshaping the Lending Landscape.* Retrieved September 30, 2025 from https://assets.ctfassets.net/ss5kfr270og3/2LceEfOxttxwSpVtkiAPUw/dd87916c5d598b4fa8950d12f896ab1d/20241203_Cash_Flow_Underwriting_Reshaping_the_Lending_Landscape_White_Paper_Plaid.pdf