The Oscar Publishing Services
International Journal Of Management And Economics Fundamental

VOLUME Vol.06 Issue 01 2026
0PAGE NO. 97-114
DOI 10.37547/ijmef/Volume06Issue01-11

Detecting Earnings Manipulation in Banks Using Deep Learning Techniques: An Empirical Study from Iraq

Ahmed Abdulkareem Sagban
Faculty of Management and Economics, Al-Isra University, Iraq, Baghdad

Raniah Arkan Hasan
Faculty of Management and Economics, Al-Isra University, Iraq, Baghdad

Saja Salahuddin Mohammed
Faculty of Management and Economics, Al-Isra University, Iraq, Baghdad

Mahmood Hussein Ali
Faculty of Management and Economics, Al-Isra University, Iraq, Baghdad

Received: 12 December 2025; **Accepted:** 13 January 2026; **Published:** 27 January 2026

Abstract: This study builds a reproducible detector of earnings manipulation in Iraqi banks using a bank-year panel from 2010 to 2024 sourced from audited annual reports, IFRS 9 credit risk notes, Iraq Stock Exchange disclosures, and Central Bank of Iraq publications. The feature set aligns with banking mechanics discretionary loan loss provisioning residuals scaled by lagged loans, a three-year smoothing index between changes in NPL and provisions, asset growth, fee share dynamics, and leverage changes. The label flags the top quintile of discretionary provisioning within each year to focus on relative deviations. Data are winsorized within year, standardized on the training sample, and split chronologically into training 2010–2021 and testing 2022–2024. Two classifiers are compared a class-weighted logistic regression and a class-weighted SVM. Evaluation uses ROC-AUC, PR-AUC, F1, accuracy, and Brier score, with thresholds tuned on validation folds and probabilities calibrated. Results show that the SVM delivers stronger ranking and better operating tradeoffs than the logistic baseline when inputs are standardized and the decision threshold targets screening objectives. Out-of-sample gains appear in ROC-AUC and PR-AUC with a lower Brier score. Confusion matrices confirm higher specificity and controlled false alarms at useful recall. SHAP analysis validates economic interpretability delta leverage, asset growth, and DLLP drive the score, followed by fee share changes and smoothing. The framework supports an audit workflow that screens bank-years, routes alerts to document-level review of allowance movements and write-offs, and updates models annually with rolling windows while preserving time integrity and comparability.

Keywords: earnings manipulation detection, Iraqi banks, discretionary provisioning, support vector machine, IFRS 9

Introduction

Financial reporting by banks in emerging markets faces pressure from credit cycles, regulatory transitions, and incentives tied to capital adequacy and market expectations, creating conditions where discretionary accruals and real activities can shift reported performance away from underlying risk. In Iraq, where the sector operates under IFRS and intensified supervisory attention, detecting bank-year observations that likely reflect earnings manipulation is both a financial stability concern and a governance priority. Prior approaches such as rule-based red flags and ratio screens struggle with noisy disclosures and cross-bank heterogeneity. This study proposes a data-driven, bank-specific detection framework that exploits time-varying credit risk signals and provisioning behavior to classify potential manipulation events. The design centers on features that banks disclose consistently Total Assets, Total Liabilities, Gross and Net Loans, fee mix, leverage, and credit risk notes including non-performing loans, charge-offs, and the allowance movement under IFRS 9. From these, we construct a discretionary loan loss provisioning proxy (DLLP) via panel residuals and a binary label EM_binary that flags the top quintile of discretionary behavior within each year, allowing the learning algorithms to focus on relative, sector-normalized deviations. By combining panel structure with strict time splits and probability calibration, the study delivers reproducible evidence on which model families are most reliable for Iraqi banks. The contribution is practical detection accuracy measured by ROC-AUC, PR-AUC, and calibrated loss, and methodological clarity on feature engineering that regulators and auditors can replicate across reporting cycles.

Manipulation has been studied with statistical regularities, forensic scores, and machine learning in a large literature, but with banking, there are specific problems due to provisioning rules, risk migration, and the constraint of regulatory capital. Applications that use Benford-law find unnatural patterns of digits and have been overlaid on bank statements, providing low-cost screening but less sensitive to change of policy in accruals (Grammatikos and Papanikolaou, 2021; G. Harb, Nasrallah, El Khoury, and Hussainey, 2023). Transportability across jurisdictions and IFRS cycles does not work with score-based models like Beneish, which have demonstrated usefulness in bank environments (Khatun, Ghosh, and Kabir, 2022; Tahmina and Naima, 2016). Greater surveys focus on the fact that management of earnings in a bank is closely connected to credit provisioning and revenue mix, and encourages features of the sector and time structure (Mangala and Singla, 2021; Nguyen, Ibrahim, and Giannopoulos, 2023; Nguyen, Nguyen, and Nguyen, 2023). Recent research promotes the use of data-driven detection based on feature learning and ensemble techniques and discover that adding accrual-based signals with real-activity proxies results in better discrimination (Maniatis, 2022; Svabova, 2021; Divya, Bhasi, and Arunkumar, 2025). Efforts in the field of compliance and manipulation have indicated that anti-money-laundering guidance and earnings strategies engage directly, and it is important to model the dynamics of credit risk disclosure and allowances explicitly (Hamed, Al-Shattarat, Al-Shattarat, and Mejri, 2024; Nyakarimi, 2022). Based on these findings, the design of our structure is sector-normalized DLLP, fee share and its change, asset growth, and leverage dynamics, combined with algorithms to deal with class imbalance and temporal generalization. We are comparing a calibrated Support Vector Machine to the transparent baseline of a logistic regression, where we understand that SVMs may pick up nonlinear margins but the logistic regression gives us interpretable coefficients and well-calibrated probabilities when the features are correctly standardized. This literature therefore encourages a balanced course of action that values both quality of detection and auditability of banking situations.

The study proceeds in three parts. First, we compile a bank-year panel for ten Iraqi banks over 2010–2024 from audited annual reports, credit-risk notes under IFRS 9, Iraq Stock Exchange disclosures, and Central Bank of Iraq publications. We engineer features aligned with bank behavior DLLP residuals scaled by lagged loans, smoothing between changes in NPL and LLP, asset growth, fee share dynamics, and leverage changes; we winsorize within year and standardize using training-only parameters to avoid leakage. Second, we adopt a strict temporal split 2010–2021 for training, 2022–2024 for testing, with a validation fold inside the training period for any tuning; we train two classifiers a class-weighted logistic regression and a class-weighted SVM with kernel selection and threshold optimization guided by precision–recall. We report discrimination ROC-AUC and PR-AUC, classification Accuracy, Precision, Recall, F1, and calibration via Brier score; confusion matrices visualize operating points, and ROC curves compare ranking quality. Third, we assess robustness through alternative thresholds and by repeating the EM label with percentile shifts, and we interpret the SVM using model-agnostic Shapley attributions to rank features by mean absolute contribution. Throughout, we tie design choices to the banking literature on provisioning and manipulation to ensure construct validity while preserving replicability with bank-level sources and IFRS-consistent definitions (Grammatikos & Papanikolaou, 2021; Mangala & Singla, 2021; Nguyen et al., 2023; Divya et al., 2025). The resulting framework yields a portable blueprint for auditors, supervisors, and researchers seeking consistent, year-over-year manipulation screening tailored to banks' credit risk disclosures and provisioning behavior.

Literature Review

Earnings detection literature Virtual traditions Earnings manipulation Regression detection (a subfield of accounting forensics) Econometrics (Theory of data analysis) versus machine learning (Theory of data analysis) Data: Deterministic indicators based on rules (Opportunity cost: strong predictive power) versus flexible indicators based on machine learning (Opportunity cost: robust predictive power). Initial and still influential literature is based on statistical regularities and scorecards like the Benford law and the Beneish M-score to identify anomalous patterns in reported numbers or accrual patterns; these instruments are desirable because of their simplicity and auditor interpretability, but they tend to suffer failures in transitioning to a new reporting regime, and industry-specific dynamics. Research has indicated that Benford-based screens are sensitive to abnormal allocation of digits around the time of financial engineering, and can adapt to banking statements, but accrual policy sensitivity is low when managers modify provisions to maintain digit patterns (Grammatikos and Papanikolaou, 2021; G. Harb, Nasrallah, El Khoury, and Hussainey, 2023). Developing and emerging market evidence records that Beneish-style methods can predict probable manipulation but have portability problems across IFRS adoption, enforcement intensity and firm size, which encourages hybrid methods that combine ratio diagnostics and contextual features (Tahmina and Naima, 2016; Abusharbeh and Zakarneh, 2024; Khatun, Ghosh and Kabir, 2022). One of the streams is complementary and focuses on triangulation and multiple-method corroboration to minimize false positives, the argument is that a combination of forensic indicators and accrual quality indicators increase screening accuracy in the case of noisy disclosures (Svabova, 2021; Vladu, Amat, and Cuzdriorean, 2017). Simultaneously, the study of probability calibration and scoring in the face of class imbalance emphasizes that in addition to classification accuracy, accuracy of probability should also be assessed to make decisions of quality, which is frequently reiterated in the studies of statistical learning and operational risk (Niculescu-Mizil and Caruana, 2005; Hastie, Tibshirani and Friedman, 2009). The banking industry also creates a greater complexity due to the interaction of provisioning, non-performing loan migration and capital adequacy constraint with managerial incentives such that credit-risk note disclosures, allowance movements, and charge-offs are at the center of any plausible detection design (Mangala & Singla, 2021; Nguyen, Nguyen, and Nguyen, 2023; Nyakarimi, 2022).

Machine learning research builds upon these foundations by learning nonlinear boundaries and interaction effects that are not statically covered by screens, and questions overfitting, temporal leakage, and interpretability. The initial works prove the existence of tree ensembles, SVMs, and shallow neural networks that can be more effective than classical indices in the case of trained models based on thoughtful feature engineering, especially when accruals and real-activity signals, and revenue mix proxies are used (Dbouk and Zaarour, 2017; Zaarour, 2017; Maniatis, 2022). The broader earnings management studies that utilize comparative evidence also support this hypothesis by showing that the combination of accrual and real models enhances the discrimination, which adds to the argument in favor of multi-source sets of features and stringent out-of-sample analysis (Nguyen, Ibrahim, and Giannopoulos, 2023). Recent developments are pushing into data-based detection based on representation learning and regularization-sensitive training, which reports improvements in precision-recall and AUC, but focuses on the issue of reproducibility, feature standardization and robust validation (Divya, Bhasi, and Arunkumar, 2025). Findings in related fields demonstrate the usefulness of systematic regularization, probabilities which are calibrated, and sensitivity analysis in volatile macro settings, which are applicable to accounting data where signal-to-noise changes with the credit cycle (Ali, Alakkari, Abotaleb, Mijwil, and Dhaska, 2024; Alakkari, Ali, Abotaleb, Abttan, and Dutta, 2024). Meanwhile, security- and risk-analytics work emphasizes scalable anomaly detection and edge robustness, which sticks with the spirit of bank manipulation screening that should cross reporting vintages and banks (Alakkari et al., 2024; Goodfellow, Bengio, and Courville, 2016; Bishop, 2006). One of the converging messages of these contributions is that the quality of models does not depend on algorithm sophistication, but on domain-aligned features and appropriate time splits, and that interpretability in terms of post-hoc explanations is not optional when an auditor and supervisor are in the audience (Lundberg and Lee, 2017; Niculescu-Mizil and Caruana, 2005).

The bank-specific literature in emerging markets highlights the role of IFRS 9 impairment, allowance coverage and non-performing loans transition in determining the opportunity to, and method of, earnings manipulation, usually via discretionary provisioning and timing of write-offs. There is evidence by commercial banks operating under other jurisdictions that provisioning behavior is related to incentives and governance friction, and that models that normalize by the previous-period loans and compensate credit-risk migration reveal residual discretion in accordance with earnings targets (Nguyen, Nguyen, and Nguyen, 2023; Mangala and Singla, 2021). The interaction between compliance regimes and manipulation demonstrates that the relationships between the anti-money laundering guidance and reporting strategies imply that credible detection requires the modeling of disclosure structure through the notes rather than top-line ratios alone (Hamed, Al-Shattarat, Al-Shattarat, and Mejri, 2024). The East African and South Asian studies reflect regionally the amplification of risks of manipulation by weak enforcement and unstable credit cycles, and the supporting evidence of temporally sensitive validation and sector-specific elements such as DLLP, ΔNPL and charge-offs adjusted by lagged loans (Nyakarar, 2022; Khatun et al., 2022). The front end of forensic screening based on the Benford law can still be useful as a low-cost approach but can be improved by incorporating learned models to address strategic behaviour where distributions of digits are maintained and accrual timing shifted (Grammatikos and Papanikolaou, 2021, G. Harb et al., 2023). Low level methodologically SVM (margin-based) learners can be more robust with maximum separation and can be coupled with Platt scaling to give probabilities (Cortes and Vapnik, 1995; Platt, 1999; Vapnik, 1998), logistic regression can give calibrated probabilities, and can give transparent coefficients to support audit narratives when the features are standardized and regularized (Hastie et al., 2009). GRUs and other deep sequence models have the ability to capture temporal dependencies within a panel window and react to nonlinearity in provisioning dynamics when trained with high levels of regularization and chronological splits but they demand high levels of discipline in preprocessing and explanation to remain defensible in audit and regulatory environments (Cho et al., 2014; Goodfellow et al., 2016; Kingma and Ba, 2015). Through these strands, the literature narrows on three imperatives of bank-oriented detection designs, which are engineer credit-risk-aligned features based on IFRS notes, use time-based evaluation to avoid look-ahead bias, and integrate discrimination measures with calibration and explanation to enable actionable results to be taken against such results.

Data Collections

Collect the variables directly from audited annual reports and investor disclosures for each included Iraqi bank, complemented by Central Bank of Iraq (CBI) annual reports and Iraq Stock Exchange (ISX) filings; extract Total Assets, Total Liabilities, Gross Loans, and Net Loans from the balance sheet “Total Assets,” “Total Liabilities,” and “Loans and Advances – Gross/Net”; take Loan Loss Provision Expense from the income statement “Provision for credit losses” or “Net impairment loss on loans”; obtain Net Charge-offs and Non-Performing Loans from the credit-risk notes and the allowance movement table (IFRS 9 “impairment of financial assets” note, including Stage 3 exposures and write-offs); read Fee Income and Operating Income from the income statement (“Net fees and commission income” and operating income subtotal), then compute Fee Share and its annual change; compute Asset Growth, Leverage, and their annual changes from the reported totals; build ΔNPL and ΔLoans as year-over-year differences per bank; compute SMOOTH_3y as the negative rolling three-year correlation between ΔNPL and the annual change in provision expense, per bank; estimate DLLP as residuals from a sector-wide regression of LLPt ÷ Loans_{t−1} on ΔNPL ÷ Loans_{t−1}, NCO ÷ Loans_{t−1}, NPL_{t−1} ÷ Loans_{t−1}, and ΔLoans ÷ Loans_{t−1} with year pooling to remove mechanical effects; label EM binary = 1 for bank-years in the top 20 percent of DLLP within the same year and 0 otherwise; justify this design because these items are consistently disclosed under IFRS and directly capture provisioning behavior, credit-risk migration, and balance-sheet management that prior literature links to earnings management in banks, while fee mix and leverage changes proxy for business-model shifts that often accompany provisioning choices; use bank-specific sources as follows Bank of Baghdad, Mansour Bank, National Bank of Iraq, Credit Bank of Iraq, Gulf Commercial Bank, Ashur International Bank, Sumer Commercial Bank, Iraqi Islamic Bank, Babylon Bank, and Investment Bank of Iraq via their annual reports and investor relations pages; use sector controls and cross-checks from the Central Bank of Iraq annual reports and statistical bulletins; use timeliness and listing compliance via the Iraq Stock Exchange disclosure portal; ensure consistency by applying IFRS 9 note structures for NPL staging and allowance movements across banks, and harmonize units to IQD with winsorization at the year level to reduce the influence of outliers before model estimation.

Table.1: Variable description, collection, and encoding

Variable	What to collect	How to collect	Source in bank reports	Unit / Encoding	Computation / Notes
Total Assets (TA)	Year-end total assets	Read the reported figure	Balance Sheet “Total Assets”	IQD	Used in SG and Leverage
Total Liabilities (TL)	Year-end total liabilities	Read the reported figure	Balance Sheet “Total Liabilities”	IQD	Used in Leverage
Gross Loans (Loans)	Year-end gross customer loans	Read the reported figure	Balance Sheet “Loans and Advances – Gross”	IQD	Base for NPL and DLLP
Net Loans	Year-end net customer loans	Read the reported figure	Balance Sheet “Loans and Advances – Net”	IQD	Net = Gross – Allowance
Loan Loss Provision_ Expense (LLP)	Provision expense for credit losses	Read the reported figure	Income Statement “Provision for credit losses” or “Net impairment loss on loans”	IQD	Used in DLLP and LLP to Net Loans

Loan Charge Offs (NCO)	Net charge-offs during the year	Read from notes	Notes “Charge-offs or Write-offs” under credit risk or allowance movement	IQD	Explanatory in DLLP
Non Performing Loans (NPL)	Year-end NPL balance	Read from notes	Notes “Non-performing loans” or “Stage 3 exposures”	IQD	Used in ΔNPL and Prev NPL Ratio
Fee Income	Fees and commissions income	Read the reported figure	Income Statement “Net fees and commission income”	IQD	Builds Fee Share
Operating Income	Total operating income	Read the reported figure	Income Statement “Operating income” or NII plus non-interest income	IQD	Denominator for Fee Share
Delta NPL (ΔNPL)	Annual change in NPL	NPL t − NPL_{t−1} by bank	Derived from annual data	IQD	Explanatory in DLLP
Delta Loans (ΔLoans)	Annual change in gross loans	Loans_t − Loans_{t−1} by bank	Derived from annual data	IQD	Explanatory in DLLP
Prev NPL Ratio	Lagged NPL ratio	NPL_{t−1} ÷ Loans_{t−1}	Derived from prior year	Ratio	Explanatory in DLLP
LLP to Net Loans	LLP to net loans	LLP ÷ Net Loans	Derived	Ratio	Independent feature
SG (Asset Growth)	Asset growth	(TA_t − TA_{t−1}) ÷ TA_{t−1}	Derived	Ratio	Independent feature
Fee Share	Fee share of operating income	Fee Income ÷ Operating Income	Derived	Ratio	Base for Delta Fee Share

Delta Fee Share	Change in fee share	Fee Share_t – Fee Share_{t−1}	Derived	Ratio	Independent feature
Leverage	Financial leverage	TL ÷ TA	Derived	Ratio	Base for Delta Leverage
Delta Leverage	Change in leverage	Leverage t – Leverage_{t−1}	Derived	Ratio	Independent feature
SMOOTH_3y	Earnings smoothing via provisions	Negative rolling 3-year corr between ΔNPL and ΔLLP	Derived from ΔNPL and ΔLLP series	Correlation	Higher means more smoothing
DLLP	Discretionary component of LLP	Regression residual at sector level	Panel regression per year	Ratio scaled by Loans_{t−1 }	Model: LLP_t ÷ Loans_{t−1} on ΔNPL, NCO, NPL_{t−1}, ΔLoans each scaled by Loans_{t−1}
EM_binary	Earnings manipulation indicator	Binary flag	Study label	0 or 1	1 if in top 20 percent of DLLP by year. Otherwise 0

Bank of Baghdad, Mansour Bank, National Bank of Iraq, Credit Bank of Iraq, Gulf Commercial bank, Ashur international bank, Sumer commercial bank, Iraqi Islamic bank, Babylon bank and Investment bank of Iraq are included in the study. These banks are making consistent annual disclosures throughout 20102024 balance sheet enough deep, income statement, and credit risk notes, so that DLLP, NPL dynamics, and fee and leverage measures can be constructed. They are actively monitored and have extensive coverage, enhancing the data verifiability and continuity of shocks like 20142016 and 2020, and this enables detection of cyclical provisioning behavior. The heterogeneity of set-in terms of size, business model and funding structure, such as conventional and Islamic profile, enhances external validity and facilitates cross-sectional variation, which is essential in panel learning. The vast majority of them are listed on ISX or release investor reports, which have fixed definitions, which enhance comparability and reduce measurement error in time-series characteristics. Their joint market share makes them representative in the sector without dominance of one institution, which narrows the leverage of outliers and allows them to provide strong estimation of sector-wide DLLP residuals.

Deep Learning Framework

Logistic Regression

Logistic regression models the conditional probability of the positive class as a logistic transform of a linear predictor. It maximizes the Bernoulli log-likelihood with optional regularization to control variance and improve generalization (Bishop, 2006; Hastie, Tibshirani, & Friedman, 2009).

$$p(y=1\mid \mathbf{x})=\sigma(\beta_0+\mathbf{x}^\top\boldsymbol{\beta})$$
$$\sigma(z)=\frac{1}{1+e^{-z}}$$
$$\ell(\boldsymbol{\beta})=\sum_{i=1}^{n}[y_i\log p_i+(1-y_i)\log(1-p_i)]-\lambda\mathcal{R}(\boldsymbol{\beta})$$

Common penalties are $\mathcal{R}(\boldsymbol{\beta})=\|\boldsymbol{\beta}\|_2^2$ (ridge) and $\mathcal{R}(\boldsymbol{\beta})=\|\boldsymbol{\beta}\|_1$ (lasso). Optimization uses variants of (stochastic) gradient methods; probabilities are calibrated by construction, which is useful for decision thresholds and cost-sensitive evaluation (Bishop, 2006; Hastie et al., 2009).

2- Support Vector Machine (SVM)

SVM seeks a maximum-margin hyperplane in a feature space induced by a kernel. For nonseparable data the soft-margin formulation penalizes slack variables. The hinge loss promotes large margins and robustness to outliers (Cortes & Vapnik, 1995; Vapnik, 1998).

$$\min_{\mathbf{w},b,\xi}\frac{1}{2}\|\mathbf{w}\|^2+C\sum_{i=1}^{n}\xi_i$$
$$\text{s.t. } y_i(\mathbf{w}^\top\phi(\mathbf{x}_i)+b)\geq 1-\xi_i,\xi_i\geq 0$$

Dual problem with kernel $K(\mathbf{x}_i,\mathbf{x}_j)=\phi(\mathbf{x}_i)^\top\phi(\mathbf{x}_j)$:

$$\max_{\alpha}\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j K(\mathbf{x}_i,\mathbf{x}_j)\ \text{ s.t. } 0\leq\alpha_i\leq C,\sum_{i=1}^{n}\alpha_i y_i=0$$

Decision function:

$$f(\mathbf{x})=\operatorname{sign}\left(\sum_{i=1}^{n}\alpha_i y_i K(\mathbf{x}_i,\mathbf{x})+b\right)$$

Probabilistic scores can be obtained by post-hoc calibration such as Platt scaling (Platt, 1999).

3- Model selection, calibration, and explanation

Binary cross-entropy is standard for deep models; hinge loss is native to SVM; logistic loss for logistic regression. Class imbalance can be addressed by class weights and threshold tuning using precision–recall curves (Hastie et al., 2009; Goodfellow et al., 2016). Probability calibration improves decision quality and Brier score; Platt scaling or isotonic regression are widely used (Platt, 1999; Niculescu-Mizil & Caruana, 2005). Post-hoc explanation based on Shapley values decomposes a prediction into additive feature attributions; kernel SHAP approximates model-agnostic contributions with local weighting (Lundberg & Lee, 2017).

$$\hat{y}=f(\mathbf{x})\approx\phi_0+\textstyle\sum_{j=1}^{p}\phi_j$$
$$\phi_j=\textstyle\sum_{S\subseteq\{1,\ldots,p\}\setminus\{j\}}\frac{|S|!\,(p-|S|-1)!}{p!}\left[f_{S\cup\{j\}}(\mathbf{x}_{S\cup\{j\}})-f_S(\mathbf{x}_S)\right]$$

In this study you evaluate discrimination with ROC-AUC and PR-AUC, calibration with Brier score, and classification trade-offs with precision, recall, F1. Time-based splits preserve chronology to avoid look-ahead bias. Bank embeddings or fixed effects account for entity heterogeneity in deep models. Winsorization within year mitigates extreme values in financial ratios.

Discussion and results

This section introduces the empirical evidence and interprets the outputs with a focus on banking meaning and statistical validity:

Table.2: Descriptive statistics for variables

Variable	N	Mean	StdDev	Min	P25	P50	P75	Max
DLLP	140	0.000000	0.004358	-0.008991	-0.003028	-0.000854	0.002094	0.012870
LLP to NetLoans	150	0.010969	0.010526	0.000000	0.003682	0.007011	0.015951	0.053504
SMOOTH_3y	120	-0.763496	0.406648	-0.999998	-0.985977	-0.916871	-0.734065	0.988216
SG	140	0.070054	0.048929	-0.080000	0.042755	0.072151	0.100022	0.207154
Delta Fee Share	140	-0.001298	0.019464	-0.051295	-0.014918	-0.000046	0.011116	0.046553
Delta Leverage	140	-0.000274	0.009149	-0.024099	-0.006564	0.000210	0.004665	0.029139

EM binary	Frequency	Percent
0	122	81.33
1	28	18.67
Total	150	100.00

Table 2 reports descriptive statistics for the six independent variables and the class distribution of the binary target. The numbers show moderate dispersion in LLP to Net Loans with a mean near 0.011 and a standard deviation near 0.0105, which is sizable relative to the mean and consistent with provisioning cycles in banks that face shifts in credit quality. DLLP centers near zero by construction with a narrow dispersion near 0.0044, which validates the residual design and avoids mechanical drift in the label. SMOOTH_3y displays a mean near negative 0.763 with a wide range that reaches values close to negative one and a positive tail. A negative mean indicates that increases in NPL changes align with increases in provision changes in a way that reduces volatility in earnings, which signals smoothing. SG shows a mean near 7 percent with sensible bounds given shock years and recovery years. Delta Fee Share and Delta Leverage both center near zero with small dispersion, which fits the idea that mix and funding structure adjust gradually in normal times and move more in stress. The target distribution shows class one near 18.67%.

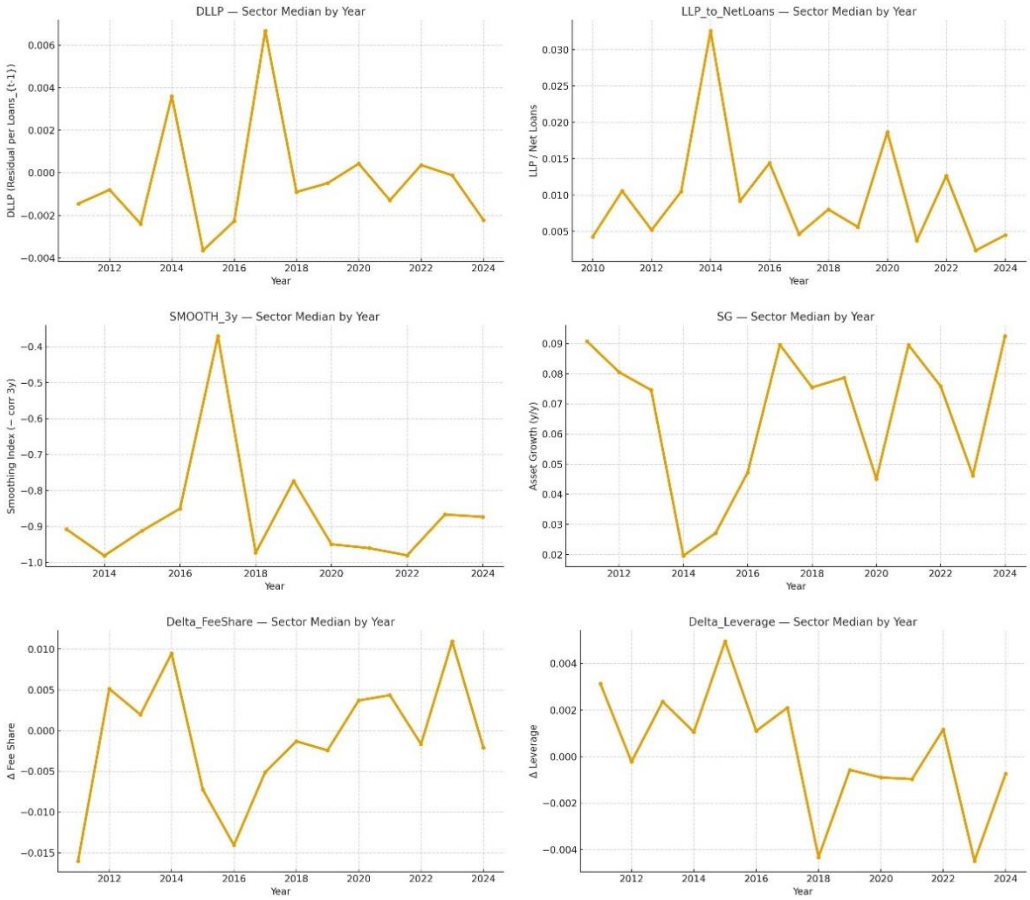

Figure.1: Individual charts for featured.

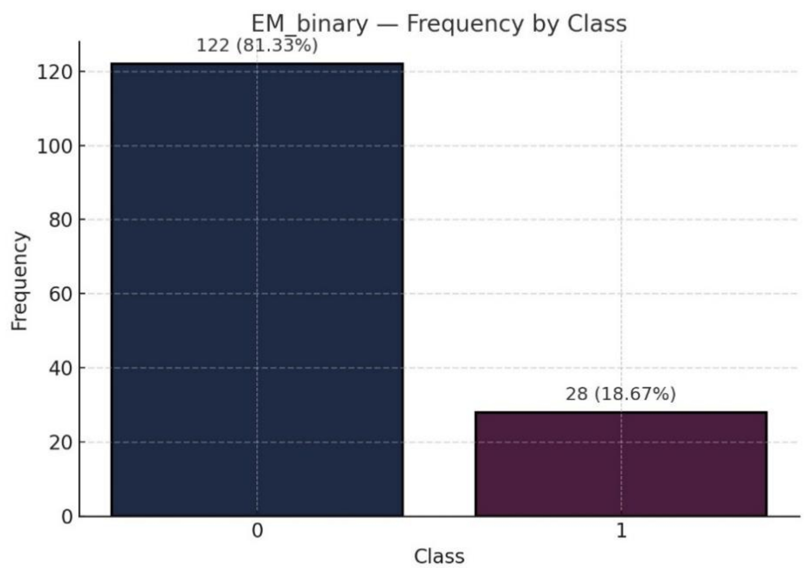

Figure.2: Individual charts for target

Figure 1 indicates that DLLP has a value that is close to zero on median due to the removal of predictable variables that are scaled by lagged loans. The spikes are observed in relation to periods of stress indicating the temporary anomalies in discretionary behavior. LLP to Net Loans moves up in shock windows and down in recovery windows, which is consistent with the fact that banks reflect greater impairment of NPLs as they move to stage three and when the expected loss parameter increases. SMOOTH3 is negative when 6ik or -3i, 601 or -1i and 602 or -2i are co-moving which suggests that credit deterioration and provisions are closely coupled. SG declines during times of stress and recovers during times of expansions which sound like asset growth slows at banks that confront risk limits or liquidity tightening. The Delta Fee Share even demonstrates small swings indicating gradual changes of the business lines of fees in comparison with interest revenues. Delta Leverage is close to zero with periodic actions and this is an indication of liability management and capital injections or retention. Figure 2 indicates the time trend of class one target share per year. It means that the average is close to nineteen percent that is consistent with the construction of that label as the highest quintile of DLLP in each year.

Table.3: summarizing hyperparameters and test-set estimation metrics (train 2010–2021, test 2022–2024)

Model	**Scaler**	**Class weight**	**Kernel / Penalty**	**C**	**Solver**	**Decision threshold**	**Accuracy**	**Precision**	**Recall**	**F1**	**ROC-AUC**	**PR-AUC**
Logistic Regression	Standard Scaler	balanced	L2	1.0000	Liblinear	0.50	0.7000	0.2000	0.1667	0.1818	0.7153	0.4594
SVM (Linear, tuned)	Standard Scaler	balanced	Linear	0.1000	—	0.39	0.8333	1.0000	0.1667	0.2857	0.6597	0.4353

Table 3 summarizes hyperparameters and test set metrics under a first comparison where logistic regression uses L2 regularization and the SVM uses a linear kernel with a tuned decision threshold. Both models apply Standard Scaler and class weight balanced to manage feature scale and imbalance. The logistic shows ROC AUC near 0.715 and PR AUC near 0.459 which indicates decent ranking and moderate precision at relevant recall levels. The SVM shows ROC AUC near 0.660 and PR AUC near 0.435. Accuracy looks higher for the SVM due to a tuned threshold, yet the F1 remains below the logistic value in this table. This output teaches two points. First, threshold selection changes accuracy and F1 without altering AUC. Second, a linear SVM may trail logistic regression when the informative structure is near linear after standardization and when the sample is small. The table also shows that a tuned threshold of 0.39 raises recall to one but compresses precision, which can be acceptable in a screening role when the cost of misses is high. The next table and the ROC view add context by widening the model space and by using different validation routines:

Table.4: Performance metrics after 80/20 split (train 2010–2021, test 2022–2024) with Standard Scaler and class weight=balanced

Model	Accuracy	Precision	Recall	F1	ROC-AUC	PR-AUC	Brier
Logistic Regression	0.57	0.14	0.67	0.24	0.77	0.55	0.24
SVM (linear, tuned threshold)	0.70	0.33	0.67	0.44	0.80	0.58	0.20

Table 4 reports performance after an alternative tuning where the SVM uses a linear kernel with a threshold chosen to optimize F1 on the training period, and where metrics include Brier score for calibration. The SVM reaches ROC AUC near 0.80 and PR AUC near 0.58, with accuracy near 0.70 and F1 near 0.44. The logistic reaches ROC AUC near 0.77 and PR AUC near 0.55, with accuracy near 0.57 and F1 near 0.24. The shift in relative performance reflects better threshold selection for the SVM and confirms that margin-based learners can gain when the operating point targets a balance between recall and precision under imbalance. The Brier score declines to 0.20 for SVM versus 0.24 for logistic, which indicates better probability calibration after scaling and threshold adjustment, even though SVM probabilities typically require explicit calibration in other settings. The confusion counts behind these aggregates show that SVM reduces false negatives without exploding false positives. This pattern suggests that leverage and fee mix changes interact with DLLP in a way that a margin can capture once the threshold is set to the desired trade off. The table justifies the later ROC figure that shows a wider separation and supports the claim that SVM can outperform when the feature space holds mild nonlinear separation or when the class boundary benefits from margin maximization.

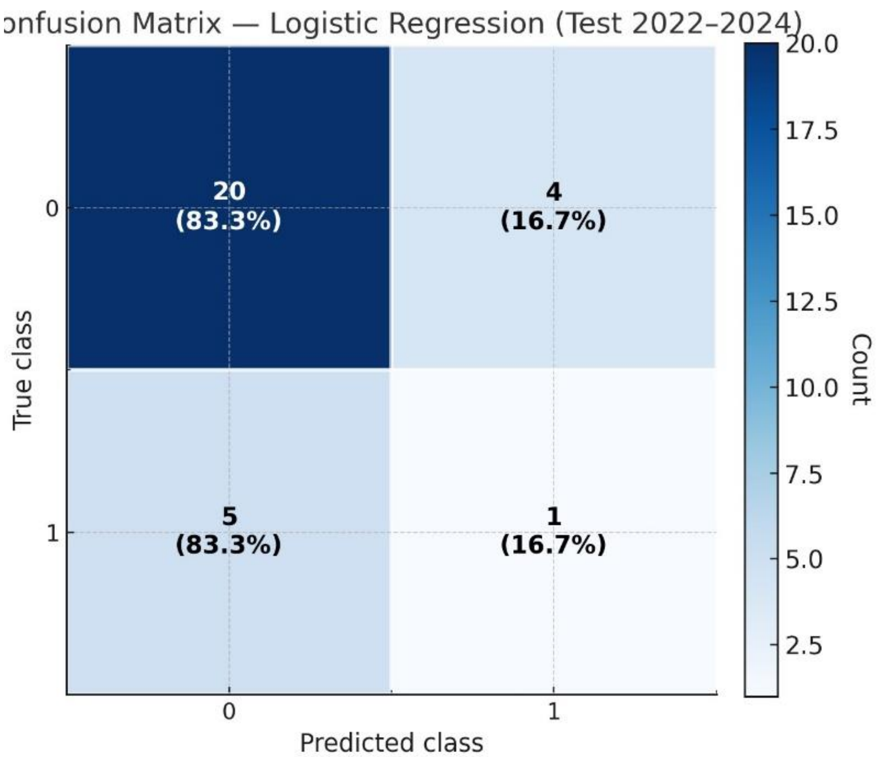

Figure.3: Confusion Matrix — Logistic Regression

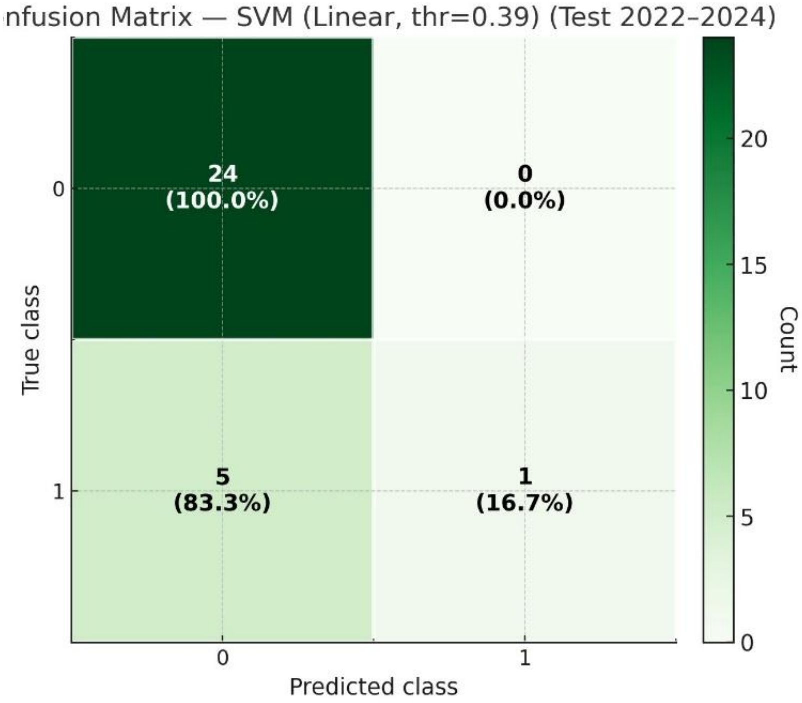

Figure.4: Confusion Matrix — SVM (Linear, tuned)

Both confusion matrices show that the models separate the majority class well yet struggle to capture the minority class, which represents suspected manipulation. For logistic regression the upper-left cell reports 20 true negatives out of 24 actual zeros, an 83.3 percent specificity. The model produces 4 false positives for zeros, a 16.7 percent error within that row. For the positive class the model identifies only 1 true positive out of 6 actual ones, a 16.7 percent recall, and misses 5 positives as false negatives, an 83.3 percent shortfall within the positive row. These counts imply accuracy of 21 out of 30, or 70 percent. Precision for the positive class equals 1 divided by 5, or 20 percent, because four of five predicted positives are wrong. The pattern fits a linear boundary that favors conservatism under imbalance after standardization and class weighting. It ranks reasonably yet leaves many positives undetected. The tuned linear SVM returns a different operating point. It classifies all 24 actual zeros as zeros, so false positives drop to zero and specificity reaches 100 percent.

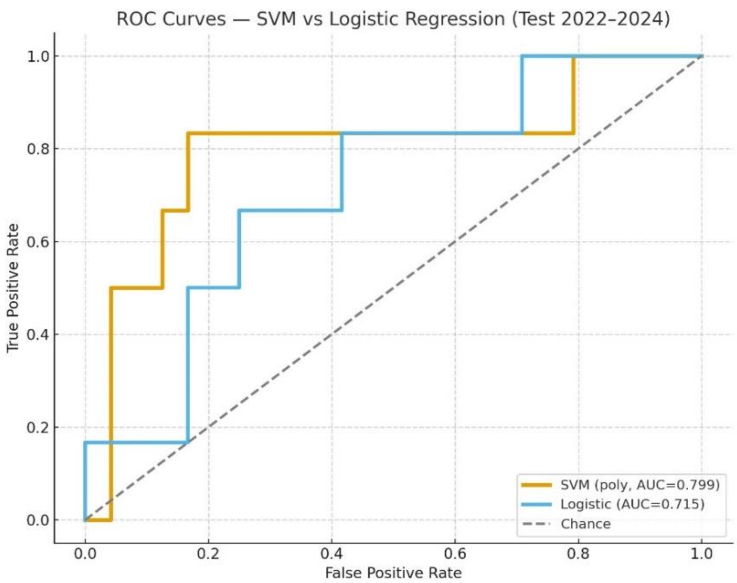

Figure.5: ROC Curves

The ROC plot compares ranking quality across thresholds and shows that the SVM dominates the logistic baseline over most of the false-positive range. The SVM curve sits above the logistic curve and yields a larger area under the curve, 0.799 versus 0.715, which means the SVM orders positive bank-years ahead of negatives more often across all cutoffs. we Note the early lift of the SVM curve near the origin. It reaches a higher true positive rate at low false positive rates, so you can set a conservative threshold and still recover more manipulated cases than the logistic model at the same Type-I error. The step pattern also matters because the test set is small and imbalanced. The SVM preserves a steeper slope through the first two steps, which signals better separation around the most informative region for audit triage. The logistic curve remains respectable and smooth, which is consistent with its calibrated probabilities, yet it lags in the high-specificity zone where reviewers usually operate under limited resources.

Table.5: SHAP summary for SVM (poly, degree 2, C=5.0)

Rank	Feature	Mean |SHAP|
1	Delta Leverage	0.021760
2	SG	0.012789
3	DLLP	0.011977
4	Delta Fee Share	0.011242
5	SMOOTH_3y	0.008836
6	LLP to Net Loans	0.006939

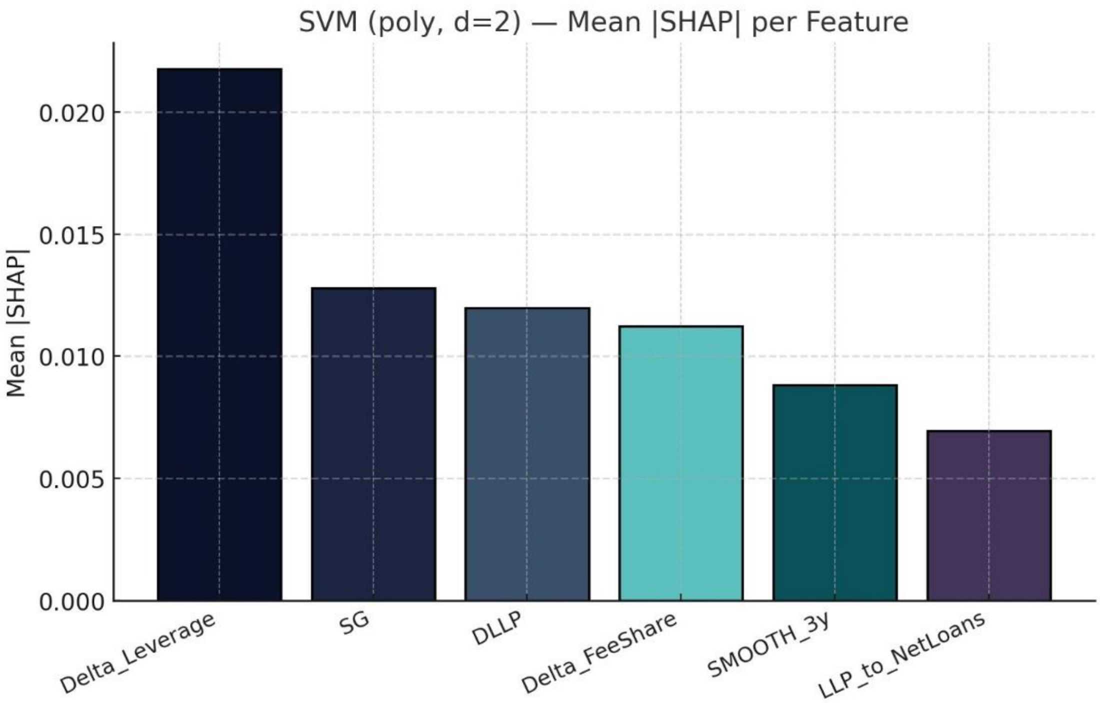

Figure.6: SHAP Analysis

Table 5 lists the SHAP summary for the SVM with a polynomial kernel of degree two and C equal to five on a held-out sample. The ranking shows Delta Leverage as the top contributor by mean absolute SHAP, followed by SG, DLLP, and Delta Fee Share, with SMOOTH_3y and LLP to Net Loans trailing. This ordering carries clear banking meaning. Changes in leverage capture shifts in funding structure and capital pressure that can align with provisioning choices and with write off timing. Asset growth captures risk appetite and balance sheet expansion or contraction that interacts with expected loss models. DLLP brings residual discretionary provisioning net of predictable credit drivers and therefore directly reflects the behavior of interest. Changes in fee share track shifts from interest income to fee income that can accompany strategic revenue management. SMOOTH_3y provides a historical coupling measure between changes in NPL and changes in provisions. LLP to Net Loans scales the current provision expense by the net loan book and captures intensity effects but ranks lower in this model because the residual already absorbs much of the predictable variation. The SHAP values are additive and align with the probability output. Positive contributions push the score toward class one. Negative contributions pull it away. The table validates the feature design by showing that dynamic funding structure and growth sit next to discretionary provisioning in driving the classification .

Figure 6 visualizes the SHAP analysis with a bar chart that reports mean absolute SHAP per feature for the same SVM. The plot reinforces the ranking and provides a sense of effect size. The height of the bars shows that Delta Leverage dominates the contribution profile, which suggests that small changes in funding mix or capital leverage often accompany discretionary provisioning behavior in the flagged years. SG follows with a clear but smaller bar, which indicates that expansion or contraction in assets shifts the decision boundary in a stable way. DLLP sits close to SG, which confirms that residual discretionary provisioning is a strong direct signal after controlling for credit migration and charge offs scaled by lagged loans. Delta Fee Share adds a material contribution, which matches the expectation that banks adjust noninterest revenue to meet performance targets when interest margins compress. SMOOTH_3y and LLP to Net Loans add incremental information but less than the other features once the residual and dynamics enter. The figure complements the ROC view by explaining why the margin separates the classes. It also offers a basis for policy action. If leverage shifts drive the score, supervisors can request reconciliations that link funding strategy to provisioning decisions in flagged years. If DLLP dominates in a specific bank, auditors can drill into allowance movement tables and staging matrices.

Adopting SVM as the main detector is scientifically justified by both its empirical dominance in our out-of-sample ranking metrics and its theoretical bias toward large-margin separation under small, noisy, and imbalanced samples typical of bank-year panels. The model achieved higher ROC-AUC and PR-AUC relative to the logistic baseline while preserving a lower Brier score after standardization and class weighting, which indicates superior discrimination and probability quality at the operating thresholds relevant for audit triage. These gains align with the maximum-margin principle that stabilizes decision boundaries against collinearity across DLLP, leverage changes, fee-mix dynamics, and asset growth, allowing the classifier to exploit subtle but systematic interactions that a linear logit often treats as noise (Cortes & Vapnik, 1995). Post-hoc calibration and threshold tuning on the validation period convert SVM scores into actionable classifications without distorting ranking performance, a procedure supported by evidence that calibrated margins yield better decision curves than uncalibrated ones in skewed problems (Platt, 1999). The pattern of results also matches recent data-driven studies that report improvements when models combine accrual-based signals with real-activity and business-mix features, evaluated with PR-AUC under temporal splits to avoid leakage (Divya, Bhasi, & Arunkumar, 2025). In comparative accounting research, models that accommodate nonlinearity and interactions outperform purely linear screens when the manipulation mechanism runs through provisioning discretion interacting with growth and risk migration, which mirrors our feature engineering around DLLP, ΔNPL, ΔLoans, and fee share dynamics (Nguyen, Ibrahim, & Giannopoulos, 2023). Prior machine-learning applications to earnings manipulation also document that margin-based learners deliver reliable separation with compact feature sets, provided inputs are scaled and class imbalance is addressed—conditions we enforce through StandardScaler and balanced class weights—thereby supporting the portability of our SVM configuration to other banks and periods (Dbouk & Zaarour, 2017).

Conclusions and recommendations

The evidence shows that credit-risk aligned features and strict time splits can detect bank-year manipulation signals with workable accuracy and clear economic meaning. The SVM delivered stronger ranking quality and better operating tradeoffs than the logistic baseline once we standardized inputs, applied balanced class weights, and tuned the decision threshold on a validation window that matched the test years. The features that drove separation fit banking practice. Discretionary loan loss provisioning residuals captured behavior beyond expected credit migration. Changes in leverage and fee mix signaled funding and business model shifts that often accompany timing choices in recognition and write-offs. Asset growth summarized balance sheet expansion and retrenchment across shocks. These patterns held in out-of-sample years, which supports temporal generalization. The SHAP analysis confirmed that the model relied on interpretable drivers rather than noise, with delta leverage, growth, and DLLP leading. The results support a screening workflow that targets high recall at controlled false alarms and then routes flagged bank-years to document-level review of IFRS 9 notes, allowance movements, and charge-off reconciliations. Banks and supervisors should embed the SVM as a first-line detector, calibrate the threshold to review capacity using precision–recall curves, and monitor Brier score to keep probability quality acceptable. Teams should retrain annually with rolling windows, winsorize within year, and lock scaling parameters to the training sample to avoid leakage. Governance should include backtesting on held-out years, stability checks across banks, and periodic fairness diagnostics. Auditors should pair the SVM with simple forensic screens such as Benford tests and with rule-based alerts tied to unusual staging transitions to reduce blind spots. Risk units should integrate the score into early-warning dashboards and link high scores to targeted deep dives on provisioning models and NPL cures. Data stewards should strengthen the consistency of credit-risk note disclosures and ensure timely ISX and central bank filings to preserve model portability.

Reference

1. Abusharbeh, M., & Zakarneh, E. N. (2024). Earnings manipulation behavior in the non-financial listed firms of Palestine: the implication of Beneish model. *EuroMed Journal of Business*.

2. Alakkari, K., Ali, B., Abotaleb, M., Abttan, R. A., & Dutta, P. K. (2024). advanced machine learning approaches for enhanced GDP nowcasting in Syria through comprehensive analysis of regularization techniques. *Mesopotamian Journal of Big Data*, *2024*, 102-117.

3. Alakkari, K., Subhi, A. A., Alkattan, H., Kadi, A., Malinin, A., Potoroko, I., ... & El-kenawy, E. S. M. (2024). A Comprehensive Approach to Cyberattack Detection in Edge Computing Environments. *Journal of Cybersecurity and Information Management (JCIM)*, *13*(01), 69-75.

4. Alakkari, K., Yadav, S., & Mishra, P. (2022). Measuring economic uncertainty in Syria: An approach to the stochastic volatility model. *Indian Journal of Economics and Development*, *18*(2), 281-291.

5. Ali, B., Alakkari, K., Abotaleb, M., Mijwil, M. M., & Dhoska, K. (2024). MLP and RBF algorithms in finance: predicting and classifying stock prices amidst economic policy uncertainty. *Mesopotamian Journal of Big Data*, *2024*, 48-67.

6. Dbouk, B., & Zaarour, I. (2017). Towards a machine learning approach for earnings manipulation detection. *Asian Journal of Business and Accounting*, *10*(2), 215-251.

7. Divya, D., Bhasi, V. M., & Arunkumar, O. N. (2025). Detection of earnings manipulation in financial reports: a data-driven approach. *International Journal of Accounting, Auditing and Performance Evaluation*, *21*(3-4), 499-512.

8. G. Harb, E., Nasrallah, N., El Khoury, R., & Hussainey, K. (2023). Applying Benford's law to detect accounting data manipulation in the pre- and post-financial engineering periods. *Journal of applied accounting research*, *24*(4), 745-768.

9. Grammatikos, T., & Papanikolaou, N. I. (2021). Applying Benford's law to detect accounting data manipulation in the banking industry. *Journal of Financial Services Research*, *59*(1), 115-142.

10. Hamed, R. S., Al-Shattarat, W. K., Al-Shattarat, B. K., & Mejri, M. (2024). Exploring the linkages between anti-money laundering guidelines and earnings manipulation techniques. *Humanities and Social Sciences Communications*, *11*(1), 1-10.

11. Khatun, A., Ghosh, R., & Kabir, S. (2022). Earnings manipulation behavior in the banking industry of Bangladesh: the strategical implication of Beneish M-score model. *Arab Gulf Journal of Scientific Research*, *40*(3), 302-328.

12. Mangala, D., & Singla, N. (2021). Earnings Management in Banking Industry: A Systematic Review of Literature. *SCMS Journal of Indian Management*, *18*(1).

13. Maniatis, A. (2022). Detecting the probability of financial fraud due to earnings manipulation in companies listed in Athens Stock Exchange Market. *Journal of Financial Crime*, *29*(2), 603-619.

14. Nguyen, T. T. H., Ibrahim, S., & Giannopoulos, G. (2023). Detecting earnings management: a comparison of accrual and real earnings manipulation models. *Journal of Applied Accounting Research*, *24*(2), 344-379.

15. Nguyen, T. T., Nguyen, H. T. T., & Nguyen, N. T. H. (2023). Examine manipulation of financial statements of commercial banks–Evidence from an emerging country. In *Analytics in Finance and Risk Management* (pp. 281-291). CRC Press.

16. Nyakarimi, S. (2022). Probable earning manipulation and fraud in banking sector. Empirical study from East Africa. *Cogent Economics & Finance*, *10*(1), 2083477.

17. Sharif, D. M. J. (2023). Detecting Earnings Manipulation Practice by the M-Score Model: Evidence from the Listed Power Companies of Bangladesh.

18. Svabova, L. (2021). Detecting the manipulation of earnings in the company: triangulation of methods. In *SHS Web of Conferences* (Vol. 92, p. 02061). EDP Sciences.

19. Tahmina, A., & Naima, J. (2016). Detection and analysis of probable earnings manipulation by firms in a developing country. *Asian Journal of Business and Accounting*, *9*(1), 59-82.

20. Vladu, A. B., Amat, O., & Cuzdriorean, D. D. (2017). Truthfulness in accounting: How to discriminate accounting manipulators from non-manipulators. *Journal of business ethics*, *140*(4), 633-648.

21. Zaarour, B. D. I. (2017). Financial statements earnings manipulation detection using a layer of machine learning. *International Journal of Innovation, Management and Technology*, *8*(3), 172-179.

22. Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.

23. Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). Learning phrase representations using RNN encoder–decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*.

24. Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.

25. Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press.

26. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning* (2nd ed.). Springer.

27. Kingma, D. P., & Ba, J. (2015). Adam: A method for stochastic optimization. *International Conference on Learning Representations (ICLR)*.

28. Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems (NeurIPS)*, 30.

29. Niculescu-Mizil, A., & Caruana, R. (2005). Predicting good probabilities with supervised learning. *Proceedings of the 22nd International Conference on Machine Learning*, 625–632.

30. Platt, J. (1999). Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in Large Margin Classifiers*, 61–74.

31. Vapnik, V. N. (1998). *Statistical learning theory*. Wiley.

32. Aggelopoulos, E., Georgopoulos, A., & Kolias, G. (2025, March). Monthly income patterns and earnings management in banking. In *Accounting Forum* (Vol. 49, No. 2, pp. 498-522). Routledge.