Model Governance Documentation
Bank Account Fraud Detection

Document ID: FCD202601

Document status: In Progress (unpublished)

Last update: 12 February 2026

Version number: 1.0

1. Overview

1.1 Purpose of the Document

Model Risk Management (MRM) is a critical aspect of any financial institution, especially those dealing with fraud detection.

Bank Account Fraud (BAF) Detection aims to develop and implement a mix of models to identify fraudulent bank accounts. This documentation aims to meet Supervisory Letter SR 11-7 standards.

Supervisory Letter SR 11-7

SR 11-7 is a set of standards and guidelines developed by the United States Securities and Exchange Commission (SEC) to ensure that financial institutions have robust systems in place for detecting, preventing, and responding to fraud. Essential components of SR 11-7 include:

- **Definition of a Model**: Any quantitative method, system, or approach using statistical or economic theories to process data into estimates.
- **Model Validation**: An independent, comprehensive review to ensure models are accurate and appropriate for their intended use.
- **Governance and Controls**: Banks must establish strong policies, procedures, and accountability mechanisms for model development, implementation, and ongoing monitoring.
- **Scope**: Applies to all institutions supervised by the Federal Reserve or OCC, including national and state banks.

Documentation Components

1 Model Identification and Purpose

   a. **Model Name and Version**: Unique identifiers that track the specific iteration of the fraud models being deployed.

   b. **Business Use Case**: A clear statement defining how the model detects specific fraud typologies within the BAF dataset, such as application and synthetic identity fraud.

   c. **Model Owner and Stakeholders**: Documentation of the individuals or teams accountable for the model's performance and regulatory compliance.

2 Data Integrity and Lineage

   a. **Data Sources and Scope**: A description of the input variables from the BAF dataset, including transaction attributes and behavioural features.

   b. **Data Quality Assessment**: Evidence of checks for missing values, outliers, and data freshness to ensure the model isn't learning from "garbage" data.

   c. **Feature Engineering Rationale**: An explanation of why specific derived features were created to capture complex fraud patterns.

(Continue to the next page)

3 Model Development and Methodology

   a. **Algorithm Selection**: Justification for choosing specific architectures (e.g. Logistic Regression and XGBoost) based on the BAF dataset's characteristics.

   b. **Model Assumptions and Limitations**: A transparent list of what the model cannot do and the conditions under which it might fail.

   c. **Training and Testing Split**: Details on how the data was partitioned to prevent data leakage and ensure the model generalizes well to unseen fraud.

4 Validation and Testing

   a. **Performance Metrics**: Results of testing using metrics like Precision, Recall, and the Area Under the Precision-Recall Curve (AUPRC), which are critical for imbalanced fraud data.

   b. **Sensitivity Analysis**: Documentation of how the model reacts to changes in input variables or shifts in fraudster behavior.

   c. **Back-testing Results**: A comparison of predicted fraud vs. actual historical outcomes within the BAF suite to prove accuracy.

5 Implementation and Monitoring

   a. **Deployment Environment**: A description of the technical infrastructure where the model resides and how it integrates with real-time payment rails.

   b. **Ongoing Monitoring Plan**: The schedule and thresholds for tracking "model drift" to ensure the model stays effective as fraud tactics evolve.

   c. **Change Management Process**: A formal protocol for how the model will be updated, retrained, or decommissioned in the future.

1.2 Document Version Control

Document version	Change date	Approver	Status
1.0	12 February 2026	George Li	Approved

2. Model Identification and Purpose

2.1 Model Name and Version

1.	General Information	
1.1	Model Code	FCD202601
1.2	Model Name	Bank Account Fraud Detection (BAF)
1.3	Model Version	1.0
2.	Technical Information	
2.1	Model Type	• Logistic Regression (regression) • XGBoost (decision tree)

2.2	Model Purpose	This solution uses a mix of logistic regression and XGBoost models to detect fraudulent bank accounts.
3	Contact Information	
3.1	Model Owner	Li Zhaozhi (email@address.com)
3.2	Model Risk Owner	George Li (email@address.com)
3.3	Model Users	First Name, Last Name – Department Name – Email Address

2.2 Business Use Case

2.3 Model Owner and Stakeholders

3. Data Integrity and Lineage

3.1 Data Sources and Scope

In a banking environment, both reliable open-source research data from trusted vendors and internal banking system data are available for data analytics and model training for fraud detection.

Overview of the Data Sets

Feedzai is an AI fraud detection platform that uses machine learning to detect fraud. Feedzai Research released anonymized data sets at NeurIPS 2022 resembling challenges in bank account fraud data such as class imbalance, missing values, and outliers.

These data sets are available in downloadable CSV format and offers numeric and categorical variables in a range of dimensions useful for analysing fraud patterns and predicting fraudulent activities such as financial data, customer profile data, credit risk data, behavioural data, digital metadata, and time series data etc.

Other dimensions useful for analysing fraud patterns and predicting fraudulent activities such as transaction data, biometric data, and fraud typologies from case data are not available in this data set.

List of Available Dimensions and Variables

- **Financial data**:

  `income, intended_balcon_amount`

- **Credit Risk data**:

  `credit_risk_score, proposed_credit_limit`

- **Digital metadata**:

  `device_os, device_fraud_count`

- **Customer Profile data**:

`name_email_similarity, customer_age, bank_months_count, phone_mobile_valid, phone_home_valid, email_is_free, employment_status, housing_status, days_since_request, zip_count_4w`

- **Behavioural data**:

  `session_length_in_minutes, keep_alive_session, foreign_request, velocity_6h, velocity_24h, velocity_4w, prev_address_months_count, current_address_month_count, source, bank_branch_count_8w, date_of_birth_distinct_mails_4w, device_distinct_emails_8w`

- **Time Series data**:

  `month`

References:

- Bank Account Fraud Dataset Suite Datasheet: Feedzai Research authored this datasheet alongside the research data to provide explanation.

The following Data Definitions Table details these variables, including variable name, definition, data type, unit, and example values.

Data Definitions Table

Num	Variable	Definition	Data Type	Unit	Example
1	`fraud_bool`	Fraud label (1: Fraud, 0: geunine)	Numerical	N/A	1
2	`income`	Annual income in quantiles	Numerical	N/A	0.3
3	`name_email_similarity`	Metric of similarity between email and applicant's name. Higher values represent higher similarity. Ranges between [0, 1].	Numerical	N/A	1
4	`prev_address_months_count`	Number of months in previous registered address of the applicant, i.e. the applicant's previous residence, if applicable. Ranges between [−1, 380] months (-1 is a missing value).	Numerical	Month	2
5	`current_address_`	Months in currently registered address of the	Numerical	Month	100

Num	Variable	Definition	Data Type	Unit	Example
	months_count	applicant. Ranges between [−1, 406] months (-1 is a missing value).			
6	customer_age	Applicant's age in bins per decade (e.g, 20-29 is represented as 20).	Numerical	N/A	30
7	days_since_request	Number of days passed since application was done. Ranges between [0, 78] days.	Numerical	Day	12
8	intended_balcon_amount	Initial transferred amount for application. Ranges between [−1, 108].	Numerical	USD	100
9	payment_type	Credit payment plan type. 5 possible (annonymized) values.	Categorical	N/A	AD
10	zip_count_4w	Number of applications within same zip code in last 4 weeks. Ranges between [1, 5767].	Numerical	App	21
11	velocity_6h	Velocity of total applications made in last 6 hours i.e., average number of applications per hour in the last 6 hours. Ranges between [−211, 24763].	Numerical	App	12
12	velocity_24h	Velocity of total applications made in last 24 hours i.e., average number of applications per hour in the last 24 hours. Ranges between [1329, 9527].	Numerical	App	1400
13	velocity_4w	Velocity of total applications made in last 4 weeks, i.e., average number of applications per hour in the last 4	Numerical	App	2779

Num	Variable	Definition	Data Type	Unit	Example
		weeks. Ranges between [2779, 7043].			
14	bank_branch_count_8w	Number of total applications in the selected bank branch in last 8 weeks. Ranges between [0, 2521].	Numerical	App	12
15	date_of_birth_distinct_emails_4w	Number of emails for applicants with same date of birth in last 4 weeks. Ranges between [0, 42].	Numerical	Emails	12
16	employment_status	Employment status of the applicant. 7 possible (annonymized) values.	Categorical	N/A	CA
17	credit_risk_score	Internal score of application risk. Ranges between [−176, 387].	Numerical	N/A	-100
18	email_is_free	Domain of application email (either free or paid).	Numerical	N/A	1
19	housing_status	Current residential status for applicant. 7 possible (annonymized) values.	Categorical	N/A	BC
20	phone_home_valid	Validity of provided home phone.	Numerical	N/A	1
21	phone_mobile_valid	Validity of provided mobile phone.	Numerical	N/A	1
22	bank_months_count	How old is previous account (if held) in months. Ranges between [−1, 31] months (-1 is a missing value).	Numerical	Month	1
23	has_other_cards	If applicant has other cards from the same banking company.	Numerical	N/A	1

Num	Variable	Definition	Data Type	Unit	Example
24	proposed_credit_limit	Applicant's proposed credit limit. Ranges between [200, 2000].	Numerical	USD	200
25	foreign_request	If origin country of request is different from bank's country.	Numerical	N/A	
26	source	Online source of application. Either browser(INTERNET) or mobile app (APP).	Categorical	N/A	Internet
27	session_length_in_minutes	Length of user session in banking website in minutes. Ranges between [−1, 107] minutes	Numerical	Minutes	12
28	device_os	Operative system of device that made request. Possible values are: Windows, Macintox, Linux, X11, or other.	Categorical	N/A	Windows
29	keep_alive_session	User option on session logout.	Numerical	N/A	1
30	device_distinct_emails_8w	Number of distinct emails in banking website from the used device in last 8 weeks. Ranges between [0, 3].	Numerical	Emails	2
31	device_fraud_count	Number of fraudulent applications with used device. Ranges between [0, 1].	Numerical	N/A	0
32	month	Month where the application was made. Ranges between [0, 7].	Numerical	Month	2

3.2 Data Quality Assessment

It is paramount that the raw data is fit-for-purpose for fraud detection data analytics and modelling following the "garbage in, garbage out" principle. This solution begins with a robust exploratory data analysis (EDA) aiming at assessing data quality and preparing the data for meaningful analysis.

3.3 Feature Engineering Rationale

4. Model Development and Methodology

4.1 Algorithm Selection

In compliance with Model Risk Management regulations and meeting the challenge of fraud detection in the banking industry, this solution applies Logistic Regression and XGBoost models to predict bank account fraud, prioritizing both explainability and accuracy.

- **Logistic Regression**: Logistic regression is a regression method that uses a supervised learning technique to predict a binary response variable based on a set of independent variables.

  The available data set has a binary, labelled response variable `fraud_bool`, representing Fraud as 1 and Not Fraud as 0 and offers 31 numeric and categorical variables available for feature engineering. This makes logistic regression an appropriate model for predicting bank account fraud using this data set.

  Logistic Regression:

$$logit(p) = log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 X_1$$

4.2 Model Assumptions and Limitations

4.3 Training and Testing Split

5. Validation and Testing

5.1 Performance Metrics

5.2 Sensitivity Analysis

5.3 Back-testing Results

6. Implementation and Monitoring

6.1 Deployment Environment

6.2 Ongoing Monitoring Plan

6.3 Change Management Process