**Y:\BSA_OFFICE_SECURED\Policy and Procedures\Templates\Logo\Screen - EWB English Logo - small.png**

**LexisNexis Fraud Intelligence (LNFI)**

**Risk & Operations**

**(Anti-Money Launder Group)**

**December/2024**

**Model Documentation Change Log**

|  |  |  |  |
| --- | --- | --- | --- |
| **Author** | **Reviewer** | **Date** | **Details of the Changes** |
| EWB |  | 12/3/2024 | Initial draft created |
| EWB |  | 12/9/2024 | Revision on the draft deliverable |
| EWB |  | 12/13/2024 | Revision on the draft deliverable |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

Table of Contents

[1. EXECUTIVE SUMMARY 5](#_Toc163230473)

[1.1 Objective and Background 5](#_Toc163230475)

[1.2 Model Purpose & Use 6](#_Toc163230476)

[1.2.1 Model Purpose 6](#_Toc163230480)

[1.2.2 Portfolio/Product/Transactions Overview 6](#_Toc163230481)

[1.2.3 Applicable Policies and Regulations 7](#_Toc163230482)

[1.2.4 Existing Models 7](#_Toc163230483)

[1.2.5 Upstream/Downstream Model Dependencies 8](#_Toc163230484)

[1.2.6 Process Flow Diagram 8](#_Toc163230485)

[1.3 Model Key Stakeholders, Change Management, & Outstanding Issues 8](#_Toc163230486)

[2. INPUT DATA INTEGRITY & APPROPRIATENESS 9](#_Toc163230487)

[2.1 MODEL DEVELOPMENT DATA 9](#_Toc163230488)

[2.1.1 Overview of Model Development Data 10](#_Toc163230489)

[2.1.2 Development Data Sources, Extraction Process, and Reconciliation 10](#_Toc163230490)

[2.1.3 Development Data Preparation 12](#_Toc163230495)

[2.1.4 Data Limitations 14](#_Toc163230497)

[2.1.5 Data Preparation Software / Platform 14](#_Toc163230498)

[2.1.6 Data Retention 14](#_Toc163230499)

[3. CONCEPTUAL SOUNDNESS 15](#_Toc163230500)

[3.1 MODEL THEORY AND ASSUMPTIONS 15](#_Toc163230501)

[3.1.1 Model Theory and Methodology 15](#_Toc163230502)

[3.1.2 Segmentation Approach 17](#_Toc163230503)

[3.1.3 Model Settings 17](#_Toc163230504)

[3.1.4 Model Assumptions 17](#_Toc163230505)

[3.1.5 Model Limitations and Weaknesses 18](#_Toc163230506)

[3.2 MODEL ESTIMATION / TRAINING AND SELECTION 20](#_Toc163230507)

[3.2.1 Estimation Methodology and Assumptions 20](#_Toc163230508)

[3.2.2 Modeling Software / Platform 20](#_Toc163230509)

[3.2.3 Hyper-parameter Tuning 21](#_Toc163230510)

[3.2.4 Feature / Variable Selection 21](#_Toc163230511)

[3.2.5 Model Estimation / Training Results 21](#_Toc163230512)

[3.2.6 Other Types of Model Estimation 23](#_Toc163230513)

[3.3 Model Development Testing 24](#_Toc163230514)

[3.3.1 Statistical and Technical Assumptions Testing 24](#_Toc163230515)

[3.3.2 Model Performance / Fit Testing 24](#_Toc163230516)

[3.3.3 Model Stability and Overfitting Testing 25](#_Toc163230523)

[3.3.4 Back-testing 26](#_Toc163230524)

[3.3.5 Model Explainability Testing 28](#_Toc163230525)

[3.3.6 Benchmarking 28](#_Toc163230526)

[3.3.7 Sensitivity Analysis 28](#_Toc163230527)

[3.3.8 Stress Testing / Scenario Analysis 29](#_Toc163230528)

[3.3.9 Other Testing 29](#_Toc163230529)

[3.3.10 Overall Performance Assessment 30](#_Toc163230530)

[3.3.11 Need for Model Overlays 30](#_Toc163230531)

[4. PRODUCTION PROCESS COMPLETENESS & ACCURACY 31](#_Toc163230532)

[4.1 Production Application Testing 31](#_Toc163230534)

[4.1.1 System Testing Approach and Results 31](#_Toc163230535)

[4.1.2 User Acceptance Testing Approach and Results 32](#_Toc163230536)

[4.2 Model Production Specifications 32](#_Toc163230537)

[4.2.1 Model Platform 32](#_Toc163230538)

[4.2.2 Data and Process Flow Diagram 32](#_Toc163230539)

[4.2.3 Input Data Specifications 33](#_Toc163230540)

[4.2.4 Model Formulas / Algorithms 33](#_Toc163230541)

[4.2.5 Model Parameters and Settings Values 33](#_Toc163230542)

[4.2.6 Model Outputs 33](#_Toc163230543)

[4.2.7 Reports 34](#_Toc163230544)

[4.3 Operational Controls 34](#_Toc163230545)

[4.3.1 Model Access and Security 34](#_Toc163230546)

[4.3.2 Production Deployment 34](#_Toc163230547)

[4.3.3 Model Usage Controls 35](#_Toc163230548)

[4.3.4 Model Backup 35](#_Toc163230549)

[4.4 Contingency Plans 35](#_Toc163230550)

[4.4.1 Disaster Recovery Plan 35](#_Toc163230551)

[4.4.2 Business Continuity Plan 35](#_Toc163230552)

[4.5 Operating Procedures / User’s Guide 36](#_Toc163230553)

[5. ONGOING MODEL GOVERNANCE & OUTCOME ANALYSIS 37](#_Toc163230554)

[5.1 Ongoing Risk & Performance Monitoring Plan 37](#_Toc163230555)

[5.2 Model Approval and Change Management Process 40](#_Toc163230556)

[5.2.1 Model Approval Process 40](#_Toc163230557)

[5.2.2 Model Change Log 40](#_Toc163230558)

[6. APPENDICES 41](#_Toc163230559)

[6.1 Appendix A 41](#_Toc163230560)

[6.2 Appendix B 41](#_Toc163230561)

# EXECUTIVE SUMMARY



## Objective and Background

Please provide a high-level description of:

* The model’s business objectives.
* Business background including history where appropriate.
* Related regulatory requirements that relate to the business objectives.
* Any other information you see appropriate.

Model Owner:

The LexisNexis Fraud Intelligence (LNFI) Bankcard model was implemented in February 2022 and is currently in production use to assist in identifying bad actors during digital bank onboarding application process.

The model calculates a fraud risk score and warning codes for an application based on the information that is provided in the score request. The data elements of the applicant's personally identifiable information (PII) that are presented in the wireless application are used to calculate a wide variety of predictive variables based on recent and historic transactions, confirmed frauds, and third-party data sources.

The model leverages these variables with proprietary LexisNexis® Risk Solutions algorithms to generate the application risk score. The model is not a credit score and is not intended, suitable, or permitted to be used for credit decisioning, evaluating the creditworthiness of an applicant, or as a basis for any adverse action.

The Fraud Intelligence model is implemented at the bank as a control to detect higher risk Social Security Number (SSN) prospective customers during onboarding through digital bank channels: web (online) and mobile (VELO app).

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



|  |  |
| --- | --- |
| **Model Name** | * *Please provide the official model-name that is used by the model owners and the MRM Group (mutually agreed).*   Model Owner: LexisNexis Fraud Intelligence (LNFI) – Model Id:048 |
| **Primary Model Owner Entity** | * *Please provide the model owner business entity name, e.g., US, China, or Hong Kong.*   Model Owner: US |
| **Primary Model Owner Group** | * *Please provide the model owner business group name.*   Model Owner: Enterprise Risk Management |
| **Model Owner** | * *Please provide the model owner names.*   Model Owner: FVP – ERM Data Manager |
| **Model Developer** | * *Please provide the model developer names (vendor name if vendor model).*   Model Owner: LexisNexis Risk Solutions (<https://risk.lexisnexis.com>) |
| **Model Production Process** | * *Please provide the model production process environment. A high-level description is encouraged.*   Model Owner: Fraud Intelligence is used in the Digital bank production onboarding flow for higher risk SSN customers that are potentially bad actors during the application process. Input data includes Personal Identifiable Information (PII) – i.e., first name, last name, Date of Birth (DOB), SSN, email. This is done through Application Program Interface (API) calls. When LexisNexis receives the PII from an applicant, they provide the LNFI score. |
| **Model User** | * *Please provide all model usernames along with business group names.*   Model Owner:Anti-Money Launder Group |
| **Portfolios the Model Applies to** | * *Please provide high level portfolio size and description that the model is applied to.*   Model Owner: SSN prospective customers being onboarded through digital channels. |
| **Model Objective** | * *Please list all model objectives at a high level.*   Model Owner: Used to detect higher risk SSN prospective customers during the onboarding through digital channels: web (online) and mobile (VELO app). |



## Model Purpose & Use



### Model Purpose

For each business purpose, discuss the following in detail:

* The overall business purposes.
* The specific role that the model output plays in business use (for example, if the model output is used as a secondary source of information in the decision-making process, this should be detailed here).
* The specific products/portfolios/customers/transactions for which the model is suitable (e.g., types of retail mortgages, types of derivatives, types of consumer transactions, etc.)
* Any restrictions on model use, for example, excluded product types within product categories or transaction size limits.

Model Owner:

The LNFI model is designed to provide predictive insights about the identity fraud risk that is associated with new applications for products or services.

The model calculates a fraud risk score (application risk score) and warning codes for an application based on the information that is provided in the score request. The data elements of the applicant's PII that are presented in the wireless application are used to calculate a wide variety of predictive variables based on recent and historic transactions, confirmed frauds, and third-party data sources.

The model does not generate a credit score and is not intended, suitable, or permitted to be used for credit decisioning, evaluating the creditworthiness of an applicant, or as a basis for any adverse action.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### Portfolio/Product/Transactions Overview

* Provide the current size of the portfolio of assets or liabilities (if applicable) and describe the history of the portfolio characteristics (e.g., the inception time frame, any notable idiosyncratic events such as mergers/acquisitions or asset sales, any notable management strategic changes, etc.)
* If the model is being applied to analyze transactions or events (e.g., debit card transactions analyzed for money laundering risk, or cyber-attacks on the Company’s infrastructure), provide the historical volumes of transactions and trends.
* Describe any specific product/customer/transaction types that are being proxied by other product types (e.g., a new product for which the model developed on a more seasoned product is applied).
* When applicable, please describe which portion of the portfolios/transactions/products that is supposed to be covered by the model (for the same business objective) but is decided to be excluded. For such portion, what business strategies are applied to ensure the same business objective is met (e.g., for BSA/AML purpose, certain transactions are monitored manually instead of using the BSA/AML model).

Model Owner:

As of December 31, 2022, digital onboarding aggregate account balances $13.08 million.

The LNFI model is applied to the SSN of prospective customers being onboarded through digital channels: web (online) and mobile (VELO app). As of 2022, only customers accounts onboarding individually. No commercial at this time. Only checking account, CDs, Prepaid card, etc.

### Applicable Policies and Regulations

* List and discuss all regulatory, accounting, legal, and/or compliance rules that are relevant to the model data, design, or use (if any).
* List and discuss all applicable internal policies relevant to the model design and use, if any.

*Note: Please provide document name including suffix.*

Model Owner:

The MRM Model vs. Non-Model Assessment form has been updated to align with the definition of a model per SR 11-7, clarifying the criteria for determining model classification and enhancing model risk governance.

**Definition of a Model per SR 11-7:**

“… the term model refers to a quantitative method, system, or approach that applies statistical, economic, financial, or mathematical theories, techniques, and assumptions to process input data into quantitative estimates.”

“A model consists of three components:

• an information input component, which delivers assumptions and data to the model;

• a processing component, which transforms inputs into estimates; and

• a reporting component, which translates the estimates into useful business information.”

Models meeting this definition might be used for analyzing business strategies, informing business decisions, identifying and measuring risks, valuing exposures, instruments or positions, conducting stress testing, assessing adequacy of capital, managing client assets, measuring compliance with internal limits, maintaining the formal control apparatus of the bank, or meeting financial or regulatory reporting requirements and issuing public disclosures. The definition of model also covers quantitative approaches whose inputs are partially or wholly qualitative or based on expert judgment, provided that the output is quantitative in nature.”

**Permitted Purpose of Data Usage**

The Fraud Intelligence model may not be used as a factor in determining eligibility for credit, insurance, employment, or another eligibility purpose that would qualify the information as a consumer report under the FCRA (Fair Credit Reporting Act).

Federal law, in conjunction with your user agreement with LexisNexis Risk Solutions, requires you to have a permissible use to view regulated personal information. The applicable laws governing model uses are the DPPA (Driver’s Privacy Protection Act) and related state laws, and the GLBA (Gramm-Leach-Bliley Act). If you do not have permissible use, then you will not be given access to the personnel information. The GLBA does not prohibit the disclosure of nonpublic personal information to protect against or preventing actual or potential fraud, unauthorized transactions, claims, or other liability; for required institutional risk control, or for resolving customer disputes or inquiries; to people holding a legal or beneficial interest relating to the consumer; or to people acting in a fiduciary or representative capacity on behalf of the consumer (see <https://www.sec.gov/about/laws/glba.pdf>).

**For more details kindly refer to** “FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf “**&** “MRM-CONTROL01 - y&n Model Assmt 048 - Albert YesM - LexisNexis LNFI.docx”.

### Existing Models

* If this model is replacing existing model(s), provide details of the existing model(s) and the rationale for the replacement.
* Discuss whether the existing model(s) will be retired once this model goes into production.

Model Owner:

The LNFI model continues to operate with the same algorithms, methodologies, and data sources since February 2022.

### Upstream/Downstream Model Dependencies

* Provide a listing and description of upstream and/or downstream models or other key systems (e.g., the Empyrean ALM model).
* Discuss the impact of known limitations of upstream models on this model.

Model Owner:

The LNFI model has an upstream dependency on the LexisNexis InstantID CVI (Comprehensive Verification Index) score. Specifically, the dependency operates as follows:

1. **InstantID CVI check**: Before LNFI is triggered, the CVI score is evaluated.

* If the CVI score equals 0 or 10, the application is auto-declined, and no further LNFI processing is performed.

1. **High Risk Indicators (HRIs)**: If HRIs are identified during the application review, LNFI is triggered directly, bypassing the need for CVI processing.
2. **Standard Workflow**: In cases where no HRIs are present, the CVI score is first generated. After obtaining the CVI, the LNFI model is triggered to perform its fraud risk evaluation.

This workflow establishes LNFI’s critical reliance on InstantID CVI for its initial decision-making process, making the CVI a key upstream dependency for LNFI operations. This ensures a streamlined and prioritized approach to fraud detection based on predefined thresholds and indicators.

### Process Flow Diagram

* Provide a process flow diagram showing how the model is used by the functional / business area. Include upstream and downstream models and systems listed in Section 1.2.5 Upstream/Downstream Model Dependencies.

 Model Owner:

The LNFI model has an upstream dependency on the LexisNexis InstantID CVI (Comprehensive Verification Index) score.

**InstantID CVI check**: Before LNFI is triggered, the CVI score is evaluated.

If the CVI score equals 0 or 10, the application is auto-declined, and no further LNFI processing is performed.

**High Risk Indicators (HRIs)**: If HRIs are identified during the application review, LNFI is triggered directly, bypassing the need for CVI processing.

**High Risk Flow (Regardless of the CVI)** If the applicant meets one or more of the below criteria:

* Introduce Risk based selfie and OCR (HRI flow) for applicants who hit the Risk Indicator (25)
* Introduce Risk based selfie and OCR for applicants who hit the Risk Indicator (27)
* Introduce Risk based selfie and OCR for applicants who hit the Risk Indicator (ER)
* Introduce Risk based selfie and OCR for applicants who hit the Risk Indicator (PR)

**\*Risk Indicators**

|  |  |
| --- | --- |
| **Risk Indicators** | **Details** |
| **25** | Unable to verify address |
| **27** | Unable to verify phone number |
| **74** | The input phone number is associated with a different name and address |
| **82** | The input name and address return a different phone number |
| **ER** | The input email address appears as high risk in the Digital Identity Network |
| **PR** | The input phone number appears as high risk in the Digital Identity Network |

**Standard flow**: In cases where no HRIs are present, the CVI score is first generated. After obtaining the CVI, the LNFI model is triggered to perform its fraud risk evaluation.

Input requirements are Personal Identifiable Information (PII) (First name, last name, DOB, SSN, and email). This is done through API calls. When LexisNexis receives the PII from an applicant, they provide the LNFI score. If the LNFI score is greater than 660, the applicant will be pushed to BSA’s manual queue for further review.

**Auto Decline (Regardless of the CVI)**

* Auto – decline applicants who hit Risk Indicators (**25** OR **27** OR **ER** OR **PR**) AND Fraud Score >= 800
* Auto – decline applicants who hit Risk Indicator (**25** OR **27** OR **ER** OR **PR**) AND if the **ID Authentication** results in a ‘**No Match**’ with **IDA Alerts** returning (**"Result": 2, "TamperResult": 2**)
* **Selfie Related -** Auto decline applicants when they fall into the “**Selfie: No match”** categories

In case of Selfie or Liveness technical failures, the applicant will be pushed to BSA Manual Queue.

## Model Key Stakeholders, Change Management, & Outstanding Issues

Describe, at minimum, the following:

1. Model output key stakeholders, review committee(s).
2. High level summary of model changes in recent time or since last model validation.
3. High level summary of the latest model related business area audit and regulatory exam results including any outstanding findings, regulatory Matter Requiring Attention (MRAs), and management self-identified issues.

Please ensure that **all of** the points mentioned above are addressed.

Model Owner:

1. The key roles and responsibilities as they pertain to the LNFI model as follows:

**Model Owner**

* Enterprise Risk Management (ERM): Coordinate with Anti-Money Launder Group and Fraud Strategy Team and on model framework, include performance monitoring and reporting and corrective measures based on MRM findings.

**Model Developer (Vendor)**

* LexisNexis Risk Solutions: Responsible for providing software enhancements, aiding in implementing system upgrade and providing support via email when requested by the Bank.

**Model User (Anti-Money Launder Group)**

* Anti-Money Launder Group: Leverages LNFI to identify fraud risk during digital onboarding review.

**Model Support User (Digital Bank Product Team)**

* Digital Bank Product Team: Support the model user on general product questions and working with vendor related to product.

2. ***Model changes***: The model continues to operate with the same algorithms, methodologies, and data sources via API.

3. ***High level summary of Audit and Regulatory Exam Results***

* *Model Inherent Risk Rating (IRR)*: The preliminary model inherent risk rating (Model-IRR) is assigned as LOW, based on a materiality threshold of less than 5%.
* *Model Owner’s Assessment:* The model owner’s assessment of the preliminary Model-IRR remains unchanged in the expert and management overlay areas, supported by the MRM group’s LOW rating, as the model serves as a secondary data source during SSN onboarding reviews.

**For more details kindly refer to** “MRM-CONTROL01 - y&n Model Assmt 048 - Albert YesM - LexisNexis LNFI.docx” & “MRM-CONTROL02 - Model-IRR Assmt 048 -L- Albert LNFI\_FINAL.docx“.



# INPUT DATA INTEGRITY & APPROPRIATENESS

## MODEL DEVELOPMENT DATA

Model Development Data refers to the data used in the research & development process to determine the model specifications. That is, the process for determining the exact mathematical formulas, algorithms, inputs, parameters, and assumptions that comprise a model.

Note: This documentation section is not applicable for those models whose structure is not determined through empirical data analysis. This includes, for example, some market risk / trading models where the model structure is based on financial theory (e.g., Black-Scholes options pricing model) or qualitative models whose structure and parameters were determined judgmentally.

**Reference Document List**

Please list all the documents referred to in this section.

|  |  |  |
| --- | --- | --- |
| **#** | **Reference Document Name** | **High Level Description and purpose of the Document** |
| 1 | FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf | It is the model reference guide. |
| 2 | MRM-CONTROL01 - y&n Model Assmt 048 - Albert YesM - LexisNexis LNFI.docx | Enterprise risk management and Model risk classification procedures. |
| 3 | MRM-CONTROL02 - Model-IRR Assmt 048 -L- Albert LNFI\_FINAL.docx | Model inherent risk rating assessment form. |
| 4 | 2023\_LNRS\_BCOverview\_Tech\_Resilience\_IT.pdf | It is the overview of business continuity technical resilience - IT |
| 5 | LexisNexis\_Business\_Continuity\_Disaster\_Recovery\_  Assessment\_Dec\_18\_2023\_13\_23 (1).pdf | It is the LexisNexis Business Continuity/Disaster Recovery Assessment Report. |

**Data Assumptions Summary**

Please list out data assumptions applied in the model development and model production process, such as missing value treatment, outlier treatment, etc.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **#** | **Assumption Name** | **Assumption Description** | **Materiality of Assumption** | **Rationales for this Assumption**  (Business driven or quantitative methodology driven) |
| 1 | **Feature Scaling** | Feature scaling was not applied because XGBoost does not require normalized or standardized input features. | Low  Impact: Not applying feature scaling has minimal impact, as XGBoost doesn’t require it for effective performance. | Quantitative methodology-driven: XGBoost handles unscaled features natively. |
| 2 | **Missing Value Treatment** | Missing values were not imputed with conventional values like mean or median but instead treated as discrete values. | Medium  Impact: It can cause inaccuracies in reporting and forecasts, potentially skewing insights. | Quantitative methodology-driven: ensures completeness of data for model training. |
| 3 | **Variable Selection** | Recursive Feature Elimination (RFE) was used to select features contributing most to the model’s predictive power while eliminating redundant or irrelevant variables. | High  Impact: Incorrect feature selection could significantly impact model performance and business decisions. | Quantitative methodology-driven: improves model performance and interpretability. |
| 4 | **Fraud Label Definition** | The target variable was based on flagged applications for fraud (fraud and non-fraud) sourced from historical data on account performance and fraud tagging processes. | High Impact: Incorrect labeling could lead to significant misclassifications, affecting the accuracy of model. | Business-driven: ensures alignment with business use case and objective. |

**Data Limitation Summary**

Please list out data limitations, their impact of business use, and ongoing monitoring program to appropriately manage the related risk.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **#** | **Limitation Name** | **Limitation Description** | **Impact on Business Use** | **Monitoring Description & Frequency** |
| 1 | **Temporal Data Gaps** | Training data was collected from specific historical periods and may not capture shifts in fraud patterns over time. | Potential decrease in predictive accuracy if fraud patterns change significantly in future periods. | Monitor the model's AUC and FDR quarterly on out-of-time (OOT) datasets to ensure consistent performance. |
| 2 |  |  |  |  |

### Overview of Model Development Data

Provide descriptive characteristics of the model development data, for example, coverage of products / portfolios / transactions, time periods, geographic distribution, etc.

The sources and flows of all the data leveraged in model development should be illustrated with a data flow diagram. The diagram should show each stage of the data preparation process from the initial data pull to the final datasets used for model development and testing including data quality assurance controls.

Model Owner:

The data used to build the model includes applications for products and services and the eventual status of each application and account regarding performance (for example, “fraud” or “not fraud”). Random samples of applications were used to select training, testing, and out-of-time validation samples.

The data set used for model development consists of 68,687,312 records spanning the time period from January 2017 to December 2019. This portfolio reflects a comprehensive representation of digital onboarding activity and provides a robust foundation for evaluating fraud risk in customer acquisition processes.

Data compiling a list of fraud point attributes, which was augmented with additional attributes from the ID Network, ensures the model had access to comprehensive data from various sources such as public records, credit bureau data, phone directories, utility data, and student directories. Specific model variables were not shared due to their proprietary nature.

From EWB’s perspective, Input data includes PII – i.e., first name, last name, DOB, SSN, email. This is done through API calls. When LexisNexis receives the PII from an applicant, they provide the LNFI score.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### Development Data Sources, Extraction Process, and Reconciliation



#### Data Sources

Identify the sources of the model development data, for example, internal data from specific corporate data warehouse tables, desktop databases, text files, or external data from third-party vendors or websites. Development data may also include the output of other upstream models or computational tools.

If both internal and external data are used in the model development, you may want to create subsections covering them separately.

Model Owner:

All the following sources that are available for consideration are used in the model:

**Tri-Credit Bureau Identity Activity**

Identity records from three national credit bureaus provide a unique perspective on identity history.

**LexisNexis Risk Solutions Customer Network**

Visibility to inquiry events provides insight on real and fraudulent identity activity.

**Online, Utility, Phone, and Other Behavioral Activity**

Frequent updates provide insight on identity events that are related to address activity, phone usage,

online activity, and email activity.

**Local, State, and Federal Government Records**

Government records provide reliable identity data that is difficult to compromise, including the

following information:

• Assigned group of SSN values

• Records of reported deceased persons by name, SSN, and DOB

• Public records of interaction with government agencies

**LexisNexis® Inquiry Identity Network**

The Inquiry Identity Network is a proprietary, cross-industry network of U.S. identity information that contains more than one trillion aggregated identity elements, more than two billion historic consumer transactions, and more than eight million reported identity fraud attempts.

The Inquiry Identity Network contains the PII of those individuals for whom transactions were submitted by clients (for example, applicants for credit card products or wireless phone service contracts). PII typically includes name, SSN, address, phone number, DOB, IP address, email address, and date of the transaction (for example, application date).

The Inquiry Identity Network helps to provide a unique cross-industry view of U.S. consumer application activity to enhance physical identity insights and fraud solutions.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



#### Data Relevance

Discuss the relevance of the development data to the modeling objective. For example, is the composition of the development data representative of the current portfolio in terms of coverage and distribution of data attributes? Is the time period selected for development data appropriate for the model’s business purpose and the statistical estimation technique?

If proxies for internal data are used, such as internal data for other products or external data from public databases or third-party services, justify and document the applicability and appropriateness of the proxy data to the specific internal portfolio / purpose.

For vendor models, document a comprehensive assessment of the vendor’s development data applicability to the Company’s internal portfolio/products/customers. This typically involves a comparison of the external and internal data for key model drivers (e.g., geographic distribution, loan/transaction size, loan/product type, etc.).

Model Owner:

The input data and assumptions are known with certainty, in context of which real-world inputs are used. The data that is used in the model is sourced nationally and is subject to monitoring and data hygiene management to ensure high levels of reliability and stability over time.

Furthermore, the vendor stated their model framework involved” … panel review process is designed to ensure model soundness, predictive power, and acknowledgement of fair lending requirements… and all LexisNexis Risk Solutions models are reviewed by an independent compliance department.”

Multiple samples were used to develop the model, the Source used is Inquiry Identity Network, the Population contains new credit card applications, and the Application dates range from January 2017 to December 2019.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



#### Data Extraction Process

Describe how the development data is extracted, either automatically or manually, or otherwise obtained.

Include references to the code or files used to extract the data or to the data files received from other individuals / departments.

Model Owner:

The data extraction process for the LexisNexis Fraud Intelligence model involved compiling a list of fraud point attributes, which was augmented with additional attributes from the ID Network. This approach ensured that the model had access to comprehensive data from various sources. The data was sourced from public records, credit bureau data, phone directories, utility data, and student directories. These diverse sources provided a robust set of attributes essential for fraud detection. To make this data suitable for model integration, it was meticulously organized into a structured, model-ready attribute set. This process allowed for the efficient extraction and preparation of data, ensuring that all relevant information was captured and could be used to assess fraud risk effectively.

This method of data extraction relied on a combination of public and proprietary data sources, creating a well-rounded dataset that incorporates multiple dimensions of identity verification and transaction history. By leveraging these different data points, LexisNexis ensured the quality and comprehensiveness of the data used in its fraud intelligence model, enabling it to detect patterns indicative of fraudulent activity.

The LexisNexis Fraud Intelligence model relies on data managed within the proprietary internal HPCC Systems Data Platform. The extraction process is designed to ensure efficient, accurate, and secure retrieval of relevant data for model input.

#### Data Reconciliation

Demonstrate that the development data has been reconciled with a source system (e.g., the general ledger) or line of business report, or alternatively, explain how the extracted data was determined to be complete and accurate.

In addition, provide a step-by-step waterfall of data counts and balances at every step in the data preparation process from the raw data extract to the final modeling dataset.

Model Owner:

For the LexisNexis Fraud Intelligence Model, data reconciliation process focused on ensuring that all records from diverse categories of data sources were properly aligned and formatted into suitable sets for modeling to produce output.

One of the following sets of elements must be provided:

First Name, Last Name, StreetAddress1, and Zip

First Name, Last Name, StreetAddress1, City and State

First Name, Last Name, and SSN

First Name, LastName, and DOB

First Name, Last Name, and Primary Phone

While the exact steps for data reconciliation were not explicitly detailed in the available materials, it is understood that during the data preparation process, raw data from different sources were transformed into a structured, model-ready format. The data sources included public records, credit bureau data, phone directories, utility data, and student directories, which were organized into a consistent set of attributes for use in fraud detection.

Regardless, the overall aim was to ensure that the data used in the fraud intelligence model was standardized, consistent, and ready for use, ensuring its quality and reliability for model’s predictions. This process is critical in maintaining the integrity of that data and optimizing it for the detection of fraud risks.

### Development Data Preparation



#### Data Quality and Treatments

Describe the raw data quality and any treatments used to address missing or erroneous values, for example, algorithms applied to impute values.

Document any analysis of data outliers and their impact on model development / outputs. Provide support for the selected approach for treating the outliers (if any).

Model Owner:

For the LexisNexis Fraud Intelligence Model, the data was carefully formatted and organized before it was used in the modeling process. Missing values were not imputed with conventional values like mean or median but instead treated as discrete values. This means that missing data were given specific codes to represent their absence, rather than being ignored or removed. For example, in the case of missing identity report times, instead of leaving the field blank, a value like “-1” was assigned, indicating that the time had never been reported for that identity. This approach allowed the model to handle missing values explicitly, preserving the integrity of the dataset without replacing the missing information with potentially misleading values.

This process ensured that the data used in the fraud detection model remained consistent and structured, with each piece of information being handled appropriately to maintain the accuracy of the model’s predictions.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



#### Data Filtering and Exclusions

Provide a detailed description of, and justification for, data filtering and significant data exclusions that may potentially introduce model bias. Where a significant number of records is excluded due to data quality or other reasons, to the extent possible, analysis should be performed and documented showing the impact of the filtering rule.

A complete waterfall from the point of raw data extract to the final development/testing data showing the impacts of each exclusion (in terms of the number of records and other key metrics) should be provided.

Model Owner:

The data filtering process for the Fraud Intelligence Model involves carefully selecting relevant and valid records to enhance model accuracy. This ensures that only records containing model-ready attributes are included in the training dataset. Data from multiple sources are processed to ensure that the records align with the model’s requirements and can be used for predictive analysis.

As everything is managed within **LexisNexis’s proprietary internal HPCC data platform**, there are no external filtering or exclusions applied. All filtering processes are handled internally within the platform, ensuring data consistency and integrity throughout the model’s development and operation. This internal management helps maintain the quality of data used in the fraud detection model.

#### Data Sampling

Provide details of statistical sampling, if any, performed to create the model development and testing datasets.

Model Owner:

Random samples of applications were used to select training, testing, and out-of-time validation datasets. The model is a collection of different clients, with each client down sampled by a different amount to approximately equalize the bad rate between clients.

The vendor stated a sampling practice that considered the following dataset:

**Training Sample**

Sample data used to fit the model

**Validation Sample**

Sample data used to provide an unbiased evaluation of a model fit on the training dataset while tuning model hyper-parameters

In LNFI model, the training and validation time periods overlapped, with the validation sample referred to as the “in-time” validation sample. To ensure no application appeared in both datasets, records were randomly divided between the two.

**Test (Out-of-Time) Sample**

Sample data used to provide an unbiased evaluation of a final model fit on the training dataset.

The test dataset provides the best practice that is used to evaluate the model. The dataset is only used after a model is completely trained.

To make the dataset more balanced and improve model performance and efficiency, non-fraud records were down sampled while all fraudulent records were kept. The amount by which non-fraud records were down sampled varied according to the client so that there were enough records from each client to make a representative sample. Final score scaling was adjusted to reflect actual good/bad odds. This sampling method is an industry-accepted practice to achieve more robust models.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



#### Data Transformations

Provide a description of, and rationale for, operations/calculations on raw data, such as scaling, forming data segments, averaging, or combining data from multiples sources (for example, to calculate charge-off rates) in order to produce model development-ready data.

Describe the composite/derived variables created out of raw data. For example, splines, Weight-of-Evidence transformations of variables, interaction terms, etc. Provide support for the technical soundness and appropriateness of the selected transformations in the context of the specific modeling approach you selected and the overall model purpose.

Specifically:

For models that utilize feature engineering, provide detailed documentation of the engineering process, including a description of the software/package used to perform the feature engineering and a discussion on the limitations of the selected engineering approach.

For models that utilize unstructured data, include detailed description of the data pre-processing of unstructured data. Provide analysis/test/comparison results with related data/scripts/outputs if any to justify the pre-processing performed.

For advanced machine learning models, also include detailed discussion on the sufficiency and appropriateness of data transformations and treatments applied with respect to the ML algorithm used (for example, standardization/normalization is required for KNN but not for Random Forests). Provide analysis/test/comparison results with related data/scripts/outputs if any to support the discussion.

Model Owner:

The Fraud Intelligence Model does not involve explicit transformation techniques like scaling, normalization or feature engineering. However, the process of organizing sources records into model-ready attributes suggests a level of data standardization to ensure consistency and suitability for modeling purposes.

Since no detailed mention of transformations is provided, it can be inferred that the primary focus was on data organization and formatting to meet model requirements, allowing the data to be processed efficiently without additional transformation steps.

#### Variable Definitions

Provide definitions of variables, including alternative transformations of variables tested. For vendor models, describe how the vendor’s definitions for inputs and outputs compare with the Bank’s internal definitions (e.g., delinquency, defaults, accounting losses, etc.).

Reference the location of the comprehensive data dictionary that lists each variable’s description, source, allowable values, and other relevant information.

Response Variable

Describe the response/performance/dependent variable that the model is designed to estimate/project.

Model Owner:

The dependent variable used by the model for training consisted of both pre-book and post-book fraud tags.

The post-book fraud tag is the same as the “confirmed fraud” indicator and corresponds with the way a client would account for a financial loss due to fraud. Fraud that is “suspect” or does not necessarily correspond to a financial loss is considered a pre-book fraud tag.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



Explanatory Variables

Describe the explanatory/independent variable candidates assessed in the model development process.

Model Owner:

Various independent variables were used for the development of the model.

The vendor stated variables that were used for development can be distributed into the following twelve (12) distinct categories:

**Component Characteristics**

The Application Component Characteristic Attributes are designed to provide a high-level and general assessment of the identity fraud risk that is related to the input identity components, including the current and the previous addresses that are linked to the identity.

**Component Correlation**

The Application Component Correlation Attributes are designed to provide insight into the identity fraud risk that is related to the corroboration of the input identity components that are reported together.

**Component Divergence**

The Application Component Divergence Attributes are designed to provide insight into the identity fraud risk that is related to the frequency of the input identity components that are linked to other identities.

**Component Validation**

The Application Component Validation Attributes are designed to provide insight into the identity fraud risk that is related to the validity of the input identity components.

**Component Velocity**

The Application Component Velocity Attributes are designed to provide insight into the identity fraud risk that is related to the search velocity of the input identity components that are seen in the search activity of LexisNexis Risk Solutions products.**Identity Associations**

The Identity Relatives and Associates Attributes are designed to provide insight into the identity fraud risk that is related to the relatives and the associates who are linked to the subject.

**Identity Overview**

The Identity Overview Attributes are designed to provide a high-level assessment of identity fraud risk.

**Identity Source**

The Identity Sources Attributes are designed to provide insight into the depth and breadth of the sources that report the identity.

**Identity Variation**

The Identity Variation Attributes are designed to provide insight into the identity fraud risk that is related to the variation of the identity components that are linked to the identity.

**Identity Velocity**

Identity Velocity Attributes are designed to provide insight into the identity fraud risk that is related to the search velocity of input identity components that are seen in the search activity of LexisNexis Risk Solutions products.

**Identity Verification**

Identity Verification Attributes are designed to provide insight into the verification of the input identity components.

**Confirmed Behavior**

Confirmed Behavior Attributes compare event information against confirmed, historic fraudulent events within the Inquiry Identity Network.

**Information is sourced from “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### Data Limitations

Provide information about known data limitations / weaknesses and an assessment of their impact on the final model’s output. For example, if the model was developed based on external data that differs notably from the Bank’s data, the differences and their potential impact must be documented. For each noted weakness / limitation, describe how the associated risk is currently being mitigated. Additionally, where longer-term remedial actions are being undertaken or planned (e.g., an initiative to clean up the existing data or collect incremental data), such actions should also be documented.

Model Owner:

The primary limitation in the data is the reliance on historically reported fraud events. The LNFI model is restricted to only the fraud cases that have been historically identified and reported, meaning that undetected fraud or non-reported fraud is not incorporated into the dataset. This limitation means that the model does not account for fraudulent activities that have not yet been discovered or reported. The model leverages historical data to predict future fraud, assuming the characteristics of previously identified fraudulent activities will remain relevant for identifying future fraud.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### Data Preparation Software / Platform

Provide information on the software and/or programming language used in the data extraction, transformation, and other steps to prepare the model development and testing data. Provide a reference to the location of the development programming codes, associated log files, and other data preparation artifacts.

Model Owner:

LexisNexis uses its proprietary internal HPCC data platform to manage and process data for the Fraud Intelligence Model. The HPCC platform is designed to handle large-scale data processing, enabling efficient Extraction, Transformation, and Loading (ETL) processes.

As everything is managed within the HPCC platform, this ensures a streamlined and secure environment for data preparation, where various data sources are integrated, cleaned, and structured into model-ready attributes. The platform is optimized for handling high volumes of data, providing the necessary computational power to support advanced analytics and predictive modelling, making it a critical part of the LNFI model’s data preparation process.

No external software or platforms are explicitly mentioned in the context of the data preparation for the LNFI model, as all activities are handled within the internal infrastructure of LexisNexis.

### Data Retention

Describe where the development data is stored (post development) and how the environment is controlled. Provide the minimum time period for data retention.

Model Owner:

Data retention focuses on securely managing data throughout its lifecycle, ensuring compliance with relevant legal and regulatory requirements. LexisNexis defines retention policies that determine how long data should be kept, ensuring compliance with regulations such as GDPR and CCPA. Data that is no longer actively used is archived using HPCC’s scalable architecture, allowing efficient access for audits or future references. Once data exceeds its retention period, it is securely deleted or anonymized to protect privacy. LexisNexis optimizes storage through tiered solutions, ensuring data is retained for the required duration and disposed securely, minimizing storage costs while adhering to compliance standards.

LexisNexis adheres to both the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA). These regulations are designed to protect consumers’ data privacy and ensure transparency in how personal information is handled. LexisNexis, which operates within the legal, corporate, and governmental sectors, provides tools and resources to help organizations comply with these laws. For instance, they offer guidance on CCPA compliance, detailing steps to protect consumer data and manage data access requests [LexisNexis](https://www.lexisnexis.com/community/insights/legal/practical-guidance-journal/b/pa/posts/the-california-consumer-privacy-act-is-in-effect-what-to-do-now), [LexisNexis](https://www.lexisnexis.com/community/insights/legal/b/industry-awareness/posts/what-the-california-consumer-privacy-act-means-to-data-protection-law). ***While the exact data retention period wasn’t specified, it is likely that the data will be kept at least until the necessary compliance checks are completed.*** This ensures proper oversight and management throughout the model development process.

# CONCEPTUAL SOUNDNESS

## MODEL THEORY AND ASSUMPTIONS

**Reference Document List**

Please list all the documents referred to in this section.

|  |  |  |
| --- | --- | --- |
| **#** | **Reference Document Name** | **High Level Description and purpose of the Document** |
| 1 | FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf | It is the model reference guide. |
| 2 | MRM-CONTROL01 - y&n Model Assmt 048 - Albert YesM - LexisNexis LNFI.docx | Enterprise risk management and Model risk classification procedures. |
| 3 | MRM-CONTROL02 - Model-IRR Assmt 048 -L- Albert LNFI\_FINAL.docx | Model inherent risk rating assessment form. |
| 4 | 2023\_LNRS\_BCOverview\_Tech\_Resilience\_IT.pdf | It is the overview of business continuity technical resilience - IT |
| 5 | LexisNexis\_Business\_Continuity\_Disaster\_Recovery\_  Assessment\_Dec\_18\_2023\_13\_23 (1).pdf | It is the LexisNexis Business Continuity/Disaster Recovery Assessment Report. |

**Model Assumption Summary**

Please list out model methodology assumptions applied in the model development and model production process, such as missing value treatment, outlier treatment, etc.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **#** | **Assumption Name** | **Assumption Description** | **Materiality of Assumption** | **Rationales for this Assumption**  (Business driven or quantitative methodology driven) |
| 1 | **Data Stability** | The model assumes that historical patterns and relationships between features and the target variable will remain stable over time. Model performance depends on the consistency of data reporting and consumer behavior patterns. Significant changes in data quality or consumer behavior may impact stability. | High: Extreme shifts in consumer behavior or reporting could lead to degraded model performance. | Quantitative methodology-driven: Ensures alignment with historical trends observed in training data. |
| 2 | **Score Application** | The score ranks individuals based on the probability of the target outcome occurring, not as an exact probability estimate. Model performance depends on the live population matching the development population and the stability of feature-target correlations over time. | High: Variations between the development and live populations can reduce model effectiveness. | Business-driven: Ensures the score maintains its purpose of ranking risk appropriately. |

**Model Limitation Summary**

Please list out model methodology related limitations, their impact of business use, and ongoing monitoring program to appropriately manage the associated risk.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **#** | **Limitation Name** | **Limitation Description** | **Impact on Business Use** | **Monitoring Description & Frequency** |
| 1 | **External Factors** | Model performance relies on external factors such as macroeconomic conditions, customers' business policies, decision-making processes, and portfolio management remain stable over time. Any significant change in these factors could degrade model performance. | The model's predictive power may decrease if external factors diverge significantly from those present during model development. | The vendor stated that as part of data monitoring process, attribute and score monitoring are conducted to evaluate day-to-day changes. In addition, ad-hoc monitoring is conducted on a weekly and monthly time-lag basis. |
| 2 | **Target Population Alignment** | The model was developed using pre-book and post-book U.S. bankcard applicant data sourced from internal LexisNexis samples. Differences between live population characteristics and the development population may impact performance. | If the live population differs significantly, the model may fail to accurately identify fraud risk. | The vendor stated that as part of data monitoring process, attribute and score monitoring are conducted to evaluate day-to-day changes. In addition, ad-hoc monitoring is conducted on a weekly and monthly time-lag basis. |
| 3 | **Fraud Tag Consistency** | The "confirmed fraud" post-book fraud tag used in model development corresponds to financial loss. However, suspect fraud cases not corresponding to financial loss are treated as pre-book fraud tags, leading to potential differences in live use. | Misalignment of fraud tagging between development and live use could impact model predictions and business decisions. | The vendor stated that as part of data monitoring process, attribute and score monitoring are conducted to evaluate day-to-day changes. In addition, ad-hoc monitoring is conducted on a weekly and monthly time-lag basis. |

### Model Theory and Methodology

#### Modeling Approach

Provide a description of the modeling approach you have selected, including the statistical estimation approach or machine learning technique, if applicable (with further details of the model construction/estimation process to be provided in Section 3.2 Model Estimation/Training and Selection).

For advanced Machine Learning (ML) models, discuss briefly whether a self-explanatory or less complex model (e.g., logistic regression, linear regression) is viable in solving the same business problem. If not, explain why not. Detailed information on this topic should be provided in Section 3.1.1.3. Alternative Approaches Explored

Model Owner:

The Fraud Intelligence model is designed to provide predictive insights about the identity fraud risk that

is associated with new applications for products or services.

The model was developed and validated using industry-standard principles and methodologies.

The modeling framework is chosen to balance two primary concerns: interpretability and predictiveness.

The model use case often dictates which algorithms may be employed. In this case, the model is not

used for adverse action and uses boosted ensembles of decision trees.

The boosted tree modeling technique produces a robust non-linear model. The training population is representative of identity fraud behavior as reported by bankcard consortium members, and the model generalizes well to all populations.

Standard analytic modeling techniques were used to build the model. This process included the use of testing and validation datasets that are separate from a training dataset. The training data is sampled for the purpose of training to effectively differentiate the two populations using an empirical comparison of the characteristics present in each population. Test-validation data is typically an out-of-time sample, which is constructed so that the data represents a period in the future of the training data. Test data may also be an out-of-time sample that occurs in the future of the training data. The outcome data from the training data is used to adjust the model to optimize the prediction of fraud risk while minimizing errors. The testing and validation data are used to validate the performance of the model by comparing the prediction made by the model with the known outcome on data that was never exposed to the model training process.

The model is trained on a blended dataset that consists of applications from a variety of members within

the Inquiry Identity Network and historical bankcard applications.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



#### Model Structure/Formulae

*Detail all relevant mathematical equations applied in the model with a clear explanation of the notation*. Describe the model inputs and outputs if not already provided in Section 2.1.3.5. Variable Definitions

Note: This section applies to all models and should be especially detailed for models that were not developed through statistical or machine learning analysis of empirical data (e.g., the market risk / trading models based on financial theory). For these models, the rationale for the particular choice of inputs (e.g., prices, interest rates, volatilities, variance/covariance matrices) should be provided.

Model Owner:

The GBDT (gradient boosted decision trees) algorithm is a stage-wise ensemble learning technique that aims to produce a strong predictor from a successive series of weak learners.

A weak learner is an estimator that produces a prediction better than a random guess. GBDT uses shallow decision trees as weak learners. At each stage of the algorithm, a shallow classification tree is built using the candidate feature set and outcome variable to predict the residuals of the entire prior ensemble. The final output of the model is the sum of all weak learner outputs in the ensemble.

LexisNexis Risk Solutions used regression trees as the base learner and minimized the binomial deviance

using the additive boosting process. The specific implementation used was XGBoost.

**Parameters**

GBDT algorithms have a set of parameters, called hyper-parameters, that control the learning process and must be tested through a process called hyper-parameter tuning.

Hyper-parameters control how the machine learning algorithm behaves. Each shallow decision tree of the GBDT model introduces a split in the model that minimizes the cumulative error of the decision tree ensemble. The hyper-parameters are the scoring coefficients in the model.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



#### Alternative Approaches Explored

Describe how the selected model theory/methodology (and estimation technique, if applicable) compares with industry practices for similar models and provide rigorous support for a selected approach that is non-standard or innovative. Provide references to industry and academic publications supporting the choice of this modeling methodology.

Describe alternative modeling approaches (including alternative estimation/numerical techniques, if applicable) that were considered and why they were not selected. Provide references to industry and academic publications discussing the alternative methodologies.

For machine learning (ML) models, provide performance comparison between the self-explanatory model and the selected ML model and a discussion on the trade-offs between model performance and transparency/interpretability. If a self-explanatory model is viable, also provide analysis/test/comparison results with related data/scripts/outputs if any to support the discussion.

Provide a comparative narrative for the selected ML model vs. other comparable/state-of-the-art methodologies with a discussion on the advantages and disadvantages of the selected ML model vs. the alternatives.

Model Owner:

The primary modeling approach for the LexisNexis Fraud Intelligence Model is Gradient Boosting Decision Trees (XGBoost), a widely recognized machine learning algorithm for its predictive accuracy and efficiency in handling complex, structured data.

Additionally, Greedy Function estimation/approximation is viewed from the perspective of numerical optimization in function space, rather than parameter space. A connection is made between stagewise additive expansions and steepest-descent minimization. A general gradient descent “boosting” paradigm is developed for additive expansions based on any fitting criterion. Specific algorithms are presented for least-squares, least absolute deviation, and Huber-M loss functions for regression, and multiclass logistic likelihood for classification. Special enhancements are derived for the case where the individual additive components are regression trees, and tools for interpreting such “TreeBoost” models are presented. Gradient boosting of regression trees produces competitive, highly robust, interpretable procedures for both regression and classification, especially appropriate for mining less than clean data. Connections between this approach and the boosting methods of Freund and Shapire and Friedman, Hastie and Tibshirani are discussed.

LexisNexis Risk Solutions used regression trees as the base learner and minimized the binomial deviance using the additive boosting process. The specific implementation used was XGBoost.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### Segmentation Approach

Describe and justify the selected model segmentation scheme (or lack thereof), including any related quantitative analyses performed and subject matter expert qualitative considerations. Provide the segmentation waterfall logic, if applicable. Assess the impact of the selected segmentation scheme on the model estimation and output.

If in-model segmentation approach was followed (rather than developing separate equations/model objects for each segment), explain this with the rationale for going the route of in-model segmentation.

Model Owner:

The segmentation approach for the Fraud Intelligence Model relies on historical data to categorize risk into fraudulent or non-fraudulent cases. Using XGBoost algorithm, the model scores each identity match by analyzing various attributes derived from the ID Network and other sources, allowing for data driven classification of potential fraud risks.

The segmentation framework does not explicitly involve hierarchical waterfall logic or pre-defined segmentation rules. Instead, the model dynamically adjusts its scoring through machine learning, ensuring adaptability and precision in identifying fraud risks across different data segments. This approach aligns with industry practices for fraud detection, emphasizing flexibility and efficiency in processing large-scale, diverse datasets.

### Model Settings

If applicable, describe model settings and parameters, including vendor model customizations. For example, a vendor model may offer alternative interest rate term structures for valuation purposes. or a vendor may recommend updated model tuning parameters (e.g., for mortgage prepayment models) to be used in place of default values. For each setting/parameter, justify the selected value relative to the other choices available.

Model Owner:

LexisNexis Risk Solutions uses an odds-doubling methodology for model score calibration, where a score

of 525 corresponds to odds of a bad rate at .0004, with odds doubling every 45 points.

For probability p (0 ≤ p ≤1), which indicates the bad rate in the sampled training data, the score is

calculated from x by the following equation:

Where the weighted probability when the odds are 0.0004, the log term becomes 0 and the score is 525. Log terms calculate how great the odds are compared to the midrate odds 0.0004 and create odds doubling every 45 points.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**

`

### Model Assumptions

List and justify the implicit and explicit assumptions associated with the model, including qualitative or quantitative expert judgments. Assess the impact of each assumption to the extent possible. For example, if a model relies on an average of historical values over the last 6 months, it may be important to test the impact on the model output of selecting alternative assumptions, e.g., 3 months, 9 months, etc.

If any assumptions are intended to be conservative, explain in what way they are conservative.

NOTE: Testing of any technical assumptions underlying the selected statistical/machine learning technique should be documented in Section 3.3.1. Statistical and Technical Assumptions Testing.

Model Owner:

The assumptions that were made in the development of the model are consistent with modeling best practices:

**Data Stability**

Predictive modeling techniques use previously observed behavior to search for patterns in behavior that may indicate the future target outcome.

The model that is produced using this technique is aligned to those past behaviors. Stable model performance depends on all data that is used to inform the score remaining stable over time and no extreme shifts in consumer behavior occurring. Changes in behavior that impact the type and quantity of data reported cannot be controlled. Data instability risk can be mitigated by regularly monitoring the score stability and performance after the model is implemented to check whether entity behavior or data quality has shifted significantly.

**Score Application**

Model performance over time is dependent on the scored population within the live environment being consistent with the population that is used for the model development.

Model performance is also dependent on the observed correlations between model variables and the target variable within the historical data that was used within the model development persisting within live data. The score was built to rank the population based on the estimated probability of the target outcome occurring, not to provide a point estimate of that probability itself.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### Model Limitations and Weaknesses

List any known model limitations and weaknesses. For each weakness / limitation, there should be a description of the associated model risk and, if applicable, the risk mitigant designed to address this risk. See the following example:

| ***Model Weakness or Limitation*** | ***Associated Model Risk(s)*** | ***Model Risk Mitigants / Remediation*** |
| --- | --- | --- |
| The model output is heavily impacted by several judgmental management assumptions, including x, y, and z. These assumptions are currently lacking empirical support. | Use of judgmental assumptions increases the risk of poor model predictions / measurements and unsupported model estimates, which may lead to inappropriate business decisions. | **Short Term Risk Mitigants**:   1. The judgmental assumptions will be subject to oversight by the governance committee X that will review and challenge the model owner's support for the assumptions on a monthly basis. 2. The model output will be benchmarked to the output from the alternative model Y on a quarterly basis. Significant divergence in the outputs will be investigated.   **Longer Term Remediation Plan**:   1. The model owner will investigate the possibility of obtaining empirical support for the assumptions x and y once an additional 6 months of data are collected. 2. The model owner will investigate the possibility of modifying the modeling approach to reduce the reliance on judgmental assumptions. |

Model Owner:

The Fraud Intelligence model has limitations.

**External Factors**

Model performance over time is dependent on external factors within the live environment being consistent with those factors that are present within the historical data that is used for model training. Stable model performance is dependent on external factors such as macroeconomic conditions, customers' business policies and decisioning processes, and portfolio management.

**Target Population**

The score was built on data that consists of both pre-book and post-book U.S. bankcard applicants that were sourced from internal customer samples from LexisNexis Risk Solutions.

The post-book fraud tag is the same as the “confirmed fraud” indicator and corresponds with the way a client would account for a financial loss due to fraud. Fraud that is “suspect” or does not necessarily correspond to a financial loss is considered a pre-book fraud tag.

|  |  |  |
| --- | --- | --- |
| ***Model Weakness or Limitation*** | ***Associated Model Risk(s)*** | ***Model Risk Mitigants / Remediation*** |
| **Dependence on External Factors** | Changes in macroeconomic conditions, customer business policies, decision-making processes, and portfolio management could negatively impact model performance. | Regular monitoring of external factors affecting the live environment. Annual reviews of macroeconomic and business policy changes. |
| **Target Population Misalignment** | Differences between the live population and the development population could result in decreased predictive accuracy and incorrect risk assessments. | Periodic analysis of the live population to ensure alignment with the development population. Quarterly reviews and adjustments as needed. |
| **Fraud Tagging Differences** | Inconsistencies in fraud tagging (e.g., confirmed fraud vs. suspect fraud) could lead to inaccuracies in model predictions and business decisions. | Routine validation of fraud tagging processes to ensure alignment with model development assumptions. Bi-annual audits of tagging practices. |
| **Data Instability Risk** | Model performance may degrade if there are significant shifts in consumer behavior or reporting practices over time, impacting input data stability. | Regular score monitoring to identify and address shifts in behavior or data quality. Monthly stability tracking and annual recalibration if necessary. |

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



## MODEL ESTIMATION / TRAINING AND SELECTION

Note: “Model estimation/training” is mostly applicable for those models that rely on statistical estimation and optimization techniques, such as regression analysis or machine learning techniques. However, this section is also relevant to some other types of models, including those that are developed using expert judgment (qualitative models).

**Reference Document List**

Please list all the documents referred to in this section.

|  |  |  |
| --- | --- | --- |
| **#** | **Reference Document Name** | **High Level Description and purpose of the Document** |
| 1 | FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf | It is the model reference guide. |
| 2 | MRM-CONTROL01 - y&n Model Assmt 048 - Albert YesM - LexisNexis LNFI.docx | Enterprise risk management and Model risk classification procedures. |
| 3 | MRM-CONTROL02 - Model-IRR Assmt 048 -L- Albert LNFI\_FINAL.docx | Model inherent risk rating assessment form. |
| 4 | 2023\_LNRS\_BCOverview\_Tech\_Resilience\_IT.pdf | It is the overview of business continuity technical resilience - IT |
| 5 | LexisNexis\_Business\_Continuity\_Disaster\_Recovery\_  Assessment\_Dec\_18\_2023\_13\_23 (1).pdf | It is the LexisNexis Business Continuity/Disaster Recovery Assessment Report. |

### Estimation Methodology and Assumptions

Describe in detail the model estimation methodology, including the assumptions that may be implicit in the estimation technique. For example, ordinary least squares estimations include assumptions about regression residuals. Describe any expert judgments related to the estimation, such as the selection of variable weighting methodologies.

For machine learning (ML) models, discuss whether monotonicity of relationships between the model features and the target variable is important or required, and whether the ML algorithm is configured to ensure such monotonicity.

Model Owner:

The LexisNexis Fraud Intelligence Model is based on the Gradient Boosting Decision Trees (GBDT) algorithm, specifically implemented using XGBoost. This methodology inherently captures complex, non-linear relationships between features and the target variable, without requiring explicit monotonic constraints.

**Monotonicity Consideration:**

In this model, monotonicity of relationships between features and the target variable is not explicitly enforced. The nature of XGBoost allows it to flexibly model both increasing and decreasing relationships as needed, optimizing for predictive accuracy rather than imposing monotonic constraints. This approach is suitable for fraud detection models, where interactions between variables and fraud risk may exhibit non-linear and non-monotonic patterns.

**Configuration for Monotonicity:**

While XGBoost provides functionality to impose monotonicity constraints on features if required, such constraints were not deemed necessary for this model. The decision not to enforce monotonicity aligns with the empirical findings during model development, where the relationships between input features and fraud risk varied in complexity and direction.

The absence of monotonic constraints ensures that the model leverages the full predictive power of the features, allowing it to adaptively fit the underlying patterns in the data. This flexibility enhances the model's performance for detecting fraudulent activity, as demonstrated by its high AUC and Fraud Detection Rate (FDR) during validation.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### Modeling Software / Platform

Provide information on the software and/or programming language used in the model estimation/training (including the version number, if applicable). If relevant, list the specific algorithms and packages used in model training.

Model Owner:

The GBDT (gradient boosted decision tree) is the standard modeling framework that is leveraged for fraud risk scoring based on rigorous, proprietary research and testing.

This model was developed with XGBoost version 1.6.2 as implemented in Python 3.9.12.

EWB is using LNFI Bank card model version 1.

### Hyper-parameter Tuning

For machine learning models, include a detailed description of the hyper-parameter tuning process, including the following information:

* The approach used for hyper-parameter tuning, including the rationale for leveraging this approach.
* The list of the hyper-parameters tuned (as well as those that are left at default values) and the range of values searched. If applicable, explain why some hyper-parameters were not tuned.
* Performance metric(s) used to select the optimal hyper-parameters and the supporting rationale.
* Sufficiently detailed discussion of the results of the tuning process and selected values, including any judgmental adjustments to the parameters, if any.

If the model also utilizes pre-training during development, provide details of the pre-training process and the pre-trained model as well as related analysis/test/comparison results.

Model Owner:

Gradient Boosting Decision Trees (GBDT) algorithms have a set of parameters, called hyper-parameters, that control the learning process and must be tested through a process called hyper-parameter tuning.

Hyper-parameters control how the machine learning algorithm behaves. Each shallow decision tree of the GBDT model introduces a split in the model that minimizes the cumulative error of the decision tree ensemble. The hyper-parameters are the scoring coefficients in the model.

The model development process did not involve any external or manual hyper-parameter tuning. Instead, the hyperparameters were set to default or pre-determined values optimized within the XGBoost framework.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### Feature / Variable Selection

Describe in detail the approaches used to select candidate and final model variables/features, including the relevant criteria/thresholds for quantitative selection criteria, as well as any expert judgments.

Describe the process for involving subject matter specialists from the line of business to obtain their views on candidate variables, including the associated economic theory/business intuition behind each variable, as well as the expectation for the directional impact of each variable on the model output.

Describe any algorithms or statistical procedures (such as correlation analysis, Information Value analysis, stepwise regression procedure, etc.) used as part of the process to select final model variables from the full set of candidate model variables.

Model Owner:

**Recursive Feature Elimination**

RFE (recursive feature elimination) is an iterative approach to variable reduction that builds successive models each with fewer predictors than the last.

At each stage, a fixed number of variables are removed from the candidate list, and a new model is trained with a fixed set of parameters. The removed variables are chosen based on a feature-importance metric that measures the contribution of each feature in a model. Each iteration records evaluation metrics on a validation sample. This process is repeated until the specified minimum number of features remain, at which point the recorded evaluation metrics can be used to determine the optimal subset of features.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### Model Estimation / Training Results

**For statistical models,** provide statistical estimation results for the final model, as well as other model structures that were considered to be strong candidates. Estimation results should include not only the estimated coefficients, but also the t-statistics and associated p-values, measures of model fit, and summary of results of the appropriate statistical diagnostic tests (detailed statistical testing should be documented in the Statistical Testing section).

In addition to providing the estimation results, explain why this model was selected (relative to other candidate models), including both quantitative and qualitative factors.

**For machine learning models:**

* Provide a listing of the full set of features included in the final model.
* Provide a feature importance chart showing the top X most important features in the final model.
* Provide information on the number of features that contribute 90%, 95%, and 99% of model fit. If the number of features providing the last 1-2% of model fit is significant, explain the rationale for their inclusion.

**For both statistical and machine learning models**, this section should contain for each feature an explanation of the economic theory/business intuition for the inclusion of this feature, as well as the assessment of the estimated directionality of the relationship between the feature and the target variable relative to the a priori expectations. For simple statistical models this assessment can be accomplished through the evaluation of the estimated coefficient signs. For complex statistical and machine learning models, use of explainability testing techniques is required (refer to Section 3.3.5 Model Explainability Testing).

Model Owner:

Random samples of applications were used to select training, testing, and out-of-time validation datasets. The model is a collection of different clients, with each client down sampled by a different amount to approximately equalize the bad rate between clients.

The sampling method is designed to maximize predictive performance and stability over time. The approach takes into consideration weighing up maximizing the bankcard contributors, performance data, data coverage, and on-going client usage.

To measure the score’s ability to predict, the following common metrics are evaluated: the AUC (area under curve) and the FDR (fraud detection rate).

The AUC score is the relationship between true positives and true negatives. The higher the AUC score, the higher the accuracy of the model.

Similarly, the FDR measures the percentage of fraud that is identified in the riskiest nth percentage depth

of file. FDR1 reports the percentage of all fraud cases that are concentrated in the highest scoring one

percent of scores. FDR3 reports fraud capture rate at three percent depth of file. FDR5 & FDR10 reports fraud capture rate at five percent & ten percent depth of file.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



#### Judgmental Adjustments

Describe and justify any judgmental overlays/overrides of statistically estimated input parameters. If any such adjustments are intended to be conservative, explain in what way they are conservative. Note: this section should not be used to detail any overlays/overrides to the model outputs (described in Section 3.3.11. Need for Model Overlays).

Model Owner:

The LexisNexis Fraud Intelligence model primarily uses a data-driven approach, relying on the training data and XGBoost algorithm to make predictions. There is no explicit mention of any judgmental adjustments or expert-driven modifications to the model’s output in the provided model reference guidelines document.

### Other Types of Model Estimation

#### Model Calibration

If applicable, describe the calibration process for models that are regularly fit to market data.

Model Owner:

LexisNexis Risk Solutions uses an odds-doubling methodology for model score calibration, where a score

of 525 corresponds to the odds of a bad rate at 0.0004, with odds doubling every 45 points.

For probability p (0 ≤ p ≤1), which indicates the bad rate in the sampled training data, the score is

calculated from x by the following equation:

Where the weighted probability when the odds are 0.0004, the log term becomes 0 and the score is 525. Log terms calculate how great the odds are compared to the midrate odds 0.0004 and create odds doubling every 45 points.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



#### Vendor Model Tuning

If applicable, describe the process and results for any customization of vendor models (e.g., tuning of vendor model behavioral model parameters to Bank portfolio credit or prepayment experience) that is analogous to a statistical estimation.

Model Owner:

**Warning Codes**

The model is augmented with warning codes.

In production, every application has particular variable values that contribute to increasing the score value. LexisNexis Risk Solutions provides a common set of warning codes that are sufficiently descriptive to handle all types of score influences but are also beneficial to determining correct courses of remedial action. By tracking all of the intermediate score changes that are caused by each variable value of an application, the most important warning codes are identified and presented in the warning code output.

**Overwrites**

In addition to the model-generated warning codes, there are several special circumstances where anomalies in the input data are communicated through a warning code, referred to as overwrites.

Overwrites take priority over warning codes that are generated by the model based on specific criteria.

**Deceased SSN**

If the SSN that is provided for the consumer is reported by the SSA (Social Security Administration) DMF (Death Master File) or by other LexisNexis Risk Solutions proprietary data sources as deceased, then the model returns a score value of 999.

Under this circumstance, all six warning codes contain a value of either 248 (Identity reported as Deceased) or 232 (SSN reported as Deceased), based on the type of match. When other types of anomalies are detected in the score request, the model overwrites the first model-generated warning code and shifts the model-generated warning codes to the subsequent warning code positions.

\*If more than one overwrites applies, then the model-generated warning codes could be overwritten entirely.

Please refer to the attached document for particulars of Warning codes and overwrites.



**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



## Model Development Testing

For each test discussed in the following subsections, include the purpose of the test, the testing methodology, the criteria used to evaluate test results (that is, the applicable metrics and thresholds), and a summary of the results with commentary and conclusions. For any anomalous results, the conclusions should include information on the impact of these results on the model outputs and business use, and whether they require any specific risk mitigant.

The level of detail for the testing documentation should be sufficient to provide a clear and definitive basis for the model owner’s conclusions about model’s performance and robustness.

**Reference Document List**

Please list all the documents referred to in this section.

|  |  |  |
| --- | --- | --- |
| **#** | **Reference Document Name** | **High Level Description and purpose of the Document** |
| 1 | FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf | It is the model reference guide. |
| 2 | MRM-CONTROL01 - y&n Model Assmt 048 - Albert YesM - LexisNexis LNFI.docx | Enterprise risk management and Model risk classification procedures. |
| 3 | MRM-CONTROL02 - Model-IRR Assmt 048 -L- Albert LNFI\_FINAL.docx | Model inherent risk rating assessment form. |
| 4 | 2023\_LNRS\_BCOverview\_Tech\_Resilience\_IT.pdf | It is the overview of business continuity technical resilience - IT |
| 5 | LexisNexis\_Business\_Continuity\_Disaster\_Recovery\_  Assessment\_Dec\_18\_2023\_13\_23 (1).pdf | It is the LexisNexis Business Continuity/Disaster Recovery Assessment Report. |

### Statistical and Technical Assumptions Testing

For statistical and any other models that include statistical and other technical assumptions, provide testing of all assumptions associated with the selected estimation technique (e.g., for Ordinary Least Squares models on time series data this includes testing for multicollinearity, heteroscedasticity, non-normality of errors, autocorrelation, non-stationarity, seasonality, etc.).

For vendor models, to the extent that the assumptions testing information is available from the vendor, include the model owner’s assessment of the testing results and any associated risks.

Model Owner:

The Fraud Intelligence model is designed to provide predictive insights about the identity fraud risk that is associated with new applications for products or services.

The model was developed and validated using industry-standard principles and methodologies.

**Framework**

The modeling framework is chosen to balance two primary concerns: interpretability and predictiveness. The model use case often dictates which algorithms may be employed. In this case, the model is not used for adverse action and uses boosted ensembles of decision trees.

The data that is used to build the model includes applications for products and services and the eventual status of each application and account regarding performance (for example, “fraud” or “not fraud”). Random samples of applications were used to select training, testing, and out-of-time validation samples.

The boosted tree modeling technique produces a robust non-linear model. The training population is representative of identity fraud behavior as reported by bankcard consortium members, and the model generalizes well to all populations.

Standard analytic modeling techniques were used to build the model. This process included the use of testing and validation datasets that are separate from a training dataset. The training data is sampled for the purpose of training to effectively differentiate the two populations using an empirical comparison of the characteristics present in each population. Test-validation data is typically an out-of-time sample, which is constructed so that the data represents a time period in the future of the training data. Test data may also be an out-of-time sample that occurs in the future of the training data. The outcome data from the training data is used to adjust the model to optimize the prediction of fraud risk while minimizing error. The testing and validation data are used to validate the performance of the model by comparing the prediction made by the model with the known outcome on data that was never exposed to the model training process.

The model is trained on a blended dataset that consists of applications from a variety of members within

the Inquiry Identity Network and historical bankcard applications

**Modeling Algorithm**

The GBDT (gradient boosted decision trees) algorithm is a stage-wise ensemble learning technique that aims to produce a strong predictor from a successive series of weak learners.

A weak learner is an estimator that produces a prediction better than a random guess. GBDT uses shallow decision trees as weak learners. At each stage of the algorithm, a shallow classification tree is built using the candidate feature set and outcome variable to predict the residuals of the entire prior ensemble. The final output of the model is the sum of all weak learner outputs in the ensemble.

LexisNexis Risk Solutions used regression trees as the base learner and minimized the binomial deviance using the additive boosting process. The specific implementation used was xgboost.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### Model Performance / Fit Testing

Provide testing of model performance / fit on the estimation and hold-out samples, including calculations of relative and absolute model errors for different population/product/portfolio risk segments and time periods. For some models, their fit can be evaluated using various additional statistical metrics and analytical techniques. This includes, for example: the K-S test, ROC curves (and AUC/Gini coefficient and similar measures of discriminatory power), lift charts, Precision/Recall, F1 score, risk profiling, etc.

For vendor models, include the model owner’s assessment of the model performance/fit testing results provided by vendor (based on vendor’s data) and any associated risks. In addition, include testing results on the Company’s internal data (or explain why it is not feasible).



#### **In-sample Performance/Fit**

Use this section for the testing of model performance/fit on the data on which the model was estimated/trained.

Model Owner:

LexisNexis Risk Solutions deploys a sampling routine that carefully considers a multitude of factors.

The sampling method is designed to maximize predictive performance and stability over time. The approach takes into consideration weighting to maximize the bankcard contributors, performance data (for the fraud tags that were used to identify the “target population”), data coverage, and on-going client usage. Fraud behaviors commonly shift in a very short period of time. To account for a vast spectrum of behavioral insights, multiple time periods were used.

The unsampled data from which the samples were built possessed the characteristics in the following table.

**Sample Performance Summary**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Population** | **Time Period** | **Good Count** | **Bad Count** | **Total Count** | **Bad Rate** |
| All Records | 1/17 to 12/19 | 68,277,081 | 410,231 | 68,687,312 | 0.60% |
| Training Sample | 1/17 to 6/19 | 34,655,016 | 197,430 | 34,852,446 | 0.57% |
| In-Time  Validation | 1/17 to 6/19 | 23,103,344 | 131,620 | 23,234,964 | 0.57% |
| Out-of-Time  Test Sample | 7/19 to 12/19 | 10,518,721 | 81,181 | 10,599,902 | 0.77% |

There were different fraud rates for different clients during different periods, so the FCR (fraud capture rate) data that is reported in this document was normalized to the long-term unsampled fraud rate of each client over the entire date range (2017 through 2019).

All samples included some down sampling of goods and higher fraud rates than are seen in the general population, so all the performance calculations display results in which the fraud rate is normalized to the underlying population bad rate.

For this analysis, records are re-weighted to reflect the population’s bad rate, so only proportions are shown in the table, rather than actual counts.

To measure the score’s ability to predict, the following common metrics are evaluated: the AUC (area under curve) and the FDR (fraud detection rate). The AUC score is the relationship between true positives and true negatives. The higher the AUC score, the higher the accuracy of the model. Similarly, the FDR measures the percentage of fraud that is identified in the bottom nth percent depth of file.

The following performance table is based on the samples that were used during model development. These samples include training. The performance table measures the predictive nature of the Fraud Intelligence model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sample Type** | **Model** | **AUC** | **FDR1** | **FDR3** | **FDR5** | **FDR10** |
| Training | Fraud  Intelligence -  Bankcard | 0.960 | 52.5% | 73.8% | 80.0% | 86.6% |

**Training Sample Performance**

Number of records: 1,931,802

AUC score: 0.960

The following table displays the performance of the training sample.

***Training Sample Performance***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Score Cutoff** | **% of Records** | **Cuml % of Records** | **Bad Rate** | **% of Bads** | **Cuml % of Bads** |
| 950 | 0.4% | 0.4% | 11.6% | 25.1% | 25.1% |
| 900 | 0.2% | 0.6% | 12.0% | 11.2% | 36.3% |
| 850 | 0.2% | 0.8% | 11.9% | 10.6% | 46.9% |
| 800 | 0.4% | 1.2% | 11.1% | 9.9% | 56.8% |
| 750 | 0.6% | 1.8% | 8.9% | 9.3% | 66.1% |
| 700 | 1.4% | 3.2% | 5.7% | 8.5% | 74.6% |
| 650 | 3.1% | 6.3% | 2.9% | 7.7% | 82.3% |
| 600 | 6.9% | 13.2% | 1.3% | 6.6% | 89.0% |
| 550 | 13.7% | 26.9% | 0.5% | 5.4% | 94.4% |
| 500 | 20.4% | 47.2% | 0.2% | 3.4% | 97.8% |
| 450 | 21.1% | 68.4% | 0.1% | 1.6% | 99.4% |
| 400 | 15.1% | 83.4% | 0.0% | 0.5% | 99.9% |
| 350 | 8.4% | 91.9% | 0.0% | 0.1% | 100.0% |
| 300 | 4.0% | 95.8% | 0.0% | 0.0% | 100.0% |
| 250 | 2.3% | 98.2% | 0.0% | 0.0% | 100.0% |
| 200 | 1.5% | 99.7% | 0.0% | 0.0% | 100.0% |
| 150 | 0.2% | 100.0% | 0.0% | 0.0% | 100.0% |
| 100 | 0.0% | 100.0% | 0.0% | 0.0% | 100.0% |
| 50 | 0.0% | 100.0% | 0.0% | 0.0% | 100.0% |
| 0 | 0.0% | 100.0% | 0.0% | 0.0% | 100.0% |

The following table displays a translation of the score cutoff to the risk depth for the training sample.

***Training Sample Score Cutoff to Risk Depth***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Risk Depth** | **Score Cutoff** | **% of Records** | **Cuml % of Records** | **Bad Rate** | **% of Bads** | **Cuml % of Bads** |
| 1% | 822 | 1.0% | 1.0% | 68.1% | 52.5% | 52.5% |
| 2% | 741 | 1.0% | 2.0% | 19.7% | 15.2% | 67.7% |
| 3% | 705 | 1.0% | 3.0% | 7.8% | 6.0% | 73.8% |
| 4% | 683 | 1.0% | 4.0% | 4.8% | 3.6% | 77.4% |
| 5% | 666 | 1.0% | 5.0% | 3.3% | 2.6% | 80.0% |
| 6% | 654 | 0.9% | 5.9% | 2.4% | 1.7% | 81.8% |
| 7% | 643 | 1.0% | 6.9% | 2.0% | 1.6% | 83.3% |
| 8% | 634 | 1.0% | 7.9% | 1.7% | 1.3% | 84.6% |
| 9% | 626 | 1.0% | 8.9% | 1.4% | 1.1% | 85.7% |
| 10% | 619 | 1.0% | 9.9% | 1.2% | 0.9% | 86.6% |
| 20% | 572 | 9.9% | 19.9% | 0.7% | 5.6% | 92.2% |
| 30% | 542 | 9.9% | 29.8% | 0.4% | 2.8% | 95.1% |
| 40% | 517 | 10.1% | 39.8% | 0.2% | 1.8% | 96.9% |
| 50% | 494 | 10.0% | 49.9% | 0.2% | 1.2% | 98.1% |
| 60% | 471 | 10.1% | 59.9% | 0.1% | 0.8% | 98.9% |
| 70% | 446 | 9.9% | 69.8% | 0.1% | 0.5% | 99.5% |
| 80% | 414 | 10.1% | 79.9% | 0.0% | 0.3% | 99.8% |
| 90% | 365 | 10.0% | 89.9% | 0.0% | 0.2% | 100.0% |
| 100% | 1 | 10.1% | 100.0% | 0.0% | 0.0% | 100.0% |

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



#### **Out-of-sample (but not out-of-time)**

Use this section for the testing of model performance/fit on data from the same time period as the in-sample estimation/training data but held out for model testing purposes.

Model Owner:

The following performance table is based on the samples that were used during model development. These samples include validation (out-of-sample). The performance table measures the predictive nature of the Fraud Intelligence model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sample Type** | **Model** | **AUC** | **FDR1** | **FDR3** | **FDR5** | **FDR10** |
| Validation | Fraud  Intelligence -  Bankcard | 0.940 | 48.1% | 69.1% | 75.5% | 82.5% |

**Validation Sample Performance**

Number of records: 1,287,865

AUC score: 0.940

The following table displays the performance of the validation sample.

***Validation Sample Performance***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Score Cutoff** | **% of Records** | **Cuml % of Records** | **Bad Rate** | **% of Bads** | **Cuml % of Bads** |
| 950 | 0.4% | 0.4% | 11.5% | 24.7% | 24.7% |
| 900 | 0.2% | 0.6% | 11.7% | 10.7% | 35.3% |
| 850 | 0.3% | 0.9% | 11.5% | 10.1% | 45.5% |
| 800 | 0.4% | 1.3% | 10.6% | 9.3% | 54.7% |
| 750 | 0.7% | 2.0% | 8.2% | 8.6% | 63.3% |
| 700 | 1.5% | 3.5% | 5.2% | 8.0% | 71.3% |
| 650 | 3.1% | 6.7% | 2.8% | 7.5% | 78.7% |
| 600 | 6.9% | 13.6% | 1.3% | 6.8% | 85.5% |
| 550 | 13.7% | 27.2% | 0.6% | 6.0% | 91.6% |
| 500 | 20.2% | 47.4% | 0.3% | 4.6% | 96.2% |
| 450 | 21.0% | 68.4% | 0.2% | 2.5% | 98.7% |
| 400 | 15.0% | 83.4% | 0.1% | 0.9% | 99.7% |
| 350 | 8.4% | 91.8% | 0.0% | 0.3% | 99.9% |
| 300 | 4.0% | 95.8% | 0.0% | 0.0% | 100.0% |
| 250 | 2.3% | 98.2% | 0.0% | 0.0% | 100.0% |
| 200 | 1.6% | 99.7% | 0.0% | 0.0% | 100.0% |
| 150 | 0.2% | 100.0% | 0.0% | 0.0% | 100.0% |
| 100 | 0.0% | 100.0% | 0.0% | 0.0% | 100.0% |
| 50 | 0.0% | 100.0% | 0.0% | 0.0% | 100.0% |
| 0 | 0.0% | 100.0% | 0.0% | 0.0% | 100.0% |

The following table displays a translation of the score cutoff to the risk depth for the validation sample.

***Validation Sample Score Cutoff to Risk Depth***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Risk Depth** | **Score Cutoff** | **% of Records** | **Cuml % of Records** | **Bad Rate** | **% of Bads** | **Cuml % of Bads** |
| 1% | 836 | 1.0% | 1.0% | 62.2% | 48.1% | 48.1% |
| 2% | 752 | 1.0% | 2.0% | 19.4% | 14.9% | 63.0% |
| 3% | 714 | 1.0% | 3.0% | 8.0% | 6.1% | 69.1% |
| 4% | 690 | 1.0% | 4.0% | 4.9% | 3.7% | 72.8% |
| 5% | 672 | 1.0% | 5.0% | 3.5% | 2.7% | 75.5% |
| 6% | 658 | 1.0% | 6.0% | 2.6% | 2.1% | 77.6% |
| 7% | 647 | 1.0% | 6.9% | 2.1% | 1.6% | 79.1% |
| 8% | 637 | 1.0% | 8.0% | 1.7% | 1.4% | 80.5% |
| 9% | 629 | 1.0% | 9.0% | 1.4% | 1.1% | 81.6% |
| 10% | 622 | 0.9% | 9.9% | 1.3% | 0.9% | 82.5% |
| 20% | 573 | 10.1% | 20.0% | 0.8% | 6.4% | 88.9% |
| 30% | 543 | 9.8% | 29.8% | 0.4% | 3.4% | 92.3% |
| 40% | 518 | 9.9% | 39.7% | 0.3% | 2.4% | 94.7% |
| 50% | 495 | 9.9% | 49.7% | 0.2% | 1.8% | 96.5% |
| 60% | 472 | 10.0% | 59.6% | 0.2% | 1.3% | 97.8% |
| 70% | 446 | 10.3% | 69.9% | 0.1% | 1.0% | 98.8% |
| 80% | 414 | 10.0% | 79.9% | 0.1% | 0.7% | 99.5% |
| 90% | 365 | 9.9% | 89.9% | 0.0% | 0.4% | 99.9% |
| 100% | 1 | 10.1% | 100.0% | 0.0% | 0.1% | 100.0% |

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



#### **Out-of-time**

Use this section for the testing of model performance/fit on data from the time period different from the in-sample data, and not used in the estimation either because it was not yet available at the time of model estimation, or because it was available but excluded from the estimation/training for the express purpose of performing out-of-time model fit testing.

Model Owner:

The following performance table is based on the out-of-time sample validation. The performance table measures the predictive nature of the Fraud Intelligence model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sample Type** | **Model** | **AUC** | **FDR1** | **FDR3** | **FDR5** | **FDR10** |
| Out-of-Time  Test | Fraud  Intelligence -  Bankcard | 0.950 | 48.6% | 71.9% | 79.1% | 86.0% |

High AUC and FDR values that are close to training dataset suggest that the model performs well on unseen data.

**Test (Out-of-Time) Sample Performance**

Number of records: 797,253

AUC score: 0.950

The following table displays the performance of the test (out-of-time) sample.

***Test (Out-of-Time) Sample Performance***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Score Cutoff** | **% of Records** | **Cuml % of Records** | **Bad Rate** | **% of Bads** | **Cuml % of Bads** |
| 950 | 0.3% | 0.3% | 11.7% | 19.9% | 19.9% |
| 900 | 0.3% | 0.6% | 11.9% | 14.0% | 33.9% |
| 850 | 0.4% | 1.0% | 11.0% | 13.2% | 47.2% |
| 800 | 0.5% | 1.4% | 9.2% | 10.5% | 57.7% |
| 750 | 0.8% | 2.2% | 6.6% | 8.9% | 66.6% |
| 700 | 1.4% | 3.6% | 4.1% | 8.0% | 74.6% |
| 650 | 2.6% | 6.2% | 2.1% | 6.9% | 81.5% |
| 600 | 5.2% | 11.4% | 0.9% | 5.4% | 86.9% |
| 550 | 9.9% | 21.3% | 0.4% | 4.5% | 91.4% |
| 500 | 16.6% | 37.9% | 0.2% | 3.6% | 95.0% |
| 450 | 19.3% | 57.2% | 0.1% | 2.5% | 97.5% |
| 400 | 14.5% | 71.8% | 0.1% | 1.3% | 98.8% |
| 350 | 10.0% | 81.8% | 0.1% | 0.6% | 99.4% |
| 300 | 7.9% | 89.6% | 0.0% | 0.4% | 99.8% |
| 250 | 6.3% | 96.0% | 0.0% | 0.2% | 99.9% |
| 200 | 3.2% | 99.2% | 0.0% | 0.1% | 100.0% |
| 150 | 0.7% | 99.9% | 0.0% | 0.0% | 100.0% |
| 100 | 0.1% | 100.0% | 0.0% | 0.0% | 100.0% |
| 50 | 0.0% | 100.0% | 0.0% | 0.0% | 100.0% |
| 0 | 0.0% | 100.0% | 0.0% | 0.0% | 100.0% |

The following table displays a translation of the score cutoff to the risk depth for the test (out-of-time) sample.

***Test (Out-of-Time) Sample Score Cutoff to Risk Depth***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Risk Depth** | **Score Cutoff** | **% of Records** | **Cuml % of Records** | **Bad Rate** | **% of Bads** | **Cuml % of Bads** |
| 1% | 844 | 1.0% | 1.0% | 40.2% | 48.6% | 48.6% |
| 2% | 761 | 1.0% | 2.0% | 13.4% | 16.2% | 64.8% |
| 3% | 718 | 1.0% | 3.0% | 5.9% | 7.1% | 71.9% |
| 4% | 690 | 1.0% | 4.0% | 3.4% | 4.2% | 76.1% |
| 5% | 669 | 1.0% | 5.0% | 2.5% | 3.0% | 79.1% |
| 6% | 653 | 1.0% | 6.0% | 1.7% | 2.0% | 81.1% |
| 7% | 640 | 1.0% | 7.0% | 1.4% | 1.7% | 82.8% |
| 8% | 629 | 1.0% | 8.0% | 1.1% | 1.3% | 84.0% |
| 9% | 620 | 0.9% | 8.9% | 0.8% | 0.9% | 84.9% |
| 10% | 611 | 1.0% | 9.9% | 0.8% | 1.0% | 86.0% |
| 20% | 556 | 9.9% | 19.8% | 0.4% | 4.9% | 90.9% |
| 30% | 522 | 10.0% | 29.8% | 0.2% | 2.6% | 93.5% |
| 40% | 495 | 10.0% | 39.9% | 0.1% | 1.8% | 95.3% |
| 50% | 470 | 9.9% | 49.8% | 0.1% | 1.3% | 96.6% |
| 60% | 443 | 9.9% | 59.7% | 0.1% | 1.1% | 97.7% |
| 70% | 408 | 10.2% | 69.9% | 0.1% | 0.9% | 98.6% |
| 80% | 361 | 10.0% | 79.8% | 0.1% | 0.7% | 99.3% |
| 90% | 298 | 10.1% | 89.9% | 0.0% | 0.5% | 99.8% |
| 100% | 1 | 10.1% | 100.0% | 0.0% | 0.2% | 100.0% |

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### Model Stability and Overfitting Testing

Provide testing to assess the stability of the model’s estimated relationships, for example:

* For statistical regression models, this involves regression coefficient stability testing and testing for structural breaks. Coefficient stability testing can be performed by repeatedly re-estimating the model on different subsets of the development sample (e.g., random sub-samples or samples representing different time periods covered by the dataset) as well as out-of-sample / out-of-time datasets. Values of regression coefficients and p-values across all samples are then assessed to evaluate the model stability.
* For machine learning models, because a comparison of model parameters is either impossible or impractical, testing of model stability generally involves a comparison of key performance statistics (e.g., K-S, AUC, Precision, Recall, F1, etc.) on different training and testing datasets. A common technique for assessing machine learning model’s stability and evaluating the risk of model overfitting is k-fold analysis. K-fold analysis should be performed in addition to testing of the model on the training, validation, out-of-sample, and out-of-time datasets.

Model Owner:

LNFI Model stability was assessed using the out-of-time (OOT) validation technique. This approach evaluates the model on dataset from a different time-period than the training dataset to test its temporal stability.

The sampling method is designed to maximize predictive performance and stability over time. The approach takes into consideration weighting to maximize the bankcard contributors, performance data (for the fraud tags that were used to identify the “target population”), data coverage, and on-going client usage. Fraud behaviors commonly shift in a very short period of time. To account for a vast spectrum of behavioral insights, multiple time periods were used.

The following performance table is based on the samples that were used during model development. These samples include training, validation and out-of-time (OOT) population. The performance table measures the predictive nature of the Fraud Intelligence model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sample Type** | **Model** | **AUC** | **FDR1** | **FDR3** | **FDR5** | **FDR10** |
| Training | Fraud  Intelligence -  Bankcard | 0.960 | 52.5% | 73.8% | 80.0% | 86.6% |
| Validation | Fraud  Intelligence -  Bankcard | 0.940 | 48.1% | 69.1% | 75.5% | 82.5% |
| Out-of-Time  Test | Fraud  Intelligence -  Bankcard | 0.950 | 48.6% | 71.9% | 79.1% | 86.0% |

As shown in the table above, the AUC of the training dataset is very close to that of the OOT test dataset. This consistency in performance indicates that the model is stable over time.

Also, when comparing the AUC values of the training and validation datasets, it is observed that the model generalizes well and is not overfitted to the training dataset. The balance between training and validation performance demonstrates the robustness of the model in predicting outcomes across different datasets.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### Back-testing

In addition to the model performance/fit testing documented in Section 3.3.2. Model Performance / Fit Testing, back-testing is highly beneficial and should be performed/documented for certain types of models. Back-testing is a class of testing techniques designed to assess the consistency of model predictions/estimations with the actual observed values, especially for different historical periods and over longer testing horizons.

These tests are designed to measure the accuracy of model performance over specified time periods. When documenting back-testing analyses, it is critically important to provide a detailed description of the test design including, for example:

* The design of the testing dataset includes the description of the time period, and information about any notable exclusions/inclusions that are inconsistent with the data used to develop the model.
* The logic for generating model predictions. For example, when back-testing a mortgage default model, the model developer would typically start with a particular historical portfolio snapshot and then use the model to generate predictions for each subsequent month/quarter without truing the model up using subsequent historical data.
* The source and nature of inputs and assumptions used in the back-test. For example, for a model that uses macroeconomic variables as inputs, the typical practice is to use actual historical values of such inputs during the back-test period (in order to isolate the error of the tested model from the error in the economic forecasts).

Use of graphical presentation of actual and predicted values is necessary in addition to any quantitative measures of model error (e.g., MAPE, MSE, etc.). This allows the model developer and reader to observe any areas of persistent model bias.

The developers should ensure that performance metrics and thresholds for acceptable performance are clearly stated and are aligned with the model’s business use. For example, for stress testing or CECL model designed to produce loss forecasts over a 2-year period, one of the error metrics should be based on the cumulative actual vs. predicted losses over a 2-year back-testing horizon.

Back-testing results should be accompanied by detailed narrative providing the model developers’ assessment of said results and their conclusions about any notable model biases or elevated error rates. Some such notable biases and performance issues may need to be noted as model weaknesses that must have associated risk mitigants.

Back-testing should be carried out for different populations. For example, when analyzing performance of residential or commercial mortgage loans, one should separately evaluate performance of the model on sub-populations that can be reasonably expected to have different behavioral characteristics. For example: different products, different vintages, or different segments of population by FICO score or by LTV or by another key risk driver.

Predictive models should also be back-tested over different economic environments, e.g., periods of stress vs. periods of economic growth. This is especially important for stress testing, CECL, and IFRS 9 models.

For vendor models, include the model owner’s assessment of the back- testing results provided by vendor (based on vendor’s data) and any associated risks. In addition, provide testing results on the Company’s internal data (or explain why it is not feasible).

In-time

Use this section for backtesting using the data from the same time period on which the model was estimated/trained.

Model Owner:

The in-time test could not be conducted as the dataset shared by the vendor is of the different vintage than of training sample. In-time testing typically assesses the model’s performance on data aligned with the same period as its training sample, providing insights into its immediate effectiveness. However, due to the difference in vintages, this assessment is not feasible.

Out-of-time

Use this section for backtesting using data from the time period different from the in-sample data.

Model Owner:

Backtesting was conducted on a current (which serves as out-of-time) dataset covering the vintage period from January 2024 to November 2024. This evaluation provided an independent assessment of the LexisNexis Fraud Intelligence Model’s ability to generalize new, unseen data. Key performance metrics, including AUC (area under curve) and FDR (fraud detection rate) were computed to validate the model’s effectiveness.

The following performance table measures the predictive nature of the Fraud Intelligence model:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sample Type** | **Model** | **AUC** | **FDR1** | **FDR3** | **FDR5** | **FDR10** |
| Current  (Jan24-Nov24) | Fraud  Intelligence -  Bankcard | 0.766 | 56.7% | 66.3% | 63.5% | 58.6% |

The AUC score demonstrates the model’s strong discriminatory power in distinguishing between high-risk and low-risk cases. Additionally, the FDR indicates that the model is capable of effectively identifying a significant proportion of fraud cases within the dataset.

The backtesting process confirms the robustness of the model and above results reinforce the model’s predictive accuracy and operational reliability.



### Model Explainability Testing

**For machine learning models**, provide sufficient information to understand the drivers of the model outputs and the directionality of their impacts. Use feature importance, Partial Dependency Plots, and a global interpretation method that explains the relationship between model inputs and outputs (e.g., SHAP feature importance, permutation-based feature importance, etc.)

For models that require generation of adverse action reason codes, testing of local interpretability using methods such as LIME is also required.

Advantages and disadvantages of the selected explainability testing methods should be discussed as well.

Model Owner:

Model Explainability Testing ensures the model’s predictions are interpretable, actionable, and aligned with business objectives. It provides insights into the rationale behind fraud risk classifications, enhancing trust and transparency.

Techniques such as Recursive Feature Elimination (RFE) were employed during the model development process to identify and keep the most impactful features, reducing redundancy and improving interpretability.

XGBoost, the primary algorithm for the model, has built-in capabilities to rank features based on their importance during training. By leveraging these rankings, the testing highlights how individual features contribute to the overall fraud risk score.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### Benchmarking

Compare model results with alternative results using other models and/or other data (if available). Describe the benchmark model or data in sufficient detail to enable an assessment of its value as a reference point. For example, a benchmark model that is also a formal Challenger model that has been independently validated (with a successful validation outcome) would be a stronger reference point than a benchmark model that may be available but that has not been extensively tested. Similarly, external peer data may be more relevant in a benchmark comparison than broader industry data. Provide a detailed narrative explaining the outcome of the comparison and any notable differences between the model outputs and benchmarks.

Model Owner:

For Benchmarking purpose, IDB9.5, the most recent flagship model for bankcard fraud detection, was utilized as a reference to evaluate LexisNexis Fraud Intelligence (LNFI).

To measure the score’s ability to predict, the following common metrics are evaluated: the AUC (area under curve) and the FDR (fraud detection rate). The AUC score is the relationship between true positives and true negatives. The higher the AUC score, the higher the accuracy of the model. Similarly, The FDR measures the percentage of fraud that is identified in the bottom nth percent depth of file.

The following performance table is based on the samples that were used during model development. These samples include training, validation, and test (out-of-time validation). The performance tables measure the predictive nature of the Fraud Intelligence model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sample Type** | **Model** | **AUC** | **FDR1** | **FDR3** | **FDR5** | **FDR10** |
| Overall | Benchmark | 0.895 | 35.9% | 54.2% | 62.1% | 71.9% |
| Overall | Fraud  Intelligence -  Bankcard | 0.952 | 50.0% | 71.7% | 78.2% | 85.1% |
| Training | Benchmark | 0.896 | 36.9% | 55.1% | 62.8% | 72.6% |
| Training | Fraud  Intelligence -  Bankcard | 0.960 | 52.5% | 73.8% | 80.0% | 86.6% |
| Validation | Benchmark | 0.897 | 37.0% | 55.1% | 62.9% | 72.7% |
| Validation | Fraud  Intelligence -  Bankcard | 0.940 | 48.1% | 69.1% | 75.5% | 82.5% |
| Out-of-Time  Test | Benchmark | 0.900 | 33.9% | 52.6% | 60.3% | 70.2% |
| Out-of-Time  Test | Fraud  Intelligence -  Bankcard | 0.950 | 48.6% | 71.9% | 79.1% | 86.0% |

The above table summarizes the AUC and FDR metrics for both the LexisNexis Fraud Intelligence (LNFI) and benchmark model (IDB9.5). Notably, LNFI outperforms the benchmark model in both metrics when evaluated across all samples, including training, validation, and test (out-of-time validation).

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### Sensitivity Analysis

Quantify the impact on model outputs of changes in the value of model inputs and assumptions (e.g., economic inputs, tuning parameters, calculation rules, and scenarios). If the model design is such that the sensitivity of the model output to changes in an individual input would depend significantly on the value of one or more of the other inputs, the impact of simultaneous changes in inputs should also be evaluated.

Model Owner:

Sensitivity analysis evaluates how changes in the model inputs and assumptions (e.g., economic inputs, tuning parameters, calculation rules, and scenarios) impact the output. This process identifies the robustness of the model and pinpoints variables with the most influence on the predictions. It is critical for understanding the degree to which model’s predictions depend on specific inputs.

For the LNFI Model, no sensitivity analysis was performed, because the model’s hyperparameters were not manually tuned but were inherently optimized through the XGBoost algorithm. Therefore, no insights are available regarding how variations in individual features or inputs could potentially affect the fraud risk scoring.

### Stress Testing / Scenario Analysis

Quantify the impact on model outputs of stressed changes in the values of inputs, including scenarios that are outside the range of ordinary expectations.

For stress testing/CECL/IFRS 9 and other models dependent on economic scenarios, assess the model forecast across benign and stressful scenarios. When evaluating model forecasts under different economic scenarios, the forecasts should be compared to historical values during similar economic conditions (to the extent that such comparison is meaningful). Any notable differences should be explained and justified. For example, if a model produces drastically lower forecasts of losses under a severe stress scenario compared to the historical losses during the Great Recession, an explanation (e.g., notable improvements in the portfolio quality) should be provided and supported with quantitative analysis, where possible.

The forecasts should also be assessed for internal consistency. For example, do the base, adverse, and severely adverse forecasts reflect incremental macroeconomic stress, or, if not, are they consistent with the unique characteristics of the scenarios and business intuition?

For vendor models, stress testing/scenario analysis should be performed on the Company’s internal data. If not feasible, include the model owner’s assessment of the stress testing and scenario analysis provided by vendor (based on vendor’s data) and any associated risks.

Model Owner:

Stress testing/Scenario analysis examines the model’s performance under extreme conditions or hypothetical scenarios. This helps assess the stability and reliability of the model’s predictions in unusual or challenging circumstances, such as dramatic shifts in data distribution or unexpected trends in fraud behavior.

For LNFI Model, no stress testing or scenario analysis was conducted. As a result, the model’s response to extreme hypothetical scenarios or atypical data patterns has not been evaluated.

### Other Testing

Describe other testing performed applicable to the selected modeling approach, **if any**.

Model Owner:

The primary performance evaluations were conducted using AUC (area under curve) and FDR (fraud detection rate) metrics. These metrics were employed to validate the model’s effectiveness in identifying fraudulent activities and ensure operational utilities.

No additional testing methodologies or analyses were conducted beyond the scope of these metrics. These evaluations confirm the model’s performance for the intended use case without necessitating further testing approaches.

### Overall Performance Assessment

Discuss overall conclusions on model performance based on the results of the testing described above.

Model Owner:

The performance of LNFI Model was evaluated on the combined (overall) population by computing the key metrics: AUC (area under curve) and FDR (fraud detection rate). The results demonstrated strong model performance, with high value of AUC indicating excellent discriminatory power in distinguishing between high-risk and low-risk cases. The FDR was also found to be satisfactory, confirming the model’s ability to effectively detect a significant proportion of fraud cases.

The following performance table is based on the samples that were used during model development. These samples include the overall population. The performance table measures the predictive nature of the Fraud Intelligence model.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Sample Type** | **Model** | **AUC** | **FDR1** | **FDR3** | **FDR5** | **FDR10** |
| Overall | Fraud  Intelligence -  Bankcard | 0.952 | 50.0% | 71.7% | 78.2% | 85.1% |

The following tables display the performance of the combined (overall) population.

**Overall Performance**

Number of records: 4,016,920

AUC score: 0.952

The following table displays the overall performance of the overall sample.

***Overall Performance***

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Score Cutoff** | **% of Records** | **Cuml % of Records** | **Bad Rate** | **% of Bads** | **Cuml % of Bads** |
| 950 | 0.4% | 0.4% | 11.6% | 24.2% | 24.2% |
| 900 | 0.2% | 0.6% | 11.9% | 11.4% | 35.6% |
| 850 | 0.3% | 0.9% | 11.6% | 10.8% | 46.5% |
| 800 | 0.4% | 1.3% | 10.6% | 9.8% | 56.2% |
| 750 | 0.7% | 2.0% | 8.3% | 9.0% | 65.2% |
| 700 | 1.4% | 3.4% | 5.2% | 8.2% | 73.4% |
| 650 | 3.0% | 6.4% | 2.7% | 7.5% | 81.0% |
| 600 | 6.6% | 12.9% | 1.2% | 6.5% | 87.5% |
| 550 | 12.9% | 25.9% | 0.5% | 5.5% | 93.0% |
| 500 | 19.6% | 45.4% | 0.3% | 3.9% | 96.9% |
| 450 | 20.7% | 66.2% | 0.1% | 2.0% | 98.9% |
| 400 | 14.9% | 81.1% | 0.1% | 0.8% | 99.7% |
| 350 | 8.7% | 89.8% | 0.0% | 0.2% | 99.9% |
| 300 | 4.8% | 94.6% | 0.0% | 0.1% | 100.0% |
| 250 | 3.1% | 97.7% | 0.0% | 0.0% | 100.0% |
| 200 | 1.9% | 99.6% | 0.0% | 0.0% | 100.0% |
| 150 | 0.3% | 99.9% | 0.0% | 0.0% | 100.0% |
| 100 | 0.1% | 100.0% | 0.0% | 0.0% | 100.0% |
| 50 | 0.0% | 100.0% | 0.0% | 0.0% | 100.0% |
| 0 | 0.0% | 100.0% | 0.0% | 0.0% | 100.0% |

The following table displays a translation of the score cutoff to the risk depth for the overall performance of the samples.

***Overall Score Cutoff to Risk Depth***

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Risk Depth** | **Score Cutoff** | **% of Records** | **Cuml % of Records** | **Bad Rate** | **% of Bads** | **Cuml % of Bads** |
| 1% | 832 | 1.0% | 1.0% | 60.4% | 50.0% | 50.0% |
| 2% | 749 | 1.0% | 2.0% | 18.6% | 15.3% | 65.4% |
| 3% | 711 | 1.0% | 3.0% | 7.8% | 6.3% | 71.7% |
| 4% | 686 | 1.0% | 4.0% | 4.6% | 3.9% | 75.6% |
| 5% | 669 | 1.0% | 5.0% | 3.3% | 2.6% | 78.2% |
| 6% | 655 | 1.0% | 6.0% | 2.4% | 2.0% | 80.3% |
| 7% | 644 | 1.0% | 6.9% | 1.9% | 1.6% | 81.8% |
| 8% | 634 | 1.0% | 8.0% | 1.6% | 1.4% | 83.2% |
| 9% | 626 | 1.0% | 8.9% | 1.3% | 1.1% | 84.2% |
| 10% | 619 | 0.9% | 9.9% | 1.2% | 0.9% | 85.1% |
| 20% | 570 | 9.9% | 19.8% | 0.7% | 5.8% | 91.0% |
| 30% | 539 | 9.9% | 29.7% | 0.4% | 3.0% | 94.0% |
| 40% | 513 | 10.2% | 40.0% | 0.2% | 2.0% | 96.0% |
| 50% | 490 | 9.8% | 49.8% | 0.2% | 1.4% | 97.4% |
| 60% | 466 | 10.2% | 59.9% | 0.1% | 1.0% | 98.4% |
| 70% | 440 | 9.8% | 69.7% | 0.1% | 0.7% | 99.1% |
| 80% | 405 | 10.2% | 79.9% | 0.1% | 0.5% | 99.6% |
| 90% | 349 | 10.0% | 90.0% | 0.0% | 0.3% | 99.9% |
| 100% | 1 | 10.0% | 100.0% | 0.0% | 0.1% | 100.0% |

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### Need for Model Overlays

Document any proposed or implemented adjustments or overlays to the model outputs and their rationale. Describe the process for derivation and application of these overlays. Provide the impact by including model results with and without these overlays. Finally, outline the overlay review & challenge/approval process, including any Senior Management / Committee reviews and approval process if applicable, and the frequency of the overlay re-evaluation.

For vendor models, discuss the need for model tuning/dialing settings to better align model outputs to the Company’s internal outcomes.

Model Owner:

Model overlays refer to post-adjustments applied to a model’s outputs to refine predictions, align with business objectives, or meet regulatory requirements. These adjustments may involve applying business rules, scaling factors, or threshold modifications based on domain knowledge or external constraints, without altering the underlying model algorithm.

For the LexisNexis Fraud Intelligence model, no manual tuning or external adjustments were applied during model development. Instead, the model’s parameters were inherently optimized through the XGBoost algorithm, which efficiently handles feature importance and model tuning during training. Performance metrics such as AUC (area under curve) and FDR (fraud detection rate) in both training and validation phases confirm that the model is performing well, as detailed in the performance assessment section.

Additionally, there is no indication or requirement in the reference guidelines document to implement overlays for this model. Based on these factors, there is no identified need for model overlays.

# PRODUCTION PROCESS COMPLETENESS & ACCURACY

This section includes procedures and information related to model testing and usage following model development or vendor model acquisition.

**Reference Document List**

Please list all the documents referred to in this section.

|  |  |  |
| --- | --- | --- |
| **#** | **Reference Document Name** | **High Level Description and purpose of the Document** |
| 1 | FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf | It is the model reference guide. |
| 2 | MRM-CONTROL01 - y&n Model Assmt 048 - Albert YesM - LexisNexis LNFI.docx | Enterprise risk management and Model risk classification procedures. |
| 3 | MRM-CONTROL02 - Model-IRR Assmt 048 -L- Albert LNFI\_FINAL.docx | Model inherent risk rating assessment form. |
| 4 | 2023\_LNRS\_BCOverview\_Tech\_Resilience\_IT.pdf | It is the overview of business continuity technical resilience - IT |
| 5 | LexisNexis\_Business\_Continuity\_Disaster\_Recovery\_  Assessment\_Dec\_18\_2023\_13\_23 (1).pdf | It is the LexisNexis Business Continuity/Disaster Recovery Assessment Report. |



## Production Application Testing

Describe the testing for accuracy of implementation of the model into production systems.

### System Testing Approach and Results

The objective of model production application testing is to ensure that computational processes implementing model calculations:

* Are consistent with the documented model specifications produced as part of the model development process. This includes source data fields, data transformation rules, mathematical equations, assumption values, etc.
* Are consistent with the documented business / user requirements.
* Are mathematically accurate and complete.
* Have been reviewed for consistency with any applicable accounting/finance specifications (e.g., GAAP and/or accounting policy requirements), stress testing requirements, or any other applicable regulatory requirements.
* Are operationally stable, repeatable, and sustainable.
* Interface accurately with both upstream and downstream systems (where applicable).

For vendor models, the purpose of the production application testing is to ensure that the models are correctly implemented on the Bank’s systems—if on-premises production process is selected, or the vendor’s model production environment is correctly connected to the Bank’s production data environment—if a cloud-based production process is selected, that the Bank’s production data inputs are consistent with the model publisher’s input specifications, and that all applicable software patches and fixes have been applied.

Describe in detail the testing plan for the individual model’s production implementation and its integration within a larger system and the vendor’s model production environment, if applicable. Include User Accepting Testing cases and scenarios, expected outcomes, and the individuals responsible for executing the test cases.

Document the results of the UAT testing execution, and the associated log of issues and subsequent resolutions.

Model Owner:

**System Testing Approach (Implementation Code Validation)**

Once the LexisNexis Fraud Intelligence Model was moved into production, it underwent extensive testing to validate that the scoring code in production matched the original developed code. A test dataset of over 100,000 records was used, including a variety of input characteristics to cover a wide distribution of model attributes. This also included rare, risky cases to ensure they were adequately represented.

**Test Execution and Validation**

The records in the test dataset are run through both the development and production scoring code. Intermediate values, final scores, and any warning codes were compared on a record-by-record basis to ensure that the production version of the scoring code is accurate.

**Results**

The testing confirmed that the production version of the scoring code accurately reflected the developed model, with no discrepancies identified. This validated the model’s readiness for operational deployment.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### User Acceptance Testing Approach and Results

Document the User Acceptance Testing approach, results, and sign-offs.

Model Owner:

User Acceptance Testing (UAT) typically refers to testing conducted by end users or business stakeholders to ensure the system meets business requirements and aligns with operational goals.

For the LexisNexis Fraud Intelligence model, no formal User Acceptance Testing (UAT) was documented or performed by end users, as the validation focused in ensuring technical accuracy and reliability through the **Implementation Code Validation** process.

This rigorous system testing ensured that the model’s scoring outputs, and warning codes were accurate and consistent, mitigating the need for additional User Acceptance Testing (UAT) at this stage.

## Model Production Specifications

The following technical specifications should cover the end-to-end operation of the model, from data inputs and assumptions to final model reports. **To avoid duplication of information, some of the following sections may refer to earlier document sections instead of repeating the information.**

### Model Platform

Describe the technologies used for running the model, for example, Python, R, Excel, etc.

Model Owner:

This model was developed with XGBoost version 1.6.2 as implemented in Python 3.9.12.

### Data and Process Flow Diagram

Provide a flow diagram showing data sources, inputs, quality assurance control points, intermediate results, outputs, and reports.

Model Owner:

**Flow Diagram (Visual Representation)**

This structure ensures clarity in demonstrating the end-to-end process and controls data quality and accuracy within the LNFI model's operation.

**1. Data Sources**

* **Internal Data:** Data from LexisNexis systems containing historical fraud cases, transactional records, and application attributes.
* **External Data:** Contributions from third-party data providers, such as identity verification data, credit history, and other public record sources.

**2. Inputs**

* **Primary Attributes:** Features like identity data (e.g., SSN, DOB, name, and address) and transactional information.
* **Pre-Processed Data:** Attributes already cleaned and standardized for input into the model.

**3. Quality Assurance Control Points**

* **Data Cleaning and Validation:**
* Missing values handled appropriately (e.g., imputation or flagging).
* Validation against business rules to ensure data accuracy.
* **Check for Data Distribution:** Confirm that attribute distributions match the expected patterns.

**4. Intermediate Processes and Results**

**Feature Engineering:**

* Recursive Feature Elimination (RFE) is used to optimize feature selection.

**Model Execution:**

* The XGBoost algorithm processes input attributes to generate a fraud risk score.
* Intermediate outputs such as decision tree splits, feature importance rankings, and cumulative error metrics are produced.

**5. Outputs**

**Fraud Risk Scores:**

* Each transaction or application receives a score indicating its likelihood of being fraudulent (on a scale of 0 to 998).

**Warning Codes:**

* Flags for specific high-risk conditions detected by the model.

**6. Reports**

* **Operational Reports**: Fraud detection rate (FDR), area under the curve (AUC), and risk score distributions are summarized.
* **Compliance and Performance Reports:** Include additional metrics to satisfy internal and external review standards.

**7. Feedback Loop**

* **Review Mechanism:** Fraud case outcomes are evaluated for continuous improvement and potential inclusion in future model iterations.

### Input Data Specifications

Provide a list of all inputs, including measurement units, a description of valid values or ranges (a full data dictionary should be attached in an appendix). Describe any data processing rules, such as filtering missing or invalid values infilling / overrides, substituting ceiling or floor values, data transformations, etc.

Model Owner:

To produce a valid score, a minimum amount of information is needed.

One of the following sets of elements must be provided:

• **First Name**, **Last Name**, **StreetAddress1**, and **Zip5**

• **First Name**, **Last Name**, **StreetAddress1**, **City**, and **State**

• **First Name**, **Last Name**, and **SSN**

• **First Name**, **Last Name**, and **DOB**

• **First Name**, **Last Name**, and **Primary Phone**

Data Sources from different categories were used to develop the model.

All of the following sources that are available for consideration are used in the model:

**Tri-Credit Bureau Identity Activity**

Identity records from three national credit bureaus provide a unique perspective on identity history.

**LexisNexis Risk Solutions Customer Network**

Visibility to inquiry events provides insight on real and fraudulent identity activity.

**Online, Utility, Phone, and Other Behavioral Activity**

Frequent updates provide insight into identity events that are related to address activity, phone usage, online activity, and email activity.

**Local, State, and Federal Government Records**

Government records provide reliable identity data that is difficult to compromise, including the following information:

• Assigned group of SSN values

• Records of reported deceased persons by name, SSN, and DOB

• Public records of interaction with government agencies

**LexisNexis® Inquiry Identity Network**

The Inquiry Identity Network is a proprietary, cross-industry network of U.S. identity information that contains more than one trillion aggregated identity elements, more than two billion historic consumer transactions, and more than eight million reported identity fraud attempts.

The Inquiry Identity Network contains the PII of those individuals for whom transactions were submitted by clients (for example, applicants for credit card products or wireless phone service contracts). PII typically includes name, SSN, address, phone number, DOB, IP address, email address, and date of the transaction (for example, application date).

The Inquiry Identity Network helps to provide a unique cross-industry view of U.S. consumer application activity to enhance physical identity insights and fraud solutions.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### Model Formulas / Algorithms

Describe detailed model formulas, algorithms, and numerical techniques, if possible.

Model Owner:

The LNFI model is built using the Gradient Boosting Decision Trees (GBDT) algorithm, with XGBoost serving as the specific implementation.

The GBDT (gradient boosted decision trees) algorithm is a stage-wise ensemble learning technique that aims to produce a strong predictor from a successive series of weak learners.

A weak learner is an estimator that produces a prediction better than a random guess. GBDT uses shallow decision trees as weak learners. At each stage of the algorithm, a shallow classification tree is built using the candidate feature set and outcome variable to predict the residuals of the entire prior ensemble. The final output of the model is the sum of all weak learner outputs in the ensemble.

LexisNexis Risk Solutions used regression trees as the base learner and minimized the binomial deviance

using the additive boosting process.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### Model Parameters and Settings Values

Provide the values for all parameters and other input assumptions, including hyper-parameters for machine learning models.

For vendor models, specify values of user-selectable settings.

Model Owner:

**Hyper Parameters**

GBDT algorithms have a set of parameters, called hyper-parameters, that control the learning process and must be tested through a process called hyper-parameter tuning.

Hyper-parameters control how the machine learning algorithm behaves. Each shallow decision tree of the GBDT model introduces a split in the model that minimizes the cumulative error of the decision tree ensemble. The hyper-parameters are the scoring coefficients in the model.

**Score Calibration**

LexisNexis Risk Solutions uses an odds-doubling methodology for model score calibration, where a score

of 525 corresponds to the odds of a bad rate at 0.0004, with odds doubling every 45 points.

For probability p (0 ≤ p ≤1), which indicates the bad rate in the sampled training data, the score is

calculated from x by the following equation:

Where the weighted probability when the odds are 0.0004, thelog term becomes 0 and the score is 525. Log terms calculate how great the odds are compared to themidrate odds 0.0004 and create odds doubling every 45 points.

**Vendor Tunings**

**Warning Codes**

The model is augmented with warning codes.

In production, every application has particular variable values that contribute to increasing the score value. LexisNexis Risk Solutions provides a common set of warning codes that are sufficiently descriptive to handle all types of score influences but are also beneficial to determining correct courses of remedial action. By tracking all of the intermediate score changes that are caused by each variable value of an application, the most important warning codes are identified and presented in the warning code output.

**Overwrites**

In addition to the model-generated warning codes, there are several special circumstances where anomalies in the input data are communicated through a warning code, referred to as overwrites.

Overwrites take priority over warning codes that are generated by the model based on specific criteria.

**Deceased SSN**

If the SSN that is provided for the consumer is reported by the SSA (Social Security Administration) DMF (Death Master File) or by other LexisNexis Risk Solutions proprietary data sources as deceased, then the model returns a score value of 999.

Under this circumstance, all six warning codes contain a value of either 248 (Identity reported as Deceased) or 232 (SSN reported as Deceased), based on the type of match. When other types of anomalies are detected in the score request, the model overwrites the first model-generated warning code and shifts the model-generated warning codes to the subsequent warning code positions.

\*If more than one overwrites applies, then the model-generated warning codes could be overwritten entirely.

Please refer to the attached document for particulars of Warning codes and Overwrites.



**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### Model Outputs

Provide a list of all model outputs, including expected values or ranges.

Model Owner:

The model returns a three-digit fraud risk score with a value that ranges from 001 to 998, up to six

warning codes, and score and warning code overwrites. A low score indicates low risk and a score of 999

indicates that the subject is deceased.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### Reports

Provide a list of all standard output files or reports and describe how they are used in the business.

Model Owner:

The LNFI score will assist in determining whether the applicant should be approved or declined.

If triggered for High-Risk Indicators (HRIs) in LexisNexis InstantID, set flow for LNFI check and push to Know Your Client (KYC) manual review.

02 – The input SSN is reported as deceased

11 - The input address may be invalid according to postal specifications

12 - The input zip code belongs to a post office box

50 - The input address matches a prison address

PO - The primary input address is a PO Box

SD – The input address State is different than the LN best address State for the input identity

CZ – Address mismatch between City/State and Zip Code

14 - The input address is a transient commercial or institutional address

CA - The primary input address is a Commercial Mail Receiving Agency

VA - The input address is a vacant address

19 - Unable to verify name, address, SSN/TIN and phone

DI - The input identity is reported as deceased

PR - The input phone appears as high risk in the Digital Identity Network.

41 - The input driver's license number is invalid for the input DL State

DV - Unable to verify driver's license number

ER - The input email address appears as high risk in the Digital Identity Network.

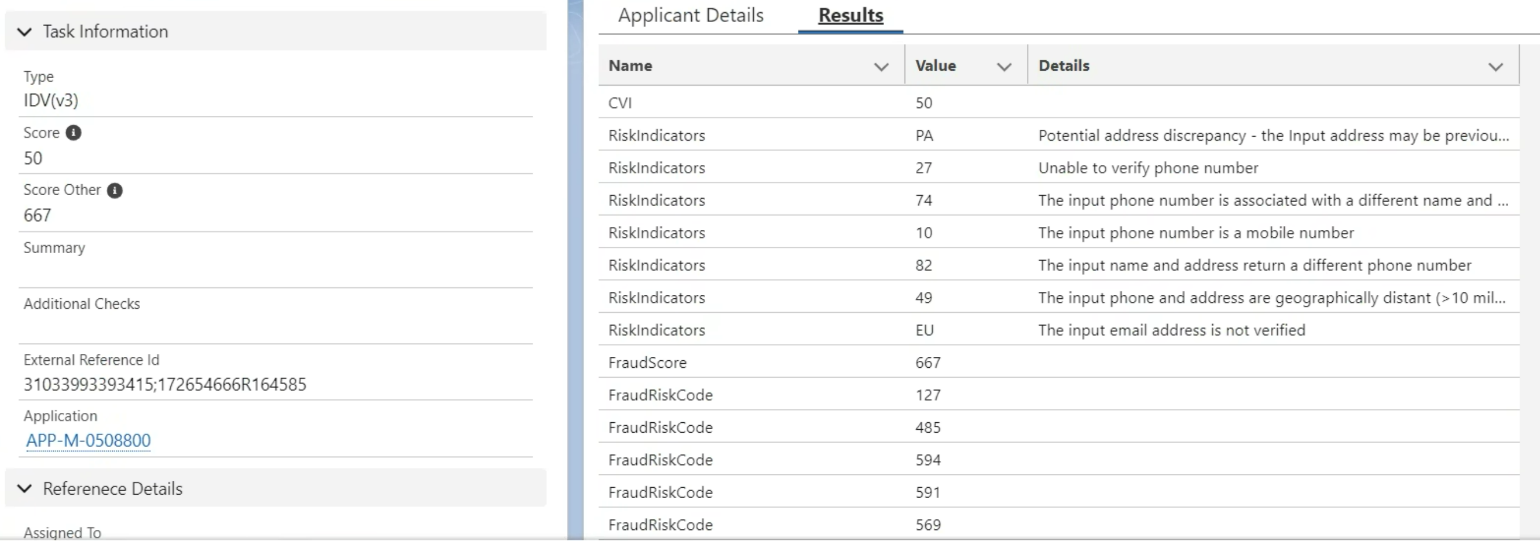
Combination HRIs (27 + 74 OR 82)

27 - Unable to verify the phone number

74 - The input phone number is associated with a different name and address

82 - The input name and address return a different phone number

Also, if LNFI scored greater than 660, the digital onboarding applicant will be send to BSA’s manual queue for additional review.



## Operational Controls

Operational controls related to the model should be in place prior to the production deployment of the model.

### Model Access and Security

Access controls prevent unauthorized changes to the production code and unauthorized operation of the model in production. Describe who has “write access” to the model and can make changes to the underlying code of the model in development and in production, who has access to run the model in production, and who controls model access rights. If there is a formal access monitoring and review process in place, describe it here. Indicate whether any model files are password protected.

If there is no technical mechanism to prevent changes to the model in production (e.g., if the model is implemented using Python code), describe any checks performed to verify that no unauthorized changes have been made since the last approved update or use of the model (such as code comparisons).

Model Owner:

The LexisNexis Fraud Intelligence (LNFI) model ensures robust access controls and security protocols to safeguard the integrity and confidentiality of the model and its associated data. Access to the model is tightly controlled through role-based permissions, which ensure that only authorized personnel can interact with the model or access sensitive data. This includes both internal staff and any third-party vendors who may be involved in the model's maintenance or use. The system is designed to restrict access to certain functionalities, such as training, model updates, or data inputs, based on a user's role and responsibility, preventing unauthorized modifications or misuse.

To further protect the model, sensitive data used in the model development and production processes is encrypted during storage and transmission. Data security is reinforced through secure authentication mechanisms, such as multi-factor authentication (MFA), which ensures that only verified users can access the model's operational environment. Additionally, all interactions with the model are logged to provide a clear audit trail of who accessed the system, what actions they performed, and when these actions took place. This audit trail is vital for monitoring and reviewing access to the model, ensuring compliance with internal and regulatory standards.

Regular security reviews and updates are conducted to keep up with emerging threats and vulnerabilities. The security measures are aligned with industry standards and best practices, such as encryption protocols and data protection regulations, ensuring that both the model’s integrity and its outputs are safeguarded. These comprehensive access and security measures are essential for maintaining the reliability and trustworthiness of the LNFI model and ensuring that it operates within a secure and compliant environment.

**Regulatory Compliance**

The model access framework adheres to industry’s best practices and complies with data protection regulations, such as the General Data Protection Regulation (GDPR) and the California Consumer Privacy Act (CCPA). These measures ensure that sensitive data is handled responsibly.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### Production Deployment

Describe the production deployment process for the new model or changed model, including related controls.

Model Owner:

When a model is moved into production, the model undergoes extensive testing to make sure that the scoring code that is implemented matches the scoring code that was developed.

A large test dataset of more than 100,000 records is used for testing. The test dataset was created by selecting input characteristics that maximize the distribution of the attributes that are used for modeling. Other records are included in the test dataset to ensure that specific risky, but rare, conditions are adequately represented. The records in the test dataset are run through the development and production scoring code. The intermediate values, final score, and warning codes are compared on a record-by-record basis to ensure that the production version of the scoring code is accurate.

The testing confirmed that the production version of the scoring code accurately reflected the developed model, with no discrepancies identified. This validated the model’s readiness for operational deployment.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### Model Usage Controls

Describe the controls related to model usage, such as verification of inputs (including reconciliation to the general ledger or other reference data, as applicable), confirmation of successful model execution (e.g., all input records were processed, output values are within valid ranges), completion of hand-offs to downstream users of the model’s outputs, etc.

Model Owner:

The LexisNexis Fraud Intelligence (LNFI) model has well-defined usage controls to ensure its proper operation, the reliability of its inputs and outputs, and smooth integration into downstream processes. Before execution, input data undergoes rigorous validation to ensure alignment with expected formats, completeness, and consistency. Any missing or anomalous values are addressed through imputation or flagged for further review, ensuring the integrity of the data entering the model.

During execution, the process is closely monitored to confirm successful completion. System logs record critical details, including start and end times, and any errors encountered trigger automated alerts to facilitate timely resolution. Following execution, outputs such as risk scores and warning codes are systematically validated. These outputs are checked for consistency and expected distributions, and summary reports are generated to confirm their accuracy and readiness for downstream use.

Model outputs are securely transmitted to downstream systems, where they are integrated into operational workflows. The integration process is monitored to ensure that output is received and usable by end-users. Standardized reports, including score distributions and warning code summaries, are reviewed and distributed, providing stakeholders with clear insights derived from the model's operation.

To ensure continuous alignment with operational requirements, periodic post-execution reviews are conducted. These reviews assess the overall process, identify any discrepancies, and suggest improvements. All model usage procedures are documented and subject to regular audits, ensuring compliance with internal guidelines and regulatory standards while maintaining the reliability and efficiency of the model's usage framework.

### Model Backup

Provide the model backup procedures, including parties involved and frequency, and describe how the model owner has determined that the procedures are functioning correctly.

Model Owner:

The model backup procedures at LexisNexis involve multiple parties, including the IT and security teams, development teams, and the Enterprise Business Continuity Office (EBCO). Critical systems and data are regularly backed up, with high availability applications having their data replicated between primary and secondary sites to ensure minimal data loss and quick recovery. The backup procedures include data replication to secondary data centers located over 500 miles away from primary sites in the US, regular testing of failover and fallback processes, and secure storage of encrypted backups. Disaster recovery plans are detailed and regularly updated, with frequent tests to ensure their effectiveness. The EBCO conducts regular reviews and audits of the backup procedures, and disaster recovery tests are performed to verify compliance with Recovery Time Objectives (RTO) and Recovery Point Objectives (RPO). Continuous monitoring and reporting provide visibility into the status of backups, and feedback from tests and real incidents is used to improve the procedures. These measures ensure that the model backups are reliable, secure, and can be restored quickly in the event of disruption.

**For more details kindly refer to** “2023\_LNRS\_BCOverview\_Tech\_Resilience\_IT.pdf” & “LexisNexis\_Business\_Continuity\_Disaster\_Recovery\_Assessment\_Dec\_18\_2023\_13\_23 (1).pdf” .



## Contingency Plans

### Disaster Recovery Plan

Provide a reference to the disaster recovery plan or describe the plan here.

Model Owner:

The disaster recovery plan for LexisNexis is comprehensive and ensures the resilience and continuity of critical systems and data. It involves regular backups and data replication to secondary data centers located over 500 miles away from primary sites in the US, with adjustments for local regulations in Europe and emerging markets. The plan includes multiple annual tests, such as failover and fallback exercises, mock disaster preparedness tests, and customer connectivity exercises, to ensure recovery viability. Data security is maintained through encryption and controlled access, and detailed disaster recovery plans outline the steps for responding to data loss or system failures. These plans are regularly updated and tested. Continuous monitoring and reporting provide visibility into backup status, and regular audits ensure compliance with Recovery Time Objectives (RTO) and Recovery Point Objectives (RPO). Feedback from tests and real incidents is used to improve the procedures, ensuring that LexisNexis can quickly and effectively respond to disruptions.

**For more details kindly refer to** “2023\_LNRS\_BCOverview\_Tech\_Resilience\_IT.pdf” & “LexisNexis\_Business\_Continuity\_Disaster\_Recovery\_Assessment\_Dec\_18\_2023\_13\_23 (1).pdf” .



### Business Continuity Plan

Provide a reference to the business continuity plan or describe the plan here. For a vendor model, provide the plan for how the model will be supported or replaced if the external vendor is no longer available to support the model or the vendor’s level of service is unsatisfactory.

Model Owner:

The business continuity plan for LexisNexis Risk Solutions Group (LNRS) includes regular backups and data replication to secondary data centers over 500 miles away, multiple annual recovery tests, and detailed disaster recovery plans. Data security is maintained through encryption and controlled access. Regular audits ensure compliance with Recovery Time Objectives (RTO) and Recovery Point Objectives (RPO). For vendor models, the plan includes provisions for supporting or replacing the model if the vendor is unavailable or their service is unsatisfactory, with contingency plans to transition to an alternative vendor or bring the model in-house.

**For more details kindly refer to** “2023\_LNRS\_BCOverview\_Tech\_Resilience\_IT.pdf” & “LexisNexis\_Business\_Continuity\_Disaster\_Recovery\_Assessment\_Dec\_18\_2023\_13\_23 (1).pdf” .



## Operating Procedures / User’s Guide

Provide step-by-step procedures for running the model, which may include:

1. Input data extraction and preparation, including data cleaning and transformations.
2. Checking the correctness of input data.
3. Setting/updating/checking model settings, assumptions, and parameter values.
4. Checking the correctness of the settings, assumptions, and parameter values.
5. Initiating the processing component of the model.
6. Checking successful completion of the model execution.
7. Extracting model outputs.
8. Checking that model outputs are valid.
9. Producing standard reports.
10. Distributing standard reports.

Note: if there is a separate operating procedural document (or User’s Guide), please list the document name below and share the document with MRM.

Model Owner:

The LexisNexis Fraud Intelligence Model V1 requires adherence to a structured, step-by-step process for operation. Below are the detailed procedures for running the model effectively:

1. **Input Data Extraction and Preparation**

* Extract relevant input data from the system, ensuring all required attributes (e.g., identity and transaction data) are available in the specified format.
* Perform data cleaning to address missing or erroneous values and apply any necessary transformations for compatibility with the model.

1. **Checking the Correctness of Input Data**

* Validate the input data for completeness and correctness using predefined checks, such as ensuring no mandatory fields are missing or contain invalid values.

1. **Model Settings and Parameter Validation**

* Ensure that all model settings, assumptions, and parameters are configured as per the standard operational guidelines.
* Verify that the settings align with the production scoring code.

1. **Initiating the Processing Component**

* Launch the processing phase of the model, using the validated input data and the scoring algorithm (XGBoost implementation).

1. **Completion Verification**

* Confirm that the model execution has been completed successfully without errors or interruptions. Log the execution status for auditing purposes.

1. **Output Extraction**

* Extract the model outputs, including fraud risk scores and associated warning codes, in the specified format.

1. **Validation of Model Outputs**

* Check the outputs to ensure they are within the expected range (e.g., fraud risk scores between 0 and 998).
* Validate specific records, especially edge cases, to confirm that the model is scoring accurately.

1. **Report Generation**

* Generate standard performance and operational reports, such as fraud detection rates (FDR), AUC metrics, and detailed attribute analyses.

1. **Report Distribution**

* Distribute the reports to the designated stakeholders, ensuring that sensitive data is handled securely and in compliance with relevant regulations.

By following these steps, users can ensure the accurate and consistent operation of the model, enabling reliable fraud risk assessment and decision-making. Adjustments or troubleshooting should be escalated to LexisNexis technical support if anomalies are detected during any step.

# ONGOING MODEL GOVERNANCE & OUTCOME ANALYSIS

**Reference Document List**

Please list all the documents referred to in this section.

|  |  |  |
| --- | --- | --- |
| **#** | **Reference Document Name** | **High Level Description and purpose of the Document** |
| 1 | FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf | It is the model reference guide. |
| 2 | MRM-CONTROL01 - y&n Model Assmt 048 - Albert YesM - LexisNexis LNFI.docx | Enterprise risk management and Model risk classification procedures. |
| 3 | MRM-CONTROL02 - Model-IRR Assmt 048 -L- Albert LNFI\_FINAL.docx | Model inherent risk rating assessment form. |
| 4 | 2023\_LNRS\_BCOverview\_Tech\_Resilience\_IT.pdf | It is the overview of business continuity technical resilience - IT |
| 5 | LexisNexis\_Business\_Continuity\_Disaster\_Recovery\_  Assessment\_Dec\_18\_2023\_13\_23 (1).pdf | It is the LexisNexis Business Continuity/Disaster Recovery Assessment Report. |

## Ongoing Risk & Performance Monitoring Plan

**Part 1** - provide an overview of the performance monitoring process, including:

1. Frequency of monitoring activities.
2. Titles/positions of individuals/teams responsible for executing performance monitoring analyses.
3. Individuals responsible for evaluating the resulting reports and documenting conclusions.
4. Stakeholders responsible for reviewing the performance reports and initiating required actions in the event that new risks or performance weaknesses are detected.

**Part 2** - provide the details of the **ongoing risk and performance monitoring plan (together, ongoing monitoring plan)** for this model. Ongoing monitoring plans should generally cover the following two types of periodic monitoring activities:

1. Model Risk Monitoring – Reassessment of the model’s risk profile. This includes but is not limited to reassessment of model weaknesses and limitations, as well as the associated risk mitigants in light of any changes in the model use, Company’s strategy, market conditions, and regulatory environment, among other things.
2. Model Performance Monitoring – Analysis of the model’s **predictive performance** and **identification of emerging model performance weakness**.

Specifically, for Model Performance Monitoring design, it is expected that all models should have some type of outcomes-based performance monitoring process in place to evaluate whether the model is meetings its designed objectives. The Model Owners must specify, as appropriate and feasible for the specific model and its individual uses, detailed plans to monitor model performance through **some combination of the following** four methods:

* Comparison of predicted outcomes to actual values (i.e., back-testing).
* Benchmarking model outputs against comparable external data points, such as observable market information, or outputs of alternative models.
* Analysis of sensitivity of model outputs to variations in model inputs, parameters, and assumptions.
* Stress testing of model predictions to extreme changes in model inputs and assumptions.

The Model Owner should define performance thresholds which, if breached, would require the Model Owner to take corresponding actions. Performance thresholds may be set based on business unit policies or procedures, judgmentally, or based on statistical methodology utilizing model performance over the development sample. In all cases, the approach for setting performance thresholds should be established during development and documented in this section.

**Guidelines** for Risk & Performance Monitoring Plan details:

* Risk Monitoring Plan Details: The risk monitoring plan should list the internal and external factors that should be considered when evaluating model risks. This may include, as applicable:
  + - * Changes in the model use.
      * Changes in the portfolio composition or characteristics of the portfolio/asset/liability/transactions to which the model is being applied.
      * Changes in the Company's strategy.
      * Industry and economic environment changes.
      * Regulatory environment changes.
      * New regulatory findings, independent model validation findings, internal audit findings, external audit findings etc. The plan should include a list of internal and external stakeholders, groups, and committees that may identify, either directly or indirectly, model-related risks through their own “ordinary course of business” activities. It is expected that the Model Owner will establish and maintain periodic communications with these stakeholders to monitor emerging risks.
    - Performance Monitoring Plan Details: The performance monitoring plan should include:
* The source(s) of data used in the performance monitoring process.
* The list of key performance metrics that will be calculated and reported along with their technical specifications.
* Description of the performance analysis that will be performed consistent with the requirements.
* Acceptable performance thresholds for each key metric, if applicable. If a specific threshold is not defined, the Model Owners should document the justification for the lack of threshold. The Model Owners’ rationale for selecting particular performance thresholds must be adequately documented. If, as is sometimes the case, an oversight committee is required by the Business Unit/Line of Business to approve model performance thresholds, then this fact must be reflected in the monitoring plan and the Model Owners must retain evidence of such approvals. Finally, the frequency of the re-evaluation of the performance thresholds should be documented.
* Procedures for communicating and escalating performance issues to appropriate stakeholders (committees, upper management, etc.).
* Procedures for responding to performance threshold breaches.
* The list of stakeholders (individuals and committees) responsible for the review of the risk and performance reports.

Part 1 – Overview

|  |  |
| --- | --- |
| Frequency of monitoring activities (e.g., monthly, quarterly, etc.) | Annually |
| Titles/positions of individuals/teams responsible for executing performance monitoring analyses | Fraud Strategy and Anti-Money Launder Group, in collaboration with Enterprise Risk Management |
| Individuals responsible for evaluating the resulting reports and documenting conclusions | Fraud Strategy and Anti-Money Launder Group, in collaboration with Enterprise Risk Management |
| Stakeholders responsible for reviewing the performance reports and initiating required actions in the event that new risks or performance weaknesses are detected | Senior Management of Risk and Operations (R&O) |

Part 2 – Risk & Performance Monitoring Plan

Model Risk Monitoring Plan Details:

Model Owner:

Model validation is the set of processes and activities that are intended to verify that models are performing as expected, in line with their design objectives and business uses.

Effective validation helps ensure that models are sound. Validation also identifies potential limitations and assumptions and assesses their possible impact.

The frequency for model validations is vaguely defined by the regulatory bodies. For sound models that are offered to clients, LexisNexis Risk Solutions highly recommend that validations are performed on an annual basis.

LexisNexis Risk Solutions performs model performance reviews at least annually to verify that no

degradation occurred in the model performance over time. Performance reviews also ensure that the

model delivers strong and consistent value to each client.

These reviews analyze input data quality, score and reason code distribution, and score performance

over time. An MPR (model performance report) that contains the results of the reviews are prepared for

clients quarterly, semi-annually, or annually, depending on the client’s requests, which meets the OCC

(Office of the Comptroller of the Currency) and FDIC (Federal Deposit Insurance Corporation) risk model

governance guidelines for annual frequency.

Model Performance Monitoring Plan Details:

Model Owner:

The vendor stated that as part of data monitoring process, attribute and score monitoring are conducted to evaluate day-to-day changes. In addition, ad-hoc monitoring is conducted on a weekly and monthly time-lag basis. Data and scores are monitored using the following methods:

• Basic statistics - Mean, standard deviation, minimum and maximum ranges, percentiles, and ratios

• Two-sample K-S (Kolmogorov-Smirnov) test

• Divergence test statistics

• PSI (population stability index)

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



## Model Approval and Change Management Process

In this section, discuss the aspects of the model approval and change management process that are specific to this model.

### Model Approval Process

Provide the names of the individuals (or a committee) involved in the approval process for this model.

Model Owner:

LexisNexis Risk Solutions adheres to a **model panel review framework** to achieve standard methodologies across business units. The panel review process is designed to ensure model soundness, predictive power, and acknowledgement of fair lending requirements.

This process includes a review of standard reports including, but not limited to, FDR reports, sampling routine, relative influence, and detailed attribute bivariate analysis.

In addition to a panel review, all LexisNexis Risk Solutions models are reviewed by an **independent compliance department**. This phase in the process is included in ensuring that all compliance standards are met during the development process. Both the data assets and the modeling process are considered as part of the review process.

More information regarding the exact approval hierarchy or steps can be explored further with LexisNexis’s modeling team.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



### Model Change Log

Provide a reference to the model Change Log. Please refer to the Bank’s 1st Line Model Risk Management Guidelines (MRM-PnP05), the MRM Procedure (MRM-PnP02), and Model Change Log Template v01.docx for detailed requirements.

Model Owner:

LexisNexis Solutions has indicated that the Fraud Intelligence Model is designed to maintain consistent performance over its lifecycle. Specifically, the score produced by the model will remain unchanged, with no modifications to the predictor variable, and no recalibration, realignment, or rescaling of the score distribution.

As a result, there is no current or anticipated need for a model change log. Presently, no change log has been associated with or shared regarding this model, as the intention is to preserve the model in its original state throughout its operational use.

**For more details kindly refer to “**FraudIntelligence\_1.0\_BankcardModel\_ReferenceGuide.pdf **”.**



# APPENDICES

## Appendix A

List and describe references to additional model-related files that have not already been referenced in the Template.

1. DocName\_1.pdf (doc, txt, xls, etc.)

Description: xxx

1. …

Model Owner:

## Appendix B

For vendor models, provide high level description of the vendor company background, qualifications, and services provided, especially relating to EWB’s purchase. In addition, please reference MRM procedure MRM-PnP04, MRM-PnP04 EWBC MRM Vendor Model Onboarding Process v01.pdf, for detailed onboarding and documentation requirements.

Model Owner:

LexisNexis Risk Solutions is a global leader in data and analytics, providing innovative solutions for managing risk, enhancing decision-making, and ensuring compliance. The company leverages vast data resources and advanced analytics to offer solutions tailored to industries such as financial services, insurance, healthcare, and government.

**Background and Qualifications**

LexisNexis Risk Solutions has a long-standing reputation for expertise in identity management, fraud prevention, and risk assessment. The company is part of RELX Group, a global provider of information-based analytics and decision tools for professional and business customers. With a foundation rooted in deep data resources, cutting-edge technology, and commitment to ethical data use, LexisNexis has developed advanced models and solutions to address modern risk challenges effectively.

**Experience**: Decades of experience in leveraging data for fraud prevention and compliance.

**Expertise**: Advanced capabilities in machine learning, data science, and artificial intelligence to provide scalable solutions.

**Compliance**: Adherence to global regulations such as GDPR and CCPA, ensuring ethical and compliant data use.

**Services Provided Related to EWB’s Purchase**

LexisNexis Risk Solutions offers services tailored to EWB’s needs, including:

**Fraud Detection and Risk Assessment:** The LexisNexis Fraud Intelligence (LNFI) model is a sophisticated tool used to evaluate fraud risk, combining identity verification, transaction monitoring, and advanced analytics.

**Identity Management:** Services include verifying the authenticity of consumer identities and detecting synthetic identities.

**Compliance Support:** Solutions to ensure adherence to regulatory requirements, including fair lending and anti-money laundering guidelines.

**Scoring Models:** Provides customized risk scoring models, like the LNFI model, to predict fraud probabilities and support business decision-making.

**Data Enrichment:** Access to extensive databases to enhance the depth and accuracy of risk assessments.

LexisNexis Risk Solutions continues to evolve its services, leveraging its vast data capabilities and innovative technologies to meet the needs of clients like EWB, offering reliable, efficient, and compliant tools for managing fraud and risk.

*Source: LexisNexis Solutions Websites, Reference Guidelines Document*