

# Mathematical Breakdown of Lexicon AI Risk Scoring for Retail Fraud Protection

Lexicon AI Team

July 12, 2024

## Introduction

Lexicon AI utilizes a sophisticated risk scoring model to protect against retail fraud. This document provides a mathematical breakdown of how the risk scores are formulated, ensuring they are unbiased and focused on critical risk vectors. Additionally, it outlines the datasets used and the weight assigned to each component.

## Risk Scoring Model

The overall risk score  $R$  is computed as a weighted sum of several key components. Each component represents a specific risk vector that contributes to the likelihood of fraud. The formula is given by:

$$R = w_1 \cdot D_v + w_2 \cdot B_v + w_3 \cdot T_m + w_4 \cdot G_a$$

where:

$D_v$  : Document Verification Score

$B_v$  : Biometric Verification Score

$T_m$  : Transaction Monitoring Score

$G_a$  : Geolocation Analysis Score

$w_1, w_2, w_3, w_4$  : Weights assigned to each component

## Document Verification Score ( $D_v$ )

This component evaluates the authenticity of identity documents. The score  $D_v$  is derived from the dataset comprising:

- Government-issued IDs
- Passport databases
- Driver's license registries

## Biometric Verification Score (Bv)

This score assesses the accuracy of biometric data such as facial recognition and fingerprint matching. The dataset includes:

- Facial recognition databases
- Fingerprint records

## Transaction Monitoring Score (Tm)

This component monitors transaction patterns for anomalies. The score  $T_m$  is based on:

- Historical transaction data
- Real-time transaction streams
- Behavioral analytics

## Geolocation Analysis Score (Ga)

This score evaluates the risk associated with the geographical location of the transaction. The dataset includes:

- IP address geolocation data
- Historical location-based fraud data

## Weight Allocation

The weights  $w_1, w_2, w_3, w_4$  are determined based on the significance of each component in detecting fraud. The allocation is as follows:

$$w_1 = 0.30, \quad w_2 = 0.25, \quad w_3 = 0.30, \quad w_4 = 0.15$$

$$R = 0.30 \cdot D_v + 0.25 \cdot B_v + 0.30 \cdot T_m + 0.15 \cdot G_a$$

## Ensuring Unbiased Scoring

To ensure the risk scores are unbiased, Lexicon AI employs:

- Regular audits of the datasets and algorithms to identify and mitigate biases.
- Continuous training of machine learning models with diverse and representative data.
- Implementation of fairness constraints in the scoring algorithms.

## **Conclusion**

Lexicon AI's risk scoring model integrates multiple risk vectors, each backed by comprehensive datasets. The weighted approach ensures a balanced and unbiased risk assessment, providing robust protection against retail fraud.