金融AI 実践ガイド

誤検知を減らす金融AIモデルの精度改善方法



データ品質
欠損・偏り・名寄せ・データ鮮度

モデル設計
特徴量・評価指標・ルール併用

判定閾値
業務別・リスク段階別の最適化

人による確認と継続改善
Human in the Loop・監視と再学習

Start Challenge Consulting co., ltd.
挑む力が、世界を照らす

発行元：Start系hallenge系onsulting系o.,系td.

想定読者：DX推進部門／リスク管理部門／審査部門／不正検知・AML担当／AI開発担当

— データ・評価指標・判定閾値・運用設計から見直す実践ガイド —
発行：2026年8月位全19ページ

Confidential — 本資料の無断転載・二次配布を禁じます。

はじめに：誤検知は「精度の問題」ではなく「業務の問題」

金融AIの誤検知は、アラート確認工数と顧客体験を同時に悪化させる。モデルの再学習だけでは収束せず、データ・評価指標・判定閾値・人による確認を一体で設計し直す必要がある。

前提｜二種類の誤り

金融AIでは、問題のある取引を取り逃がす「見逃し」と、正常な取引・顧客を異常と判定する「誤検知」の双方が業務リスクとなる。

どちらか一方だけを最適化すると、もう一方が必ず悪化する。

課題｜誤検知の業務負荷

- アラート確認件数の増加
- 審査・調査担当者の負荷増大
- 正常な決済・送金の停止
- 審査期間の長期化
- 重要アラートの埋没

論点｜精度だけでは解けない

誤検知の要因は、学習データの偏りや正解ラベルの誤り、一律の判定閾値、業務知識を反映しない特徴量、そして運用側の確認プロセスに分散して存在する。

モデル差し替えでは解消しない。

本資料で分かること

- 誤検知の発生原因の構造整理
- 業務別に定めるべき許容水準
- 8ステップの改善アプローチ
- 精度を維持する監視設計
- 推進体制とチェックリスト

本資料の立場：誤検知削減は「モデル改善プロジェクト」ではなく、データ品質・評価指標・判定閾値・人による確認・継続監視を対象とする業務改善プロジェクトとして設計する。

本資料は特定製品の紹介ではなく、金融業務における誤検知の抑制を目的とした一般的な設計手順を整理したものです。各節は「現状分析額→額指標設定→額データ／ラベル／特徴量額→額閾値・ルール額→額人の確認額→額監視」の順に構成しています。

金融AIにおける誤検知とは

誤検知（偽陽性）と見逃し（偽陰性）は同じ判定閾値の上に並ぶ表裏の誤りであり、片方だけを削減する目標設定は成立しない。まず混同行列で現状の分布を確定させることが出発点となる。

混同行列による判定結果の整理

実際の状態	AIが異常と判定	AIが正常と判定
異常	正しい検知（真陽性）	**見逃し（偽陰性）**
正常	**誤検知（偽陽性）**	正しい判定（真陰性）

偽陽性＝誤検知

正常な取引や顧客を異常と判定する誤り。確認工数の増加、決済停止、審査遅延という形で業務側に負担が現れる。

偽陰性＝見逃し

問題のある取引を正常と判定する誤り。損失や当局対応、レピュテーションに直結し、許容度は誤検知より厳しい。

誤検知と見逃しのトレードオフ

判定閾値（スコアのカットオフ）を動かす

閾値を下げる
検知件数が増え、見逃しは減るが、誤検知と確認工数が増える。

閾値を上げる
誤検知と工数は減るが、見逃しが増えるリスクが高まる。

同一モデルでは、両者を同時に改善する余地は限定的

改善余地は「データ品質・特徴量・ルール併用・人の確認」を含めた設計側にある

実務上の含意：「誤検知率を下げる」という単独目標ではなく、「見逃し率を業務許容水準以内に維持したまま誤検知を削減する」という制約付きの目標設定が必要となる。許容水準の定義はP6で業務別に整理する。

誤検知が金融業務に与える影響

誤検知はアラート確認工数の増加に始まり、顧客接点の停止・審査長期化を経て、信頼低下と重要アラートの埋没という経営リスクへ連鎖する。工数問題として局所対応すると、後段の影響が残存する。

発生事象	直接原因	顧客接点への影響	業務への影響	経営への影響
正常な取引・顧客が異常と判定される ・アラート確認件数の増加 ・確認対象リストの膨張 ・同一顧客の反復検知	確認処理が人手に依存する ・審査・調査担当者の負荷増大 ・1件あたり確認時間の増加 ・優先順位付けの不在	正常な取引が止まる ・カード決済・送金の停止 ・口座開設手続の中断 ・問い合わせ・苦情の増加	処理サイクルが伸びる ・融資審査・口座開設の長期化 ・繁忙期の処理積み残し ・要員計画の見直し圧力	顧客満足度と信頼が低下する ・解約・他行利用への流出 ・重要アラートの埋没による検知漏れ ・当局対応リスクの増大

最も見落とされやすいのは右端の「重要アラートの埋没」である。確認容量を超えるアラートが発生すると、実質的な確認品質が低下し、誤検知の増加が結果として見逃しを生む。誤検知対策は見逃し対策と同一線上にある。

事実：誤検知は確認工数と顧客接点の両方に作用する／解釈：影響は5階層で連鎖する／仮説：確認容量超過時に確認品質が低下する（仮）／提案：経営影響まで含めて効果を測る

金融AIで誤検知が発生する主な原因

誤検知の原因は「データ」「モデル」「運用」の3領域に分散して存在し、モデル領域の対策だけでは全体の一部にしか届かない。改善はデータ領域から着手する順序が有効である。

Ⅰ. データに起因する原因

・学習データの偏りや不足
　不正事例が極端に少数で、正常側の多様性も不足している

・正解ラベルの誤り
　調査未完了の案件が「正常」として学習されている

・データ欠損・表記揺れ・重複
　同一顧客が別人物として扱われ、通常行動が学習されない

Ⅱ. モデル・設計に起因する原因

・業務知識を反映していない特徴量
　金額・回数のみで、顧客の通常行動との差分を見ていない

・一律の判定ルールや閾値
　顧客属性・商品・金額帯の違いを考慮していない

・精度指標の選定ミス
　不均衡データにAccuracyを用い、誤検知の実態を見誤る

Ⅲ. 運用に起因する原因

・過去データと現在の取引傾向の変化
　決済手段や送金行動の変化が学習時点と乖離している

・モデル単体で最終判断している
　人の確認範囲とAIの判断範囲が設計されていない

・判定結果が蓄積・還流されない
　担当者の判断が再学習に反映されず、同じ誤検知が続く

着手順序：データ品質・正解ラベル（Ⅰ）を整えないままモデル・閾値（Ⅱ）を調整すると、誤った教師情報に最適化された結果が固定化する。P9～P11でⅠ、P12～P13でⅡ、P14～P15でⅢを扱う。

利用業務別に異なる誤検知の特徴

誤検知の「許容できる水準」は業務によって異なる。改善の前提として、業務ごとに許容できる誤検知率と、絶対に許容できない見逃しを文書で定義しておく必要がある。

利用業務	誤検知の主な現れ方	許容できない見逃し	許容水準の考え方
不正送金・カード不正利用検知	正常な決済・送金の即時停止	高額・連続の不正利用	即時判定のため顧客影響が直接発生。停止ではなく追加認証への振替を検討
マネーロンダリング・AML監視	大量アラートの発生と調査滞留	届出対象となる疑わしい取引	規制対応上、見逃し許容度は極小。誤検知は「削減」ではなく「優先順位付け」で対処
法人融資・個人ローン審査	正常先の過剰なリスク判定・否決	返済能力の重大な誤評価	誤検知は機会損失として計測可能。与信コストとの比較で閾値を決める
保険金不正請求検知	正当な請求の支払保留・調査	組織的・反復的な不正請求	支払遅延が顧客対応上の重大リスク。調査対象の絞り込み精度が要点
本人確認・eKYC	正当な申込の差戻し・手続中断	なりすまし・偽造書類の通過	差戻しは申込離脱に直結。自動承認・要確認・却下の三段構成が前提
市場取引・コンプライアンス監視	正常な発注・通信の照会対象化	相場操縦・情報漏えいの兆候	監視対象が広く母数が大きい。シナリオ単位で許容水準を分けて設定

共通する考え方：誤検知の許容水準は「確認に投入できる工数」と「顧客影響の大きさ」で決まり、見逃しの許容水準は「損失額」と「規制上の要請」で決まる。両者を数値で定義しないまま改善を始めると、成否の判断ができない。

改善ステップ1：誤検知の現状を可視化する

誤検知は総量で語ると打ち手が決まらない。商品・顧客・チャネル・アラート理由の軸で分解し、確認工数と顧客影響の大きい少数のパターンから着手する。

可視化の6手順（実施順）

1. 誤検知件数と発生率の把握
アラート総数・確認済件数・そのうち正常判定となった件数を月次で確定させる

2. 商品・顧客・チャネル別の分析
カード／送金、法人／個人、店頭／アプリ等の軸で発生率の偏りを特定する

3. アラート理由別の分類
検知シナリオ・ルール・スコア寄与要因ごとに件数を集計し、上位要因を抽出する

4. 担当者による判定結果の収集
正常／不正／判断保留の最終判定と、その判断理由を構造化して蓄積する

5. 誤検知による確認工数の算出
1件あたり平均確認時間 × 誤検知件数で、月次の投入工数を金額換算する

6. 改善優先順位の決定
影響度（工数＋顧客影響）と対処難易度の2軸で、着手対象を3件以内に絞る

現状把握シートの記載項目（例）

対象業務	カード不正利用検知
月間アラート件数	［自社実績を記入］
うち正常と判定	［自社実績を記入］
1件あたり確認時間	［自社実績を記入］
上位アラート理由	上位3シナリオを記載
顧客影響の有無	決済停止・手続中断の件数
改善優先度	高／中／低

数値が取得できない項目は、まず記録の取得方法を整備することが最初の改善施策となる。

留意点：多くの現場では「誤検知件数」自体が記録されていない。判定結果を後から集計できる形で残す仕組みの整備が、精度改善の前提条件となる。

可視化の完了条件は「誤検知の発生率が業務軸で分解され、上位3要因と月次確認工数が数値で示されている状態」とする。

事実：誤検知は業務軸で偏在する／解釈：総量管理では打ち手が特定できない／提案：上位3要因に絞って着手する（記載数値は自社実績で置換）

改善ステップ2：評価指標を正しく設定する

金融AIでは異常データが極端に少数のため、Accuracyは高い値のまま実態を隠す。不均衡データを前提に、PrecisionとRecallを併せて確認し、最終判断は業務コストで行う。

Accuracy（正解率）だけでは不十分な理由

不正取引が全体のごく一部である場合、すべてを「正常」と判定するモデルでもAccuracyは高い値を示す。異常検知の性能をまったく持たないモデルが、指標上は良好に見えるため、Accuracyは金融AIの主要指標として使用しない。

Precision（適合率）

AIが異常と判断した結果の正確さ。
低いほど誤検知が多く、確認工数が増える。
誤検知削減の主指標。

Recall（再現率）

実際の異常を検知できた割合。
低いほど見逃しが多く、損失・規制リスクが増える。
維持すべき制約条件。

F1スコア

PrecisionとRecallのバランスを1つの値で示す。
モデル間の比較には有効だが、業務上の非対称なコストは反映されない。

PR-AUC ／ ROC-AUC の使い分け

不均衡データではPR-AUCが実態を反映しやすい。ROC-AUCは陰性が大量にある場合、良好に見えやすい。

業務コストを含めた評価方法

指標の良否ではなく、**誤検知1件あたりの確認コスト**と**見逃し1件あたりの想定損失**を用いて、閾値ごとの総コストを比較する。

総コスト ＝
（誤検知件数 × 確認単価）＋（見逃し件数 × 想定損失額）

確認単価・想定損失額は自社実績から算出する。実績がない場合は（仮定値）と明示し、感度分析で幅を確認する。

報告時の原則
単一指標でなく、Precision・Recall・確認工数・顧客影響件数を併記して意思決定する。

AMLのように見逃し許容度が極小の業務では、Recallを制約条件として固定し、その条件下でPrecisionを最大化する形で目標を設定する。

改善ステップ3：データ品質を改善する

誤検知の多くは、顧客の通常行動が正しく再現されていないことに起因する。名寄せ不備・表記揺れ・データ鮮度の欠落を先に解消することが、モデル調整より効果が大きい。

Ⅰ.誤検知に直結するデータ品質の論点

欠損値・異常値・重複データの処理

欠損を0で補完すると「取引がなかった」と解釈され、通常行動の推定が歪む。補完方法は特徴量の意味に応じて選択し、欠損自体を情報として扱う設計も検討する。

顧客・口座・取引データの名寄せ

同一顧客が複数IDに分散していると、過去の利用実績が参照できず、通常の取引が「初回の異常行動」として検知される。名寄せは誤検知削減の基盤施策となる。

データ定義と入力ルールの統一

加盟店名・送金先名義・住所などの表記揺れは、同一先を別先として扱わせる。正規化ルールを定義し、入力段階と加工段階の双方で適用する。

Ⅱ.学習データの鮮度・偏り・統制

古いデータや不要なデータの除外

決済手段や取引慣行が変化した期間のデータは、現在の正常行動を表さない。対象期間の方針を明文化する。

不正・正常データの偏りへの対応

サンプリングや重み付けを行う場合、検証データは実際の発生比率を維持し、評価が楽観化しないようにする。

データリネージュによる加工履歴の管理

どの元データがどの加工を経て特徴量になったかを追跡可能にし、精度低下時に原因箇所を特定できる状態にする。

個人情報を考慮した匿名化・仮名化

分析・開発環境での利用範囲と手法を、社内規程および関連法令に基づき事前に定める。

優先順位：名寄せと表記揺れの解消は、モデル変更を伴わずに誤検知を削減できる領域であり、最初に着手する対象として適している。

事実：名寄せ不備は履歴参照を不可能にする／解釈：通常行動が再現できない状態が誤検知を生む／提案：データ整備をモデル調整より先に実施する

改善ステップ4：正解ラベルを見直す

誤ったラベルで学習したモデルは、誤りを忠実に再現する。金融業務では「調査未完了」と「正常」が混在しやすく、ラベル定義の統一が誤検知削減の前提条件となる。

Ⅰ.ラベル品質が損なわれる典型パターン

誤った正解ラベルがモデルに与える影響

調査打ち切り案件が「正常」として学習されると、類似の不正パターンを正常と判断する方向に最適化される。逆に確認不十分な案件を「不正」とすれば誤検知が増幅する。

調査担当者による判定基準のばらつき

同一事象でも担当者・拠点・時期によって判定が異なると、モデルは一貫性のない教師情報を学習する。判定基準の統一はデータ整備と同等の重要度を持つ。

後日判明した結果が反映されない

チャージバック確定や当局照会の結果など、後から判明する事実が学習データへ戻されない場合、モデルは過去の誤判定を修正できない。

Ⅱ.ラベル定義と品質管理の設計

ラベル区分	定義と学習上の取り扱い
不正	事実として確認された不正。証跡の要件を明記し、学習の正例とする
正常	調査を完了し問題なしと確認された案件のみ。未着手案件を含めない
判断保留	情報不足・調査未完了。学習から除外し、保留件数を別途監視する

- ・重要案件は複数担当者でレビュー
- ・判定が分かれた案件は上位者が裁定
- ・付与ルールを文書化し改定履歴を管理
- ・判断理由を必須入力とする

- ・定期的にサンプル抽出して再判定
- ・担当者間の判定一致度を測定
- ・後日判明結果との整合を確認
- ・是正結果を学習データへ反映

要点：「判断保留」を独立した区分として設けることが、誤ったラベルの混入を防ぐ最も実務的な対策である。保留を正常に丸め込む運用は避ける。

改善ステップ5：特徴量を改善する

金額や回数の絶対値だけで判定すると、高額でも通常どおりの取引が検知される。「その顧客にとって通常か」を表す相対的・時系列的な特徴量が誤検知削減の中核となる。

As-Is｜絶対値のみの特徴量

・取引金額が一定額以上
・一定期間内の取引回数
・特定国・特定業種との取引
・深夜帯の取引

課題

普段から高額決済を行う顧客、出張・旅行時の利用、給与・賞与時期の送金増加が、いずれも同一に検知される。

To-Be｜通常行動との差分

・顧客属性と通常行動の組み合わせ
・過去の取引パターンとの差分
・取引時間・場所・端末・送金先の変化
・口座・企業・取引先のネットワーク情報

効果

同じ金額でも顧客ごとに異常度が変わり、通常利用の範囲内である取引を検知対象から除外できる。

設計上の要点

一時点ではなく時系列で判断する

直近数日・数週間・数か月の移動平均や変化率を併用し、単発の外れ値と傾向変化を区別する。季節性・給与日等の周期も考慮する。

業務担当者の知見を特徴量へ反映する

調査担当者が「正常と判断した根拠」を聞き取り、その判断材料を特徴量として定義する。誤検知の削減余地が最も大きい領域である。

業務知見を特徴量化する進め方（4ステップ）

1. 誤検知事例の棚卸し
正常と判定された案件を抽出し、判断根拠を分類する

2. 判断材料の言語化
「過去に同一店で利用実績あり」等を業務用語で記述する

3. データ化可能性の判定
既存データで再現可能か、追加取得が必要かを切り分ける

4. 効果検証と採否
追加前後でPrecisionとRecallの変化を検証し採否を決める

留意点：特徴量の追加は説明可能性とのバランスを要する。判定理由を業務側へ説明できない特徴量は、調査工程で活用されず、結果として確認工数の削減につながらない。

事実：絶対値ベースの特徴量は顧客差を表現しない／解釈：通常行動との差分が異常判定の本質／提案：業務担当者の判断根拠を特徴量として定義する

改善ステップ6：判定閾値を最適化する

全顧客に同一閾値を適用すると、低リスク層に誤検知が集中し、高リスク層の見逃しが残る。顧客・商品・金額帯ごとに閾値を分け、二値判定から段階判定へ移行する。

Ⅰ. 閾値設定の構造と一律運用の問題

閾値を下げると
検知件数が増え見逃しは減るが、誤検知と確認工数が比例して増加する。

閾値を上げると
誤検知と工数は減るが、検知されない不正が増える可能性が高まる。

全顧客に同じ閾値を使う問題

取引規模や利用頻度が大きく異なる顧客に同一基準を適用すると、活動量の多い顧客層に誤検知が偏り、活動量の少ない層では異常が閾値に達しない。

調査コストと損失額を考慮した最適化

閾値候補ごとに「誤検知件数 × 確認単価 ＋ 見逃し件数 × 想定損失額」を算出し、総コストが最小となる水準を選ぶ。単価・損失額が未確定な場合は（仮定値）を明示する。

Ⅱ. 段階判定への移行（判定例）

リスク区分	処理方法	顧客への影響
高リスク	即時保留・担当者確認	取引停止（影響大）
中リスク	追加認証・追加情報の取得	一手間追加（影響中）
低リスク	通常処理・継続監視	影響なし

段階判定の効果

従来「停止」としていた中間帯を追加認証へ振り替えることで、顧客影響を伴う誤検知の件数を抑えつつ検知範囲を維持できる。

閾値の分割軸

顧客セグメント／商品・チャネル／取引金額帯／シナリオ単位。分割数は運用管理可能な範囲に留める。

運用上の前提：閾値は一度決めて固定するものではなく、モニタリング結果（P15）に基づき定期的に見直す対象として、変更手続と承認者を定めておく。

事実：閾値操作は誤検知と見逃しを同時には改善しない／解釈：一律閾値は誤検知を特定層に偏在させる／提案：セグメント別の段階判定へ移行する（判定例は設計例）

改善ステップ7：モデルとルールを組み合わせる

明確な禁止条件はルールで確実に判定し、複雑なパターンはAIでスコアリングする。役割を分けることで、ルールの過剰適用による誤検知とAIの説明困難性を同時に抑制できる。

Ⅰ. AIモデルと業務ルールの役割分担

判定手段	担当させる判定内容
業務ルール	制裁対象・禁止取引・法令上の明確な要件など、判断が一意に定まる条件
AIモデル	複数要素の組み合わせや通常行動との差分など、条件化が困難なパターン
アンサンブル	複数モデルのスコアを統合し、単一モデルの偏りによる誤検知を平準化

顧客単位と取引単位の二段階判定
取引単位のスコアに加え、顧客単位のリスク評価を組み合わせる。低リスク顧客の単発の異常は監視に留め、高リスク顧客の同一事象は即時確認へ回す。

誤検知理由を確認できる説明可能AI
スコアの寄与要因を提示できれば、担当者は確認範囲を絞り込める。説明機能は精度そのものではなく、確認工数と誤検知の是正速度に効く。

Ⅱ. 判定フローの構成（To-Be）

① 取引発生 ／ 顧客手続の受付

② ルール判定：禁止条件に該当する取引を確実に停止
条件は一覧管理し、重複・失効ルールを定期棚卸し

③ AIスコアリング：通常行動との差分を数値化
複数モデルのスコアを統合し寄与要因を併せて出力

④ 顧客リスク評価との突合・段階判定
高／中／低の区分に応じて処理方法を分岐（P12）

⑤ 担当者確認 ／ 追加認証 ／ 通常処理（P14）

管理上の注意　ルールの追加のみを続けると条件が相互に重複し、誤検知の発生源が特定できなくなる。追加時は既存ルールとの重複確認と失効期限の設定を必須とする。

要点：ルールは「確実に止めるべきもの」に限定し、判断の余地がある領域はAIスコアと段階判定に委ねる。この切り分けが、ルール肥大化による誤検知の再発を防ぐ。

改善ステップ8：人による確認を設計する

AIの出力を最終判断とせず、判断範囲・提示情報・記録・還流までを一体で設計する。担当者の判断結果が再学習に戻る構造をつくることで、誤検知の削減が継続的に進む。

Human in the Loop の基本構成（5工程の循環）

① AI判定	② 優先順位付け	③ 担当者確認	④ 記録	⑤ 再学習へ還流
スコアと寄与要因を出力し、高・中・低のリスク区分を付与する	リスクと顧客影響で確認順序を決め、確認容量に合わせて配分する	提示された判断材料に基づき、正常・不正・保留を判定する	判断理由・対応結果・所要時間を構造化データとして残す	判定結果をラベルとして反映し、同種の誤検知の再発を抑える

AIが判断すべき範囲と人が判断すべき範囲

AIが担う	大量取引の一次スクリーニングと優先順位付け
人が担う	顧客への影響を伴う最終判断・当局対応を要する判定
責任の所在	最終判断責任は人および所管部門にあることを規程で明確化

担当者に提示する判断材料

- スコアと、スコアを高めた寄与要因
- 当該顧客の通常の取引パターンと今回の差分
- 過去の検知履歴と、その際の判定結果
- 関連する口座・取引先のネットワーク情報
- 確認すべき論点と、判定に必要な最小情報

要点：担当者の負荷を下げる最短の手段は検知件数の削減ではなく、1件あたりの確認時間の短縮である。判断材料の提示設計は、誤検知そのものの削減と同等の効果を持つ。

モデル精度を維持するモニタリング設計

改善した精度は放置すれば劣化する。データ・判定結果・工数の3系統を定点観測し、再学習の実施条件とロールバック基準を事前に定めておくことが運用設計の要件となる。

Ⅰ.監視対象

データドリフトの監視
入力データの分布変化を検知する。取引金額帯・チャネル比率・欠損率の推移を定期比較する。

コンセプトドリフトの監視
入力と正解の関係自体の変化を検知する。不正手口の変化により、同じ特徴量の意味が変わる。

誤検知率・見逃し率の定点観測
顧客層別・商品別に分解して比較し、全体では見えない特定セグメントの劣化を早期に把握する。

Ⅱ.監視指標（KPI）

誤検知率
アラート件数に対する正常判定の割合。
月次・セグメント別

見逃し率
後日判明した事案のうち未検知の割合。
四半期・累積

確認工数
1件あたり確認時間と総投入工数の推移。
月次

顧客影響件数
決済停止・手続中断・苦情の発生件数。
月次

報告体系
月次はリスク管理部門と業務部門、四半期は経営会議へ報告する。閾値変更・再学習の実施は所定の承認手続を経る。

Ⅲ.再学習と切戻しの基準

再学習を実施する条件
・監視指標が事前設定の許容範囲を超過
・入力データ分布の有意な変化を検知
・新たな不正手口の発生を確認
・定期実施（例：四半期ごと）

モデル更新前後の比較テスト
同一の検証データで新旧モデルを比較し、Precision・Recall・確認工数の変化を承認資料として記録する。

問題発生時のロールバック
旧モデルへ戻す判断基準・実施手順・所要時間・承認者を事前に定義し、定期的に手順の実効性を確認する。

設計の要点：監視の目的は異常の検知ではなく「行動の起動」である。各指標に対して、超過時に誰が何を実施するかを事前に紐付けておかなければ、監視は報告作業に留まる。

金融AIモデルの精度改善事例

3つの事例に共通するのは、モデルの差し替えではなく「顧客・企業の通常状態を表す情報を追加した」点である。誤検知削減の打ち手は、アルゴリズムより情報設計に集中している。

事例1｜クレジットカード不正利用検知

課題（As-Is）
金額と地域の絶対値で判定していたため、旅行時の利用や高額購入が一律に検知され、正常な決済が停止していた。

打ち手
顧客ごとの通常利用パターン（利用店舗・地域・金額帯・時間帯）を特徴量として追加し、通常範囲との差分で判定する構成へ変更した。

効果
旅行・高額購入に起因する誤検知を削減し、決済停止による顧客影響を抑制。

事例2｜AMLアラート

課題（As-Is）
単一シナリオの金額基準でアラートが大量に発生し、調査が滞留。重要な取引の確認が後回しになっていた。

打ち手
顧客属性・送金先・取引ネットワーク情報を組み合わせてリスクを段階化し、確認対象を高リスク取引へ集中させた。

効果
検知範囲を維持したまま調査の優先順位が明確化し、滞留の解消につながった。

事例3｜法人融資審査

課題（As-Is）
決算期の財務情報のみで評価していたため、一時的な数値悪化が過剰なリスク判定につながっていた。

打ち手
入出金推移などの月次の資金動態と、業種特性（季節性・回収サイクル）を評価情報に追加した。

効果
一時的要因による過剰判定を抑制し、審査の妥当性と説明力が向上した。

共通する成功要因：「絶対値による一律判定」から「対象の通常状態との差分による判定」への転換。追加した情報は、いずれも業務担当者が従来から確認していた判断材料である。

注記：本ページは特定企業の実績値ではなく、金融業務で一般的に用いられる改善アプローチの類型を整理したものです（定量効果は自社実績での検証が必要）。

精度改善プロジェクトの進め方

許容値とKPIを先に合意し、過去データによる検証と限定範囲での並行稼働を経て本番導入する。開発部門単独では許容値の設定と最終判断の承認が成立しないため、体制設計が前提となる。

実施ステップ（8段階）

1. 対象業務と改善目的の明確化
対象シナリオを限定し、削減対象の誤検知を定義する

2. 誤検知・見逃しの現状分析
発生率・上位要因・確認工数を数値で確定させる（P7）

3. 業務上の許容値とKPIの設定
許容誤検知率と禁止見逃しを業務部門と合意する（P6・P8）

4. データ・ラベル・特徴量の改善
名寄せ・ラベル定義・通常行動の差分特徴量を整備（P9～P11）

5. モデルと閾値の再設計
ルールとの役割分担と段階判定を設計する（P12・P13）

6. 過去データを使った検証
期間を分けて検証し、時系列の情報混入を避ける

7. 限定範囲での並行稼働
現行モデルと並走させ、判定差分を担当者が検証する

8. 本番導入と継続モニタリング
監視指標・再学習条件・切戻し手順を運用へ移管（P15）

関係部門の役割分担

部門	主な役割	意思決定事項
業務部門（審査・調査）	誤検知の実態提供、判断根拠の言語化、並行稼働時の検証	許容誤検知率・確認運用の承認
リスク管理部門	見逃し許容度の設定、モデルリスク管理、監視結果の評価	閾値変更・モデル更新の承認
開発部門（DX・AI）	データ整備、特徴量設計、検証実施、監視基盤の構築	技術方式・検証計画の決定
法務・コンプライアンス	法令・規制要件の確認、個人情報の取扱方針、説明責任の担保	データ利用範囲・記録要件の承認

事実：許容値の設定と承認は業務・リスク管理部門の権限に属する／解釈：開発部門単独では改善の可否判断ができない／提案：ステップ3の合意を着手条件とする

精度改善チェックリスト・まとめ

データ・モデル・運用の3領域を同時に点検する。いずれか1領域のみが未達の場合、他の2領域の改善効果は運用開始後に失われる。

データ

- ☐ 学習データに偏りがないか（正常側の多様性も確認）
- ☐ 正解ラベルの定義が統一されているか（不正／正常／判断保留の三区分）
- ☐ 現在の取引傾向が反映されているか（対象期間・データ鮮度の方針）
- ☐ 名寄せと表記揺れが解消されているか

モデル

- ☐ Accuracy以外の指標を確認しているか（Precision・Recall・PR-AUC）
- ☐ 判定理由を説明できるか（スコアの寄与要因を提示）
- ☐ 業務別に閾値を調整しているか（顧客・商品・金額帯・段階判定）
- ☐ ルールとの役割分担が整理されているか

運用

- ☐ 人による確認プロセスがあるか（AIと人の判断範囲の切り分け）
- ☐ 担当者の判断結果を記録しているか（理由・対応結果・所要時間）
- ☐ ドリフトを継続的に監視しているか（データ／コンセプト双方）
- ☐ 再学習・再検証・ロールバックの基準があるか

まとめ

金融AIの誤検知削減は、モデルの変更だけでは達成できない。データ品質・評価指標・判定閾値・人の判断・運用監視を一体で改善し、業務部門とリスク管理部門が許容水準を合意したうえで継続的に運用することが要件となる。

次のアクション：対象業務を1つ選び、誤検知の現状把握（P7）と許容水準の合意（P6・P8）から着手する。

お問い合わせ

本資料の内容に関するご質問、および自社業務への適用に向けたご相談を承ります。まずは対象業務を1つ選定し、誤検知の現状把握（P7）と許容水準の合意（P6・P8）からご検討ください。

お問い合わせ先

Start Challenge Consulting co., ltd.

担当　平出

E-mail　info@startchg.com

お問い合わせの際は、対象業務（不正検知／AML／審査 等）と現在の課題を併記いただくと、より具体的なご回答が可能です。

挑む力が、世界を照らす

Start Challenge Consulting co., ltd.

ご相談いただける主なテーマ

誤検知の現状分析

発生率・上位要因・確認工数の可視化と優先順位付け

評価指標・許容水準の設計

Precision／Recallと業務コストに基づくKPI設定

データ・ラベル・特徴量の改善

名寄せ、ラベル定義の統一、通常行動との差分設計

運用・監視体制の構築

Human in the Loop、ドリフト監視、再学習・切戻し基準

本資料は一般的な設計手順を整理したものであり、個別の業務・規制要件への適用可否については、貴社の業務部門・リスク管理部門・法務コンプライアンス部門とのご確認をお願いいたします。

Confidential — 本資料の無断転載・二次配布を禁じます。 © Start Challenge Consulting co., ltd.