A Strategic Blueprint for Fintech Innovation: Solving for Next-Generation Credit Risk and Security

Executive Summary

This report provides a comprehensive strategic blueprint for developing winning solutions to two critical hackathon challenges: alternative data credit scoring and advanced mobile banking security. For credit risk, we move beyond mere predictive accuracy to propose an architecture rooted in **Explainable AI (XAI)** and **algorithmic fairness**, addressing the core market need for transparent and inclusive financial products. This "glass box" model is not only technically superior but also regulatorily compliant and ethically sound. For mobile security, we architect an **Adaptive Authentication Framework** that shifts the paradigm from static, high-friction checks to a continuous, invisible defense. This system ingests a multitude of real-time signals—including device integrity, location intelligence, and behavioral biometrics—to generate a dynamic risk score that orchestrates a seamless, passwordless user experience while proactively neutralizing threats like SIM swapping and account takeover. This document serves as an end-to-end guide, providing deep analysis, innovative architectural designs, and a curated set of actionable resources to enable the rapid development of a sophisticated and competitive prototype.

Part I: Reimagining Credit Risk with Alternative Data

Section 1: Deconstructing the Credit Scoring Challenge: The Imperative for Financial Inclusion

1.1 The Systemic Failure of Traditional Credit Models

The foundational challenge presented is that traditional credit scoring models are fundamentally inadequate for assessing the creditworthiness of a large and growing segment of the global population.[1] These models, epitomized by systems like the FICO score, are built upon a narrow set of historical financial data, primarily sourced from credit reports and bank statements.[1] This reliance creates a systemic barrier for individuals and small businesses who lack an extensive credit history, a group often referred to as "credit invisible" or "thin-file" applicants.[2] This cohort is not a niche demographic; it includes vast populations in emerging markets, as well as significant segments in developed economies such as young adults, recent immigrants, freelancers, and the elderly.[2]

The consequences of this exclusionary model are profound. It perpetuates and amplifies existing socio-economic inequalities by systematically denying access to essential financial services to those who may be perfectly capable of repaying a loan but cannot demonstrate it through conventional channels.[5] This creates a paradox where individuals must have credit to build credit, a catch-22 that traps many outside the formal financial system.[4] The very structure of traditional scoring rewards participation in the formal credit system itself, effectively functioning as a "debt score" rather than a true measure of financial responsibility.[4] This is not merely a data gap but a fundamental paradigm flaw; the system is not designed to measure an individual's holistic "willingness and ability to repay" but rather their "history of formal borrowing." For lenders, this translates into a significant missed opportunity. By overlooking these vast, untapped markets, they forgo substantial interest revenue, lose the chance to build long-term relationships with the industry magnates of tomorrow, and risk brand damage by appearing outdated or discriminatory.[6]

From a technical standpoint, these legacy systems are often brittle and ill-suited for the modern data landscape. They frequently operate on outdated infrastructure with rigid, inflexible algorithms that cannot account for unconventional financial profiles.[6] Furthermore, the data they use is often a lagging indicator, with credit files typically refreshed only once a month.[8] This fails to capture the dynamic, real-time financial behaviors of consumers, a particularly acute problem when assessing younger generations like Gen Z, whose financial lives are radically different from those of previous generations and are often conducted through mobile payment apps and subscription services that are invisible to traditional credit bureaus.[2]

1.2 The Market Opportunity: Serving the Underserved

The inadequacy of traditional models creates a clear and substantial market opportunity. In the United States alone, the Federal Reserve Board reports that 6% of adults are unbanked and another 13% are underbanked, collectively representing 19% of the population that traditional models struggle to serve.[2] Alternative data has the potential to make an additional 19 million US adults scorable.[8] This issue is magnified in the emerging markets targeted by the hackathon, where the unbanked and underbanked populations constitute a majority.[1]

Fintech companies have been quick to recognize and exploit this gap, disrupting the lending industry by leveraging alternative data to build more accurate, inclusive, and dynamic credit assessment models.[1] These new approaches are not just a social good; they are a primary business driver. Financial inclusion unlocks new revenue streams, expands the total addressable market, and builds brand loyalty among previously excluded customer segments.[6] The desired outcome of this hackathon challenge is therefore not simply to build another predictive model, but to engineer a new credit risk framework that fundamentally improves the accuracy and, critically, the *representation* of credit assessments. The goal is to create a system that is transparent, scalable, and compliant with data privacy regulations, thereby empowering lenders to extend credit to these underserved populations confidently and with minimal risk.[1]

Section 2: The Alternative Data Arsenal: Current Solutions and Data Sources

To build a more holistic and accurate profile of a borrower, particularly one with a thin credit file, modern lenders are turning to a diverse array of non-traditional data sources. These "alternative data" streams provide a real-time, multifaceted view of an individual's financial responsibility, stability, and character. They can be broadly categorized into three main types: financial and transactional data, digital footprint and behavioral data, and personal and psychometric data.

2.1 A Taxonomy of Alternative Data

Financial & Transactional Data: This is the most direct and powerful category of alternative data, offering a clear window into a person's current financial health.

- **Bank Account & Cash Flow Data:** Through open banking APIs, lenders can (with user consent) analyze bank account activity to understand income stability, spending habits, savings patterns, and average account balances. This provides a real-time, dynamic view of a borrower's ability to manage cash flow, a far more current indicator than a monthly credit report.[8] Companies like Plaid have become central to this ecosystem, providing the secure infrastructure for these connections.[8]
- **Utility, Rent, and Telecom Payments:** A consistent history of on-time payments for essential services like rent, electricity, water, gas, and mobile phone bills is a strong predictor of financial responsibility. These regular payments demonstrate an ability to manage recurring obligations, yet they are historically ignored by traditional credit bureaus.[3]
- **Alternative Loan Repayments:** The repayment history on modern forms of credit, such as "Buy Now, Pay Later" (BNPL) services, paycheck advances, or rent-to-own agreements, can be a valuable data point. While not traditional loans, they demonstrate a borrower's ability to manage and repay debt, which is particularly relevant for younger consumers who may use these services frequently.[8]

Digital Footprint & Behavioral Data: This category analyzes a user's digital life to infer patterns of behavior and stability.

- **Telco Data:** In many developing economies, a person's mobile phone is their primary connection to the digital and financial world. As such, telco data—including billing history, call and text patterns, data usage, SIM card age, and top-up habits—is a highly predictive data source for creditworthiness.[9]
- **Social Media & Web Behavior:** This is a more nascent and controversial area. Lenders can analyze publicly available information from social media profiles, such as professional history on LinkedIn, to verify employment or education.[3] Other signals include the age of an email account, IP address geolocation, and even "clickstream" data, which analyzes how an applicant navigates a lender's

website.[2] However, using social media content itself raises significant privacy, ethical, and regulatory concerns, as it can be easily manipulated and may correlate with protected characteristics, leading to bias.[3]

- **E-commerce & App Usage:** Data on purchasing behavior, the types of apps installed and used, and patterns of in-app purchases can offer insights into a user's lifestyle, financial habits, and overall digital savviness.[3]

Personal & Psychometric Data: This category moves beyond observable financial and digital behaviors to assess an individual's character and potential.

- **Public Records & Demographics:** Information such as verified employment history, educational attainment, and professional licenses can provide additional context and verify information provided by the applicant.[10]
- **Psychometrics:** This is an innovative technique designed for individuals with little to no digital or financial footprint. It involves using scientifically designed questionnaires, interactive games, and puzzles to measure personality traits that are strongly linked to creditworthiness, such as conscientiousness, integrity, and forward-thinking.[15] Case studies in Ethiopia and Indonesia have shown that psychometric scoring can successfully identify reliable borrowers who would otherwise be denied credit, thereby improving financial inclusion.[17]

2.2 Data Providers and Aggregators

A growing ecosystem of companies facilitates access to this alternative data. Open banking platforms like **Plaid** are key enablers for accessing consented bank account data.[8] Specialized vendors such as

RiskSeal, **credolab**, and **FinScore** focus on aggregating and analyzing digital footprints, telco data, and other non-traditional sources to provide comprehensive risk profiles.[11] Even the traditional credit bureaus are adapting;

Experian, for example, launched "Experian Boost," a service that allows consumers to voluntarily add their positive utility and telecom payment histories to their credit files to potentially increase their scores.[12]

A hierarchy of data value and risk is evident. Consented transactional data is the most powerful, defensible, and regulatorily accepted source. The digital footprint offers predictive signals but must be navigated with extreme care to avoid privacy violations and algorithmic bias. Psychometrics provide a powerful solution for the truly "invisible" but require rigorous validation to ensure they are culturally unbiased and effective.[19] The most robust models will not replace traditional data but will fuse it with alternative data. A FICO study confirmed that credit scoring models incorporating both traditional and alternative data are more powerful than models using either type alone.[8] A successful hackathon solution should be designed as a modular engine capable of augmenting a traditional score when available or operating independently when it is not. The following table provides a strategic overview of key alternative data sources.

Table 1: Alternative Data Sources for Credit Scoring

Data Type	Key Metrics	Predictive Power	Data Acquisition Method	Privacy/Bias Risk	Example Providers/Sources
Bank Account Data	Income stability, avg. balance, expense categories, overdraft frequency	High	Open Banking API (user-consented)	Low	Plaid, Tink [8]
Utility Payments	On-time payment history for electricity, water, gas	High	User-provided, data aggregators	Low	Experian Boost, Esusu [10]
Rental Payments	On-time payment history, housing stability	High	User-provided, rent reporting services	Low	Fannie Mae, Plaid [8]
Telco Data	SIM card age, top-up frequency, data usage patterns, call	High	Telco partnerships, data vendors	Medium	FinScore, Trusting Social [9]

	patterns				
BNPL History	Repayment history, credit utilization, transaction frequency	Medium-High	BNPL providers, data aggregators	Low	Affirm, Klarna (via partnerships) [8]
Psychometrics	Scores for conscientiousness, integrity, fluid intelligence	Medium-High	Direct user assessment (surveys, games)	Medium	Entrepreneurial Finance Lab (past) [15]
Digital Footprint	Email age, IP geolocation, web browsing behavior	Medium	Data vendors, on-site analytics	High	RiskSeal, credolab [2]
Social Media Data	Professional history (LinkedIn), network size (metadata)	Low-Medium	Public scraping, user-provided	High	(Controversial, few mainstream providers) [3]

Section 3: The Winning Edge: An Explainable, Hybrid Scoring Architecture

To gain a competitive edge, a modern credit scoring solution must move beyond simply achieving high predictive accuracy. It must be transparent, fair, and trustworthy. This requires a sophisticated architecture that combines powerful machine learning models with state-of-the-art Explainable AI (XAI) techniques, creating a "glass box" system that can justify its decisions.

3.1 The Technology Stack: Machine Learning at the Core

The financial industry is increasingly adopting advanced machine learning models that can capture complex, non-linear relationships in data far more effectively than traditional logistic regression.

- **Model Selection:** For tabular data, which characterizes most credit scoring problems, ensemble methods have proven to be exceptionally effective. Models like **Random Forest** and gradient boosting machines—particularly implementations like **XGBoost** and **LightGBM**—are industry favorites due to their high performance and robustness.[13] These models combine the predictions of many individual decision trees to produce a final output that is more accurate and less prone to overfitting than any single tree.
- **Handling Imbalanced Data:** A critical challenge in credit risk modeling is that default is a rare event, resulting in highly imbalanced datasets where non-defaulters vastly outnumber defaulters. Training a model on such data can lead to it simply predicting the majority class every time. To counteract this, specialized techniques are essential. Over-sampling methods like **SMOTE** (Synthetic Minority Over-sampling Technique) create synthetic examples of the minority class (defaulters), while under-sampling methods randomly remove examples from the majority class to create a more balanced dataset for training.[23]
- **Infrastructure:** Production-grade credit scoring systems are built on scalable cloud infrastructure. Platforms like Amazon Web Services (AWS), Google Cloud Platform (GCP), or Microsoft Azure provide the necessary tools for data storage, processing, model training, and real-time deployment via APIs.[3]

3.2 The "Glass Box" Imperative: Explainable AI (XAI)

The superior predictive power of models like XGBoost comes at a cost: they are inherently complex and operate as "black boxes," making their internal decision-making processes opaque. In the high-stakes, highly regulated world of finance, this opacity is a critical failure point. Explainable AI (XAI) is a field of machine learning dedicated to making these models transparent, and its integration is non-negotiable for several reasons:

- **Regulatory Compliance:** Financial regulations in many jurisdictions, such as the Equal Credit Opportunity Act (ECOA) in the United States, mandate that lenders must provide consumers with specific, accurate reasons for adverse actions like a loan denial.[16] It is impossible to comply with this "right to explanation" using an uninterpretable black box model.[24]

- **Trust and Transparency:** For a model to be adopted, it must be trusted by all stakeholders. Model owners within the financial institution need to understand how the model works to manage risk and validate its logic. Customers who are denied credit deserve a clear explanation to understand the decision and learn how to improve their financial standing.[26]
- **Bias Detection and Algorithmic Fairness:** This is perhaps the most critical role of XAI in this context. Alternative data sources, while powerful, can inadvertently introduce biases if they correlate with protected characteristics like race, gender, or age. XAI techniques are essential tools for auditing models to ensure they are not making discriminatory decisions, a paramount concern for regulators and society.[28]

Two XAI techniques have emerged as the de facto industry standard for explaining complex models on tabular data [32]:

- **SHAP (SHapley Additive exPlanations):** Based on principles from cooperative game theory, SHAP calculates the marginal contribution of each feature to every individual prediction.[33] It provides a unified framework for both *local* explanations (why was this specific person denied a loan?) and *global* explanations (what are the most important features for the model overall?). SHAP values can be used to generate highly intuitive visualizations, such as force plots that show the features pushing a prediction up or down, and summary plots that rank global feature importance.[33]
- **LIME (Local Interpretable Model-agnostic Explanations):** LIME explains a single prediction by creating a simple, interpretable model (like a linear regression) that is trained on a small, perturbed dataset centered around the specific instance being explained.[35] This simple model provides a locally faithful approximation of the complex model's behavior, highlighting the key features that drove that particular decision.

3.3 Proposed Architecture: A Hybrid, Explainable Scoring Engine

A winning solution should not just be a predictive model, but a complete decision intelligence platform. The architecture should be modular and transparent by design.

1. **Data Ingestion & Feature Engineering Module:** A robust pipeline to ingest and

process diverse alternative data sources (simulated for the hackathon). This module will engineer features that capture key signals of creditworthiness, such as income stability (e.g., standard deviation of monthly income), payment consistency (e.g., percentage of on-time utility payments), and digital responsibility (e.g., age of primary email account).

2. **Core Predictive Module:** An XGBoost or Random Forest model trained on the engineered features. This model's primary output is the raw probability of default (P(default)) for a given applicant.
3. **XAI Layer (SHAP Explainer):** This is the crucial differentiating component. An integrated SHAP explainer runs in parallel with the predictive model. For every prediction of P(default), this layer calculates the SHAP values for each input feature, quantifying its exact contribution to that specific outcome.
4. **Output & Justification Module:** The system's final output is multifaceted, providing not just a score but a narrative. It should deliver:
   - The final credit score or probability of default.
   - A human-readable list of the top 3-5 factors that *positively* influenced the score (e.g., "Consistent on-time rent payments," "Stable monthly income").
   - A human-readable list of the top 3-5 factors that *negatively* influenced the score (e.g., "High debt-to-income ratio," "Lack of long-term payment history").
5. **Fairness Module (Advanced Implementation):** To demonstrate a sophisticated understanding of the ethical challenges, this module would incorporate techniques to ensure algorithmic fairness. This could involve pre-processing methods like reweighing the training data to give more importance to underrepresented groups, or post-processing checks to ensure that key fairness metrics, such as demographic parity (equal approval rates across groups) and disparate impact, fall within acceptable, legally defined thresholds.[28] The most advanced approach involves using causal inference models to ensure that alternative data features are not acting as illegal proxies for protected attributes.[29]

This architecture elevates the solution from a simple prediction tool to a comprehensive decision intelligence platform. While any team can train a model to generate a score (the "what"), a system that can explain the "why" behind that score provides immense value to business users, compliance officers, and customers alike. It directly addresses the core business and regulatory requirements of transparency and accountability, making it a far more complete and compelling product. By explicitly designing for fairness, the model becomes more robust, defensible, and ethically sound, transforming a potential constraint into a key competitive feature.

Section 4: Implementation Roadmap: Data, Code, and Papers

To build a prototype of the proposed explainable credit scoring engine, the team will need access to relevant datasets for training, code repositories to accelerate development, and foundational research to inform the model's design.

4.1 Acquiring and Simulating Data

Real-world alternative credit data is proprietary and not publicly available. Therefore, a hybrid strategy of using public datasets for the core architecture and simulating alternative data features is the most practical approach for a hackathon.

- **Public Datasets for Prototyping:** Several public datasets, while primarily containing traditional credit variables, are excellent for building and testing the end-to-end pipeline, including the machine learning model and the crucial XAI layer.
    - **Kaggle** is the primary resource for such datasets. Notable examples include the classic "German Credit Risk" dataset [41], which contains demographic and loan information; the "Credit Card Approval Prediction" dataset [42]; and various "Credit Score Classification" datasets that provide a range of financial features.[43]
- **Strategy for Simulating Alternative Data:** The core of the innovation lies in demonstrating the value of alternative data. The team should start with a base dataset from Kaggle and then programmatically generate synthetic columns that logically correlate with the existing features and the target variable (default). This simulation should be guided by the data taxonomy in Section 2. For instance:
    - Create a utility_payment_ontime_pct feature, where a higher percentage is negatively correlated with the default label.
    - Generate a telco_sim_age_months feature, where a longer SIM card age is also negatively correlated with default.
    - Create a psychometric_conscientiousness_score on a scale of 1-100, where higher scores are associated with a lower probability of default.

This approach allows the team to build a model that uses these simulated features and, more importantly, to use the XAI layer to show how these new features are influencing the model's predictions.

4.2 GitHub Repositories for Accelerated Development

Leveraging open-source code is essential for rapid prototyping in a hackathon environment. The following repositories provide excellent starting points and essential libraries.

- **End-to-End Credit Scoring Frameworks:** These repositories offer a complete project structure, saving valuable time on boilerplate code.
  - **nafiul-araf/Credit-Risk-Modeling-End-to-End-Project** [23]: A comprehensive project that serves as an excellent template. It uses XGBoost, handles class imbalance with SMOTE, and includes a Streamlit web application for an interactive user interface.
  - **Pratik94229/Credit-Card-Default-Prediction-End-to-End-Project** [47]: Another strong end-to-end example that covers the full lifecycle from data validation and preprocessing to model training and deployment.
  - **fedeghigo/Credit-Risk.Machine-learning-application** [48]: This repository focuses on hyperparameter tuning for model optimization and demonstrates deployment of a web app using Streamlit and Heroku.
- **Core Machine Learning & XAI Libraries:** These are the essential building blocks for the proposed architecture.
  - **scikit-learn:** The foundational library for data preprocessing, baseline models (e.g., Logistic Regression, Random Forest), and evaluation metrics.[50]
  - **xgboost / lightgbm:** High-performance libraries for implementing the core gradient boosting model.
  - **shap/shap** [33]: The definitive library for implementing the XAI layer. Its GitHub repository is rich with documentation, tutorials, and examples for generating various explanation plots.
  - **interpretml/interpret** [52]: A comprehensive interpretability toolkit from Microsoft that includes SHAP, LIME, and other advanced techniques like Explainable Boosting Machines (EBMs).
  - **cloudera/CML_AMP_Explainability_LIME_SHAP** [53]: A repository that provides clear, side-by-side examples of implementing and comparing LIME and SHAP, which can be useful for understanding their respective strengths.

4.3 Foundational Research Papers

Basing the project on solid academic and industry research will add depth and credibility to the final presentation.

- **On Fairness and Bias Mitigation:**
    - "**Debiasing Alternative Data for Credit Underwriting Using Causal Inference**" (arXiv:2410.22382) [29]: This is a cutting-edge paper that addresses the most sophisticated concern with alternative data: ensuring it does not act as an illegal proxy for protected attributes. Citing and implementing a simplified version of this concept in the "Fairness Module" would be highly impressive.
    - General research on algorithmic fairness outlines the three primary approaches to bias mitigation: **pre-processing** (adjusting the data before training), **in-processing** (modifying the learning algorithm), and **post-processing** (adjusting the model's predictions).[28]
- **On Explainable AI in Finance:**
    - "**Explaining the Unexplainable: A Systematic Review of Explainable AI in Finance**" (arXiv:2503.05966) [30]: This paper provides a comprehensive justification for why XAI is critical in the financial sector and confirms that techniques like SHAP are state-of-the-art.
    - "**Shapley values as an interpretability technique in credit scoring**" [27]: This article provides a direct application of SHAP to the credit scoring domain, demonstrating its relevance and comparing it to traditional metrics like Weight of Evidence (WOE).
- **On Alternative Data:**
    - "**The Value of Big Data for Credit Scoring: Enhancing Financial Inclusion using Mobile Phone Data...**" (arXiv:2002.09931) [56]: A key study that empirically demonstrates the significant predictive power of telco and mobile phone data for credit scoring.
    - "**Alternative Credit Scoring of Micro-, Small and Medium-sized Enterprises**" (HKMA White Paper) [57]: This white paper from the Hong Kong Monetary Authority provides a strong business and technical case for using alternative data, particularly for lending to small and medium-sized enterprises (MSMEs).

Part II: Fortifying Mobile Banking with Advanced Authentication

Section 5: The Modern Fraud Landscape in Digital Banking

The second problem statement addresses a critical vulnerability in the digital banking ecosystem: fraudulent registrations and account takeovers. As financial services migrate to mobile and internet platforms, fraudsters have developed sophisticated methods to exploit weaknesses in traditional authentication processes, leading to significant financial losses and erosion of customer trust.

5.1 The Core Challenge: Differentiating Customer from Fraudster

The central challenge is to reliably differentiate between a registration or login attempt initiated by a legitimate customer and one initiated by a fraudster who has impersonated them.[1] The problem statement highlights a common and highly damaging scenario: a fraudster, having obtained a customer's credentials through social engineering, registers their own device for mobile banking. Once this registration is successful, they gain control of the account and can swiftly transfer funds to accounts they control.[1] This initial registration event is the pivotal moment where a strong defense is most needed.

5.2 Key Attack Vectors

Modern banking fraud is rarely a simple case of a stolen password. It is a multi-stage process that combines psychological manipulation with technical exploits to bypass security controls.

- **Social Engineering:** This remains the primary entry point for the vast majority of attacks. Fraudsters employ a range of deceptive tactics to trick users into divulging sensitive information. They often create a sense of urgency or authority by impersonating trusted entities such as the user's bank, a courier service, customs officials, or even the police.[1] Common social engineering techniques include:

  - **Phishing:** Sending fraudulent emails with malicious links designed to harvest credentials from fake login pages.[1]
  - **Smishing (SMS Phishing):** A mobile-centric version of phishing where the malicious link is delivered via text message.[1]
  - **Vishing (Voice Phishing):** Deceptive phone calls where a fraudster coaxes a victim into revealing information like debit card details, login credentials, or one-time passcodes (OTPs).[1]
- **SIM Swap Fraud:** This is one of the most potent attacks in the current landscape because it directly undermines a widely used security measure: SMS-based two-factor authentication (2FA). The attack involves a fraudster using stolen personal information to convince a mobile network operator to transfer the victim's phone number to a new SIM card in the fraudster's possession.[58] Once the swap is complete, the victim's phone loses service, and all incoming calls and SMS messages—including password reset links and OTPs—are redirected to the attacker.[58] Research has shown that the authentication procedures used by many mobile carriers to authorize a SIM swap are often insecure and can be bypassed with easily obtainable personal information, making this a highly effective attack vector.[60] The reliance on SMS for authentication is a fundamental vulnerability; security frameworks like those from NIST have long warned against its use due to these risks.[61] Any proposed solution that continues to rely on SMS OTP as a primary security factor fails to address the core of the modern threat.
- **Account Takeover (ATO):** This is the ultimate objective of the fraudster. After successfully compromising credentials and bypassing 2FA, often via a SIM swap, the attacker gains control of the user's online banking account. They can then change the password to lock the legitimate user out, add new payees, and execute unauthorized transactions, often draining the account before the victim is even aware of the breach.[58]

The multi-stage nature of these attacks reveals a critical insight: a robust defense cannot be a single, static checkpoint at the point of login. The attack unfolds over time, from initial information gathering to the final fraudulent transaction. Therefore, an effective security system must be a continuous, dynamic process that monitors for anomalies throughout the entire user journey, from registration to session activity and transaction execution.

Section 6: A Multi-Layered Defense: Current Solutions

To combat the multifaceted threats of modern banking fraud, a defense-in-depth strategy is required, incorporating solutions that operate at the device, user behavior, and authentication protocol layers. Each layer provides a distinct set of signals that, when combined, can create a robust and resilient security posture.

6.1 Device and Network Layer

This layer focuses on verifying the integrity and context of the device and network being used to access the banking application.

- **SIM Binding & Device Fingerprinting:** Basic SIM binding, as mentioned in the problem statement, involves capturing the mobile number, IP address, and device ID.[1] However, this is susceptible to spoofing. Advanced, cryptographically-secure SIM binding offers a much stronger defense. It leverages the unique, tamper-resistant identifier on the SIM card itself, the International Mobile Subscriber Identity (IMSI), to create a hard link between a user's account and their physical SIM card.[59] This method is highly resilient to SIM swap fraud because even if the phone number is ported to a new SIM, the IMSI will not match the one registered to the account, immediately flagging the activity as fraudulent.[59] Complementing this is device fingerprinting, which creates a unique signature based on dozens of hardware and software attributes (e.g., OS version, screen resolution, installed fonts), making it possible to recognize a trusted device and flag new, unrecognized ones.[64]
- **Location Intelligence:** Analyzing geolocation data provides powerful contextual clues for fraud detection. This is a specific requirement of the problem statement, which calls for an algorithm to identify the distance between the customer's known location and the location of the person initiating a registration.[1] Key techniques include:
    - **Impossible Travel Detection:** This is a cornerstone of location-based fraud detection. The system flags two consecutive access attempts from

    geographically distant locations that occur in a time frame too short for legitimate travel. For example, a login from California followed five minutes later by a transaction attempt from New York is a clear indicator of a compromised account.[66]
  - **Risky Location and Cluster Analysis:** The system can maintain a database of high-risk locations, such as IP addresses associated with known botnets, anonymous proxies (VPNs, Tor), or geographic regions with high fraud rates.[66] Cluster analysis can also identify "device farms," where a large number of distinct user accounts are being accessed from a single, concentrated physical location, a strong signal of organized fraud.[66]

6.2 User Behavior Layer

This layer moves beyond the "what" (device, location) to the "how" (the user's unique interaction patterns), providing a dynamic and deeply personal layer of security.

- **Behavioral Biometrics:** This is a passive and continuous form of authentication that analyzes a user's unique, subconscious patterns of interaction with their device. It builds a behavioral profile that is extremely difficult for a fraudster to mimic. Key modalities include:
  - **Keystroke Dynamics:** The rhythm, speed, and pressure of a user's typing.[68]
  - **Mouse and Touch Dynamics:** The patterns of mouse movement, the velocity and curvature of screen swipes, scrolling speed, and the amount of pressure applied to a touchscreen.[68]
  - **Device Handling:** The unique angle at which a user typically holds their phone, captured by the device's accelerometer and gyroscope sensors.[70]
  - **Gait Analysis:** The pattern of a user's walk as they use their device, which can also be captured by motion sensors.[70]
- **Strengths and Weaknesses:** The primary strength of behavioral biometrics is that it is both frictionless for the user (operating invisibly in the background) and provides continuous security throughout an entire session.[68] It can detect anomalies like account takeover even *after* a successful login, for instance, if a fraudster's typing rhythm differs from the legitimate user's.[62] However, its effectiveness depends on having a sufficient baseline of the user's normal behavior. It must also account for natural variations, or "behavioral drift," over time and address potential privacy concerns through data anonymization and encryption.[68]

- **Multimodal Biometrics:** Research consistently shows that combining multiple behavioral modalities (e.g., analyzing keystroke dynamics *and* swipe patterns *and* device handling simultaneously) results in a significantly more accurate and robust authentication system than relying on any single trait alone.[75]

6.3 Authentication Protocol Layer

This layer concerns the cryptographic methods used to verify a user's identity, with a strong industry-wide push to move beyond vulnerable passwords.

- **Passwordless Authentication (FIDO & Passkeys):** This is the new gold standard for secure and user-friendly authentication. Spearheaded by the FIDO Alliance, passkeys aim to replace passwords entirely.[77]
  - **How it Works:** The system is based on public-key cryptography. When a user registers for a service, a unique cryptographic key pair is generated on their device. The **private key** is stored securely within a hardware-backed environment (like a Trusted Platform Module or Secure Enclave) and *never leaves the device*. The corresponding **public key** is sent to the service's server for storage.[79] To authenticate, the server sends a challenge to the device. The user authenticates to their device using its built-in screen lock method (e.g., a fingerprint scan, facial recognition, or PIN). This action authorizes the device to sign the server's challenge with the private key. The signed challenge is sent back to the server, which verifies it using the stored public key.[77]
  - **Key Benefits:** This architecture is inherently **phishing-resistant** because the cryptographic keys are bound to the specific website or app they were registered with; a user cannot be tricked into using their passkey on a fraudulent site.[77] There are no shared secrets (like passwords) stored on the server, drastically reducing the impact of a data breach. For the user, the experience is faster, easier, and removes the burden of remembering complex passwords.[78]

The most effective security posture does not treat these layers as independent silos. Instead, it uses them in concert. Signals from the device and behavior layers should inform the decisions made at the protocol layer. This leads to an "orchestrated" security model where, for example, a low-risk user on a trusted device might be authenticated seamlessly with a passkey, while a user exhibiting anomalous behavior on a new device is challenged with additional verification steps. This intelligent orchestration is the core of an adaptive authentication framework.

Section 7: The Winning Edge: An Adaptive, Behavior-Driven Authentication Framework

The most innovative and effective solution to the problem of impersonation and fraud is not a single new technology, but rather an intelligent framework that integrates multiple layers of defense into a cohesive, risk-aware system. This approach, known as Adaptive Authentication or Risk-Based Authentication, moves security from a static, one-size-fits-all model to a dynamic, context-sensitive one that balances robust security with a frictionless user experience.

7.1 The Core Concept: Adaptive Authentication

The fundamental principle of adaptive authentication is to dynamically adjust the required level of security based on a real-time risk assessment of the user's context and actions.[81] Instead of subjecting every user to the same, often cumbersome, multi-factor authentication process for every login, the system evaluates the risk of each interaction. A legitimate user logging in from their usual device and location during normal hours represents a low-risk event and should be granted access seamlessly. Conversely, an attempt to register a new device from an unrecognized, high-risk IP address immediately after a SIM swap has been detected is a high-risk event that should be blocked and trigger an alert.[85] The goal is to achieve "frictionless security," where security measures are stringent when risk is high but invisible when risk is low, thereby enhancing both safety and user satisfaction.

7.2 Proposed Architecture: A Real-Time Risk Engine

A robust adaptive authentication framework can be architected as a three-layer system: a signal ingestion layer, a risk scoring layer powered by machine learning, and a policy and orchestration engine that translates risk into action.

1. **Signal Ingestion Layer:** This layer acts as the sensory nervous system of the framework, continuously collecting a wide array of data points in real-time during every user session. These signals provide the raw input for the risk assessment.
   - **Device & Network Signals:** This includes the device's unique fingerprint, the status of the cryptographic SIM binding (verifying the IMSI), the IP address and its reputation (e.g., known proxy, Tor node), and the device's current geolocation.[67]
   - **Behavioral Biometric Signals:** This involves the passive collection of behavioral data, such as keystroke dynamics, swipe and scroll patterns, and device handling characteristics (e.g., phone angle from gyroscope data).[69]
   - **Contextual & Historical Signals:** This includes the context of the current action (e.g., time of day, transaction type, monetary value) and compares it against the user's established historical patterns (e.g., typical login times, common transaction amounts, frequently used payees).[86]
2. **Risk Scoring Layer (Machine Learning Model):** This is the analytical core of the system. The ingested signals are fed as features into a machine learning model that calculates a real-time **risk score** for the current action, typically on a scale such as 0-100. For detecting novel fraud patterns, an **unsupervised anomaly detection model** like an Isolation Forest or an Autoencoder is highly effective.[88] These models are trained on data from legitimate user sessions and learn to identify what "normal" behavior looks like. Any significant deviation from this learned norm is flagged as an anomaly and results in a higher risk score. Alternatively, if sufficient labeled historical data is available, a **supervised classifier** like XGBoost can be trained to predict the probability of fraud based on the input signals.[83]
3. **Policy & Orchestration Engine:** This is the decision-making brain of the framework. It takes the numerical risk score from the ML model and translates it into a concrete security action based on a predefined set of rules. This engine orchestrates the user's authentication journey.
   - **Low Risk (e.g., Score 0-30):** The action is deemed safe. For a login, this would trigger a seamless FIDO/Passkey authentication, allowing the user to sign in instantly with a biometric scan or device PIN.
   - **Medium Risk (e.g., Score 31-70):** The action is suspicious and requires additional verification. This triggers a "step-up" authentication challenge. The user might be prompted for a second factor, such as an OTP from a secure authenticator app (critically, *not* from SMS) or a live biometric check like a face scan.[81]

- **High Risk (e.g., Score 71-100):** The action is deemed highly likely to be fraudulent. The system should automatically block the action, potentially lock the account to prevent further damage, and immediately send an alert to the legitimate customer through a secure, out-of-band channel like a push notification to a trusted device or a registered email address.[81] The event should also be flagged for immediate review by a human fraud analyst.

7.3 Applying the Framework to the Problem Statement

This adaptive architecture directly addresses the hackathon's challenges:

- **New Device Registration:** When a user attempts to register a new device, the system ingests signals like the new device fingerprint, its IP geolocation, and crucially, it can query a (simulated) telecom API to check for recent SIM swap activity associated with the user's phone number. A combination of a new device, an unusual location, and a recent SIM swap would generate a very high risk score, leading the orchestration engine to block the registration and alert the true user.
- **Passwordless Login with Continuous Authentication:** A user initiates a login using a secure and convenient passkey. This is a low-risk event. However, the system continues to monitor their behavior passively throughout the session. If, before attempting a large fund transfer, the user's typing rhythm or swipe patterns suddenly deviate significantly from their established profile, the risk score will increase in real-time. The orchestration engine would detect this spike and, before allowing the transfer to proceed, trigger a step-up challenge, such as a quick face scan, to re-verify that the legitimate user is still in control of the session.[73]

This architecture transforms security from a single, brittle checkpoint into a continuous, ambient state of vigilance. It provides persistent protection throughout the entire user session, addressing the fundamental weakness of traditional authentication systems where a compromised login grants an attacker unrestricted access. Furthermore, it creates a powerful data feedback loop: every user interaction, every generated risk score, and every confirmed fraud event becomes valuable training data that can be used to continuously retrain and improve the risk model, allowing the system to adapt and stay ahead of evolving fraud tactics.[65]

Table 2: Adaptive Authentication Policy Matrix

Risk Signal	Risk Score Range	Action	Justification
Known device, known location, normal behavior	**Low (0-30)**	**Allow Seamlessly** (e.g., FIDO Passkey)	High confidence in user identity. Prioritize frictionless experience.
New device OR unusual location	**Medium (31-70)**	**Step-Up Authentication** (e.g., Authenticator App OTP, Biometric Scan)	Context has changed, requiring additional proof of identity before proceeding.
Impossible travel detected	**High (71-100)**	**Block Action + Alert User**	Geographically impossible for the legitimate user to be performing this action. High probability of ATO.
Significant behavioral biometric anomaly	**High (71-100)**	**Step-Up or Block** (depending on context)	User's interaction pattern does not match their profile, suggesting a different person is using the device.
SIM Swap flag from Telco API	**High (71-100)**	**Block Action + Alert User**	A recent SIM swap is a primary indicator of an imminent account takeover attempt.
High-risk IP (Proxy/Tor)	**Medium (31-70)**	**Step-Up Authentication**	Anonymizing services are often used to mask fraudulent activity.
Multiple high-risk signals combined	**Critical (>90)**	**Block + Freeze Account + Alert User**	Overwhelming evidence of fraudulent activity. Immediate action required to prevent loss.

Section 8: Implementation Roadmap: Data, Code, and Papers

To build a prototype of the adaptive authentication framework, the team will need datasets for training the risk model, open-source code to accelerate development, and foundational research to guide the architectural design.

8.1 Acquiring and Simulating Data

Real-world user behavior and fraud data is highly sensitive and not publicly available. The team must therefore use public transaction fraud datasets as a base and simulate the additional required signals.

- **Public Datasets for Fraud Detection:**
  - **Kaggle:** The **IEEE-CIS Fraud Detection** dataset is a large-scale, real-world dataset of e-commerce transactions with fraud labels, making it an excellent choice for training a core anomaly detection or classification model.[91] The **Synthetic Financial Fraud Detection Dataset** provides mobile money transactions and is another valuable resource.[92]
  - **Amazon FDB (Fraud Dataset Benchmark):** This is a curated collection of publicly available datasets covering various fraud types, which can be useful for exploring different fraud patterns.[93]
- **Strategy for Simulating Behavioral & Contextual Data:** The key to demonstrating the adaptive framework is to create synthetic signals that can be fed into the risk engine.
  - **Process:** Start with a transaction dataset like the IEEE-CIS one. For each unique user ID, create a "normal profile" by calculating baseline statistics (e.g., average transaction amount, most common login locations, typical time of day). Then, for transactions labeled as fraudulent, programmatically introduce anomalies into the synthetic signal columns. For example, for a fraudulent transaction, set the geolocation to a country different from the user's profile, set the sim_swap_flag to True, and generate a behavioral_biometric_score that is a significant outlier compared to a simulated normal distribution for that user. This allows the team to train the

risk model to recognize these combined signals as high-risk.

8.2 GitHub Repositories for Accelerated Development

Leveraging existing open-source projects is crucial for building a complex system within a hackathon's time constraints.

- **Real-Time Anomaly/Fraud Detection Systems:**
  - **Paulescu/end-2-end-real-time-ml** [94]: A valuable tutorial for building a real-time fraud detection system from the ground up, covering the setup of a feature pipeline using Docker.
  - Tutorials on building real-time systems using **Apache Kafka** (for data streaming) and **Apache Spark** (for real-time processing) provide the architectural blueprint for the signal ingestion and risk scoring layers.[95]
- **Anomaly Detection Libraries:**
  - **yzhao062/pyod**: A comprehensive Python library for outlier detection. It contains implementations of numerous algorithms (e.g., Isolation Forest, Autoencoders) that can be used for the risk scoring model.
- **Behavioral Biometrics (Open-Source Examples):** While commercial solutions are complex, several open-source projects provide a proof-of-concept for capturing and analyzing behavioral data.
  - **Agisthemantobeat/OpenAI-GPT-Powered-Behavioral-Biometrics** [96]: This repository contains a simple but effective Python script for capturing keystroke dynamics (key hold time, release time, typing speed) and training a Random Forest classifier on them. This is an excellent starting point for building a basic behavioral biometric component.
  - **IIEHU/Behavioral-Biometrics-Collection** [97]: This project includes an Android application for collecting sensor data from a phone's accelerometer and gyroscope. It demonstrates the practical methods for capturing device handling and motion data.
  - **mikeroyal/Biometrics-Guide** [98]: A curated guide that lists a wide range of open-source tools, libraries, and algorithms related to biometrics, which can be a useful reference.

8.3 Foundational Research Papers

Grounding the project in established research will provide credibility and a deeper understanding of the underlying principles.

- **On SIM Swap Fraud:**
  - "**An Empirical Study of Wireless Carrier Authentication for SIM Swaps**" (IEEE) [61]: This is a seminal academic paper that provides empirical evidence of the vulnerabilities in mobile carrier authentication procedures that enable SIM swap attacks. It is a critical piece of literature for justifying the need to move beyond SMS-based security.
  - "**A Study of the Emerging Trends in SIM Swapping Crime and Effective Countermeasures**" [99]: This paper provides a broader overview of SIM swap attack trends and discusses various countermeasures, including the potential for AI-based detection systems.
- **On Behavioral Biometrics:**
  - "**User Identification and Authentication using Multi-Modal Behavioral Biometrics**" [75]: A key research paper demonstrating that fusing data from multiple behavioral modalities (keyboard, mouse, GUI) leads to significantly higher authentication accuracy than using a single modality. This supports the proposed architecture of ingesting multiple behavioral signals.
  - "**Multi-Modal Behavioural Biometric Authentication for Mobile Devices**" [76]: This paper focuses specifically on mobile devices and shows that combining modalities like keystroke dynamics and linguistic profiling improves performance, directly relevant to the mobile banking context.
- **On Adaptive Authentication Architecture:**
  - "**Adaptive Authentication and NIST AAL: Building a Secure and Resilient Cybersecurity Framework**" [83]: This paper discusses the architecture of adaptive systems, including the process of feature engineering from authentication logs and training machine learning models to act as risk assessment engines.
  - Various industry articles and white papers provide high-level architectural diagrams and best practices for implementing risk-based authentication by combining signals from location, device, and user behavior to inform authentication decisions.[80]

Conclusion: A Unified Strategy for a Winning Prototype

This report has dissected two of the most pressing challenges in modern financial technology—financial inclusion and digital security—and has laid out a strategic blueprint for developing innovative, competitive, and comprehensive solutions. The proposed architectures for an **Explainable AI (XAI) Credit Scoring Engine** and an **Adaptive Authentication Framework** are designed not as isolated solutions, but as components of a unified, data-driven approach to understanding and protecting the digital customer.

The true strategic advantage lies in the synergy between these two systems. The rich alternative data and behavioral signals gathered to create a more inclusive credit risk model—such as cash flow patterns, digital footprint characteristics, and even psychometric indicators—are the very same signals that can be repurposed as features for the adaptive authentication risk engine. A user's financial habits are intrinsically linked to their digital behavior. By leveraging this unified data pool, a financial institution can build a holistic, 360-degree view of its customers. This allows for a powerful feedback loop: a deeper understanding of a customer's creditworthiness enhances the ability to recognize their legitimate behavior, and a more robust security profile protects the very financial access that the credit model enables.

Ultimately, the most successful hackathon projects will be those that move beyond a narrow focus on a single algorithm or feature. A high-accuracy predictive model is expected; a secure login is a baseline requirement. The winning edge will be demonstrated by teams that present a cohesive platform built on a clear strategic vision. This vision must prioritize transparency through explainability, ensure fairness by actively mitigating bias, and deliver a superior user experience by making security both intelligent and invisible. By building a prototype that embodies these principles, a team can showcase not just technical proficiency, but a mature understanding of the business, regulatory, and ethical landscape of modern finance.**P**

Works cited

1. Topic.pdf
2. Traditional Vs. Alternative Credit Scoring Methods - RiskSeal, accessed July 26, 2025, https://riskseal.io/blog/what-is-alternative-credit-scoring-and-how-does-it-differ-from-the-traditional
3. Alternative Credit Scoring Fintech & Alternative Credit Data - Django Stars, accessed July 26, 2025, https://djangostars.com/blog/alternative-credit-scoring/

4. Criticism of credit scoring systems in the United States - Wikipedia, accessed July 26, 2025, https://en.wikipedia.org/wiki/Criticism_of_credit_scoring_systems_in_the_United_States
5. How traditional credit scoring can be a barrier for many consumers, accessed July 26, 2025, https://www.kansascityfed.org/ten/how-traditional-credit-scoring-can-be-a-barrier-for-many-consumers/
6. The Limitations of Traditional Credit Scoring Systems - Medium, accessed July 26, 2025, https://medium.com/@phindulo60/the-limitations-of-traditional-credit-scoring-systems-e92833fdfa8a
7. Why is the traditional credit rating system losing steam? - Credolab, accessed July 26, 2025, https://www.credolab.com/blog/why-is-the-traditional-credit-rating-system-losing-steam
8. 6 types of alternative credit data for better loan decisions | Plaid, accessed July 26, 2025, https://plaid.com/resources/lending/alternative-credit-data/
9. The Use of Alternative Data in Credit Risk Assessment: Opportunities, Risks, and Challenges - World Bank Documents and Reports, accessed July 26, 2025, https://documents1.worldbank.org/curated/en/099031325132018527/pdf/P179614-3e01b947-cbae-41e4-85dd-2905b6187932.pdf
10. Types of Alternative Data for Credit Scoring | Brankas, accessed July 26, 2025, https://www.brankas.com/types-of-alternative-data-for-credit-scoring
11. The Top 20 Alternative Data Providers for Credit Risk Analysis, accessed July 26, 2025, https://riskseal.io/blog/top-alternative-data-providers-that-serve-the-credit-industry
12. Alternative Data in Financial Services | Congress.gov, accessed July 26, 2025, https://www.congress.gov/crs-product/IF11630
13. Alternative Data For Credit Scoring: Enhancing Credit Scoring - Eagle Alpha, accessed July 26, 2025, https://www.eaglealpha.com/2024/05/08/alternative-data-for-credit-scoring/
14. Sourcing new data for richer credit-risk decisions - Visa, accessed July 26, 2025, https://corporate.visa.com/content/dam/VCOM/corporate/services/documents/vca-sourcing-new-data-for-credit-risk-vf.pdf
15. How to Use Alternative Data in Credit Risk Analytics - FICO, accessed July 26, 2025, https://www.fico.com/blogs/how-use-alternative-data-credit-risk-analytics
16. Alternative credit data 101: What it is and what it's used for - Stripe, accessed July 26, 2025, https://stripe.com/resources/more/alternative-credit-data-101-what-it-is-and-what-its-used-for
17. can psychometric credit-scoring address collateral constraints for women entrepreneurs? - World Bank Documents and Reports, accessed July 26, 2025, https://documents1.worldbank.org/curated/en/099440203162337439/pdf/IDU046

d2582b04f4b047d6086a408f375dfc12ae.pdf

18. Psychometric Credit Scoring in Indonesia Microfinance Industry: A Case Study in PT Amartha Mikro Fintek - ResearchGate, accessed July 26, 2025, https://www.researchgate.net/publication/333809157_Psychometric_Credit_Scoring_in_Indonesia_Microfinance_Industry_A_Case_Study_in_PT_Amartha_Mikro_Fintek
19. Character Counts: Psychometric-Based Credit Scoring for Underbanked Consumers - MDPI, accessed July 26, 2025, https://www.mdpi.com/1911-8074/17/9/423
20. Character Counts: Psychometric-Based Credit Scoring for Underbanked Consumers, accessed July 26, 2025, https://www.researchgate.net/publication/384321330_Character_Counts_Psychometric-Based_Credit_Scoring_for_Underbanked_Consumers
21. max-fitzpatrick/Credit-scoring-model: Credit scoring machine learning algorithm which predicts probability of default - GitHub, accessed July 26, 2025, https://github.com/max-fitzpatrick/Credit-scoring-model
22. Real-Time Fraud Detection Using Machine Learning - Scientific Research Publishing, accessed July 26, 2025, https://www.scirp.org/journal/paperinformation?paperid=133190
23. GitHub - nafiul-araf/Credit-Risk-Modeling-End-to-End-Project, accessed July 26, 2025, https://github.com/nafiul-araf/Credit-Risk-Modeling-End-to-End-Project
24. Explainable AI in finance - University of Twente Student Theses, accessed July 26, 2025, http://essay.utwente.nl/100969/1/Pantov_BA_EEMCS.pdf
25. Explainable AI Policy: It Is Time to Challenge Post Hoc Explanations, accessed July 26, 2025, https://www.cigionline.org/documents/2696/no.296.pdf
26. Exploring Explainable AI in the Financial Sector: Perspectives of Banks and Supervisory Authorities - ResearchGate, accessed July 26, 2025, https://www.researchgate.net/publication/357756214_Exploring_Explainable_AI_in_the_Financial_Sector_Perspectives_of_Banks_and_Supervisory_Authorities
27. Shapley values as an interpretability technique in credit scoring - Journal of Risk Model Validation, accessed July 26, 2025, https://www.risk.net/journal-of-risk-model-validation/7958697/shapley-values-as-an-interpretability-technique-in-credit-scoring
28. Algorithmic decision making methods for fair credit scoring - arXiv, accessed July 26, 2025, https://arxiv.org/html/2209.07912
29. Debiasing Alternative Data for Credit Underwriting Using Causal Inference - arXiv, accessed July 26, 2025, https://arxiv.org/pdf/2410.22382
30. A Systematic Review of Explainable AI in Finance Abstract 1. Introduction - arXiv, accessed July 26, 2025, https://arxiv.org/pdf/2503.05966
31. Written evidence submitted by the Working Group on fAIr Credit, Credit Research Centre, University of Edinburgh - UK Parliament Committees, accessed July 26, 2025, https://committees.parliament.uk/writtenevidence/140316/pdf/
32. Explainable AI in finance - University of Twente Student Theses, accessed July 26, 2025, http://essay.utwente.nl/100969/
33. shap/shap: A game theoretic approach to explain the output ... - GitHub,

accessed July 26, 2025, https://github.com/shap/shap

34. An Introduction to SHAP Values and Machine Learning Interpretability - DataCamp, accessed July 26, 2025, https://www.datacamp.com/tutorial/introduction-to-shap-values-machine-learning-interpretability
35. Mastering LIME for Deep Learning Models - Number Analytics, accessed July 26, 2025, https://www.numberanalytics.com/blog/mastering-lime-deep-learning-models
36. Machine Explainability: A Guide to LIME, SHAP, and Gradcam | by Suryansh Raghuvanshi, accessed July 26, 2025, https://suryansh-raghuvanshi.medium.com/machine-explainability-a-guide-to-lime-shap-and-gradcam-60f6265f365f
37. www.irjmets.com, accessed July 26, 2025, https://www.irjmets.com/uploadedfiles/paper//issue_3_march_2025/71478/final/fin_irjmets1744210205.pdf
38. Algorithmic discrimination in the credit domain: what do we know about it? - DSpace@MIT, accessed July 26, 2025, https://dspace.mit.edu/bitstream/handle/1721.1/150785/146_2023_Article_1676.pdf?sequence=1&isAllowed=y
39. Debiasing Alternative Data for Credit Underwriting Using Causal Inference - arXiv, accessed July 26, 2025, https://arxiv.org/html/2410.22382v1
40. Debiasing Alternative Data for Credit Underwriting Using Causal Inference - arXiv, accessed July 26, 2025, https://arxiv.org/pdf/2410.22382?
41. German Credit Risk - Kaggle, accessed July 26, 2025, https://www.kaggle.com/datasets/uciml/german-credit
42. Credit Card Approval Prediction - Kaggle, accessed July 26, 2025, https://www.kaggle.com/datasets/rikdifos/credit-card-approval-prediction
43. Credit Score - Kaggle, accessed July 26, 2025, https://www.kaggle.com/datasets/conorsully1/credit-score
44. Credit Score Classification | Kaggle, accessed July 26, 2025, https://www.kaggle.com/competitions/1056lab-credit-score-classification
45. Credit_Scoring_Data | Kaggle, accessed July 26, 2025, https://www.kaggle.com/datasets/cs49adityarajsharma/credit-scoring-data
46. Credit score classification - Kaggle, accessed July 26, 2025, https://www.kaggle.com/datasets/parisrohan/credit-score-classification
47. Pratik94229/Credit-Card-Default-Prediction-End-to-End-Project - GitHub, accessed July 26, 2025, https://github.com/Pratik94229/Credit-Card-Default-Prediction-End-to-End-Project
48. fedeghigo/Credit-Risk.Machine-learning-application.: This work aim to, on one side, the backend code develop on Python Notebook (code_credit_scoring_np.ipynb), for reason of readability, to find the best Hyperparameter in order to achieve the best forecast for the credit scoring with algorithm we analize during - GitHub, accessed July 26, 2025, https://github.com/fedeghigo/Credit-Risk.Machine-learning-application.

49. credit-scoring · GitHub Topics, accessed July 26, 2025, https://github.com/topics/credit-scoring?l=jupyter+notebook
50. MRobalinho/ML-Credit_Score: Using machine learning to create a credit score to customers, accessed July 26, 2025, https://github.com/MRobalinho/ML-Credit_Score
51. This is ML project which is based on Classification of Credit Score - GitHub, accessed July 26, 2025, https://github.com/Prem07a/Credit-Score-Classification
52. interpretml/interpret: Fit interpretable models. Explain blackbox machine learning. - GitHub, accessed July 26, 2025, https://github.com/interpretml/interpret
53. cloudera/CML_AMP_Explainability_LIME_SHAP: Learn how to explain ML models using LIME and SHAP. - GitHub, accessed July 26, 2025, https://github.com/cloudera/CML_AMP_Explainability_LIME_SHAP
54. Fair and unbiased algorithmic decision making: Current state and future challenges - EconStor, accessed July 26, 2025, https://www.econstor.eu/bitstream/10419/227696/1/1675751455.pdf
55. [2503.05966] Explaining the Unexplainable: A Systematic Review of Explainable AI in Finance - arXiv, accessed July 26, 2025, https://arxiv.org/abs/2503.05966
56. [2002.09931] The Value of Big Data for Credit Scoring: Enhancing Financial Inclusion using Mobile Phone Data and Social Network Analytics - arXiv, accessed July 26, 2025, https://arxiv.org/abs/2002.09931
57. Alternative Credit Scoring of Micro-, Small and Medium-sized Enterprises - Hong Kong Monetary Authority, accessed July 26, 2025, https://www.hkma.gov.hk/media/eng/doc/key-functions/financial-infrastructure/alternative_credit_scoring.pdf
58. Understanding SIM Swap Attacks: Detection, Risks & Protection - Group-IB, accessed July 26, 2025, https://www.group-ib.com/resources/knowledge-hub/sim-swap/
59. How does SIM-based device binding prevent phishing and fraud? - IDlayr, accessed July 26, 2025, https://idlayr.com/blog/device-binding-prevent-fraud/
60. An Empirical Study of Wireless Carrier Authentication for SIM Swaps - USENIX, accessed July 26, 2025, https://www.usenix.org/system/files/soups2020-lee.pdf
61. An Empirical Study of Wireless Carrier Authentication for SIM Swaps, accessed July 26, 2025, https://www.ieee-security.org/TC/SPW2020/ConPro/papers/lee-conpro20.pdf
62. What Is Behavioral Biometrics? - BioCatch, accessed July 26, 2025, https://www.biocatch.com/blog/what-is-behavioral-biometrics
63. Sim Binding | Mohit Kanwar's App : My Cents, accessed July 26, 2025, https://mohitkanwar.com/blogs/sim-binding/
64. Mobile Fraud Detection: Types, Techniques & Best Practices - FOCAL, accessed July 26, 2025, https://www.getfocal.ai/blog/mobile-fraud-detection
65. Fraud Detection With Machine Learning: 5 Steps to Build One - Designveloper, accessed July 26, 2025, https://www.designveloper.com/guide/fraud-detection-machine-learning/
66. How Does Bureau Leverage Location Intelligence to Detect Fraud , accessed July 26, 2025,

https://www.bureau.id/blog/how-does-bureau-leverage-location-intelligence-to-detect-fraud

67. What is Geolocation and How It Helps with Fraud Prevention - Decentro, accessed July 26, 2025, https://decentro.tech/blog/geolocation/
68. What Is Behavioral Biometrics: How Does It Work Against Fraud - Feedzai, accessed July 26, 2025, https://www.feedzai.com/blog/behavioral-biometrics-next-generation-fraud-prevention/
69. What is Behavioral Biometrics - LexisNexis Risk Solutions, accessed July 26, 2025, https://risk.lexisnexis.com/insights-resources/article/what-is-behavioral-biometrics
70. What is Behavioral Biometrics? - OneSpan, accessed July 26, 2025, https://www.onespan.com/topics/behavioral-biometrics
71. Behavioral Biometrics Authentication: Use Cases and Benefits - ASEE Cybersecurity, accessed July 26, 2025, https://cybersecurity.asee.io/blog/what-is-behavioral-biometrics-authentication/
72. Behavioral authentication for security and safety, accessed July 26, 2025, https://sands.edpsciences.org/articles/sands/full_html/2024/01/sands20230028/sands20230028.html
73. Here's How Behavioral Biometrics Work With FIDO Standards - Mobile ID World, accessed July 26, 2025, https://mobileidworld.com/behavioral-biometrics-with-fido-standards-903084/
74. Biometric Authentication: Advantages and Disadvantages - miniOrange, accessed July 26, 2025, https://www.miniorange.com/blog/biometric-authentication-advantages-disadvantages/
75. User Identification and Authentication using Multi-Modal Behavioral Biometrics | Request PDF - ResearchGate, accessed July 26, 2025, https://www.researchgate.net/publication/261102469_User_Identification_and_Authentication_using_Multi-Modal_Behavioral_Biometrics
76. Multi-Modal Behavioural Biometric Authentication for Mobile Devices | Request PDF, accessed July 26, 2025, https://www.researchgate.net/publication/266615172_Multi-Modal_Behavioural_Biometric_Authentication_for_Mobile_Devices
77. Passkeys: Passwordless Authentication - FIDO Alliance, accessed July 26, 2025, https://fidoalliance.org/passkeys/
78. Discover Secure & Convenient Passwordless Authentication, accessed July 26, 2025, https://www.pingidentity.com/en/resources/identity-fundamentals/authentication/passwordless-authentication.html
79. Passkeys - Google for Developers, accessed July 26, 2025, https://developers.google.com/identity/passkeys
80. FIDO Authentication: Why This Security Professional Believes It's the Future of Digital Identity | by Abhijith Soman | Medium, accessed July 26, 2025, https://medium.com/@4bhijithsoman/fido-authentication-why-this-security-prof

essional-believes-its-the-future-of-digital-identity-0063394db316

81. What is Adaptive Authentication and How Does It Work? - LoginTC, accessed July 26, 2025, https://www.logintc.com/types-of-authentication/adaptive-authentication/
82. Advanced Authentication Techniques - Number Analytics, accessed July 26, 2025, https://www.numberanalytics.com/blog/advanced-authentication-techniques-network-security
83. (PDF) Adaptive Authentication and NIST AAL: Building a Secure and Resilient Cybersecurity Framework - ResearchGate, accessed July 26, 2025, https://www.researchgate.net/publication/389167014_Adaptive_Authentication_and_NIST_AAL_Building_a_Secure_and_Resilient_Cybersecurity_Framework
84. What is Adaptive Authentication? How does it work? - AuthX, accessed July 26, 2025, https://www.authx.com/blog/what-is-adaptive-authentication/
85. Adaptive Authentication | Entrust IAM Portfolio, accessed July 26, 2025, https://www.entrust.com/products/iam/capabilities/adaptive-authentication
86. Adopt Adaptive Authentication: 8 Best Practices for Secure Access - TrustBuilder, accessed July 26, 2025, https://www.trustbuilder.com/en/adaptive-authentication-best-practices-zero-trust/
87. How machine learning works for payment fraud detection and prevention - Stripe, accessed July 26, 2025, https://stripe.com/resources/more/how-machine-learning-works-for-payment-fraud-detection-and-prevention
88. Anomaly detection for fraud prevention - Advanced strategies, accessed July 26, 2025, https://www.fraud.com/post/anomaly-detection
89. Fraud Detection with Machine Learning: Identifying Suspicious Patterns in Financial Transactions | by Zhong Hong | Medium, accessed July 26, 2025, https://medium.com/@zhonghong9998/fraud-detection-with-machine-learning-identifying-suspicious-patterns-in-financial-transactions-8558f3f1e22a
90. Real-Time Fraud Detection — ML System Design | by Ajay L R Sharma - Medium, accessed July 26, 2025, https://medium.com/@ajaylrsharma/real-time-fraud-detection-ml-system-design-4379d925f7ef
91. IEEE-CIS Fraud Detection | Kaggle, accessed July 26, 2025, https://www.kaggle.com/competitions/ieee-fraud-detection
92. Financial Fraud Detection Dataset - Kaggle, accessed July 26, 2025, https://www.kaggle.com/datasets/sriharshaeedala/financial-fraud-detection-dataset
93. FDB: Fraud Dataset Benchmark - Amazon Science, accessed July 26, 2025, https://www.amazon.science/code-and-datasets/fdb-fraud-dataset-benchmark
94. Paulescu/end-2-end-real-time-ml: Let's build a real time fraud detection system using TurboML - GitHub, accessed July 26, 2025, https://github.com/Paulescu/end-2-end-real-time-ml
95. Build Realtime Fraud Detection AI from Scratch - End to End Machine Learning

Project - Part 1 - YouTube, accessed July 26, 2025, https://www.youtube.com/watch?v=ve5xTvsvots&pp=0gcJCfwAo7VqN5tD

96. Agisthemantobeat/OpenAI-GPT-Powered-Behavioral-Biometrics: Behavioral biometrics is a cutting-edge field that focuses on analyzing user behavior to enhance security and authentication processes. This repository serves as a hub for everything related to this innovative technology. Read more at - GitHub, accessed July 26, 2025, https://github.com/Agisthemantobeat/OpenAI-GPT-Powered-Behavioral-Biometrics
97. IIEHU/Behavioral-Biometrics-Collection: This is the behavioral biometric data collection App implementation of the paper Behavioral Biometrics-based Continuous Authentication Using A Lightweight Latent Representation Masked One-Class Autoencoder - GitHub, accessed July 26, 2025, https://github.com/IIEHU/Behavioral-Biometrics-Collection
98. mikeroyal/Biometrics-Guide - GitHub, accessed July 26, 2025, https://github.com/mikeroyal/Biometrics-Guide
99. A Study of the Emerging Trends in SIM Swapping Crime and Effective Countermeasures, accessed July 26, 2025, https://www.researchgate.net/publication/364038169_A_Study_of_the_Emerging_Trends_in_SIM_Swapping_Crime_and_Effective_Countermeasures
100. User Identification and Authentication using Multi-Modal Behavioral Biometrics, accessed July 26, 2025, https://scholar.afit.edu/facpub/1148/