definity.

Strategic Planning for Fraud Detection Innovation

Huan He
Enterprise Fraud Data Scientist
2025-05-26

TABLE OF CONTENT

Context

- **Role**: Data Scientist
- **Team**: Enterprise Fraud Analytics & Intelligence
- **Team Status**: In the early stages of building a strong, scalable *fraud detection function*.
- **Vision**: Reimagine fraud detection across the insurance lifecycle.
- **Methods**: Leverage advanced analytics, machine learning, and AI.
- **Priorities**: *Accurate, fast, efficient* fraud detection model, ensure collaboration with *investigative, underwriting, and claims* teams.

Strategic Approach

Understanding Our Starting Point

Understand the Business Context

- Key stakeholders' pain points, expectations, and operational constraints.
- Products and fraud types.

Review Existing Processes & Tools

- Fraud detection workflows
- Cross-functional communication protocols
- Data entry and management procedures.

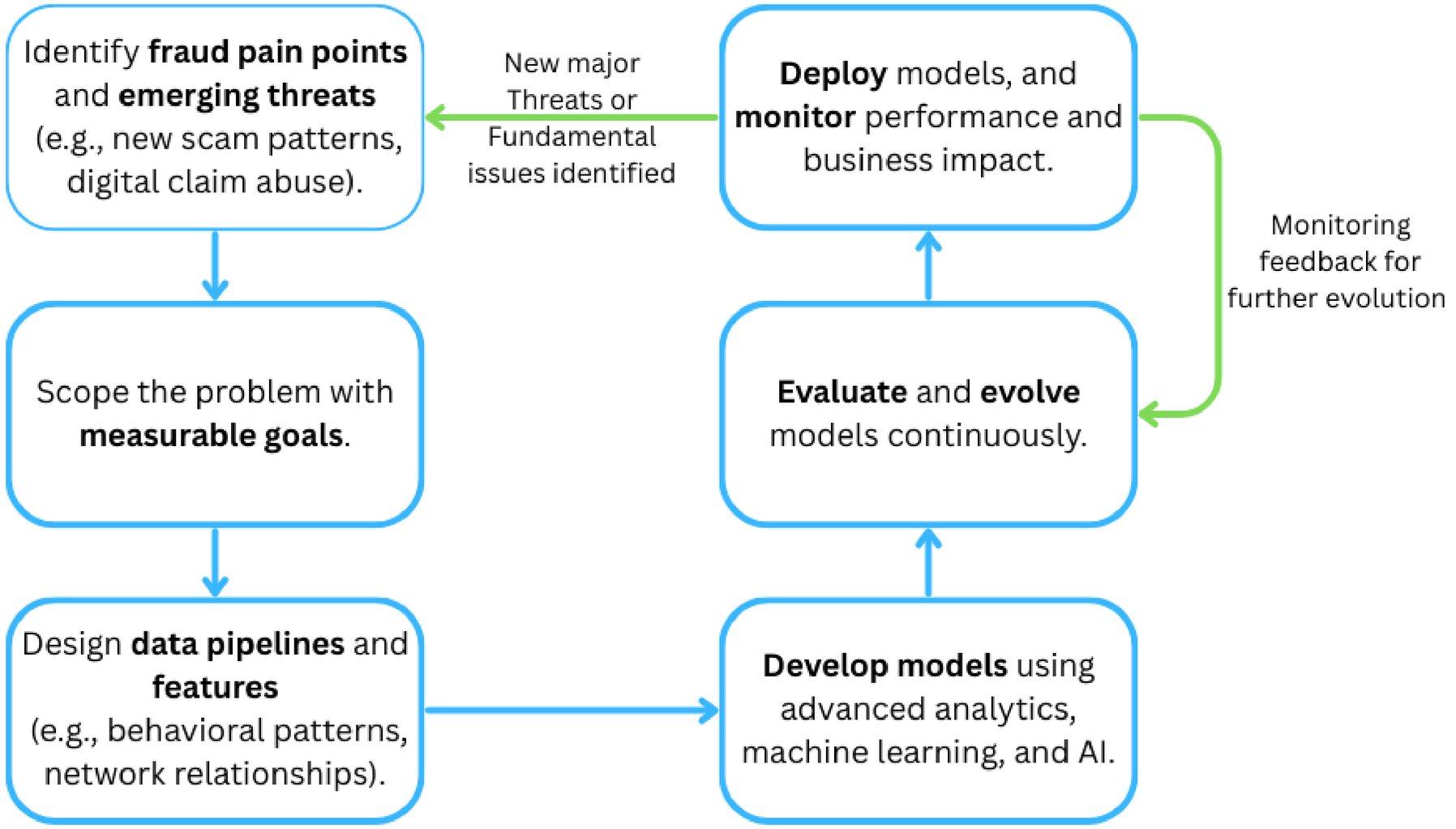
Assess Data & Models

- Data sources, availability, quality, labeling strategy.
- Model type, explainability and interpretability.

Identify Gaps & Opportunities

- Scalability.
- Detection performance.
- Automation potential.

Methodology for Innovation



Balancing Exploration (R&D) with Execution (Impact Delivery)

	Explorations (R&D)	Execution (Impact Delivery)
Goals	Investigate new techniques, technologies, and fraud patterns (long-term).	Deliver measurable, business-aligned outcomes (short-term).
Cycle	Monthly or quarterly planning	Daily or Weekly planning
Methods	Explore cutting-edge AI/ML approaches (e.g., GenAI, advanced NLP applications).	Implement well-understood ML techniques, optimize existing rules, improve integration with operations, etc.
Effort	20% ~ 30%	70% ~ 80%

Roadmap The 60-120-180 Day Plan

01 Learning and Landscape Assessment

- Comprehensive understanding.
- Build relationships.
- Pinpoint quick wins.

02 Experimentation & Value Validation

- Test hypotheses.
- Quantify business value.
- Showcase ML/AI benefits.

03 Scaling & Strategic Positioning

- Operationalize successful solutions.
- Strengthen feedback loops.
- Establish long-term contribution.

Phase 1: Learning & Landscape Assessment

Access Requirements

- **Data**: Info from claims, policies, underwriting, customers, and investigations; past fraud cases, investigation notes, fraud alerts, risk scores, and outside data sources.
- **Tools**: Database systems, programming environments, data visualization tools, and version control systems/code repositories.
- **People**: Enterprise Fraud Analytics & Intelligence, investigative, underwriting, claims, IT and Legal/Compliance teams.
- **Processes**: How fraud is currently detected, how claims are handled, underwriting rules, investigation steps, and data management policies.

Critical Questions

- What are the most challenging types of insurance fraud?
- What are the most significant pain points in the fraud detection process and information flow?
- What are the baseline metrics for accuracy, speed of detection, operational efficiency?
- What are the perceived technical or organizational barriers to adopting more advanced analytical techniques or AI/ML solutions for fraud detection?

Quick Wins

- Automate currently manual, repetitive data extraction or reporting tasks to free up analyst time.
- Conduct an analysis of historical confirmed fraudulent claims to identify common characteristics.
- Develop interactive dashboards.

Phase 2: Experimentation and Value Validation

Hypotheses to Explore

- We can increase fraud detection accuracy using a supervised ML model trained on historical confirmed fraud cases.
- Unstructured data (e.g., claim notes, adjuster comments) holds untapped signals for fraud detection.
- Certain rules are triggering too frequently and overwhelming investigators with false positives.

Success Metrics

- **Model Performance**: Recall, Precision, F1-Score, and Area Under the ROC Curve (AUC-ROC).
- **Operational Impact**: Reduction in false positives, increase in the true positive detection rate, a decrease in the average time taken to detect fraud, and estimated cost savings from prevented fraudulent payouts.
- **Stakeholder Feedback**: Qualitative input from investigators and claims adjusters regarding the utility, actionability, and interpretability of the outputs.

Cross-Organizational Collaboration

- **Investigative team** to validate results and get real-world feedback.
- **Underwriting team** to improve upfront risk assessment processes.
- **Claims team** to ensure integration into their workflows.
- **IT team** to help with pipelines.
- **Legal/Compliance** if sensitive features like demographic data are involved.

Phase 3: Scaling and Strategic Positioning

Scaling Learnings

- Deploy validated model into production.
- Automate data pipelines.
- Build a feedback loop to capture case outcomes and retrain models.
- Create thorough documentation for all developed models, associated codebases, and operational processes.

Strategic Contributions

- Mature fraud detection models with a proven return on investment.
- Automated reporting dashboards that provide actionable insights to business users.
- Refined alert generation systems that integrate smoothly into investigative workflows.

Long-Term Embedding

- Cultivate deep expertise in specific fraud domains.
- Regularly search for emerging technologies, novel data sources, and innovative methodologies.
- Actively share knowledge, mentor junior team members.
- Consistently ensure all analytical work and proposed initiatives are clearly linked to the business objectives of the company.
- Design and implement robust feedback loops to enhance the performance of fraud detection capabilities.

Technology, Tools, Technique

Supervised Learning

- **Techniques:** Gradient Boosting Machines (e.g., XGBoost), Random Forests, Logistic Regression.
- **Rationale:** Highly effective for classification tasks when sufficient labeled historical fraud data is available.

Structured Data

Unsupervised Learning

- **Techniques:** Isolation Forests, One-Class Support Vector Machines (SVM), Clustering algorithms (e.g., K-Means, DBSCAN).
- **Rationale:** Essential for identifying novel or unusual patterns that do not conform to known fraud signatures.

Structured Data

Natural Language Processing

- **Techniques:** Hugging Face Transformers, LLMs (GPT-4 Turbo in retrieval mode).
- **Rationale:** NLP unlocks insights from unstructured text, revealing nuanced indicators of potential fraud that structured data alone cannot capture.

Unstructured Data

Building a Fraud Model from Scratch

1. Problem Definition & Goal Setting
2. Data Collection and Preparation
3. Exploratory Data Analysis (EDA)
4. Feature Engineering
5. Model Selection

6. Model Training & Validation
7. Thresholding & Scoring
8. Interpretability & Explainability
9. Deployment & Monitoring
10. Feedback Loop & Iteration

Innovation Suggestion

Network Analysis is a method used to study the relationships and interactions between entities (nodes) connected by links (edges). These entities can be people, organizations, accounts, devices, etc., and the connections between them represent interactions like communications, transactions, or shared attributes.

Why Network Analysis?

- Increased Accuracy.
- Improved Speed of Detection.
- Enhanced Operational Efficiency

A policyholder claims an accident involving a repair shop and a medical clinic. Network analysis shows:

- The same repair shop is involved in 40% of suspicious claims.
- The same doctor signs off on most of them.
- All claimants are connected via a social network.

Key Takeaways

- Reiterate the vision for a transformed fraud detection function.
- Emphasize the strategic, phased approach (60-120-180 plan).
- Highlight the power of combining proven ML with innovative techniques like Network Analysis.
- Reinforce commitment to accuracy, speed, efficiency, and collaboration.

THANK YOU