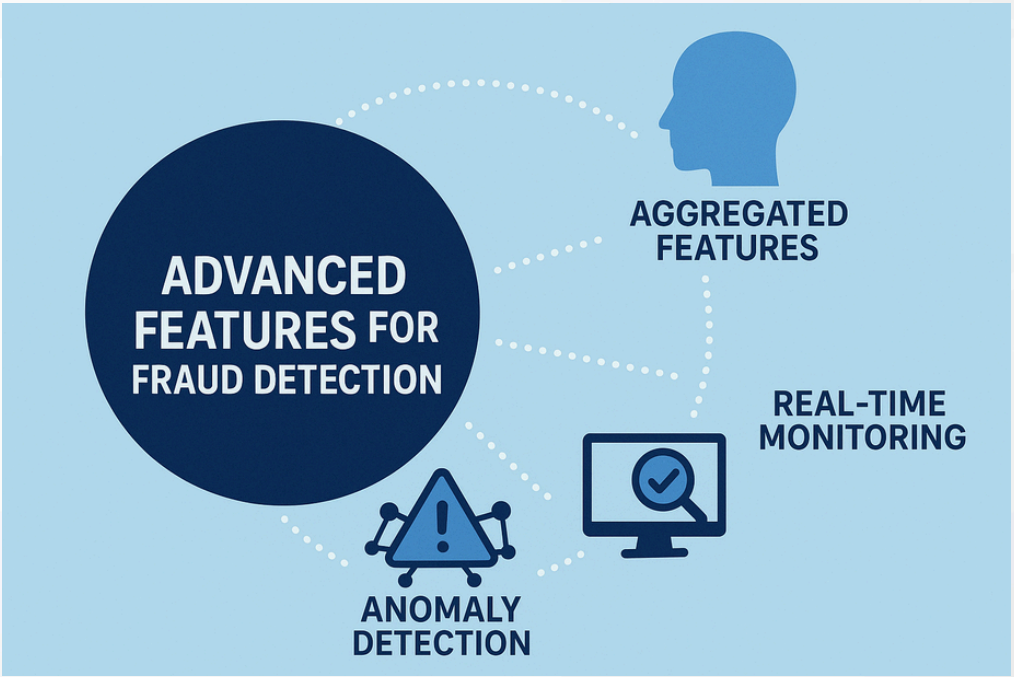
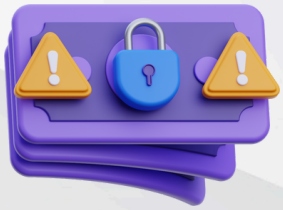
Feature Engineering for Fraud: Crafting the Signals that AI Can Understand - Part 2

- While fundamental and channel-specific features lay the groundwork for fraud detection, the true power to unearth sophisticated and evolving fraudulent activities lies in **advanced feature engineering.**
- This crucial stage goes beyond the obvious, crafting new, highly informative features that capture **intricate patterns, temporal nuances, and the complex web of relationships hidden** within transaction data.



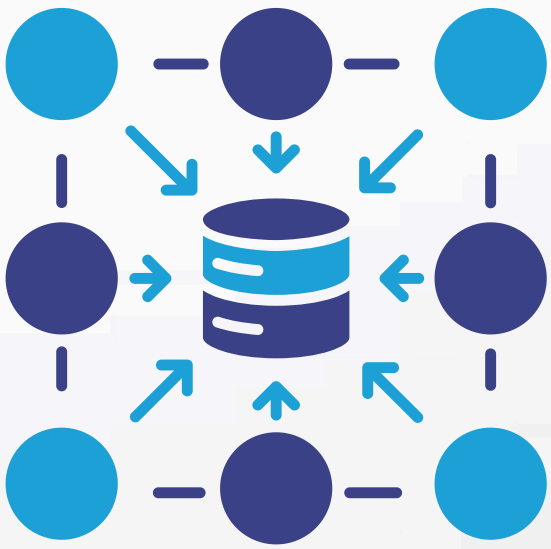


Feature Engineering for Fraud: Crafting the Signals that AI Can Understand - Part 2

- Data scientists, armed with analytical prowess and domain expertise, employ a range of sophisticated techniques to illuminate the unseen, creating signals that AI can understand and act upon.
- Think of **advanced feature engineering** as moving beyond individual clues to understanding the broader narrative of a potential crime.
- It involves not just observing a single suspicious transaction but analyzing its context within a user's history, the wider network of transactions, and temporal trends.

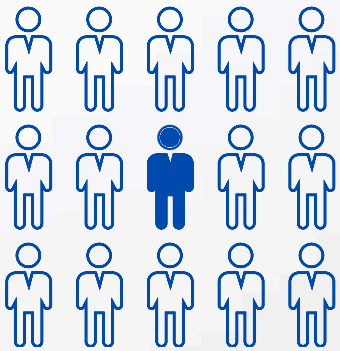
Crafting Deeper Insights: Advanced Techniques in Action

- **Aggregated Features**: Painting a Historical Picture. Summarizing a user's past behavior over defined time windows provides a vital historical context for evaluating the legitimacy of a current transaction. These aggregated features highlight deviations from established norms, offering a powerful lens for anomaly detection.
- **Transaction Count:** Instead of just looking at the current transaction, we can calculate the number of transactions a user has made in the last hour, day, or week. If a user typically makes no more than 5 online purchases a day, a sudden spike to 20 transactions raises a significant red flag.



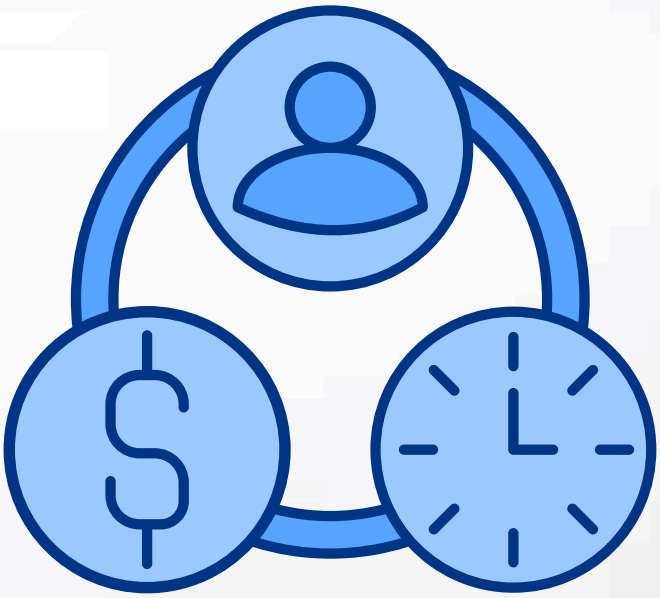
Crafting Deeper Insights: Advanced Techniques in

- **Average Spending:** Calculating the average transaction amount for a user over the past month can help identify transactions that are significantly higher or lower than their usual spending. A user with an average online purchase of $30 attempting a $500 transaction at an unfamiliar electronics store would be flagged due to this deviation. Similarly, a sudden series of micro-transactions could also be anomalous.
- **Distinct Merchants Visited:** Tracking the number of unique merchants a user interacts with within a specific timeframe can reveal unusual behavior. A user who typically shops at 2-3 familiar online retailers showing transactions at 10 different new e-commerce sites in a single day could indicate a compromised account.



Crafting Deeper Insights: Advanced Techniques in Action

- **Time-Series Features:** Unraveling Temporal Anomalies. Explicitly capturing the temporal dimension of transaction data allows AI models to identify patterns and anomalies related to when transactions occur. These features go beyond simple timestamps to understand the rhythm of a user's activity.
- **Recency of Last Transaction:** The time elapsed since a user's last legitimate transaction can be informative. A transaction occurring mere seconds after a failed login attempt or a password reset might be suspicious.

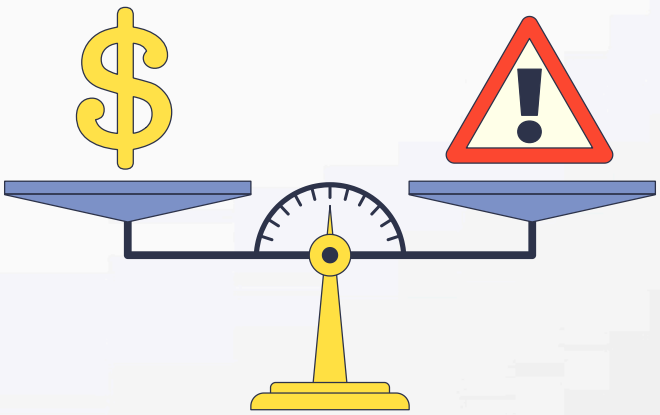


Crafting Deeper Insights: Advanced Techniques in Action

- **Transaction Frequency:** Analyzing the rate of transactions within specific time windows can reveal velocity attacks. A user making multiple high-value transactions within a very short period, such as three $1000 purchases within 5 minutes, is highly unusual for most individuals.
- **Time-of-Day/Day-of-Week Patterns:** Creating features that indicate if a transaction occurs during a user's typical active hours or on unusual days can be revealing. A transaction at 4 AM local time for a user who consistently shops between 9 AM and 5 PM on weekdays is anomalous. Similarly, unusual activity patterns on weekends or holidays could also be flagged.

Crafting Deeper Insights: Advanced Techniques in Action

- **Interaction Features**: Uncovering Hidden Synergies. Combining two or more basic or even previously engineered features can uncover synergistic effects that individual features might miss. These interactions can reveal non-linear relationships that are strong indicators of fraud.
- **Amount x Merchant Risk:** Multiplying the transaction amount by a pre-calculated risk score associated with the Merchant Category Code (MCC) can create a more nuanced risk indicator. A $200 transaction at a high-risk MCC (e.g., online electronics retailer with a history of fraud) might be flagged as higher risk than a $500 transaction at a low-risk MCC (e.g., a well-established grocery store).

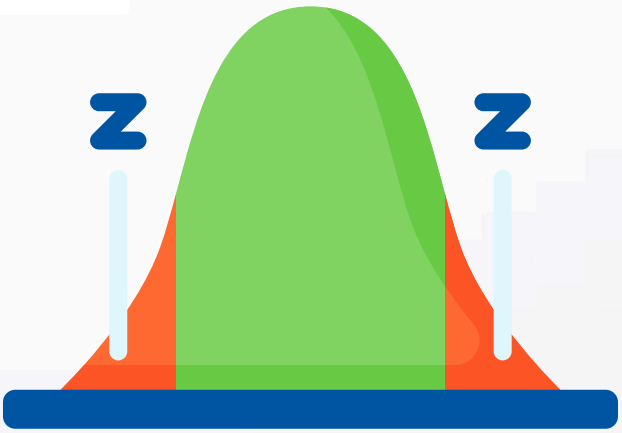


Crafting Deeper Insights: Advanced Techniques in Action

- **Location Discrepancy** x **Time Since Last Login:** Combining the geographic distance between the current transaction and the user's last known location with the time elapsed since their last account login can highlight suspicious scenarios. A transaction originating from a foreign country just minutes after a user logged in from their home IP address could indicate account takeover.
- **Transaction Amount Change Rate:** Calculating the percentage change in transaction amount compared to the user's previous few transactions can highlight sudden and drastic shifts in spending behavior. A 500% increase in transaction amount compared to the user's average over the last week could be a strong fraud signal.

Crafting Deeper Insights: Advanced Techniques in Action

- **Anomaly/Deviation Features**: Quantifying Unusualness. These features directly measure how "out of the ordinary" a transaction is, either in the context of a single user's history or compared to the broader population of users.
- **Z-Score of Transaction Amount:** Calculating the Z-score of the transaction amount relative to the user's historical average and standard deviation can quantify how many standard deviations away from their norm the current transaction is. A transaction with a Z-score of +5 or -5 would be a significant outlier.



Crafting Deeper Insights: Advanced Techniques in Action

- **Ratio to Median Price:** Comparing the transaction amount to the median purchase price for the specific Merchant Category Code (MCC) or for the user's past purchases within that category can highlight unusually priced items. A transaction for significantly more than the typical price for that type of product at that merchant could be suspicious.
- **Peer Group Comparison:** Comparing a user's current transaction behavior (e.g., spending at a specific type of merchant) to that of a defined peer group (users with similar demographics and spending habits) can identify deviations. If a user suddenly starts spending heavily at online gambling sites, while their peer group shows no such activity, it could be a red flag.

Crafting Deeper Insights: Advanced Techniques in Action

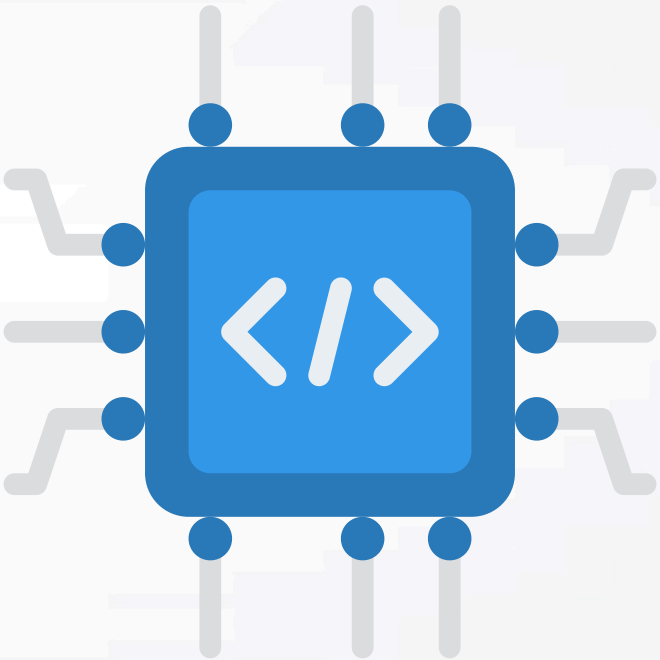
- **Network-based Features**: Mapping the Web of Deception. Representing transactions and entities (users, merchants, devices) as a graph and deriving features from this network structure can be invaluable for detecting sophisticated fraud rings and collusive behavior that might be invisible when analyzing transactions in isolation.
- **Shared IP Addresses/Devices:** Identifying multiple different user accounts originating transactions from the same unusual IP address or device could indicate a coordinated fraud attempt. If five different credit cards are used to make purchases all originating from the same suspicious IP address in a foreign country, it suggests a potential fraud ring.

Crafting Deeper Insights: Advanced Techniques in Action

- **Transaction Links:** Analyzing patterns of transactions between specific users and merchants can reveal suspicious connections. A newly created user making multiple high-value transactions with a merchant known for facilitating fraudulent activities could be flagged.
- **Community Detection:** Algorithms can identify clusters or communities within the transaction graph. Users or merchants belonging to a dense cluster of high-risk activity might be indicative of a fraud ring. Identifying a group of users who frequently transact with each other at a set of obscure online retailers could reveal a coordinated scheme.

Crafting Deeper Insights: Advanced Techniques in Action

- **Graph Embeddings:** Using Graph Neural Networks to learn low-dimensional vector representations (embeddings) of users, merchants, and devices can capture complex relational information. Users with similar embedding vectors might exhibit similar fraudulent behaviors, even if their direct transaction patterns appear different.



Summary

- Mastering these advanced feature engineering techniques requires a deep understanding of both the intricacies of transaction data and the multifaceted ways in which fraud can manifest.
- Data scientists act as skilled artisans, carefully crafting these powerful signals that empower AI models to see beyond the obvious and effectively combat the ever-evolving threat of credit and debit card fraud.
- Their ability to hypothesize, create, and rigorously validate these advanced features is a critical differentiator in the ongoing fight to protect the financial ecosystem.

Special Thanks to ChatGPT and Gemini for Content support