



Machine Learning and the Evolution of Employee Fraud Detection

David Speights, PhD
Chief Data Scientist, Appriss

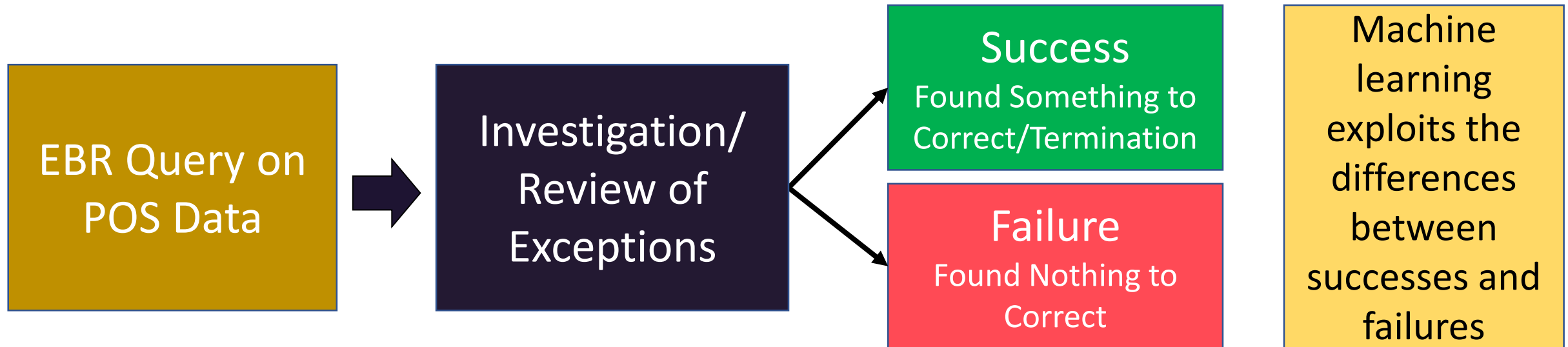
May 2017

Agenda



- Current state of exception reporting
- Collective intelligence – combining the knowledge gained from many retailers
- What is machine learning, predictive modeling, artificial intelligence, and deep learning?
- Parallels to credit card fraud detection

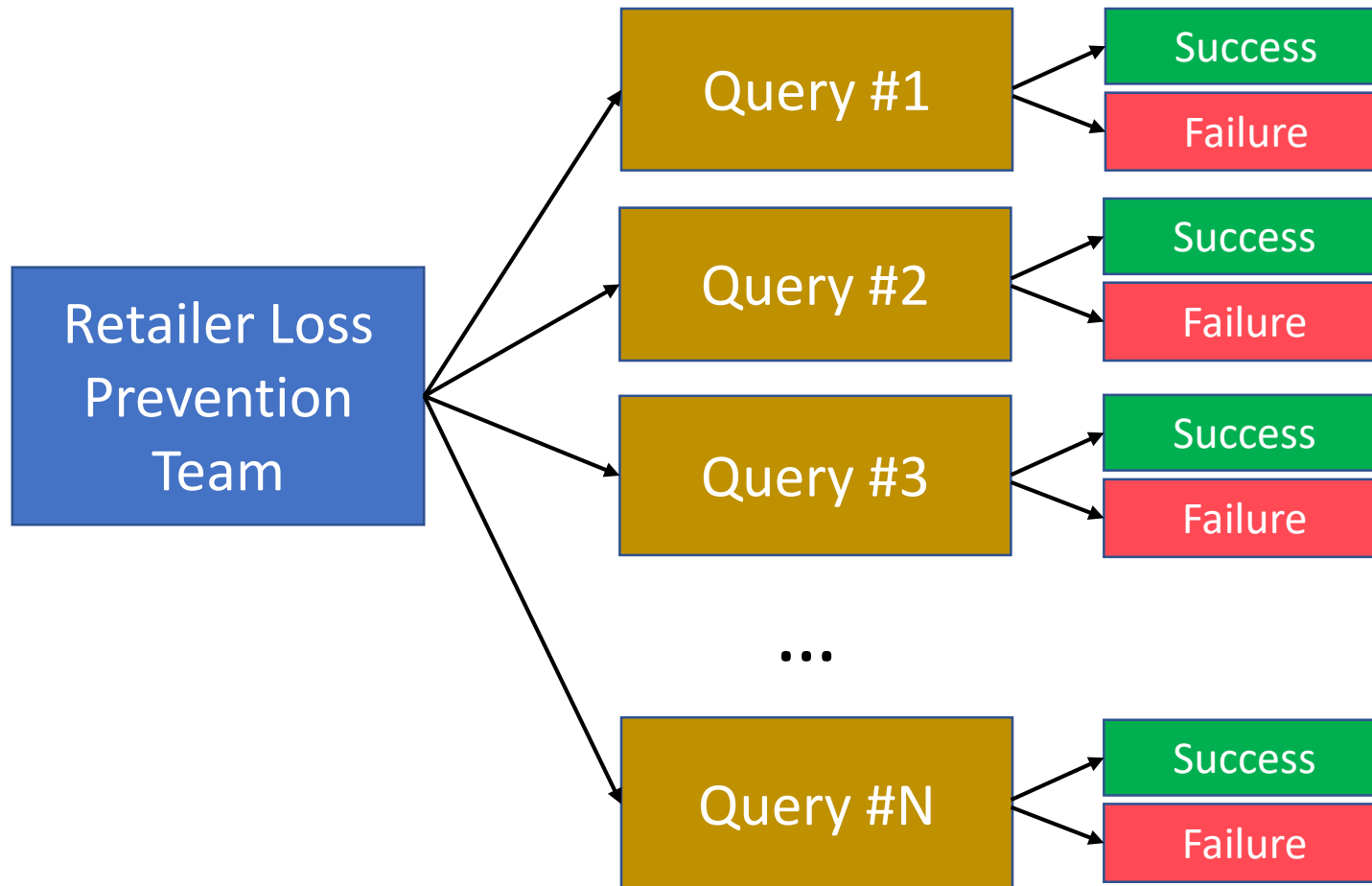
Current Flow for EBR Investigations for a Single Query



- Queries tend to focus on:
 - Transaction anomalies, unusual sequences of transactions, etc.
 - Usually based on a small number of variables
- Machine learning will use hundreds of variables to identify patterns of fraud based on experience from many different retailers and has the ability to learn/evolve over time as new fraud scams arise

Loss Prevention Teams will have a Collection of Queries

Looking for Fraudulent or Problematic Employee Activity

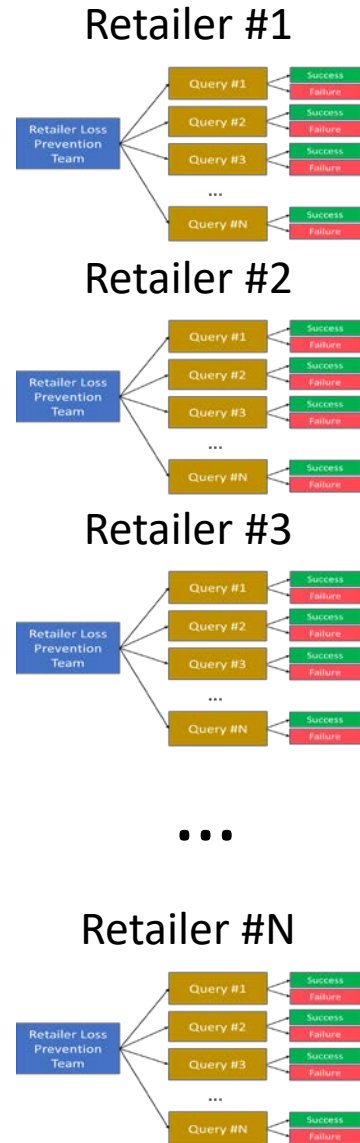


How this feeds into machine learning

- Queries are converted into a collection of variables
- Query based variables are combined with many other variables:
 - Other retailers successful queries
 - SRAs (sales reducing activities)
 - Control variables: tenure, position, store location

Collective Intelligence

The Best Machine Learning Models will Use Many Retailer's Strategies



Combining retailer strategies leads to better detection

- Variables which result from shared questions
- Models built on many successful strategies

Appriss Retail's Base for Collective Intelligence

Sysrepublic + The Retail Equation Serves 90+ Retailers



Advance
Auto Parts

Bath
& Body
Works

Dobbies
GARDEN CENTRES
Brighten your day

ICA

Michaels
Where Creativity Happens™

Sainsbury's

TESCO

ANN TAYLOR

BEST
BUY

DELHAIZE

IKEA

NASCAR

Schnucks

TESCO
Lotus

Argos

bloomingdale's

EVERY SEASON STARTS AT
DICK'S
SPORTING GOODS

jcp

NIKE

7
ELEVEN

THE
HOME
DEPOT

ASDA

BON-TON
COME TO THE RIGHT PLACE®

Finish Line

JUMBO
supermarkten

RadioShack®

SMYTHS

Wegmans

AutoZone®

Burlington
coat factory

GAP

KOHL'S

RITE
AID

Stein Mart®

WinCo
FOODS

AVIS

charlotte
russe

GYMBOREE

كارفور
Carrefour

ROSS
DRESS FOR LESS

SCOTMID
co-operative

WH
S

BARNES & NOBLE
BOOKSELLERS

coles

Hallmark

YOUR M&S

SAFEWAY
Ingredients for life.™

Toys R Us

VICTORIA'S
SECRET

Developing a Machine Learning System



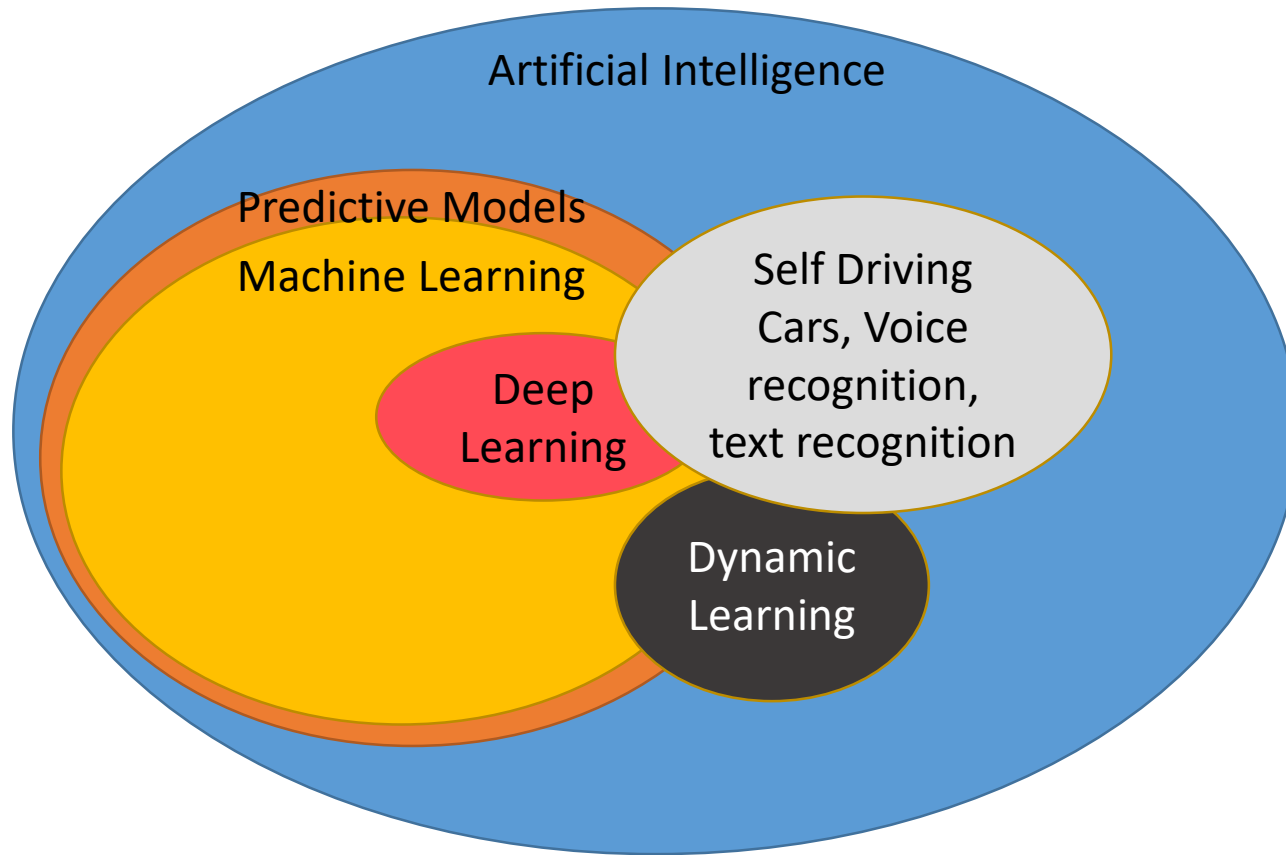
Data Breadth

- Number of employees in database = 5+ Million
- Number of transactions = 45+ Billion
- Number of stores = 100,000+
- Incarceration data and the relationship with Appriss Public Safety = 125+ Million Incarcerations (US)

Model Development Efforts for Machine Learning Model Development

- Currently soliciting retailers to be involved in testing/model refinement
- 3 major U.S. retailers now participating
- Number of terminated employees used in modeling 10,000
- 3,000+ variables available for modeling
- Slated for general availability in late 2017

Artificial Intelligence, Machine Learning, and Predictive Models...



Artificial Intelligence (AI)

- Computer software/algorithms that emulate human intelligence or make automated decisions
- Predictive models and machine learning are tools/subsets of AI

Predictive Models

- Using data to estimate models which make predictions
- Originated by statisticians

Machine Learning

- Using data to "train" models which make predictions
- Originated by computer scientists

Deep learning

- Uses highly complex model structures to learn patterns in the data without any human intervention.
- Very hard to fit these models.

Dynamic/Recursive Learning

- Systems that re-fit or update the machine learning/predictive models dynamically as new data presents itself

Machine learning and Predictive models are very close to being the same thing

From Queries to Machine Learning

Variable Derivation is the Foundation

Objective – create a comprehensive list of model variables which can best predict problematic employees

Survey of Many Retailers top questions



Employee Fraud Study with Retailer #1

Employee Fraud Study with Retailer #2

Created list of core variables for model development

Key Variable Categories

- Voids
- Discounts
- Employee transactions
- Refunds
- Collusion
- No Sales
- Key sequences of transactions

More than 3,000 variables created

Machine Learning Approach

Information Flowchart

Employee Variables

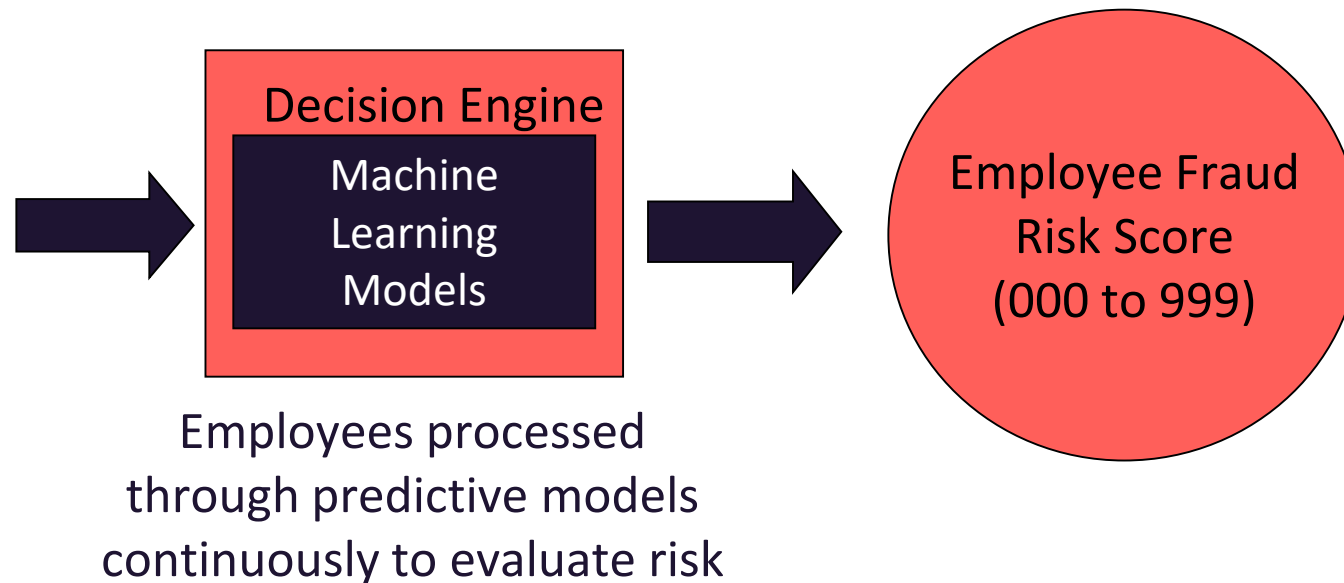
Employee Transactional
Histories

SRAs (Voids, Returns,
Markdowns)

Consumer / Employee
Collusion Patterns

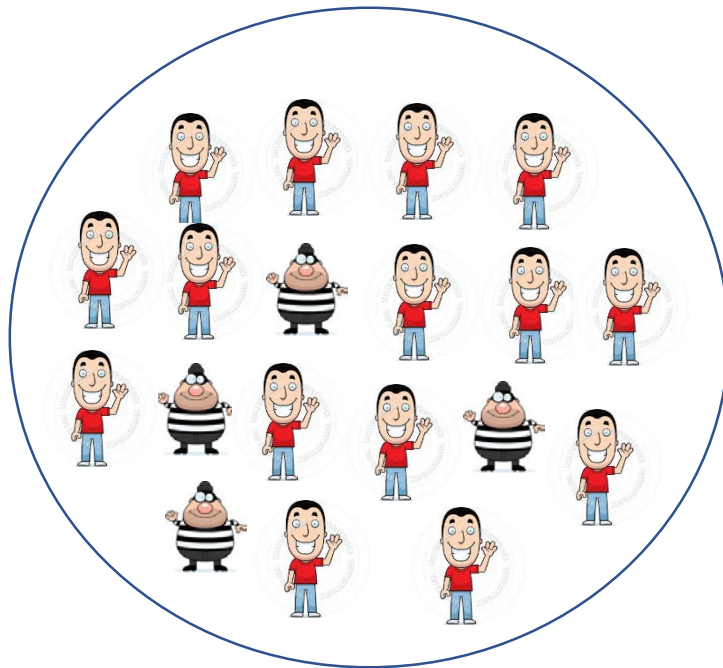
Employee Self Run
Transactions

High Risk Behavioral
Patterns

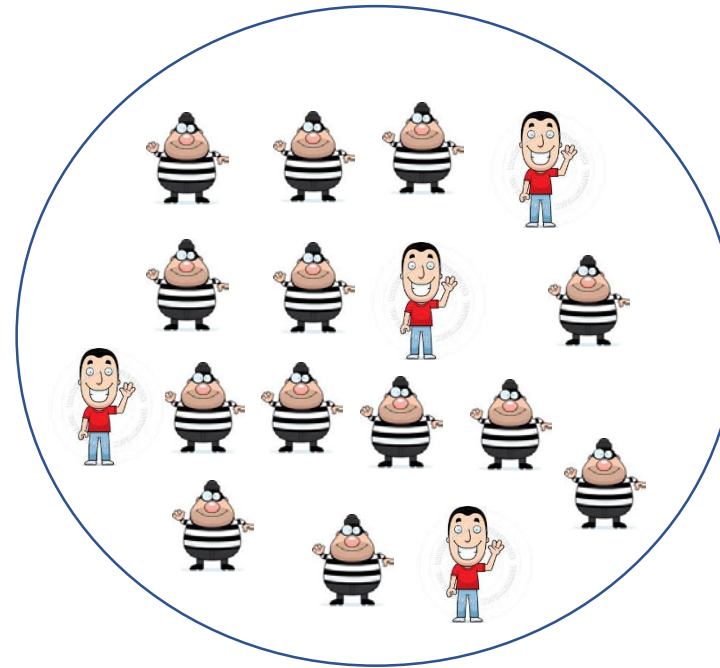


Employee Deviance Detection

Query Based Investigations

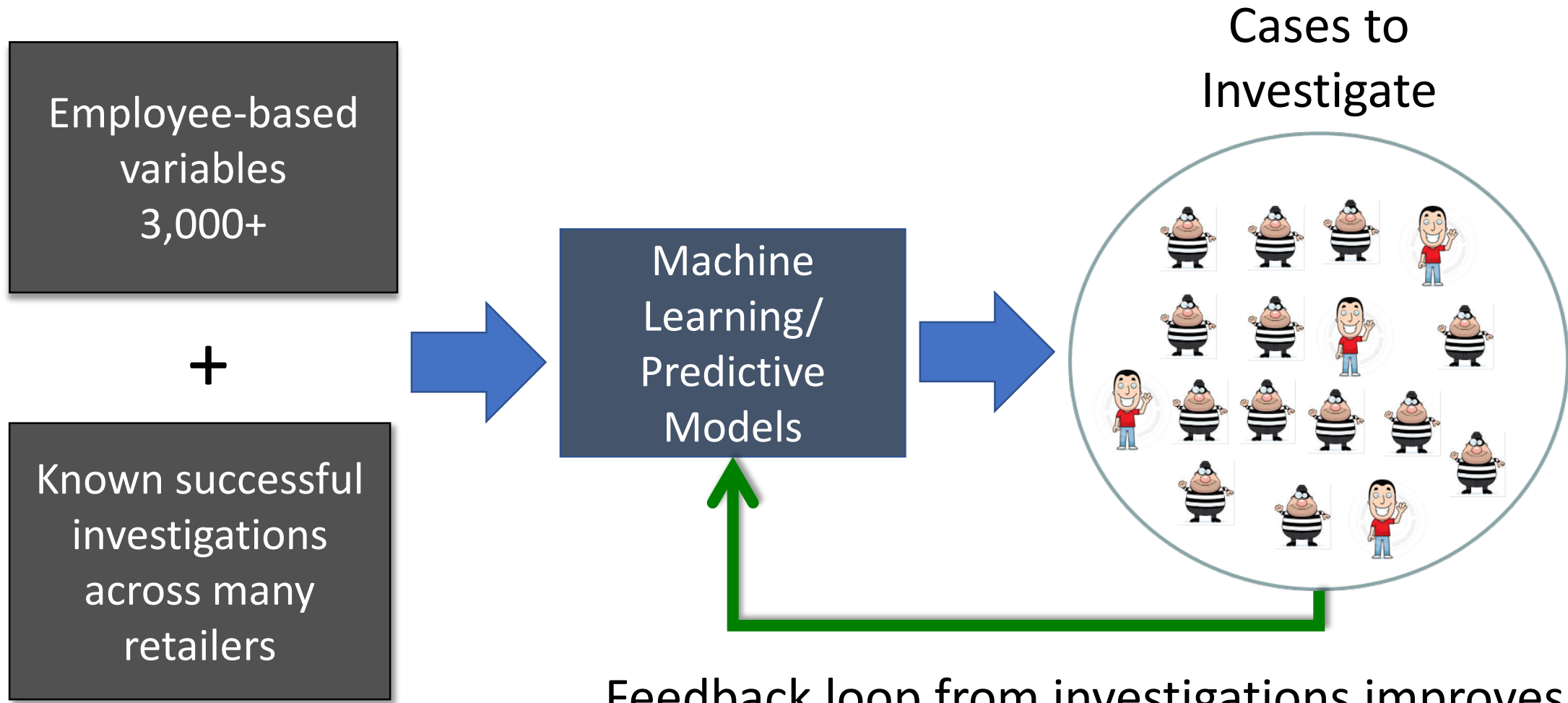


Machine Learning Based Investigations



A machine learning based approach will have a higher success rate than traditional approaches

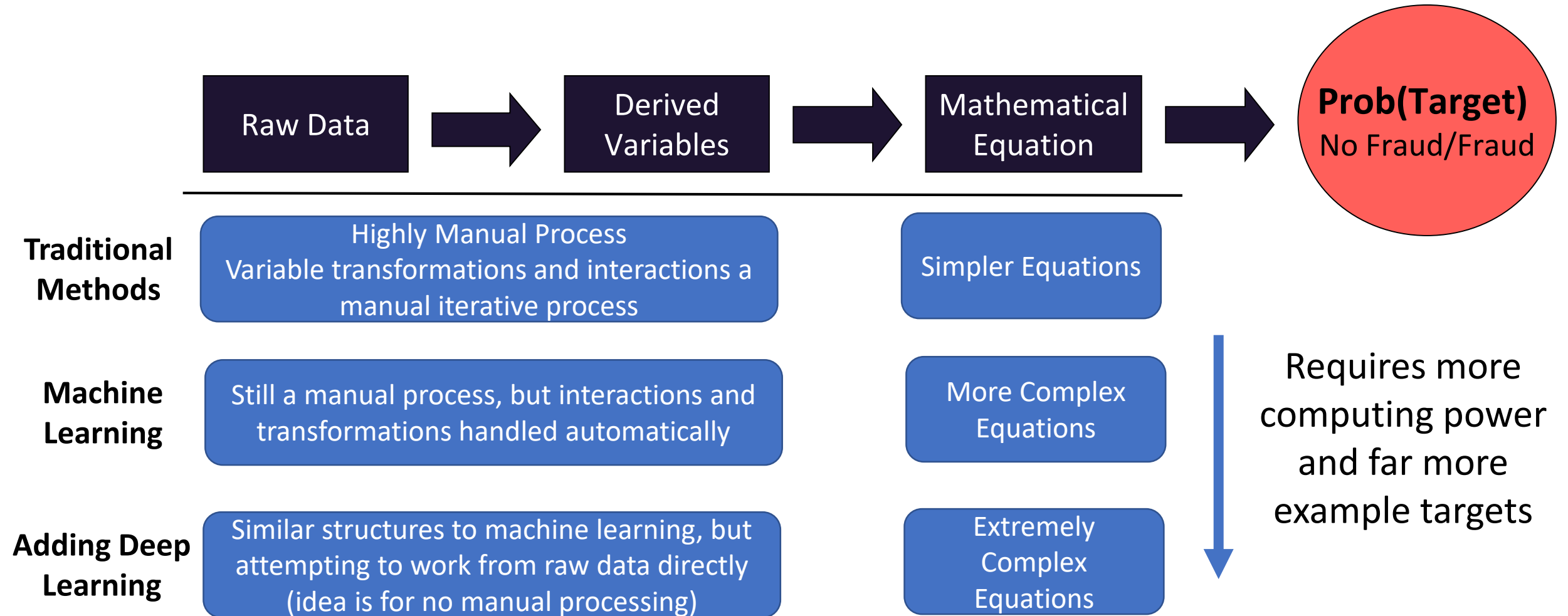
Dynamic Learning



Feedback loop from investigations improves models & keeps up with new fraud

Key Components of Newer Technology

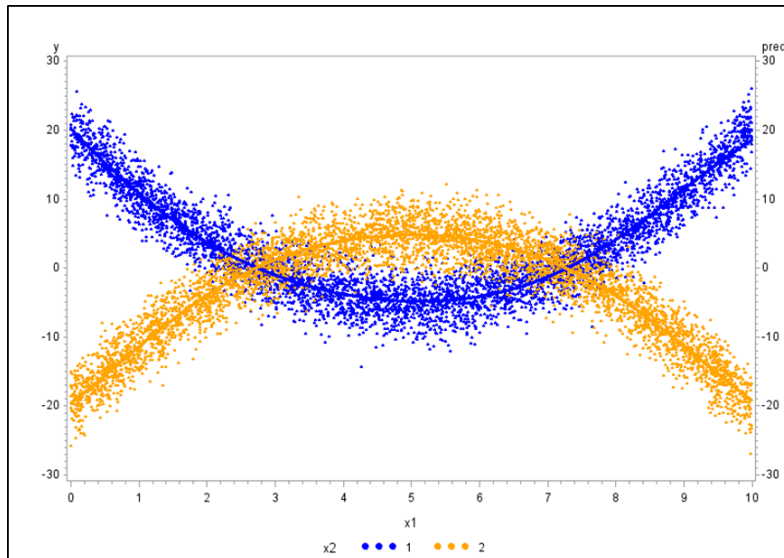
Identifying Variables' Relationship to Target



Machine Learning at Work

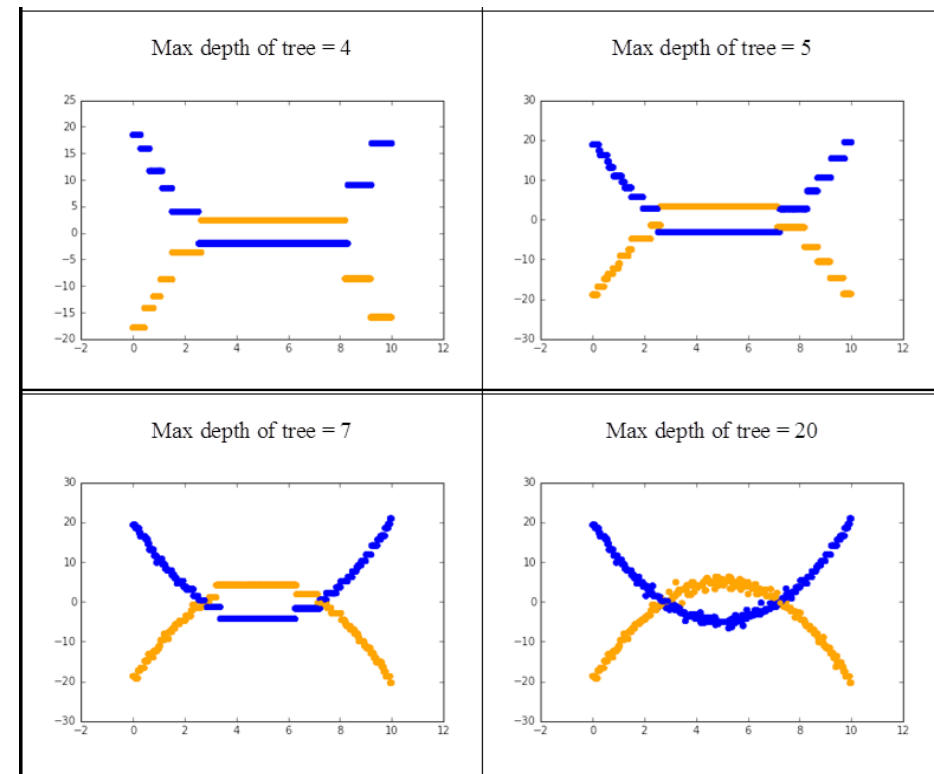
Example Problem Below:

- Non-linear relationship with interaction.
- Traditional methods require analyst to discover these facts

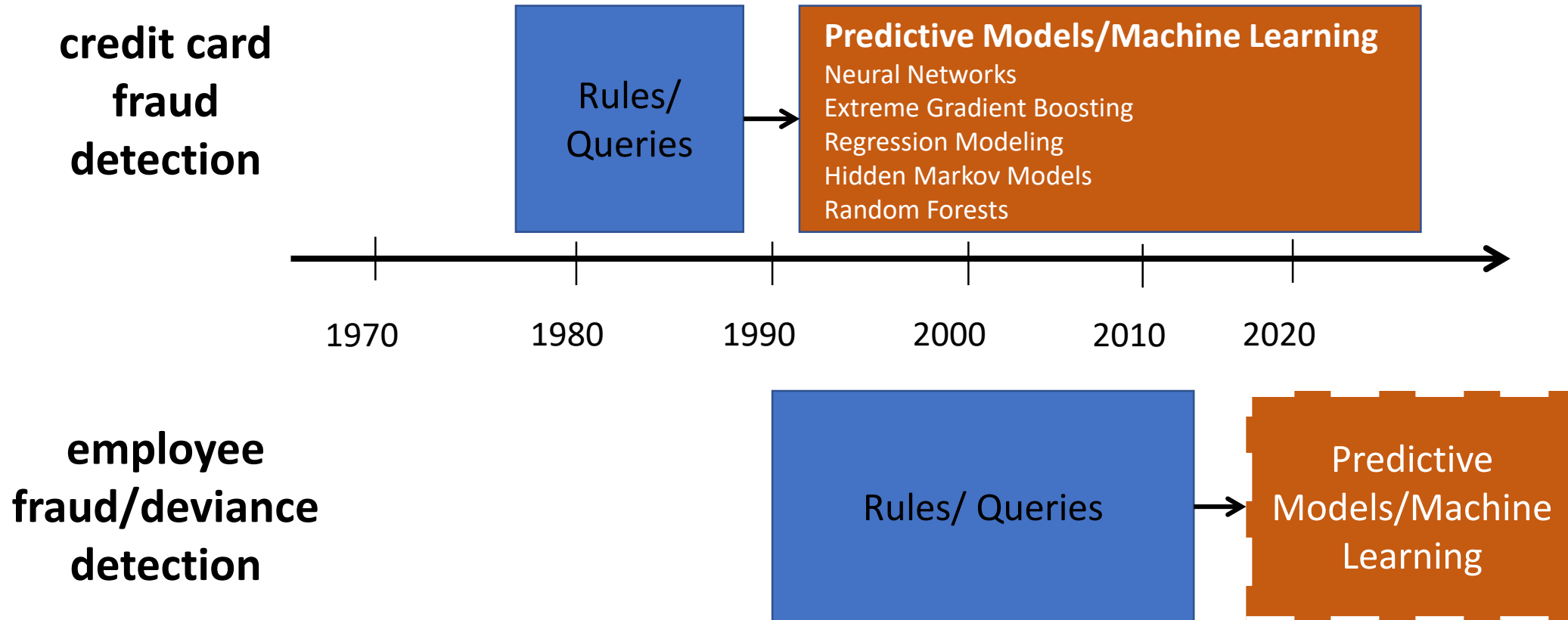


Machine Learning Automatically fits this relationship:

- Machine learning models detect complex relationships automatically



Employee Deviance: Predictive Modeling Evolution (credit card industry versus retail)





Machine Learning and the Evolution of Employee Fraud Detection

David Speights, PhD
Chief Data Scientist, Appriss

May 2017