Ai+ TRAINING

Advanced Fraud Modeling & Anomaly Detection

Part 2

Dr. Aric LaBarr

Associate Professor of Analytics

www.ariclabarr.com

Course Outline – Part 1 & Part 2

- Introduction
- Data Preparation
- Supervised Modeling
- Implementation / Deployment
- Conclusion

Course Outline – Part 1 & Part 2

- Introduction
  - Who am I?
  - What is Fraud?
  - Fraud Detection Analytical Framework
- Data Preparation
- Supervised Modeling
- Implementation / Deployment
- Conclusion

Course Outline – Part 1 & Part 2

- Introduction
- Data Preparation
    - Feature Engineering
    - Fraud Data
    - Anomaly Detection with Statistical Techniques
    - Anomaly Detection with Machine Learning Techniques
    - Sampling Concerns
- Supervised Modeling
- Implementation / Deployment
- Conclusion

Course Outline – Part 1 & Part 2

- Introduction
- Data Preparation
- Supervised Modeling
  - Interpretable Models
  - Naïve Bayes Models
  - More Advanced Models
  - Model Evaluation
  - **NOT**-fraud Model
- Implementation / Deployment
- Conclusion

Course Outline – Part 1 & Part 2

- Introduction
- Data Preparation
- Supervised Modeling
- Implementation Deployment
    - Clustering Revisited
    - Interpretability
    - Long-term Fraud Strategy
    - Chance & Loss Models
- Conclusion

Coding in Action

Example

Introduction

Introduction

- Introduction
  - Who am I?
  - What is Fraud?
  - Fraud Detection Analytical Framework

Introduction

Who Am I?

- Introduction
  - Who am I?
  - What is Fraud?
  - Fraud Detection Analytical Framework

Who Am I?

- 4-time North Carolina State University graduate:
  - BS in Statistics
  - BS in Economics
  - MS in Statistics
  - PhD in Statistics with minor in Economics

Who Am I?

- 4-time North Carolina State University graduate
- Former Senior Data Scientist and Director at Elder Research Inc.
    - Passionate about helping people solve challenges using their data.
    - Mentored a team of data scientists to work closely with clients and partners to solve problems in predictive modeling, advanced analytics, forecasting, and risk management.

Who Am I?

- 4-time North Carolina State University graduate
- Former Senior Data Scientist and Director at Elder Research Inc.
- Associate Professor of Analytics at Institute for Advanced Analytics at NC State University
    - Nation's first master of science in analytics degree program
    - Helped design the innovative program to prepare a modern work force to wisely communicate and handle a data-driven future.
    - Developed and taught courses in statistics, mathematics, finance, risk management, and operations research.

Who Am I?

- 4-time North Carolina State University graduate
- Former Senior Data Scientist and Director at Elder Research Inc.
- Associate Professor of Analytics at Institute for Advanced Analytics at NC State University
- Find me online:
  - https://www.linkedin.com/in/ariclabarr/
  - https://www.youtube.com/c/AricLaBarr/
  - https://www.ariclabarr.com/

Introduction

What is Fraud?

- Introduction
  - Who am I?
  - What is Fraud?
  - Fraud Detection Analytical Framework

What is an Anomaly?

anomaly

noun

/əˈnäməlē/

something that **deviates** from what is **standard**, **normal**, or **expected**

Why Detect Anomalies?

- Anomalies in data can lead to incorrect or out of date decisions to be made.
- Need to find these **outliers** before they become too much of a problem.
- Anomaly detection techniques used in variety of areas:
    - Cleaning data
    - Monitoring health of computer systems
    - Cybersecurity threats
    - Fraudulent claims or transactions

Why Detect Anomalies?

- Anomalies in data can lead to incorrect or out of date decisions to be made.
- Need to find these **outliers** before they become too much of a problem.
- Anomaly detection techniques used in variety of areas:
  - Cleaning data
  - Monitoring health of computer systems
  - Cybersecurity threats
  - Fraudulent claims or transactions

What is Fraud?

fraud

noun

/frôd/

Wrongful or criminal **deception** intended to result in financial or personal **gain**

Fraud Characteristics

1. Uncommon
2. Concealed and trying to be avoided
3. Ever changing and adapting
4. Thought out and organized
5. Doesn’t all look the same

Fraud Problem – Uncommon

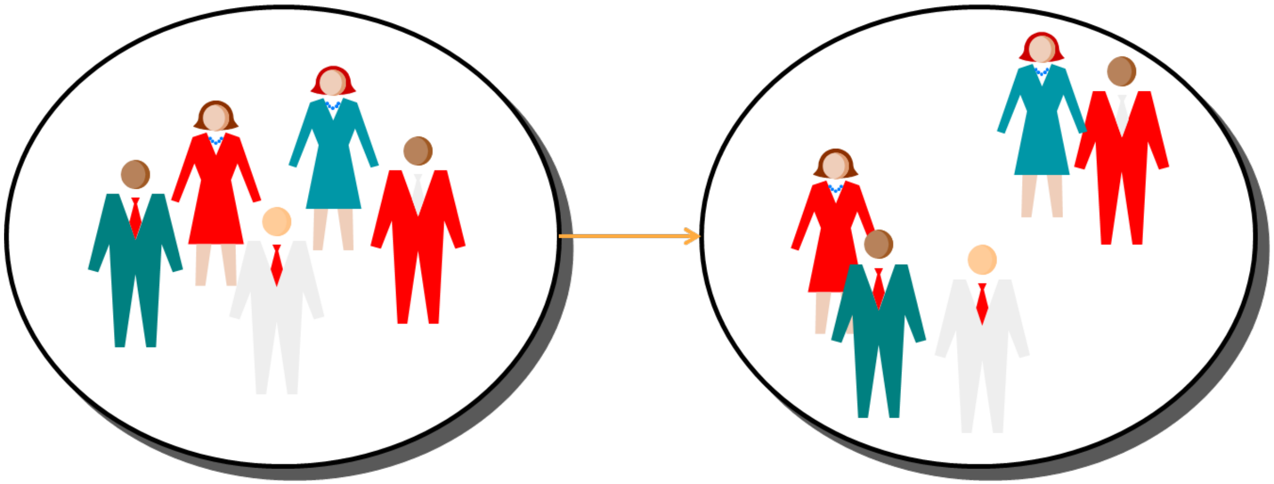
- In 2022, the ACFE (Association of Fraud Examiners) estimated that organizations lose approximately 5% of their revenues to fraud.
- Based on 2022 world GDP (IMF estimates) this would mean approximately $5.08 trillion is lost each year due to fraud.

Fraud Problem – Cat & Mouse Game

- In fraud data sets, observations are **trying to not be analyzed** or discovered – blending in.
  - Planned ahead of time – otherwise easier to detect in modeling.
  - Models have short shelf lives and are adapted often

Fraud Problem – Sociometry

- J L Moreno founded a social science called sociometry, where sociometrists believe that society is made up of individuals and their social, economic, or cultural ties.



Fraud Problem – Sociometry

- J L Moreno founded a social science called **sociometry**, where sociometrists believe that society is made up of individuals and their social, economic, or cultural ties.

- Fraud is often an organized crime.
    - No independence
    - Copycat
    - Homophily: “Birds of a feather flock together.”

Fraud Characteristics

1. Uncommon
2. Concealed and trying to be avoided
3. Ever changing and adapting
4. Thought out and organized
5. Doesn’t all look the same

- Because of these characteristics, fraud is a tough anomaly problem to solve.
- Data science can help aid in this problem!

Introduction

Fraud Detection Analytical Framework

- Introduction
    - Who am I?
    - What is Fraud?
    - Fraud Detection Analytical Framework

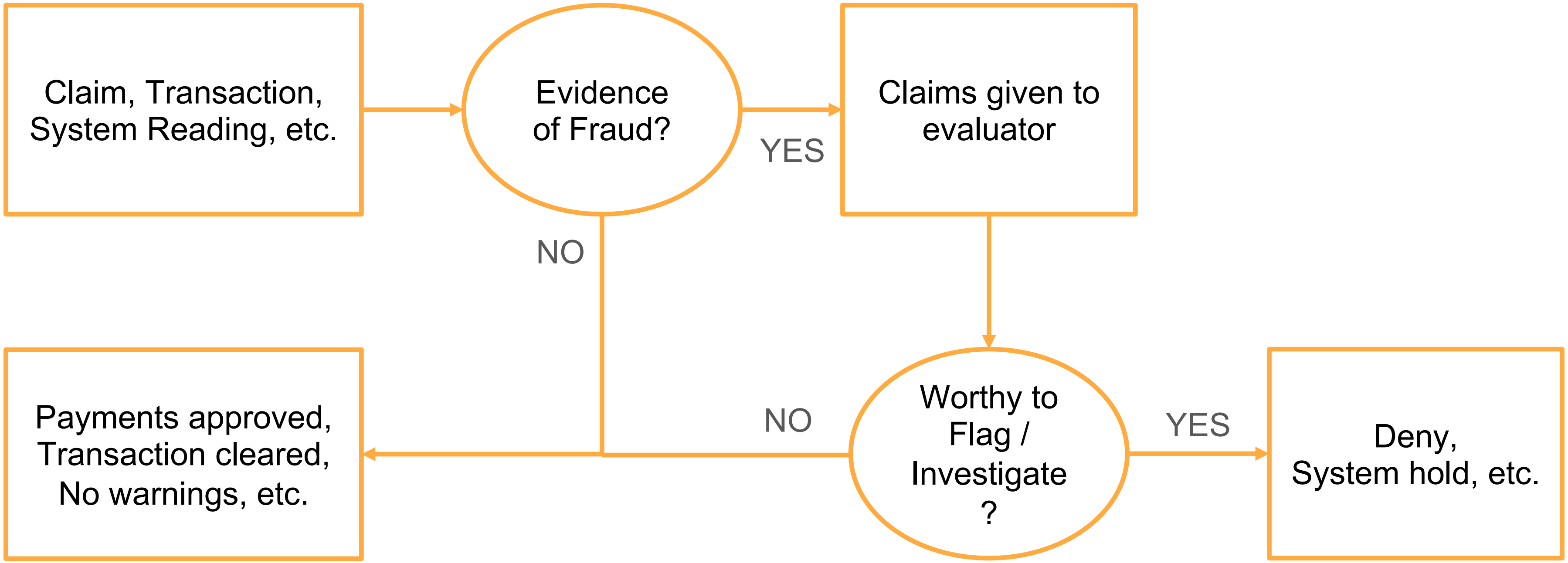
Anomaly Detection Systems

- Regardless of the industry, two things are important for any anomaly detection solution or system:

  1. **DETECTION** – able to identify current anomalies in the system

  2. **PREVENTION** – able to flag potentially new anomalies in the system

Anomaly Detection Systems



Anomaly Detection Maturity – Card Transaction

- New / young anomaly detection solutions are based on **business rules**.
- Example:
    - IF:
        - Amount of transaction above threshold
    - THEN:
        - Flag as suspicious AND
        - Alert evaluator

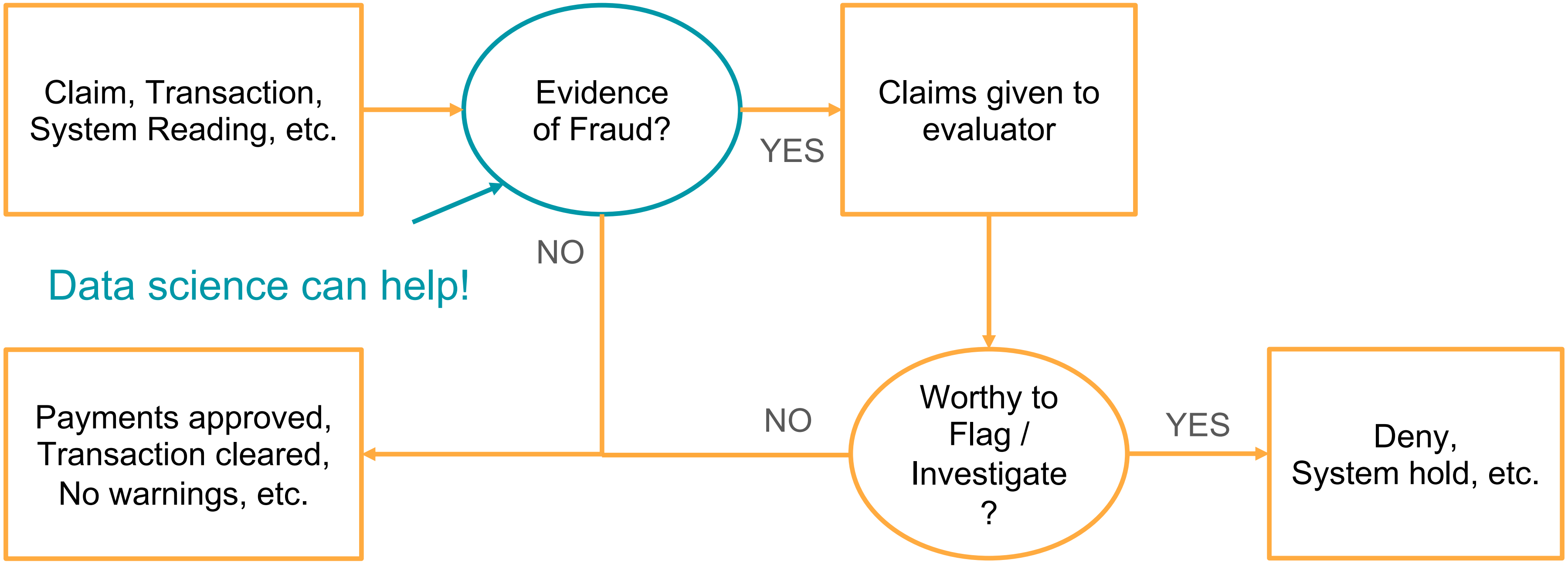
Anomaly Detection Maturity – Insurance Fraud

- New / young anomaly detection solutions are based on **business rules**.
- Example:
  - IF:
    - Severe injury but no doctor report
  - THEN:
    - Flag as suspicious AND
    - Alert evaluator

Business Rule Approach

- Advantages:
  - Simple
  - Easy to implement
- Disadvantages:
  - Expensive
  - Difficult to maintain and manage
  - Completely historical
  - Threats discover rules

Anomaly Detection Systems



Ai+ TRAINING

Analytical Fraud Detection Framework

- Advantages
  1. **Precision**
     - Increased detection power
     - More information used in decisions
     - More anomalies evaluated

Analytical Fraud Detection Framework

- Advantages
    1. **Precision**
    2. **Efficiency in Operations**
        - Automated processing of claims
        - Ranked cases for evaluators

Analytical Fraud Detection Framework

- Advantages
  1. **Precision**
  2. **Efficiency in Operations**
  3. **Efficiency in Costs**
     - Cheaper to long-run maintain
     - Quicker evaluation
     - Higher return on evaluations

Introduction

Conclusion

- Introduction
  - Who am I?
  - What is Fraud?
  - Fraud Detection Analytical Framework

Supervised Modeling

Supervised Modeling

- Supervised Modeling
    - Interpretable Models
    - Naïve Bayes Model
    - More Advanced Models
    - Model Evaluation
    - **NOT**-fraud Model

Fraud Data

- There are 3 common scenarios when it comes to fraud detection data sets:
    1. No previous data on fraudulent cases.

Fraud Data

- There are 3 common scenarios when it comes to fraud detection data sets:
  1. No previous data on fraudulent cases.
  2. Previous data on fraudulent cases, but can not use it.
     - Organizational structure prohibits exchange of information.
     - Retrieving data is too time consuming or expensive.
     - Fraudulent transactions can not be mapped to master database of important information.

Fraud Data

- There are 3 common scenarios when it comes to fraud detection data sets:
  1. No previous data on fraudulent cases.
  2. Previous data on fraudulent cases, but can not use it.
  3. Previous data on fraudulent cases that is fully integrated into company databases and structure.

Fraud Data

- There are 3 common scenarios when it comes to fraud detection data sets:

  1. No previous data on fraudulent cases.
  2. Previous data on fraudulent cases, but can not use it.
  3. Previous data on fraudulent cases that is fully integrated into company databases and structure.

How to handle these situations?

Anomaly Detection

- When no known fraud cases exist, we can find anomalous observations to serve as proxies.
- Anomaly detection techniques:
    - Probabilistic and Statistical Approaches
        - Benford's Law, Z-scores, IQR Rule, Mahalanobis Distances
    - Machine Learning Approaches
        - k-NN, Local Outlier Factor, Isolation Forests, CADE, One-class SVM

Anomaly Detection

- When no known fraud cases exist, we can find anomalous observations to serve as proxies.
- 2 Paths from here:

  1. Wait for SIU to investigate anomalies and slowly gather data over time.
  2. Bring in subject matter experts (SME’s) to help with continuing modeling process.

Supervised Learning

- Supervised learning techniques are techniques where you know the values of the target value.
- The model will classify the individuals into one of two groups – suspected fraud or not.
- Models do this through scoring.

Scoring

- Models will produce a score for each individual between 0 and 1.
- A cut-off value is derived for the score where anything above the cut-off is suspected of fraud and anything below is not.
- Cut-off values are best determined through time and cost calculations.

Supervised Modeling

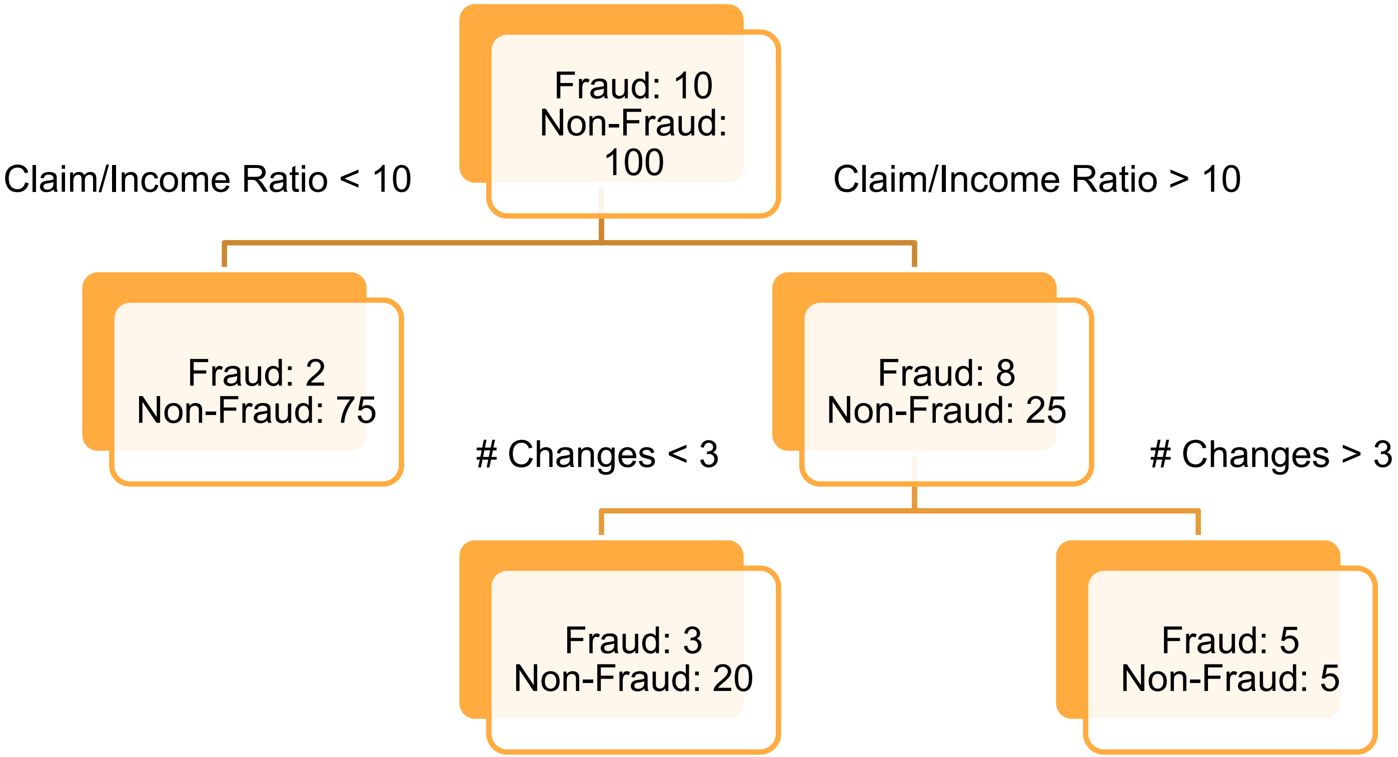
Interpretable Models

- Supervised Modeling
  - Interpretable Models
  - Naïve Bayes Model
  - More Advanced Models
  - Model Evaluation
  - **NOT**-fraud Model

Decision Trees

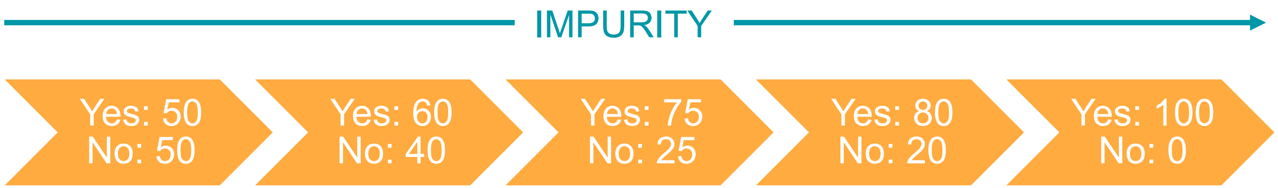
- A tree is built by recursively splitting the data into successively **purer** subsets of data.
- Splitting is done according to some condition.

Decision Trees



Decision Trees – Selecting the Split

- Variety of measures used to select the best split, but all look at **impurity** of a node.



- Entropy, Gini, Classification Error

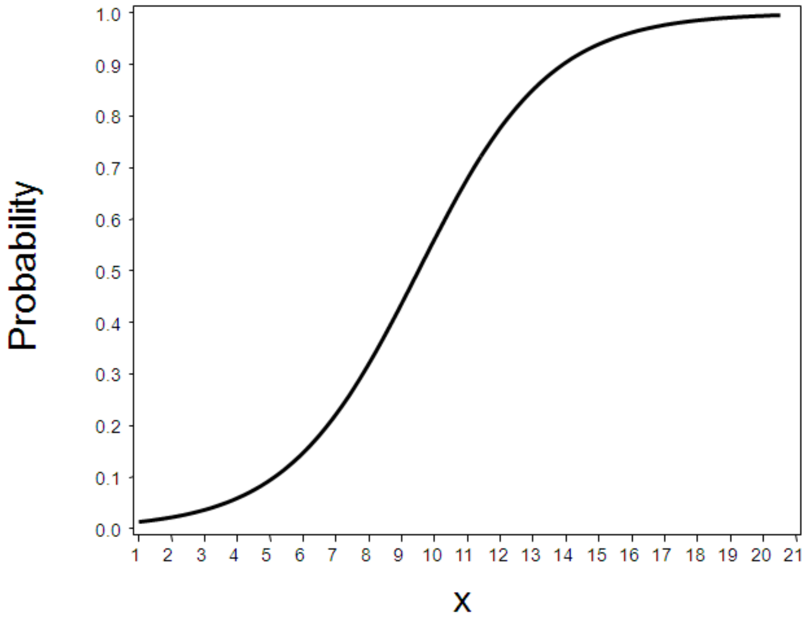
Coding in Action

Supervised Modeling – Interpretable Models – Decision Trees

Logistic Regression

- A statistical model used to calculate the probabilities of an event occurring based on input variables.

$$p_i = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_{1,i})}}$$



Logistic Regression

$$\text{logit}(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_k x_{k,i}$$

- To create a linear model, a link function (logit) is applied to the probabilities.
- Interpretation: If x_1 goes up by 1 unit, the odds of the outcome increases by $100\times(e^{\beta}-1)\%$.

Coding in Action

Supervised Modeling – Interpretable Models – Logistic Regression

Supervised Modeling

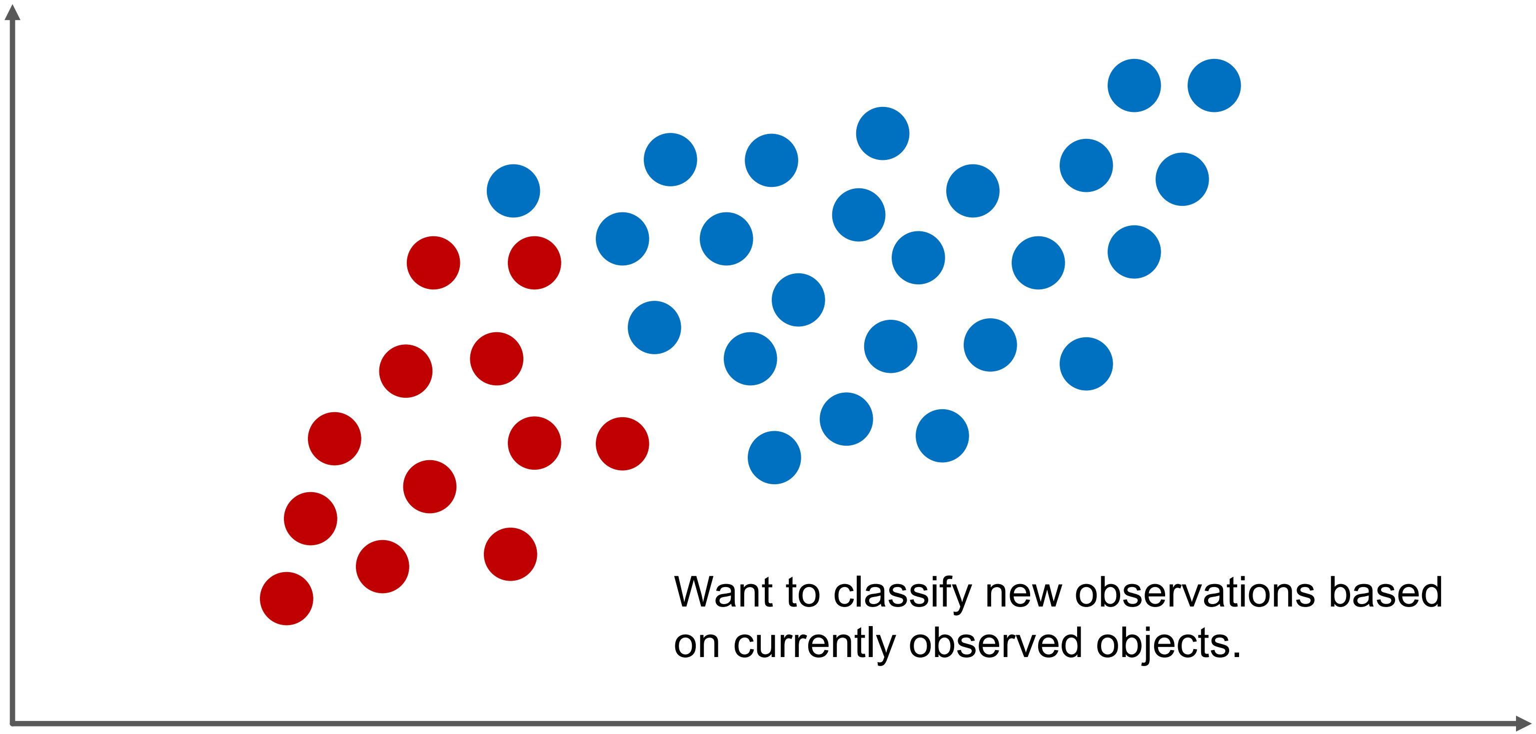
Naïve Bayes Model

- Supervised Modeling
    - Interpretable Models
    - Naïve Bayes Model
    - More Advanced Models
    - Model Evaluation
    - **NOT**-fraud Model

Naïve Bayes Classification

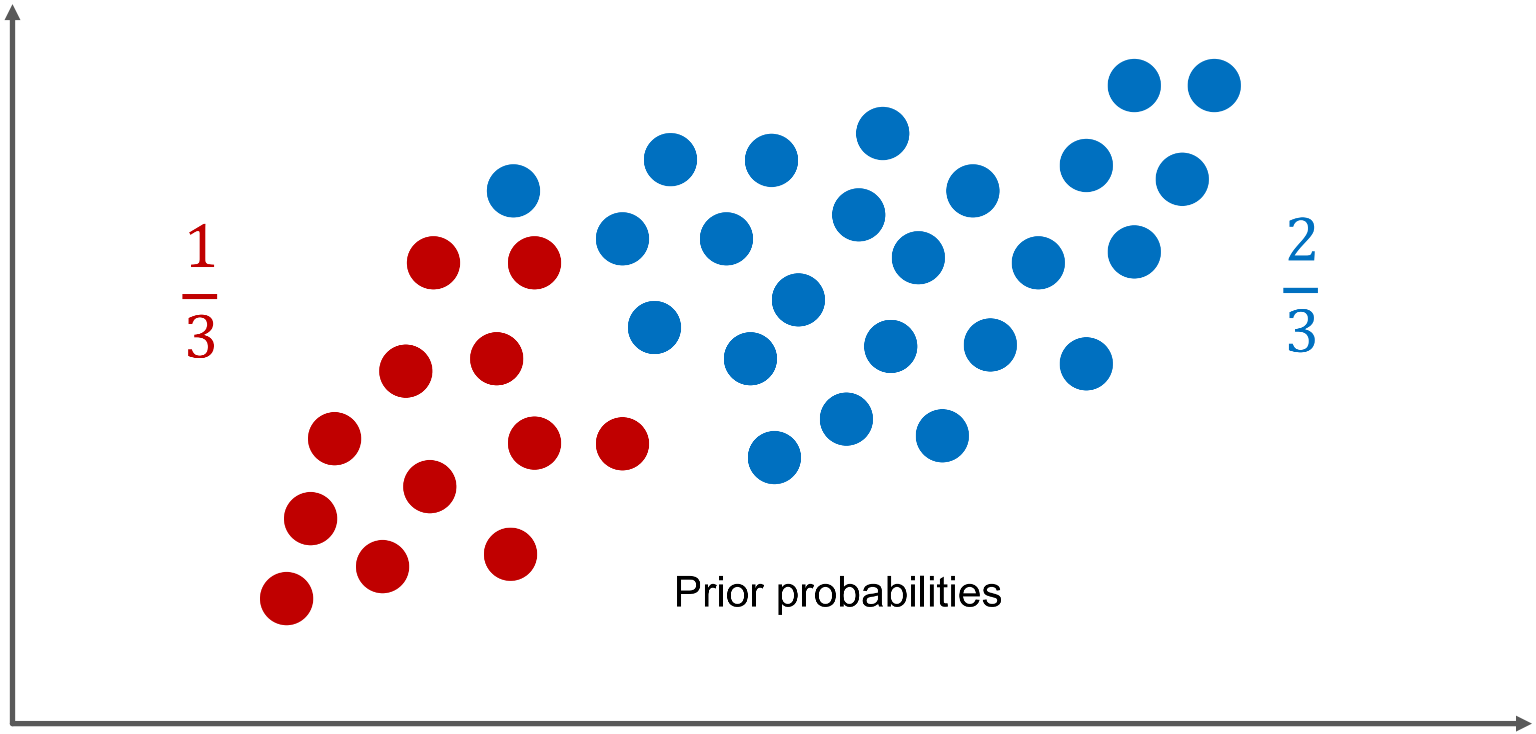
- When we need to classify variables there are two different sources of evidence:
    1. Similarity to each other based on certain metrics.
    2. Past decisions on classifications of observations like it.

Naïve Bayes Classification

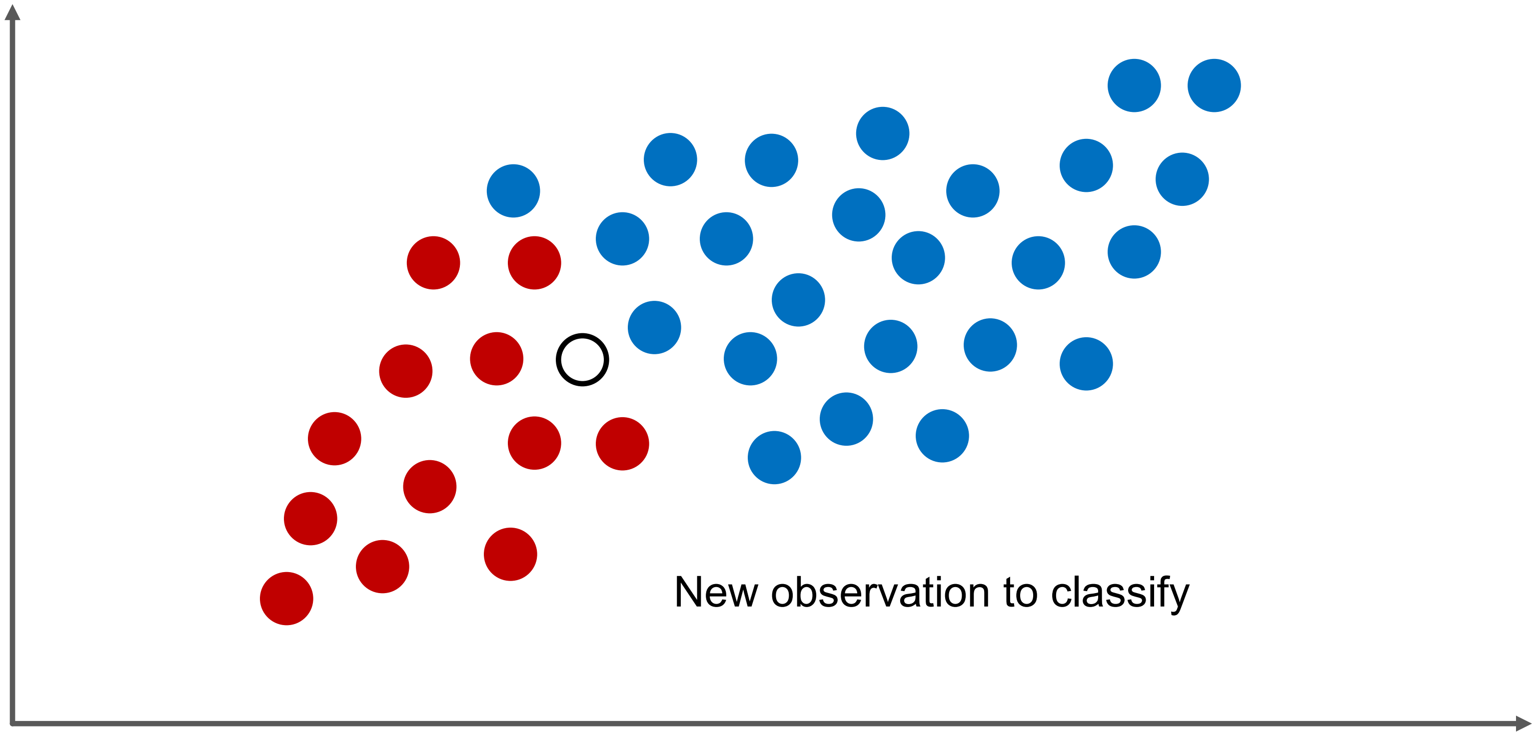


Want to classify new observations based on currently observed objects.

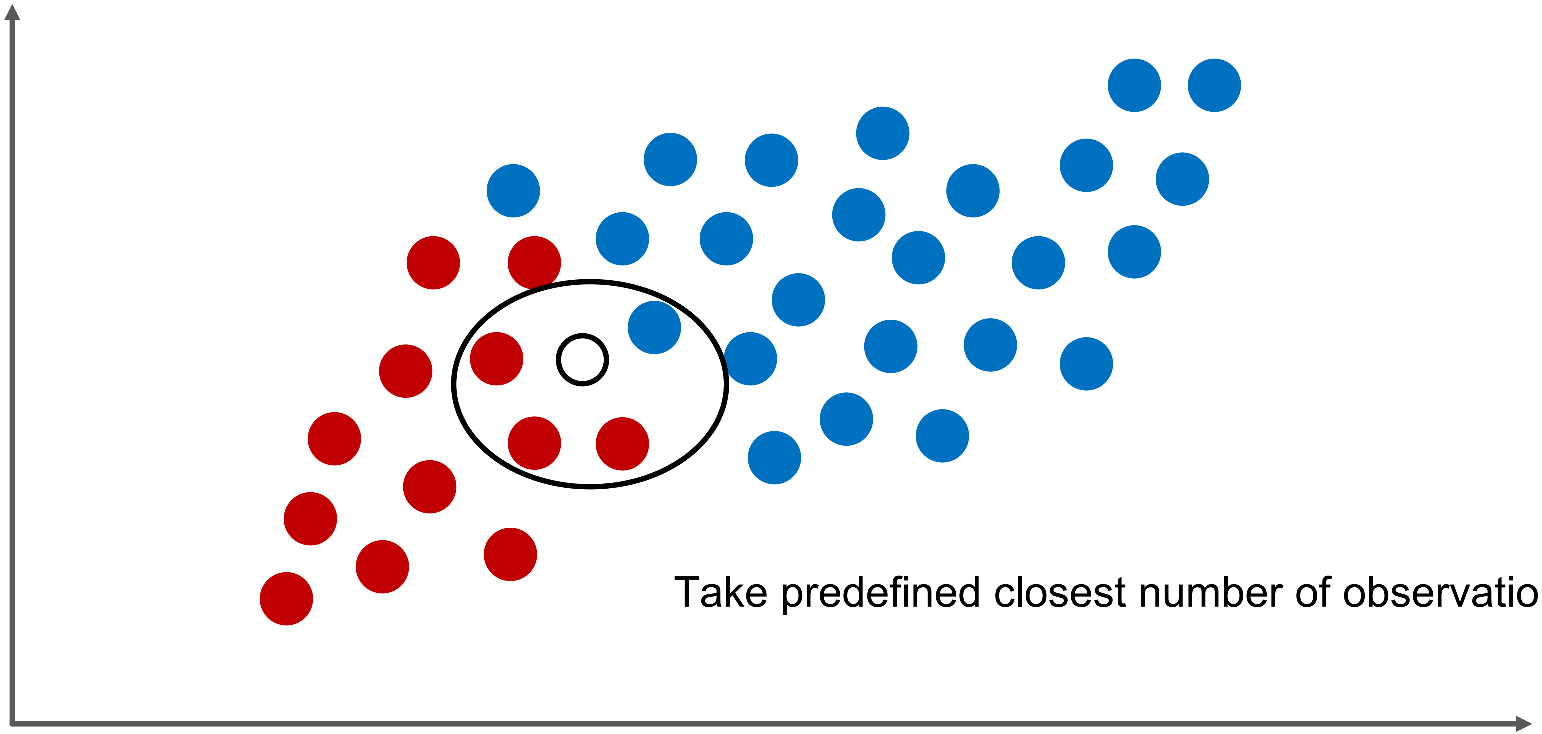
Naïve Bayes Classification



Naïve Bayes Classification

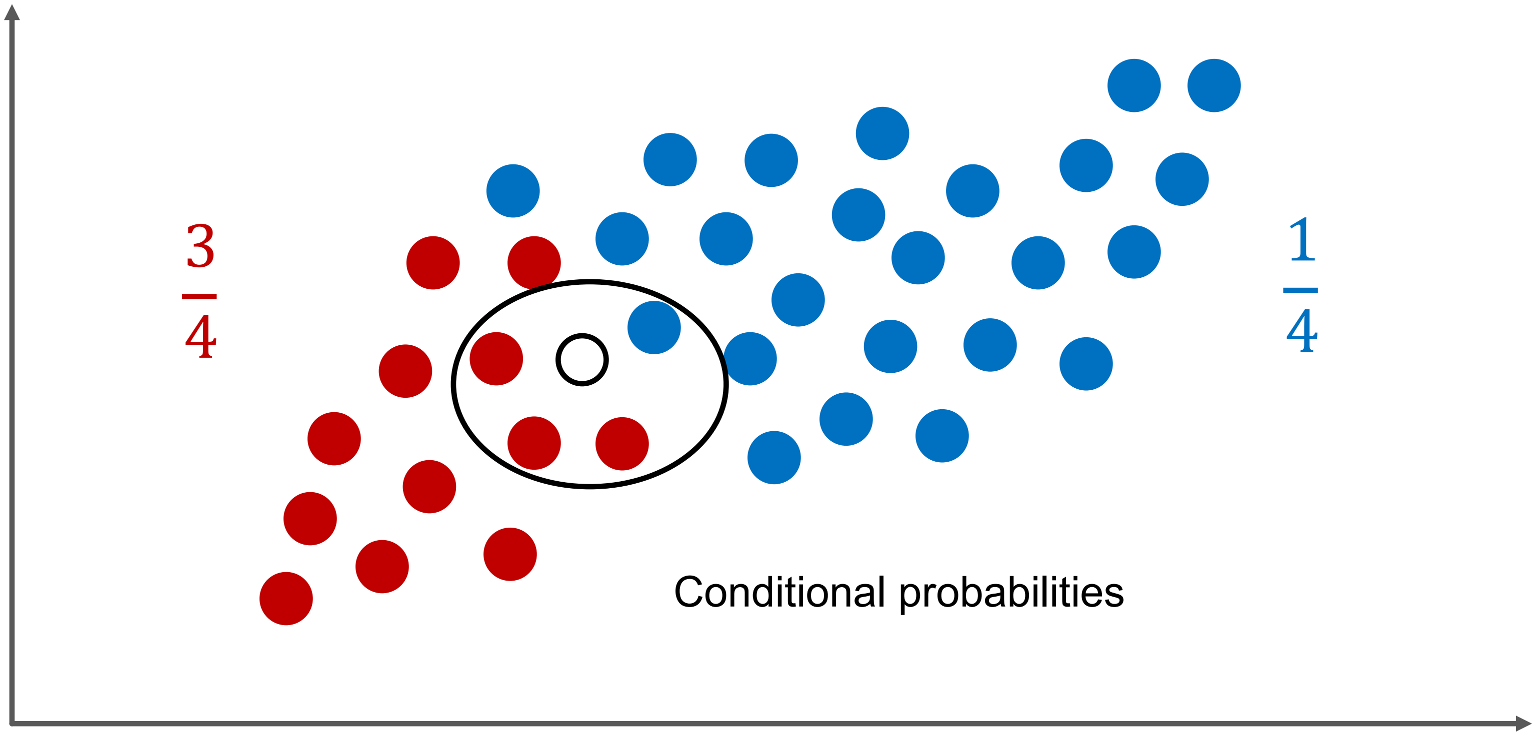


Naïve Bayes Classification

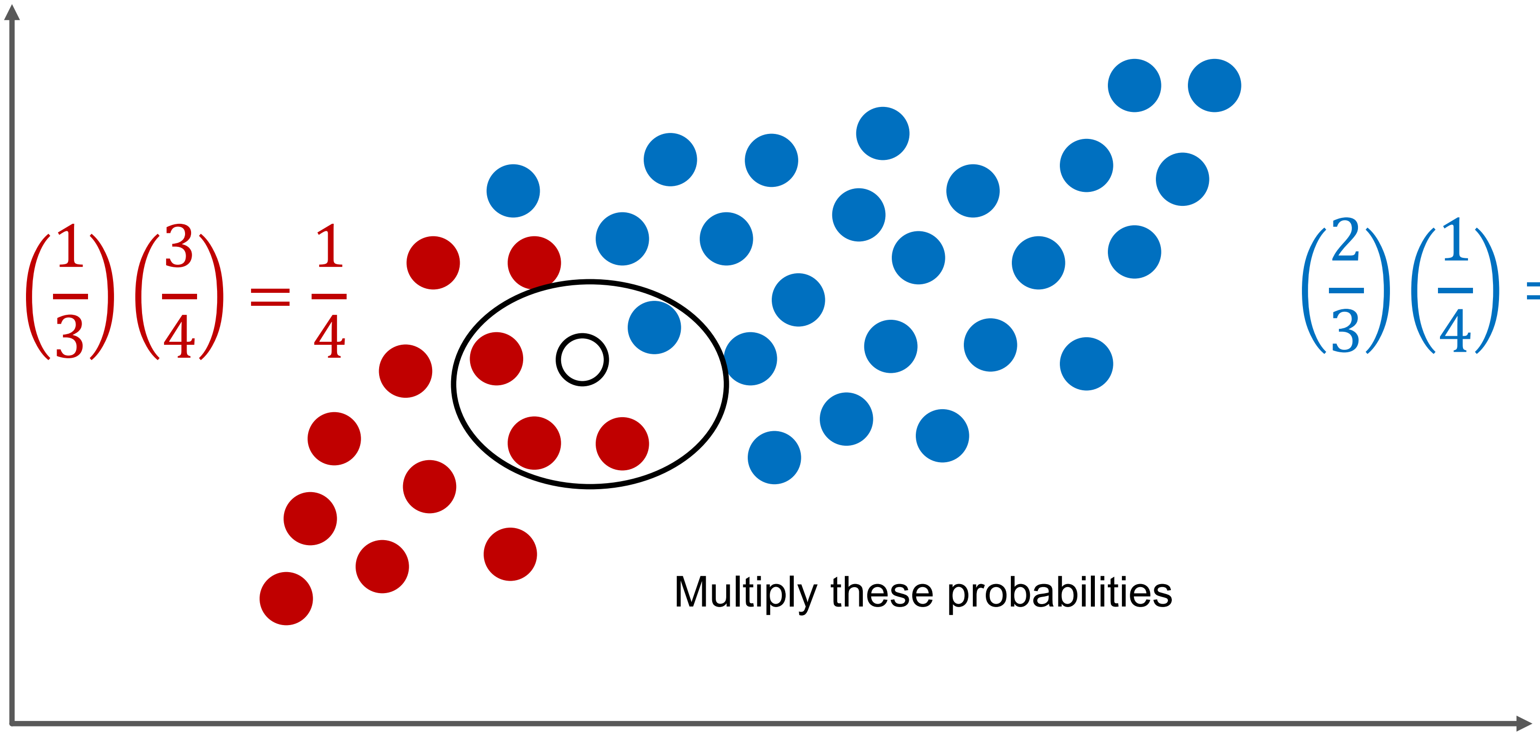


Take predefined closest number of observations

Naïve Bayes Classification



Naïve Bayes Classification

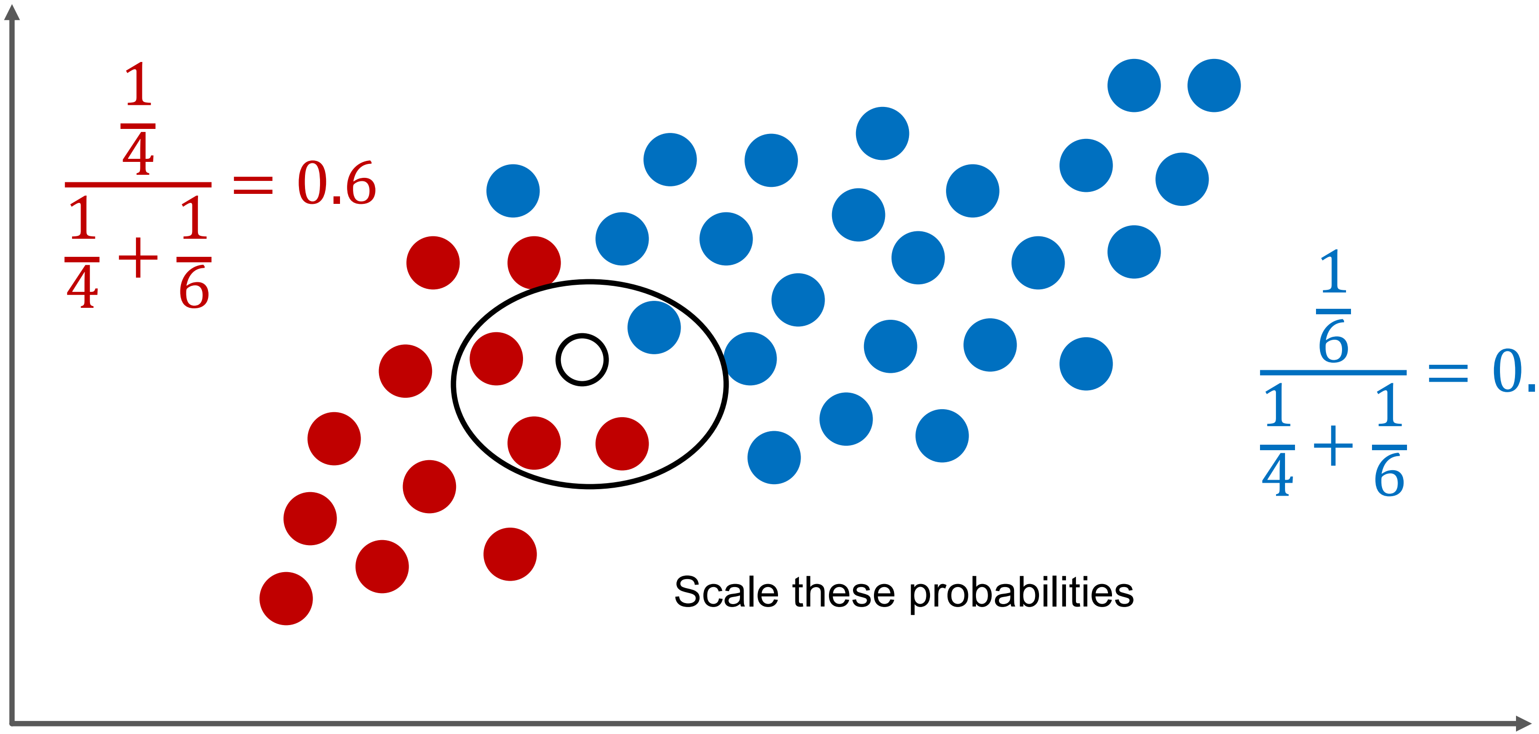


$$\left(\frac{1}{3}\right)\left(\frac{3}{4}\right) = \frac{1}{4}$$

$$\left(\frac{2}{3}\right)\left(\frac{1}{4}\right) = \frac{1}{6}$$

Multiply these probabilities

Naïve Bayes Classification



$$\frac{\frac{1}{4}}{\frac{1}{4}+\frac{1}{6}} = 0.6$$

$$\frac{\frac{1}{6}}{\frac{1}{4}+\frac{1}{6}} = 0.4$$

Naïve Bayes Assumption

- One of the big assumptions of the Naïve Bayes Classification method is one of the hardest things to accept:
  - Predictor variables are independent in their effects on the classification.

- This is a rather "naïve" assumption.
- Assumption doesn't seem to bother posterior probabilities too greatly in case studies.

Coding in Action

Supervised Modeling – Naïve Bayes Model

Supervised Modeling

More Advanced Models

- Supervised Modeling
    - Interpretable Models
    - Naïve Bayes Model
    - More Advanced Models
    - Model Evaluation
    - **NOT**-fraud Model

Random Forest

- Random forests are combinations of many decision trees that are **ensemble** together.
- Each tree is built on a sample of data (with replacement) **and** a subset of features (not all) are considered at each split.
- The results from the trees are ensemble into one voting system.

Random Forest

- Advantages
  - Computationally fast
  - Very accurate
  - Handles missing data
  - Variable importance possible
- Disadvantages
  - No interpretability in final model
  - Possible overfitting
  - Lots to tune

Coding in Action

Supervised Models – More Advanced Models – Random Forest

Gradient Boosting

- Build a simple model to predict target:

$$y = f_1(x) + \varepsilon_1$$

- Model has error. What if we tried to predict this error?

$$\varepsilon_1 = f_2(x) + \varepsilon_2$$

- This model has error too…

Gradient Boosting

- Can do this repeatedly over and over…

$$y = f_1(x) + \eta * f_2(x) + \eta * f_3(x) + \cdots + \eta * f_k(x) + \varepsilon_k$$

Original model

Predict error from model 1

Predict error from model 2

- The η is used to dampen the effects of the error models to prevent overfitting.

Coding in Action

Supervised Models – More Advanced Models – Gradient Boosting

Supervised Modeling

Model Evaluation

- Supervised Modeling
    - Interpretable Models
    - Naïve Bayes Model
    - More Advanced Models
    - Model Evaluation
    - **NOT**-fraud Model

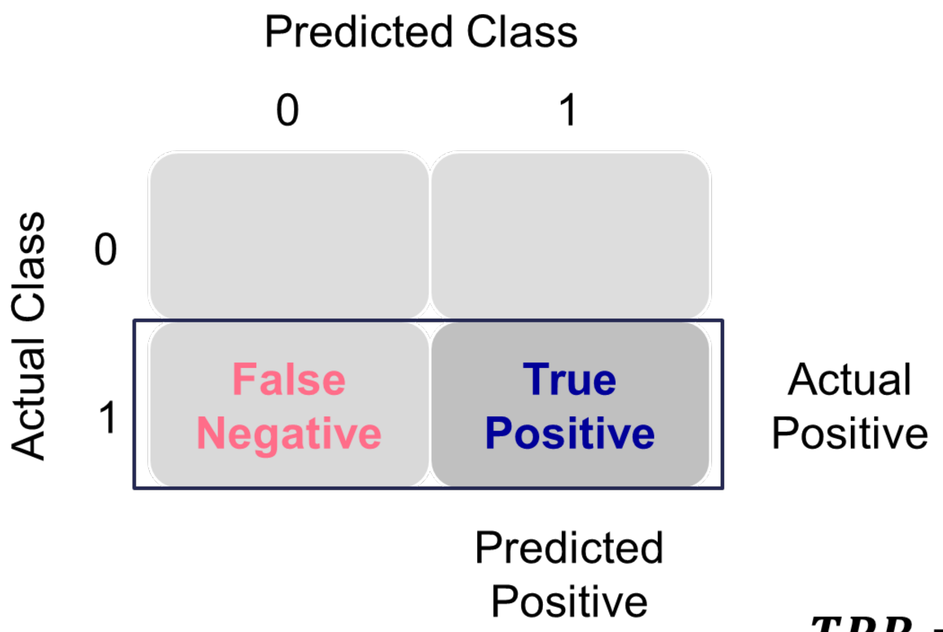
Classification

- Want model to correctly classify events and non-events.
- **Classification** forces the model to predict $\hat{y}_i = 1$ or $\hat{y}_i = 0$ based on whether the predicted probability exceeds some threshold – for example, $\hat{y}_i = 1$ if $\hat{p}_i > 0.5$.
- Strict classification-based measures completely discard any information about the actual quality of the model's predicted probabilities.

Classification Table

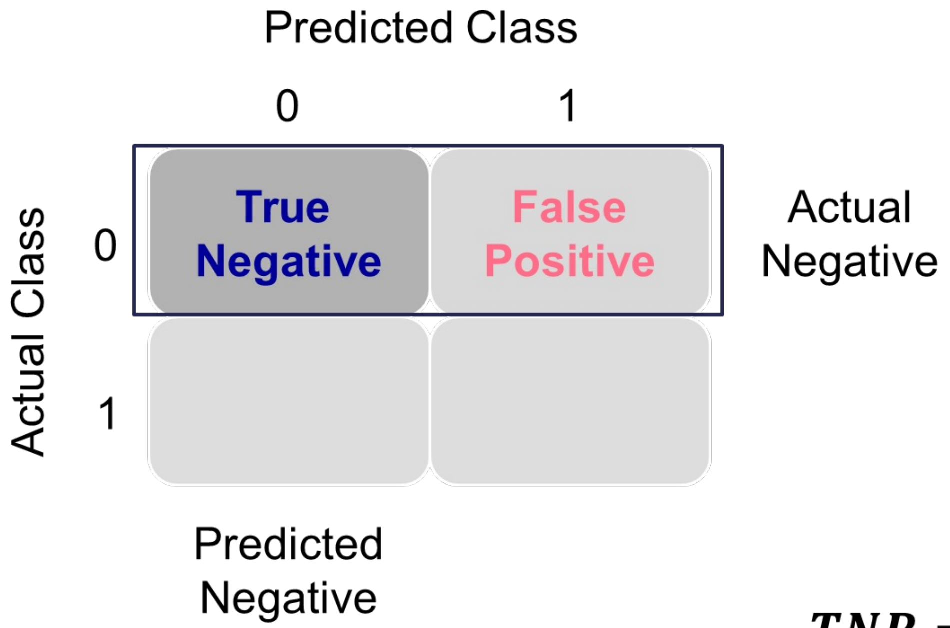
		Predicted Class	
		0	1	
Actual Class	0	**True Negative**	**False Positive**	Actual Negative
	1	**False Negative**	**True Positive**	Actual Positive
		Predicted Negative	Predicted Positive	

Sensitivity / Recall



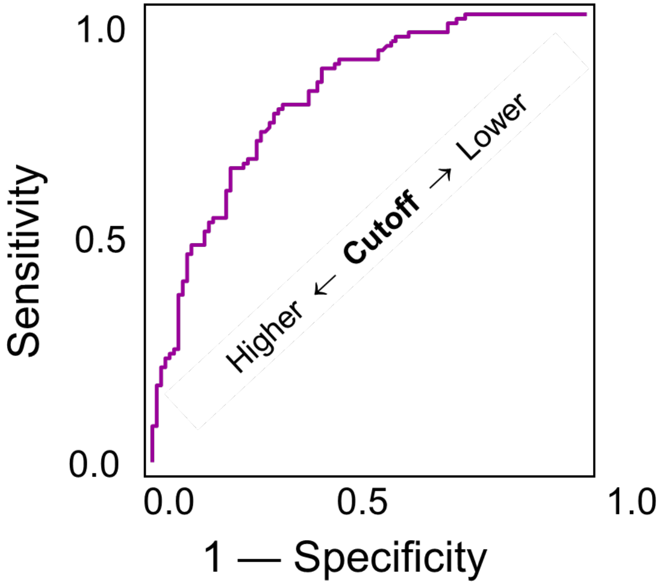
$$TPR = \frac{TP}{TP + FN}$$

Specificity



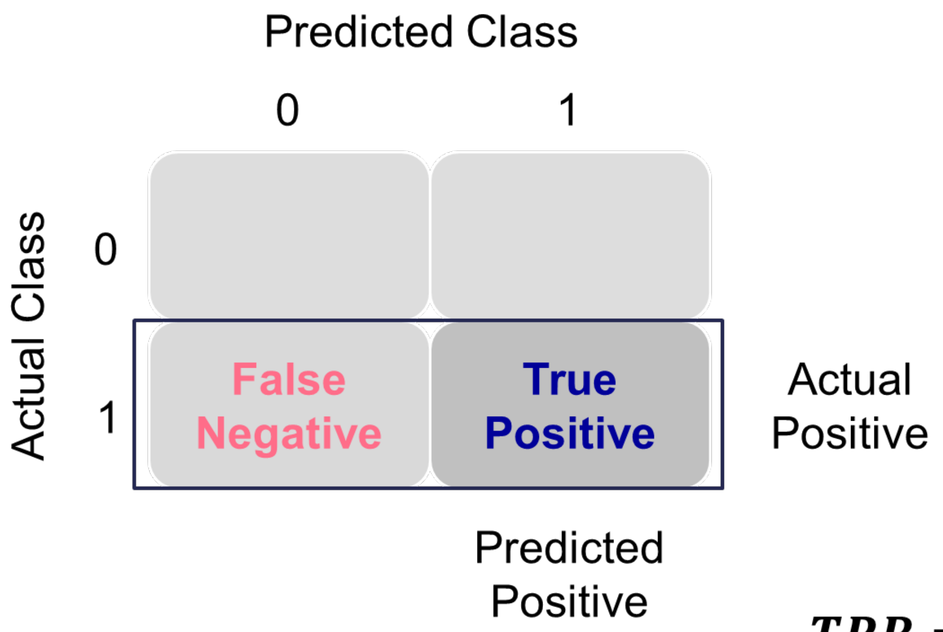
$$TNR = \frac{TN}{TN + FP}$$

ROC Curve



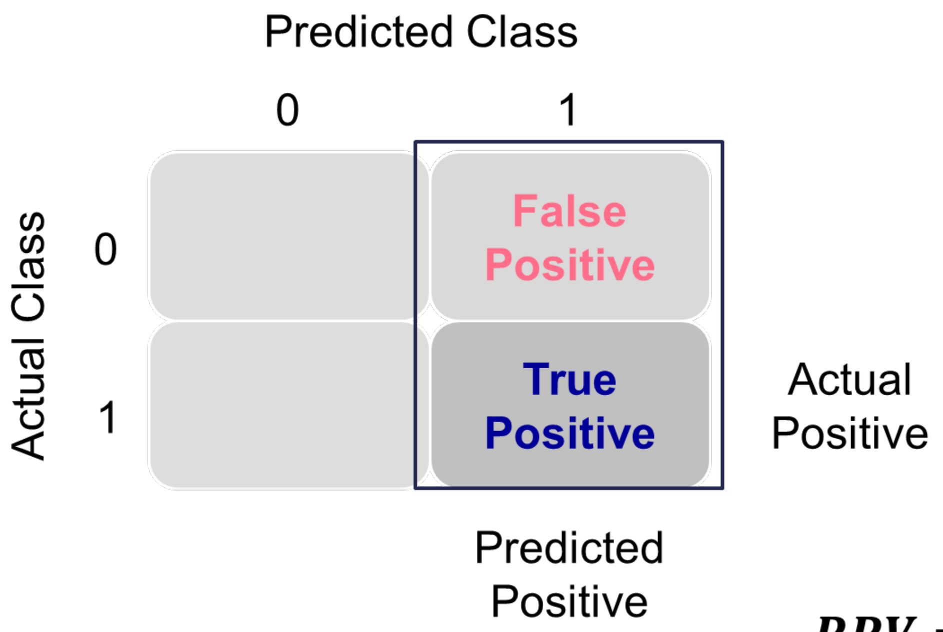
- **ROC curve** plots *TPR* vs. *FPR* for a grid of thresholds.
- **Area under the curve** (AUC or AUROC) summarizes the overall quality of ROC curve – equivalent to c-statistic.
- Want high sensitivity and high specificity.

Sensitivity / Recall



$$TPR = \frac{TP}{TP + FN}$$

Precision



$$PPV = \frac{TP}{TP + FP}$$

Best Cut-off?

- Many different techniques to “optimal” cut-off.
- **Youden J statistic** (or **Youden’s index**):

$$J = \text{sensitivity} + \text{specificity} - 1$$

- “Optimal” – false positives and false negatives are weighed equally , so select cut-off that produces highest Youden *J* statistic.

Best Cut-off?

- Many different techniques to “optimal” cut-off.
- $\boldsymbol{F_1}$ **score** (precision-recall version of Youden’s Index):

$$F_1 = 2\left(\frac{\text{precision}\times\text{recall}}{\text{precision} + \text{recall}}\right)$$

- “Optimal” – precision and recall are weighed equally, so select cut-off that produces highest F_1 score.

Balancing Unbalanced Costs

- Even the best fraud models catch about 25-35% of fraud initially.
- Models should be evaluated more on costs/savings than accuracy in fraud models.
    - May be **very** accurate due to correctly identifying non-fraud.

Balancing Unbalanced Costs

	True Non-Fraud	True Fraud
Predicted Non-Fraud	No Cost	Cost = Amount Paid
Predicted Fraud	Cost = Investigation	Cost = Investigation

Coding in Action

Supervised Models – Model Evaluation

Supervised Modeling

NOT-fraud Model

- Supervised Modeling
  - Interpretable Models
  - Naïve Bayes Model
  - More Advanced Models
  - Model Evaluation
  - **NOT**-fraud Model

Balancing Unbalanced Costs

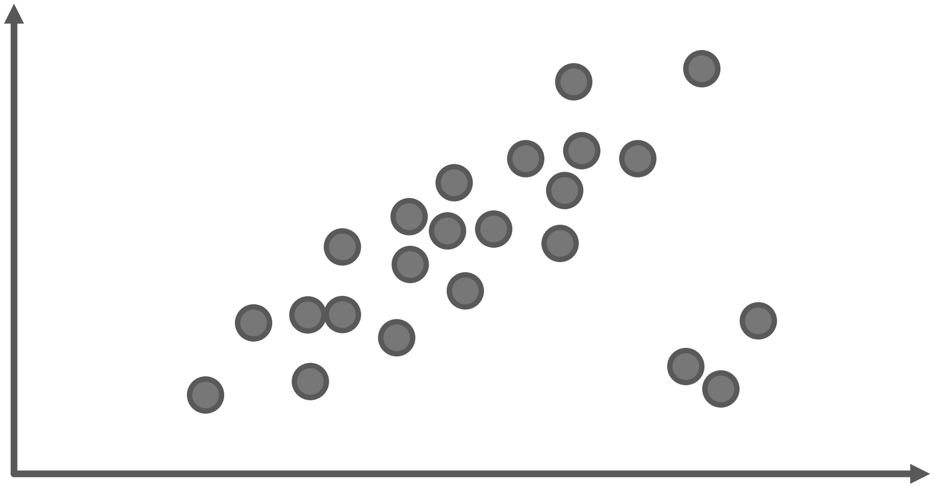
- Regardless of the industry, two things are important for any fraud detection solution:

  1. **DETECTION** – Observing **known** fraudulent observations to determine patterns that may assist in finding other fraudulent observations.

  2. **PREVENTION** – Observing behavior and identifying suspicious actions that might be fraudulent – lead to further investigation and identification of **new** fraudulent observations.

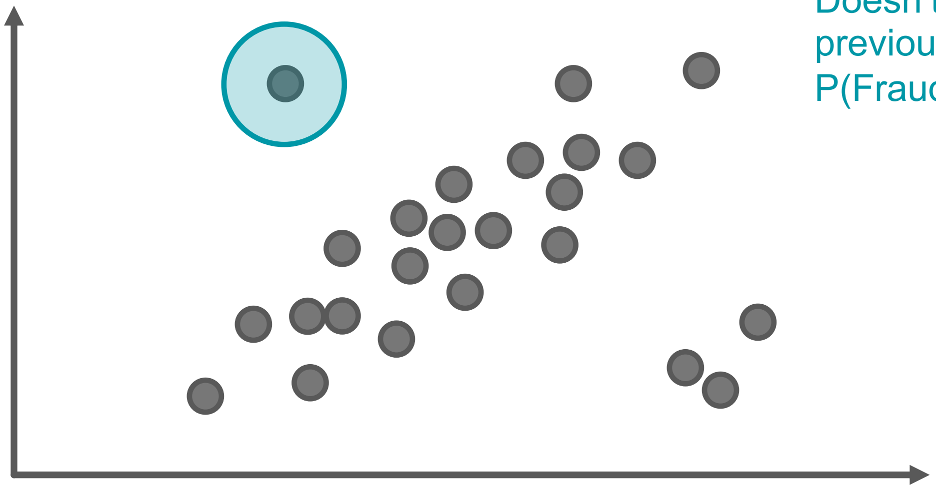
NOT-Fraud Supervised Model

- Predicting previous known cases of fraud works for fraud detection.
- Predicting previous known cases of **not**-fraud works for prevention of new fraud.



NOT-Fraud Supervised Model

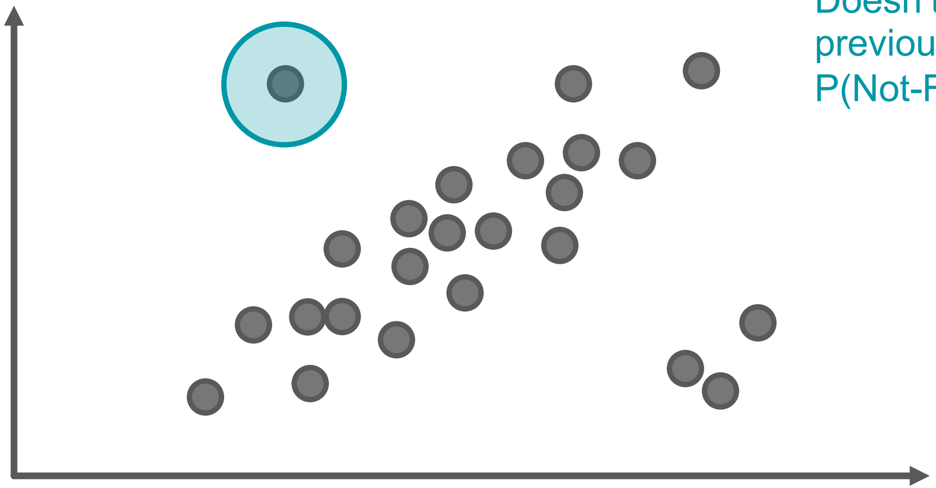
- Predicting previous known cases of fraud works for fraud detection.
- Predicting previous known cases of **not**-fraud works for prevention of new fraud.



Doesn't look like previous fraud.
P(Fraud) = **LOW**

NOT-Fraud Supervised Model

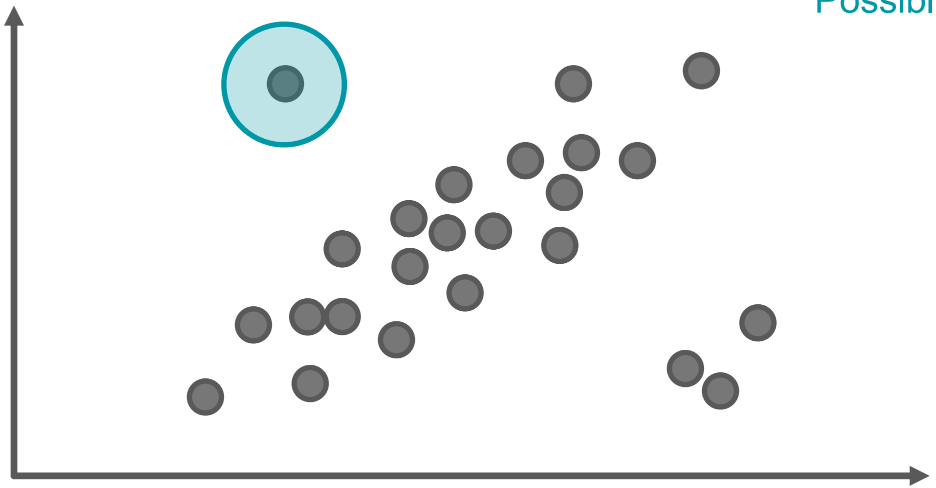
- Predicting previous known cases of fraud works for fraud detection.
- Predicting previous known cases of **not**-fraud works for prevention of new fraud.



Doesn't look like
previous not-fraud.
P(Not-Fraud) = **LOW**

NOT-Fraud Supervised Model

- Predicting previous known cases of fraud works for fraud detection.
- Predicting previous known cases of **not**-fraud works for prevention of new fraud.



Coding in Action

Supervised Models – **NOT**-fraud Model

Supervised Modeling

Conclusion

- Supervised Modeling
    - Interpretable Models
    - Naïve Bayes Model
    - More Advanced Models
    - Model Evaluation
    - **NOT**-fraud Model

Implementation / Deployment

Implementation / Deployment

- Implementation / Deployment
  - Clustering Revisited
  - Interpretability
  - Long-term Fraud Strategy
  - Chance & Loss Models

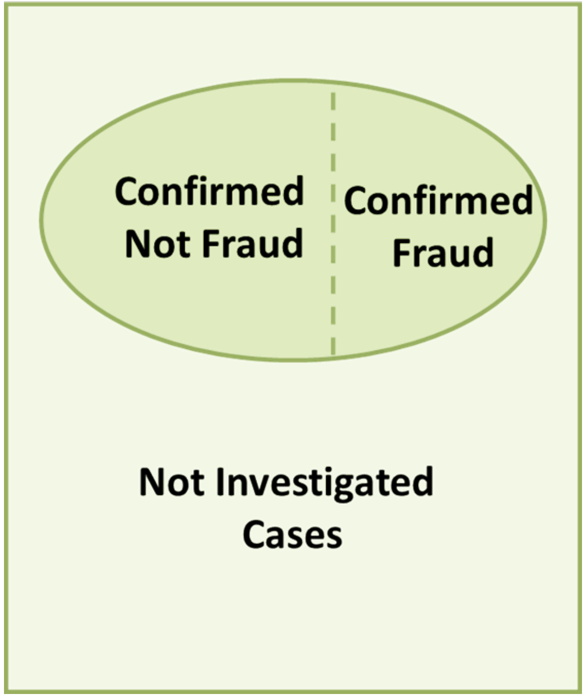
Implementation / Deployment

Clustering Revisited

- Implementation / Deployment
  - Clustering Revisited
  - Interpretability
  - Long-term Fraud Strategy
  - Chance & Loss Models

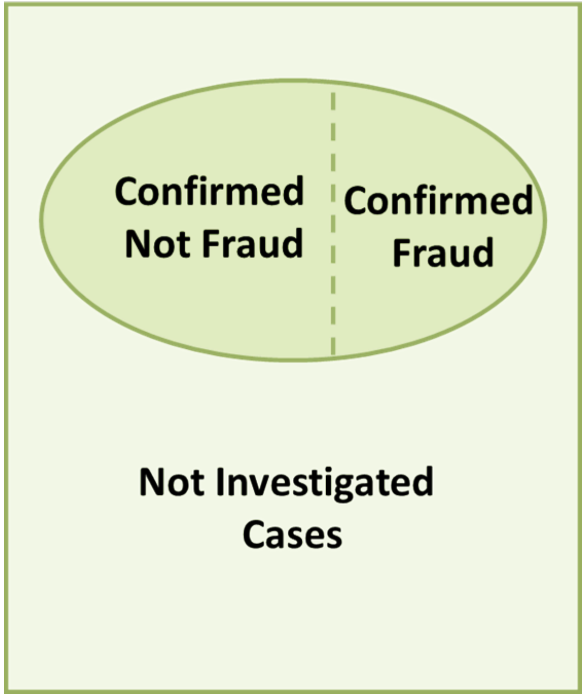
Universe of Potential Fraud Cases

- Even if fraud data exists, a majority of the fraud data has a typical value of “Unknown.”
- While a claim that has never been investigated is most likely not fraud compared to fraud, it is still impossible to correctly label.



Fraud Model, Not-Fraud Model, …

- After identifying both the fraud and not-fraud models from the known data, turn attention to **unknown** data.
- Trying to find the unique instances of observations that aren't like previous fraud **and** not like previous not-fraud.



Unknown **Scored** Observations

- Possibly too many to investigate, so how do I prioritize the ones I need.
- Instead of just giving highest scoring observations, sometimes we take same approach as when we didn't have data:

  1. Anomaly models
  2. Clustering

Unknown **Scored** Observations

- Find the collections of scored observations that might represent new groups of fraud.
- Then same process with SME's as before:

1. SME's will look through the anomalies (clusters) for possible fraud.
2. Tag suspected fraud groups based on expert domain knowledge.
3. Treat these possible fraud groups as if they had committed fraud and other groups as if they have not.
4. Ideally, have SME's also identify small set of legitimate claims in non-anomaly data.

Unknown **Scored** Observations

- One of 2 paths:
    1. **IDEALLY**, investigators trust your process and investigate new types of fraud based solely on the SME recommendations.
    2. **MIGHT** have to put these tagged “possible new fraud” claims into the modeling process and let the model results tell the investigators to act.

Implementation / Deployment

Interpretability

- Implementation / Deployment
    - Clustering Revisited
    - Interpretability
    - Long-term Fraud Strategy
    - Chance & Loss Models

Fraud End Users

- Typically, the user of a fraud system is an investigator:
  - Former/current law enforcement
  - Years of experience in investigations
  - Succeeded in their job without analytics
  - Have a current process in place
  - Need to be sold on why they might change

Listening

- VERY IMPORTANT
- Listening requires two things:

  1. Desire

  2. Humility

- Research ahead of time – YES!
- Be biased ahead of time – NO!
- Ask many questions to help understand – YES!

Beneficial to Investigators

- Fits into their current process
  - Dashboard?
- Where should I start the investigation?
  - Important variables that drove model to pick this person as potential fraud

Scorecard Models

Variable	Level	Scorecard Points
Pay Time	$x < 10$	100
Pay Time	$10 \leq x < 15$	120
Pay Time	$15 \leq x < 25$	185
Pay Time	$x \geq 25$	200
Report	Yes	225
Report	No	110
Ratio	$x < 1$	225
Ratio	$1 \leq x < 2.5$	200
Ratio	$2.5 \leq x < 5$	180
Ratio	$5 \leq x < 7$	140
Ratio	$x \geq 7$	120

Traffic Light Indicators

Variable	Level	Scorecard Points
Pay Time	$x < 10$	100
Pay Time	$10 \leq x < 15$	120
Pay Time	$15 \leq x < 25$	185
Pay Time	$x \geq 25$	200
Report	Yes	225
Report	No	110
Ratio	$x < 1$	225
Ratio	$1 \leq x < 2.5$	200
Ratio	$2.5 \leq x < 5$	180
Ratio	$5 \leq x < 7$	140
Ratio	$x \geq 7$	120

Traffic Light – Example

Variable	Level	Scorecard Points
Pay Time	$x < 10$	100
Pay Time	$10 \leq x < 15$	120
Pay Time	$15 \leq x < 25$	185
Pay Time	$x \geq 25$	200
Report	Yes	225
Report	No	110
Ratio	$x < 1$	225
Ratio	$1 \leq x < 2.5$	200
Ratio	$2.5 \leq x < 5$	180
Ratio	$5 \leq x < 7$	140
Ratio	$x \geq 7$	120

Implementation / Deployment

Long-term Fraud Strategy

- Implementation / Deployment
  - Clustering Revisited
  - Interpretability
  - Long-term Fraud Strategy
  - Chance & Loss Models

Classification

- Claims are referred to the SIU for investigation and classified as fraud or no fraud.
- Investigated claims are labeled “Yes” or “No”.
- Non-investigated claims are labeled “Maybe”.
    - Classified based on unsupervised learning techniques previously discussed.
- All claims are then merged into supervised prediction model.

False Negatives?

- Claims that are labeled as no fraud should occasionally be investigated as well.
- Determine how many low scoring claims can be checked under the budget constraints.
- Randomly select low scoring claims to be passed on to SIU.
- This provides an idea for the false negative rate in the modeling process.

Implementation / Deployment

Chance & Loss Models

- Implementation / Deployment
  - Clustering Revisited
  - Interpretability
  - Long-term Fraud Strategy
  - Chance & Loss Models

Chance & Loss

- In fraud it is not only important if someone will commit fraud, but how much the fraud will cost the company.
- Want to calculate two things with regards to fraudulent claims:

1. Probability of fraud occurring
2. Monetary losses if the fraud occurs

Chance & Loss

- In fraud it is not only important if someone will commit fraud, but how much the fraud will cost the company.
- Want to calculate two things with regards to fraudulent claims:

1. Probability of fraud occurring
2. Monetary losses if the fraud occurs

$$Score = P(Fraud) \times E(Loss|Fraud)$$

Chance & Loss

- In fraud it is not only important if someone will commit fraud, but how much the fraud will cost the company.
- Want to calculate two things with regards to fraudulent claims:

1. Probability of fraud occurring
2. Monetary losses if the fraud occurs

$$Score = \underbrace{P(Fraud)}_{\text{Binary}} \times \underbrace{E(Loss|Fraud)}_{\text{Continuous}}$$

Common Approach

$$Score = P(Fraud) \times E(Loss|Fraud)$$

- Estimate the probability of fraud and the expected loss given fraud as two separate models followed by multiplying them together.
- Possible models:
  - Multiple Regression
  - Regression Trees
  - Survival Analysis

Survival Analysis

- Type of modeling when loss amounts are not fully available – monthly payments over time as long as injury remains.
- Helpful for open claims in the system since survival analysis can handle **censored observations**.
- Censored observations are values you don’t know the full value of yet.

- Survival analysis is typically used for fraud modeling to determine the expected loss over time for a claim.
- More common in other types of fraud compared to life insurance.

Implementation / Deployment

Conclusion

- Implementation / Deployment
  - Clustering Revisited
  - Interpretability
  - Long-term Fraud Strategy
  - Chance & Loss Models

Conclusion

Course Outline – Part 1 & Part 2

- Introduction
- Data Preparation
- Supervised Modeling
- Implementation / Deployment
- Conclusion

Course Outline – Part 1 & Part 2

- Introduction
  - Who am I?
  - What is Fraud?
  - Fraud Detection Analytical Framework
- Data Preparation
- Supervised Modeling
- Implementation / Deployment
- Conclusion

Course Outline – Part 1 & Part 2

- Introduction
- Data Preparation
  - Feature Engineering
  - Fraud Data
  - Anomaly Detection with Statistical Techniques
  - Anomaly Detection with Machine Learning Techniques
  - Sampling Concerns
- Supervised Modeling
- Implementation / Deployment
- Conclusion

Course Outline – Part 1 & Part 2

- Introduction
- Data Preparation
- Supervised Modeling
  - Interpretable Models
  - Naïve Bayes Models
  - More Advanced Models
  - Model Evaluation
  - **NOT**-fraud Model
- Implementation / Deployment
- Conclusion

Ai+ TRAINING

Course Outline – Part 1 & Part 2

- Introduction
- Data Preparation
- Supervised Modeling
- Implementation Deployment
    - Clustering Revisited
    - Interpretability
    - Long-term Fraud Strategy
    - Chance & Loss Models
- Conclusion

Where Am I?

- Find me online:
  - https://www.linkedin.com/in/ariclabarr/
  - https://www.youtube.com/c/AricLaBarr/
  - https://www.ariclabarr.com/

Thank you

Ai+ TRAINING