Altra Insurance

Brittania Phoebe Lee
Denissa Ting
Samuel Benedict
Thomas Aryawan

Problems

685
Fraudulent Claims from 2015 - 2017 (5.92%)

Worth around
$17.2 million

Huge Financial Loss due to Fraudulent Claims

- Methods to identify and prove frauds are expensive
- Merger with Rixen expanded the number of customers and frauds
- Values customer satisfaction, avoid additional charges on customer

Objectives

- Minimize the number of frauds.
- Predict and identify frauds without resorting to expensive and extensive methods.
- Eliminate financial loss without increasing premiums

Decision Criteria

Strategy	Identify Frauds/ Fraudsters	Deter Frauds	Not Involving Premium Increase	Feasible	Innovative	Decision
Usage-based insurance (PAYD)	✗	✓	✓	✓	✗ (old-fashioned)	✗
Employers training on fraudulent claims	✗ (inefficient)	✗	✓	✓	✗	✗
Utilizing blockchain for secure claiming	✓	✓	✗ (highly expensive)	✗	✓	✗
Enhanced Risk-Scoring System	✓	✗	✓	✓	✓	✓
IoT Sensor Partnership	✓	✓	✓	✓	✓	✓
Screening Check-list	✓	✗	✓	✓	✓	✓

Recommendations

Decision Tree Method and Excel Table

Identify the probable characteristics of a fraudulent claim

Pull specific information
based on these characteristics

Underwriting

Enhanced risk-scoring system and identify high-risk clients

Policy Restructuring

Biz4Intellia IoT sensor are required for high-risk clients

Claims

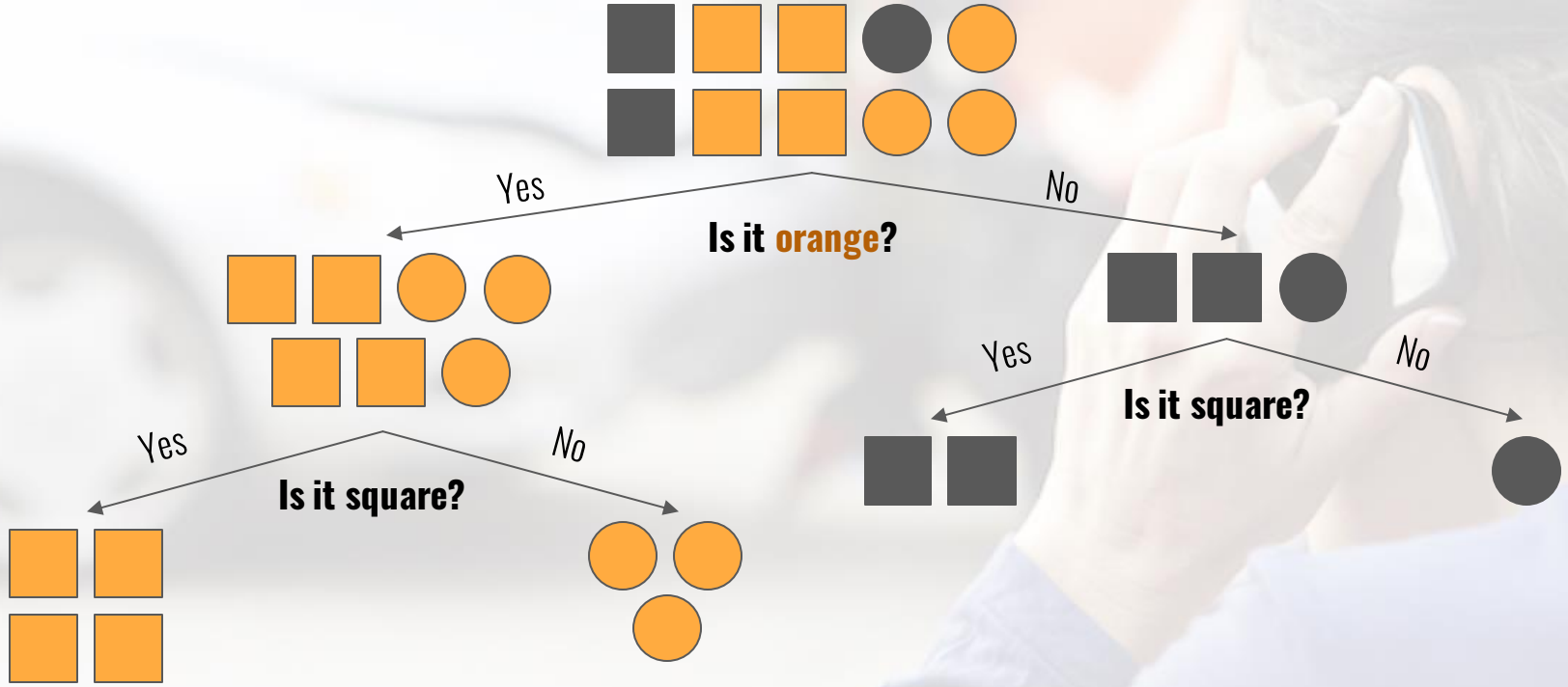
Claims that meet the characteristics are flagged and undergo risk scoring

Minimize frauds cost-efficiently ⇨ Reduce financial loss due to false claims ⇨ **No increase in premiums**

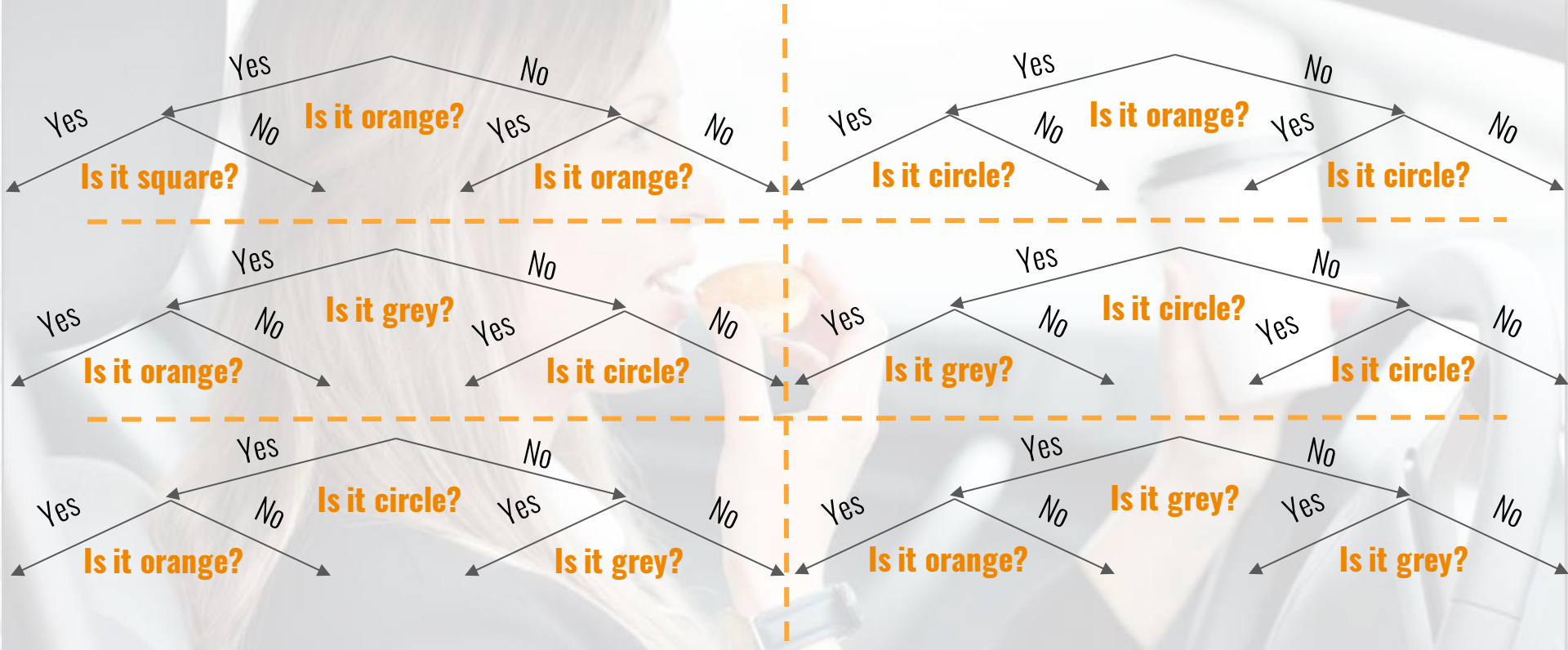
Big Data Analysis

Step 1: Refine the dataset.
Step 2: Apply Random Forest Classifier.
Step 3: Find the most important variables.
Step 4: Analyze the variables.

Decision Trees



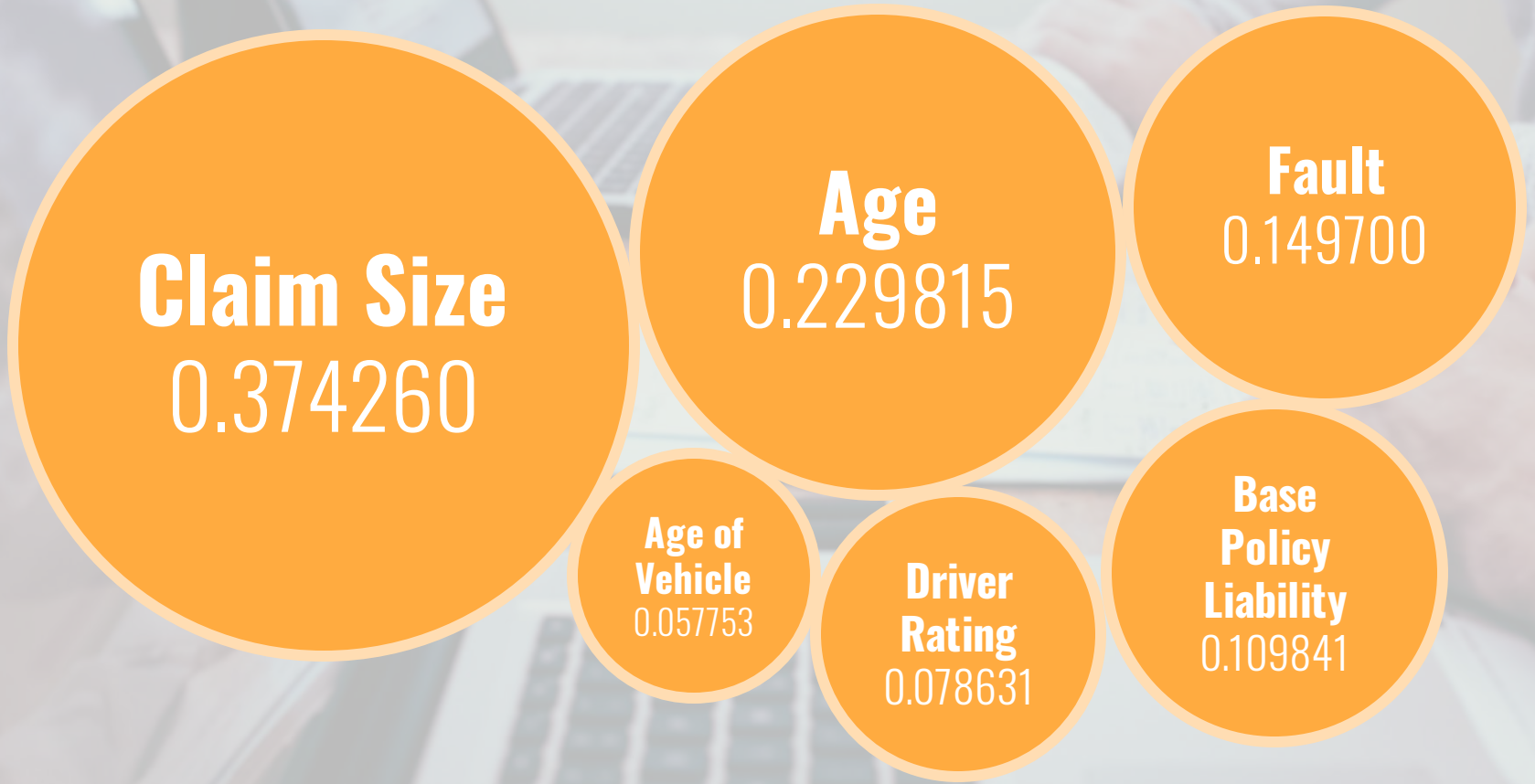
Random Forest Classifier

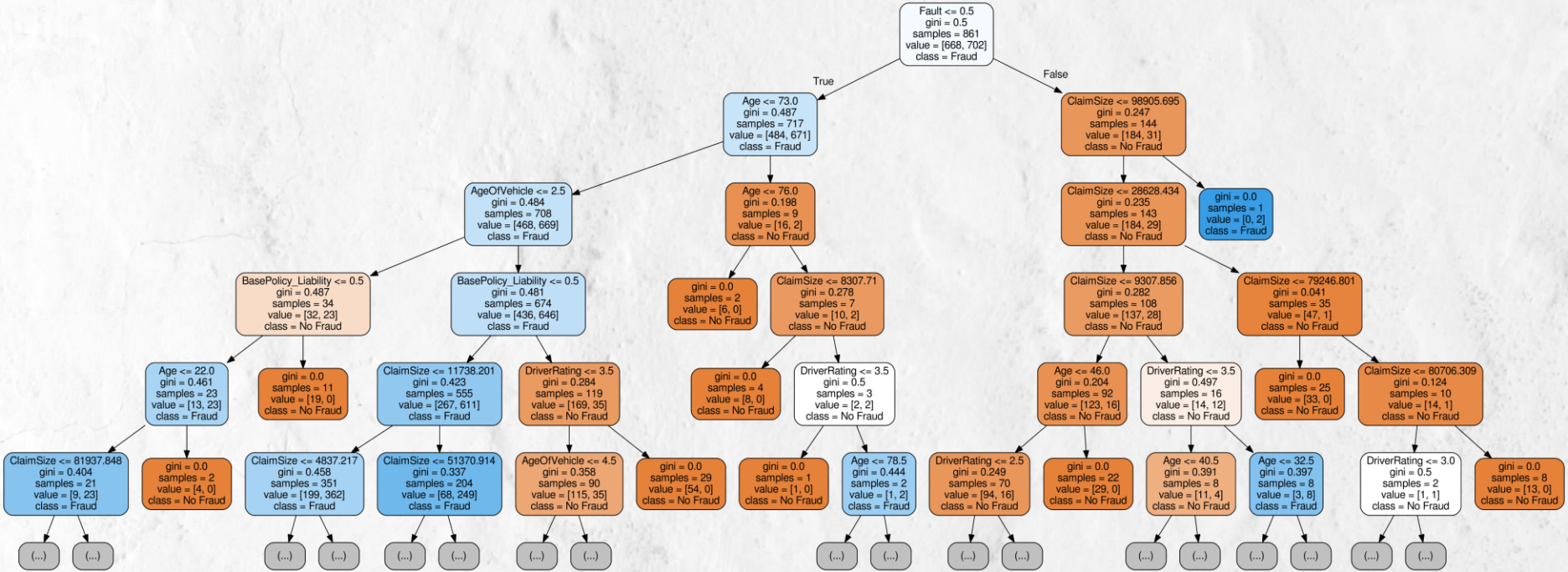


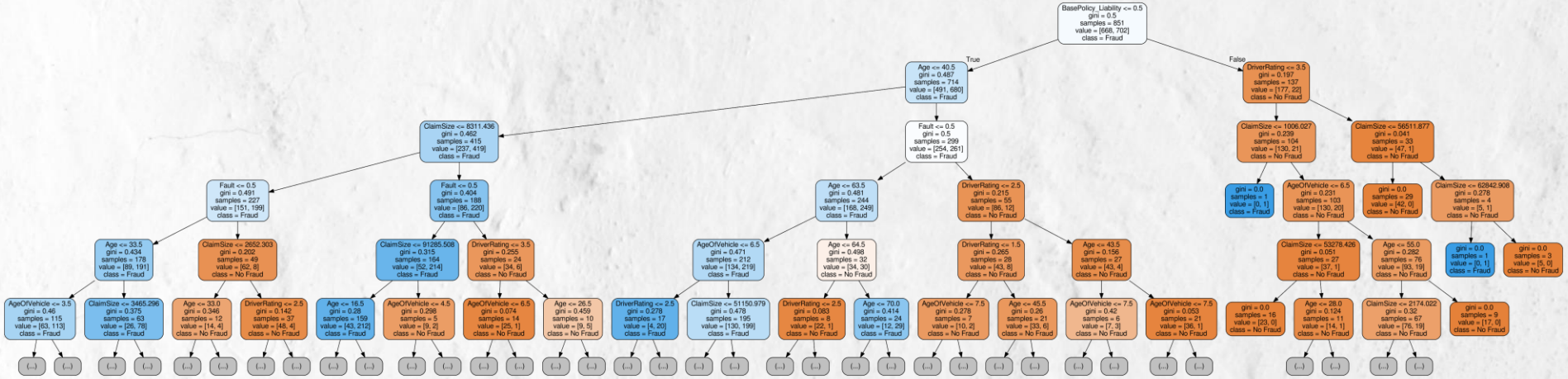
Variable Importance

Fault	**0.121828**
ClaimSize	**0.109072**
Age	**0.088589**
BasePolicy_Liability	**0.042835**
DriverRating	**0.042304**
AgeOfVehicle	**0.042197**
AgeOfPolicyHolder	0.036275
VehiclePrice	0.033492
VehicleCategory_Sport	0.027890

Only Top 6 Variables With Decision Trees




Fault <= 0.5
gini = 0.5
samples = 861
value = [668, 702]
class = Fraud
True
False
Age <= 73.0
gini = 0.487
samples = 717
value = [484, 671]
class = Fraud
ClaimSize <= 98905.695
gini = 0.247
samples = 144
value = [184, 31]
class = No Fraud
AgeOfVehicle <= 2.5
gini = 0.484
samples = 708
value = [468, 669]
class = Fraud
Age <= 76.0
gini = 0.198
samples = 9
value = [16, 2]
class = No Fraud
ClaimSize <= 28628.434
gini = 0.235
samples = 143
value = [184, 29]
class = No Fraud
gini = 0.0
samples = 1
value = [0, 2]
class = Fraud
BasePolicy_Liability <= 0.5
gini = 0.487
samples = 34
value = [32, 23]
class = No Fraud
BasePolicy_Liability <= 0.5
gini = 0.481
samples = 674
value = [436, 646]
class = Fraud
gini = 0.0
samples = 2
value = [6, 0]
class = No Fraud
ClaimSize <= 8307.71
gini = 0.278
samples = 7
value = [10, 2]
class = No Fraud
ClaimSize <= 9307.856
gini = 0.282
samples = 108
value = [137, 28]
class = No Fraud
ClaimSize <= 79246.801
gini = 0.041
samples = 35
value = [47, 1]
class = No Fraud
Age <= 22.0
gini = 0.461
samples = 23
value = [13, 23]
class = Fraud
gini = 0.0
samples = 11
value = [19, 0]
class = No Fraud
ClaimSize <= 11738.201
gini = 0.423
samples = 555
value = [267, 611]
class = Fraud
DriverRating <= 3.5
gini = 0.284
samples = 119
value = [169, 35]
class = No Fraud
gini = 0.0
samples = 4
value = [8, 0]
class = No Fraud
DriverRating <= 3.5
gini = 0.5
samples = 3
value = [2, 2]
class = No Fraud
Age <= 46.0
gini = 0.204
samples = 92
value = [123, 16]
class = No Fraud
DriverRating <= 3.5
gini = 0.497
samples = 16
value = [14, 12]
class = No Fraud
gini = 0.0
samples = 25
value = [33, 0]
class = No Fraud
ClaimSize <= 80706.309
gini = 0.124
samples = 10
value = [14, 1]
class = No Fraud
ClaimSize <= 81937.848
gini = 0.404
samples = 21
value = [9, 23]
class = Fraud
gini = 0.0
samples = 2
value = [4, 0]
class = No Fraud
ClaimSize <= 4837.217
gini = 0.458
samples = 351
value = [199, 362]
class = Fraud
ClaimSize <= 51370.914
gini = 0.337
samples = 204
value = [68, 249]
class = Fraud
AgeOfVehicle <= 4.5
gini = 0.358
samples = 90
value = [115, 35]
class = No Fraud
gini = 0.0
samples = 29
value = [54, 0]
class = No Fraud
gini = 0.0
samples = 1
value = [1, 0]
class = No Fraud
Age <= 78.5
gini = 0.444
samples = 2
value = [1, 2]
class = Fraud
DriverRating <= 2.5
gini = 0.249
samples = 70
value = [94, 16]
class = No Fraud
gini = 0.0
samples = 22
value = [29, 0]
class = No Fraud
Age <= 40.5
gini = 0.391
samples = 8
value = [11, 4]
class = No Fraud
Age <= 32.5
gini = 0.397
samples = 8
value = [3, 8]
class = Fraud
DriverRating <= 3.0
gini = 0.5
samples = 2
value = [1, 1]
class = No Fraud
gini = 0.0
samples = 8
value = [13, 0]
class = No Fraud
(...)
(...)
(...)
(...)
(...)
(...)
(...)
(...)
(...)
(...)
(...)
(...)
(...)
(...)
(...)
(...)
(...)
(...)
(...)
(...)


BasePolicy_Liability <= 0.5
gini = 0.5
samples = 851
value = [668, 702]
class = Fraud
True
False
Age <= 40.5
gini = 0.487
samples = 714
value = [491, 680]
class = Fraud
DriverRating <= 3.5
gini = 0.197
samples = 137
value = [177, 22]
class = No Fraud
ClaimSize <= 8311.436
gini = 0.462
samples = 415
value = [237, 419]
class = Fraud
Fault <= 0.5
gini = 0.5
samples = 299
value = [254, 261]
class = Fraud
ClaimSize <= 1006.027
gini = 0.239
samples = 104
value = [130, 21]
class = No Fraud
ClaimSize <= 56511.877
gini = 0.041
samples = 33
value = [47, 1]
class = No Fraud
Fault <= 0.5
gini = 0.491
samples = 227
value = [151, 199]
class = Fraud
Fault <= 0.5
gini = 0.404
samples = 188
value = [86, 220]
class = Fraud
Age <= 63.5
gini = 0.481
samples = 244
value = [168, 249]
class = Fraud
DriverRating <= 2.5
gini = 0.215
samples = 55
value = [86, 12]
class = No Fraud
gini = 0.0
samples = 1
value = [0, 1]
class = Fraud
AgeOfVehicle <= 6.5
gini = 0.231
samples = 103
value = [130, 20]
class = No Fraud
gini = 0.0
samples = 29
value = [42, 0]
class = No Fraud
ClaimSize <= 62842.908
gini = 0.278
samples = 4
value = [5, 1]
class = No Fraud
Age <= 33.5
gini = 0.434
samples = 178
value = [89, 191]
class = Fraud
ClaimSize <= 2652.303
gini = 0.202
samples = 49
value = [62, 8]
class = No Fraud
ClaimSize <= 91265.508
gini = 0.315
samples = 164
value = [52, 214]
class = Fraud
DriverRating <= 3.5
gini = 0.255
samples = 24
value = [34, 6]
class = No Fraud
AgeOfVehicle <= 6.5
gini = 0.471
samples = 212
value = [134, 219]
class = Fraud
Age <= 64.5
gini = 0.498
samples = 32
value = [34, 30]
class = No Fraud
DriverRating <= 1.5
gini = 0.265
samples = 28
value = [43, 8]
class = No Fraud
Age <= 43.5
gini = 0.156
samples = 27
value = [43, 4]
class = No Fraud
ClaimSize <= 53278.426
gini = 0.051
samples = 27
value = [37, 1]
class = No Fraud
Age <= 55.0
gini = 0.282
samples = 76
value = [93, 19]
class = No Fraud
gini = 0.0
samples = 1
value = [0, 1]
class = Fraud
gini = 0.0
samples = 3
value = [5, 0]
class = No Fraud
AgeOfVehicle <= 3.5
gini = 0.46
samples = 115
value = [63, 113]
class = Fraud
ClaimSize <= 3465.296
gini = 0.375
samples = 63
value = [26, 78]
class = Fraud
Age <= 33.0
gini = 0.346
samples = 12
value = [14, 4]
class = No Fraud
DriverRating <= 2.5
gini = 0.142
samples = 37
value = [48, 4]
class = No Fraud
Age <= 16.5
gini = 0.28
samples = 159
value = [43, 212]
class = Fraud
AgeOfVehicle <= 4.5
gini = 0.298
samples = 5
value = [9, 2]
class = No Fraud
AgeOfVehicle <= 6.5
gini = 0.074
samples = 14
value = [25, 1]
class = No Fraud
Age <= 26.5
gini = 0.459
samples = 10
value = [9, 5]
class = No Fraud
DriverRating <= 2.5
gini = 0.278
samples = 17
value = [4, 20]
class = Fraud
ClaimSize <= 51150.979
gini = 0.478
samples = 195
value = [130, 199]
class = Fraud
DriverRating <= 2.5
gini = 0.083
samples = 8
value = [22, 1]
class = No Fraud
Age <= 70.0
gini = 0.414
samples = 24
value = [12, 29]
class = Fraud
AgeOfVehicle <= 7.5
gini = 0.278
samples = 7
value = [10, 2]
class = No Fraud
Age <= 45.5
gini = 0.26
samples = 21
value = [33, 6]
class = No Fraud
AgeOfVehicle <= 7.5
gini = 0.42
samples = 6
value = [7, 3]
class = No Fraud
AgeOfVehicle <= 7.5
gini = 0.053
samples = 21
value = [36, 1]
class = No Fraud
gini = 0.0
samples = 16
value = [23, 0]
class = No Fraud
Age <= 28.0
gini = 0.124
samples = 11
value = [14, 1]
class = No Fraud
ClaimSize <= 2174.022
gini = 0.32
samples = 67
value = [76, 19]
class = No Fraud
gini = 0.0
samples = 9
value = [17, 0]
class = No Fraud
(...)

Four Main Variables

from Random Forest

Claim Size
Frauds occur more often at small claim sizes

Age
Frauds occur more between ages 25-55

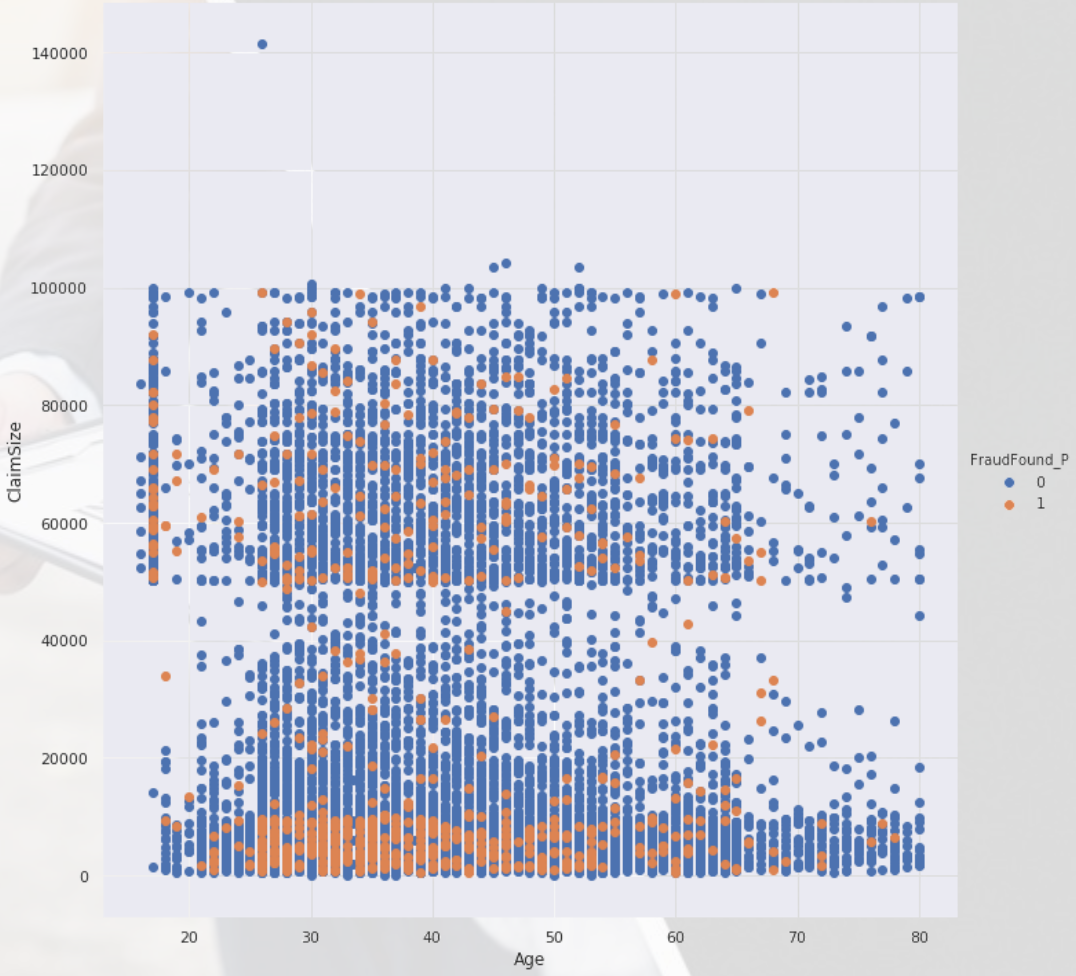
Policy Type
"Liability" policies have significantly less fraud

Fault
Frauds occur more often when the policyholder is at fault

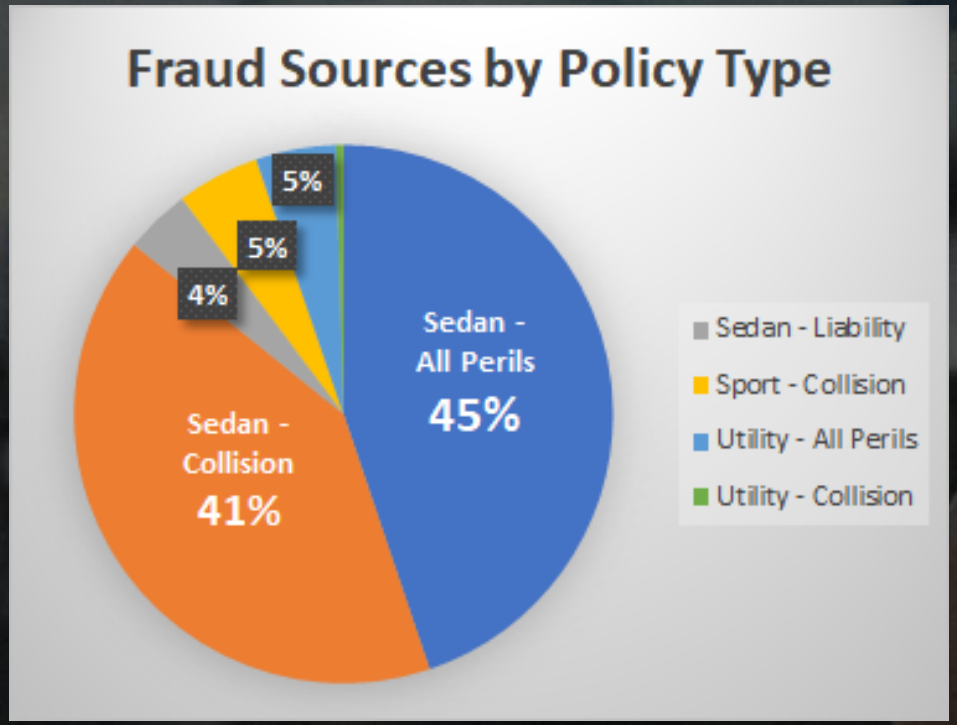
Age, Claim Size, and Fraud

Most fraudulent claims are **Under $10,000**

Most fraudulent claims are **Ages 25-55**



Policy Type



Proportion of Fraud in each Policy Type:

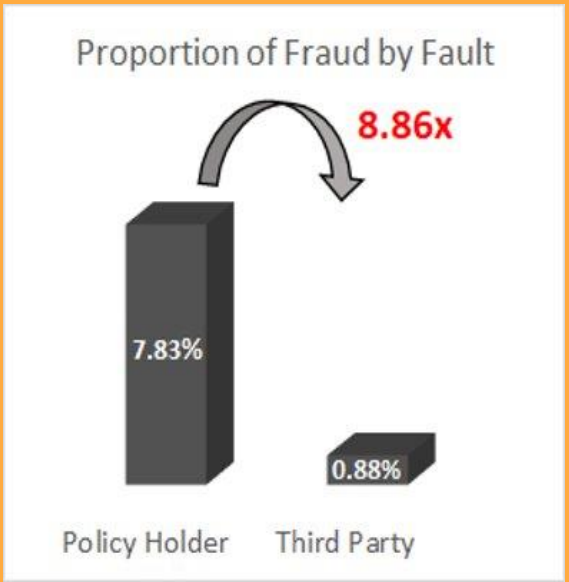
10.01% of Sedan All Perils

6.71% of Sedan Collision

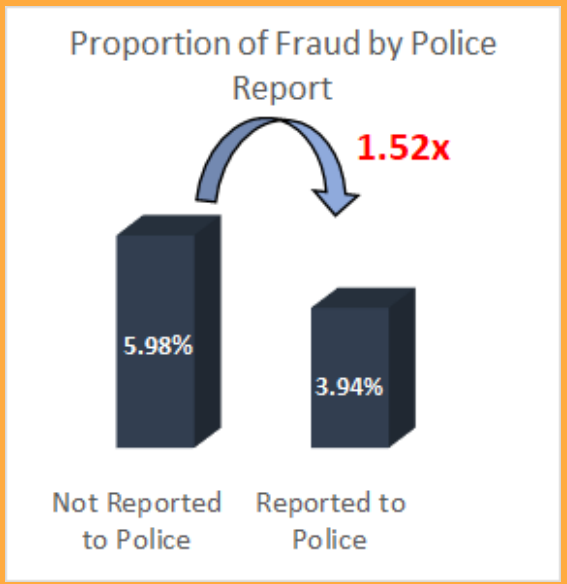
13.42% of Utility All Perils

12.98% of Sport Collision

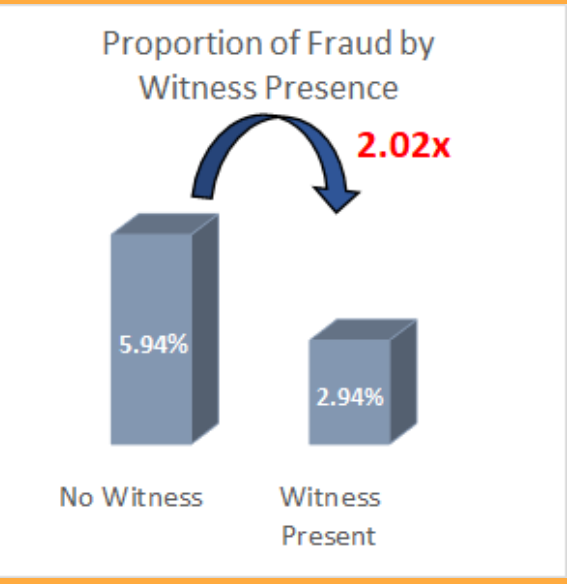
Fault and Other Fraud Insights



Intended self-loss

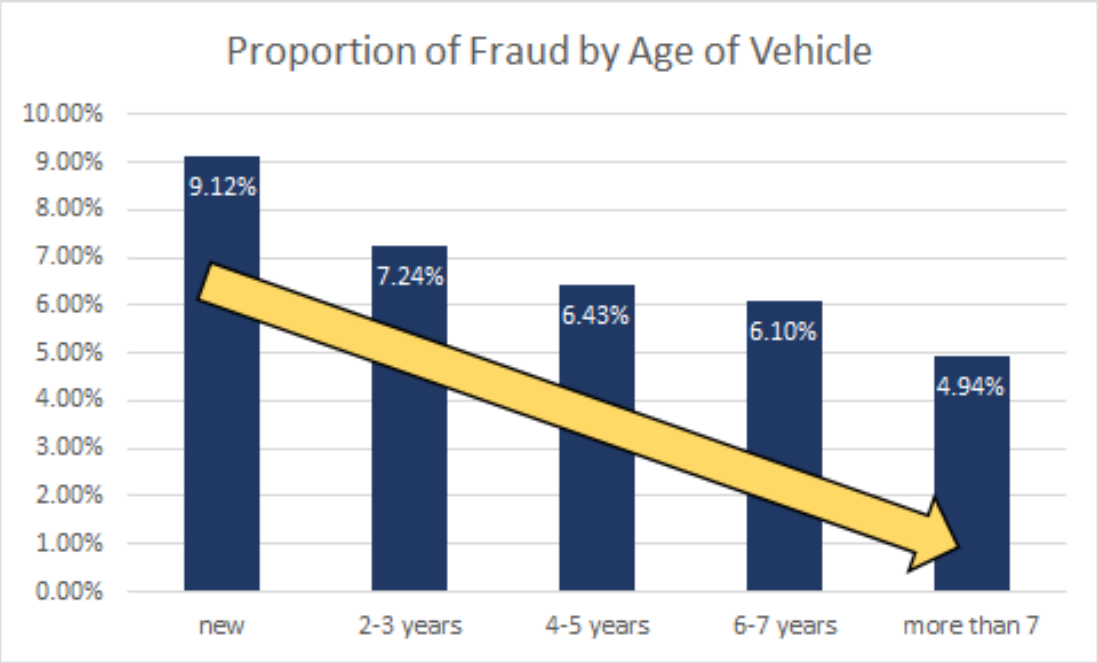


Unreported accidents



Claims without witnesses

Other Fraud Insights



3 out of 4

claims with an address change of less than six months are frauds

14.41%

of claims with an address change of less than 3 years are frauds

Strategy 1: Enhanced Risk-Scoring System

What?

Create or adjust an existing risk-scoring system based on the most important variables given by the Random Forest classifier

Why?

Increased chance of detecting fraudsters and classifies high-scoring clients as high risk

How?

- Determine the variables that need adjustments
- Adjust the scoring weight of the variable
- Integrate the new findings to the risk-scoring systems
- Dedicate data scientists to predictive modelling

Weight Adjustments for Enhanced Risk Scoring

Variable	Risk Weight Adjustment
Age: 25 to 55	↑↑
Claim Size: Less than $10,000	↑↑
Policy Type: Liability	↓↓↓
Policy Type: All Perils	↑
Fault: Policyholder	↑↑
Police Report: Not filed	↑
Witness: Not present	↑
Age of Vehicle: New Vehicles	↑

Strategy 2: Partnering with an IoT Sensor Company (Biz4Intellia)

What?

Partner with Biz4Intellia to install sensors in clients' vehicles

Why?

Data obtained from sensors will be used to cross-check claims during investigation

How?

Real-time sensor data will be used to help identify fraud during the investigation process

After an initial testing period, policies may be expanded to all clients

Negotiate with Biz4Intellia to collaborate

Restructure policies to include the sensor requirements for high-risk clients

Sensor data also serves as an external data source for further data analysis and research

Biz4Intellia

Top 50
IoT Companies
by CIOReview

Biz Intellia
An end-to-end IoT Solution

Only
$2
per device

Est.
2017

Low dealing fee

Problems | Objectives | Decision Criteria | Recommendation | Data | Analysis | Implementation Plan | Contingency Plan

Biz Intellia
An end-to-end IoT Solution

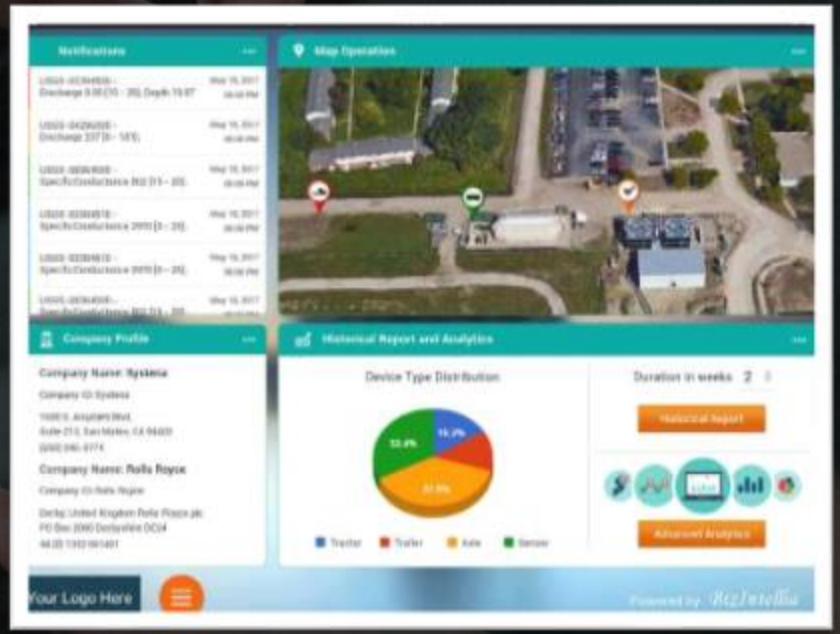
Information provided by sensor:

Driving Time	Idle Time	Distance Travelled
Average Speed	Last Location	Collisions Record

Possible benefits from this partnership

Promotes good driving behavior , as drivers know their activity will be 'monitored'. Sensors send alerts when drivers exceed speed limits.

External data source for future data analytics proce dure.



Strategy 3: Screening Checklist for Claims

What?

A check-list with the indicators that were deemed as significant by the Random Forest process

Why?

This filters a significant number of claims so that only the suspicious ones will be reviewed; this saves labour, time and money

How?

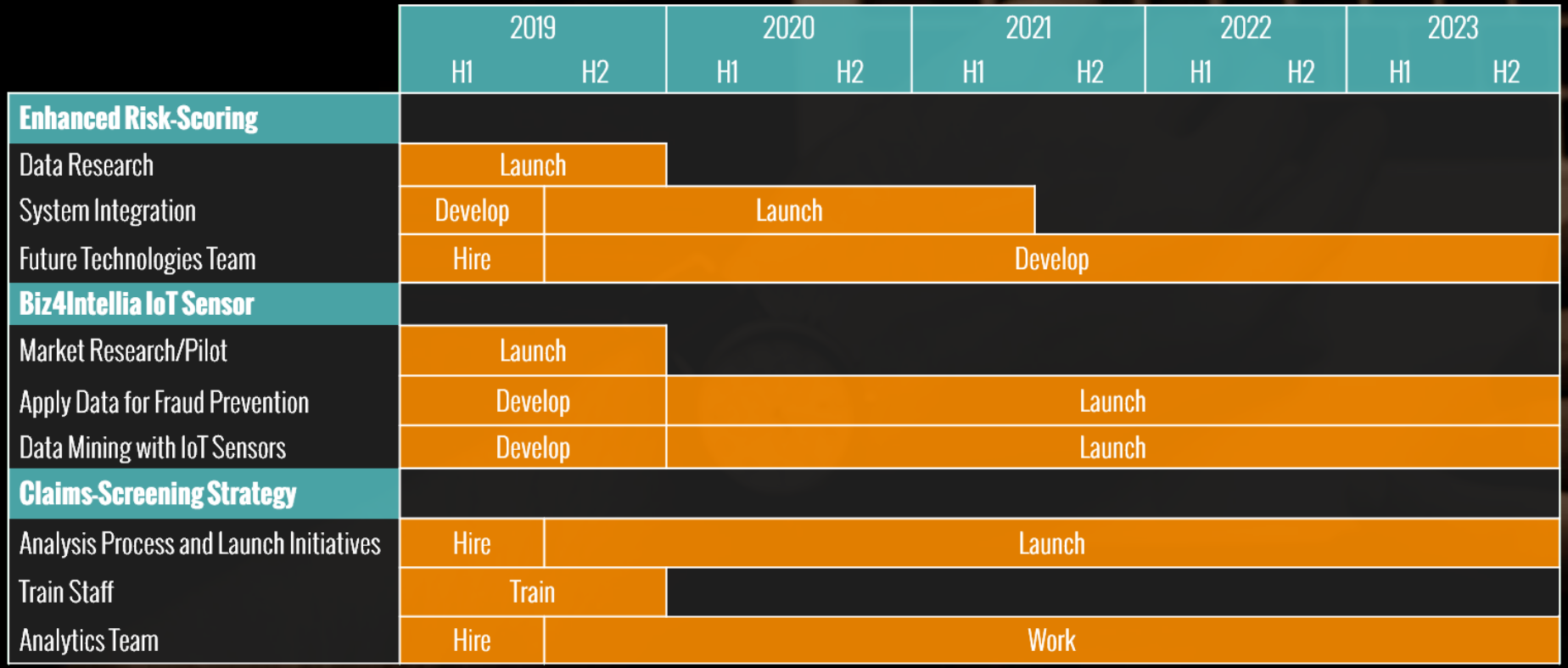
Create a checklist with the significant indicators

All claims undergo the checklist process

Claims that are found to be suspicious will be further reviewed

The checklist will be continually updated with additional data

Implementation Timeline



Contingencies

Anticipated Risks	Mitigations
Predictive modelling presented by Random Forest may be made more accurate	Explore other strategies as the field of data science advances; hire data scientists
Predictive modelling presented by Random Forest may not represent future trends	Collect up-to-date data and continually update the model with new data
Policy restructuring may not be well-received by customers	Marketing strategies by marketing team
Enhanced risk-scoring and screening checklist do not deter all types of frauds	Update techniques with more information and Complement with anti-fraud campaigns


Thank You

References

Biz4Intellia. (2019). *IoT in Transportation.* Retrieved from Biz4Intellia: https://www.biz4intellia.com/iot-in-transportation/

Capgemini. (2017, December). *Top 10 Trends in Property & Casualty Insurance 2018.* Retrieved from Capgemini Insurance: https://www.capgemini.com/wp-content/uploads/2017/12/property-and-casuality-insurance-trends_2018.pdf

IMA: The Association of Accountants and Financial Professionals in Business. (2007). *Enterprise Risk Management: Tools and Techniques for Effective Implementation.* Montvale: IMA.

Sibal, N. (2019). *Using Advanced Analytics to Identify Fraud in Property and Casualty Insurance*. Retrieved from EXL Service: https://www.exlservice.com/using-advanced-analytics-to-identify-fraud-in-property-and-casualty-insurance