CIS of Reported Child Abuse and Neglect

Practical Data Science

P. Boily, Y. Huang

Contents

Project Description

Project Description
Context

- The *Canadian Incidence Study of Reported Child Abuse and Neglect* (CIS) is a national surveillance program dedicated to the **health of children in Canada** which examines incidences of reported child maltreatment and characteristics of the children and families investigated by Canadian child welfare sites from all 13 subnational jurisdictions

- Minor methodological changes were introduced over the years:
  - increased sample size every cycle
  - differences in jurisdictional oversampling strategies
  - increased sample from First Nations agencies, etc.

- Prior to the CIS 3rd cycle (2008), prevalence estimates were for 5 types of child maltreatment:
  - **physical** abuse
  - exposure to **family violence**
  - **neglect**
  - **sexual** abuse
  - **emotional** abuse

- From 2008 onwards, a 6th type was added: **risk** of maltreatment

Project Description
Context

- Child protection workers (CPWs) reported on whether any type of abuse was
  - **Substantiated** (evidence indicated it happened)
  - **Unfounded** (did not happen)
  - **Suspected** (may have happened but cannot be definitively proven)
- CPWs could indicate that a maltreatment allegation was substantiated for risk (2008+)
- Possible outcomes for risk of future maltreatment are recorded as
  - **Significant** (or Confirmed)
  - **Not Significant** (or Denied)
  - **Unknown**
- The changes were brought about because:
  - there could be a concern that a maltreatment incident may have occurred (which would be reported as **Substantiated** or **Suspected**);
  - but even if such an incident was not substantiated or suspected, there may be **Significant** risk of **future** maltreatment

Project Description
Context

Maltreament		Future Risk	
Substantiated	36%	Confirmed	5%
Suspected	8%	Unknown	4%
Unfounded	30%	Denied	17%
Total:	74%	**Total:**	26%

Distribution of CIS Investigation Types (2008)

- Prior to 2008, the CIS variables did not include the type of investigation for specific cases (but the overall distribution was reported)
- The *Public Health Agency of Canada* (PHAC) is interested in **determining what the distribution of investigation types would have been for the CIS 2nd Cycle (2003)**, had Future Risk been reported – a **classification** task.

Project Description
Data

- The working dataset contains 16,372 investigations from the CIS 2008 data (some of which have been imputed). Quebec data not included in the working dataset due to methodological differences in 2003.

- Dataset contains 201 variables (original and derived) from 5 explanatory and 2 response categories:
  - **Caregiver** – sex of primary caregiver, attended residential school, etc.
  - **Child** – attachment issues, inappropriate sexual behavior, academic difficulties, etc.
  - **Household** – home overcrowded, accessible drug paraphernalia, etc.
  - **Intake** – referral from custodial parent, from community or social services personnel, etc.
  - **Services** – placement during investigation, in-home family/parent consulting, etc.
  - **Investigation** – type of investigation, previous report for suspected maltreatment, etc.
  - **Maltreatment** – primary, secondary, tertiary; investigated, suspected, substantiated for physical, exposure, neglect, sexual, emotional, risk

- Investigation variables and Maltreatment variables are nearly in **alignment**. The model then should predict what kind of investigation was conducted, as well as the investigation's output (what happens otherwise?)

Modeling

Modeling
Assumptions

- For the majority of the data elements, the collection strategies have not changed significantly over the cycles; CPWs would record and collect the **same answers** to the **same questions** in **similar situations** in each cycle.

- Within a cycle, the data does not differ significantly from jurisdiction to jurisdiction; there are **no essentially different substantiation patterns in different jurisdictions**.

- Our final assumption is that the investigators and questionnaire designers are not introducing **systematic bias** into the dataset.

- In practice, there are bound to be small discrepancies from cycle to cycle, and even from investigator to investigator within the same cycle, and perhaps even from investigation to investigation for a given investigator, but that is an internal matter.

Modeling
Methodology

- After some preliminary experiments and discussions with the client, **conditional inference decision trees** (with recursive partitioning) augmented by a **boosting strategy** were selected as the modeling approach, because decision trees:
  - easily lend themselves to **interpretation** and statistical analysis;
  - require minimal data preparation (compared to other methods);
  - easily accommodate various data types and missing observations;
  - perform "well" with large datasets, and
  - are robust against small data departures from theoretical assumptions.

- **On the other hand**, they may
  - fall prey to **over-fitting**
  - require **manual pruning**
  - be biased in favour of attributes with a **large number of categories**

Modeling
Methodology

- Abstractly, any decision tree is grown as follows:
  1. a stopping criterion determines if the tree is to be grown further from a given **branch** or if that branch's **leaf** been reached
  2. if required, a **branching variable** (node) is selected
  3. an appropriate **splitting level** is selected to partition the data on the branching node
  4. Steps 1., 2. and 3. are repeated until the stopping criterion is met for all branches

- CI Trees can help overcome some of the limitations, as the stopping criterion, branching variables and splitting levels are **computed automatically from statistical properties** of the data.

- As a result, overfitting is unlikely to be an issue, and manual pruning is not needed.

- CI Trees are implemented in the R package `party`'s function `ctree()`.

Modeling
Training and Testing Sets

- The first act consists in splitting the dataset upon which the model is built into **training** and **testing** sets.

- There are no hard and fast rule regarding the size of these sets.

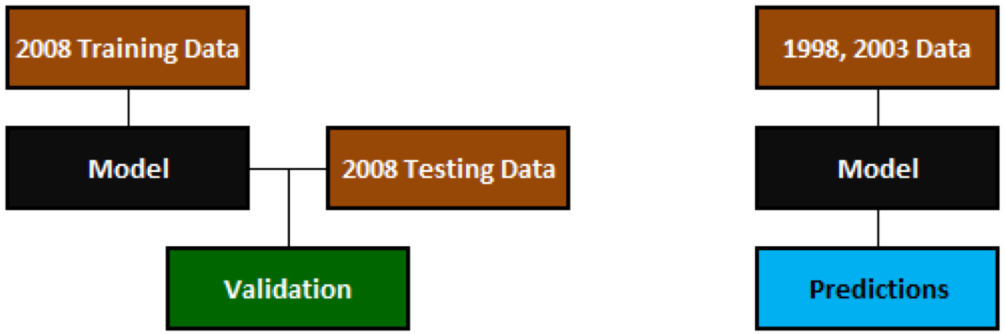
- A basic experimental principle is that using too large a training set can lead to overfitting, whereas using too small a training set may not allow the model to capture the essential signal in the data.

- The boosting strategy requires **numerous training-testing pairs.**

- We generate them by giving each observation a 70% chance of being part of the training set.

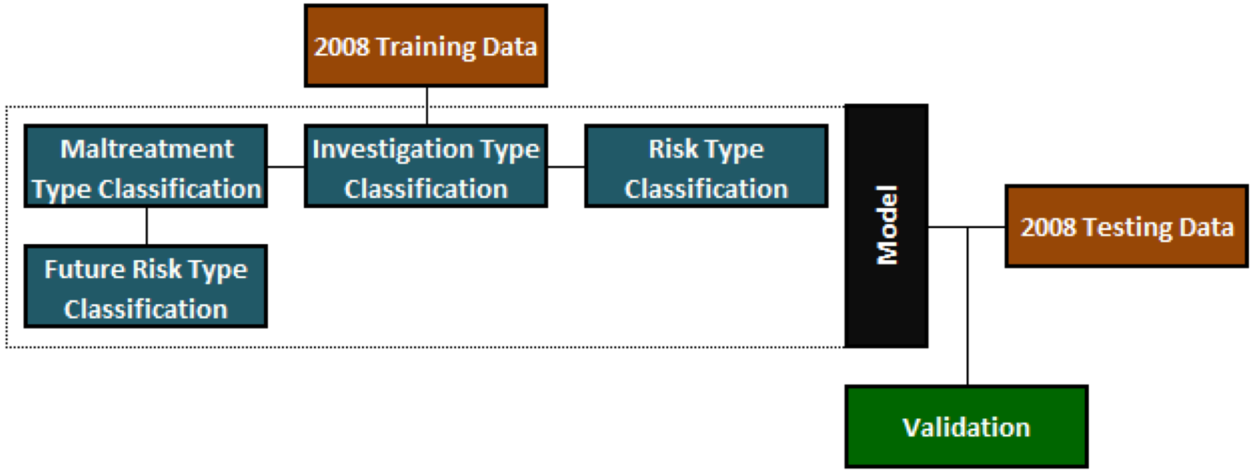
- Given that there are 16,372 cases in total, we would expect the training sets to contain
$$0.7 \times 16{,}372 \; = \; 11{,}460.4$$
cases on average, while the average size is 4,911.6 for testing sets.

Modeling
Workflow



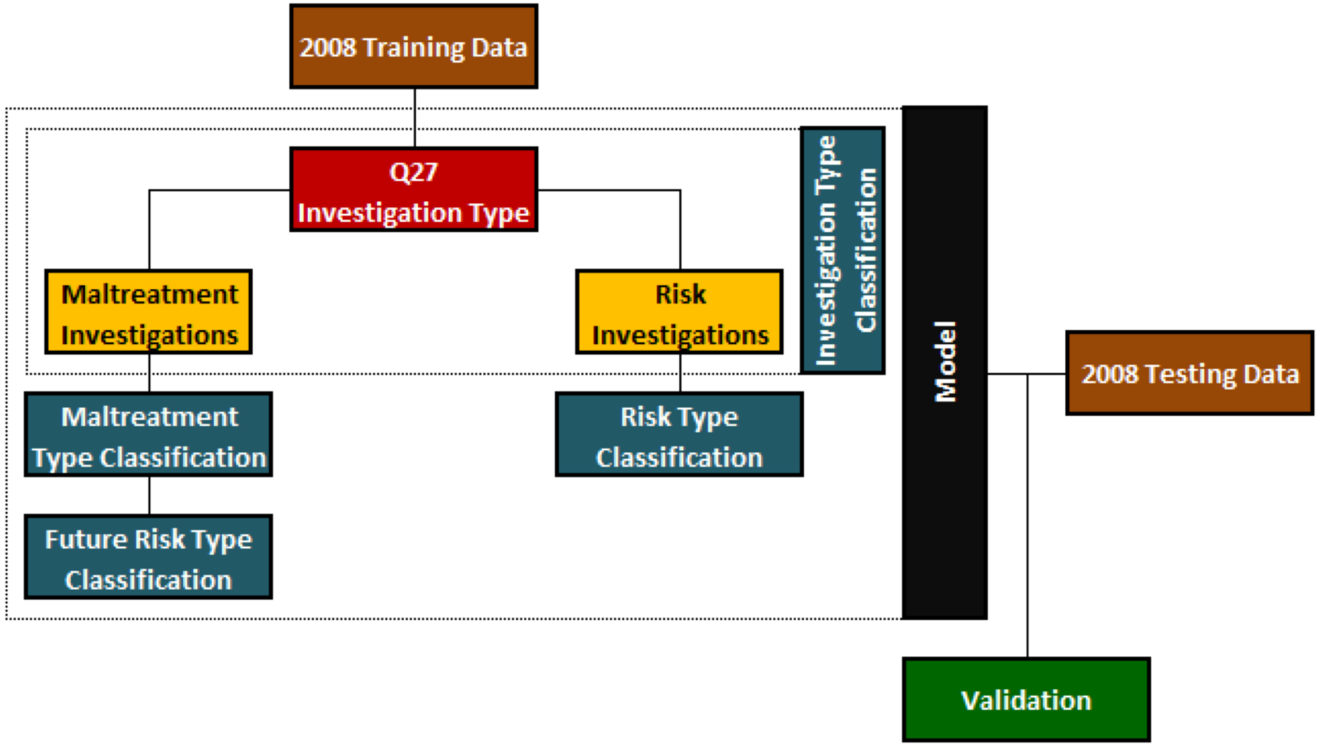
- The schematics of each classifier are shown above.
- The model is built using a training set and the testing set is used to validate the classification results, by comparing them with the actual classification (which is known but not used to build the model).
- This process is repeated multiple times and the results are “averaged” together (to be discussed further).

Modeling
Workflow



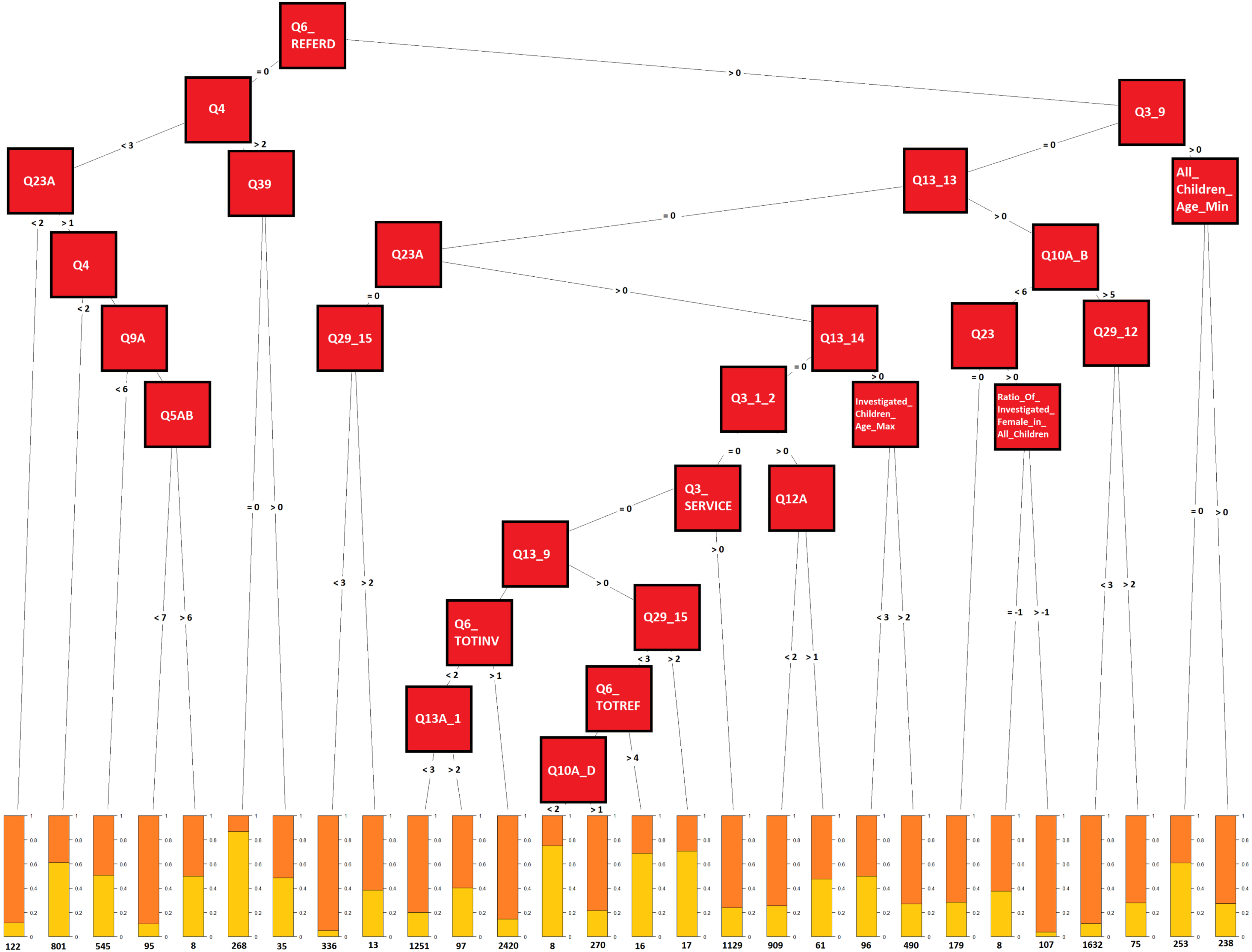
- The **Model** can be expanded further into sub-classifiers.
- We use the training data to predict the **Investigation Type**, and then use that prediction, together with the training data, to predict **Maltreatment Type** and **Risk Type**.
- We further use Maltreatment Type, with the training data and Investigation Type to predict **Future Risk Type**.

Modeling
Workflow

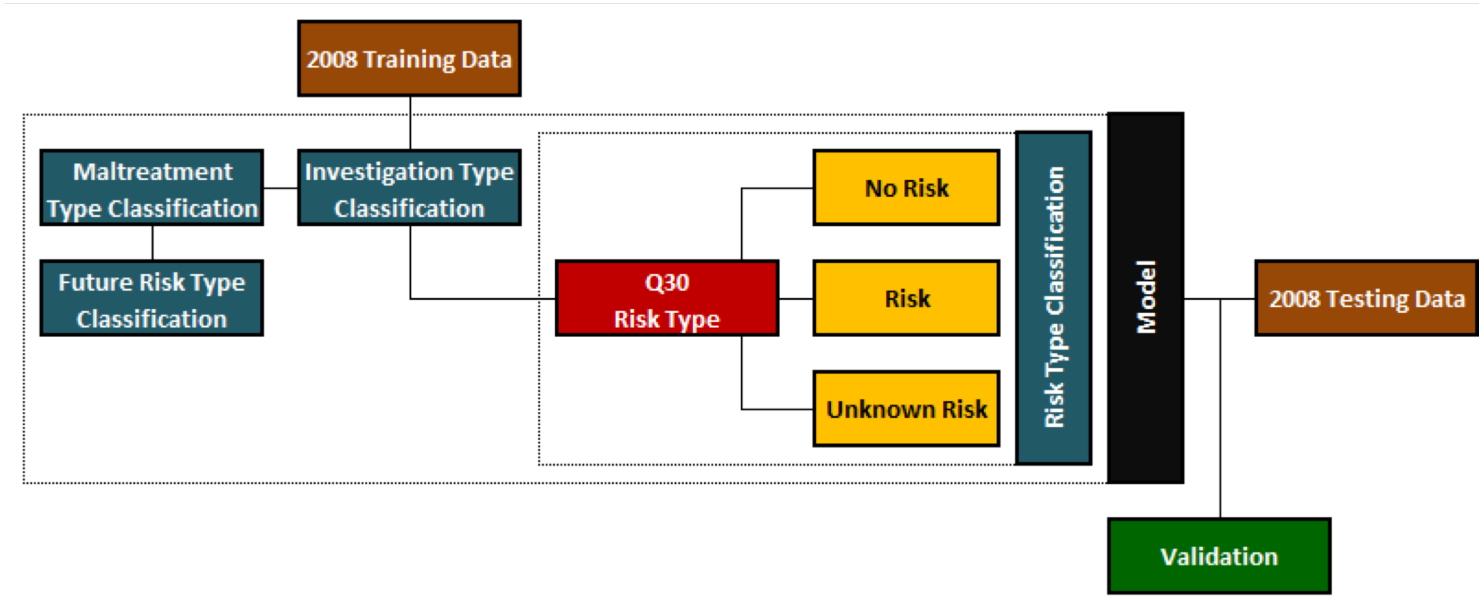


CI Tree for Q27 – Investigation Type (1 run)

Maltreatment, Risk of Maltreatment

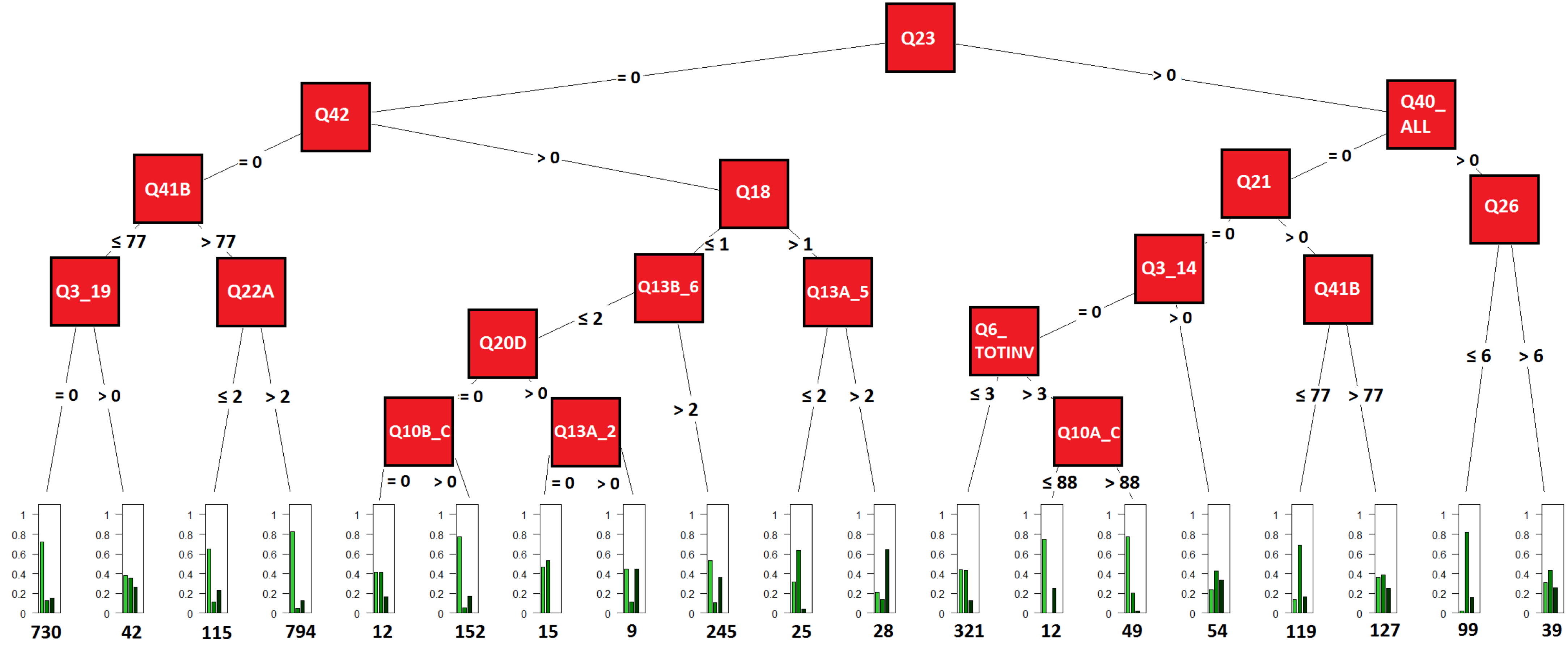


Modeling
Workflow

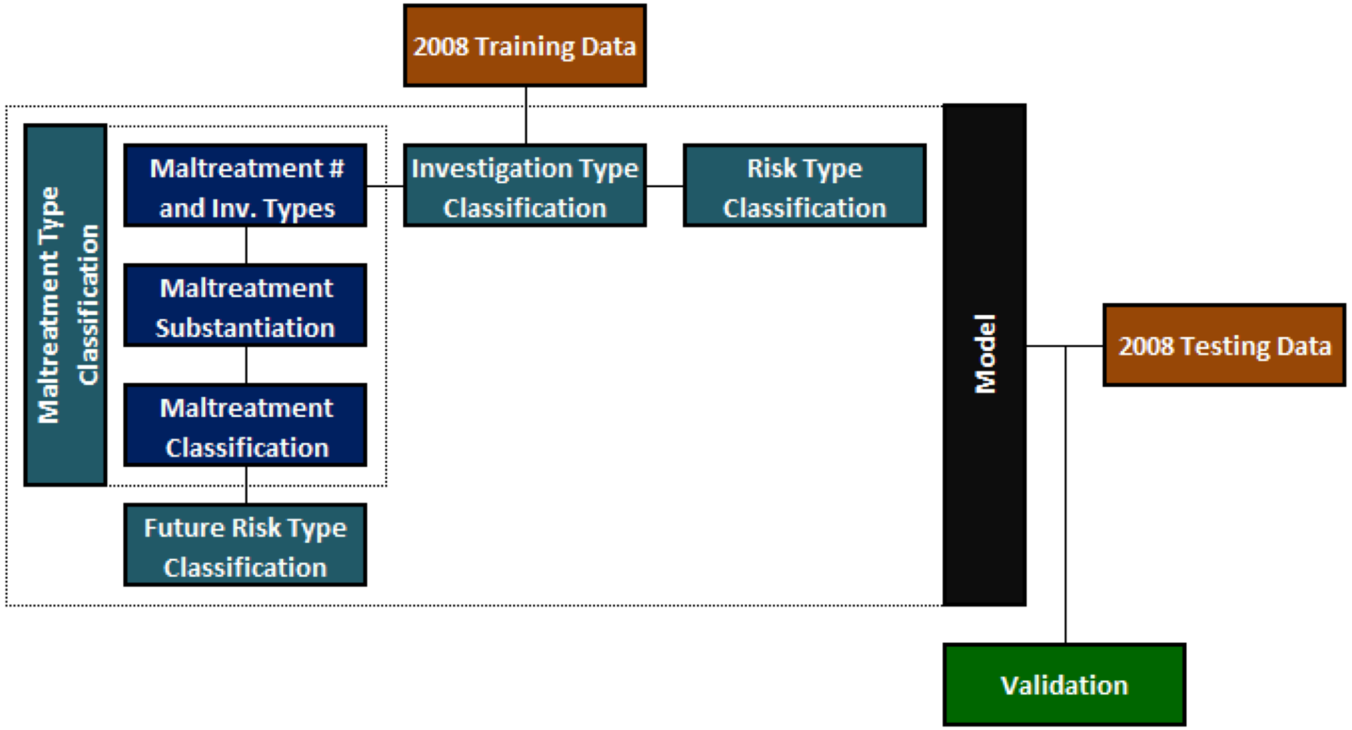


CI Tree for Q30 – Risk Type (1 rep)

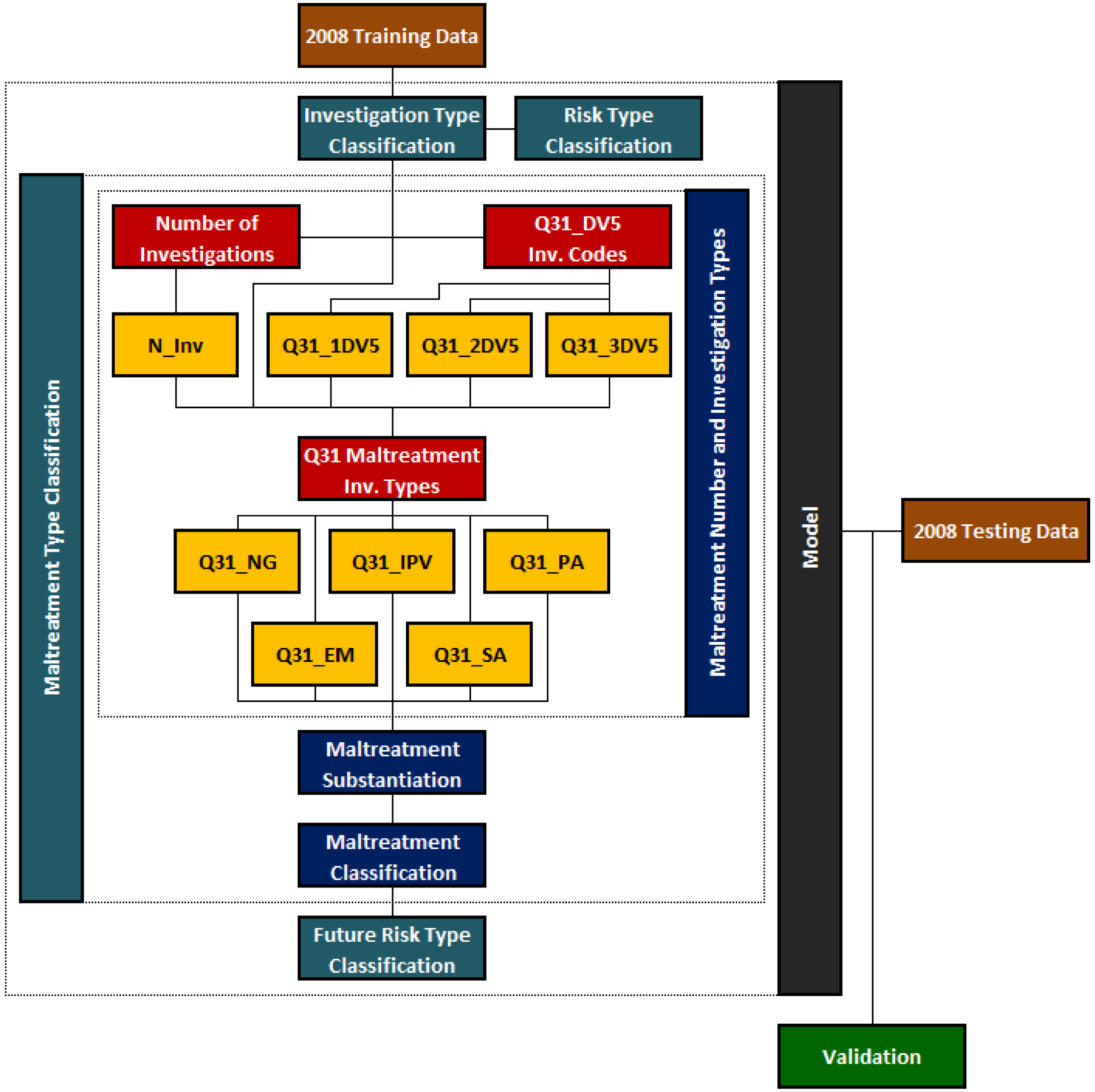
No Risk, Future Risk, Unknown



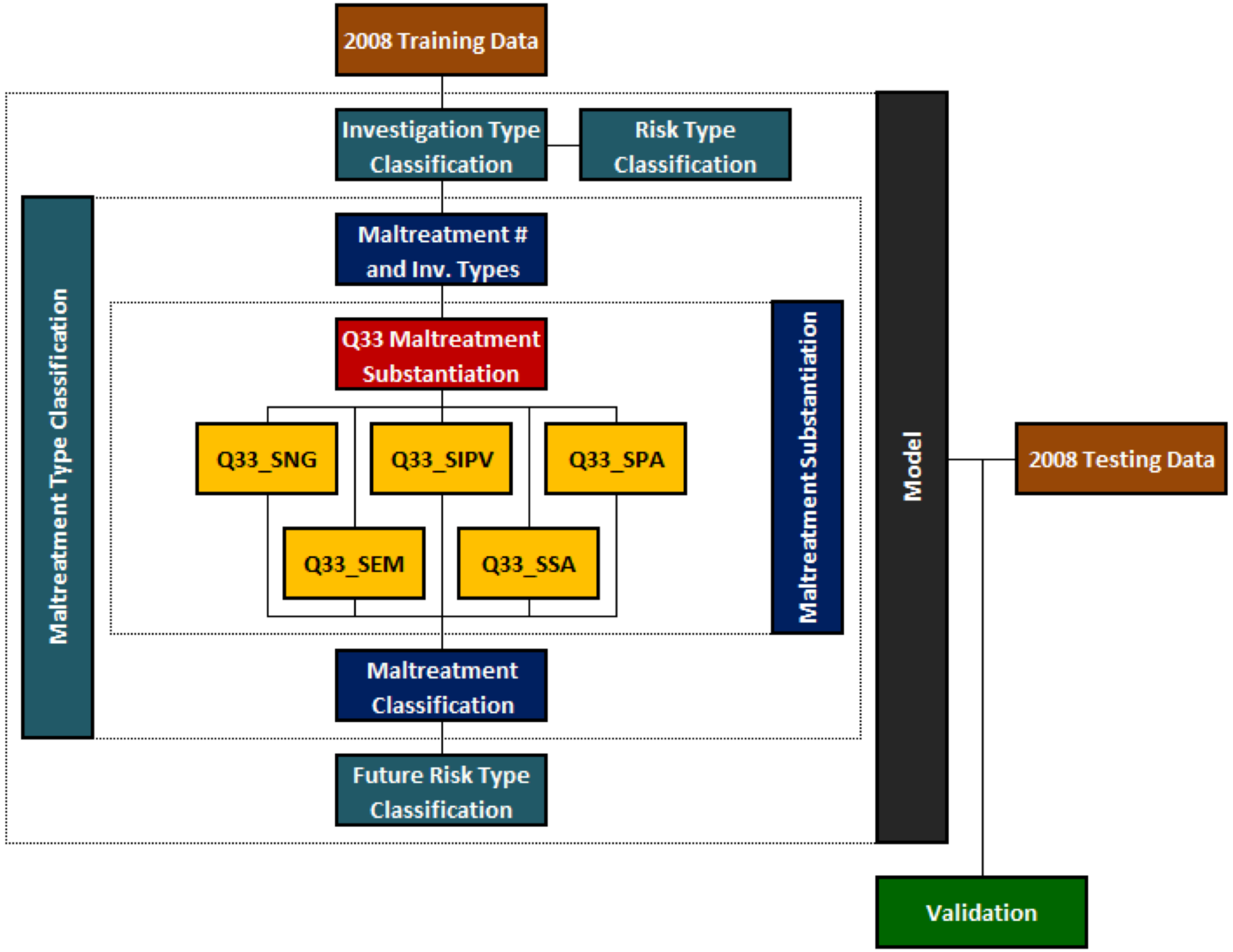
Modeling
Workflow



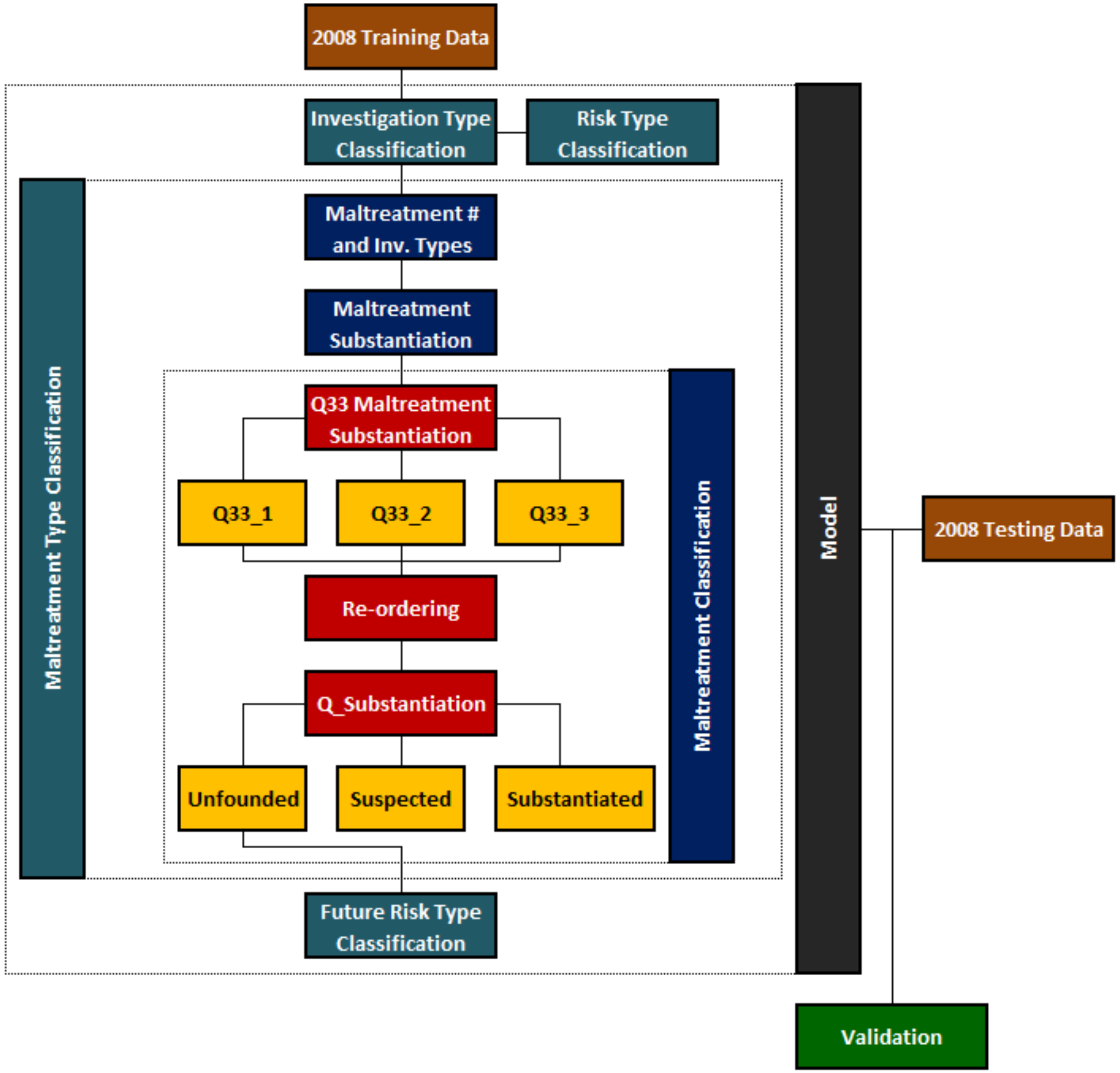
Modeling
Workflow



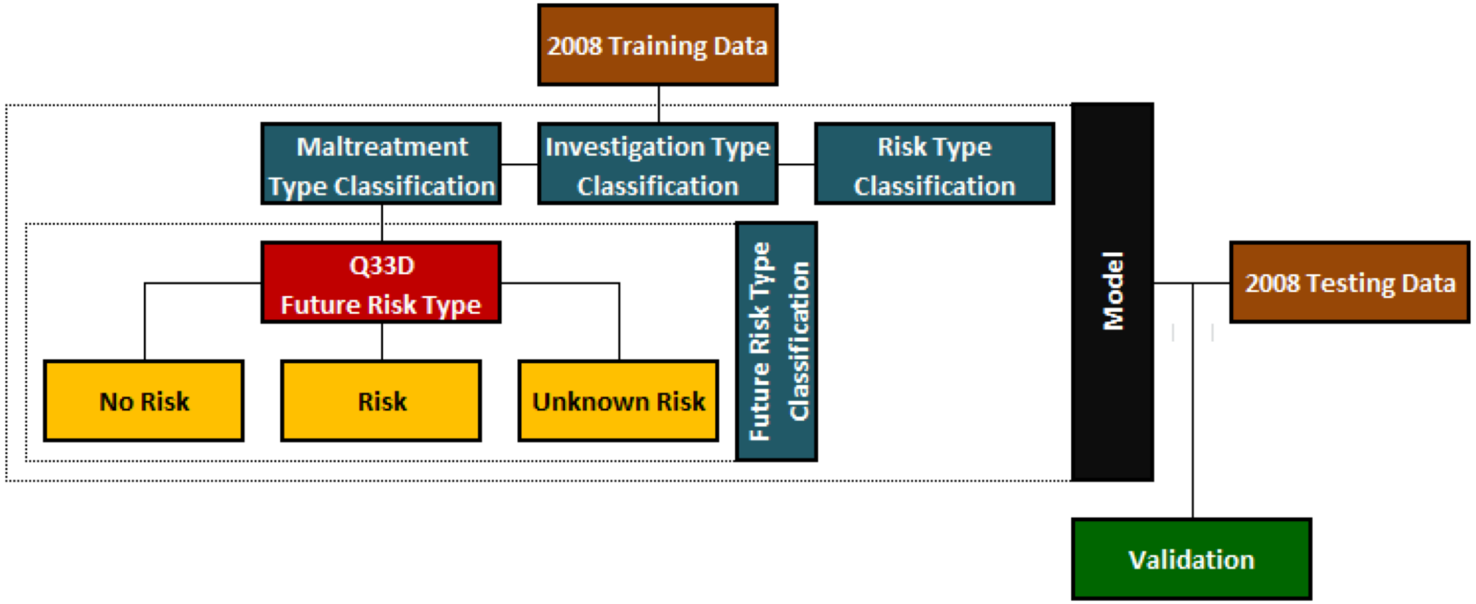
Modeling
Workflow



Modeling
Workflow



Modeling
Workflow



Results

Results
2003 Distribution

- We run the model on 50 different training-testing pairs, and produce a probability vector of classification for each observation in the 2003 dataset

Maltreament		Future Risk	
Substantiated	41.6%	Confirmed	5.3%
Suspected	6.4%	Unknown	4.5%
Unfounded	26.6%	Denied	15.6%
Total:	74.5%	**Total:**	25.5%

Results

Confusion Matrices – 2018 Data

MCC: 31.7%
Accuracy: 50.5%
Pearson: 0.28486
Hist: 47.6%

			Predicted: Maltreatment – Unfounded – No Risk	Maltreatment – Unfounded – Future Risk	Maltreatment – Unfounded – Unknown Risk	Maltreatment – Suspected	Maltreatment – Substantiated	Risk – No	Risk – Yes	Risk – Unknown	*Total*	
Actuals: Maltreatment	Unfounded	No Risk	2,097	0	47	-	712	349	4	-	*3,209*	*19.8%*
		Future Risk	84	-	26	-	146	17	8	2	*283*	*1.7%*
		Unknown Risk	413	-	173	-	465	54	2	-	*1,106*	*6.8%*
	Suspected		353	-	32	-	540	61	7	1	*995*	*6.1%*
	Substantiated		1,073	-	150	-	4,832	235	47	3	*6,339*	*39.2%*
Risk	No		791	-	76	-	793	946	23	4	*2,632*	*16.3%*
	Yes		108	1	24	1	525	118	103	2	*880*	*5.4%*
	Unknown		185	1	42	-	288	189	19	22	*745*	*4.6%*
Total			*5,103*	*1*	*569*	*1*	*8,300*	*1,969*	*212*	*34*	***16,189***	
			31.5%	*0.0%*	*3.5%*	*0.0%*	*51.3%*	*12.2%*	*1.3%*	*0.2%*		

- Prediction prepared only from models in which the observation did not appear in the training set (~15, on average).

Results

Confusion Matrices – 2018 Data

MCC: 32.8%
Accuracy: 53.8%
Pearson: 0.19586
Hist: 37.0%

			Predicted							
			Maltreatment			Risk				
			Unfounded	Suspected	Substantiated	No	Yes	Unknown	Total	
Actuals	Maltreatment	Unfounded	2,908	-	1,427	431	14	2	*4,781*	*29.2%*
		Suspected	385	-	540	61	7	1	*995*	*6.1%*
		Substantiated	1,223	-	4,832	235	47	3	*6,339*	*38.7%*
	Risk	No	867	-	793	946	23	4	*2,632*	*16.1%*
		Yes	132	1	525	118	103	2	*880*	*5.4%*
		Unknown	228	-	288	189	19	22	*745*	*4.6%*
		Total	*5,742*	*1*	*8,404*	*1,980*	*212*	*34*	***16,372***	
			35.1%	*0.0%*	*51.3%*	*12.1%*	*1.3%*	*0.2%*		

- Prediction prepared only from models in which the observation did not appear in the training set (~15, on average).

Results

Confusion Matrices – 2018 Data

MCC: 34.8%
Accuracy: 56.0%
Pearson: 0.17276
Hist: 37.0%

			Predicted					
			Maltreatment					
			Unfounded	Suspected	Substantiated	Risk	*Total*	
Actuals	Maltreatment	Unfounded	2,908	-	1,427	446	*4,781*	*29.2%*
		Suspected	385	-	540	70	*995*	*6.1%*
		Substantiated	1,223	-	4,832	285	*6,339*	*38.7%*
	Risk		1,226	1	1,606	1,425	*4,257*	*26.0%*
Total			*5,742*	*1*	*8,404*	*2,225*	***16,372***	
			35.1%	*0.0%*	*51.3%*	*13.6%*		

- Prediction prepared only from models in which the observation did not appear in the training set (~15, on average).

Results

Confusion Matrices – 2018 Data

MCC: 21.2%
Accuracy: 73.2%
Pearson: 0.21498
Hist: 40.9%

			Predicted: Unfounded – No Risk	Predicted: Unfounded – Future Risk	Predicted: Unfounded – Unknown Risk	Predicted: All Others	*Total*	
Actuals	Unfounded	No Risk	667	-	1	2,542	*3,209*	*19.6%*
		Future Risk	12	-	-	272	*283*	*1.7%*
		Unknown Risk	82	-	14	1,011	*1,106*	*6.8%*
	All Others		464	-	11	11,300	*11,774*	*71.9%*
Total			*1,224*	-	*26*	*15,123*	***16,372***	
			7.5%	*0.0%*	*0.2%*	*92.4%*		

- Prediction prepared only from models in which the observation did not appear in the training set (~15, on average).

Results

Confusion Matrices – 2018 Data

MCC: 30.5%
Accuracy: 75.6%
Pearson: 0.10314
Hist: 24.8%

			Predicted					
				Risk				
			Maltreatment	No	Yes	Unknown	*Total*	
Actuals	Maltreatment		11,314	727	68	6	*12,115*	*74.0%*
	Risk	No	1,660	946	23	4	*2,632*	*16.1%*
		Yes	657	118	103	2	*880*	*5.4%*
		Unknown	516	189	19	22	*745*	*4.6%*
Total			*14,147*	*1,980*	*212*	*34*	***16,372***	
			86.4%	*12.1%*	*1.3%*	*0.2%*		

- Prediction prepared only from models in which the observation did not appear in the training set (~15, on average).

Results

Confusion Matrices – 2018 Data

ROC: 26.9%
MCC: 34.4%
Accuracy: 77.8%
Pearson: 0.08003
Hist: 24.8%

		Predicted: Maltreatment	Predicted: Risk	*Total*	
Actuals	Maltreatment	11,314	801	*12,115*	*74.0%*
	Risk	2,833	1,425	*4,257*	*26.0%*
Total		*14,147*	*2,225*	***16,372***	
		86.4%	*13.6%*		

- Prediction prepared only from models in which the observation did not appear in the training set (~15, on average).

Consulting Post-Mortem

Consulting Post-Mortem

- Client was hoping that classification was going to be near perfect…
  - but the confusion matrices on the **training set** were not even that great
  - expectations were not managed appropriately
  - client was ultimately disappointed
- This project was not actually run well on the consultant side
  - left too many things in the hands of fate
  - analytical approach was sub-optimal in many ways
  - contract value too small (to try to get client to agree)
- Topic was distressing
  - we underestimated the effect of working with such depressing data
- All in all, not our finest hour…