Machine Learning for Rail Safety Incident Classification

Ben Merange
RGB Assurance

ASSC 2019
23rd May

Overview

- Quantitative risk models are valuable

  But often difficult to quantify

- Safety-related datasets are plentiful

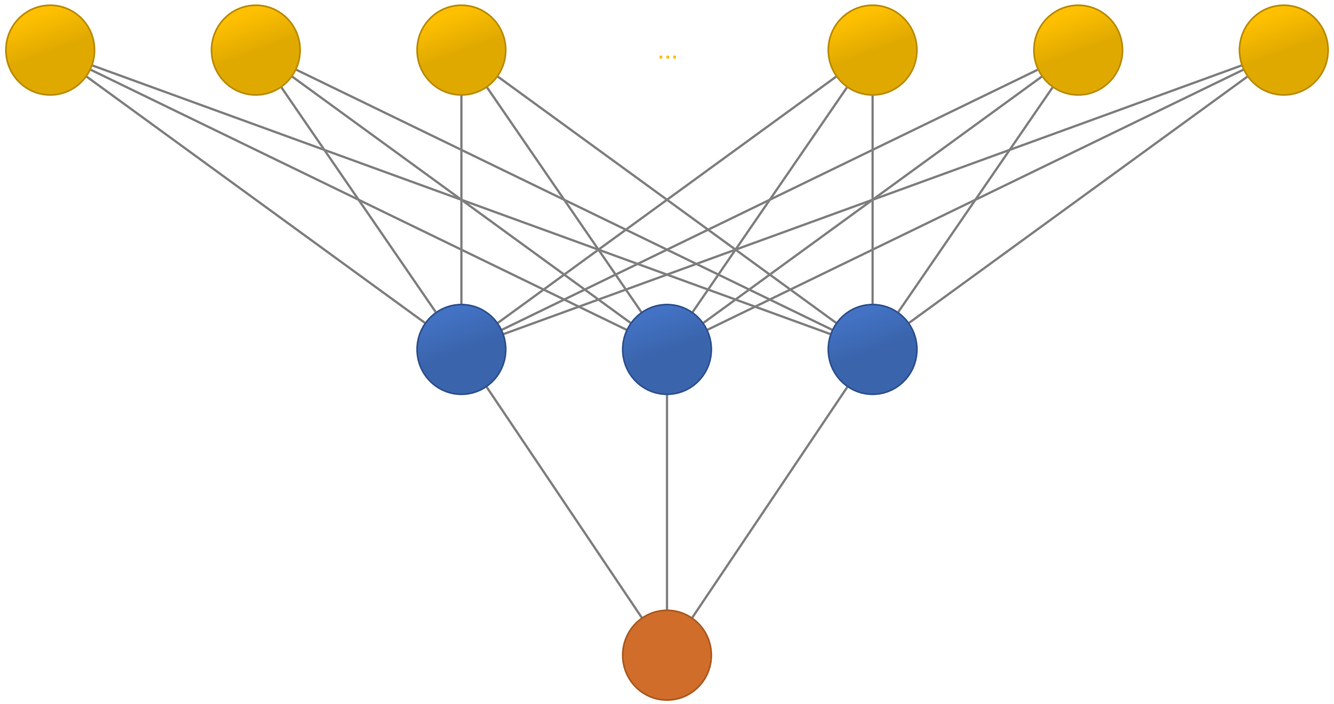
  But often unwieldy

- Machine learning is an effective tool

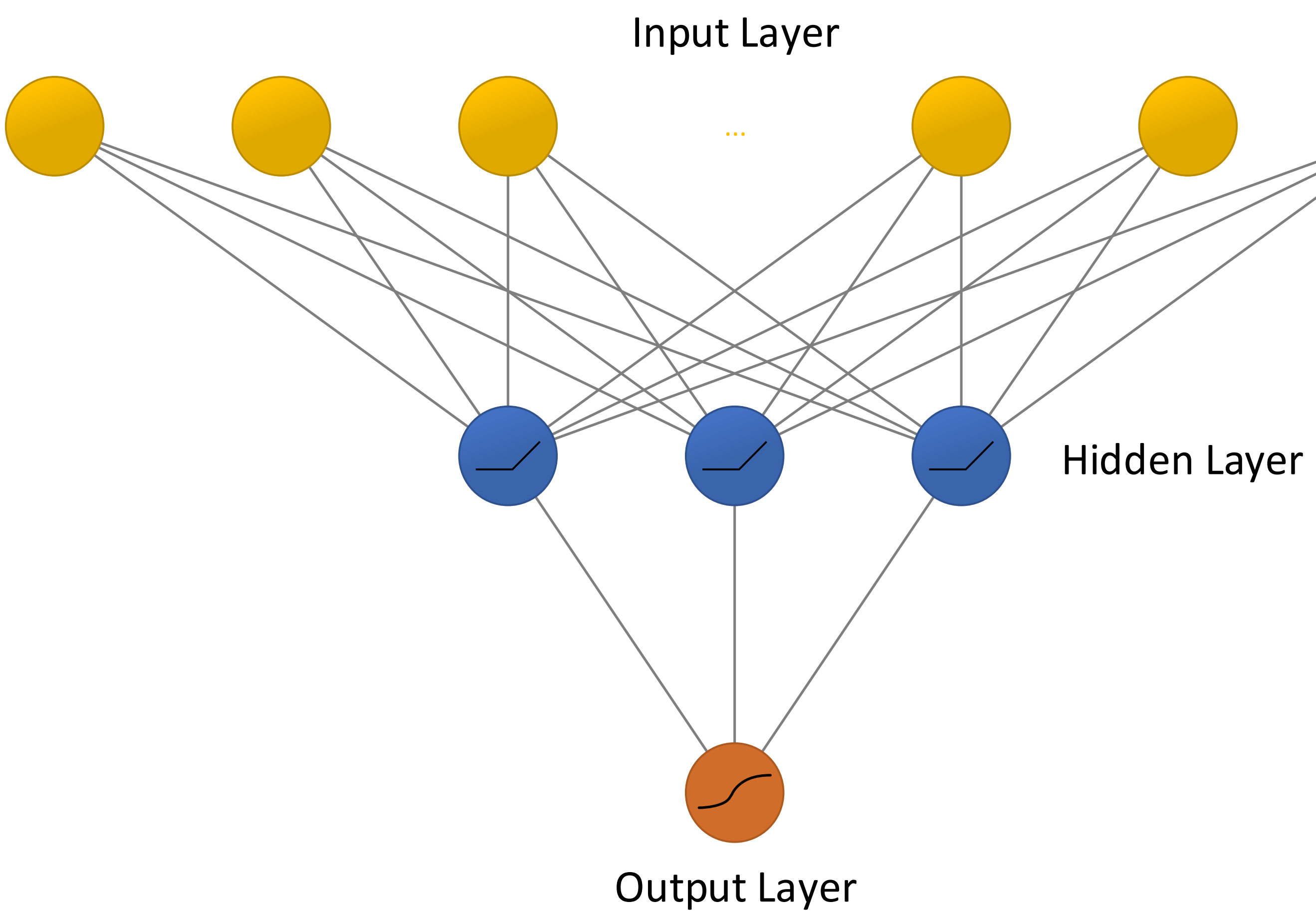
  Wrangle data to quantify risk models

Overview

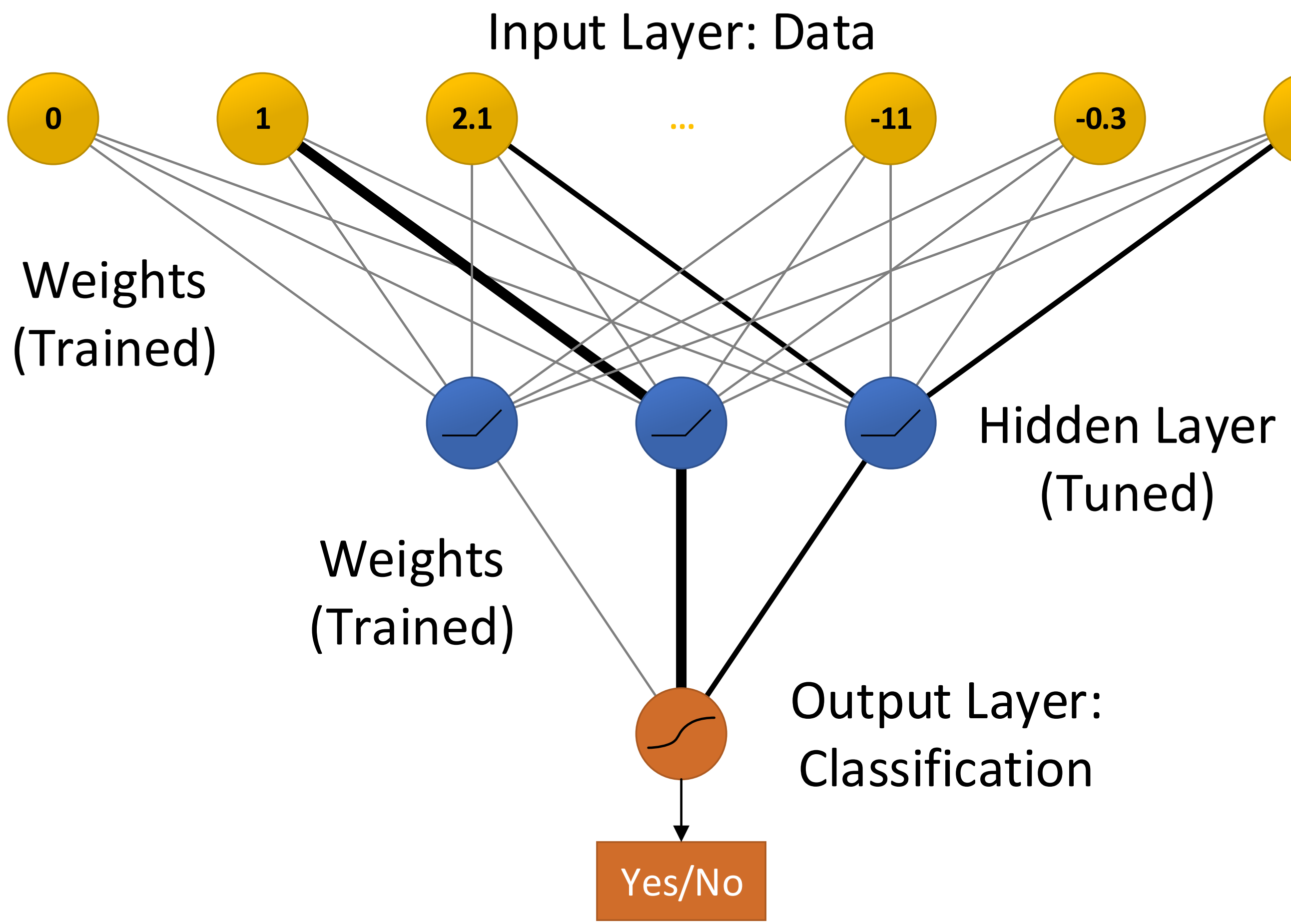
1) Neural networks
2) Example Application
   – Movie reviews
3) Our Applications
   – Rail OCG1 Categories
   – Risk modelling



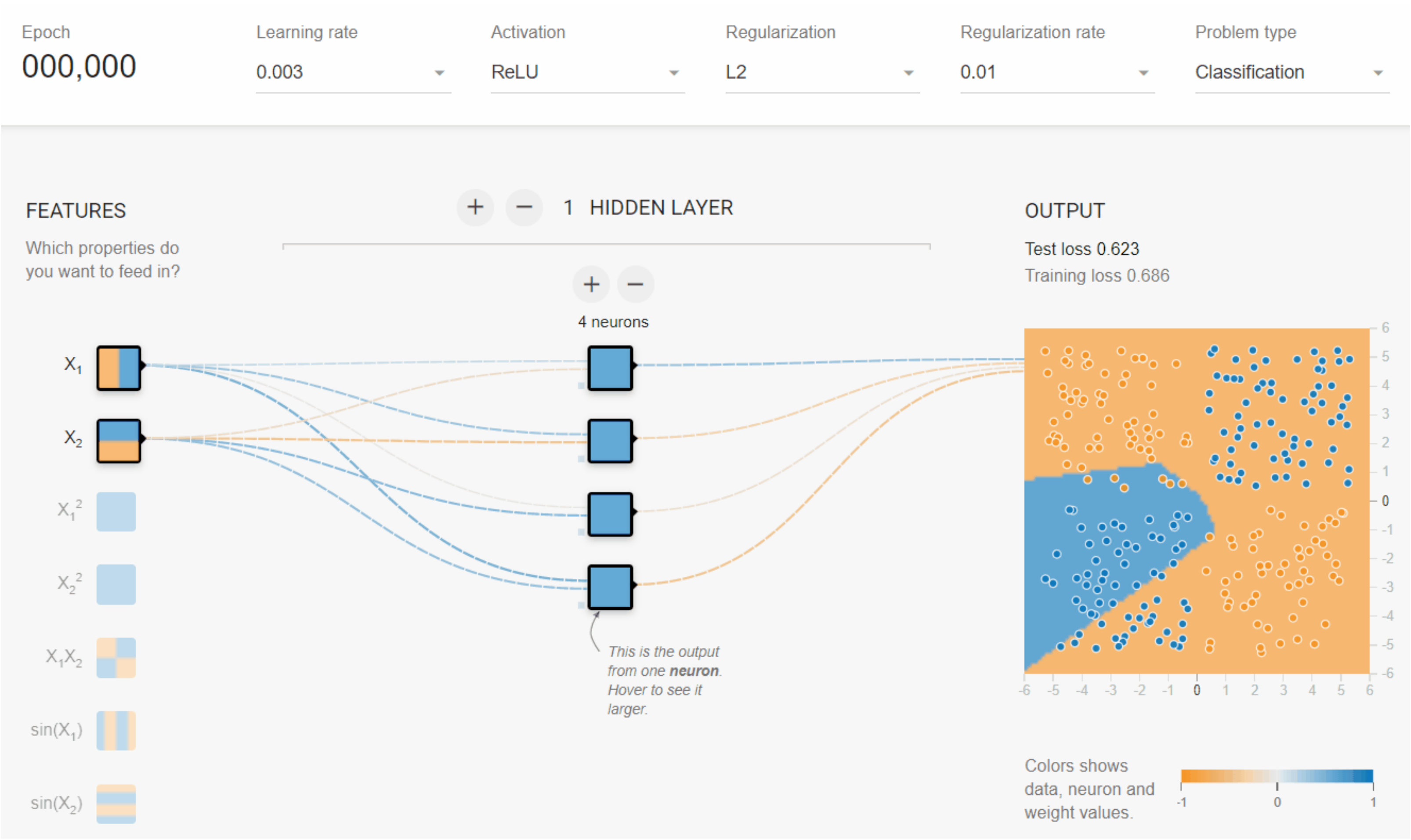
Neural Networks



Neural Networks



Neural Networks



playground.tensorflow.org

Movie Reviews

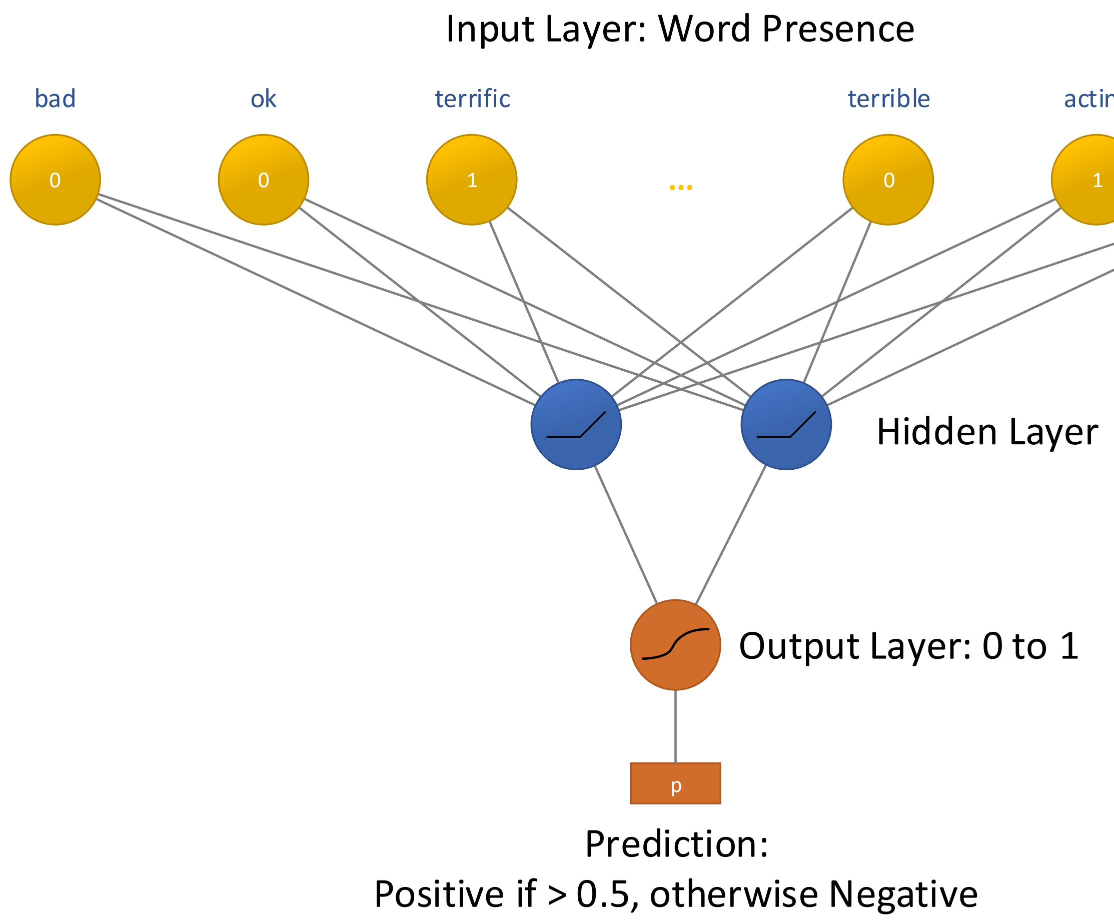
- Classify movie reviews from IMDb
  - Either positive or negative
- 50,000 textual reviews
  - Each labelled as positive or negative
  - 25,000 reviews for training
  - 25,000 reviews for evaluation
- Is there a correlation between words and labels?

Movie Reviews

Some examples:

- “This movie is terrible but it has some good effects…”
  - Negative
- “He is excellent is this film, he makes a fascinating woman…”
  - Positive
- “If you haven‘t seen this it‘s terrible. It is pure trash…”
  - Negative
- “This guy is a real genius! The movie is of excellent quality and…”
  - Positive

Movie Reviews

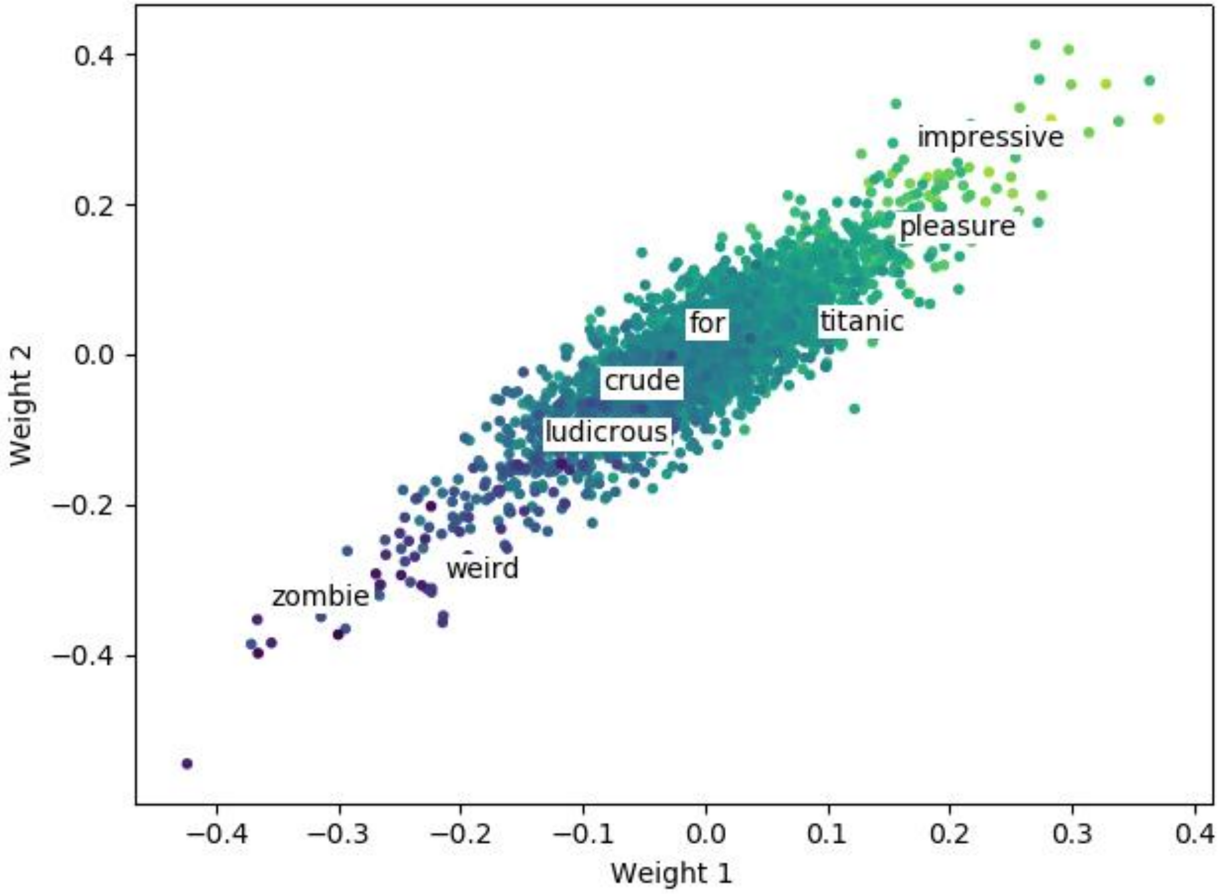


Movie Reviews

- 3 passes over the data
- Trained in 3.5 seconds (on my laptop)
- Prediction correct ~87% of the time
  - – Across 25,000 reviews

- What do the weights represent?
  - – In this case, we can plot the weights for each word in 2D

Movie Reviews



Our Applications

- For ONRSR: *Office of the National Rail Safety Regulator*
  - Map rail safety incidents to OCG1 Categories (*Occurrence Classification Guideline*)
  - Using ONRSR's dataset

- For the ARRM: *Australian Rail Risk Model*
  - Map rail safety incidents to the risk model
  - Using ARRM dataset

Rail OCG1 Categories

- Classify rail safety incidents
  - Use ONRSR's OCG1 categories
- 23 categories
  - Collision
  - Derailment
  - Fire
  - Level Crossing
  - Slip, Trip or Fall
  - …
- Used to monitor rail safety trends

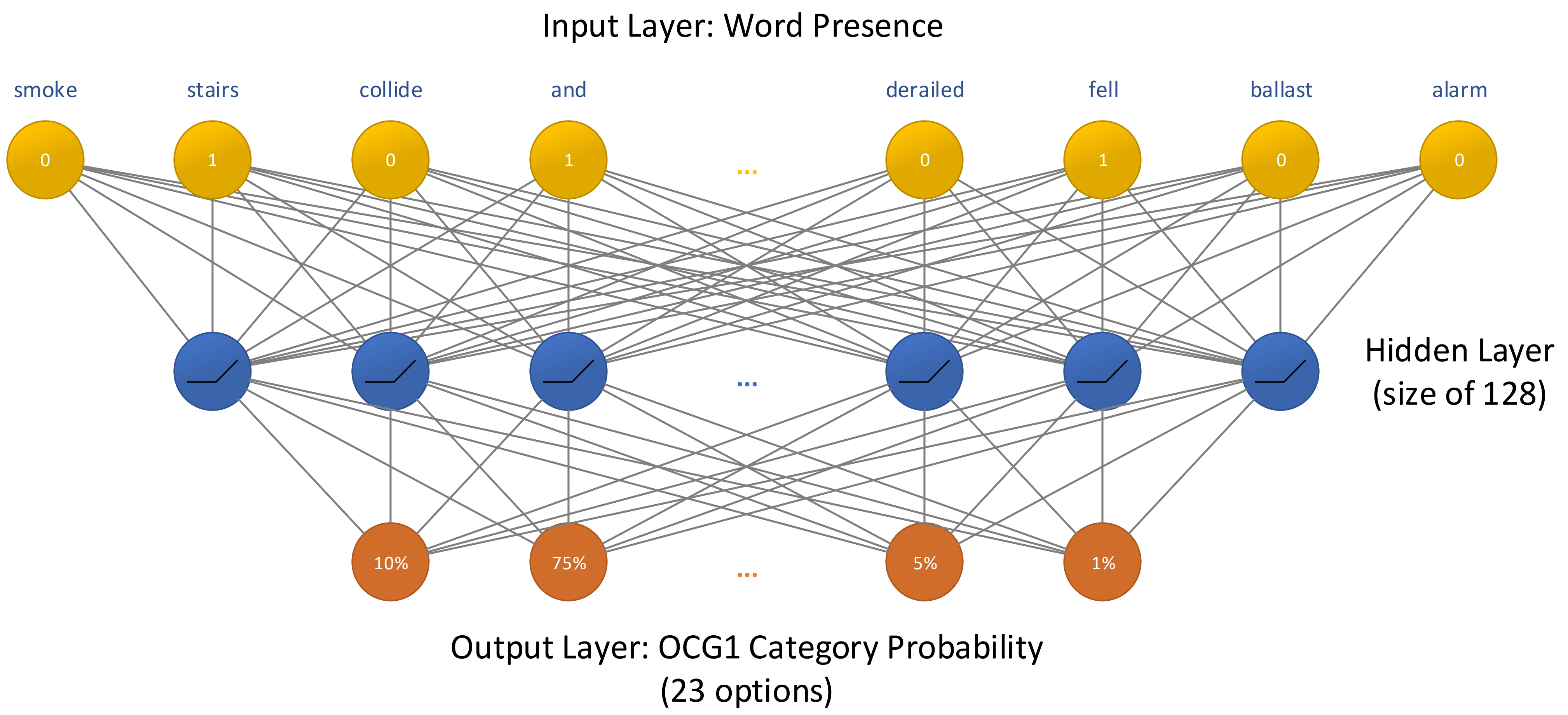
Rail OCG1 Categories

- Database of ~370,000 incidents
  - Spelling and grammatical errors
  - Domain specific terminology
- Keep 20% of data for evaluation

Example incident descriptions:

- “A person fell down the stairs leading from the concourse to platform 1. Sustained swelling to left knee. First aid applied.”
- “Points are not detecting in reverse. FCC advised. 0924hrs repairs completed. Fault caused by build up of sand in points.”

Rail OCG1 Categories



Rail OCG1 Categories

- Trained in < 1 minute (on my laptop)
- Prediction correct ~87% of the time
  - Across ~74,000 incidents
- A random guess would only be correct ~4.4% of the time

Can provide prediction confidence:

- "A person fell down the stairs leading from the concourse to platform 1. Sustained swelling to left knee. First aid applied."
  - Predicted "*Slip, Trip or Fall*" with 99.8% confidence

Rail OCG1 Categories

Can be used to detect human error:

- "Points are not detecting in reverse. FCC advised. 0924hrs repairs completed. Fault caused by build up of sand in points."
  - Incorrectly in dataset as "*Rolling stock Irregularity*"
  - Correctly predicted as "*Track and Civil Infrastructure Irregularity*" with 92% confidence

Quality and quantity:

- Data quality impacts accuracy and confidence
- Data quantity impacts ability to generalise

Rail OCG1 Categories

What do the weights represent?

- Terms are clustered to best map to OCG1 categories
- As a side effect, synonyms are learnt

Terms with similar weights:

- Collision
  - Struck, collided, hit, damage
- Slip
  - Fall, trip, fell, slipped, collapsed, tripped
- Derail
  - Derailment, derailed

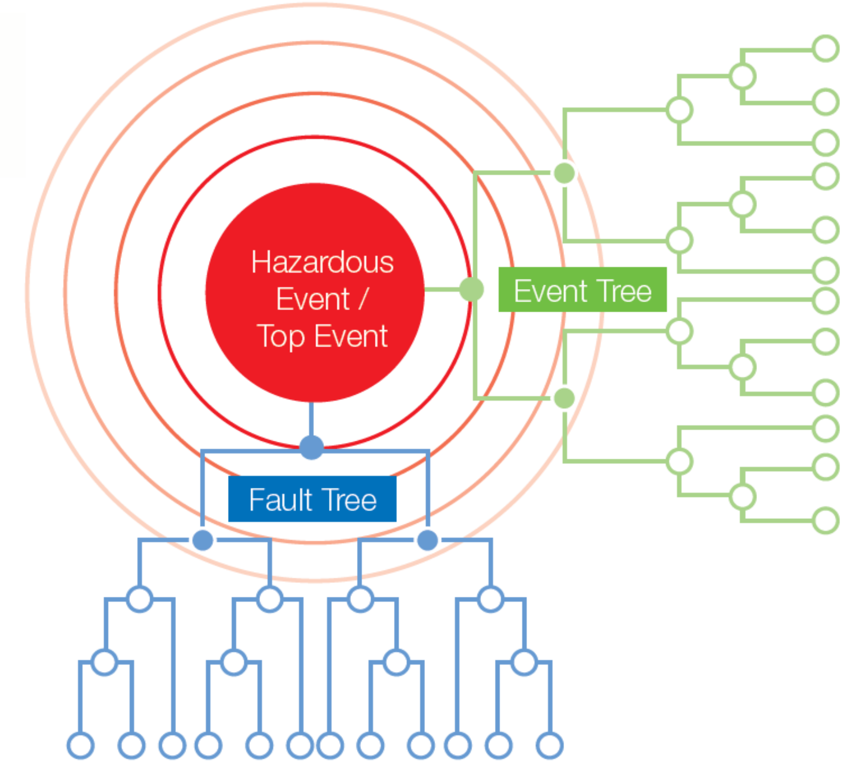
Rail OCG1 Categories

Occurrence description...

Risk Modelling

For the Australian Rail Risk Model:

- Classify rail safety incidents
  - Use events in the risk model
- Large risk model
  - 143 hazardous events
  - ~1,200 basic events
- Used by industry



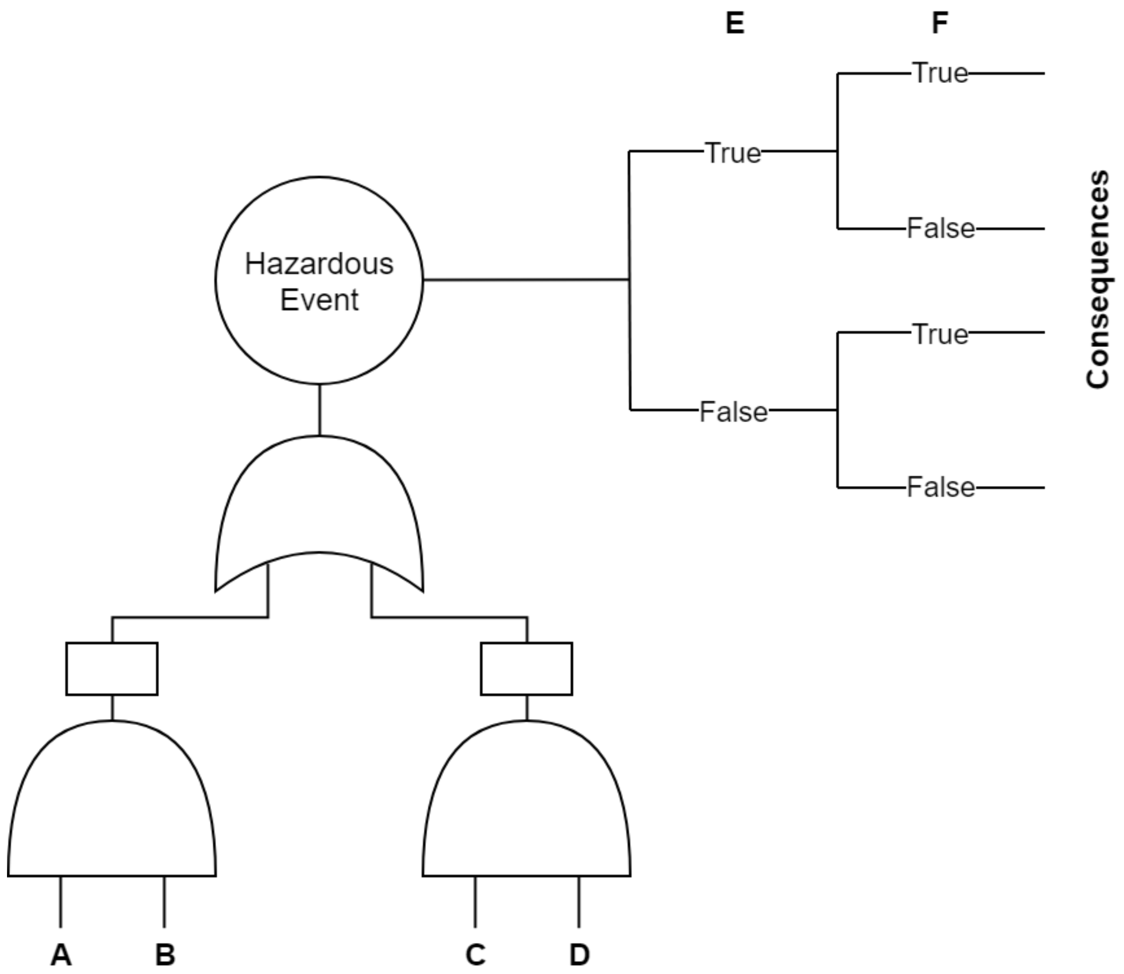
Risk Modelling

Concept of a Modelled Incident:

- Set of basic events
- Does not always realise the hazard
- Does not always have a consequence

For example:

MI_1: { A, B, ¬E, F }
MI_2: { B, C }
MI_3: { A, F }



Risk Modelling

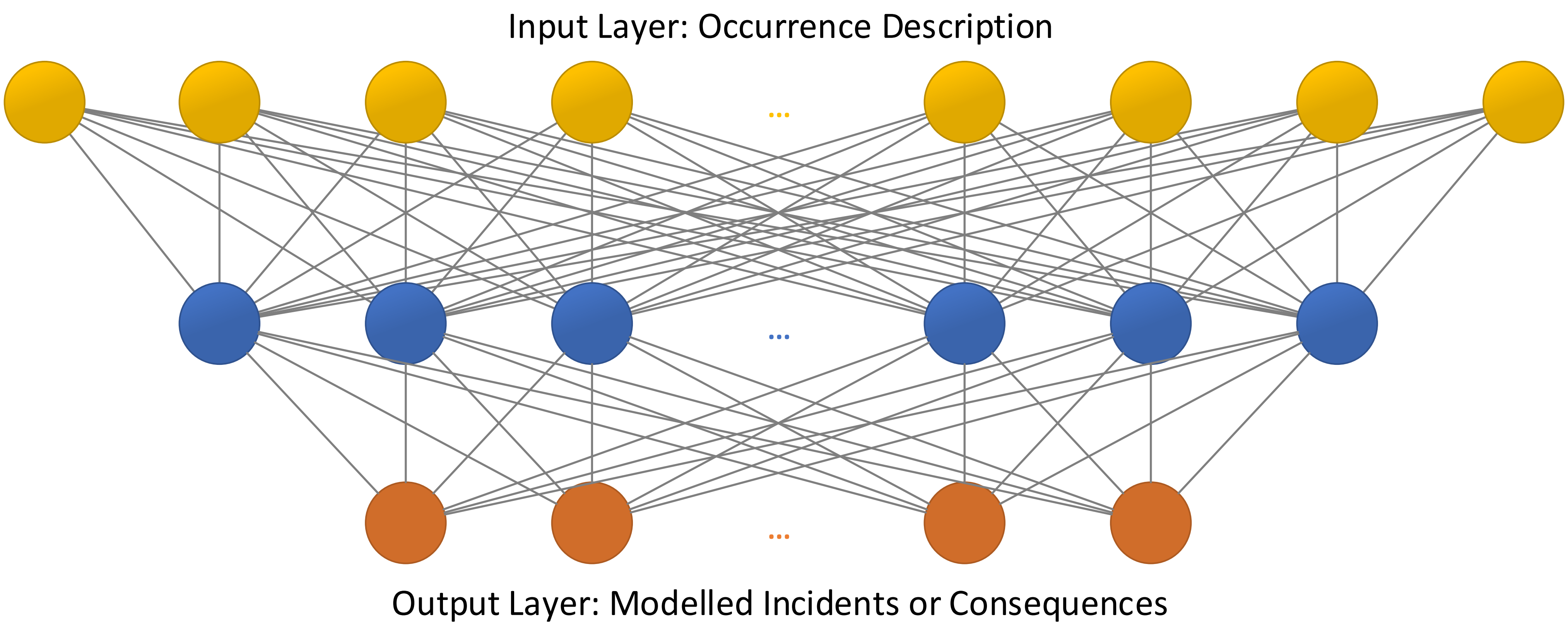
After 2 years of the ARRM, we had:

- A database of ~40,000 occurrences
  - Each mapped to a Modelled Incident
- Mappings to 874 unique Modelled Incidents

Can we, given a textual incident description:

- Predict the Modelled Incident (set of basic events)?
- Predict the consequence?
- Provide confidence of each prediction?

Risk Modelling



Risk Modelling

For Modelled Incidents:

- 74% of predictions were correct
- A random guess would only be correct ~0.1% of the time
- 44% of predictions could be confidently predicted
  - With a confidence of at least 95%

For consequences:

- 98% of predictions were correct
- 98% of predictions could be confidently predicted
  - With a confidence of at least 99%

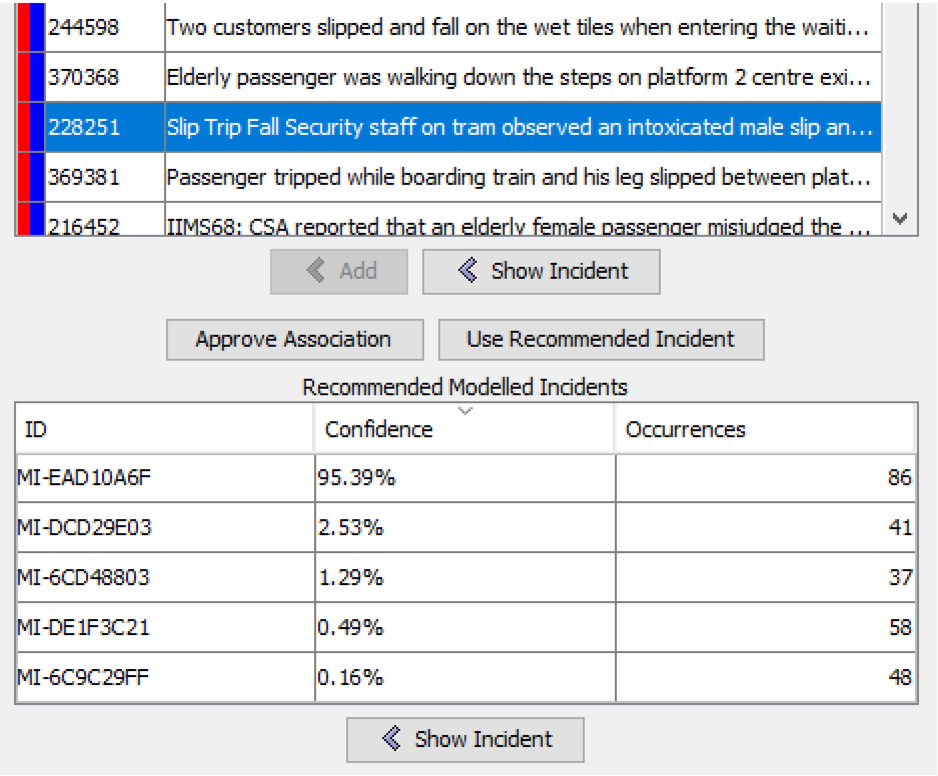
Risk Modelling

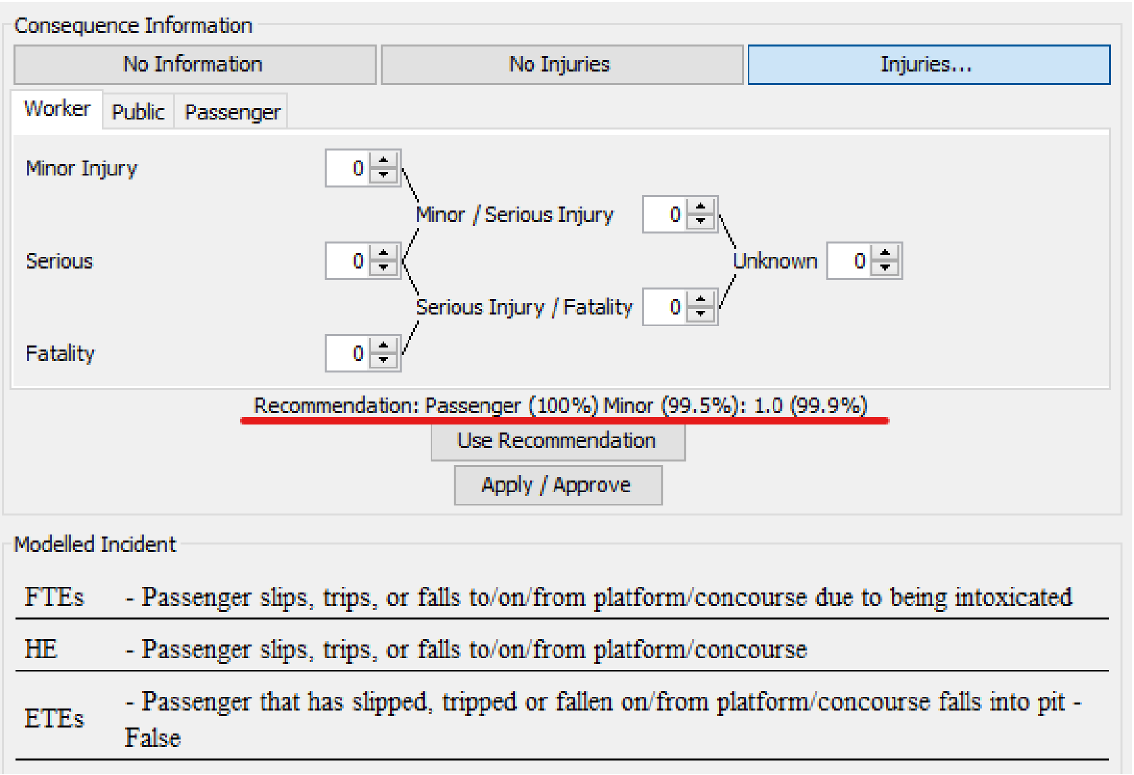
How we use it:

- Evaluated with K-fold cross-validation
- Confirmed by humans
- Not used for rare events
- Used to identify human error

xkcd.com/1838

Risk Modelling





Risk Modelling

- Led to a 40% time saving in 2019

Modelled Incidents:

- 72% used the classifier’s recommendation (top suggestion)
- 94% used one of the suggestions offered by the classifier

Consequences:

- 98% used the classifier’s recommendation

Risk Modelling

Limitation:

- Restricted to modelled incidents that have been observed

For example, “Ben climbed Mount Everest”:

- All Modelled Incidents low confidence
- Highest is MI-38C9045F with 4.5% confidence
  - Miscellaneous event on rail infrastructure
  - Miscellaneous events that have been reported but are not associated with a hazardous event

Summary

- Quantitative risk models are valuable

  But often difficult to quantify

- Safety-related datasets are plentiful

  But often unwieldy

- Machine learning is an effective tool

  Wrangle data to quantify risk models