How Artificial Intelligence (AI) Has Transformed the Insurance Industry

Summer National Meetings // NAIC - CIPR
New York City, NY // 5th August, 2019

https://www.halosinsurance.com/
https://www.datarobot.com/

HOW do we build AI
WHY do we trust AI

The Speaker

Satadru Sengupta

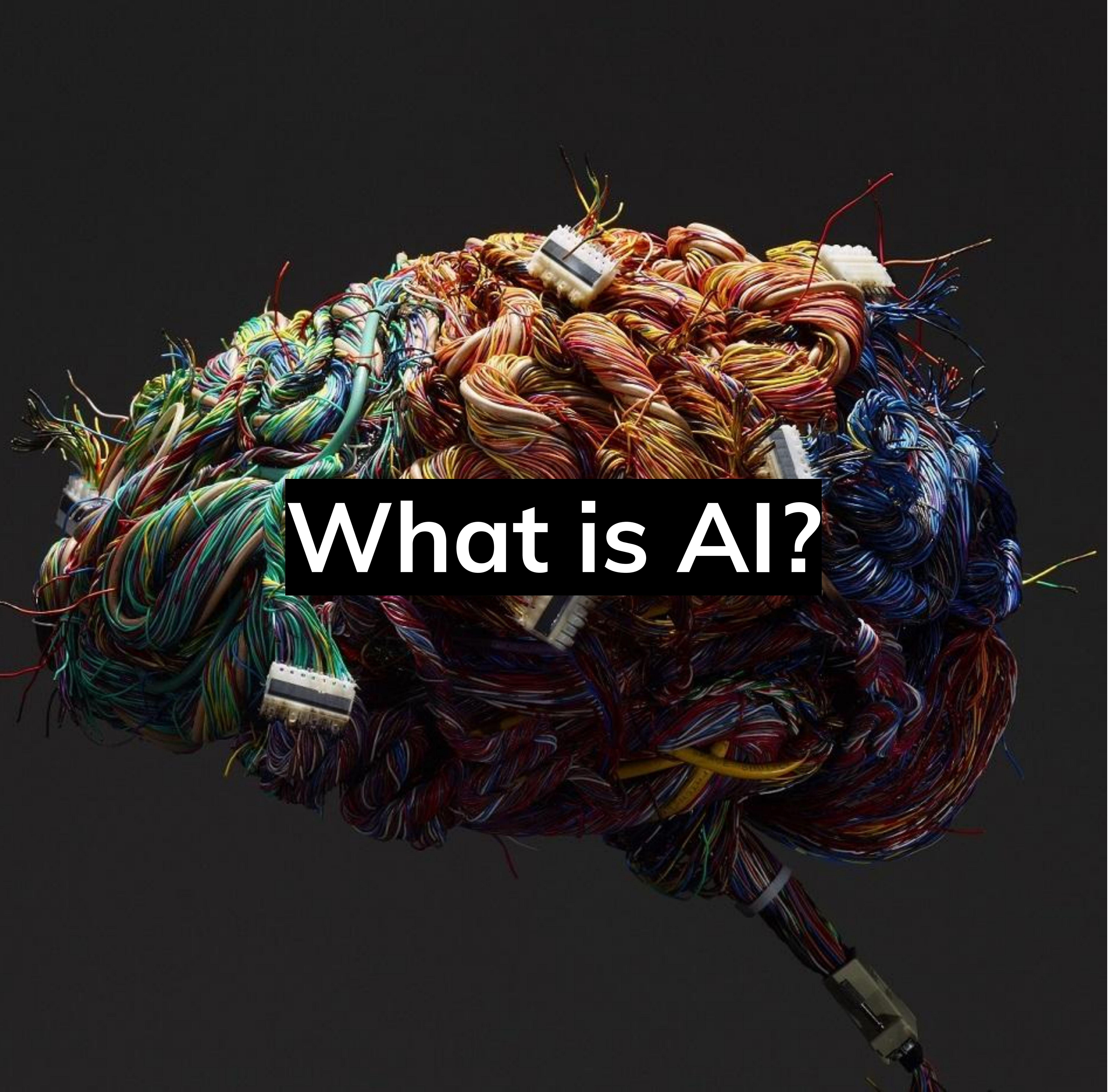
Founder & CEO at Halos

- built multi-million dollar insurance business from pre-revenue at AI pioneer DataRobot: 2015-2018
- built & operationalized multiple AI applications at AIG
- formerly, actuarial data scientist at Liberty Mutual & Deloitte
- CSPA designee Casualty Actuarial Society

HALOS DataRobot AIG Liberty Mutual INSURANCE Deloitte.

HOW do we build an AI application

let's use an example to understand the key concepts and workflow



computer
systems
able to do tasks
that require
human
intelligence

EXAMPLE

Predicting claims fraud in auto insurance

1

Problem statement

predict the "likelihood of fraud" of an incoming claim based on policy data and claims data at FNOL

2

Scope of use:

to be used to triage claims and help SIU in targeted investigation

3

Key Objectives:

- accurate predictions
- prediction explanations

QUESTION # 1

WHY do we need a computer to do a human task

- processing large, unstructured data
- an objective way of making decision
- fast and automated

DATASET

NUMERIC

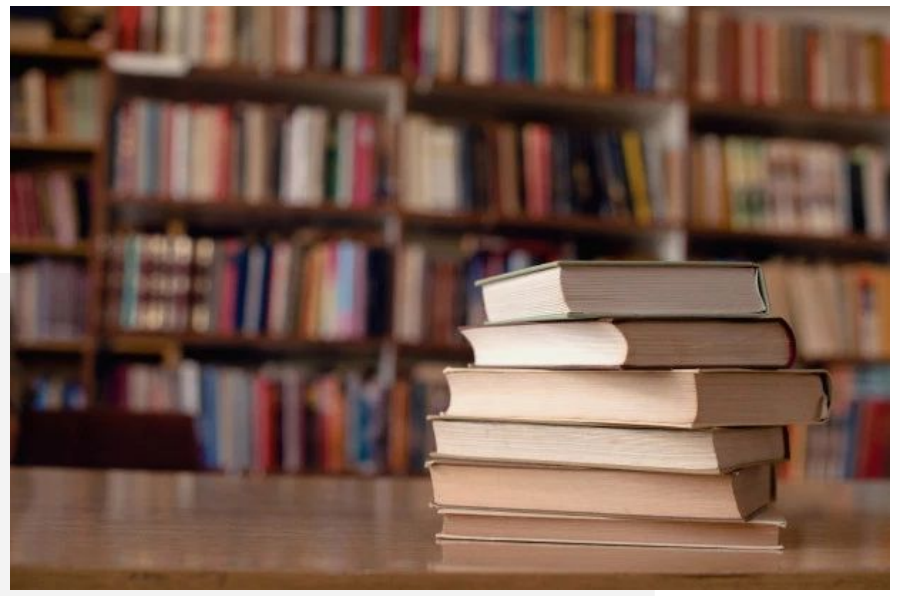
CATEGORICAL

TEXT

ID	FRAUD	DISTINCT_PARTIES_ON_CLAIM	CLM_AFTER_RNWL	CLAIM_DESCRIPTION
1	0	4	0	this via others themselves inc become within ours slow parking lot f
2	0	4	0	would less bottom de what then find cry motorbike brakes van sudd
3	1	21	0	indeed none you to somehow call whereas anyhow driving left scho
4	1	5	0	am not fire same now over whence therein right left not indicating c
5	0	2	0	formerly by fifteen again are please four bottom caravan motorbike
6	0	1	1	nor put see not seems serious is herself motorbike caravan parking
7	0	1	0	not others into who its these else during car sun right school driving
8	1	2	0	describe except yourself what whom every because within slow ma
9	0	1	0	more being third us part but found neither not indicating windscreen
10	0	5	0	would couldnt etc or wherever her may this carpark van sun parking
11	0	3	0	have co further three cant found whereafter nevertheless mall round
12	0	2	1	cant still front among whom wherein serious part not indicating rour

QUESTION # 2

HOW do we teach the computer?

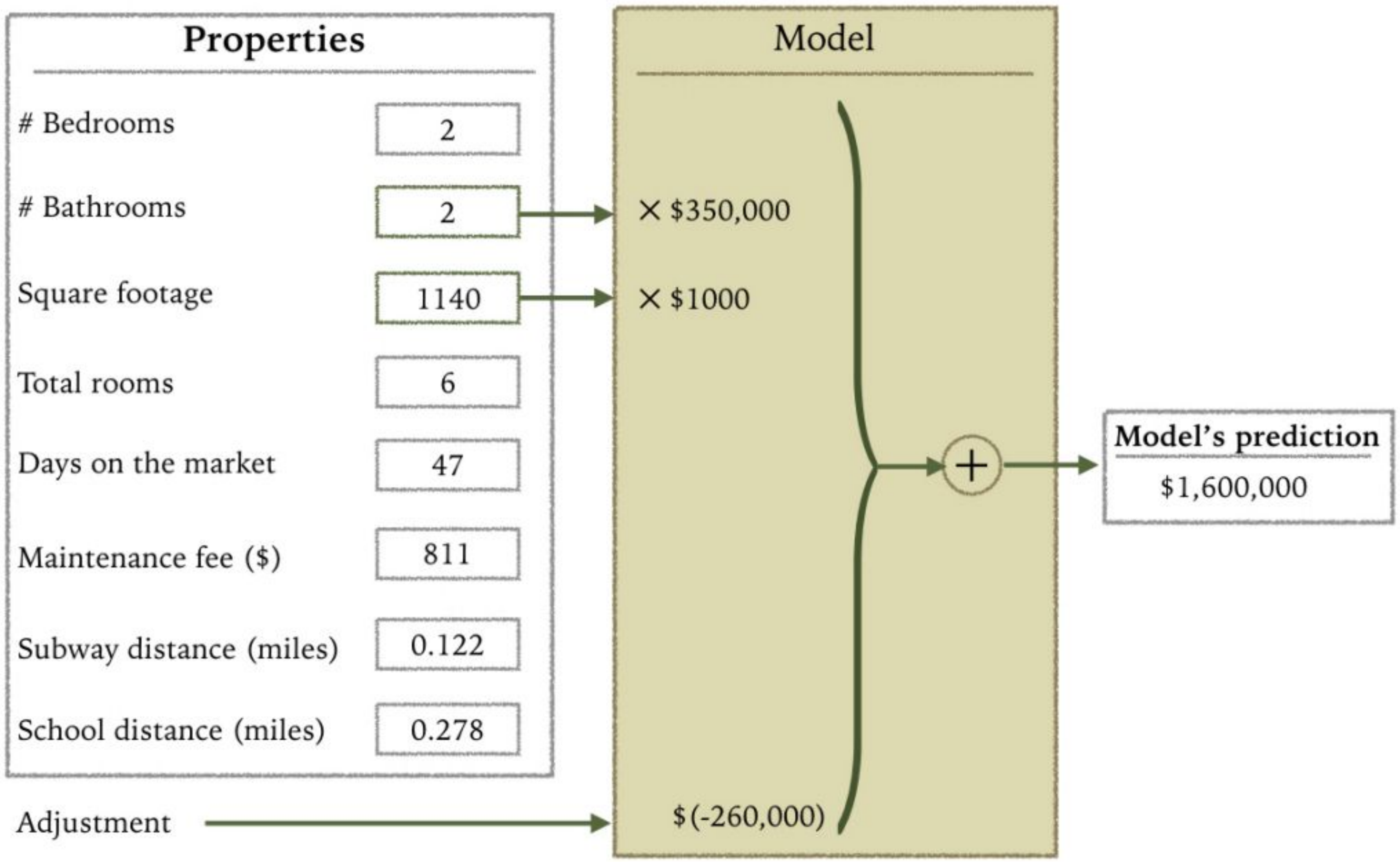


1 A historical dataset

2 Machine learning algorithms

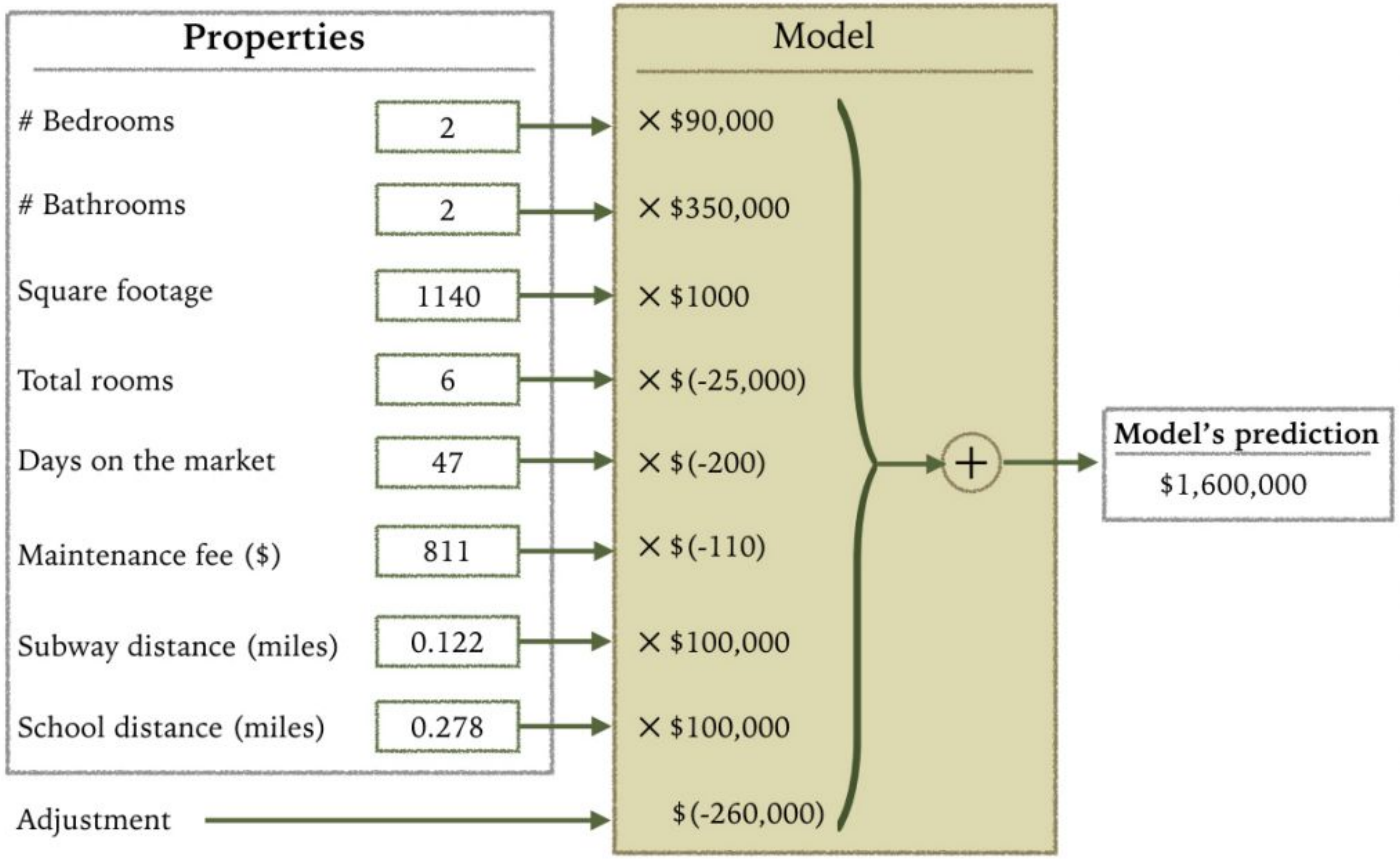
MACHINE LEARNING

The ability to learn from the past to predict the future without being explicitly programmed



(a) Clear, two-feature condition (CLEAR-2).

Same dataset but a different algorithm, more complex this time



(c) Clear, eight-feature condition (CLEAR-8).

Algorithms matter

If the model is inaccurate, there could be terrible outcome: bad customer experience to insolvencies

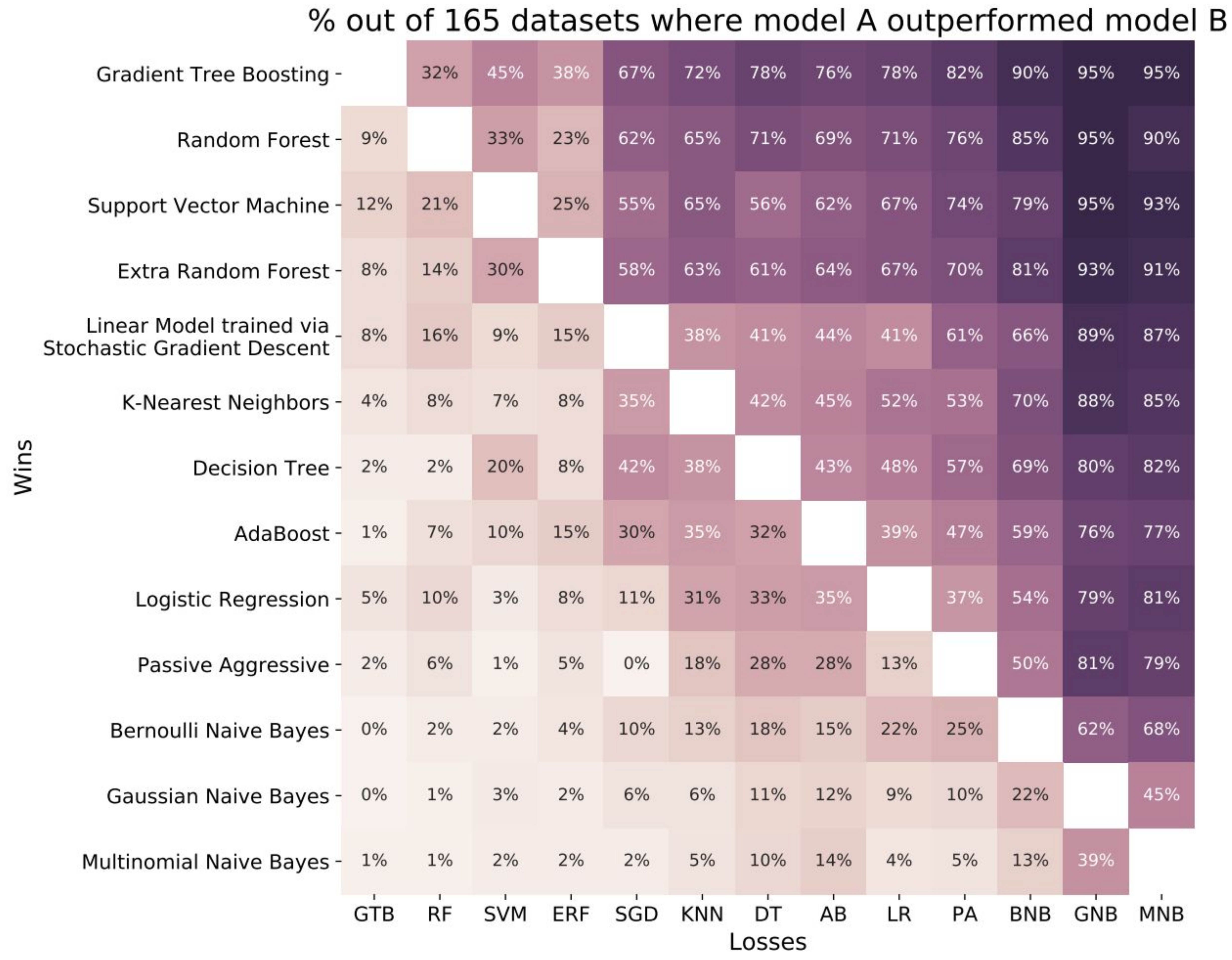


Fig. 2. Heat map showing the percentage out of 165 datasets a given algorithm outperforms another algorithm in terms of best accuracy on a problem. The algorithms are ordered from top to bottom based on their overall performance on all problems. Two algorithms are considered to have the same performance on a problem if they achieved an accuracy within 1% of each other.

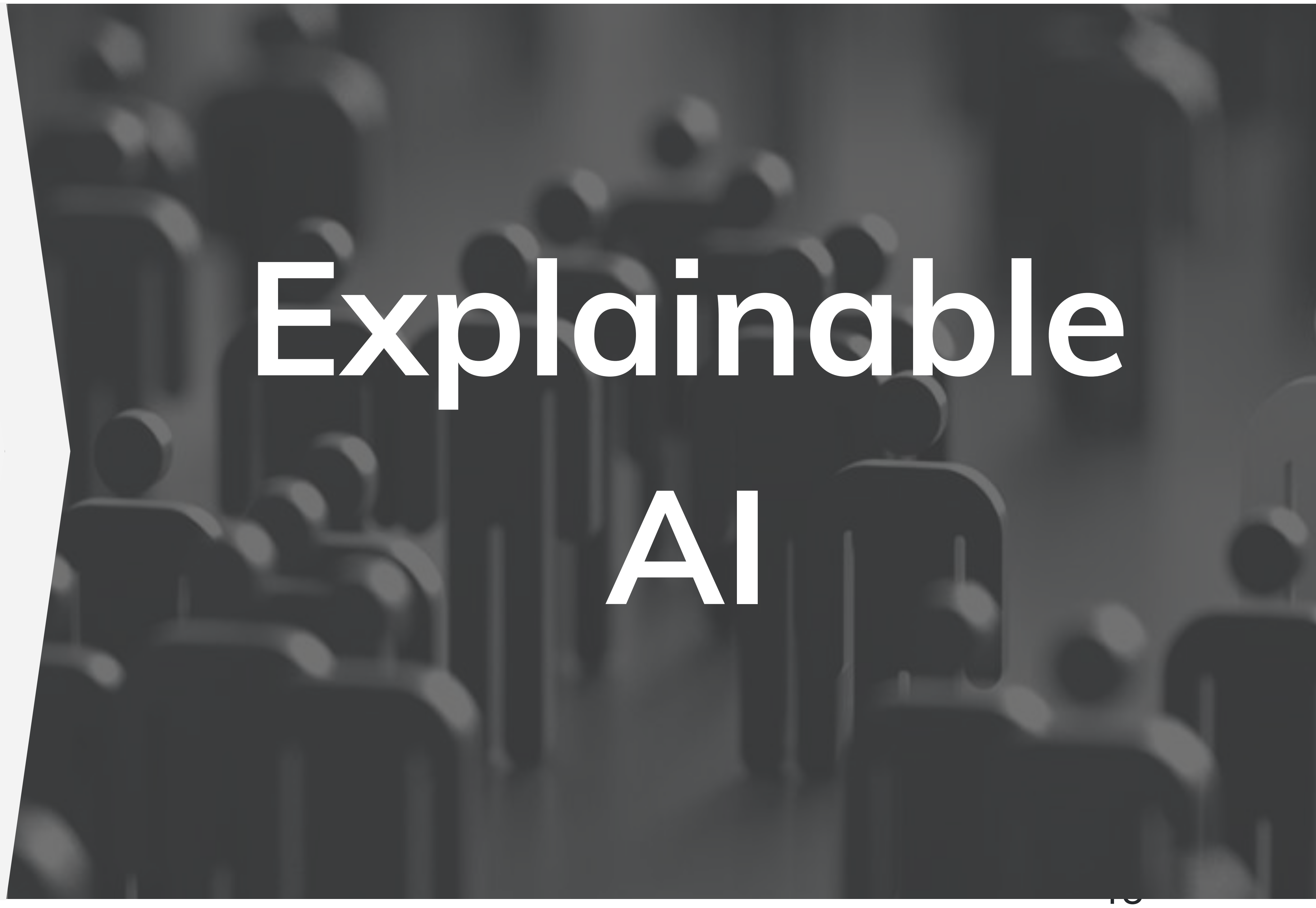
FOCUS AREA

GREAT NEWS!!!

We have explanation tools.

These tools are algorithm agnostic.

- Most impactful features
- Directionality of the feature
- Explain every prediction



Feature importance:
which predictors drive the model performance

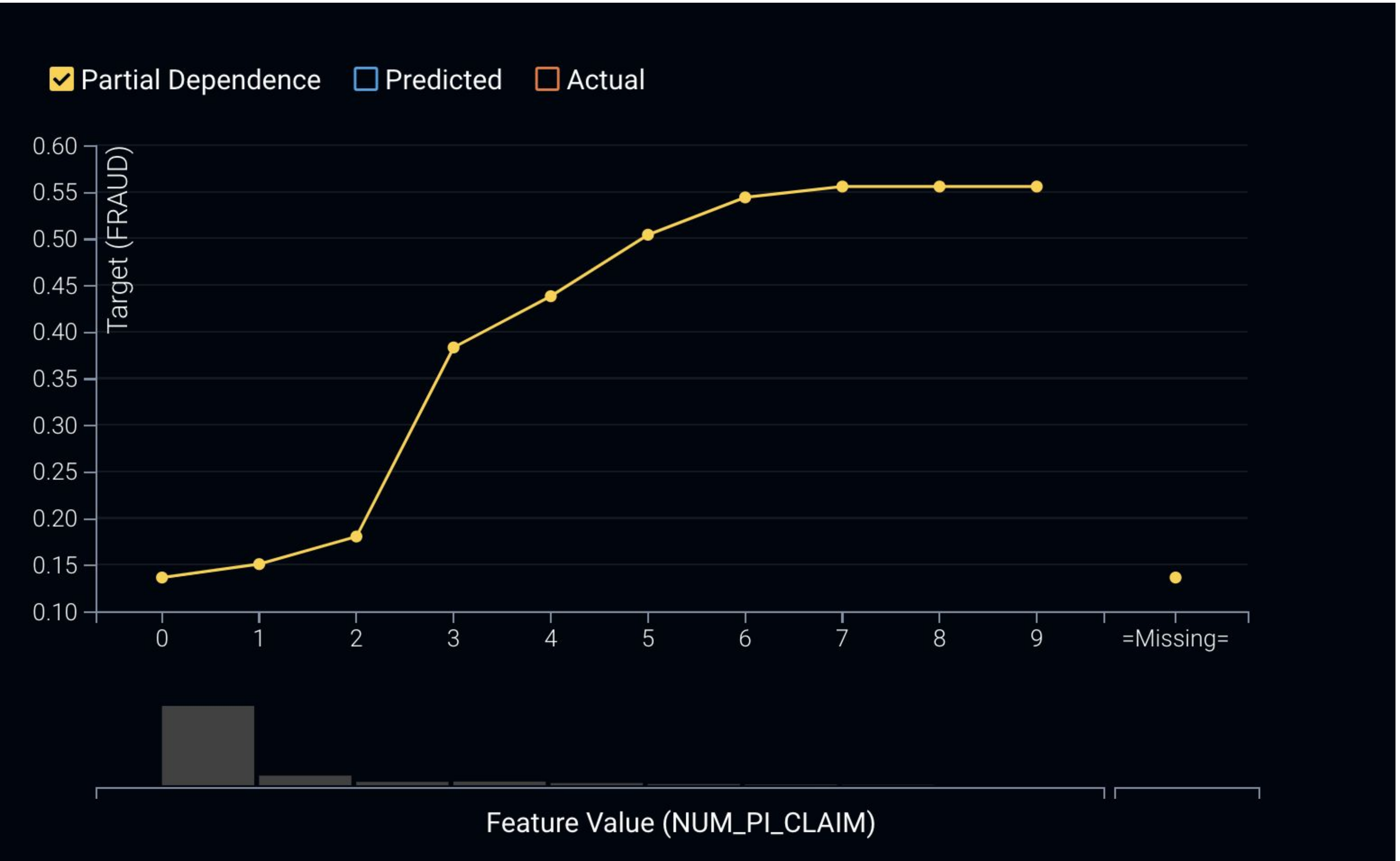
A few top predictors:

prior claims,

claim type,

people involved

Directionality of the feature:
How a predictor influences overall outcome

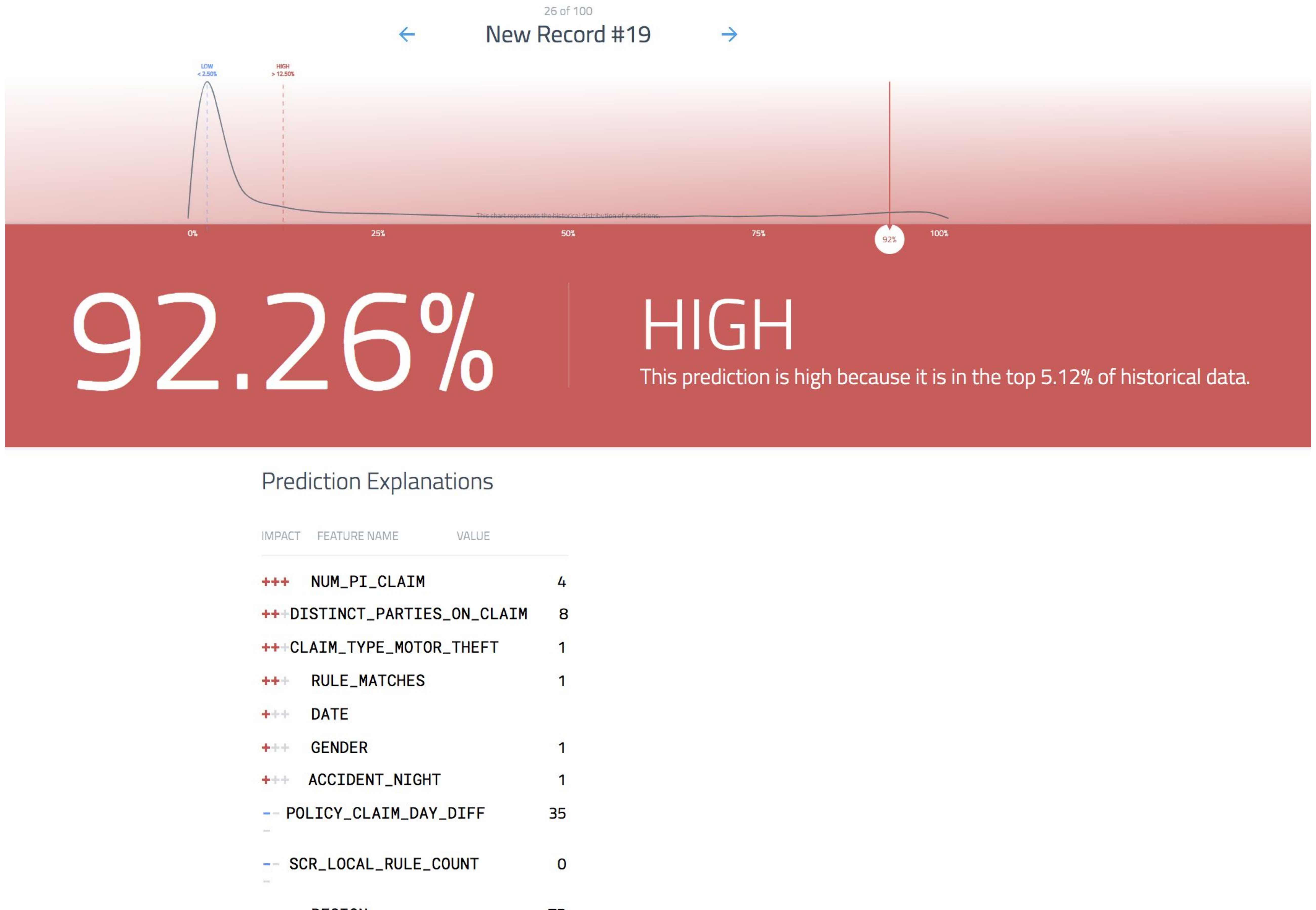


 # prior claims:

3 or more prior claims ⇒ higher chances of fraud

Prediction explanation (AI storytelling):

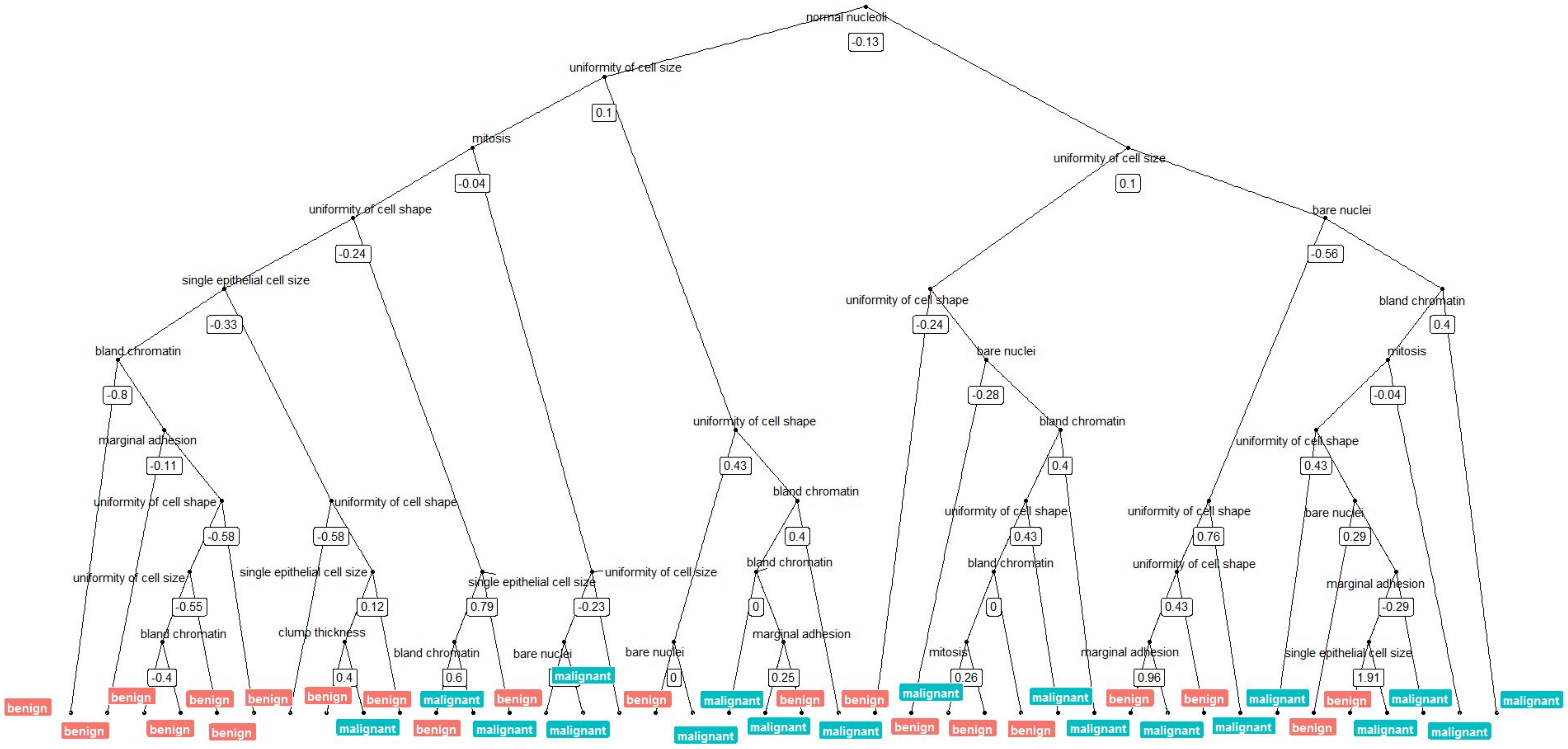
Going forward, once implemented, we can tell what are the factors behind a prediction



A male claimant reported a theft involving 8 people within a month of buying the policy. He had 4 claims in the last 5 years.

We generated these explanation from a fairly complex model:

XGBoost: a very complex and powerful algorithm



Few things that you need to build an AI application

1

Team:

- data engineers
- data scientists
- domain experts
- users (in this case, fraud/ SIU analyst)

2

A dataset:

We are using a dataset with 10,000 claims from past with fraud indicator and 40 possible predictors (aka: features)

3

An AI platform:

we are using DataRobot Automated Machine Learning platform

DATASET

What happened in the past

Features or variables accompanied the historical outcome

Many historical examples...

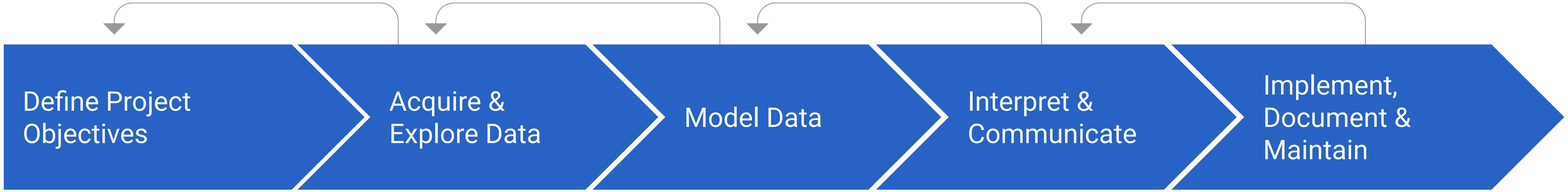
ID	FRAUD	DISTINCT_PARTIES_ON_CLAIM	CLM_AFTER_RNWL	CLAIM_DESCRIPTION
1	0	4	0	this via others themselves inc become within ours slow parking lot fast vehicle
2	0	4	0	would less bottom de what then find cry motorbike brakes van suddenly not inc
3	1	21	0	indeed none you to somehow call whereas anyhow driving left school motorbik
4	1	5	0	am not fire same now over whence therein right left not indicating car carpark s
5	0	2	0	formerly by fifteen again are please four bottom caravan motorbike not indicatin
6	0	1	1	nor put see not seems serious is herself motorbike caravan parking lot car righ
7	0	1	0	not others into who its these else during car sun right school driving not indicat
8	1	2	0	describe except yourself what whom every because within slow mall vehicle wi
9	0	1	0	more being third us part but found neither not indicating windscreen vehicle bra
10	0	5	0	would couldnt etc or wherever her may this carpark van sun parking lot slow le
11	0	3	0	have co further three cant found whereafter nevertheless mall roundabout stop
12	0	2	1	cant still front among whom wherein serious part not indicating roundabout car
13	0	5	0	formerly rather it but might former neither done mall roundabout brakes fast inc
14	0	4	0	become no being throughout someone twelve part whole motorbike slow round
15	0	2	0	without among each none system who many well vehicle right slow left school
16	0	2	0	two its was already in this somehow fify school carpark parking lot indicating ro
17	0	0	1	though too full no take together a seem parking lot vehicle caravan windscreen
18	0	1	0	due describe hundred therefore became bottom others so vehicle fast brakes c
19	1	12	0	under whence co only therefore eg no around sun parking lot motorbike school
20	0	6	0	six whereupon please nothing interest noone often several slow stopped carav

DataRobot

Data Science Iron Man

The AI Life Cycle

The AI Life Cycle

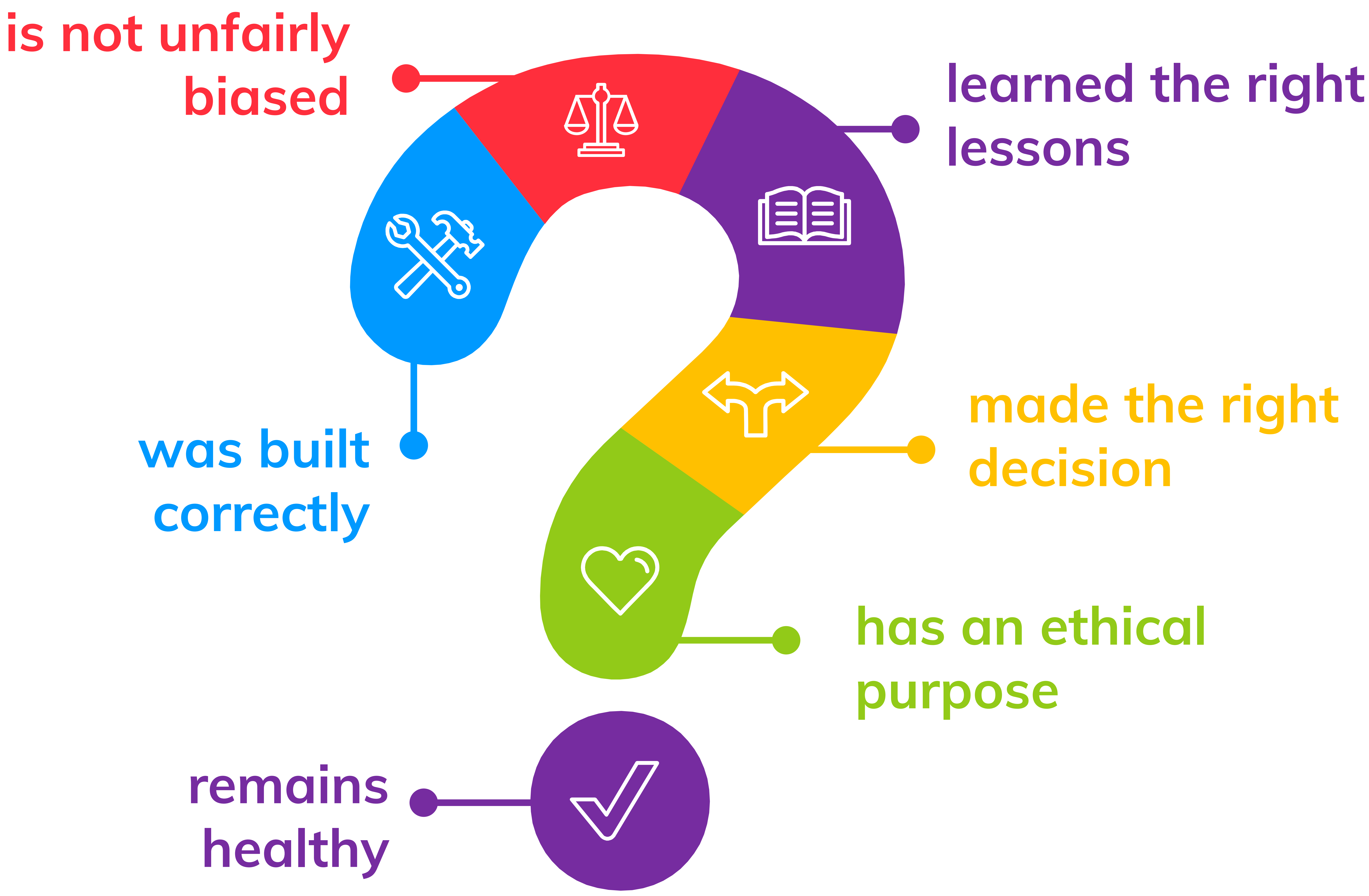


1. **Define Project Objectives**
   - [ ] Specify business problem
   - [ ] Acquire subject matter expertise
   - [ ] Define unit of analysis and prediction target
   - [ ] Prioritize modeling criteria
   - [ ] Consider risks and success criteria
   - [ ] Decide whether to continue
2. **Acquire & Explore Data**
   - [ ] Find appropriate data
   - [ ] Merge data into single table
   - [ ] Conduct exploratory data analysis
   - [ ] Find and remove any target leakage
   - [ ] Feature engineering
3. **Model Data**
   - [ ] Variable selection
   - [ ] Build candidate models
   - [ ] Model validation and selection
4. **Interpret & Communicate**
   - [ ] Interpret model
   - [ ] Communicate model insights
5. **Implement, Document & Maintain**
   - [ ] Set up batch or API prediction system
   - [ ] Document modeling process for reproducibility
   - [ ] Create model monitoring and maintenance plan

Why Should We TRUST Artificial Intelligence (AI)

“Sometimes attaining the deepest familiarity with a question is our best substitute for actually having the answer.”

BRIAN GREENE
THEORETICAL PHYSICIST & MATHEMATICIAN, COLUMBIA UNIVERSITY


is not unfairly biased
learned the right lessons
was built correctly
made the right decision
has an ethical purpose
remains healthy


"Man is fallible, but maybe men are less so."
Atul Gawande

THANK YOU

Satadru Sengupta

email: satadru@halosinsurance.com

text/ call: +1.617.301.2471

https://www.datarobot.com/

https://www.halosinsurance.com/