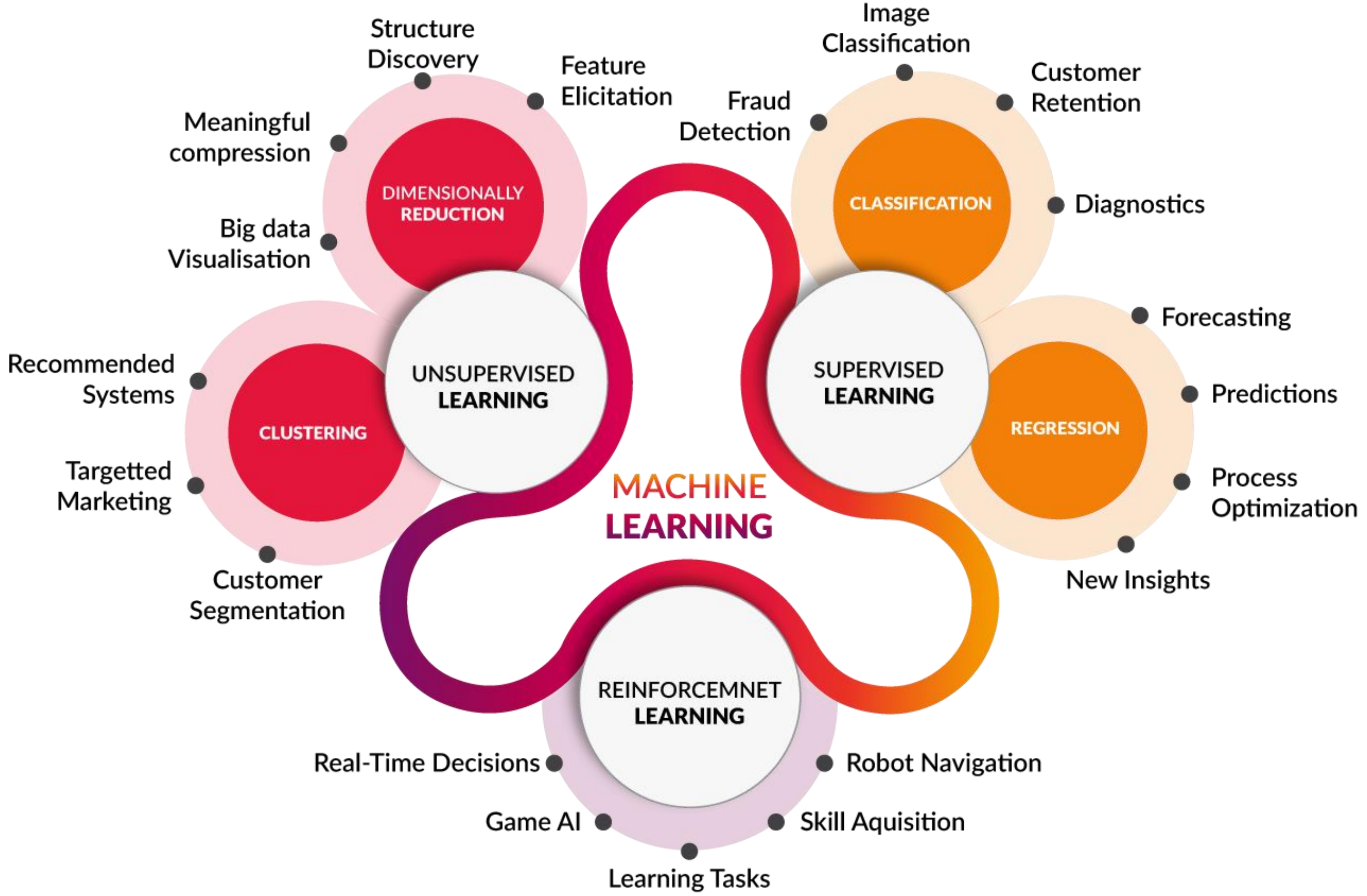
Application of ML to the Estimation of Intact Rock Strength from Core Logging Data: A Case Study at the Newcrest Cadia East Mine

Chris Thielsen, Jason Furtney, Maria Elena Valencia,
Matt Pierce,
Cristian Orrego, Peter Stonestreet, David Tennant

Outline

- Background
- Project Motivation
- Data & Preprocessing
- Machine Learning Model
- Analysis
- Conclusions

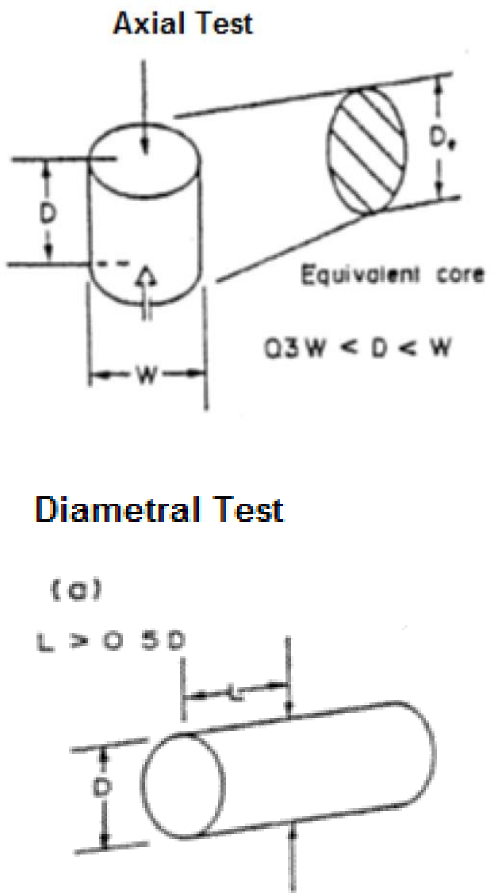


Background

- Located near Orange, New South Wales, Australia
- Operated by Newcrest Mining
- An underground mine, Cadia East is part of a series of gold and copper mines located in the Cadia Valley
- Cadia is home to Australia's largest underground gold mine and one of the largest gold and copper deposits in the world
- Feasibility study began in 2020, to expand panel caving operations at the mine

Project Motivation

- Strength information required to complete a feasibility study at the mine
- Large amount of data logged from ~1300 drillholes (~590,000 m)
  - Mineralogy, Alteration, Lithology, Geotechnical, etc.
- Determining point load strength at every meter along each core is desired, to be used for the construction of geospatial rock strength models
- Although simple and relatively cheap, point load testing would be too expensive for this number of points
- Solution: conduct point load testing on a smaller amount of core and use machine learning to predict the remaining point load strength (I_{s50})



Data & Preprocessing

- Many core logs record irregular segments along the drillhole (e.g., 4m – 7m)
- The data was homogenized such that each sample represented a 1m section of core

Mineralogy

0.00 – 3.27, …
3.27 – 5.84, …
4.39 – 8.79, …

Alteration

0.00 – 2.85, …
5.67 – 6.15, …
1.52 – 14.20, …

Lithology

0.00 – 7.05, …
7.05 – 13.48, …
13.48 – 20.09, …

Geotech.

0.00 – 1.15, …
1.15 – 3.98, …
3.98 – 6.44, …

→

Min, Alt, Lith, Geotech, etc.

0.00 – 1.00, …
1.00 – 2.00, …
2.00 – 3.00, …
3.00 – 4.00, …
4.00 – 5.00, …
5.00 – 6.00, …
6.00 – 7.00, …
7.00 – 8.00, …
…

Data & Preprocessing

- 14 raw features
  - RQD
  - Fracture frequency
  - Mineralized veins per meter
  - Rock density
  - ISRM rock strength
  - Color
  - Lower contact
  - Rock texture
  - Rock type
  - Selvage mineralization
  - Minerals present
  - Rock structure
  - Rock fabric
  - PLT machine

- 8 engineered features
  - Fracture spacing
  - Mineralization percent sum
  - Mineralization strength index (MSI)
  - Alteration strength index (ASI)
  - Number of joint sets (JSI)
  - Fracture feature index (FFI)
  - Discontinuity Frequency index (DFI)
  - Weighted DFI (DFIw)

Data & Preprocessing

- For the model to predict point load strengths, some training data is necessary
  - Systematic point load testing was performed on 19 drillholes, resulting in ~7700 total training samples
- Regression was attempted, but yielded poor results because of noise in the dataset
  - I_{s50} (MPa) values were converted to 5 classes, using the Jenks natural breaks algorithm

Is50 Class	Class Bounds [MPa]	Class Mean [MPa]	Training Count
1	0.00 – 1.29	0.21	3,344
2	1.29 – 3.65	2.38	1,796
3	3.65 – 6.32	4.93	1,382
4	6.32 – 9.59	7.70	998
5	> 9.59	11.51	194
		Total:	7,687

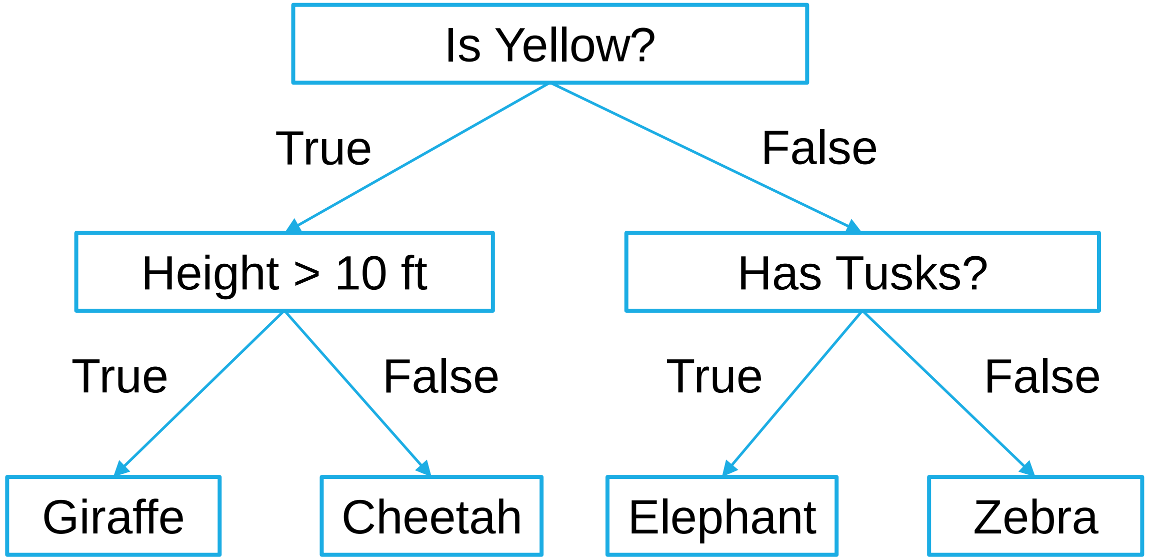
NEWCREST MINING LIMITED

PIERCE ENGINEERING THE SUPPORTING MECHANISM

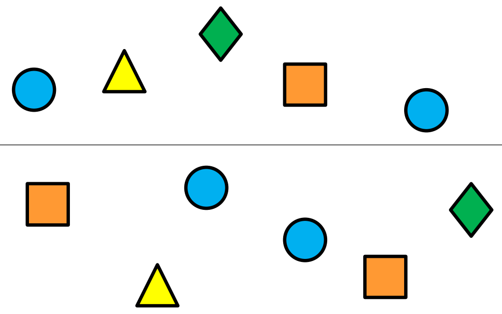
ITASCA™

Decision Tree Model

- Simple, flowchart like model
- Tree is created from training data
- Splits are determined using a purity metric
- Terminal nodes contain samples belonging to the same class



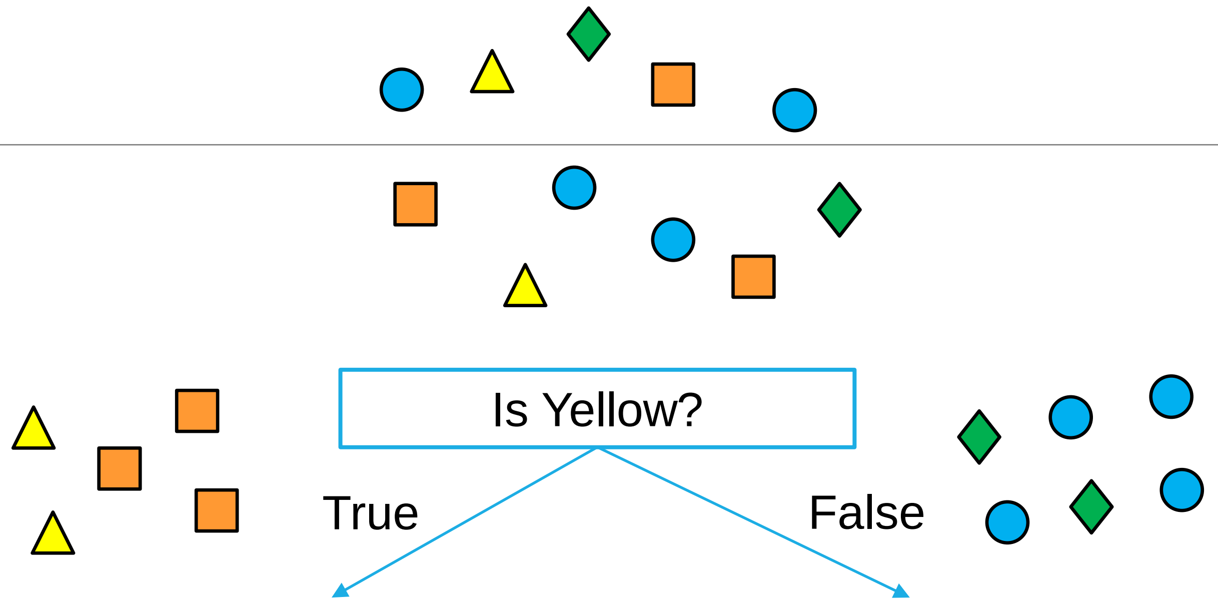
Decision Tree Model



- Labels/Targets:
  - Giraffe
  - Cheetah
  - Elephant
  - Zebra
- Features:
  - Is Yellow? (Boolean)
  - Height (Numerical)
  - Has Tusks? (Boolean)

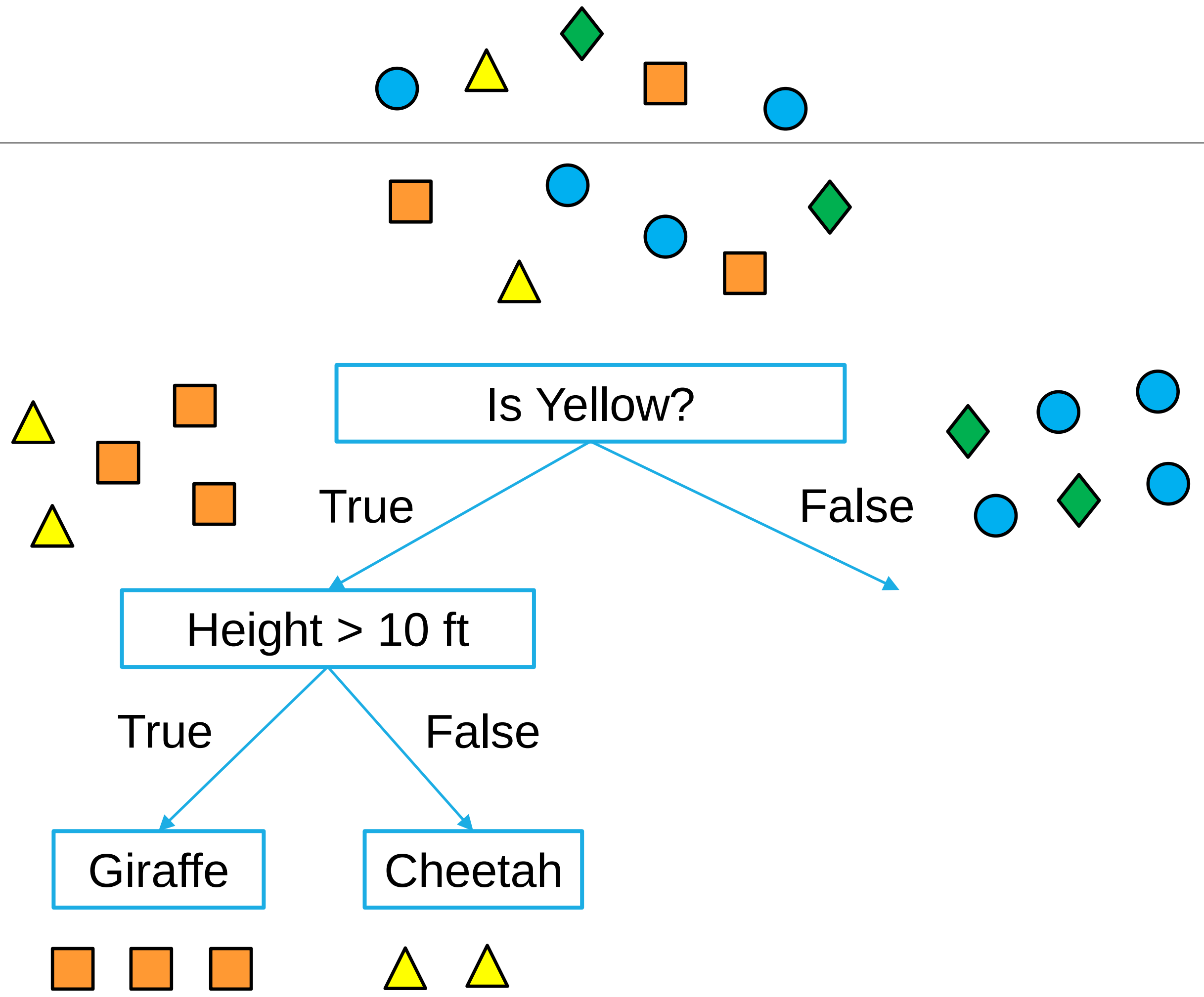
Decision Tree Model

- Labels/Targets:
  - Giraffe
  - Cheetah
  - Elephant
  - Zebra
- Features:
  - Is Yellow? (Boolean)
  - Height (Numerical)
  - Has Tusks? (Boolean)



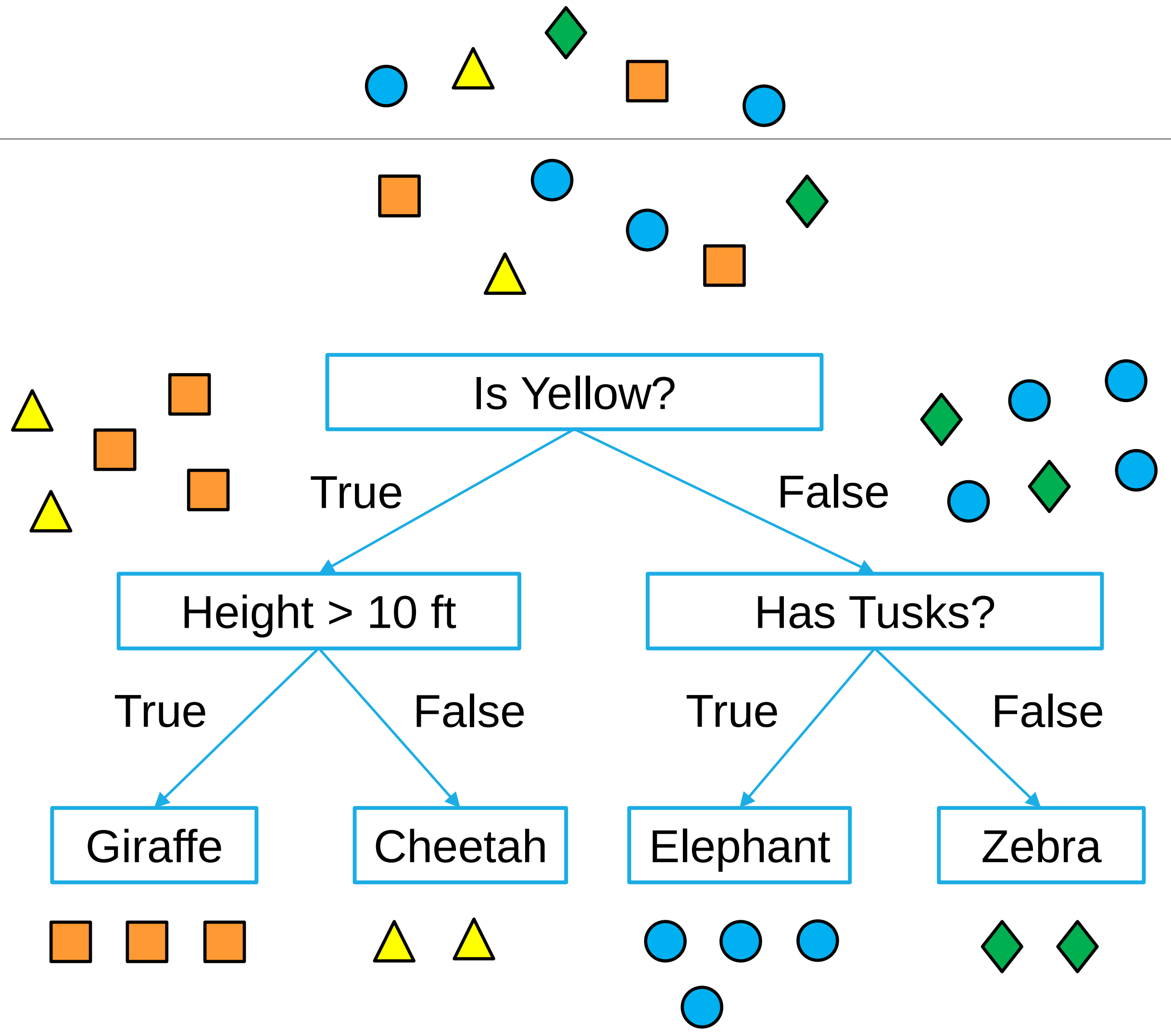
Decision Tree Model

- Labels/Targets:
  - Giraffe
  - Cheetah
  - Elephant
  - Zebra
- Features:
  - Is Yellow? (Boolean)
  - Height (Numerical)
  - Has Tusks? (Boolean)



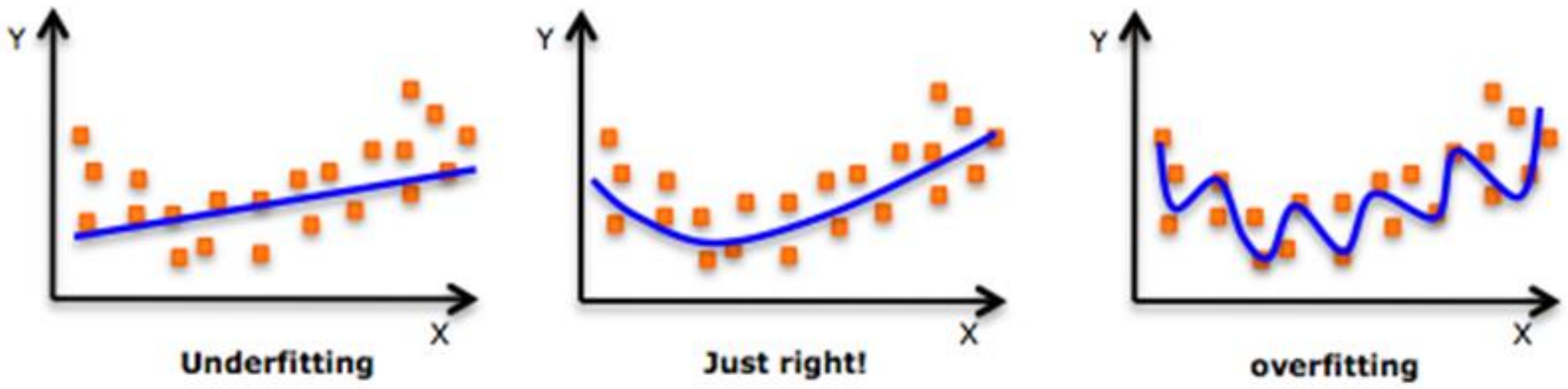
Decision Tree Model

- Labels/Targets:
  - Giraffe
  - Cheetah
  - Elephant
  - Zebra
- Features:
  - Is Yellow? (Boolean)
  - Height (Numerical)
  - Has Tusks? (Boolean)



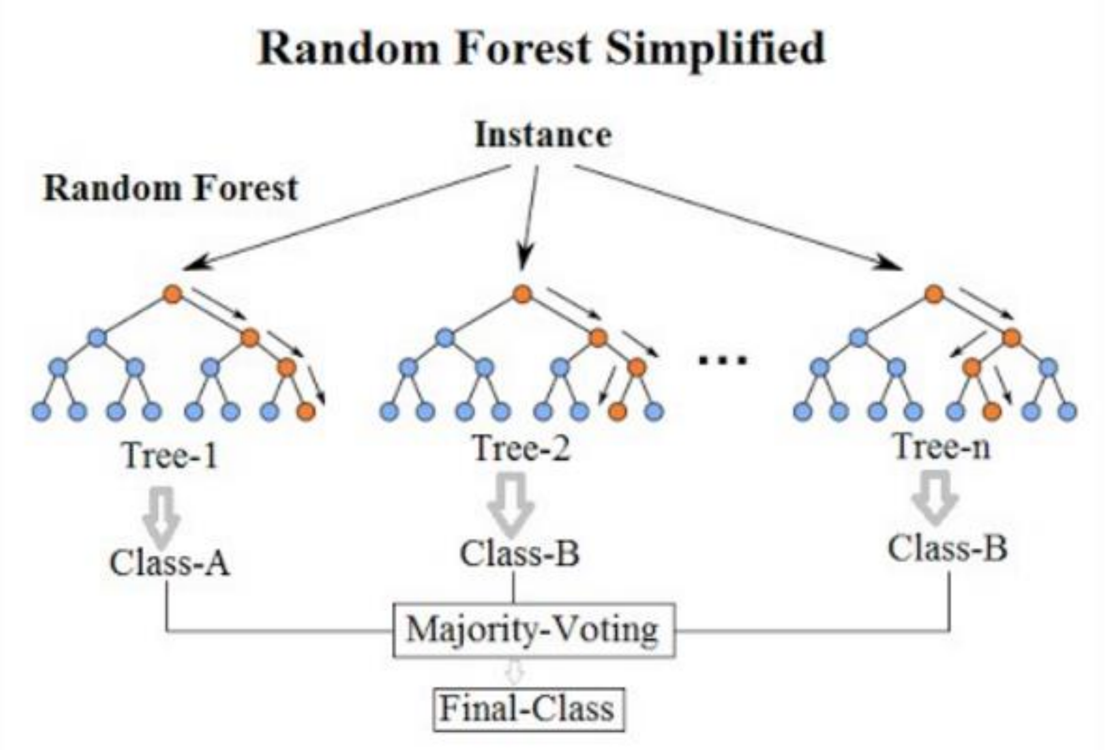
Decision Tree Model

- Decision trees can quickly get large and complex
  - Tree depth of 40 → $2^{40+1} - 1 \cong 2.20 * 10^{12}$ nodes
  - While they are useful, decision trees tend to overfit to your training set



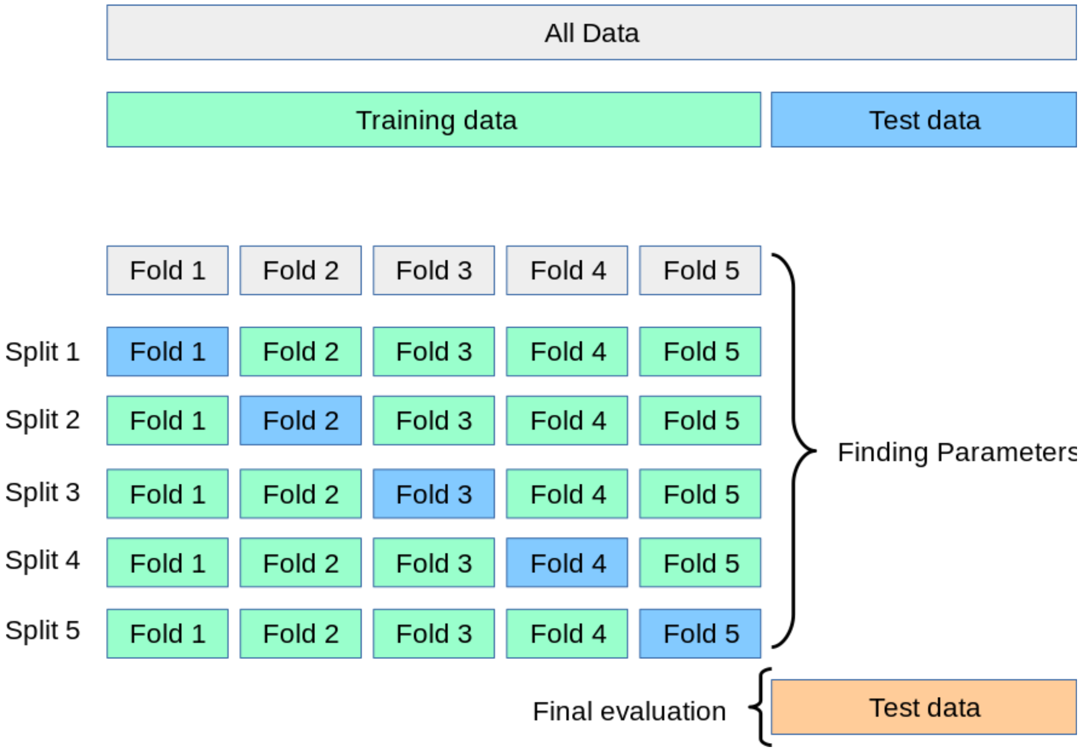
Random Forest Model

- Trees take random subset of training data as input → bootstrap aggregating = “bagging”
- Random subset of features considered at each split
- Prediction by majority voting of individual trees
- Generalizes better by correcting overfitting
- Class bias is corrected by adjusting class weights inversely proportional to class frequencies



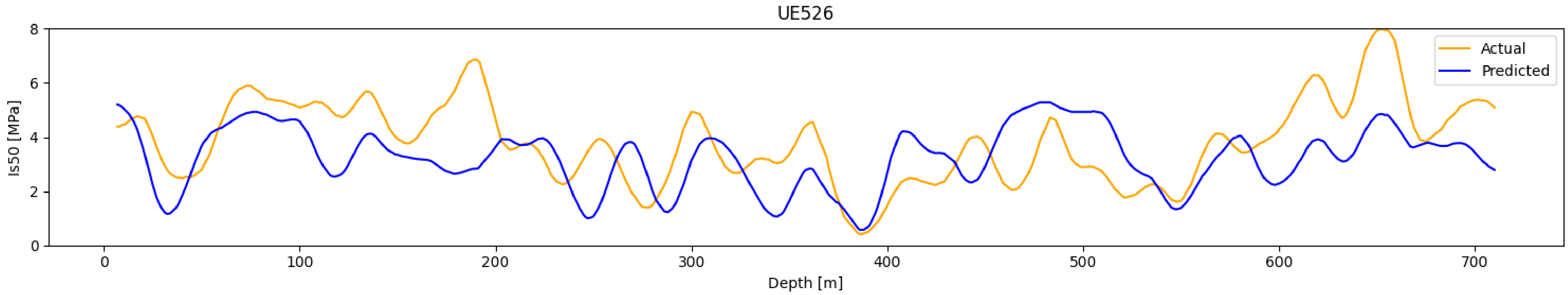
Machine Learning Evaluation

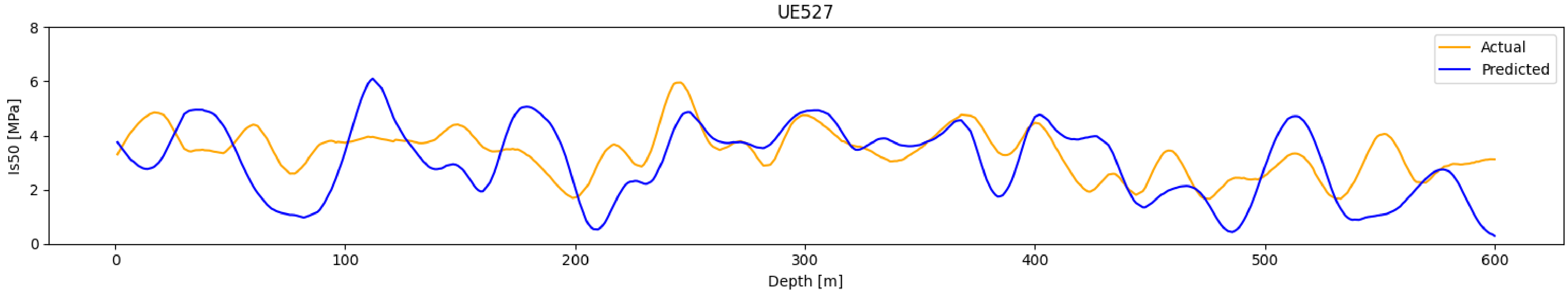
- To tune the model, n-fold validation was performed, and prediction plots were made for each drill hole
- This helps prevent overfitting the parameters to a specific drill hole
- The data was visualized by calculating the rolling mean on the predicted and actual class labels, converted to the mean Is50 value for each class
- ‘Rolling Mean Score’ defined as the percentage of the rolling mean curves within 1 MPa of each other



Random Guess Accuracy	Correct Class Accuracy	Accuracy within 1 Class	Accuracy within 1 MPa
20%	39%	81%	48%

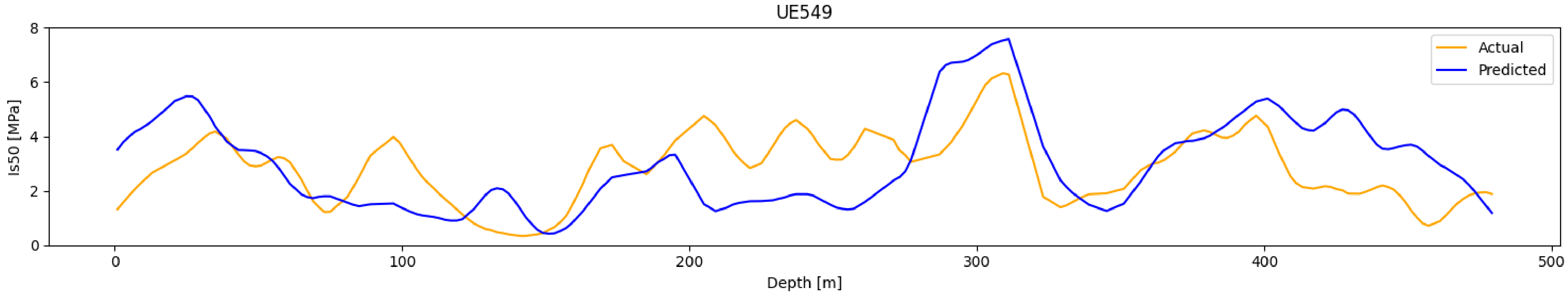
Evaluation Plot Examples

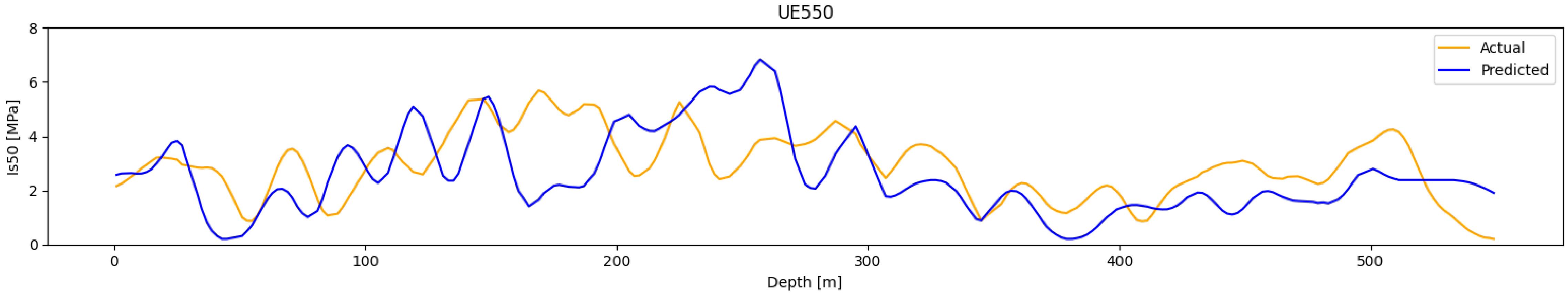




Average Rolling Mean Score: 48%

Evaluation Plot Examples

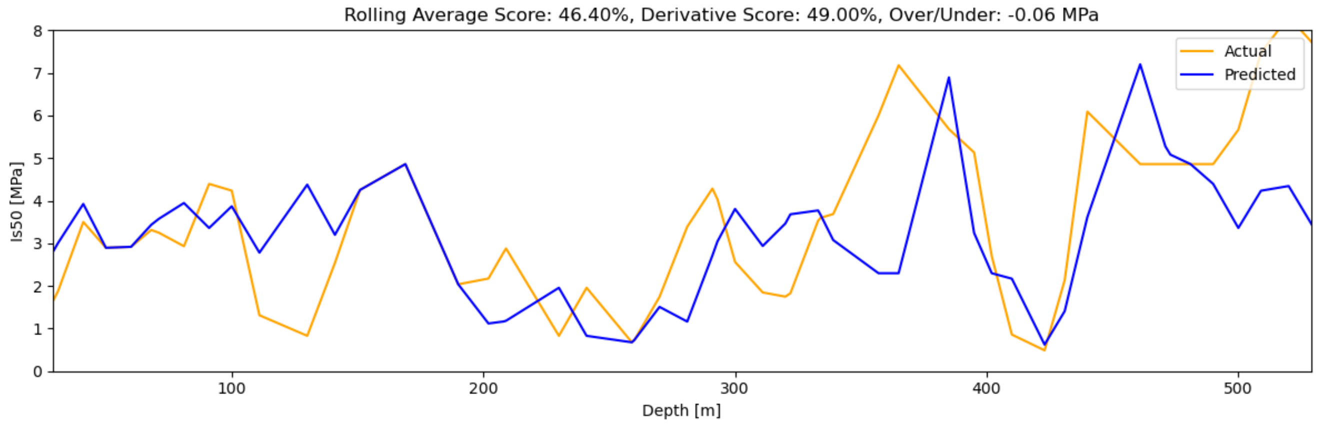


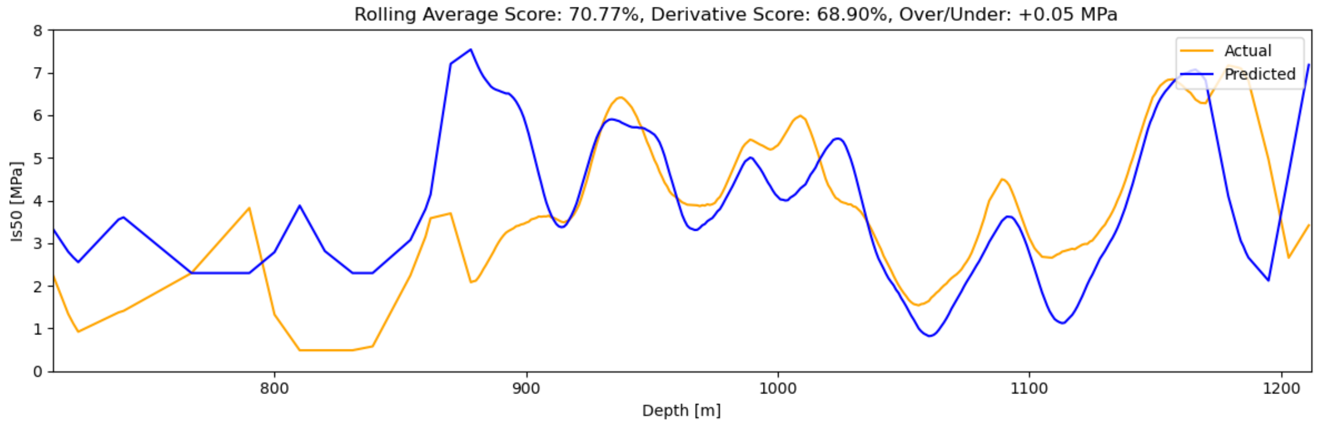


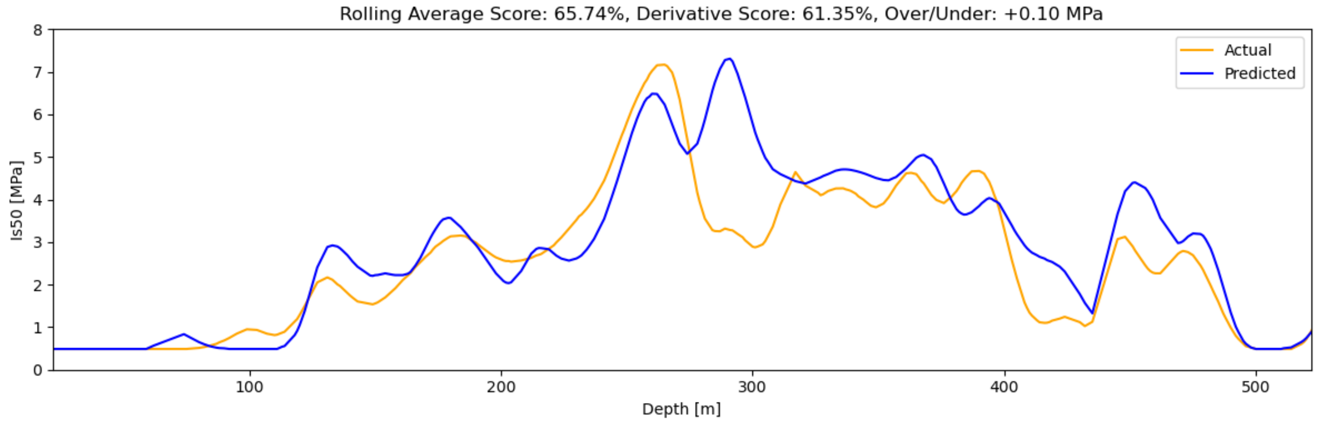
Average Rolling Mean Score: 48%

Evaluation Plot Examples

- Evaluation plots for a random forest model trained at a different mine with more PLT data (~13,000 training points)
- We have since used this approach at multiple mines with improved results
- As you add more training data, the model's predictive capability improves







Feature Importance

- Once the model was tuned, the top features of the model could be found using permutation importance
- This is done by permuting the samples of each respective feature and seeing how much the score changes from the baseline score
- Surprisingly, rock type isn't among the top 10 features, but instead many of the geotech and engineered features are near the top

Feature	Is50 Feature Importance
FFI	2.78%
Fracture frequency	1.92%
RQD	1.76%
Density	1.46%
Fracture spacing	1.27%
ASI	1.13%
MSI	0.70%
Minerals Present	0.68%
MSI percent sum	0.56%
Weighted DFI	0.50%

Predictions & Confidence Calculation

- Predictions made at every meter along 590 km of core, resulting in 590,000 predictions of Is50
- A confidence metric was calculated, defined as the percentage of features present, weighted by the feature importance

Confidence Range	Percentage of Is50 Predictions	Prediction Length [km]
> 75%	34%	200.0
50% - 75%	36%	211.8
25% - 50%	12.4%	72.95
< 25%	17.6%	103.5

Conclusions

- Data preprocessing is crucial
  - How you choose to represent and encode your data
  - Handling bias and noise; balancing between overfitting and underfitting
- Important to apply domain knowledge to your model. Engineering of new features from these logging data is essential to create an accurate predictive model of Is50.
- Traditional intact/defected strength domaining may use lithology and alteration facies to set strength domains. This work demonstrates that many factors contribute to rock strength, and at Cadia East, geotechnical factors are stronger indicators of strength than lithology.

Questions

Thank you for listening