2025 API PIPELINE CONFERENCE & EXPO

PIPELINE, CONTROL ROOM AND CYBERNETICS

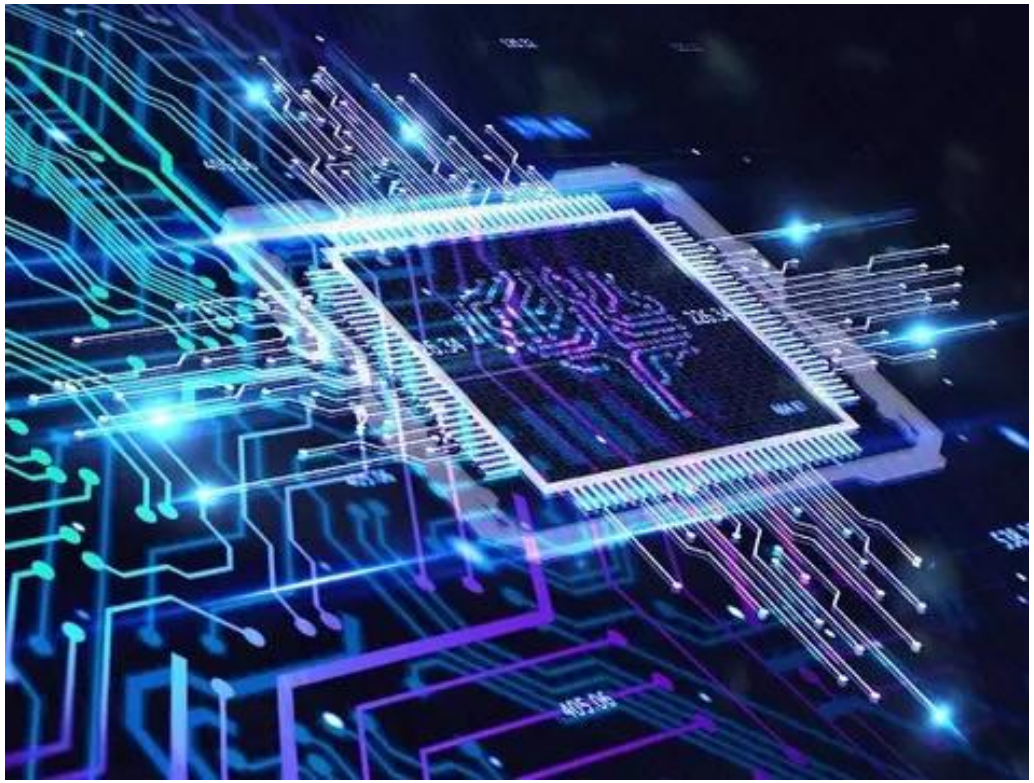
Machine Learning for Pipeline Risk

Overview & Panel Discussion

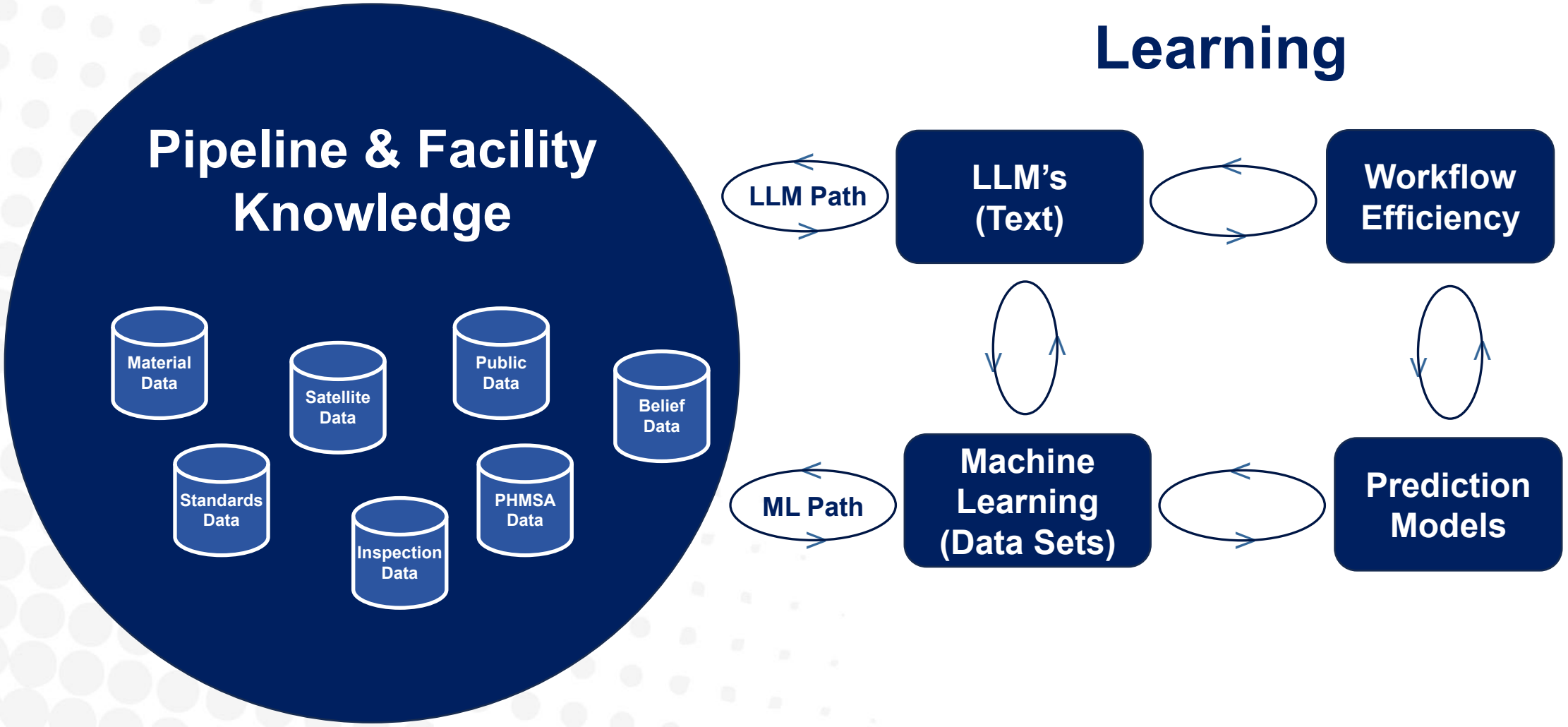
APRIL 28-30, 2025
Marriott Austin Downtown

Machine Learning for Risk

- AI & Machine Learning
- Risk Use Case
- Panel Discussion



Big Picture


Pipeline & Facility Knowledge
Material Data
Satellite Data
Public Data
Belief Data
Standards Data
Inspection Data
PHMSA Data
Learning
LLM Path
LLM's (Text)
Workflow Efficiency
ML Path
Machine Learning (Data Sets)
Prediction Models

Big Picture – LLM's

The equation for pipeline risk is probability of failure (PoF) x Consequence of Failure (CoF)

Human Enters "Truth"

Token count
33

Truth Converted to Tokens

```
<|im_start|>system<|im_sep|>The equation for pipeline risk is probability of failure (PoF) x Consequence of Failure (CoF)<|im_end|><|im_start|>user<|im_sep|><|im_end|><|im_start|>assistant<|im_sep|>
```

Training Data Learns (Weights) this Sequence

```
200264, 17360, 200266, 976, 42006, 395, 29412, 7031, 382, 34933, 328, 14090, 350, 25010, 37, 8, 1215, 25365, 6975, 328, 59925, 350, 8037, 37, 8, 200265, 200264, 1428, 200266, 200265, 200264, 173781, 200266
```

Big Picture – LLM's

what is the equation for pipeline risk?

DeepSearch Think Attach

Human Asks Question

LLM Returns "Best" Response based on Context & Weights

Risk = Probability of Failure (PoF) × Consequence of Failure (CoF)

Breakdown:

1. **Probability of Failure (PoF):**
   - This represents the likelihood that the pipeline will fail due to factors like corrosion, external damage, material defects, or operational errors.
   - It's often calculated using historical data, inspection results, or statistical models.
   - Example factors: pipe age, wall thickness, pressure, environmental conditions.
2. **Consequence of Failure (CoF):**
   - This measures the potential impact if a failure occurs, such as environmental damage, economic loss, or human safety risks.
   - It's often quantified in terms of cost (e.g., dollars), area affected, or population at risk.

Big Picture – LLM's & Data Stores

write code to calculate pipeline risk against an MS SQL table

DeepSearch Think Attach

```R
# Load required libraries
library(DBI)
library(odbc)

# Establish connection to MS SQL Server
con <- dbConnect(odbc::odbc(),
                 Driver = "ODBC Driver 17 for SQL Server",  # Adjust driver name if different
                 Server = "your_server",                    # e.g., "localhost" or IP address
                 Database = "your_db",                      # e.g., "PipelineDB"
                 UID = "your_user",                         # e.g., "sa"
                 PWD = "your_password",                     # Your SQL Server password
                 Port = 1433)                               # Default MS SQL port

# Check if connection is successful
if (dbIsValid(con)) {
  cat("Connected to MS SQL Server database successfully!\n")
} else {
  stop("Failed to connect to the database.")
}

# Function to calculate pipeline risk
calculate_pipeline_risk <- function(pof, cof) {
  return(pof * cof)
}

# Query to fetch pipeline data
query <- "SELECT segment_id, pof, cof FROM dbo.pipeline_data"
pipeline_data <- dbGetQuery(con, query)

# Calculate risk for each segment
pipeline_data$risk <- mapply(calculate_pipeline_risk, pipeline_data$pof, pipeline_data$cof)
```

Human Requests Code to Apply to Data Store since Data Stores are not Tokenized for Learning

LLM Returns Code to Apply to Data Sets

Code Requires Function Inputs, Review and Verification by Human or System

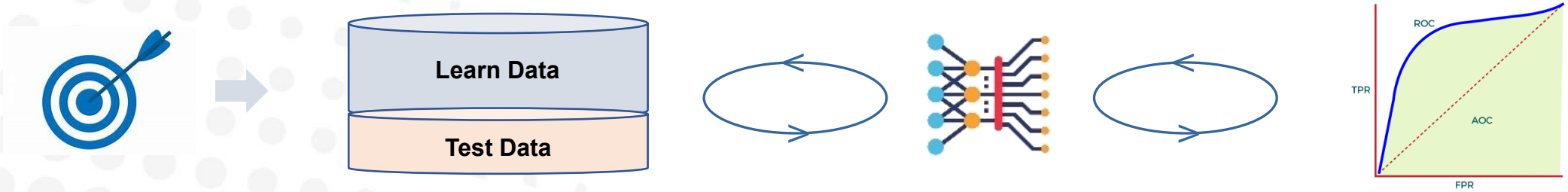
Big Picture - Machine Learning (ML)

Learning Target
(Threats, Consequence, Risk)

Training Data
(Observations)

Learned Model
(Methods, Tuning)

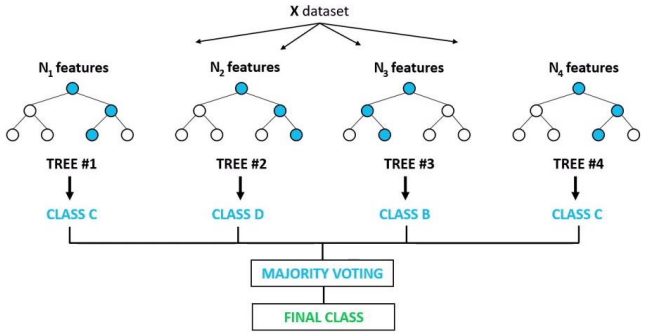
Performance & Insights
(Validation & Acceptance)



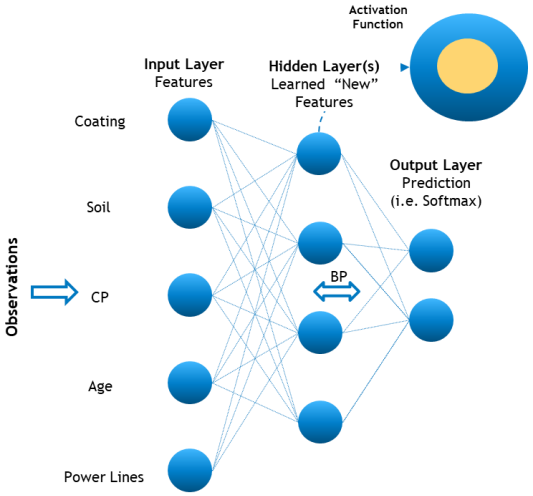
“Use LLM to Create Code to Support Machine Learning Process”

Learning Methods

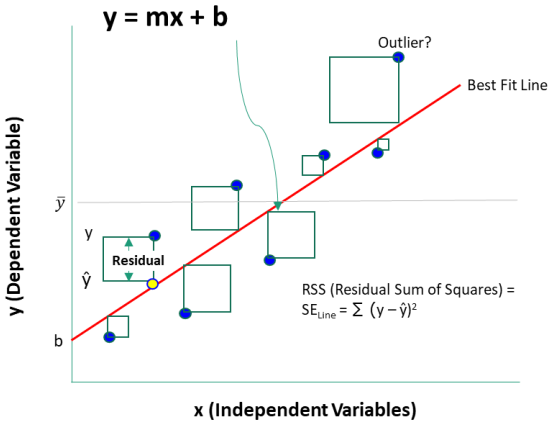
Tree Bagging



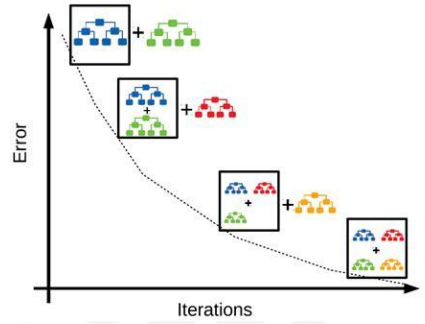
Neural Net



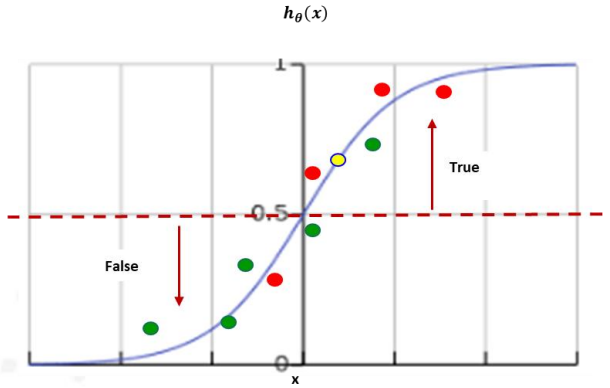
Linear Regression



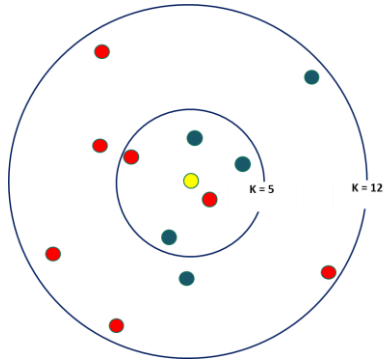
Tree Boosting



Logistic Regression



KNN



Learning Tasks

Classification

- Probability of Cracking
- Probability of Third-Party Damage
- Probability of Pipe Manufacturer

Regression

- Prediction of Corrosion Growth Rates
- Prediction of Inspection Costs
- Simulation of Deterministic Results

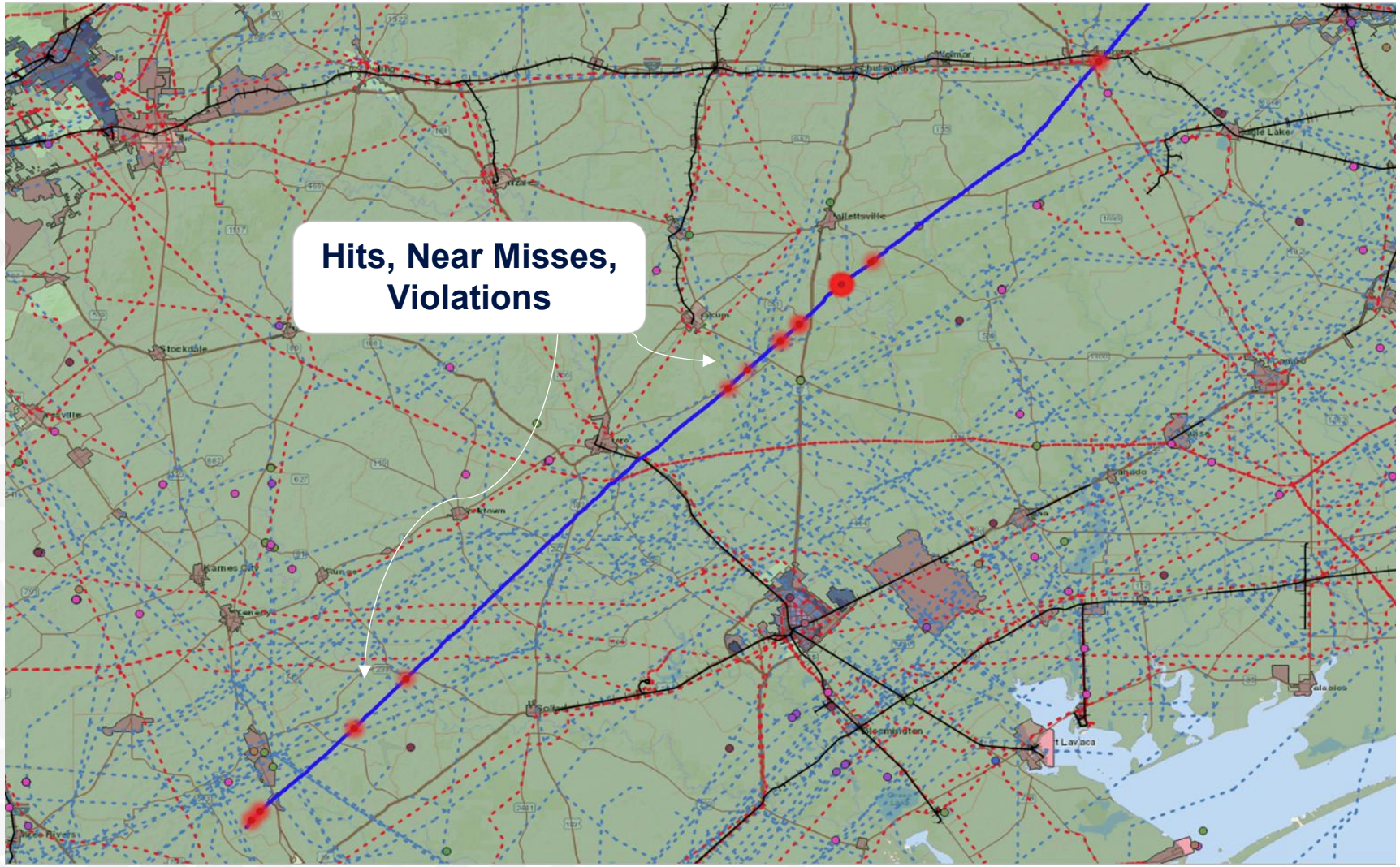
Time Series

- Prediction of CP Readings
- Prediction of Ground Bed Life
- Simulation of Seasonal Patterns

Common Questions

- Are Machine Learned Models an Improvement Over Deterministic Models?
- Do I have Enough of the Right Data?
- Are Patterns Inferential or Predictive? What's the Difference?
- Does the Model Meet Domain Expert Review?
- What Assets can I Apply the Learned Model?
- Is Performance Acceptable for Production Use?

Third Party Damage Risk



Training Data

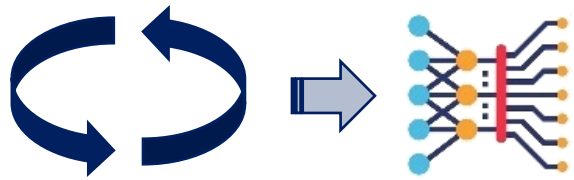
- Hits & Near-Misses
- One-Calls
- Pipe Properties
- Depth Cover
- Activity
- Land-Use
- Crossings
- Structures
- Patrol
- Public Awareness

Learning Target

Predictors

		Class	Diameter	DOC	Farmland	Install_Yr	LineMark	PatroFreq
All	A	A	All	A	All	All	All	All
No_Evidence	F	1.00	8.00	24.00	Not_Farmland	1,980.00	Line_of_Site	Semi-Annual
No_Evidence	F	1.00	8.00	24.00	Not_Farmland	1,980.00		Semi-Annual
No_Evidence	F	1.00	8.00	25.00	Not_Farmland	1,980.00	Line_of_Site	Semi-Annual
No_Evidence	F	1.00	8.00	33.00	Not_Farmland	1,980.00	Line_of_Site	Semi-Annual
No_Evidence	F	1.00	8.00	30.00	Not_Farmland	1,980.00	Line_of_Site	Semi-Annual
One_Call_Violation	T	1.00	8.00	26.00	Not_Farmland	1,980.00	Line_of_Site	Semi-Annual
One_Call_Violation	T	2.00	8.00	26.00	Not_Farmland	1,980.00	Line_of_Site	Semi-Annual
Near_Miss	T	2.00	8.00	29.00	Farmland	1,980.00	Line_of_Site	Semi-Annual
One_Call_Violation	T	2.00	8.00	24.00	Farmland	1,980.00	Line_of_Site	Semi-Annual
Near_Miss	T	2.00	8.00	28.00	Farmland	1,980.00	Line_of_Site	Semi-Annual
Near_Miss	T	2.00	8.00	34.00	Farmland	1,980.00	Line_of_Site	Semi-Annual
Near_Miss	T	2.00	8.00	41.00	Farmland	1,980.00	Line_of_Site	Semi-Annual
No_Evidence	F	2.00	8.00	31.00	Farmland	1,980.00	Line_of_Site	Bi-Weekly
No_Evidence	F	3.00	8.00	24.00	Farmland	1,980.00	Line_of_Site	Bi-Weekly

Training Data



Machine Learning Process

Learned Model

Learned Model Classification Performance

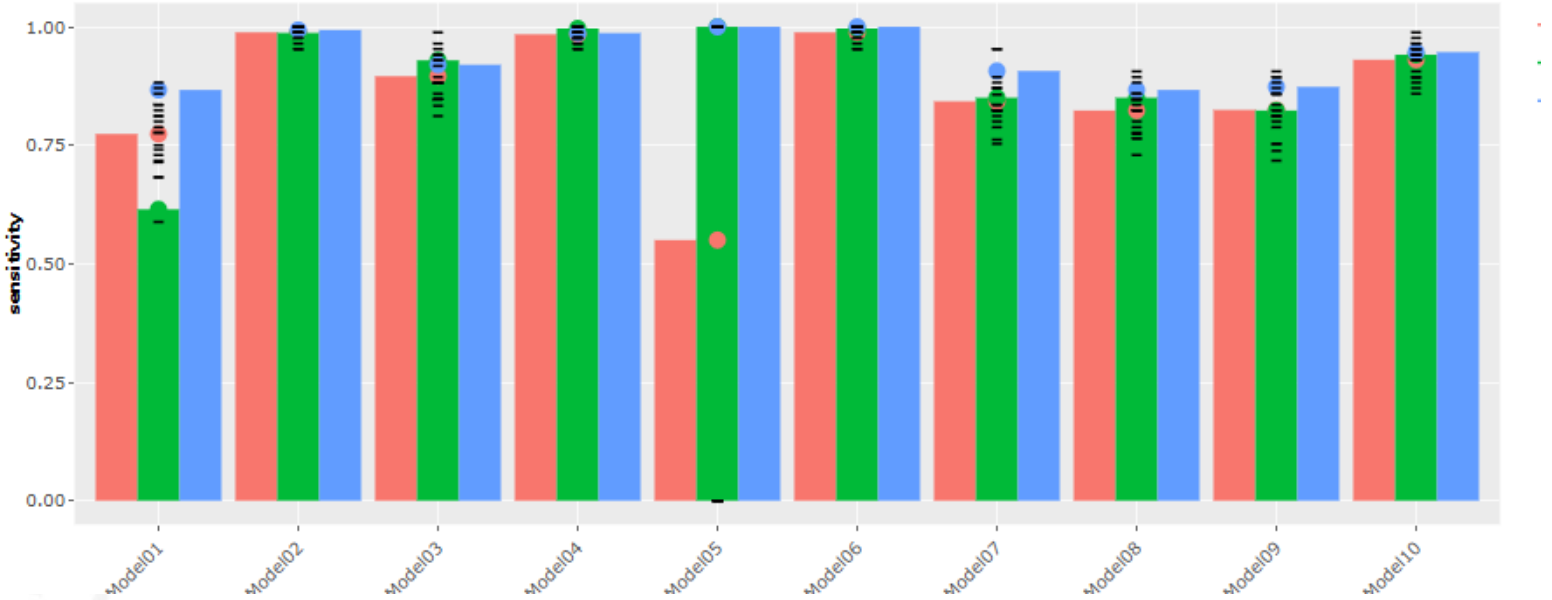
Metrics

- Accuracy
- Sensitivity
- Specificity
- AUC

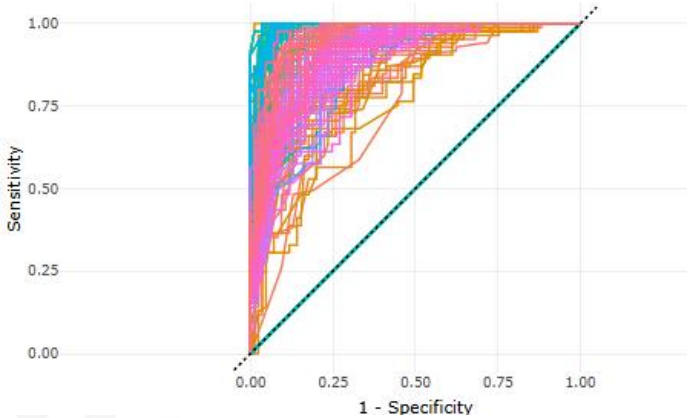
Model

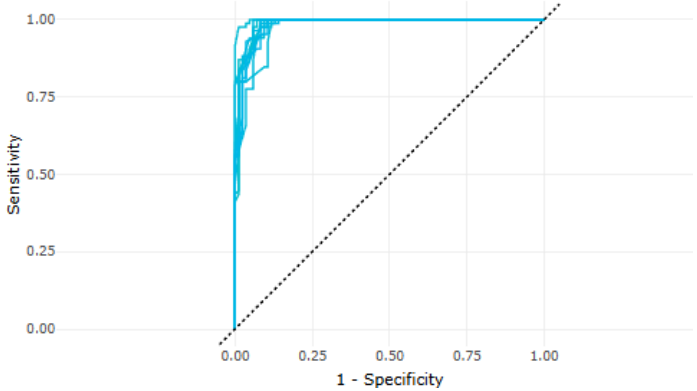
- Xgboost Method
- 2000 Trees
- 5 Depth
- 2 Min Obs
- .0001 Loss

Candidate Model Performance



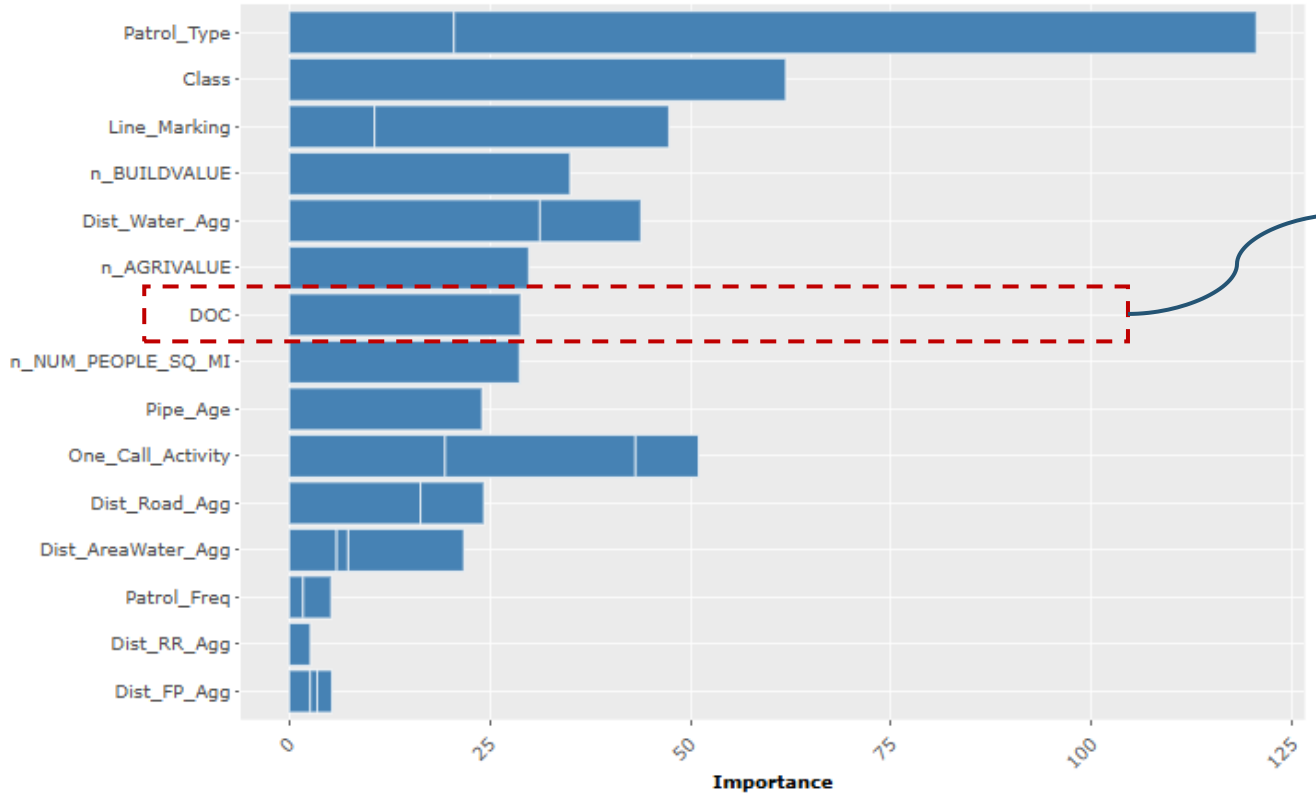
Candidate Model ROC's



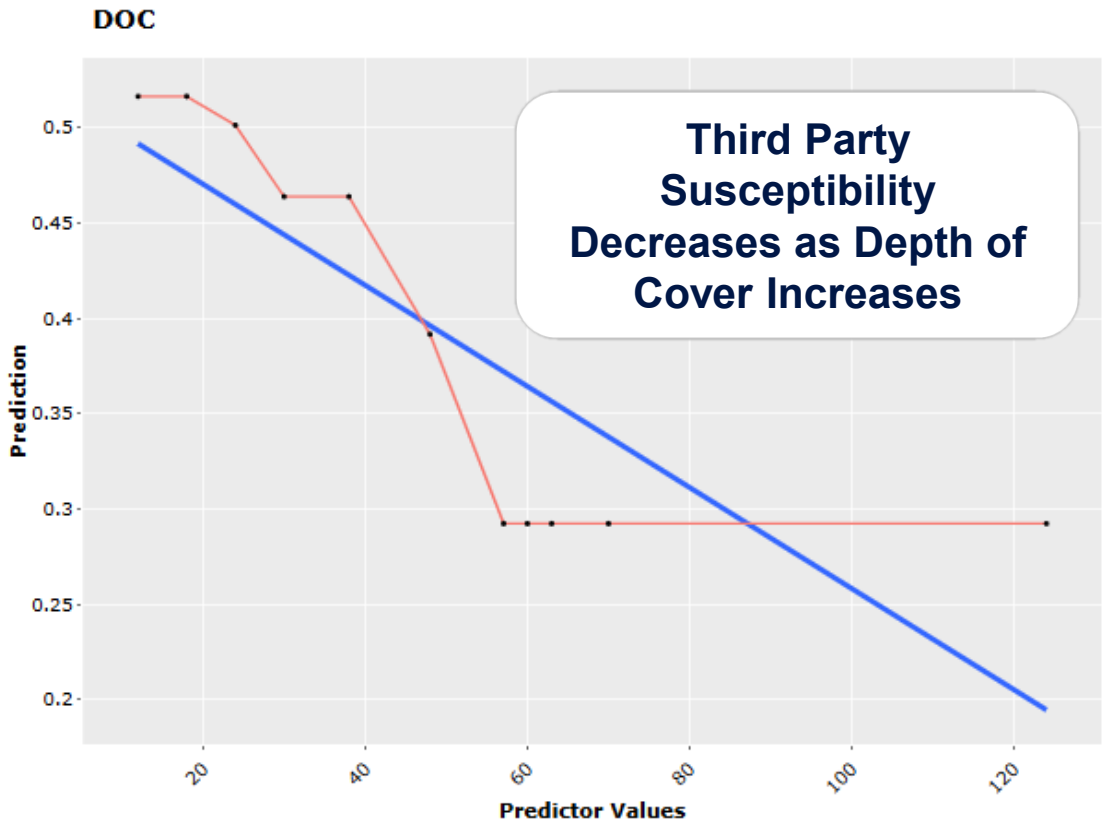


Learned Model – Global Weights

Model Predictor Importance

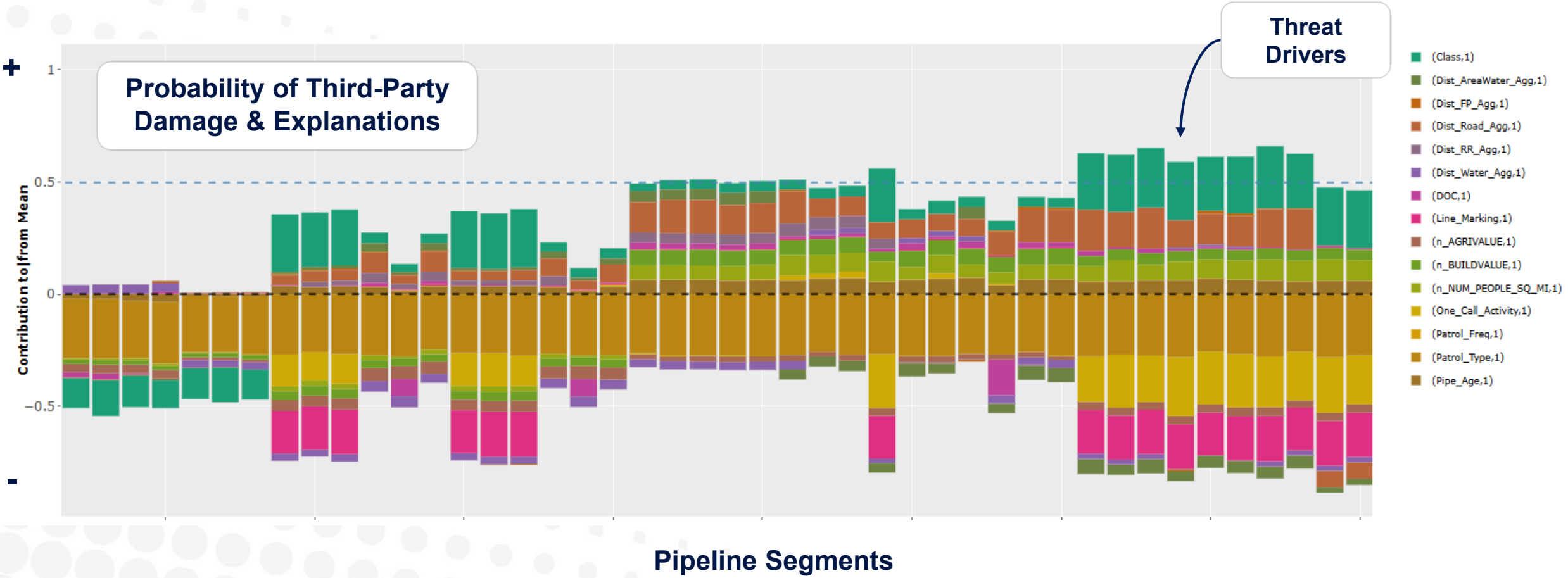


Model Predictor Directionality

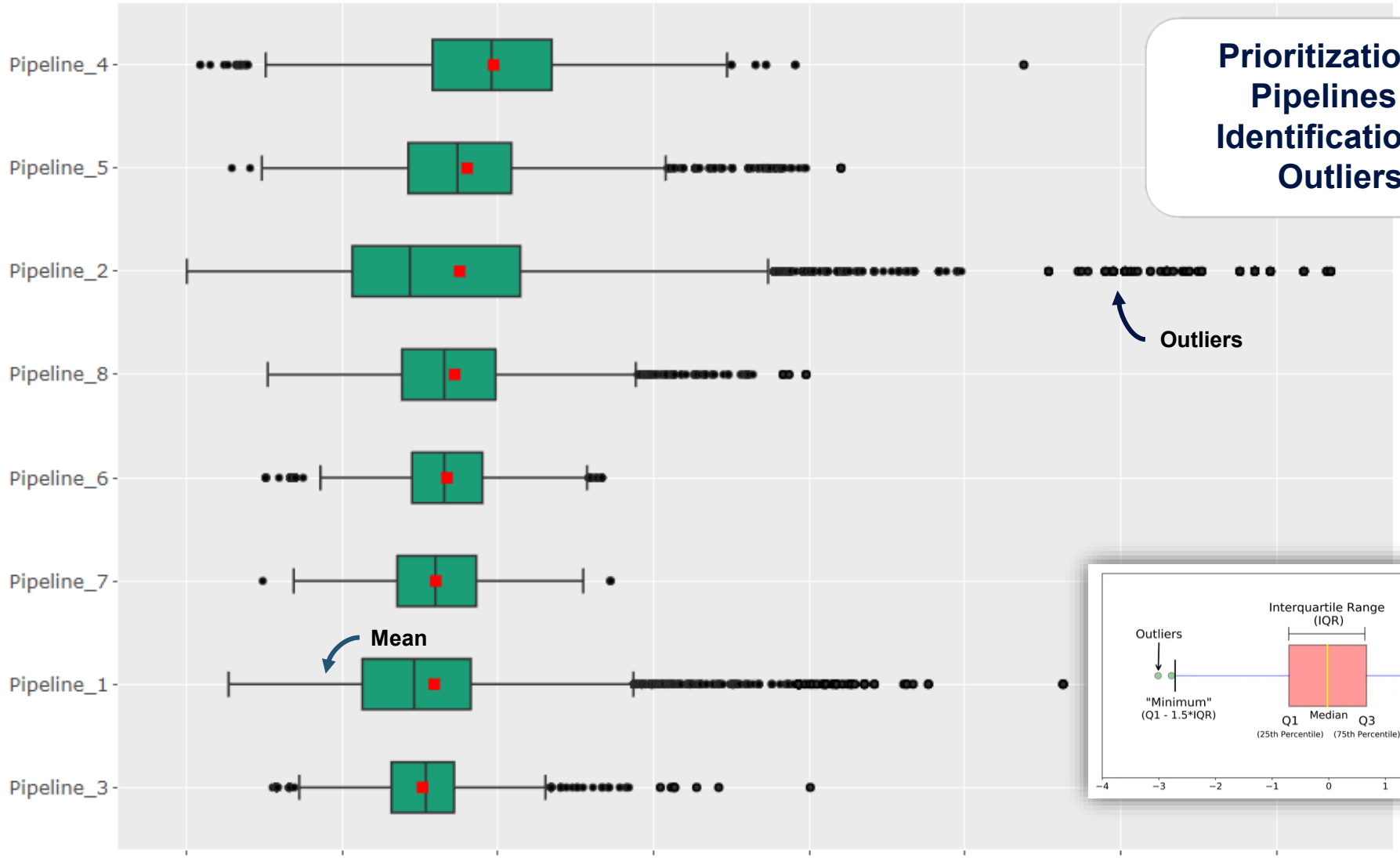


Third Party Susceptibility Decreases as Depth of Cover Increases

Model Application & Explanation



Model Application



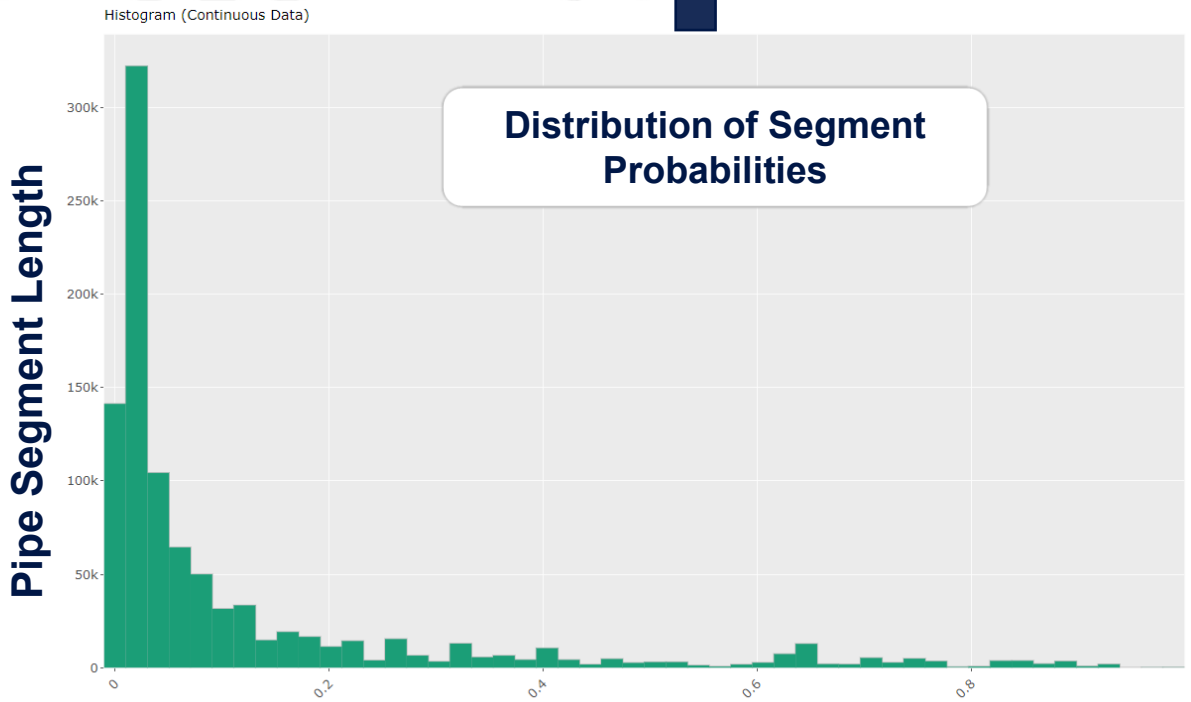
Prioritization of Pipelines & Identification of Outliers

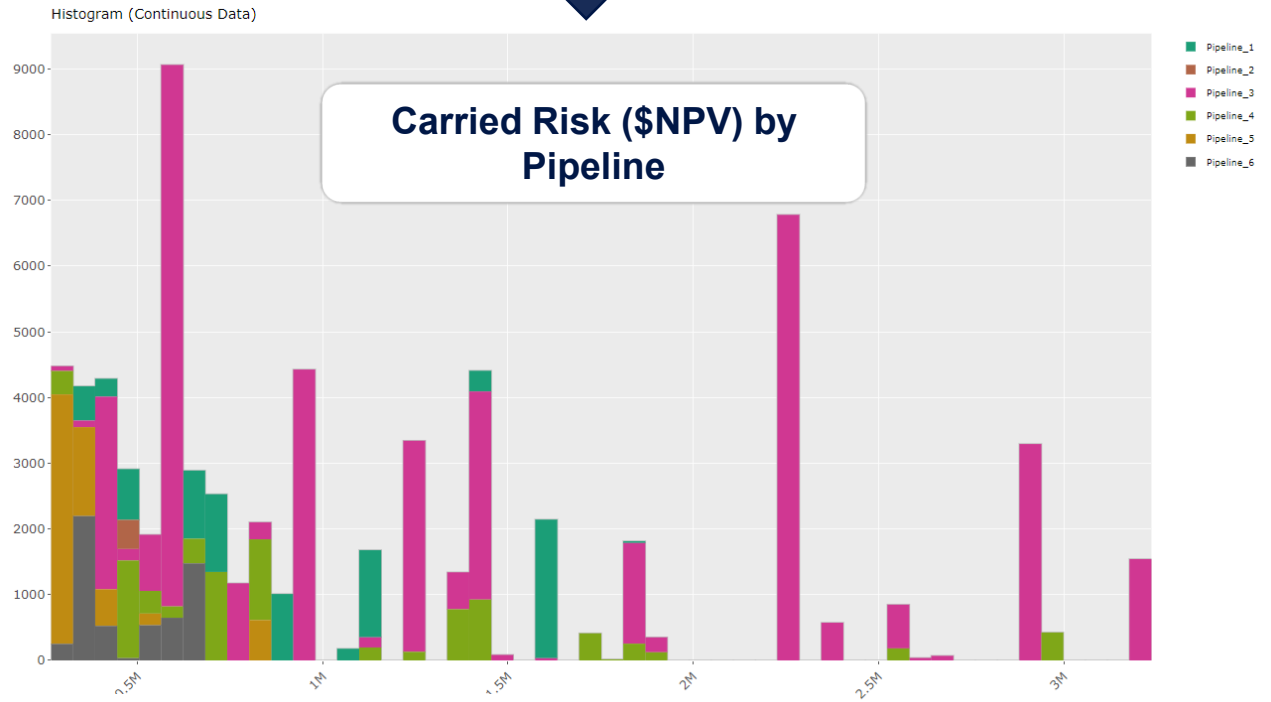
Third Party Probability by Pipe Segment

Monetized Risk
Time Independent Threats

QRA

- Consider resistance (pipe WT, toughness)
- Normalize to incident distributions (P50\P99)





Machine Learned Results

QRA Monetized Risk

Summary

Machine Learning has an Important Role in Risk

- **Leverage Existing Data**
- **Find Actionable Patterns based on Actual Observations**
- **Validate & Explain Prediction Results**
- **Align with AI Technology Progression & Adoption**

Panel Discussion

- Michael Gloven, President, Pipeline-Risk (PLR)
- Victor Anisi, phd Candidate, Colorado School of Mines
- Joel Bassani,
- Syed Aijaz, Pipeline Integrity Engineer, TC Energy

2025 API PIPELINE CONFERENCE & EXPO

PIPELINE, CONTROL ROOM AND CYBERNETICS

Thank You

Michael Gloven, PE

michaelgloven@pipeline-risk.com

APRIL 28-30, 2025
Marriott Austin Downtown