Machine Learning for Pipeline Integrity

Mike Gloven, PE
Managing Partner
Pipeline-Risk
www.pipeline-risk.com

Do You Want to Get More from Your Inspection Investments?

Seven Machine Learning Methods

1) Identify Outliers
2) Measure Data Importance
3) Impute Missing Data
4) Find Natural Groupings
5) Learn Threat Susceptibility Patterns
6) Learn Threat Severity Patterns
7) Measure Uncertainty

1 - Identify Outliers

Corrosion Observations

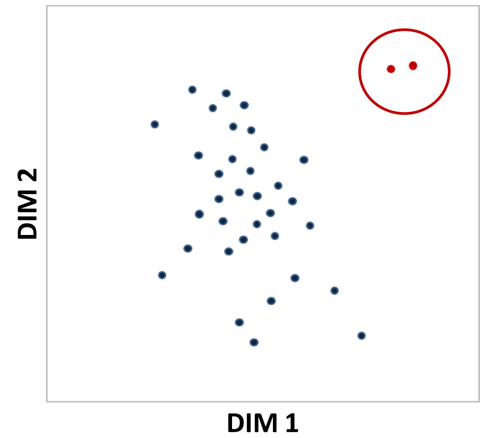
Observe	CP	Coating	Power	Corrosivity
T	-0.90	TGF	-	High
T	-0.85	TGF	-	High
F	-0.90	TGF	-	Medium
F	-0.85	TGF	-	Low
T	-0.90	TGF	-	Low
F	-0.80	TGF	-	Low
F	-0.85	TGF	-	High
T	-0.85	TGF	Yes	High
F	-0.97	TGF	Yes	Medium
F	-0.95	FBE	Yes	Low
F	-0.90	FBE	-	Low
T	-0.85	FBE	-	Low
F	-0.94	FBE	-	High
T	-0.85	FBE	-	High
F	-0.85	FBE	-	Medium
F	-0.95	FBE	-	Low
T	-0.85	FBE	-	Low
F	-0.90	FBE	-	Low
F	-0.90	FBE	-	High
T	-0.85	FBE	-	High

Method

t-SNE*

DIM 1	DIM 2
-6	5
-4	-1
19	30
-2	3
-5	4
2	2
25	31
1	8
-3	7
2	-2
4	6

Outliers



* t-SNE (t-Distributed Stochastic Neighbor Embedding) is an unsupervised, non-linear technique, which takes the original data set in many dimensions and reduces it to a low dimensional graph

1 - Identify Outliers

Dig Repair Costs

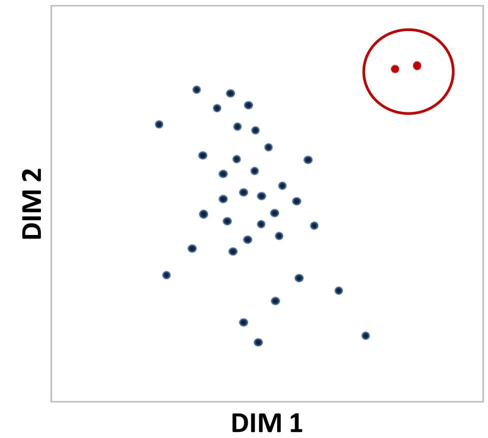
Cost	WT	Install	Diameter	Grade
25,000	-	A	12	C
75,000	-	A	24	B
1,000,000	Y	C	36	B
150,000	Y	B	8	A
35,000	-	B	8	A
15,000	-	A	8	C
95,000	-	A	8	C
25,000	-	C	8	B
75,000	-	C	12	D
81,000	Y	B	24	C
150,000	-	D	24	A
35,000	-	C	12	C
15,000	-	A	12	B
95,000	Y	B	12	B
25,000	-	B	6	A
81,000	-	D	8	A
150,000	-	D	20	C
35,000	-	A	8	C
15,000	-	C	8	B
95,000	-	B	10	D

Method

t-SNE*

DIM 1	DIM 2
-6	5
-4	-1
19	30
-2	3
-5	4
2	2
25	31
1	8
-3	7
2	-2
4	6

Outliers



* t-SNE (t-Distributed Stochastic Neighbor Embedding) is an unsupervised, non-linear technique, which takes the original data set in many dimensions and reduces it to a low dimensional graph

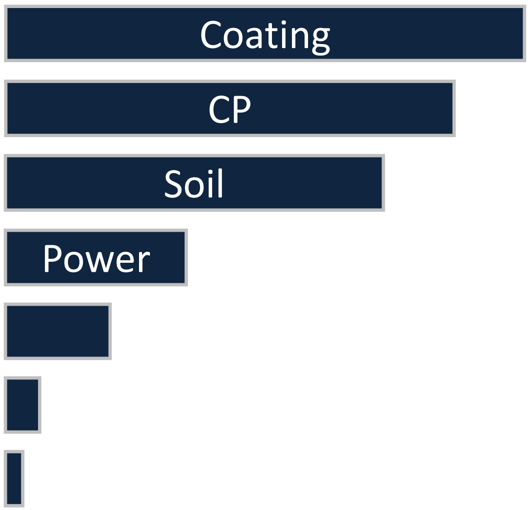
2 - Measure Data Importance

Corrosion Observations

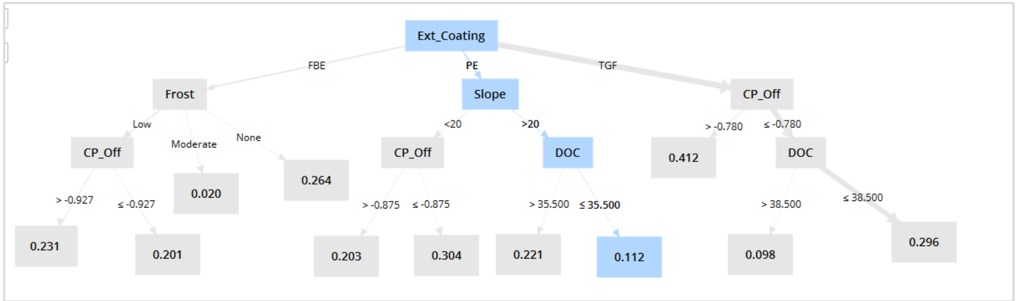
Observe	CP	Coating	Power	Corrosivity
T	-0.90	TGF	-	High
T	-0.85	TGF	-	High
F	-0.90	TGF	-	Medium
F	-0.85	TGF	-	Low
T	-0.90	TGF	-	Low
F	-0.80	TGF	-	Low
F	-0.85	TGF	-	High
T	-0.85	TGF	Yes	High
F	-0.97	TGF	Yes	Medium
F	-0.95	FBE	Yes	Low
F	-0.90	FBE	-	Low
T	-0.85	FBE	-	Low
F	-0.94	FBE	-	High
T	-0.85	FBE	-	High
F	-0.85	FBE	-	Medium
F	-0.95	FBE	-	Low
T	-0.85	FBE	-	Low
F	-0.90	FBE	-	Low
F	-0.90	FBE	-	High
T	-0.85	FBE	-	High

Method

Entropy & Information Gain



Shannon's Entropy $H = -p_a * log_2(pa) - p_b * log_2(pb)$

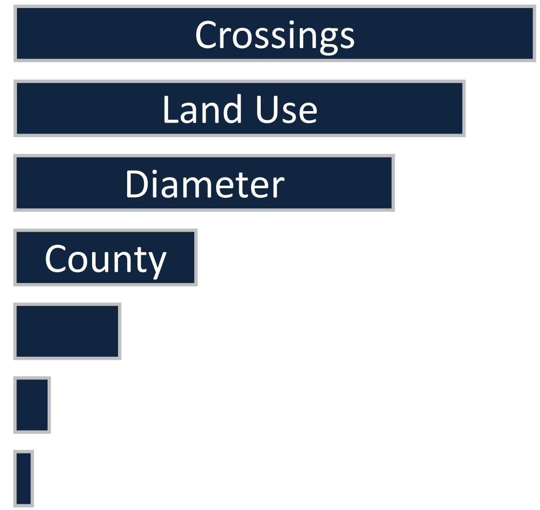


Dig Repair Costs

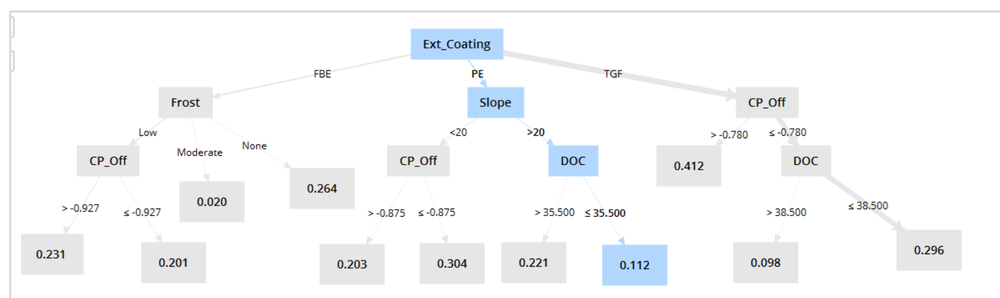
Cost	Crossings	Land Use	Diameter	County
25,000	-	A	12	C
75,000	-	A	24	B
1,000,000	Y	C	36	B
150,000	Y	B	8	A
35,000	-	B	8	A
15,000	-	A	8	C
95,000	-	A	8	C
25,000	-	C	8	B
75,000	-	C	12	D
81,000	Y	B	24	C
150,000	-	D	24	A
35,000	-	C	12	C
15,000	-	A	12	B
95,000	Y	B	12	B
25,000	-	B	6	A
81,000	-	D	8	A
150,000	-	D	20	C
35,000	-	A	8	C
15,000	-	C	8	B
95,000	-	B	10	D

Method

Entropy & Information Gain



Shannon's Entropy $H = -p_a * log_2(pa) - p_b * log_2(pb)$



3 - Impute Missing Data

Impute Pipe Toughness

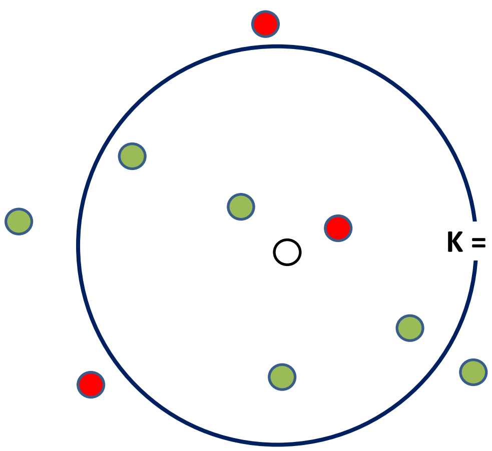
Toughness	WT	Install	Diameter	Grade
35	.281	1975	12	B
38	.250	1981	24	X-52
MISSING	.322	1981	36	X-52
40	.312	1965	8	B
42	.500	1972	8	X-52
42	.188	1962	8	X-52
45	.322	1990	8	X-42
35	.279	1975	8	X-60
39	.250	1981	12	B
50	.281	1981	24	X-52
52	.250	1965	24	X-52
48	.322	1972	12	X-42
47	.312	1962	12	X-52
42	.500	1990	12	X-52
38	.188	1975	6	X-42
35	.322	1981	8	X-60
32	.279	1981	20	X-42
50	.250	1965	8	X-60
52	.281	1972	8	B
48	.250	1962	10	X-52

Method

K-Nearest Neighbor

Imputed Data

Toughness	WT	Install	Diameter	Grade
35	.281	1975	12	B
38	.250	1981	24	X-52
36	.322	1981	36	X-52
40	.312	1965	8	B
42	.500	1972	8	X-52
42	.188	1962	8	X-52
45	.322	1990	8	X-42
35	.279	1975	8	X-60
39	.250	1981	12	B
50	.281	1981	24	X-52



K-Nearest Neighbor

K = 5 (80% Confident Green)

4 - Find Natural Groupings

Digs & Limited Observations of Cracks

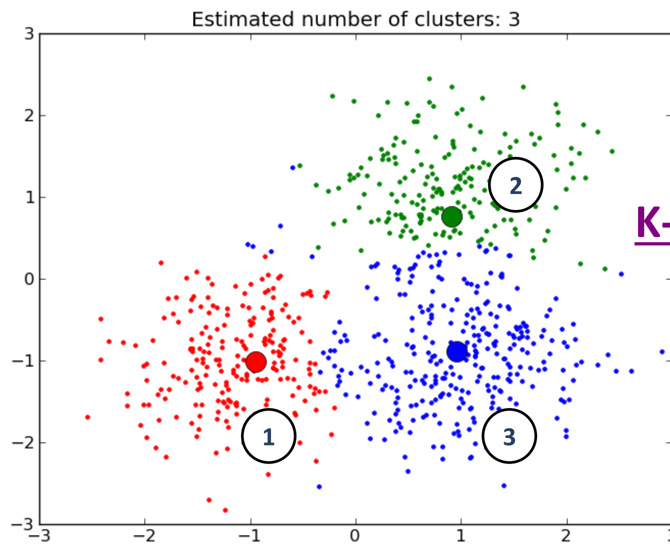
Crack	CP	Coating	Power	Corrosivity
-	-0.90	TGF	-	High
-	-0.85	TGF	-	High
-	-0.90	TGF	-	Medium
-	-0.85	TGF	-	Low
-	-0.90	TGF	-	Low
evidence	-0.80	TGF	-	Low
-	-0.85	TGF	-	High
-	-0.85	TGF	Yes	High
-	-0.97	TGF	Yes	Medium
-	-0.95	FBE	Yes	Low
-	-0.90	FBE	-	Low
-	-0.85	FBE	-	Low
-	-0.94	FBE	-	High
-	-0.85	FBE	-	High
-	-0.85	FBE	-	Medium
-	-0.95	FBE	-	Low
-	-0.85	FBE	-	Low
-	-0.90	FBE	-	Low
-	-0.90	FBE	-	High
-	-0.85	FBE	-	High

Method

K-Means Clustering

Cluster
1
2
3
1
1
2
3
3
2

← **Investigative Dig Finds Evidence**



K-Means Clustering

5 - Learn Classification Patterns - Susceptibility

What is the Potential **Susceptibility** of the Threat to the Asset?

External Corrosion Observations

Observe	CP	Coating	Power	Corrosivity
T	-0.90	TGF	-	High
T	-0.85	TGF	-	High
F	-0.90	TGF	-	Medium
F	-0.85	TGF	-	Low
T	-0.90	TGF	-	Low
F	-0.80	TGF	-	Low
F	-0.85	TGF	-	High
T	-0.85	TGF	Yes	High
F	-0.97	TGF	Yes	Medium
F	-0.95	FBE	Yes	Low
F	-0.90	FBE	-	Low
T	-0.85	FBE	-	Low
F	-0.94	FBE	-	High
T	-0.85	FBE	-	High
F	-0.85	FBE	-	Medium
F	-0.95	FBE	-	Low
T	-0.85	FBE	-	Low
F	-0.90	FBE	-	Low
F	-0.90	FBE	-	High
T	-0.85	FBE	-	High

Learning Data (first 14 rows)

Held Back Test Data (last 6 rows)

Learned Model

Model Performance

Actual	Probability True	Prediction*	Performance
F	.46	F	True Negative
F	.55	T	False Positive
T	.34	T	False Negative (miss)
F	.14	F	True Negative
F	.23	F	True Negative
T	.98	T	True Positive

** 50% Threshold*

Key Terms

Sensitivity - % Correct True Positives

Specificity - % Correct True Negatives

6 - Learn Regression Patterns - Severity

What is the Potential **Severity** of the Threat to the Asset?

External Corrosion Observations

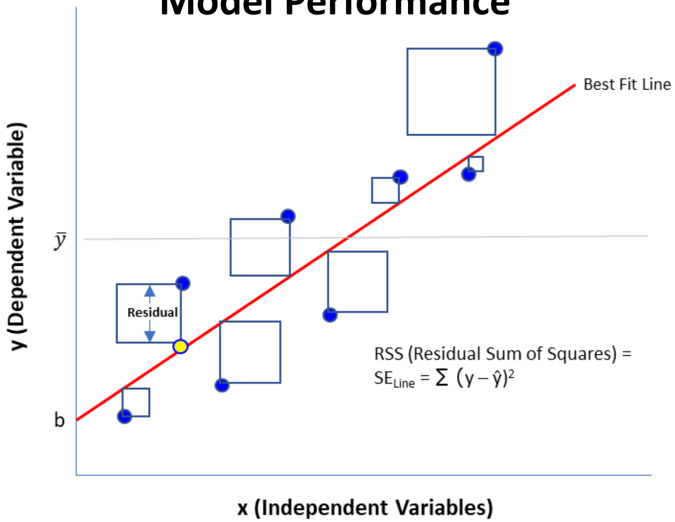
Observe	CP	Coating	Power	Corrosivity
.23	-0.90	TGF	-	High
.43	-0.85	TGF	-	High
.29	-0.90	TGF	-	Medium
.75	-0.85	TGF	-	Low
.54	-0.90	TGF	-	Low
.32	-0.80	TGF	-	Low
.31	-0.85	TGF	-	High
.41	-0.85	TGF	Yes	High
.17	-0.97	TGF	Yes	Medium
.19	-0.95	FBE	Yes	Low
.22	-0.90	FBE	-	Low
.27	-0.85	FBE	-	Low
.33	-0.94	FBE	-	High
.42	-0.85	FBE	-	High
.20	-0.85	FBE	-	Medium
.30	-0.95	FBE	-	Low
.25	-0.85	FBE	-	Low
.41	-0.90	FBE	-	Low
.49	-0.90	FBE	-	High
.41	-0.85	FBE	-	High

Learning Data

Held Back Test Data

Learned Model

Model Performance



Key Terms

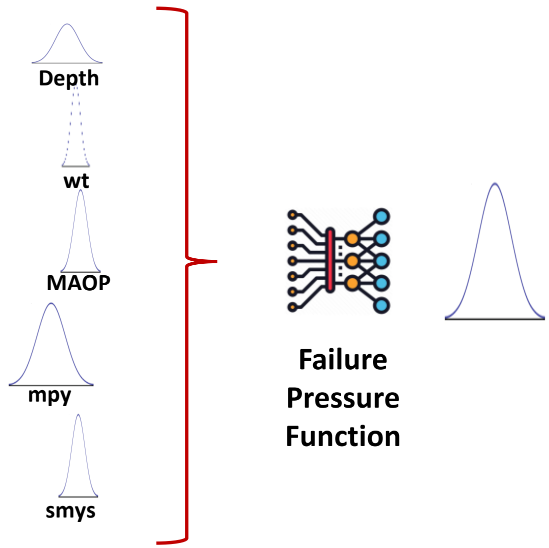
RMSE – Root Mean Square Error

R2 – How well Change in Predictors Explain Change in Predictions

7 - Measure Uncertainty

What is the Probability a Calculated Failure Pressure might exceed Safe Operating Pressure when Considering Uncertainties?

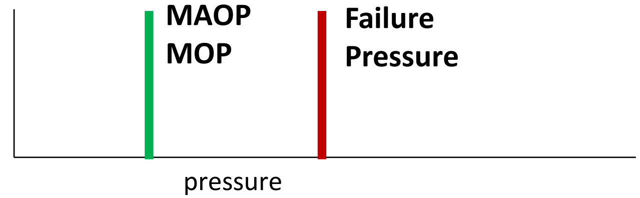
Data Inputs as Sampling based on Uncertainties



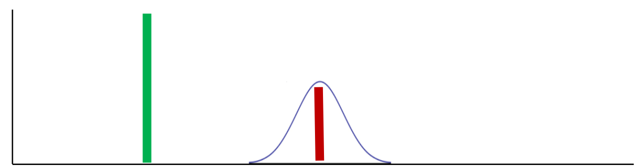
Data	Distribution	Standard Deviation	Lower Limit	Upper Limit
Reported Depth %	Normal	7.8%	Non-Negative	-
MAOP|MOP	Normal	10%	0%	10%
mpy	Gumbel	10%	-10%	20%
Wall Thickness	Gumbel	10%	-12.5%	12.5%
SMYS	Gumbel	15%	0%	10%
Outside Diameter	Gumbel	.3%	0%	.3%

For Example Purposes Only *

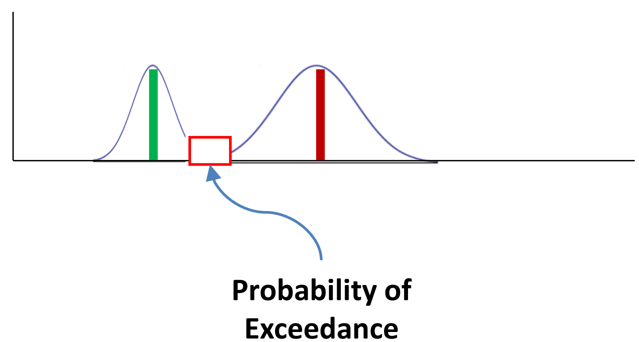
No Uncertainty



Inspection Reading Uncertainty



Consider All Function Input Uncertainties



Machine Learning Models

“All Models are Wrong, but Some Models are Useful”

- George E.P. Box

Famous Quality Control Mathematician

Leading Open-Source Machine Learning Platforms

R Studio

R Project

Scikit-learn

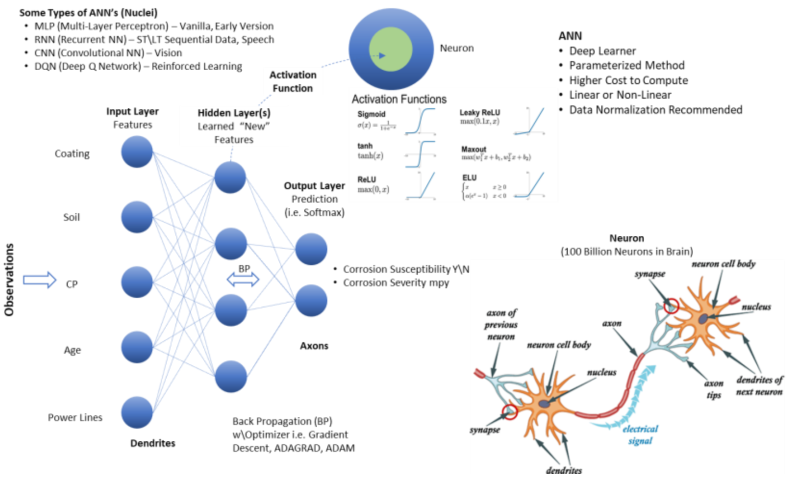
Python

Keras

TensorFlow

Machine Learning for Pipeline Integrity

Deep Learning – Artificial Neural Network (ANN)



Mike Gloven, PE
Managing Partner
Pipeline-Risk
www.pipeline-risk.com

The Course

Practical Application of Machine Learning - Asset Integrity & Risk

1/2 Day Virtual Cass

What You will Learn

You will learn the basics of machine learning & how the process is used to support integrity and risk management. Strategic & technical learning includes feature analysis, regression, classification, clustering, outlier detection, model validation, basic inferential statistics & data quality assurance.

Testimonials

"Instructor did a great job going over important concepts"

"Very good course, got me excited to utilize machine learning for my program"

"Great overview of machine learning"

"Very interesting and well delivered"

Interactive Software

An interactive educational version of our software is provided including course content and demo data. The secure web based application allows the use of your data to experience how machine learning brings value to asset integrity management.

Schedule & Agenda

We hold the course on the first Tuesday of each month. Duration is 8 - 12:00 noon US Mountain Time. The first three hours are instructional with 10 minute breaks at top of the hour, the last hour is for open & hands-on questions. All instructional material, software access and demonstration data is sent prior to the course after payment.

Register

Cost of the course which includes continuous access to the software is $495/person. To register, please contact michaelgloven@pipeline-risk.com and indicate your preferred course date. We also hold courses on a per company basis to fit your schedule if there are at least six participants. In either case, please contact us by e-mail and we will provide further instructions.

Prerequisites

This is a basic machine learning course presenting fundamental elements where you can quickly work thru the process with data familiar to you. Prerequisites are familiarization with asset integrity & inspection concepts, proficiency in math & basic statistics, and ability to work with data (rows-columns).