Application of Data Analytics to Mine Nuclear Plant Maintenance Data

Dave Olack, Principal Technical Leader
Nuclear Sector – Plant Engineering
Charlotte, NC

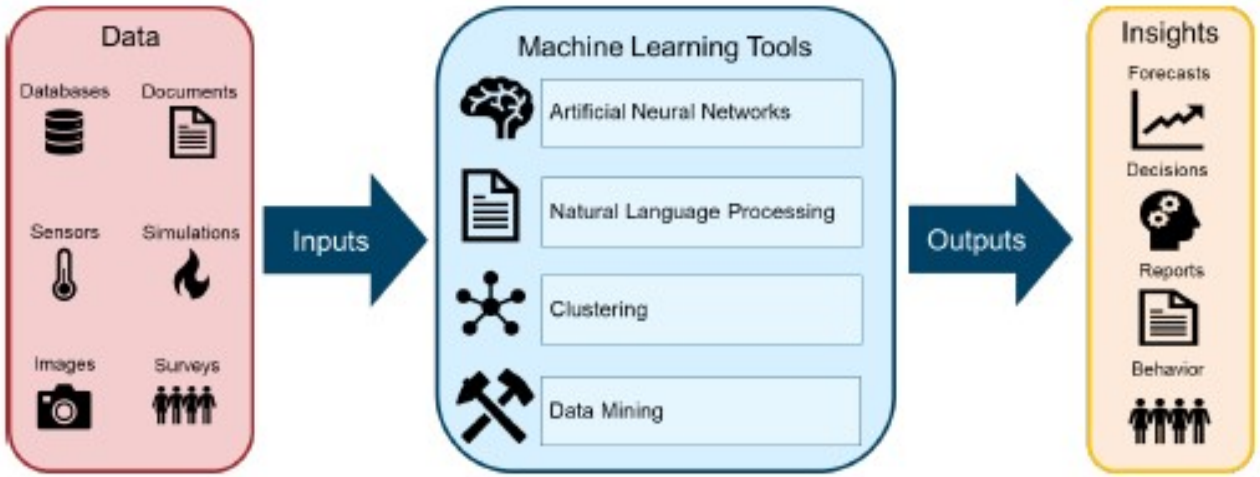
U.S. NRC AI Workshop
August 2021

Background

- Commercial nuclear power utilities have large amounts of equipment maintenance records captured over many decades
- Due to a combination of advancements in computational capabilities and external market financial pressures on the nuclear power industry, EPRI has engaged in a project to analyze and more effectively utilize maintenance data in order to implement more cost-effective preventative maintenance (PM) strategies
- Some utilities have applied a combination of natural language processing (NLP) and an artificial neural network to evaluate similar plant process data to improve the administration and evaluation of programmatic data to reduce the required labor resources.

Project Objective

- Utilizing machine learning (ML) and data analytics (DA), determine to what extent these analysis tools can analyze large volumes of equipment data and provide insights leading to improving plant equipment reliability and/or reduce significant equipment related events
  - EPRI has collected approximately 18 million maintenance work order records from 10 utilities over the last few years
- Using NLP, compare the work order history of similar components across a number of different utilities and plants
  - Compare statistical annual costs of each matching (similar) component with existing PM strategy
- Evaluate the impact of different PM strategies based on total CM and PM costs (both labor and material)



Technical Details

- **From the 10 EPRI utility members that routinely provide WO data, 4 were selected to be used for the project due to:**
  - the volume of work order data details
  - commonality of work order data fields
- **Prepare work order data -> Component ID Dataset**
  - Remove/Link duplicate work order entries
  - Concatenate text entries and sum hours, costs for each component id
    - Total PM and CM hours & costs

Data Quality

Table 3-1
Data quality assessment. The IDs of the member utilities whose data were used in this project are highlighted in yellow.

Utility ID	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
AP-928 Job Type																		
AP-928 Work Type																		
Actual Hours																		
As Found Condition Codes																		
Component #																		
Component Description																		
Component ID																		
Component Type																		
Criticality																		
E_Code																		
Equipment #																		
Equipment Name																		
Equipment Status																		
Equipment Type																		
Manufacturer																		
Material Cost																		
Model #																		
MonthYear																		
PM Category Code																		
PM Description																		
PM Frequency Code																		
SPV																		
Serial #																		
Service Condition																		
System																		
Work Order #																		
Work Order Description																		
Count >= 2	24	27	24	25	25	25	22	23	23	23	22	24	24	20	21	21	22	17
Count >= 10	24	19	21	20	19	19	21	20	19	19	20	17	17	18	18	16	16	16
Count Avg	24	23	23	23	22	22	22	22	21	21	21	21	21	19	20	19	19	17

Note: SPV stands for single-point vulnerability

Matching Component ID's

Process (NLP) for matching similar Component IDs

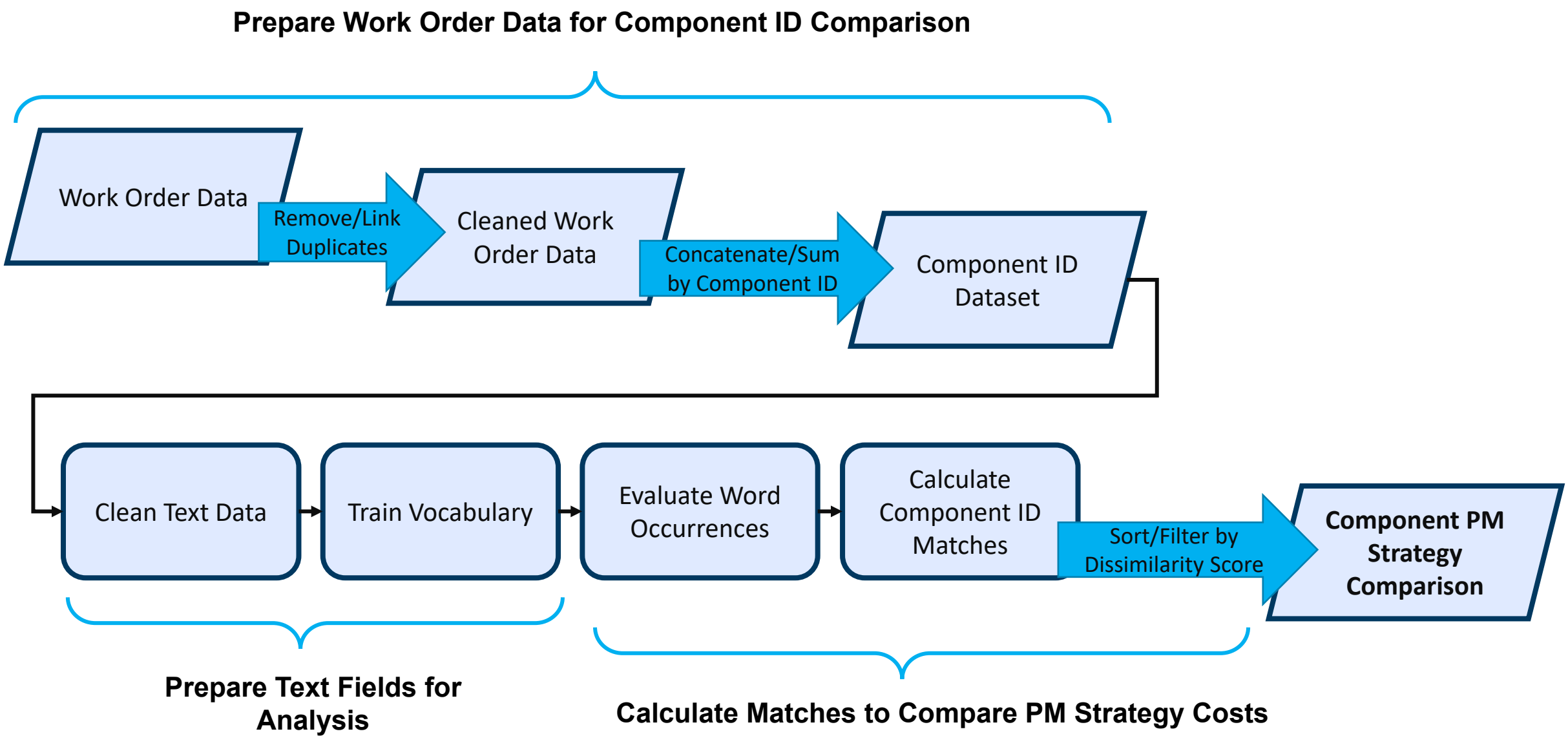
1. Collect work orders for each Component ID and combine into the **Component ID Dataset**
2. Train word models to obtain vocabulary in the dataset for both text fields
3. Evaluate word and phrase occurrences for each Component ID in the dataset
4. Compute the pairwise cosine distance for each Component ID to other Component IDs in the dataset
   a) Separate dissimilarity score for each field $\left(\mathrm{D} = 1 - \frac{u \cdot v}{\|u\|_2 \|v\|_2}\right)$, which ranges from 0 (same) to 1 (different)
   b) Potential matches have $D_{COMP\ DESC} < 0.5$
   c) Matches are sorted based on $D_{WO\ Activity}$

Example Component Description Matching

Table 4-4
Example scores for Component Description field matches. As the similarity of the text string improves, the score decreases.

Component Description, Cleaned and Concatenated	Score
safety injection motor	(Ref, 0)
safety injection system safety injection pump motor	0.1296
safety injection pump motor train	0.2254
safety injection tank outlet valve motor	0.2928
high pressure safety injection pump motor motor heater breaker lh	0.3333

PM Strategy Comparison Overview



Clean Data Text & Train Vocabulary

- Preparing Text Fields for Analysis
  - Creating a list of stopwords streamlines the text analysis process
  - Acronym translation matrix improves matches and larger volumes of acronyms will lead to more accurate text matches

B
LIST OF STOPWORDS

The table below lists the stopwords used in this analysis. Common names and surnames were also removed.

a	aa	able	about	above	according	accordingly
across	actually	after	afterwards	again	against	ain't
all	allow	allows	almost	alone	along	already
also	although	always	am	among	amongst	an
and	another	any	anybody	anyhow	anyone	anything
anyway	anyways	anywhere	apart	appear	appreciate	appropriate
ar	are	aren't	around	as	a's	aside
ask	asking	associated	at	available	away	awfully
b	be	became	because	become	becomes	becoming
been	before	beforehand	behind	being	believe	below
beside	besides	best	better	between	beyond	both
brief	brwkj	but	by	c	ca	came
can	cannot	cant	can't	cause	causes	ccf
cdt	certain	certainly	changes	clearly	c'mon	co
com	come	comes	comment	comments	concerning	consequently

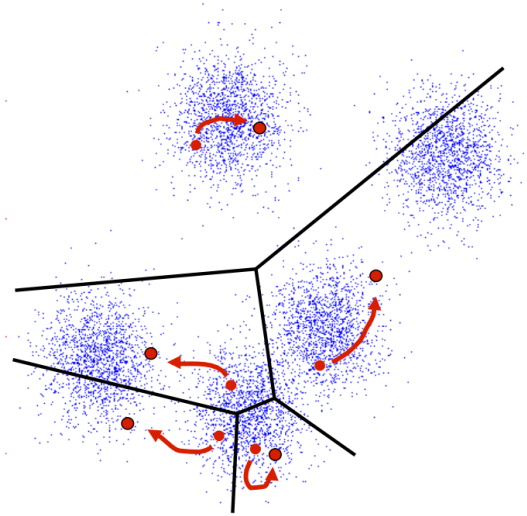
C
ACRONYM TRANSLATION MATRICES

Table C-1
Pump and motor Component Description acronym translation matrix

Acronym	Replacement Text
add	addition
alt	alternate
alum	aluminum
aux	auxiliary
auxilary	auxiliary
afw	auxiliary feedwater
asw	auxiliary service water
bckup	backup
bkup	backup
brg	bearing
bdb	beyond design basis
bd	blowdown
bldn	blowdown
blr	boiler
boilers	boiler
bcw	boiler circulating water
bwst	borated water storage tank
batp	boric acid transfer pump
brkr	breaker
bld	building

Data Analysis of the Project

- Analysis Steps
  - Develop and test the computational architecture and algorithms to be used to perform the data analytics
  - Created K-mode clustering algorithms and applied to an example dataset to establish initial data clustering and to identify data centroids
  - Created an acronym translation matrix and applied to a sample set of the dataset
  - Processing of text data fields and incorporated results into clustering analysis
  - Correlation of text field phrases with actual labor hours and costs



Data Analysis of the Project (Test Dashboard)

- Statistical Analysis of Work Orders
  - Developed K-mode clustering approach to identify similar work orders
  - Performed statistical assessment of clusters to identify trends in material and labor costs
- PM Strategy Comparison
  - Developed approach to identify similar equipment at different sites and utilities
  - Developing the ability to examine the impact of different PM strategies on the overall maintenance costs

Annualized Material Costs and Labor Hours

Component ID	Site	Component Description	PM Hours	CM Hours	All Hours	PM Costs	CM Costs	All Costs
01-FP-P-2-PUMP		diesel driven fire protection pump	19.30	179.00	198.30	38.80	5,436.30	5,475.10
01-FP-P-10-PUMP		warehouse diesel fire pump	15.20	31.90	47.10	0.00	629.90	629.90
M2P82P		fire pump	17.70	5.10	22.90	28.80	0.00	28.80
0FP03PB-PMPA-03PB-P30-<		pump diesel driven fire pump	64.30	47.00	111.30	1,017.80	5,091.80	6,109.50
01-FP-P-1-PUMP		motor driven fire pump	3.80	15.80	19.50	0.50	132.40	132.90

Queried Component ID

PM Strategies for each selected Component ID

PUMP-01 PM Strategy

PMID	PMFREQ	PMDESC	PMHRSAVG	PMCSTAVG
RE500071	364	Pump Packing Inspection and Adjustment (Annual PM)	7.84	0
RE500401	364	SERVICE BATTERY CHARGER	5.96	0
RE500067	728	Oil CNG in Diesel Driven FP (Angle Drive) & lube U-Joint that connects AD to ENG	8.36	77.99

Alternative Pump 02 PM Strategy

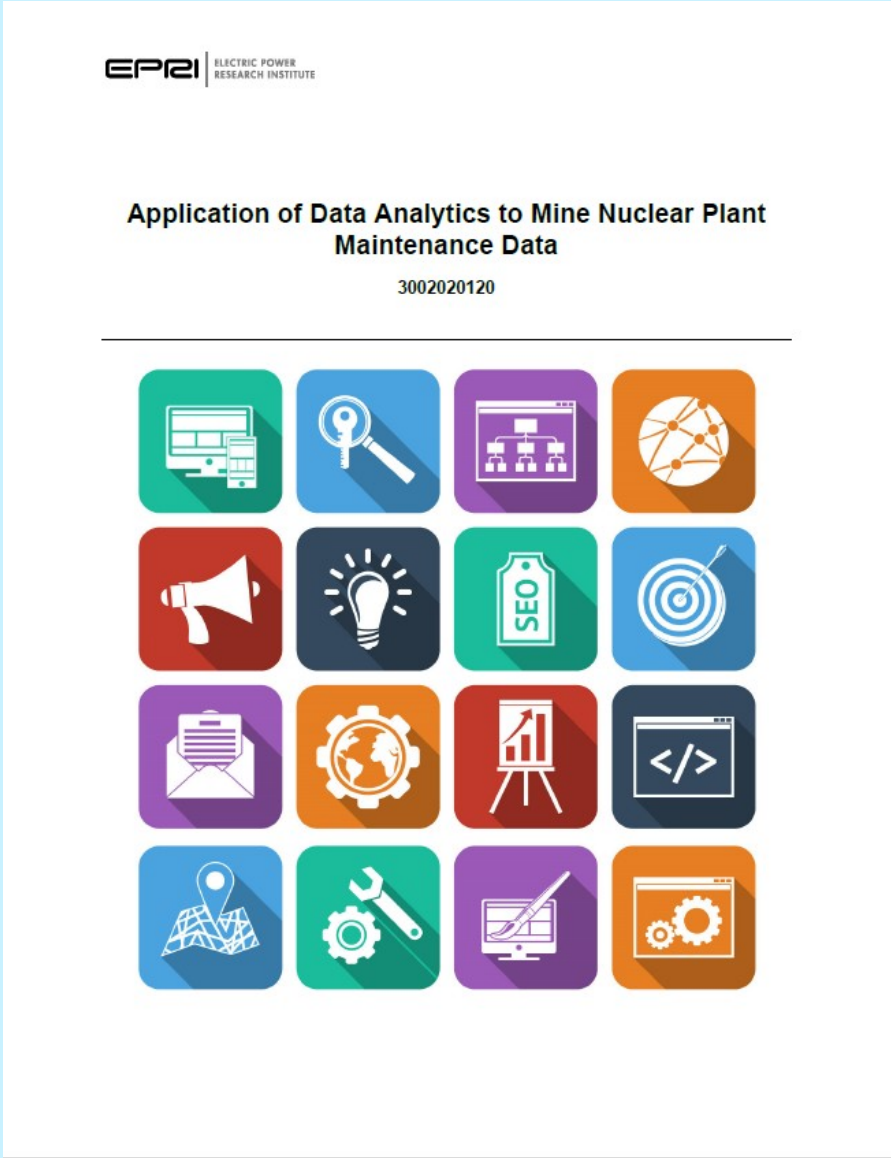
PMID	PMFREQ	PMDESC	PMHRSAVG	PMCSTAVG
RE500178	182	LUBRICATE BEARINGS eWP	3.09	0
RE500070	364	Perform Annual Pump Maintenance - Packing Inspection and Adjustment	8.22	0

Challenges

- Data Quality
  - These records are in a variety of host database software programs
  - There is not a standard set of data fields utilized by all utilities
  - Due to the variation of original plant architect engineers, system and component IDs vary
  - Within the industry there is not a standard set of acronyms
  - High dollar value and negative values for select labor hours and material costs require further text field review for resolution

EPRI Technical Update

- Product 3002020120
- Published March 2021



Conclusions and Next Steps

- Although the results from this project were successful, additional insights could be gained from a broader selection of utility data
- The NLP analysis approach demonstrated that high-quality comparisons of similar component systems/functions from different utilities and sites is possible
- Technologies developed in this project would be of interest to utilities, but additional work would be required to facilitate that direct member access = utility personnel would need to be capable of data analysis techniques
- Pre-processing software would need to be updated in order to apply it to the entire group of utility member datasets
- EPRI would need to establish the extent to which data would be shared amongst utilities due to the sensitive nature (resources & material costs) of the data records