DATA SCIENCE AND ARTIFICIAL INTELLIGENCE REGULATORY APPLICATIONS WORKSHOPS

WORKSHOP 2: CURRENT TOPICS
OPENING REMARKS

Jeremy Groom, Managing Director
EMBARK Venture Studio
Office of Nuclear Reactor Regulation

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

EMBARK VENTURE STUDIO

WELCOME

- Over 200 registered attendees with participants from five different countries.
- Second of Three Workshops with a focus on Current Topics
  - Morning Sessions – NRC Topics and Initiatives
  - Afternoon Sessions – Industry Topics and Initiatives

EPRI ELECTRIC POWER RESEARCH INSTITUTE

DEPARTMENT OF ENERGY · UNITED STATES OF AMERICA ·

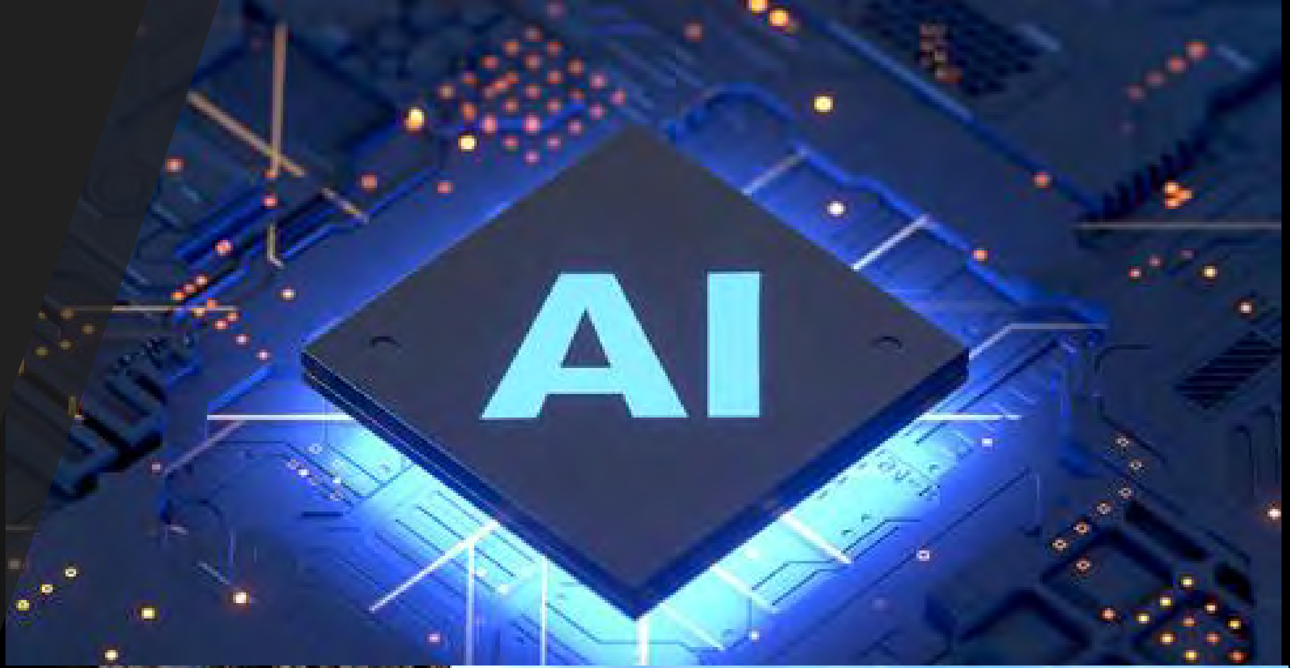
Exelon.

NNS

Regulatory Purpose

- NRC recognizes a need to use data analytics and AI for regulatory enhancements as part of its effort to become a modern, risk-informed regulator[1]

- The nuclear industry is investigating and using AI applications; therefore, the NRC must be prepared to understand and evaluate the technology



[1] "The Dynamic Futures for NRC Mission Areas," (ML19022A178)

Upcoming Workshop #3: Future Focused Initiatives

September/October 2021

https://www.nrc.gov/public-involve/conference-symposia/data-science-ai-reg-workshops.html

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

EMBARK
VENTURE STUDIO

Autonomous Control Algorithms to Simulate Boiling Water Reactor Cycle Depletion

NRC Team:

Nate Hudson, Ph.D. Nathanael.Hudson@nrc.gov

Nazila Tehrani, Ph.D. Nazila.Tehrani @nrc.gov

Peter Yarsky, Ph.D. Peter.Yarsky@nrc.gov

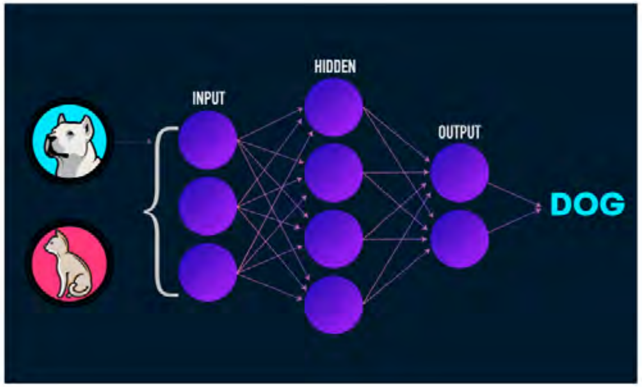
Regulatory Purpose

NRC then

- Reactive
- Plant specific
- Traditional methods

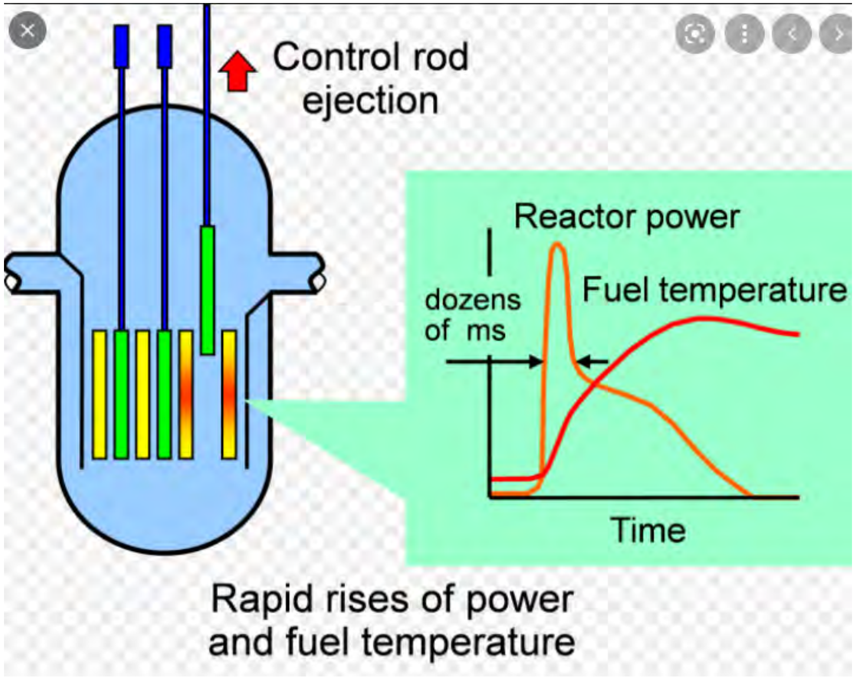
NRC now

- Proactive
- Generic model
- Advanced methods



NRC is exploring innovative tools

- Perform independent transients/accidents analysis
- Increase staff efficiency
- Identify/focus safety significant issues
- Boost confidence in licensee results
  - uncertainty high
  - margin low
- Address emergent issues



Motivation

- Automate LWR core/cycle design
- Create models not associated with any specific licensing action
- Use autonomous control methods

RES develops generic LWR TRACE models for accident/transient analysis

Boiling water reactor

- Peter Yarsky (Project Lead, Senior Reactor Systems Engineer, Code And Reactor Analysis Branch)
- Nate Hudson (Reactor Systems Engineer, Code And Reactor Analysis Branch)
- Nazila Tehrani (Reactor Systems Engineer, Accident Analysis Branch)

Pressurized water reactor

- Andy Bielen (Project Lead, Nuclear Engineer (Fuels/Neutronics), Fuel & Source Term Code Development Branch)
- Mike Rose (Reactor System Engineer (Neutronics Analyst, Fuel & Source Term Code Development Branch)
- Alice Chung (Reactor System Engineer (Fuel Analyst), Fuel & Source Term Code Development Branch)

BWR cores are complex to design

Control excess reactivity during cycle

- burnable poisons
- control blades
- flow control windows

Competing goals

- meeting desired cycle energy
- maintaining safety margins
- minimizing duty related fuel failures

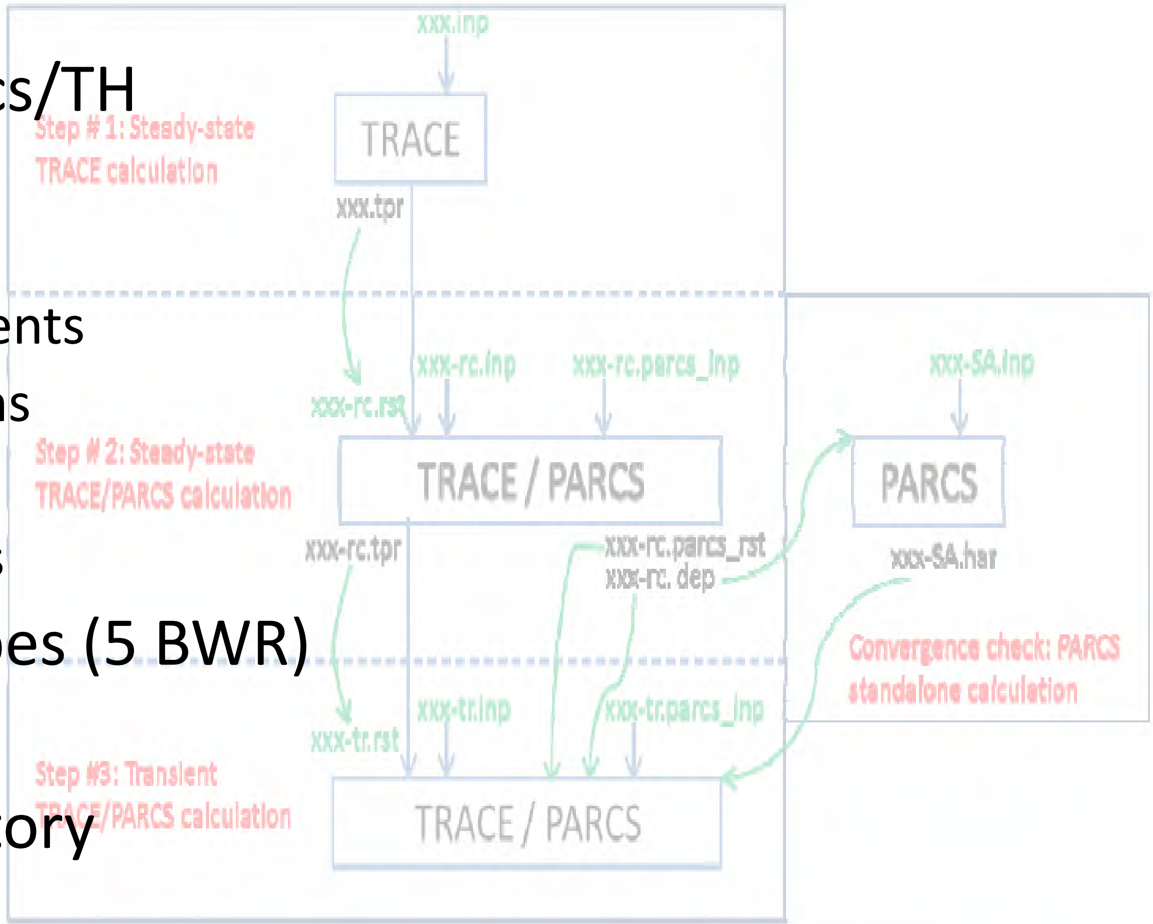
High-dimensionality of BWR cores

- issues with symmetry
- very large number of types of fuel elements

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

TRACE/PARCS models of PWR/BWR transients/accidents analysis

- Coupled neutronics/TH transient
  - normal
  - anticipated transients
  - accident conditions
  - current reactors
  - advanced reactors
- 12 major plant types (5 BWR)
- Standard models
- Efficient confirmatory calculations



Necessity

Traditional NRC

- NRC used PARCS to simulate BWR cycle depletion
- calculate cycle power and burnup distributions
- transient analyses need burnup-dependent
  - rod patterns
  - flow rate
  - EOC bundle shuffle sequence

Innovative NRC

- develop an alternative approach for BWR core designs
- generate a BWR equilibrium cycle
  - theoretical concept
  - operate a "typical" plant
  - given fuel design
  - over a long time

Autonomous control algorithms

Literature review

- Proposed micro-reactor
- sense reactor conditions
- sense reactor coolant system
- judge qualification of signals
- evaluate current state of system
- make decisions about actions
- implement actions for operation

NRC goals

- PARCS models
- dynamically adjust
  - fuel loading between cycles
  - control rod pattern
  - flow rate during cycle
- yield all statepoint information over full cycle

Bayesian networks for dynamic (PRA)

Kim, et al.

Literature Review

Annealing method

Hays and Turinsky

DNN to optimize core loading pattern for BWR

Saleem, et al.

Literature Review

Bayesian networks for dynamic (PRA)

- studies evolution of risk during postulated events
- makes decisions during a transient for reducing risk
- artificial reasoning rely on surrogate models
- NRC wants to create surrogate models

Annealing method

- use core simulator
- use sampling of design choices similar to particles distribution at a temperature
- iteratively lowering temperature, algorithm finds optimal solution
- ~100,000 core simulator runs
- NRC wants to optimize loading patterns

DNNs to optimize core loading pattern for BWR

- DNNs trained against core simulator
- artificial reasoning makes decisions about core loading patterns
- meeting power peaking limits
- cycle energy demand
- reasonable computational expense/accuracy
- used to find optimal designs
- ~10,000 simulations to train
- NRC wants to optimize loading patterns

Work to be Performed

- Autonomous control for BWR core/cycle design feasible**?**
  - Apply combination of existing decision-making methods
    - Bayesian
    - Neural Networks
    - Machine Learning
    - etc.
  - Approximate core loading design/control rod sequence
- Contingency (Traditional Methods)
- Need feedback

Definitions

- BWR: Boiling water reactor
- DNN: Deep neural networks
- EOC: End of cycle
- LWR: Light water reactor
- PARCS: NRC Reactor Kinetics code
- PRA: Probabilistic risk assessment
- PWR: Pressurized water reactor
- RES: Office of Nuclear Regulatory Research
- TH: Thermal-Hydraulics
- TRACE: NRC Thermal-Hydraulics code

Findings of a Literature Survey on Machine Learning for Nondestructive Examination

P. Ramuhalli, H. Sun, D. Womble, R. Jacob*

Oak Ridge National Laboratory

* Pacific Northwest National Laboratory

Second Data Science and AI Regulatory Applications Workshop

Aug 18, 2021

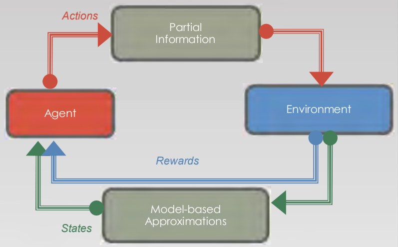
Sponsored by US NRC / Office of Research under IAA 31310020F0038.
Carol Nove, NRC COR

ORNL Strategic Directions in AI/ML

Data

- Facilities operation and control
- Experimental design
- Data curation and validation
- Compressed sensing

Learning

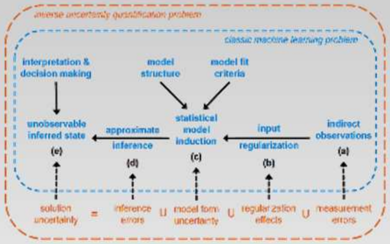


- Physics informed
- Accelerating learning
- Stability and robustness
- Foundations of ML formulations - RL, GANs, GNNs, BNNs
- Dimension reduction and encoding

Scalability

- Algorithms, complexity and convergence
- Levels of parallelization
- Mixed precision arithmetic
- Communication
- Implementations on accelerated-node hardware

Assurance



- Uncertainty quantification
- Robustness
- Explainability and interpretability
- Validation and verification
- Causal inference and hypothesis generation

Workflow

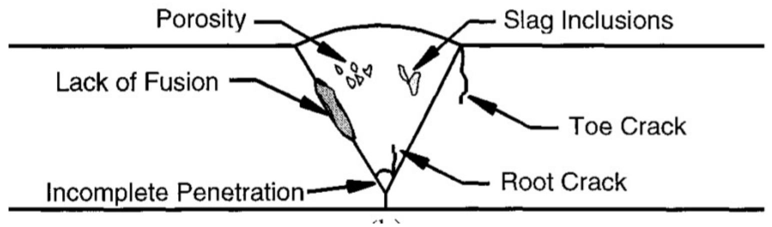
- Edge AI
- Compression
- Online learning
- Federated learning
- Infrastructure
- Augmented intelligence and Human-Computer Interface

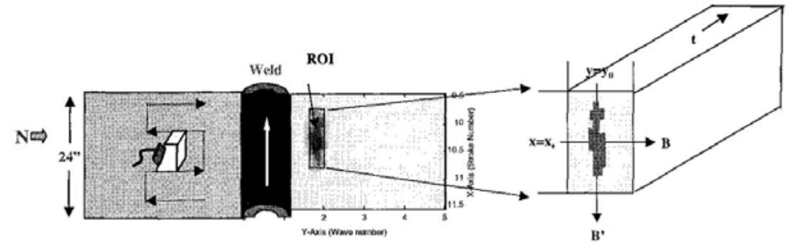
Outline

- Overview
  - Nondestructive examination (NDE)
  - Artificial intelligence (AI)/machine learning (ML)
- Machine learning for nondestructive examination
  - Background
  - Objectives
- Key findings from literature assessment
- Summary and next steps

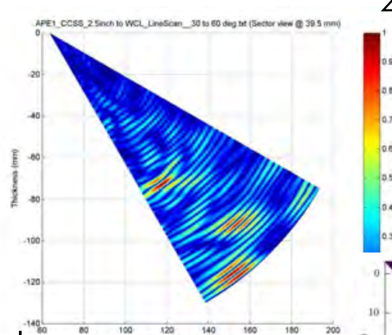
Nondestructive Examination (NDE) in Nuclear Power

- Detect surface or internal anomalies that could compromise the ability of a component to perform its function
  - Examination methods generally classified as volumetric, surface and visual
- Inservice inspection (ISI) of nuclear power plant components required by10CFR50.55a which incorporates by reference Sections III and XI of the ASME Boiler and Pressure Vessel Code
- Analysis of NDE examination data typically performed manually by qualified inspectors
- Increased interest in machine learning (ML) for flaw detection in ASME Code-required inspections
  - Anticipated cost savings, time savings, and expected future shortage of qualified inspectors
  - Potential for future Code activities in application of ML, and licensee submittals

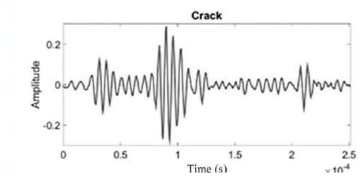




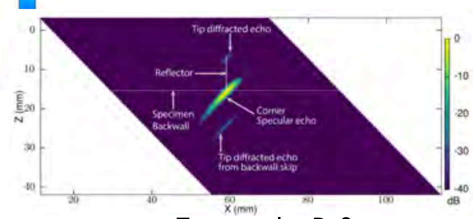
Weld Inspection Example (From J. Kim et al, QNDE 2001)



Example Sector Scan



Example A-Scan



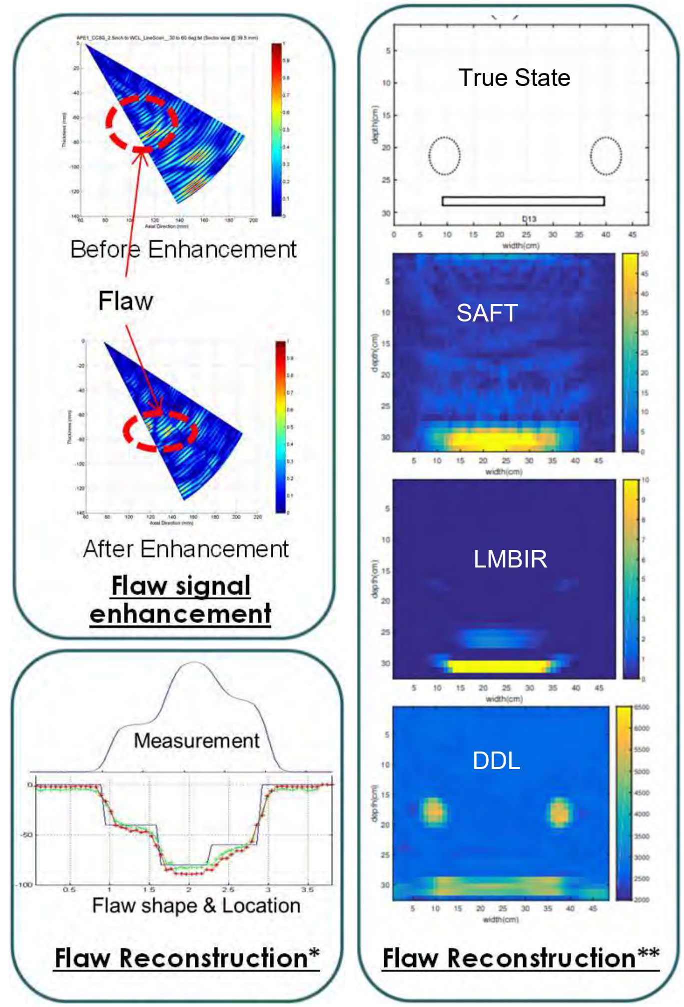
Example B-Scan (From PNNL-26336)

What is the impact of ML on NDE reliability?

Machine Learning for NDE

- Objectives
  - Assess current capabilities of ML and automated data analysis for improving NDE reliability
  - Provide technical basis to support regulatory decisions regarding reviews of relief requests and Code actions that implement automated data analysis for NDE of nuclear power plant components
- Expected outcomes
  - Identify capabilities and limitations of ML for ultrasonic NDE applications
  - Identify factors influencing ML performance and their impact on NDE reliability
  - Recommend verification and validation (V&V) approaches and methods for qualifying ML for nuclear power NDE
  - Identify gaps in existing codes and standards relative to ML for ultrasonic NDE

Examples

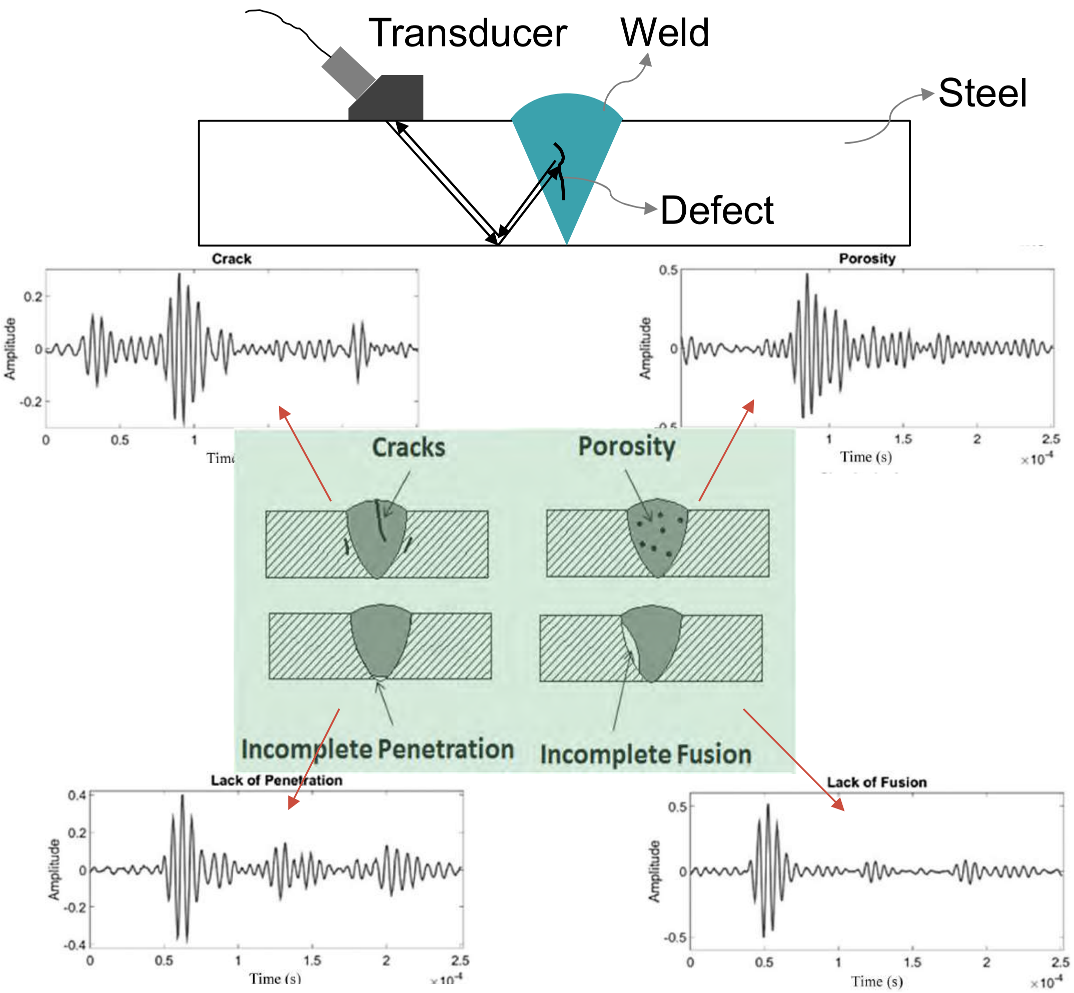


*T. Khan 2011 (PhD Dissertation, Michigan State U.)

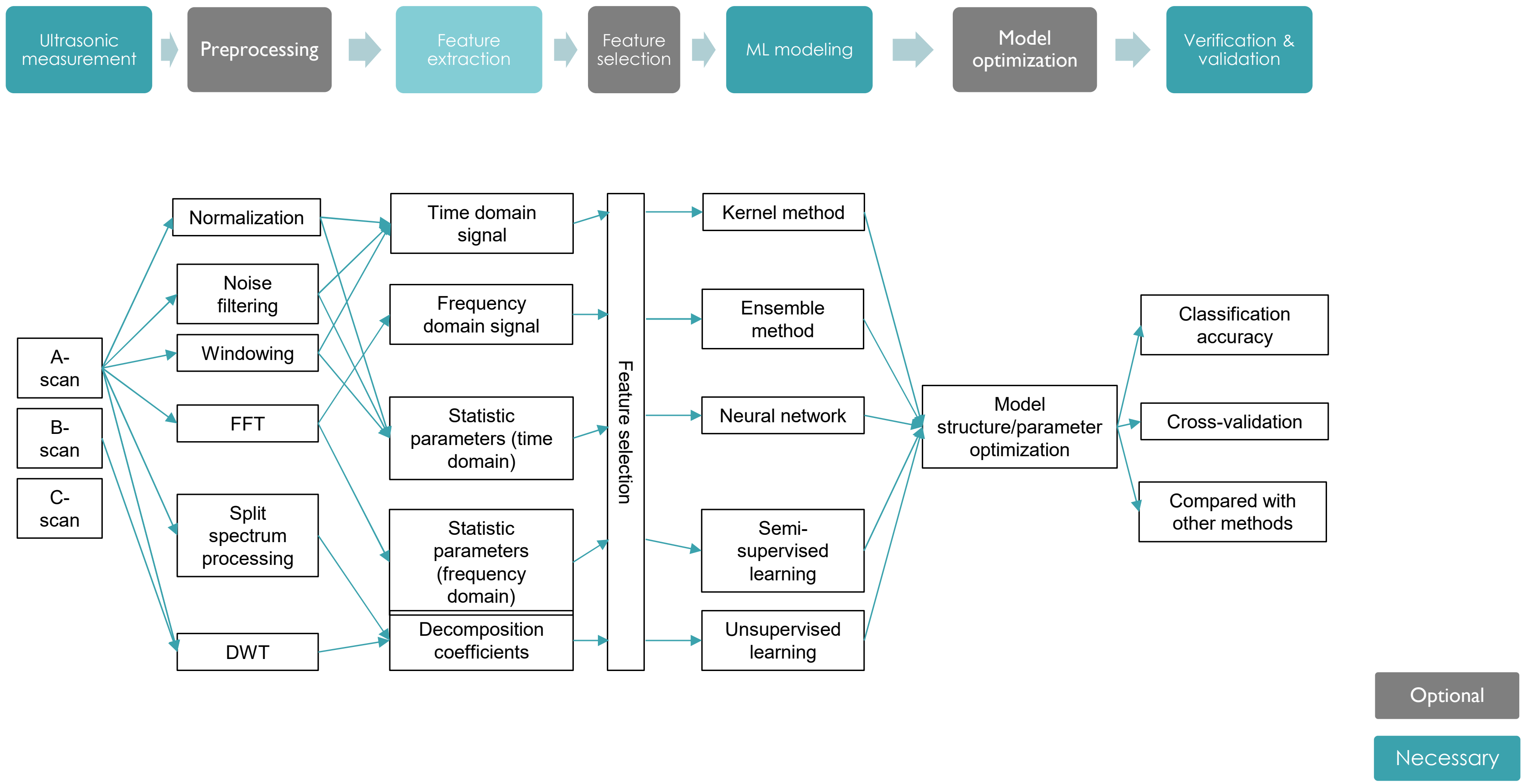
**H. Almansouri, S. V. Venkatakrishnan, G. T. Buzzard, C. A. Bouman and H. Santos-Villalobos, 2018 IEEE GlobalSIP; 10.1109/GlobalSIP.2018.8646704.

Focus: Ultrasonic NDE, Data-driven Learning Algorithms

- Limited to NDE classification problems with data from weld inspections
  - Materials: Steel (carbon, austenitic, cast,…), nickel alloys
  - Flaw types: *thermal fatigue, stress corrosion cracking, weld fabrication flaws*
  - Inspection setup assumed to be appropriate for weld inspections
- Approach: **Literature review** followed by empirical studies
  - Literature set identified is large but not exhaustive

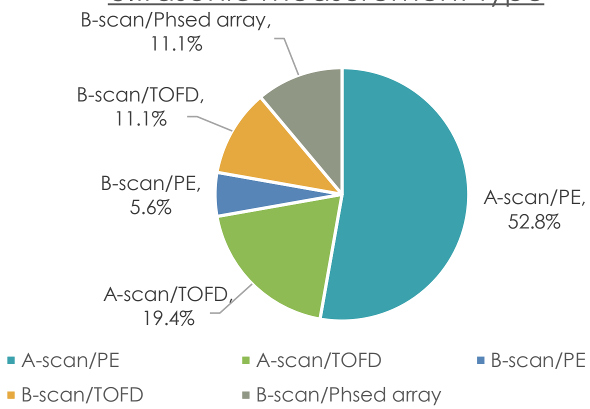


Data Flow in ML for Ultrasonic NDE

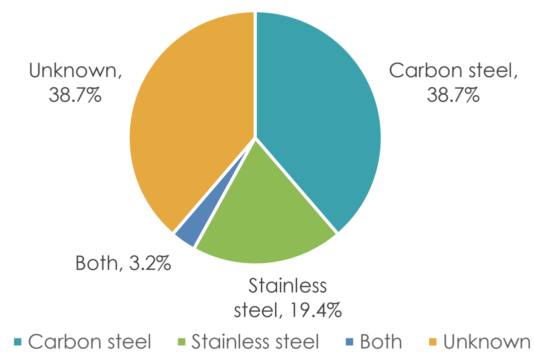


Summary of Literature Data

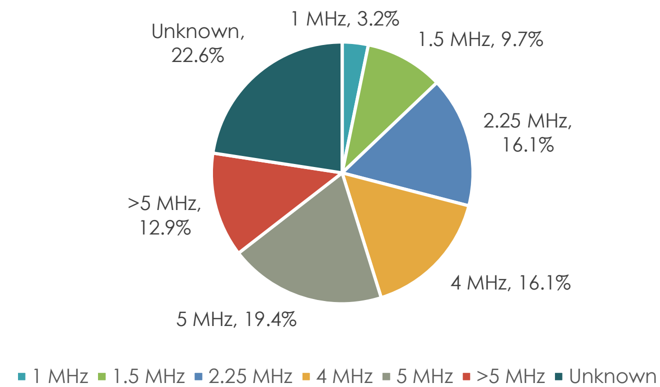
Ultrasonic measurement type



Specimen materials

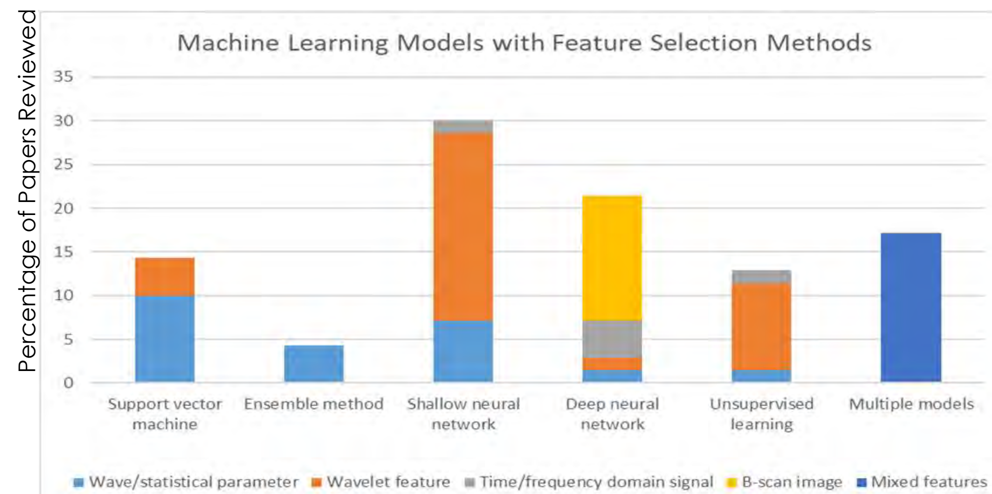


Ultrasonic NDE: Frequencies



Total instances	Physical specimens	Number of crack flaws	Number of non-crack flaws	RT verification?	Ref.
50	Steel specimen	15	35	N/A	[19]
273	Simulated flaws	73	100	N/A	[5]
240	10 steel specimens	N/A	N/A	Yes	[28]
282	1 steel specimen	N/A	N/A	N/A	[29]
100	100 steel specimens	0	100	N/A	[30]
61	Bearing steel samples	N/A	N/A	N/A	[15]
438	438 specimens	0	217	N/A	[14]
239	Steel specimen	N/A	N/A	N/A	[43]
135	135 specimens	45	90	Yes	[7]
246	EPRI database	N/A	N/A	N/A	[16]
293	EPRI database	N/A	N/A	N/A	[16]
90	90 steel specimens	25	44	N/A	[31]
120	6 aluminum specimens	N/A	N/A	N/A	[6]
240	12 steel specimens	N/A	N/A	N/A	[8]
200	12 steel specimens	N/A	N/A	N/A	[8]

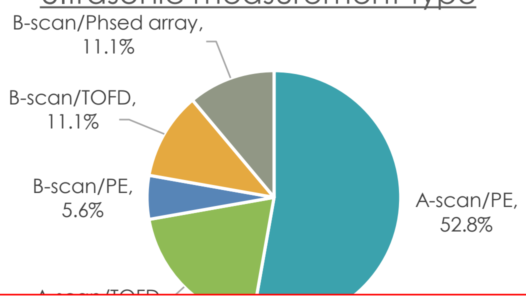
Examples of Data Distribution



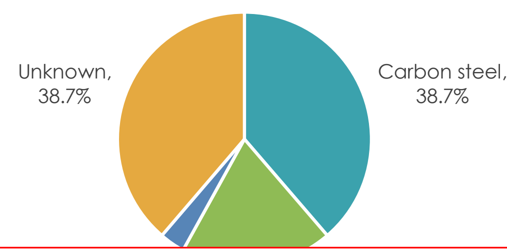
ML Models and Feature Selection

Summary of Literature Data

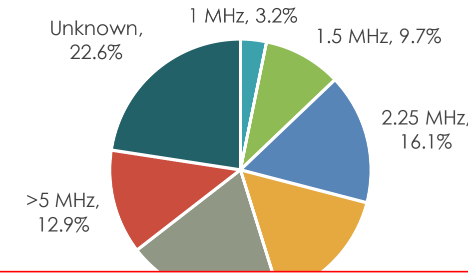
Ultrasonic measurement type



Specimen materials



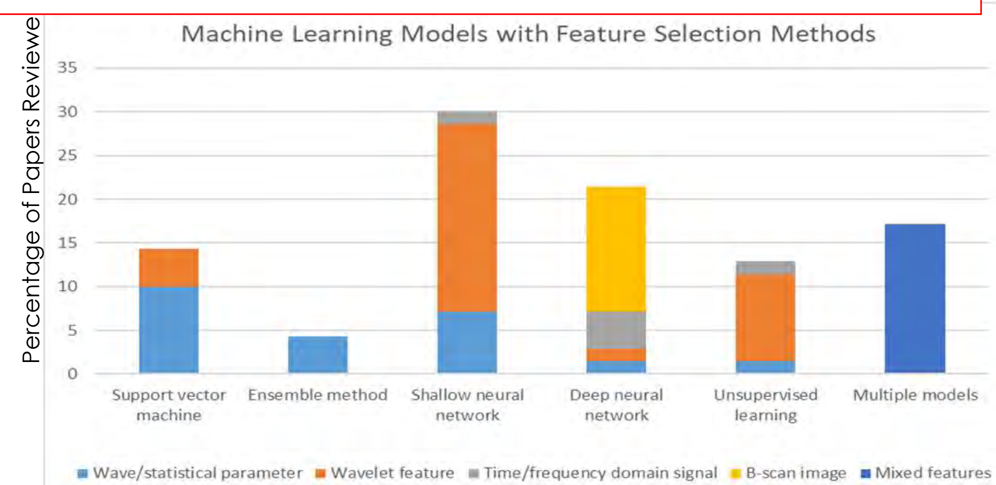
Ultrasonic NDE: Frequencies



Lack of common data sets and diversity in methods/data sets challenge direct comparisons, though general insights into capabilities possible.

Total instances	Physical specimens	Number of crack flaws	Number of non-crack flaws	RT verification?	Ref.
50	Steel specimen	15	35	N/A	[19]
273	Simulated flaws	73	100	N/A	[5]
240	10 steel specimens	N/A	N/A	Yes	[28]
282	1 steel specimen	N/A	N/A	N/A	[29]
100	100 steel specimens	0	100	N/A	[30]
61	Bearing steel samples	N/A	N/A	N/A	[15]
438	438 specimens	0	217	N/A	[14]
239	Steel specimen	N/A	N/A	N/A	[43]
135	135 specimens	45	90	Yes	[7]
246	EPRI database	N/A	N/A	N/A	[16]
293	EPRI database	N/A	N/A	N/A	[16]
90	90 steel specimens	25	44	N/A	[31]
120	6 aluminum specimens	N/A	N/A	N/A	[6]
240	12 steel specimens	N/A	N/A	N/A	[8]
200	12 steel specimens	N/A	N/A	N/A	[8]

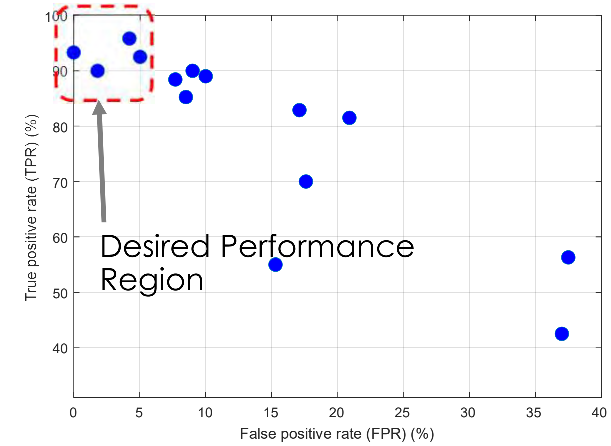
Examples of Data Distribution



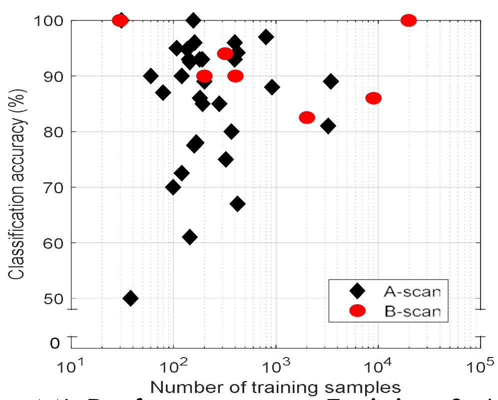
ML Models and Feature Selection

Examples of Reported ML Performance in the Literature

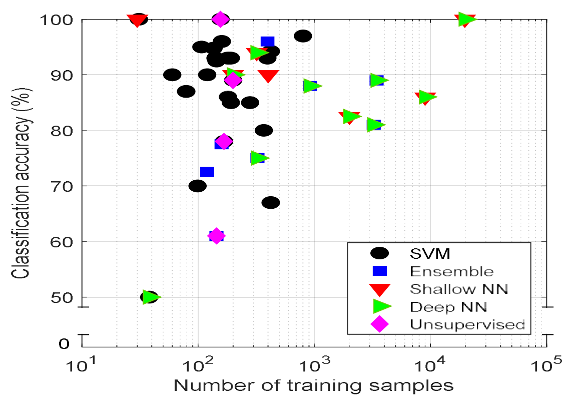
What factors influence the performance of machine learning (ML) and automated data analysis techniques when applied to NDE data?



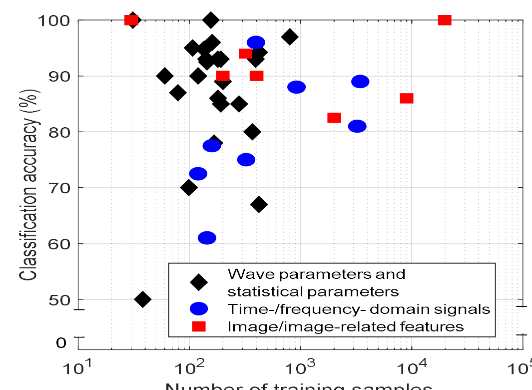
True Positive Rate (TPR) vs False Positive Rate (FPR)



ML Performance vs Training Set Size, Sorted by Type of Data



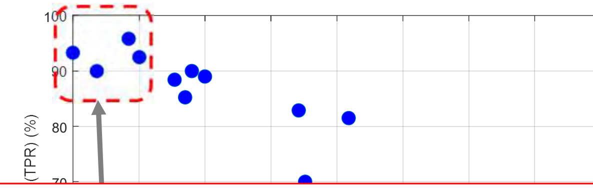
ML Performance vs Training Set Size, Sorted by ML Method



ML Performance vs Training Set Size, Sorted by Feature Type

Examples of Reported ML Performance in the Literature

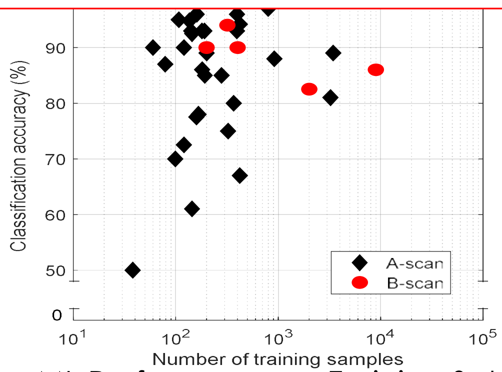
What factors influence the performance of machine learning (ML) and automated data analysis techniques when applied to NDE data?



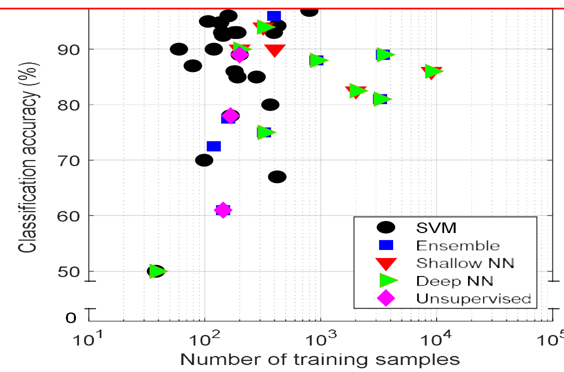
High classification accuracy (high true positive rate and low false positive/negative rate) is possible with ML applied to ultrasonic NDE data

Most ML methods are likely to be capable of good classification performance, with performance depending on the data used for model training and model parameter tuning

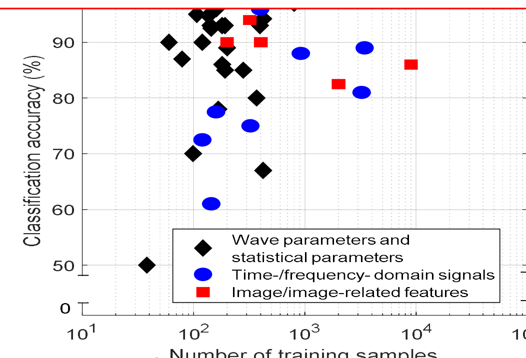
There may be a need for common data sets to compare across methods/solution providers



ML Performance vs Training Set Size, Sorted by Type of Data



ML Performance vs Training Set Size, Sorted by ML Method

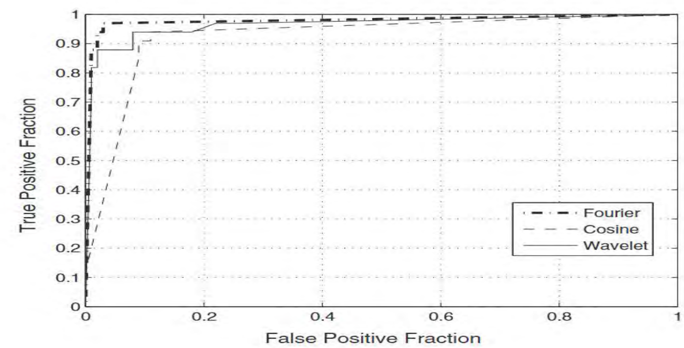


ML Performance vs Training Set Size, Sorted by Feature Type

ML and Ultrasonic NDE Reliability

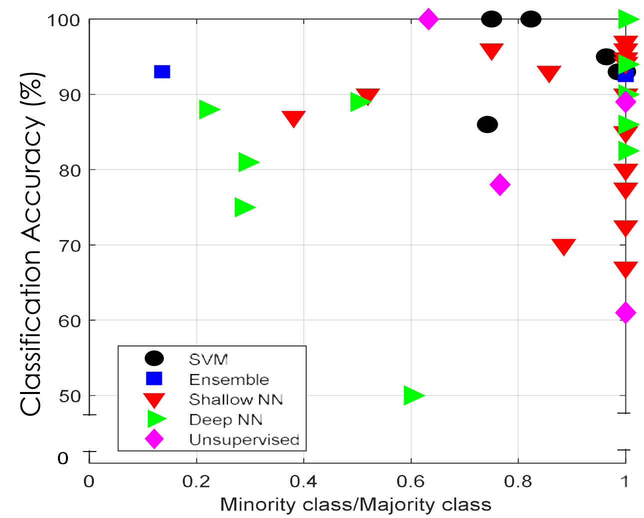
- Limited information in literature on:
  - Sensitivity of classification performance to various factors
  - Demonstrating confidence in generalization performance
  - Quantifying impact of ML on ultrasonic NDE probability of detection (POD)
- Methods for demonstrating confidence in ML performance being studied in other applications and as part of Standards development activities

*Literature on ML application to other NDE techniques also shows promise though not all studies address the above factors



ROC Curve Comparison for Defect/No Defect Classification

(From Cruz et al, Ultrasonics 73, pp 1-8 (2017))



Classification accuracy vs Data Imbalance

Open slide master to edit

A Need for Representative, Common, Public NDE Data Sets

- Sample size and representativeness seem to be a limiting condition in most ML for ultrasonic NDE studies
  - Data augmentation approaches have been applied in some studies to mitigate sample size concerns
  - Unclear whether data augmentation helps with generalization performance
- Representative, common data sets
  - Enable comparison between methods
  - Support V&V approaches to demonstrate impact of ML on NDE reliability
  - Enable reproducibility of ML research results

Robustness of ML Solution

- Sensitivity studies relative to model parameters are likely to be important to improving confidence in the reported results
  - Impact of noise in the data on the results is part of the assessment
  - Model tuning should be a standard part of the methodology for developing ML solutions for NDE
- Effective V&V approaches to quantify confidence in ML solution necessary
- Robustness assessment/V&V of ML will need information on software tools and development environment
  - Enables assessment of potential limitations with tools
  - Increases reproducibility of results
  - Simplifies maintainability of code-base

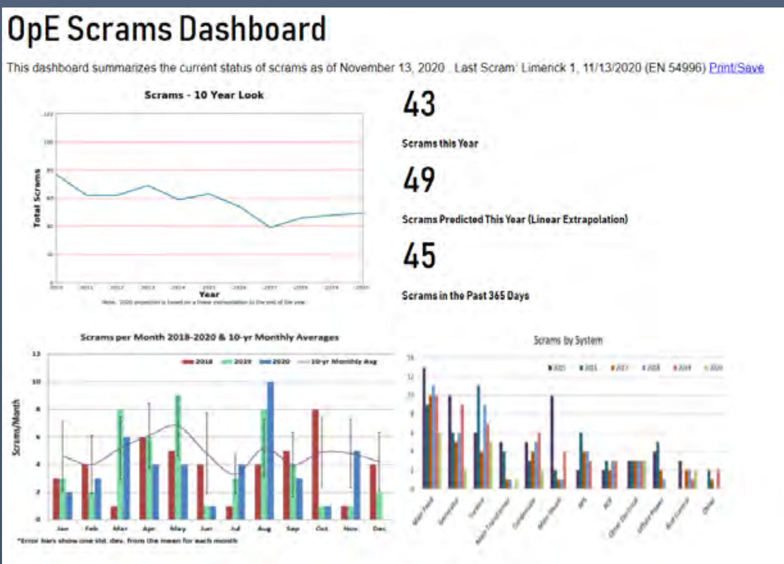
Summary and Future Plans

- Assessment of literature demonstrates the potential of ML for automating ultrasonic NDE data analysis
  - Literature survey to assess the state of art in ML for ultrasonic NDE being finalized for publication
- Literature review identified several open questions related to the impact of factors that influence ML performance, and the contribution of ML to increasing NDE reliability
- Recommendations being formulated for addressing these questions and developing the technical bases to support regulatory decisions regarding reviews of relief requests and Code actions that that include ML
- Future plans: compilation of reference data sets and empirical studies to address open questions from literature review

Questions?

NRC Operating Experience Artificial Intelligence Workshop

An Overview

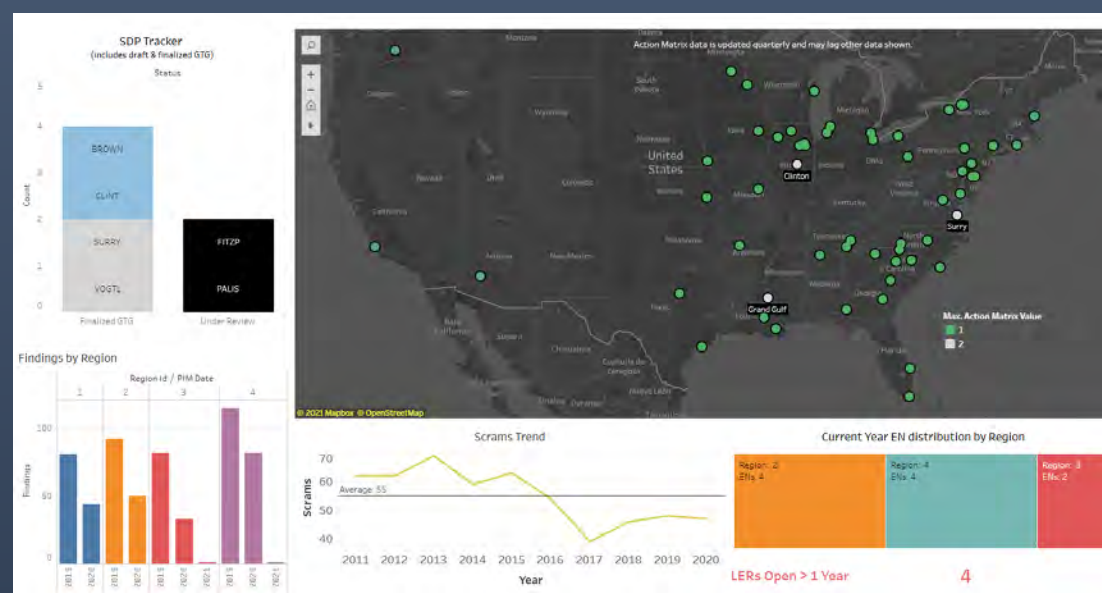


Team

Jason Carneal
Chris Speer
Julie Winslow
Rebecca Sigmon
Lisa Regner

Focus

- Machine learning / natural language processing applications for operating experience
- Progress to date
- Impacts to the reactor oversight process



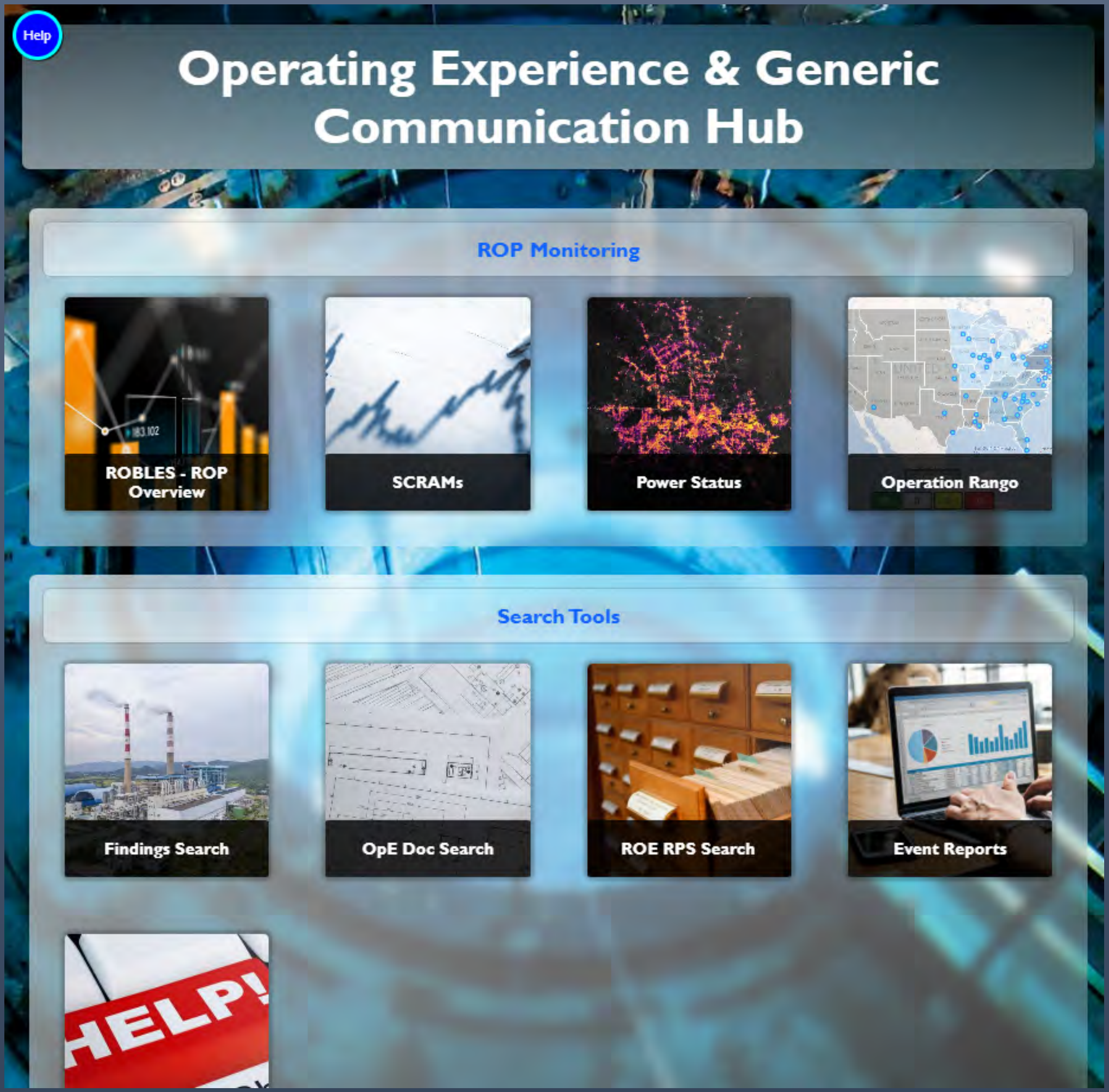
OpE Artificial Intelligence Projects

The operating experience branch is developing machine learning / natural language processing algorithms to make our processes more efficient.

Automation of operating experience processing

Extending existing search tools to allow association of reports to inspection procedure, system, and available risk information

OpE Hub – Deployed Products



Consolidation of Deployed Products

- Website portal for NRC users
- One stop shop for all OpE products
- Easy to navigate
- Facilitates user interaction and support

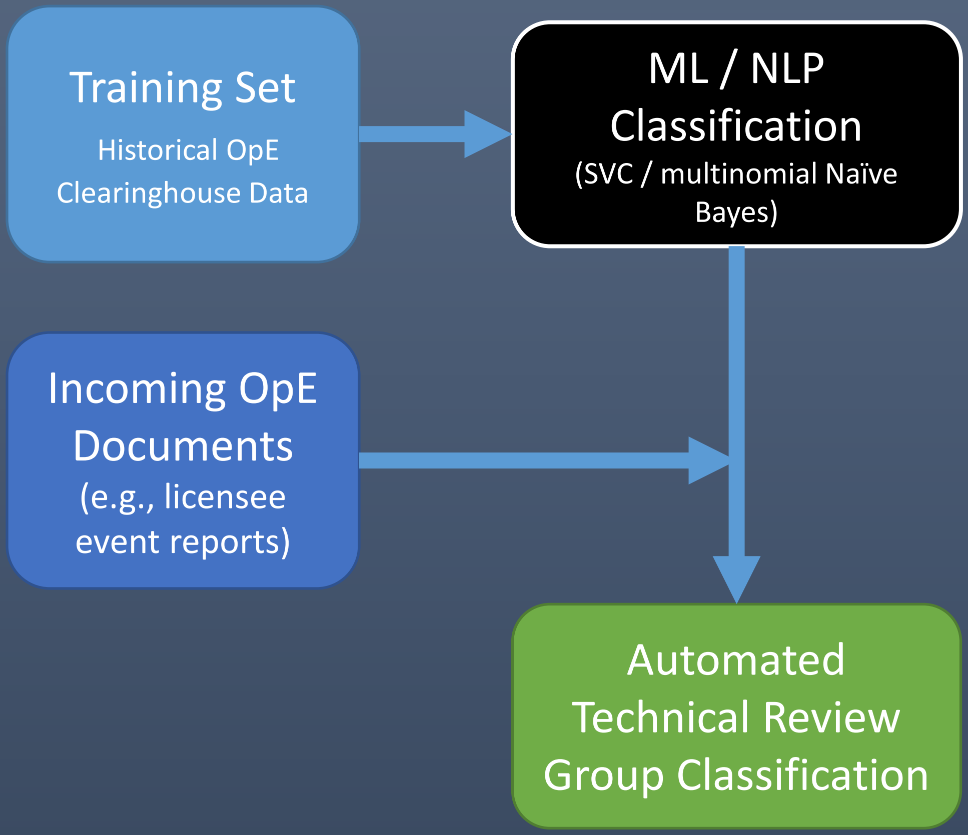
ROP Machine Learning Case 1
Classification of Operating Experience Documents by Technical Review Group

Objectives

- Build a classifier that can sort incoming OpE documents by technical review group
- Automate certain aspects of operating experience workflow

Progress

- Working classifier for event notifications (ranges from 60%-90% accuracy depending on technical area)
- Exploring extension to licensee event reports



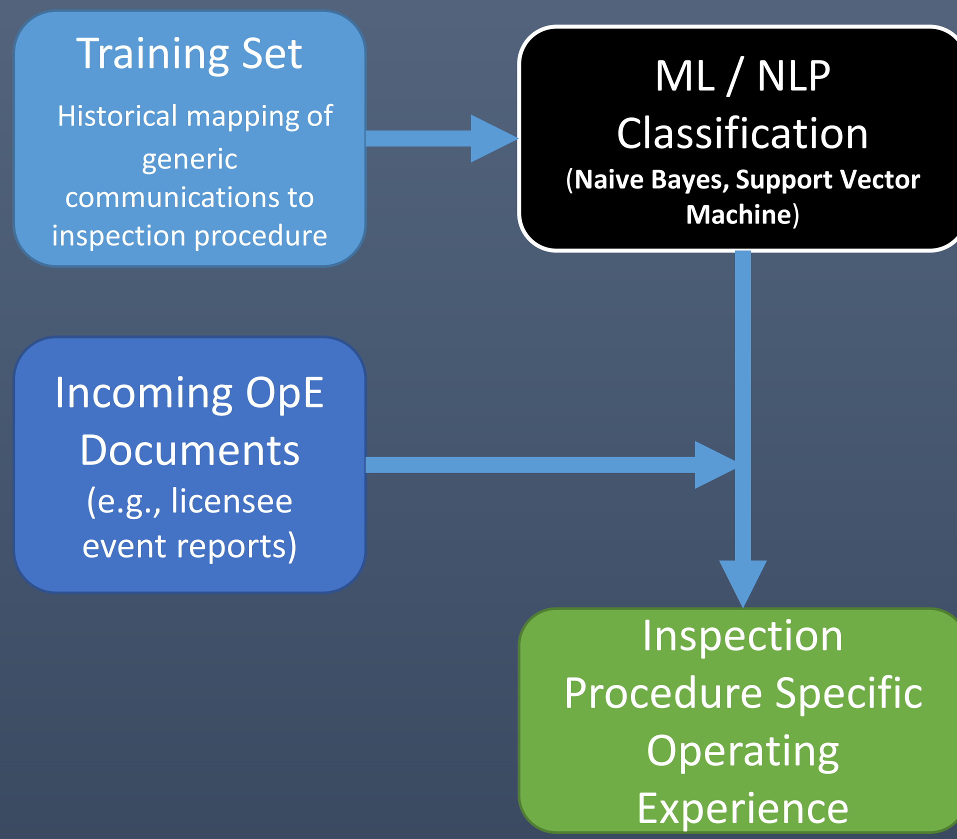
ROP Machine Learning Case 2
Classifying Operating Experience Documents by Inspection Procedure

Objectives

- Build a classifier that maps OpE documents to applicable inspection procedures
- Expand to include different forms of classification (equipment types, failure modes, etc.)
- Deploy advanced search capability

Progress

- Generic Communication classifier built (ranges from 65-70% accuracy)
- Exploring extension to licensee event reports and other operating experience documents



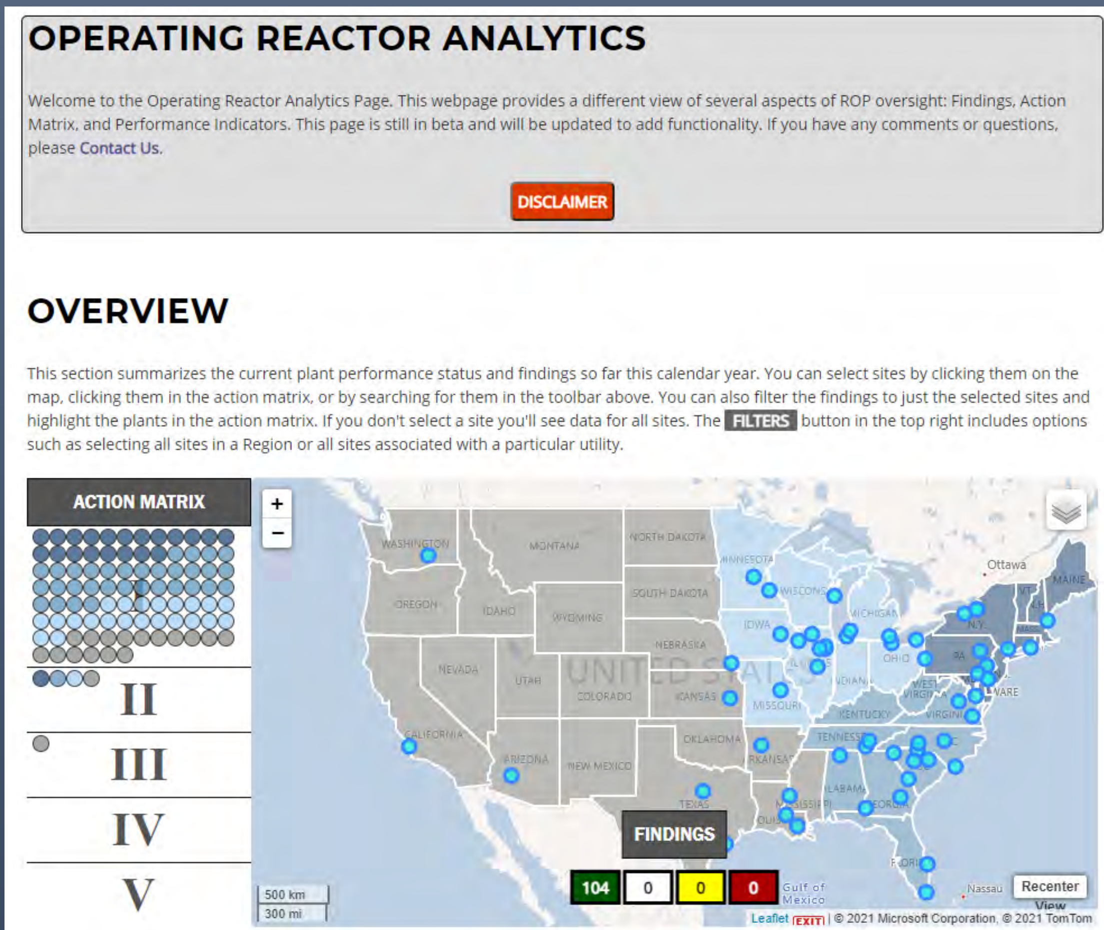
Offsetting redundant tasks

- Eliminating repetitive manual reports on popular topics
- Tools for inspectors / NRC staff / management to review data of interest
- Lowering bar of access to data for both internal and external users

Improving data-driven decision-making

- Consolidating and democratizing access to sparse and difficult to access data
- Deploying tools that allow users to explore data on their own
- Sharing insights previously difficult to ascertain

Operating Reactor Analytics Public Site



Public Site for Reactor Oversight Process Information

- Action Matrix
- Performance Indicators
- Findings

Lead Contact:
Reed Anzalone
NRR/EMBARK

https://www.nrc.gov/reactors/operating/oversight/analytics.html

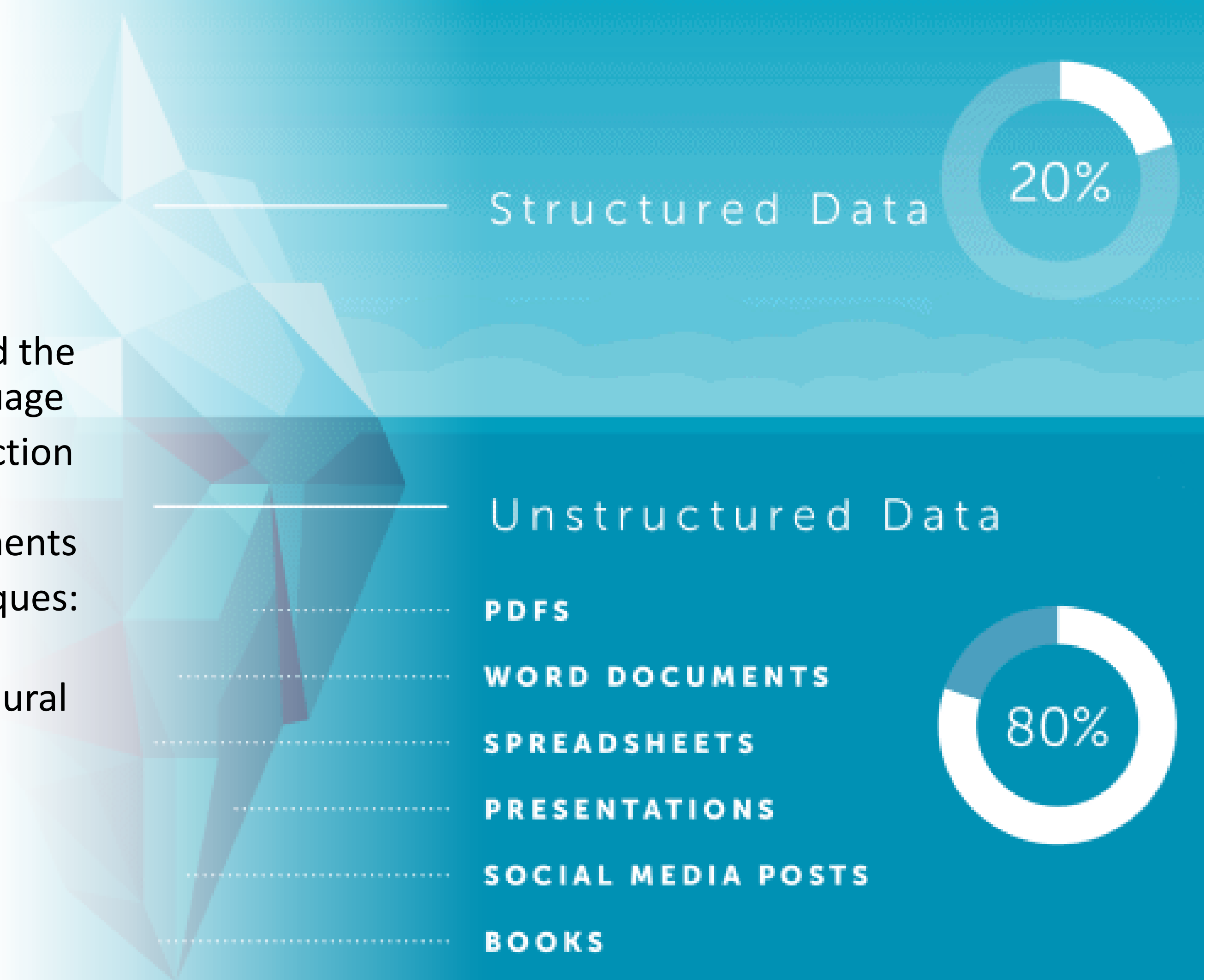
Resource Prediction Using Natural Language Processing

Trey Hathaway
U.S. Nuclear Regulatory Commission
RES/DSA/AAB
August 18, 2021

NRC Data Science and Artificial Intelligence Regulatory Applications Workshops: Current Topics

Natural Language Processing

- Techniques that allow computers to understand the contents of natural language
  - Allows for the extraction of information and insights from documents
  - Collection of techniques:
    - Rule-based, statistical, or neural



U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Use Cases Goals

- Apply Natural Language Processing techniques to NRC data and use cases
- Demonstrate Successes

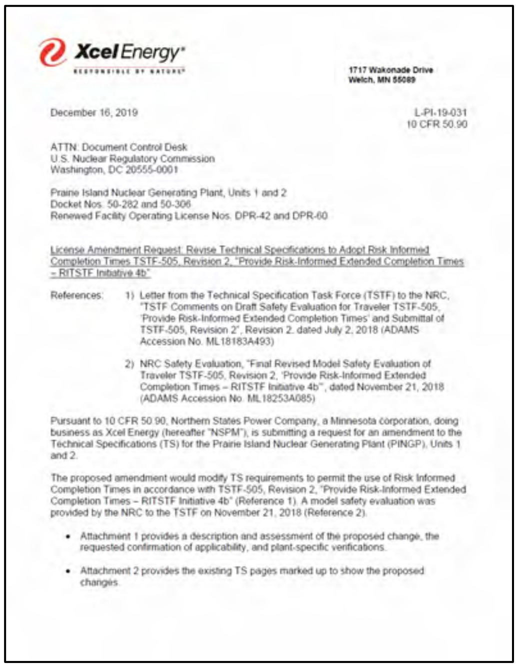
Resource Prediction

- ***Challenge:*** Deviations between resource estimates to complete a licensing review and the actual hours charged
- ***Goal:*** Create tool to assist project managers in formulating resource estimates
  - Leverage historical data
  - Find historically similar reviews
- ***Method:*** Use term frequency-inverse document frequency vectors to represent documents and perform similarity calculations
  - Rank documents based on similarity

Resource Prediction

- **Term Frequency-Inverse Document Frequency (tf-idf)**
  - Weighting factor for words
  - Product term frequency and inverse document frequency
- **Term Frequency (tf)**
  - How frequency a word appears in a document
  - Importance of word
- **Inverse document Frequency (idf)**
  - How frequently a word appears in a collection of documents

Term Frequency-Inverse Document Frequency
(Vector Representation)



Xcel Energy®
RESPONSIBLE BY NATURE®

1717 Wakonade Drive
Welch, MN 55089

December 16, 2019

L-PI-19-031
10 CFR 50.90

ATTN: Document Control Desk
U.S. Nuclear Regulatory Commission
Washington, DC 20555-0001

Prairie Island Nuclear Generating Plant, Units 1 and 2
Docket Nos. 50-282 and 50-306
Renewed Facility Operating License Nos. DPR-42 and DPR-60

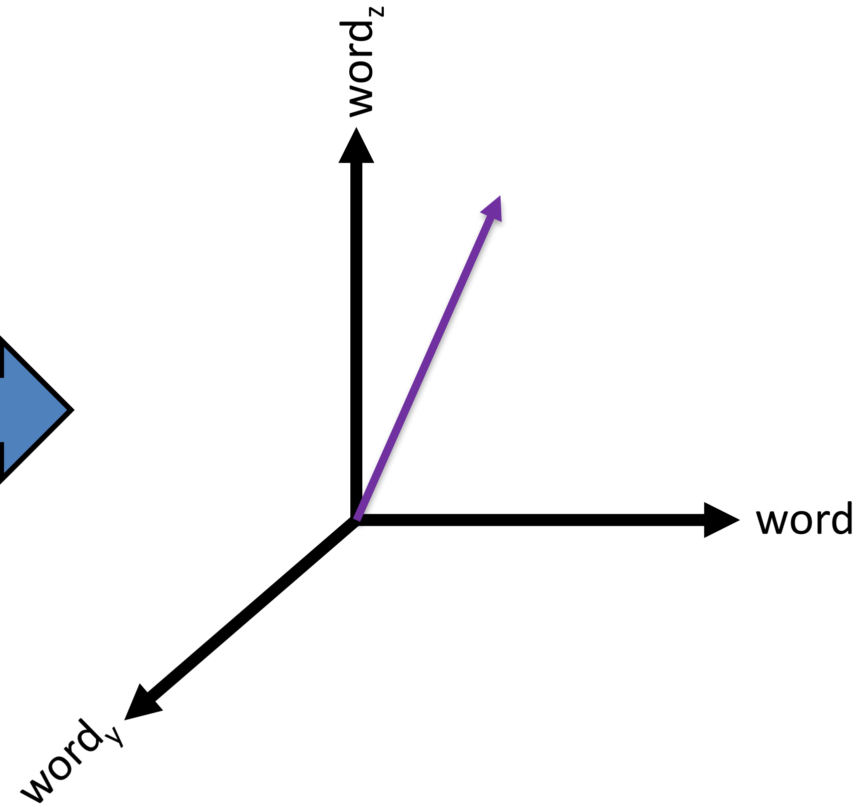
License Amendment Request: Revise Technical Specifications to Adopt Risk Informed Completion Times TSTF-505, Revision 2, "Provide Risk-Informed Extended Completion Times – RITSTF Initiative 4b"

References:
1) Letter from the Technical Specification Task Force (TSTF) to the NRC, "TSTF Comments on Draft Safety Evaluation for Traveler TSTF-505, 'Provide Risk-Informed Extended Completion Times' and Submittal of TSTF-505, Revision 2", Revision 2, dated July 2, 2018 (ADAMS Accession No. ML18183A493)
2) NRC Safety Evaluation, "Final Revised Model Safety Evaluation of Traveler TSTF-505, Revision 2, 'Provide Risk-Informed Extended Completion Times – RITSTF Initiative 4b'", dated November 21, 2018 (ADAMS Accession No. ML18253A085)

Pursuant to 10 CFR 50.90, Northern States Power Company, a Minnesota corporation, doing business as Xcel Energy (hereafter "NSPM"), is submitting a request for an amendment to the Technical Specifications (TS) for the Prairie Island Nuclear Generating Plant (PINGP), Units 1 and 2.

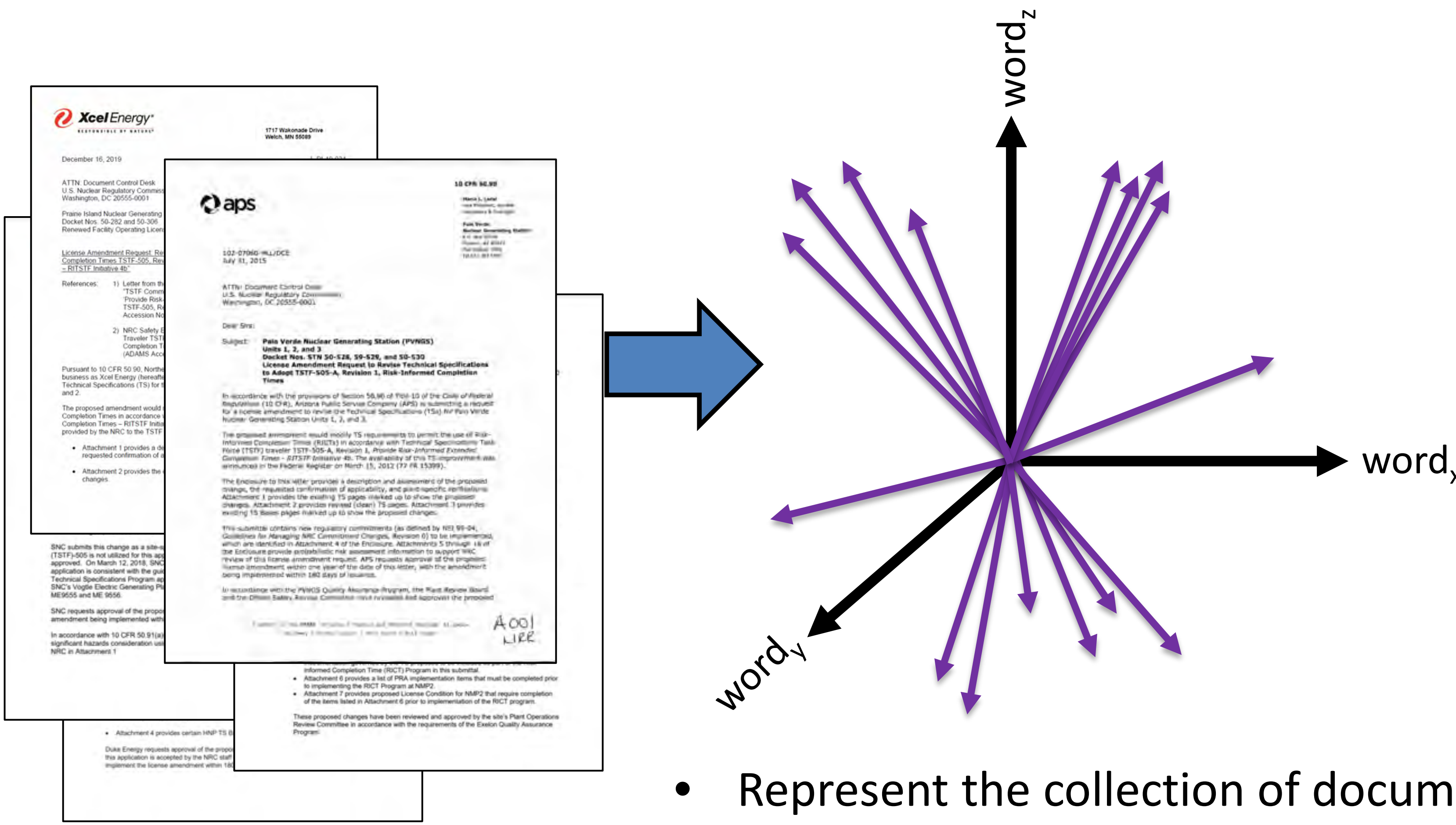
The proposed amendment would modify TS requirements to permit the use of Risk Informed Completion Times in accordance with TSTF-505, Revision 2, "Provide Risk-Informed Extended Completion Times – RITSTF Initiative 4b" (Reference 1). A model safety evaluation was provided by the NRC to the TSTF on November 21, 2018 (Reference 2).

- Attachment 1 provides a description and assessment of the proposed change, the requested confirmation of applicability, and plant-specific verifications.
- Attachment 2 provides the existing TS pages marked up to show the proposed changes.



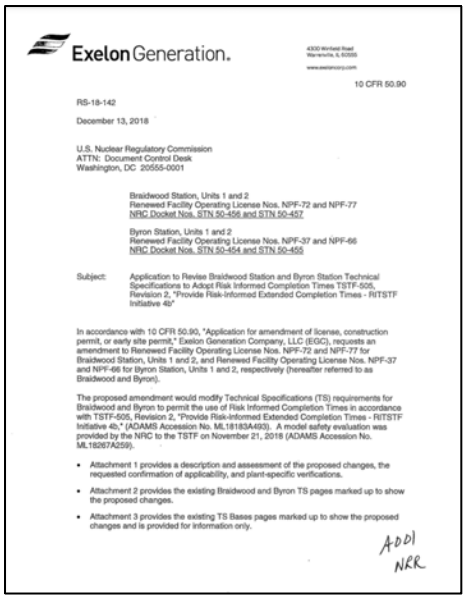
- Represent a document as a vector
  - The vector reflects the word usage in the document
  - The vector will have 1000's of dimensions

Term Frequency-Inverse Document Frequency
(Vector Space Corpus)



- Represent the collection of documents as vectors
  - Create a vocabulary of all words used in the collection

Term Frequency-Inverse Document Frequency
(Similarity Calculations)



Exelon Generation.

10 CFR 50.90

RS-18-142

December 13, 2018

U.S. Nuclear Regulatory Commission
ATTN: Document Control Desk
Washington, DC 20555-0001

Braidwood Station, Units 1 and 2
Renewed Facility Operating License Nos. NPF-72 and NPF-77
NRC Docket Nos. STN 50-456 and STN 50-457

Byron Station, Units 1 and 2
Renewed Facility Operating License Nos. NPF-37 and NPF-66
NRC Docket Nos. STN 50-454 and STN 50-455

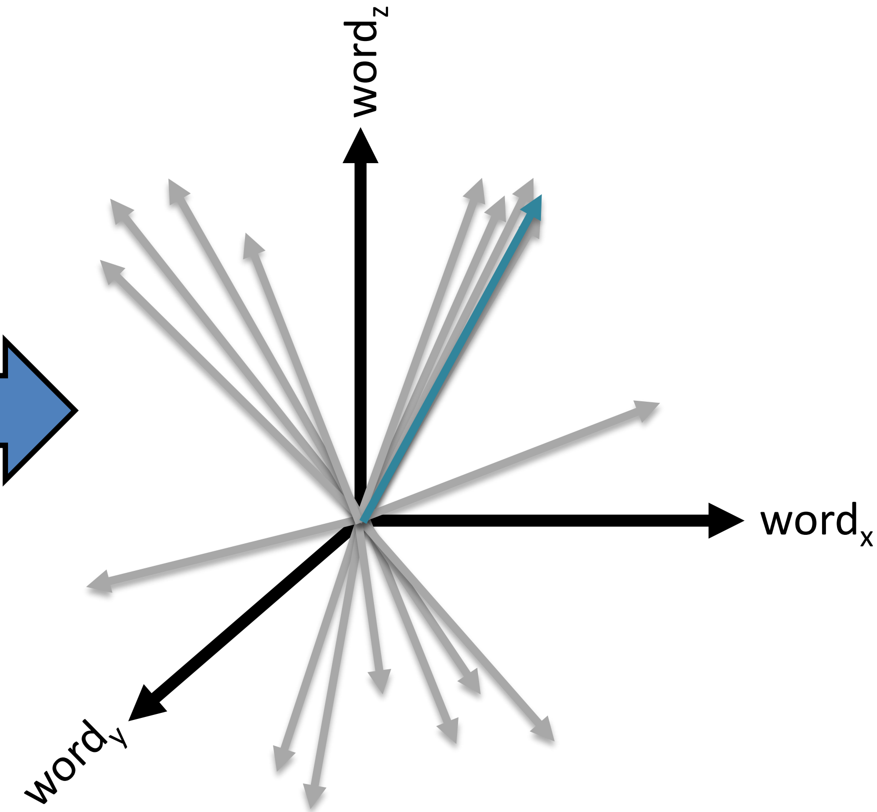
Subject: Application to Revise Braidwood Station and Byron Station Technical Specifications to Adopt Risk Informed Completion Times TSTF-505, Revision 2, "Provide Risk-Informed Extended Completion Times - RITSTF Initiative 4b"

In accordance with 10 CFR 50.90, "Application for amendment of license, construction permit, or early site permit," Exelon Generation Company, LLC (EGC), requests an amendment to Renewed Facility Operating License Nos. NPF-72 and NPF-77 for Braidwood Station, Units 1 and 2, and Renewed Facility Operating License Nos. NPF-37 and NPF-66 for Byron Station, Units 1 and 2, respectively (hereafter referred to as Braidwood and Byron).

The proposed amendment would modify Technical Specifications (TS) requirements for Braidwood and Byron to permit the use of Risk Informed Completion Times in accordance with TSTF-505, Revision 2, "Provide Risk-Informed Extended Completion Times - RITSTF Initiative 4b," (ADAMS Accession No. ML18183A493). A model safety evaluation was provided by the NRC to the TSTF on November 21, 2018 (ADAMS Accession No. ML18267A259).

- Attachment 1 provides a description and assessment of the proposed changes, the requested confirmation of applicability, and plant-specific verifications.
- Attachment 2 provides the existing Braidwood and Byron TS pages marked up to show the proposed changes.
- Attachment 3 provides the existing TS Bases pages marked up to show the proposed changes and is provided for information only.

ADDI
NRR



- A new document is converted to a vector based on the vocabulary of the collection of documents
  - The similarity (angle between vectors) is calculated as the dot product between vectors
  - Documents ranked by similarity score

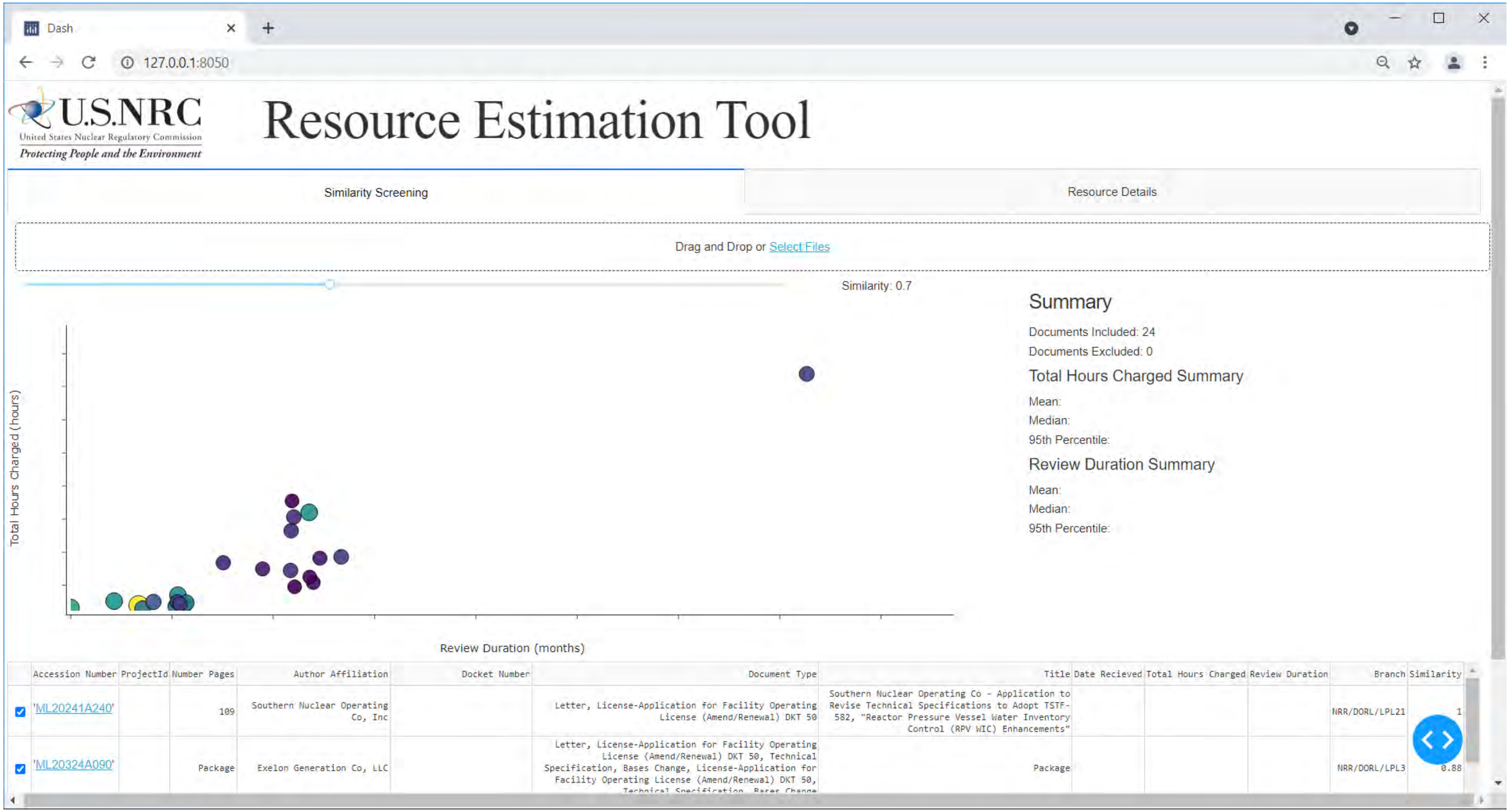
Resource Prediction

Approach

- Acquire historical licensing actions and resource requirements
- Extract text data from pdf files
- Clean data
- Create tf-idf matrix
- Create User Interface
  - Extracts text data
  - Performs similarity calculations

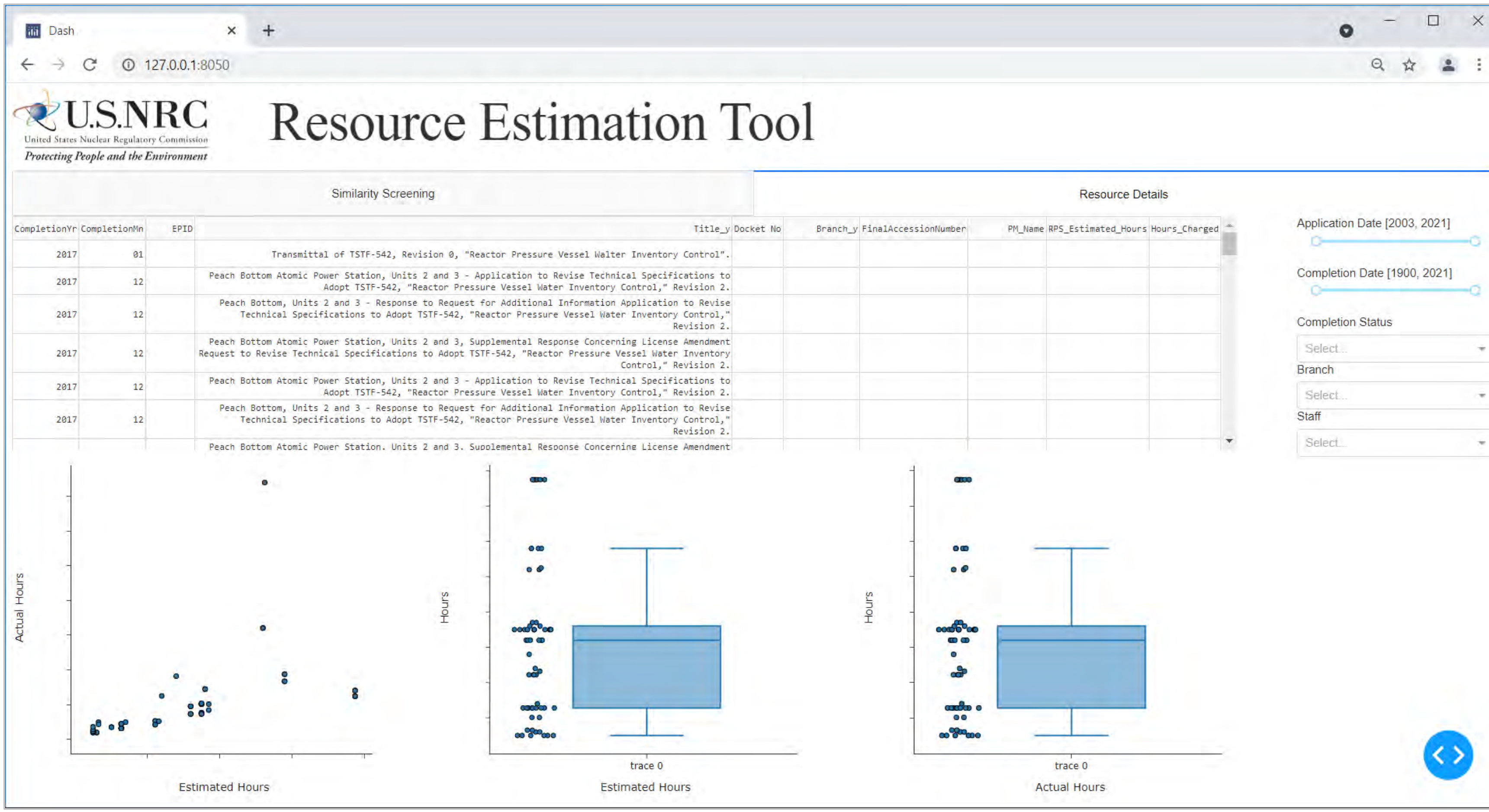
U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Resource Estimation Tool



U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Resource Estimation Tool



U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Current Status and Follow-on Work

- Preliminary acceptance testing complete
  - Historical data provides reasonable estimates of required resources and review durations
- NRR/EMBARK and NRR/DORL coordinating to finalize visualizations
- Develop and deploy final User Interface
- Potential Follow-on Work:
  - Search capabilities
  - Predict Branch assignments
  - Predict Standard Review Plan
  - Predict which Regulatory Guide(s) was used for the licensing action

Regulatory Named Entity Recognition

- ***Challenge:*** Title 10 of the Code of Federal Regulations (CFR), and other regulatory documents, reference sections of 10 CFR
  - Revisions to 10 CFR could impact other sections
- ***Goal:*** Create a tool to find and extract 10 CFR references from documents
- ***Method:*** Use Named Entity Recognition (NER) to label text as regulations and extract that text

Named Entity Recognition

(2 CARDINAL) On or before July 26, 1990 DATE , each holder of an operating license for a production or utilization facility in effect on July 27, 1990 DATE , shall submit information in the form of a report as described in 10 CARDINAL CFR 50.75 of this part, indicating how reasonable assurance will be provided that funds will be available to decommission the facility. [21 CARDINAL FR 355, Jan. 19, 1956 DATE , as amended at 35 FR 19660 DATE , Dec. 29, 1970 DATE ; 38 CARDINAL FR 3956, Feb. 9, 1973 DATE ; 45 CARDINAL FR 55408, Aug. 19, 1980 DATE ; 49 CARDINAL FR 35752 WORK_OF_ART , Sept. 12, 1984 DATE ; 53 CARDINAL FR 24049 DATE , June 27, 1988 DATE ; 69 CARDINAL FR 4448 WORK_OF_ART , Jan. 30, 2004 DATE ; 72 CARDINAL FR 49490 CARDINAL , Aug. 28, 2007] DATE 4 Emergency planning zones (EPZs) are discussed in NUREG–0396 DATE , EPA ORG 520/1–78–016, Planning Basis for the Development of State and Local Government Radiological Emergency Response Plans NORP in Support of Light-Water Nuclear Power Plants, December 1978 DATE .

SpaCy Default Entities

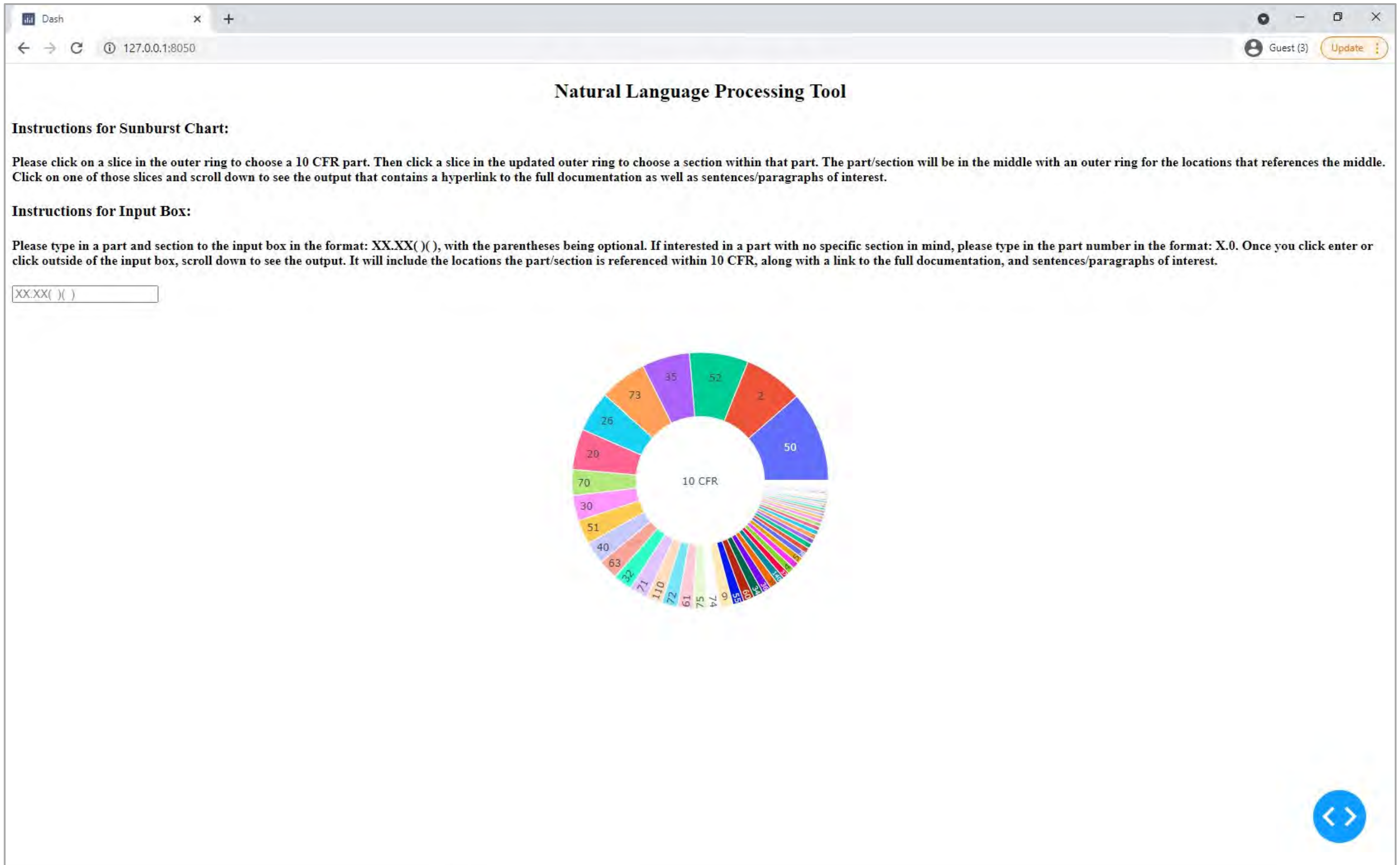
Addition of NRC Specific Language Patterns

(2 CARDINAL) On or before July 26, 1990 DATE , each holder of an operating license for a production or utilization facility in effect on July 27, 1990 DATE , shall submit information in the form of a report as described in 10 CFR 50.75 REG of this part, indicating how reasonable assurance will be provided that funds will be available to decommission the facility. [21 FR 355 FR , Jan. 19, 1956 DATE , as amended at 35 FR 19660 FR , Dec. 29, 1970 DATE ; 38 FR 3956 FR , Feb. 9, 1973 DATE ; 45 FR 55408 FR , Aug. 19, 1980 DATE ; 49 FR 35752 FR , Sept. 12, 1984 DATE ; 53 FR 24049 FR , June 27, 1988 DATE ; 69 FR 4448 FR , Jan. 30, 2004 DATE ; 72 FR 49490 FR , Aug. 28, 2007] DATE 4 Emergency planning zones (EPZs) are discussed in NUREG–0396 NUREG , EPA ORG 520/1–78–016, Planning Basis for the Development of State and Local Government Radiological Emergency Response Plans NORP in Support of Light-Water Nuclear Power Plants, December 1978 DATE .

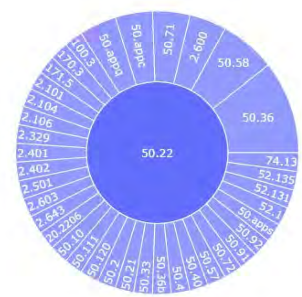
U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

- ***Used Python package Spacy***

10 CFR Reference Identification Tool



10 CFR Reference Identification Tool



Please type in a part and section to the input box in the format: XX.XX( )( ), with the parentheses being optional. If interested in a part with no specific section in mind, please type in the part number in the format: X.0. Once you click enter or click outside of the input box, scroll down to see the output. It will include the locations the part/section is referenced within 10 CFR, along with a link to the full documentation, and sentences/paragraphs of interest.

Output from the Sunburst Chart:

Click the link below to see the full documentation:

10 CFR 50.58

Below is a snippet with the reference of interest (each paragraph is separate):

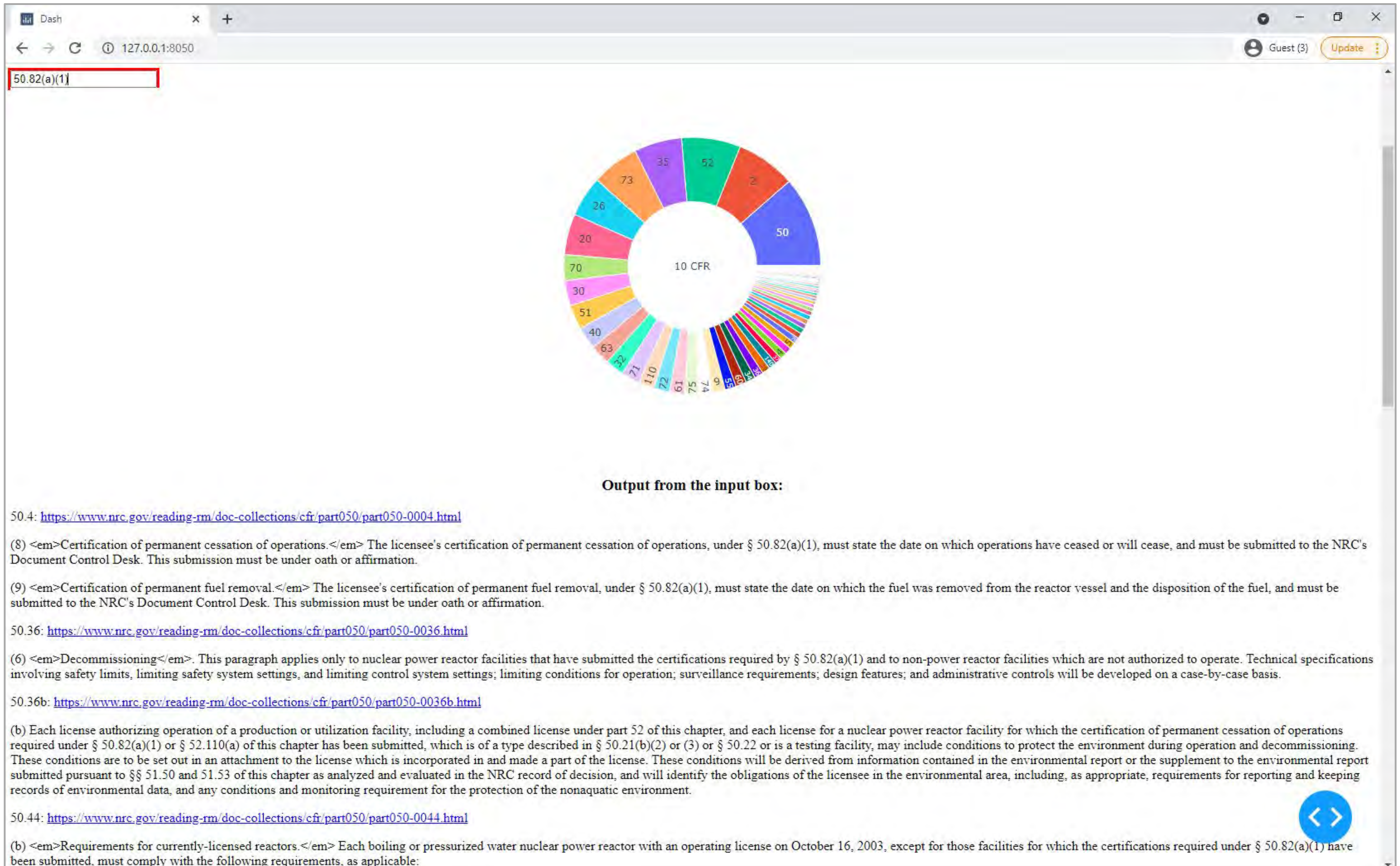
(a) Each application for a construction permit or an operating license for a facility which is of a type described in § 50.21(b) or § 50.22, or for a testing facility, shall be referred to the Advisory Committee on Reactor Safeguards for a review and report. An application for an amendment to such a construction permit or operating license may be referred to the Advisory Committee on Reactor Safeguards for review and report. Any report shall be made part of the record of the application and available to the public, except to the extent that security classification prevents disclosure.

(b)(1) The Commission will hold a hearing after at least 30-days' notice and publication once in the FEDERAL REGISTER on each application for a construction permit for a production or utilization facility which is of a type described in § 50.21(b) or § 50.22, or for a testing facility.

(5) The Commission will use the standards in § 50.92 to determine whether a significant hazards consideration is presented by an amendment to an operating license for a facility of the type described in § 50.21(b) or § 50.22, or which is a testing facility, and may make the amendment immediately effective, notwithstanding the pendency before it of a request for a hearing from any person, in advance of the holding and completion of any required hearing, where it has determined that no significant hazards consideration is involved.

U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

10 CFR Reference Identification Tool



U.S.NRC
United States Nuclear Regulatory Commission
Protecting People and the Environment

Conclusions

- Natural Language Processing is a powerful tool to leverage unstructured data in historical documents
- Deploying these tools would increase efficiency of staff by reducing time required for manual searches
  - Staff can leverage historical data in informing decisions

NRC CLOUD INFRASTRUCTURE

PUSHPA JAYAPAL AND STEVE SCHRADER

CLOUD STRATEGY

- The Agency’s objectives are:
  - Improve security, cost effectiveness, efficiency, agility, and scalability in delivering IT services;
  - Align with the OMB’s “Cloud Smart” policy and the Federal Cloud Computing Strategy;
  - Accomplish appropriate system and application migrations to cloud services as part of compliance with Federal DCOI mandates;
  - Establish consistent cloud solution planning and migration practices; and
  - Reduce risks to IT delivery, availability, and performance through a more distributed and consistent infrastructure and platform environment.
- To maximize cloud services benefits, the NRC will use the following strategies:
  - Leverage Software-as-a-Service (SaaS) first to support a low-code deployment approach and optimize functional requirements to take full advantages of SaaS benefits.
  - Leverage Platform-as-a-Service (PaaS) to drive technology standardization for modernized systems and applications that require customization.
  - Plan to acquire and support standardized PaaS platforms.
  - Increase application refactoring activities to rearchitect applications from monolithic and tightly integrated applications to loosely-coupled, cloud-based microservices focused on activities and workflows.
  - Adopt Infrastructure-as-a-Service (IaaS) only by exception.

MAJOR COMPONENTS

- Azure Commercial (IaaS and PaaS)
- AWS, NRC RES-Managed
- Other SaaS

AZURE INFRASTRUCTURE

- ExpressRoute connected stub network
  - No direct Internet Access—All access through the NRC TIC connection
  - 2-Gbps Connection through Equinix
  - Equinix will be the TIC 3.0 connection for NRC
  - Supports Cloud EDTE, Production, and DMZ zones
- Currently supporting several systems using the following PaaS (not exhaustive):
  - Azure Web Apps
  - Azure SQL Database
  - Azure Functions
  - Azure Bot and QnA Maker
  - Azure Search
  - Azure Cognitive Speech Service

CLOUD SECURITY

- When possible, all cloud systems use Private IP Space
  - In Azure, SaaS and PaaS Services use PrivateLink to provide for Private IP usage
  - In Azure and AWS IaaS, no public IP addresses are assigned to VMs
- Cloud Access Security Broker (CASB)
  - Provides policy enforcement regardless of what sort of device is attempting to access cloud services
- Azure Defender
  - Currently monitors Azure SaaS and PaaS Configurations
  - Can be configured to remediate identified issues
- "Standard" Network Security approaches, e.g. Splunk, Firewalls, IDS, AV

CURRENT AND FUTURE PROJECTS

- Application Migration Efforts – EIE, ILDC, NITA, Data Warehouse, ADAMS, TTC ColdFusion, RPS, ALM, Azure VDI
- New Capabilities – ActiveNav
- PaaS Implementations – Containers, Azure Security Center, Site Recovery, Mobile Apps, Logic Apps
- 3WFN Data Center Consolidation
- Evaluating and Scheduling all NRC FISMA Systems Cloud Migrations
  - Goal is to migrate all systems which can be migrated to the cloud by Dec 2026

THANK YOU

JAYAPAL, PUSHPA
PUSHPARANI.JAYAPAL@NRC.GOV

STEVE SCHRADER
STEVEN.SCHRADER@NRC.GOV

U.S. Nuclear Industry Survey on Artificial Intelligence and Machine Learning in Operating Nuclear Plants

Authors:
Zhegang Ma, Han Bao, Sai Zhang, Andrea Mack, INL
Min Xian, Univ of Idaho, Idaho Falls

J.C. Lane, NRC Project Manager

Division of Risk Analysis
Office of Nuclear Regulatory Research
U.S. Nuclear Regulatory Commission
NRC Agreement Number 31310019N0006
Task Order Number 31310019F0045

Project Objectives

- **Explore potential uses and applications of advanced computational tools and techniques, such as artificial intelligence (AI) and machine learning (ML), for operating nuclear plants**
- **Review nuclear DATA sources that could be applied by advanced computational tools and techniques**
  - **Generic-national & international data**
  - **Plant specific operating experience data**
- **Introduce widely used AI/ML algorithms in both supervised and unsupervised learning**
- **Review applications of advanced computational tools and techniques**
  - **reactor system design and analysis**
  - **plant operation and maintenance**
  - **nuclear safety and risk analysis**
- **Present insights on the potential applicability of AI/ML techniques to:**
  - **improve advanced computational capabilities**
  - **contribute to understanding of safety and risk**
  - **help decision-makers make better decisions**

Project Tasks

Task 1: Literature search for advanced computational tools and techniques appropriate for operating nuclear plants

Task 2: Survey to assess the current and potential applications of advanced computational tools in the commercial nuclear industry

Task 3: Explore potential applications of advanced computational tools and techniques to advanced reactors

Nuclear Industry Modernization Is On The Way

- **New approaches will support more efficient compliance with regulatory requirements in 10 CFR 50.65, “Requirements for Monitoring the Effectiveness of Maintenance at Nuclear Power Plants”**
- **NEI 18-10* industry guidance is a departure from the current preventative maintenance assessment paradigm (e.g., establishing structure, system and component performance criteria) and is intended to allow for a more dynamic assessment of maintenance effectiveness based on the use of data and risk trending analytics**
- **NRC resident inspectors will need better understanding of the underlying technologies employed in these new approaches, e.g., AI, ML, and data analytical tools**

***Nuclear Energy Institute, “Monitoring the Effectiveness of Nuclear Power Plant Maintenance,” NEI 18-10, 2018**

FRN NRC-2021-0048

- **Issued April 21 2021 (86FR20744)**
- **Posed 11 question to the public at large (ML21104A056)**
- **Designed to elicit industry and public perception regarding the benefits of using AI in nuclear plant operations**
- **12 Responses received (also available on Regulations.com site)**

No.	Participant	Response Accession Number
1	Anonymous	ML21113A083
2	Southern Research Institute (SRI)	ML21126A011
3	Florida Power & Light Company (FPL)	ML21139A103
4	Electric Power Research Institute (EPRI)	ML21141A184
5	Xcel Energy (Xcel)	ML21141A185
6	ForHumanity	ML21145A363
7	Blue Wave AI Labs (Blue Wave)	ML21145A364
8	X-energy	ML21145A366
9	Insight Enterprises, Inc. (IEI)	ML21145A367
10	Nuclear Energy Institute (NEI)	ML21145A369
11	Framatome Inc. (Framatome)	ML21153A056
12	Westinghouse Electric Company LLC (WEC)	ML21211A077

Survey Response Matrix

No.	Participants	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Responses Beyond Question Scope
1	Anonymous	N	N	N	N	N	N	N	N	N	N	N	Referenced two publications (Kortelainen et al. 2020, Suman 2020)
2	SRI	N	N	Y	N	Y	N	N	N	N	N	Y	None
3	FPL	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	None
4	EPRI	N	N	N	N	N	N	N	N	N	N	N	Referenced four EPRI-authored publications (EPRI 2020b, EPRI 2020a, EPRI 2021b, EPRI 2021a)
5	Xcel	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	None
6	ForHumanity	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Included an introduction of AI-related work conducted by the For Humanity
7	Blue Wave	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	None
8	X-energy	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	None
9	IEI	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	None
10	NEI	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	None
11	Framatome	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	None
12	WEC	Y	Y	Y	Y	N	Y	N	N	N	N	Y	None

High-Level Benefits of AI

Design, Operational Automation, Preventive Maintenance Trending, and Staff Productivity

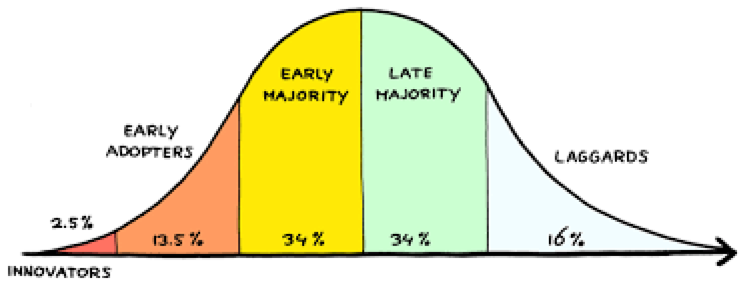
- **Increasing design-process efficiency**
- **Enabling data collection and analysis at a larger scope and faster speed**
- **Identifying patterns unnoticed by humans**
- **Suggesting control strategies not necessarily thought of beforehand**
- **Automating labor-intensive work**
- **Optimizing resource allocation**
- **Streamlining maintenance scheduling**

Common Refrains from Survey

- **Nuclear power may benefit more from AI/ML applications compared to other power generation sources**
  - **Due to the high cost of the workforce and regulatory requirements**
  - **Benefits may be concentrated on high-impact events**
    - **nuclear fuel**
- **But the cost of developing and implementing AI/ML is a challenge**
  - **Costs are usually high and upfront**
  - **Benefits are neither timely nor guaranteed**
- **Survey participants that have completed AI/ML applications said they were able to balance the development costs against expected plant improvements**

Common Refrains (con't)

- Confusion as to whether the nuclear power industry is:
  - Early adopter
  - Majority
  - Laggard



- Licensees foresee using a combination of in-house AI/ML talents and support from external entities such as vendors, national laboratories, and universities
- Most frequently mentioned AI/ML methods:
  - Artificial Neural Network
  - Clustering algorithms
  - Natural Language Processing

Artificial Neural Networks

- **ANNs are the most well-known methods of supervised learning and have the capabilities to be applied in broad areas, including regression analysis, classification, data preprocessing, and robotics**

- **ANN is composed of three types of layers: input, hidden, and output layers. Each layer consists of a set of nodes called neurons. A typical ANN has one input layer, one output layer, and multiple hidden layers**

- **The connections between nodes in different layers are associated with the weights that define the connection strength and are adjusted as learning proceeds**

- **ANN with more than three hidden layers is called a deep NN**

Clustering Algorithms

- **Clustering Analysis is used to identify data subgroups and clusters in such a way that data samples from the same cluster are more similar to each other than to those from different clusters**
- **There are many clustering algorithms because the notion of a ‘cluster’ is not easy to clearly define**
- **The most appropriate algorithm for a particular task needs to be chosen experimentally**
- **It works well for medium size datasets and a small number of clusters**
- **Hierarchical Clustering (HC) builds clusters by recursively partitioning data samples using merging or splitting strategies:**
    - **HC introduces a linkage criterion to decide which cluster pairs should be merged or should be split. The linkage criterion defines the dissimilarity and distance between sets of data samples.**

AI vs Regulatory Efficiency

AI/ML applications can improve NRC regulatory efficiency and effectiveness

Direct Approaches to Improve Regulatory Oversight

- **Use AI/ML to automate NRC staff labors such as reviewing plant documentation**
- **Use NLP to make the NRC ADAMS data more searchable**
- **Use surrogate modeling to verify and run simulation models submitted by the licensees**
- **Adopt advanced oversight methods to streamline regulatory process such as coordinating diagnostics data with risk-informed categorization**

Indirect Approaches to Improve Regulatory Oversight

- **AI/ML applications have potential to lead to safer plants with fewer events and thus a reduction in the number of regulatory activities**
- **Integrating AI/ML into regulatory activities can be a learning process and although decreased efficiency and increased cost might be a side effect initially, it is hoped to be temporary so that costs will eventually stabilize, if not decline**

Data Security

Major Concerns

- **Cyber intrusion**
- **Proprietary information leakage**
- **Loss of export control Several survey participants mentioned that their organizations**

Data Security Defenses

- **Data may have “inherent security” because it may be difficulty to draw significant insights from the stolen data, unless the intruders had access to the original model and software**
- **Some organizations are exploring AI platforms with co-located hardware and inspection systems to process data locally and minimize the need for data transfer**

Applications of AI

<u>Many</u> **applications are under concept exploration or strategic consideration:**

- **System and component monitoring (3 votes)**
- **Predictive maintenance (2 votes)**
- **Digital twins (1 vote)**
- **NDE inspections (1 vote)**
- **Automating human labor (1 vote)**
- **Cybersecurity (1 vote)**
- **Design support (1 vote)**
- **Fuel management (1 vote)**
- **Outage reduction (1 vote)**

Applications of AI (con't)

A *small number* are currently under development

- **Textual report analysis**
- **Predictive maintenance**
- **Work management**
- **Fuel cycle management**
- **Reactor operation and control**
- **Surrogate model development**
- **Focused on:**
  - **Corrective Action Program**
  - **Non-Destructive Examinations**
  - **Root Cause Analysis**

Tools Up & Running

Very few **are already up and running**

- **Customized tools developed for plant improvements:**
  - **Tool to predict moisture carryover in BWRs (MCO.ai developed by the Blue Wave AI Labs)**
  - **Tool to predict the BWR eigenvalue evolution for future fuel cycles (Eigenvalue.ai developed by the Blue Wave AI Labs)**
  - **Tool to determine root causes derived from symptoms (Metroscope developed by the Électricité de France)**
  - **Tool to evaluate multiple regression-based AI/ML algorithms to find trends in the data and select the optimal algorithm-Westinghouse**

Off-the-shelf products:

- **Commercial software tools--IBM Watson**
- **Tools not quite ready for prime time but close:**
  - **Xcel Energy’s CAP Intelligence Advisor (targeting late 2021 for the first deployment)**
  - **X-Energy’s Xe-100 Digital Twin (targeting 2025-2027 for the first deployment)**

Top Down Mgmt of AI

Both the top-down approach and the case-by-case approach for developing and implementing AI/ML are deemed to have pros and cons

- **No strong preference is demonstrated by the survey participants**

- **Commonly-mentioned advantages of top-down approach:**
  - **Enables a holistic and standardized framework**
  - **Easier to generalize and save repetitive work**
  - **Easier to share knowledge and experience**
  - **Increasing business efficiency**

- **Commonly-mentioned disadvantages of top-down approach include:**
  a. **Difficulty to adapt the framework to a changing technology landscape**
  b. **Challenge in developing a catchall strategy accommodating diverse applications**
  c. **Potential loss of innovative human inputs**
  d. **Uncertainty in oversight from NRC on requirements for top-down guidance**

FRN Questions

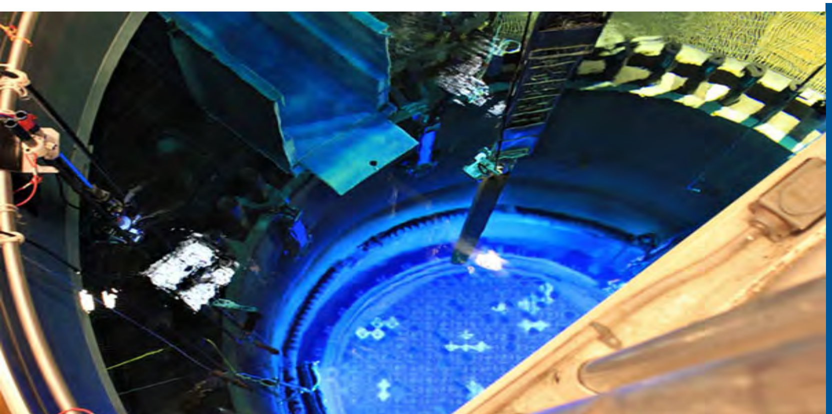
- What is status of the commercial nuclear power industry development or use of AI/ML tools to improve aspects of nuclear plant design, operations or maintenance or decommissioning? What tools are being used or developed? When are the tools currently under development expected to be put into use?
- What areas of commercial nuclear reactor operation and management will benefit the most, and the least, from the implementation of AI/ML? Possible examples include, but are not limited to, inspection support, incident response, power generation, cybersecurity, predictive maintenance, safety/risk assessment, system and component performance monitoring, operational/maintenance efficiency and shutdown management.
- What are the potential benefits to commercial nuclear power operations of incorporating AI/ML in terms of (a) design or operational automation, (b) preventive maintenance trending, and (c) improved reactor operations staff productivity?
- What AI/ML methods are either currently being used or will be in the near future in commercial nuclear plant management and operations? Example of possible AI/ML methods include, but are not limited to, artificial neural networks (ANN), decision trees, random forests, support vector machines, clustering algorithms, dimensionality reduction algorithms, data mining and content analytics tools, gaussian processes, Bayesian methods, natural language processing (NLP), and image digitization.
- What are the advantages or disadvantages of a high-level, top-down strategic goal for developing and implementing AI/ML across a wide spectrum of general applications versus an ad-hoc, case-by-case targeted approach?
- With respect to AI/ML, what phase of technology adoption is the commercial nuclear power industry currently experiencing and why? The current technology adoption model characterizes phases into categories such as: the innovator phase, the early adopter phase, the early majority phase, the late majority phase, and the laggard phase.
- What challenges are involved in balancing the costs associated with the development and application of AI/ML, against plant operational and engineering benefits when integrating AI/ML applications into operational decision-making and workflow management?
- What is the general level of AI/ML expertise in the commercial nuclear power industry (e.g. expert, well-versed/skilled, or beginner)?
- How will AI/ML effect the commercial nuclear power industry in terms of efficiency, costs, and competitive positioning in comparison to other power generation sources?
- Does AI/ML have the potential to improve the efficiency and/or effectiveness of nuclear regulatory oversight or otherwise affect regulatory costs associated with safety oversight? If so, in what ways?
- AI/ML typically necessitates the creation, transfer and evaluation of very large amounts of data. What concerns, if any, exist regarding data security in relation to proprietary nuclear plant operating experience and design information that may be stored in remote, offsite networks?

Memorandum of Understanding (MOU) Between US NRC and DOE on

Cooperation in the area of Operating Experience and Applications of Data Analysis

Matthew Humberstone, PhD
Office of Nuclear Regulatory Research
Nuclear Regulatory Commission

Felix Gonzalez, P.E.
Office of ES&H Reporting and Analysis
Department of Energy/AU



Background

- The NRC and DOE have been
  - leveraging data analytics technology
  - collecting/analyzing range of operating experience (OpE) data

- Started meeting in July 2020
  - Information Exchange meetings
  - NRR, RES, NMSS, INL, etc.

- Determined that leveraging an MOU to share tools, data, and experiences would be beneficial

NRC

Collects and Analyze data

Anticipates and resolves potential safety significance issues

Develops technical bases to support regulatory positions

DOE

Office of Environmental Protection and ES&H Reporting and Analysis

Collects and Analyzes data

Continuous improvements in environment, health, and safety

Track safety indicators

Resolve safety significant issues

MOU

- Formalized June 2021 (ADAMS ML21069A328)
  - The Objective is to cooperate through sharing data, technical information, lessons learned, and tools.
  - First meeting July 16th, 2021
  - MOU is for a period of 5 yrs

- Engagement Plan
  - Quarterly meetings to exchange recent developments on activities
  - One annual workshop to bring awareness across the agencies
  - Explore additional activities and model uses

- Box account established to share data and tools

Safety, Analytics, Forecasting, and Evaluation Reporting (SAFER)

Presentation at the

NRC Data Science and AI Workshop

August 18, 2021

NNSA Infrastructure

- NNSA's vital national security missions are dependent upon a safe, reliable, and modern Nuclear Security Enterprise infrastructure. This is a vast and complex network of facilities that include cutting-edge scientific, experimental, and engineering structures located throughout the nation.

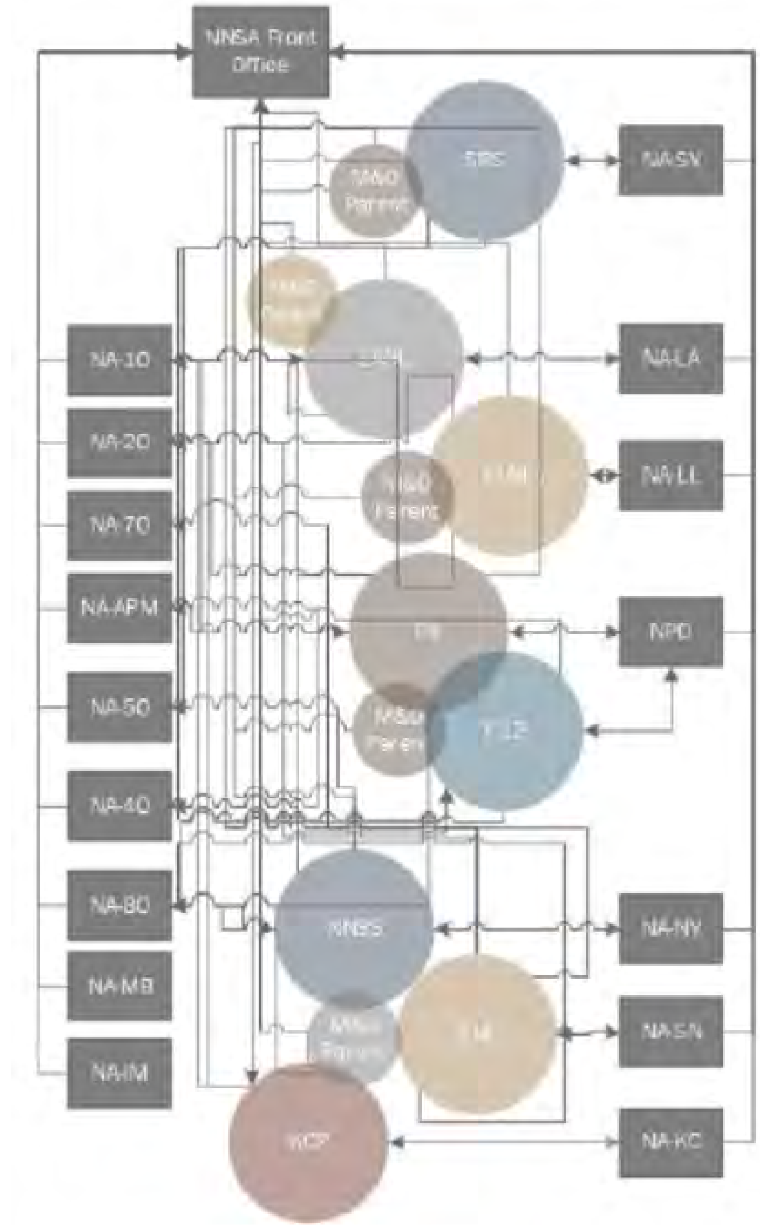
NNSA's Office of Safety, Infrastructure, and Operations

- NNSA's Office of Safety, Infrastructure, and Operations ensures that the existing architecture is safely operated, effectively managed, and that current and new facilities are adequately maintained to meet mission needs.

- To carry out this mission, the office has the responsibility to manage and implement the programs, policies, processes, and procedures for assuring effective integration of activities across the enterprise, working closely with other NNSA program offices.

NNSA Oversight

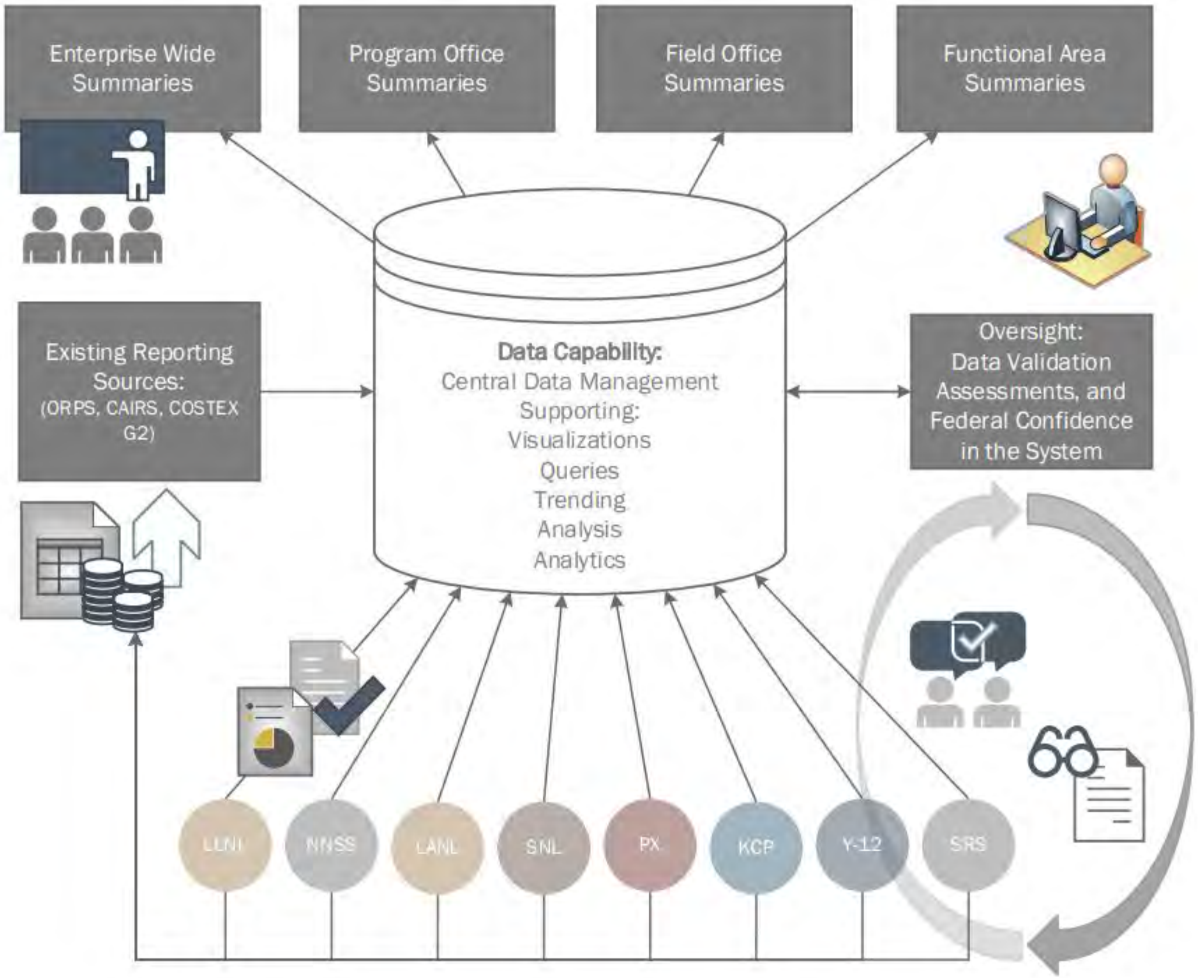
- NNSA Safety Oversight is performed at varying levels of maturity with a multitude of static, lagging, and uncoordinated data sources that limit data analysis and decision makers' ability to manage risks.



Simplified Model of the Current State of Information Flow

Desired State

NNSA is deploying new data-driven, risk informed, tools aimed at improving our communication, including the data, analysis, and visualizations we use to inform decision makers.



Desired Future State of Central Data Management and Increased Data Capability

HOW DOES SAFER FIT IN

WHAT IS IT?

- **An innovative software platform that easily integrates Departmental databases.**
  - Commercial Item, configured to meet the needs of NNSA.
  - Contains a suite of data analysis, data visualization and reporting capabilities.

WHAT WILL IT DO?

- **Improve Information Management and Sharing**
  - Improve our ability to understand the health of safety management programs.
  - Improve knowledge management
  - Improve communications throughout to NNSA complex

- **Improve Data-Informed Decision Making**
  - Improve our ability to use the information we currently and better understand what information should be collected.
  - Support planning and deploying resources in support of NNSA mission accomplishment.

Getting the right information, to the right person, when they need it.

SAFER -- Current Status

- **Maintenance Pilot** – First Functional Area incorporated into SAFER
  - Maintenance data from five NNSA sites integrated into the SAFER Platform
  - Data visualizations and user homepages established

- **User Interface Pages**
  - Field Office subject matter expert (SME),
  - Field Office Manager
  - NA-50 homepage, NA-50 Functional Area Leads, Topic Specific
  - Management and Operating Contractors (M&Os) -- Beginning initial planning for developing M&O SME page

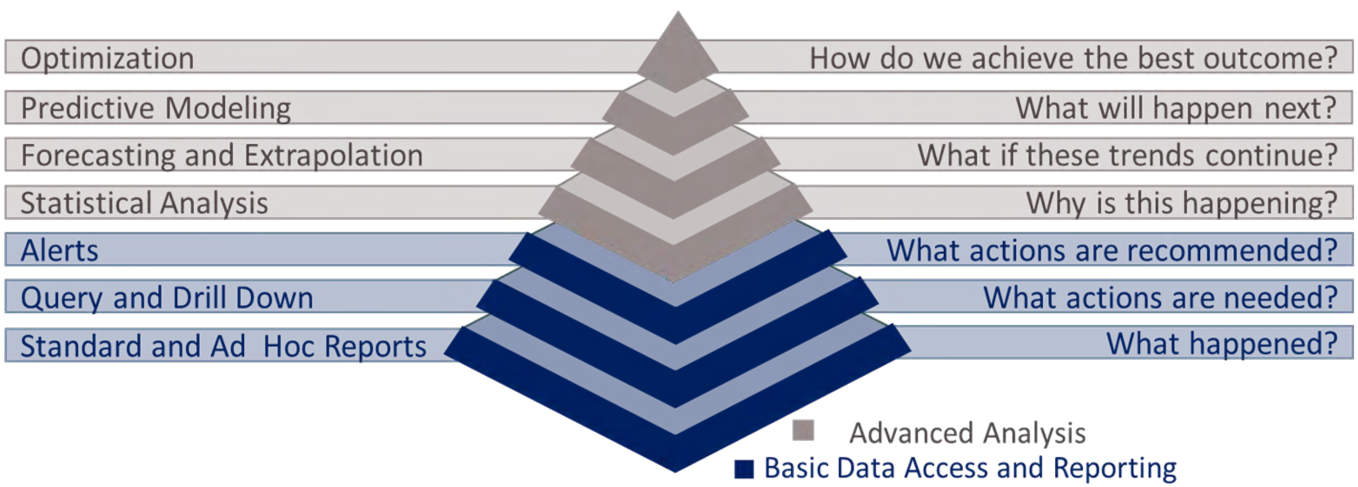
SAFER -- NEXT STEPS

WHERE WE ARE GOING

- Near Term:
  - Expanding - Addition of 4 Functional Areas in FY21
    - Electrical Safety, Fire Protection, Radiation Protection, Safety Basis,,
  - Begin interacting with M&O on development of M&O "Use Cases"

- Longer Term
  - Adding 4 more Functional Areas in FY22

- Continual: Adding data analytic tools including Alerts and Analysis



Summary

- SAFER is a modern data management and analysis tool to support NNSA safety and mission.
- Cooperative effort with HQ, Field Office and M&Os.
- Demonstrated value and potential in Maintenance Pilots.
- Value increases exponentially with adding of Functional Areas.

SAFER Team/Contact

SAFER Team

- NA-50 (Office of Safety, Infrastructure and Operations)
- NNSA Field Offices
- Palantir

NA-50 POC

For more information contact:

Jim O'Brien at james.Obrien@nnsa.doe.gov

U.S. Department of Energy
Office of ES&H Reporting and Analysis:
Similarity Search
Use Cases and Applications

Presentation for the U.S. Nuclear Regulatory Commission
[Data Science and Artificial Intelligence Regulatory Applications Workshops]
August 18, 2021

Felix Gonzalez, P.E.
Office of ES&H Reporting and Analysis
U.S. Department of Energy
Felix.Gonzalez@hq.doe.gov
301-903-9311

Office of Environment, Health, Safety and Security

Presentation Agenda

- Overview of similarity search ranking process and Natural Language Processing (NLP)
- Applications and Use cases
  - Q&A's database
  - Query complex Environment, Safety and Health (ES&H) related topics
- Lessons Learned and Concluding remarks

Process Summary: Query and Similarity Ranking

Database
(structured and
unstructured data)

Data Filtering

Unstructured data (i.e., text) processing

Formulate query (e.g., input text, keywords, question, etc.)

Submit data filters and text query

Input text processing

BOW Model with TFIDF

Vectorization

Similarity Ranking (i.e., cosine)

Sorting Results

Ranked Results

User Interface

Natural Language/Text Processing

- Text processing and normalization:
  - Lower-case (red)
  - Removes special characters, numbers, 2-character words, etc. (Yellow)
  - Remove stop-word (underlined)
  - *Lemmatization or Stemming*

- Model and metrics used:
  - Bag of Words (BoW) model
  - Term Frequency-Inverse Document Frequency (TFIDF)

- BoW and TFIDF used to calculate the cosine similarity metric

Sample Text Normalization and BoW Matrix

"Deficiencies in FY 2020 Funding and deficient cooling air caused the motor Fire."

Lemmatization

deficiency funding deficient cool air cause motor fire

Stemming

defici fund defici cool air caus motor fire

BoW Matrix

defici	fund	cool	air	caus	motor	fire
2	1	1	1	1	1	1

Search Query Application: Q&A's database

- DOE's COVID Hotline has answered questions from staff since the start of the pandemic
  - Q&A'S were initially tracked via spreadsheet in a shared drive
  - Hotline representatives searched the spreadsheet for answers
- As the spreadsheet grew it became challenging to find answers to questions
- An application was developed to show potential of Chat Bots to support the Hotline operations
- Hotline representatives requested the application instead show the top results which would improve their efficiency in evaluating questions and obtaining an answer quickly
- Evolved into a similarity search application that was integrated into Hotline's existing framework

Q&A Database Example
Text Normalization (1/2)

Sample Q&A's in the database

Question	Answer
Will DOE be collecting personal information upon building re-entry and, if so, how will it be protected?	We are not currently collecting personal or health information, but if it is determined to become necessary, any personal and health information collected by DOE or its contractors will be protected in accordance with applicable laws.
What advanced notice can I expect before returning to work?	We are working with supervisors and managers to give employees a reasonable amount of time to plan prior to being recalled to the workplace.
Where should face coverings or mask be worn?	DOE is following guidance published by the Centers for Disease Control and Prevention (CDC).
Do I need to wear a mask outside of a building?	DOE is following guidance published by the Centers for Disease Control and Prevention (CDC).

NLP

Normalized Question (no stemming)
collecting personal information upon entry protected
advanced notice expect returning
face covering mask worn
need wear mask outside

Vectorization

BoW model wTFIDF

Sample input question

"Do I need to wear a mask when inside a building?"

NLP

need wear mask inside

*no stemming

Vectorization

BoW model wTFIDF

Similarity Ranking

Q&A Database: Example Ranking (2/2)

- The ranking column specifies how similar is the "input question" to the questions in the database.

Question

"Do I need to wear a mask when inside a building?"

Question	Ranking* (= 1 - Cosine)
Do I need to wear a mask outside of a building?	0.84
Where should face coverings or mask be worn?	0.12
Will DOE be collecting personal information upon building re-entry and, if so, how will it be protected?	0
What advanced notice can I expect before returning to work?	0

*Ranking score of 1.0 would be a perfect match while 0 is no similarity.

- Model accuracy continued to be improved by adding different ways to ask a question to the Q&A database.

Similarity Search: Complex Safety Topics

- DOE Data Analytics and Machine Learning Tools used to analyze ES&H data
  - Search algorithms
  - Data visualization and trending
  - Topic modeling
  - Text clustering

- Leverage the Q&A application to obtain insights in ES&H data and perform more efficient searches

Use Case: Reports related to Oxygen Deficient Atmosphere

- DOE maintains several ES&H databases that are used to:
  - Extract insights from past related events
  - Increase awareness of hazards (e.g., thru safety communications)

- Recent events related to workers accessing oxygen deficient atmosphere (e.g., nitrogen inerted cabinet or room) and passing out or asphyxiating.

- Current tools are limited in how keywords are considered in the searches

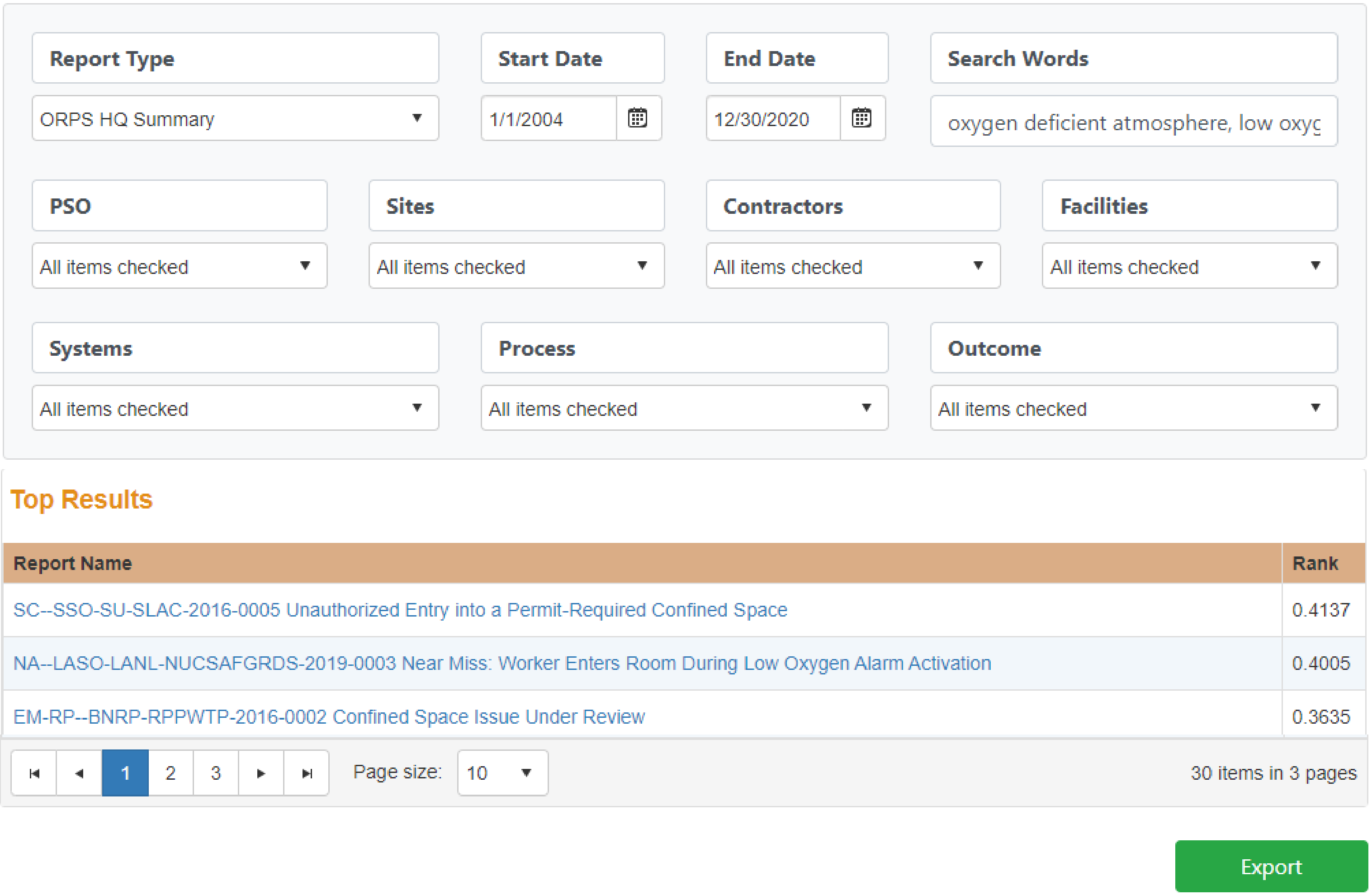
Occurrence Reports Search Approaches

- Topic categorization relies on identified issues of interest (140+ topics are currently tracked)
- Advance information retrieval and search approaches can benefit current systems
- Categorization of occurrence reports help drill down
- Similarity based ranking that relies on the text can be used with multiple keywords or full text of an event description

Similarity Search

- Similarity search used to find and rank reports:
  - Using topic keyword search “oxygen deficient atmosphere, low oxygen alarm, nitrogen inert, confined spaces, halon”
  - Using text of a report of interest
- Testing different approaches:
  - Lemmatization
  - Stemming
  - Importance weighting

Similarity Search Dashboard Sample Screen Shot (1/2)



Similarity Search Dashboard Sample Screen Shot (2/2)

Report Type	Start Date	End Date	Search Words
ORPS HQ Summary	1/1/2004	12/30/2020	On April 21, 2016, a Facility and Operat

PSO	Sites	Contractors	Facilities
All items checked	All items checked	All items checked	All items checked

Systems	Process	Outcome
All items checked	All items checked	All items checked

Top Results

Report Name	Rank
SC--SSO-SU-SLAC-2016-0005 Unauthorized Entry into a Permit-Required Confined Space	0.9958
EM-RL--PHMC-PFP-2006-0018 241-Z D-4 Tank Pit entry prior to completion of atmosphere sampling	0.7233
EM-RP--BNRP-RPPWTP-2016-0002 Confined Space Issue Under Review	0.7014

1 2 3 Page size: 10 30 items in 3 pages

Export

Similarity Search Lessons Learned

- Avoid removing/ignoring words important to the corpus
  - Develop custom stop-words list
  - Do not ignore terms using document frequency parameters
    - max_df =1.0
    - min_df = 0

- Computational costs affected by
  - Size of data
  - Size of BoW model matrix
    - Stop-words
    - N-grams (co-occurring words)
    - Larger values of max_df (up to 1.0)
    - Lower values of min_df

- Stemming is computationally faster than lemmatization and recommended when users don’t need to see the normalized text.

CAP Automation and Informed Inspection Preparation Project

Tim Alvey, Manager, Exelon Nuclear Innovation Group

Ahmad Al Rashdan, Ph.D. Senior Research and Development Scientist, Idaho National Laboratory

Jonathan Hodges, PhD., Global Service Lead for Data Analytics, Jensen Hughes

August 18, 2021, NRC Workshop

Exelon Generation

Agenda

- Introduction – Tim
  - Vision
  - Incentive
- Technical Approach - Jonathan
  - Challenges
  - Text Confidence Scores
  - Neural network architecture
  - Measuring success
- Broader Industry Potential - Ahmad
  - Integrating data from multiple plants
  - Data-driven keywords
- Future Work and Concluding Remarks – Tim

Vision

- Explore artificial intelligence and machine learning techniques to improve use of plant information
- Leverage rapidly advancing technologies/methods
- Opportunities to improve process (e.g., CAP)

Incentive for Change … Why CAP?

- Cornerstone of Reactor Oversight Process (ROP)
- Streamline and improve corrective action program (CAP) and process
- Better inform the information provided for NRC inspection planning and support purposes

Exelon Generation®

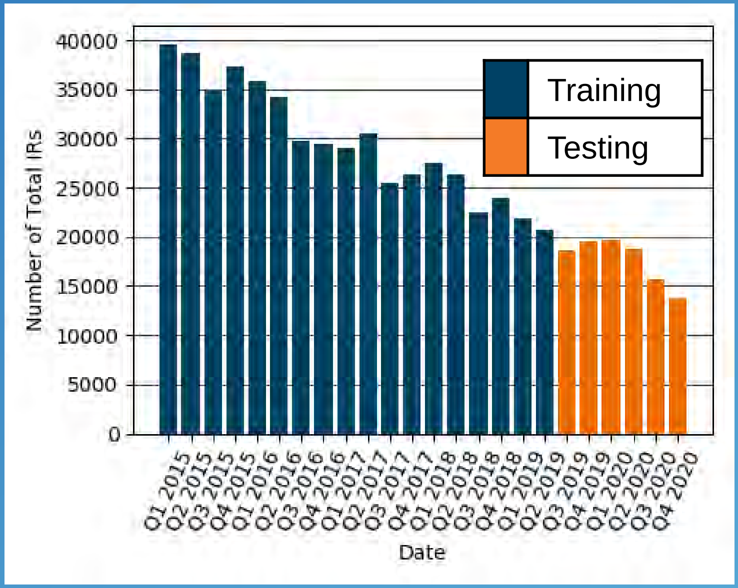
Challenges – Available Data

Category	Field	Description
Identifiers	FACILITY	Site affected by the incident
	IR_NUMBER	Numeric identifier
	ORIGINATION_DATE	Date the incident report was written
	SYSTEM_CODE	Which system was affected
	UNIT	Which unit was affected
Initial Text Description	IR_SUBJECT	Subject line describing the incident
	CONDITION_DESCRIPTION	Primary text field describing the incident.
	IMMEDIATE_ACTIONS_TAKEN	Describes immediate actions responding to the incident.
	RECOMMENDED_ACTIONS	Describes actions recommended by the reporter
Initial Screening Questions	HAS_EQUIPMENT	Was the incident associated with a specific piece of equipment?
	INITIAL_SCREENING_1	Is the equipment located in the Vital Area, Protected Area, or other owner controlled properties?
	INITIAL_SCREENING_2	Procedure or process issues with the potential to affect compliance with TS or license conditions?
	INITIAL_SCREENING_3	Potential reportability concerns?
	INITIAL_SCREENING_4	Analysis or setpoint deficiencies that impact onsite or offsite dose or dose rates?
	INITIAL_SCREENING_5	Nuclear safety issue?
	INITIAL_SCREENING_6	Significant Industrial Safety Issue (i.e.; excluding First Aids, non-work related issues, PPE Issues, etc?
	INITIAL_SCREENING_7	Personnel injury requiring offsite medical attention?
	INITIAL_SCREENING_8	Tampering, vandalism or malicious mischief?
Shift Review Questions	EQUIPMENT_FUNCTIONAL	Binary field - Did the equipment lose functionality due to the event represented by IR?
	EQUIPMENT_OPERABLE	Binary field - Was the equipment operable at the time the incident occurred?
	EVENT_REPORTABLE	Binary field - Does the incident represent a reportable incident?
	FUNCTIONAL_BASIS	Text describing why the incident represents a loss of functionality.
	OPERABLE_BASIS	Text describing why the incident represents a loss of operability
	REPORTABILITY_BASIS	Text describing why the incident represents a reportable incident
	HAS_WORK_REQUEST	Is there a work request associated with the incident report?
Station Ownership Committee (SOC) Review	IR_PRIOIRTY	Investigation class of an event, based on risk impact and risk of recurrence.
	IR_SEVERITY	Significance level of an event, based on consequence of what happened and could have happened.
	MRFF	Does the event qualify as a maintenance rule functional failure.

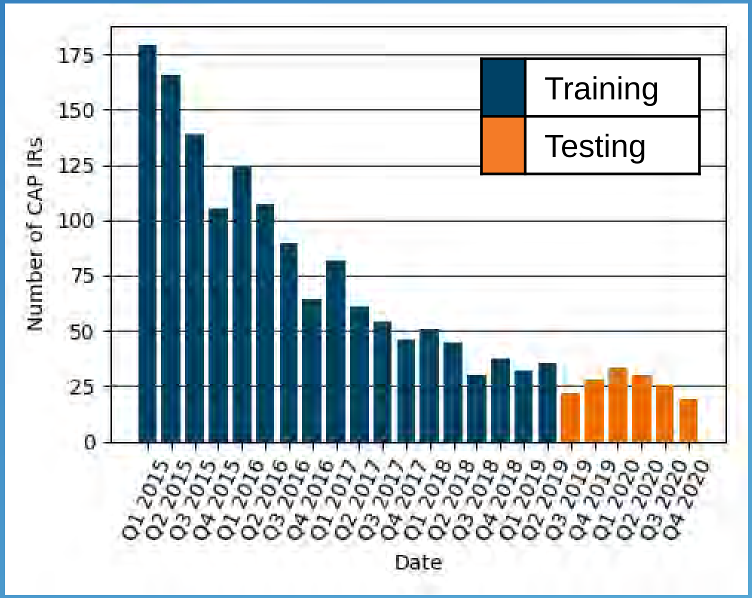
Exelon Generation

Challenges – IR Statistics

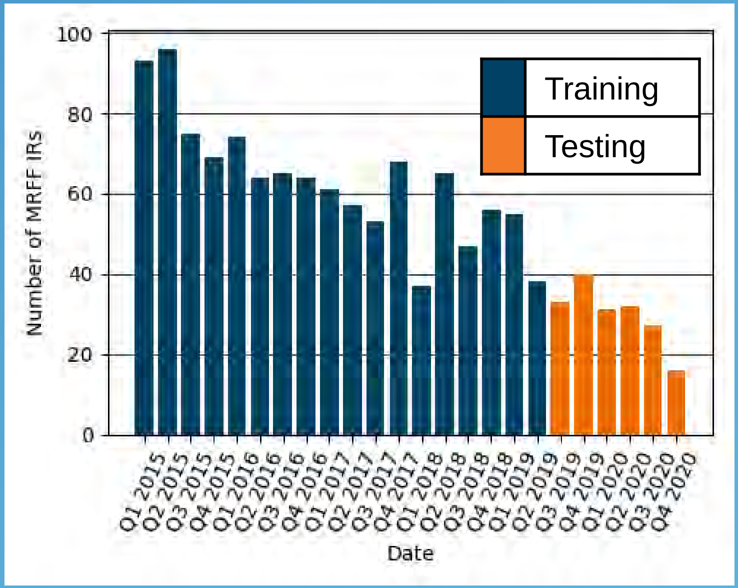
- Highly skewed datasets
- Adverse to Quality IRs ~0.1-0.2% of data



Total IRs

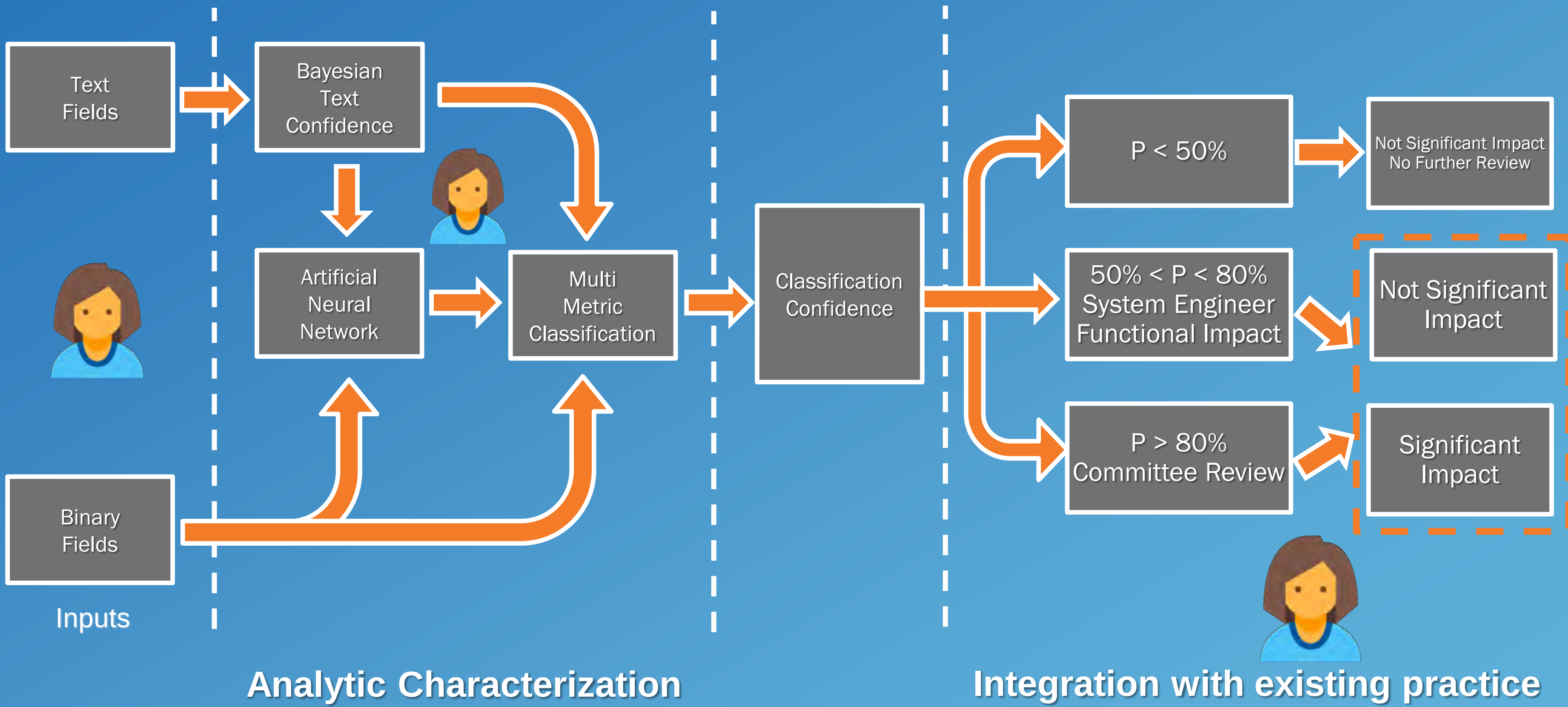


CAP IRs



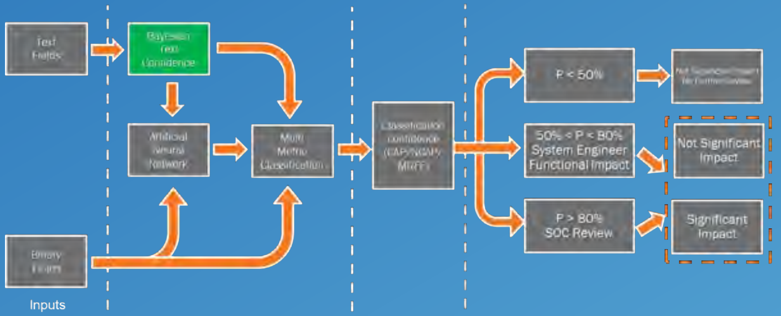
MRFF IRs

The Approach

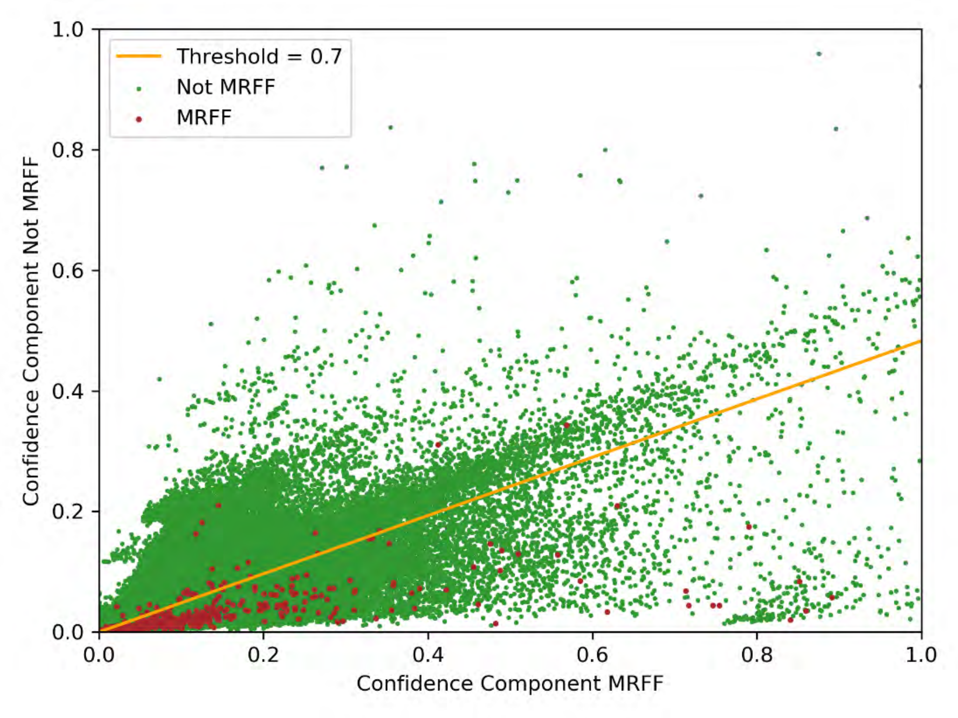


Exelon Generation.

Text Confidence Scores

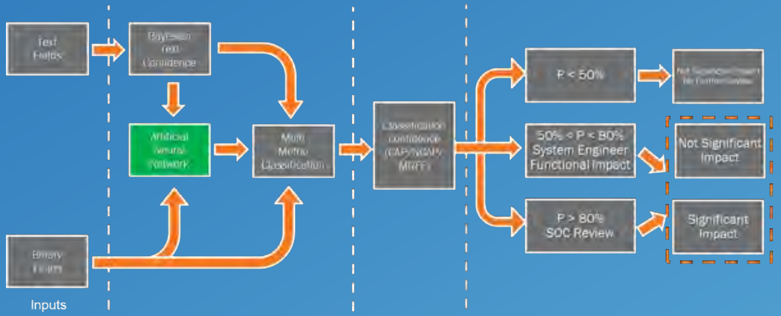


- Bag of words approach to Natural Language Processing (NLP)
- Split each text field into 1-word, 2-word, and 3-word phrases
- Bayesian inference uses conditional probability of class 1 versus class 2



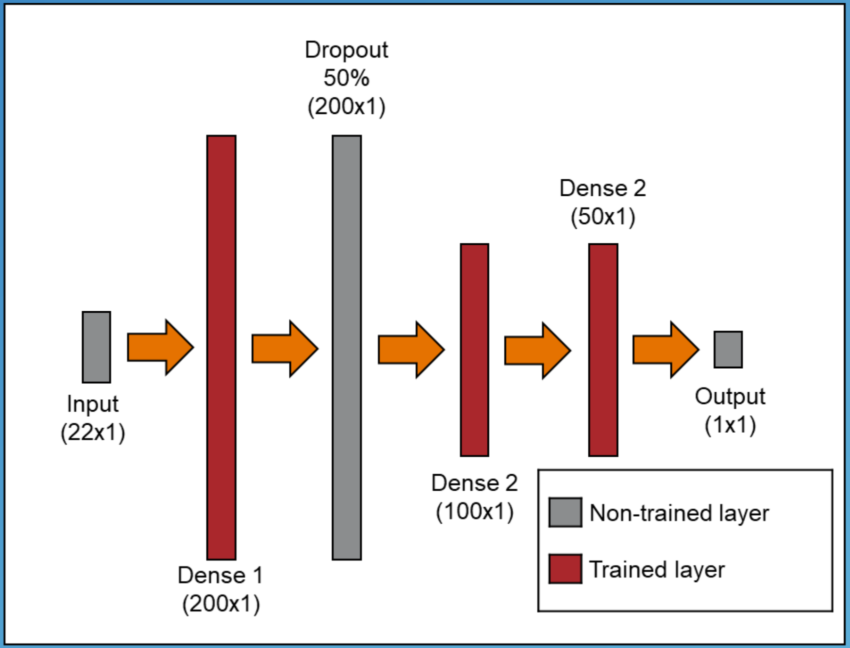
Exelon Generation.

Artificial Neural Network



SUBJECT-CONFIDENCE
CONDITION_DESCRIPTION-CONFIDENCE
IMMEDIATE_ACTIONS_TAKEN-CONFIDENCE
RECOMMENDED_ACTIONS-CONFIDENCE
OPERABLE_BASIS-CONFIDENCE
REPORTABLE_BASIS-CONFIDENCE
FUNCTIONAL_BASIS-CONFIDENCE
SOC_COMMENTS-CONFIDENCE
EQUIPMENT_OPERABLE
EQUIPMENT_FUNCTIONAL
EVENT_REPORTABLE
UNIT
INITIAL_SCREENING_1
INITIAL_SCREENING_2
INITIAL_SCREENING_3
INITIAL_SCREENING_4
INITIAL_SCREENING_5
INITIAL_SCREENING_6
INITIAL_SCREENING_7
INITIAL_SCREENING_8
HAS_EQUIPMENT
HAS_WORK_REQUEST_NUMBER

Text Confidence
Numeric/Binary Data



Neural Network Confidence

Inputs

Network Architecture

Output

Exelon Generation

Measuring Success

Misses:	Potential regulatory impacts
False Positives:	Process efficiency impacts
System Bias:	False Positives > Misses

- Accuracy
  - Bad metric for skewed data
  - 99.8% accurate by predicting NO system issues

$$Accuracy = \frac{(TP+TN)}{(TP+FP+TN+FN)}$$

- False Negative Rate (FNR)
  - Fraction of real issues which may have regulatory implications depending on the significance

$$FNR = \frac{FN}{(TP+FN)}$$

- False Discovery Rate (FDR)
  - Fraction which will need to be evaluated by plant personnel due to false alarms

$$FDR = \frac{FP}{(TP+FP)}$$

	Ground Truth	
Model	Issue	Not Issue
Issue	True Positive (TP)	False Positive (FP)
Not Issue	False Negative (FN)	True Negative (TN)

Dataset	Training		Testing	
Metric	FDR	FNR	FDR	FNR
ANN Alone	2%	0%	3%	6%
Multi Metric Class.	15%	0%	20%	2%

Exelon Generation

Broader Industry Potential

- Integrate data from multiple plants to improve AI/ML model performance
- Create industry scalable model for CR data-mining
- Validate plant AI/ML models via benchmarking

How can data from the broader industry be used to improve model results?

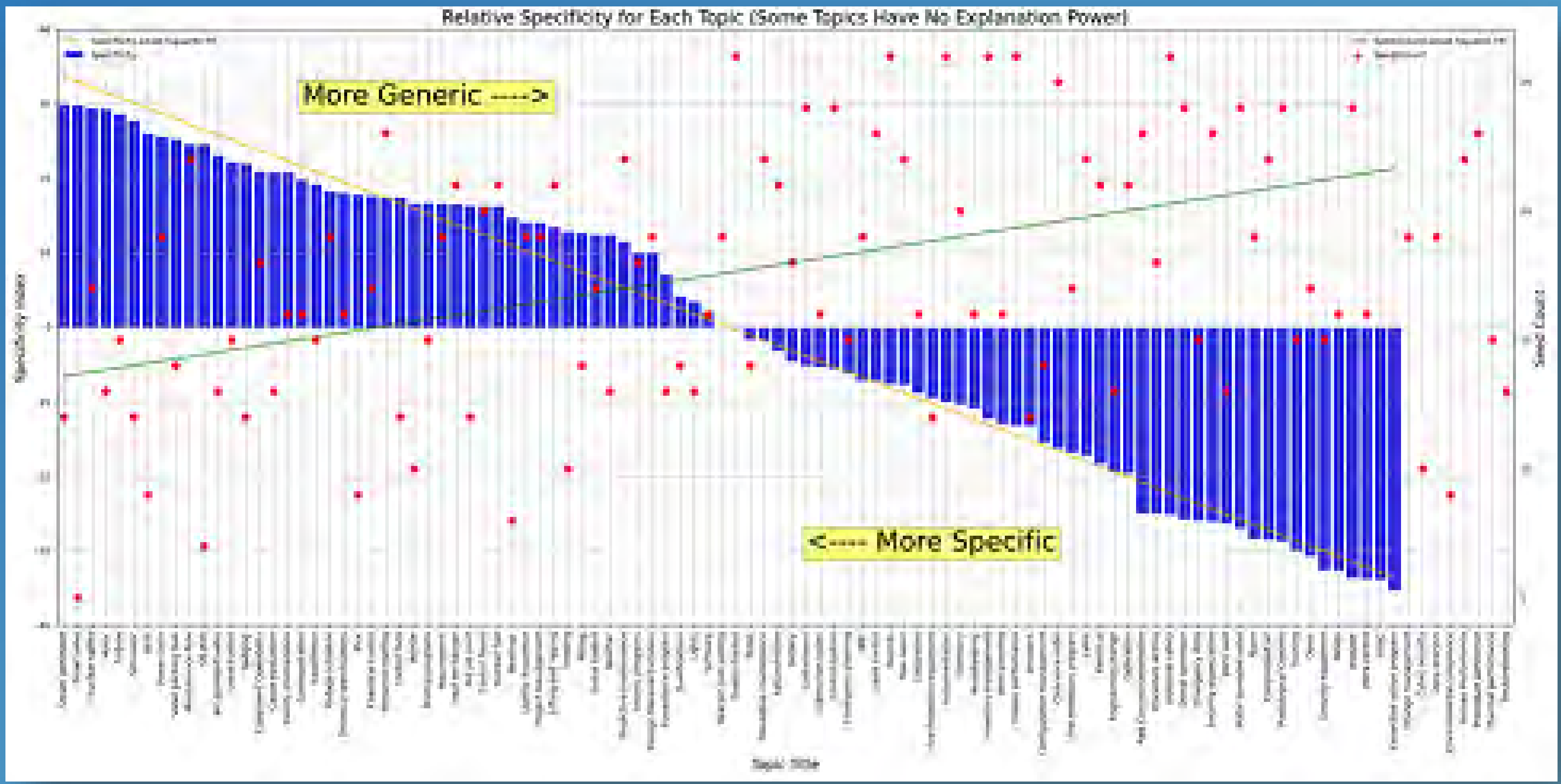
MIRACLE
(*Machine Intelligence for Review and Analysis of Condition Logs and Entries*)

	Utility 1 Model	Utility 2 Model	Combined Model *using fewer fields*
Utility 1 Data (large dataset)	84%	75%	>85%
Utility 2 Data (medium dataset)	77%	90%	>90%

Idaho National Laboratory LWRS LIGHT WATER REACTOR SUSTAINABILITY Exelon Generation

Broader Industry Potential

Create data-driven keywords using industry data to standardize usage for industry-wide trending



Future Work

- Validate plant models independently via benchmarking
- Enhance assessments and inform inspections
  - Streamline information sharing through an inspection data portal
  - Develop data-driven metrics to support inspection outcomes
  - Inform these processes though automation
- Develop tools to automate and identify risk contributors
  - Components and/or operator actions
  - Programmatic and predictive trends
- Deploy open-source tools for broad industry use

Concluding Remarks

- AI/ML will strengthen Corrective Action Program
- Improve Exelon's internal governance and oversight
- Technologies and methods are improving rapidly
- Integration of similar applications with NRC (e.g., pilot project) presents the opportunity for a powerful outcome

Questions?

Tim Alvey
Manager
Exelon Nuclear Innovation Group
Tim.Alvey@exeloncorp.com

Jonathan Hodges
Service Line Leader in Advanced Modeling
Jensen Hughes
jhodges@jensenhughes.com

Ahmad Al Rashdan
Senior R&D Scientist
Idaho National Laboratory
Ahmad.alrahdan@inl.gov

Application of Data Analytics to Mine Nuclear Plant Maintenance Data

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

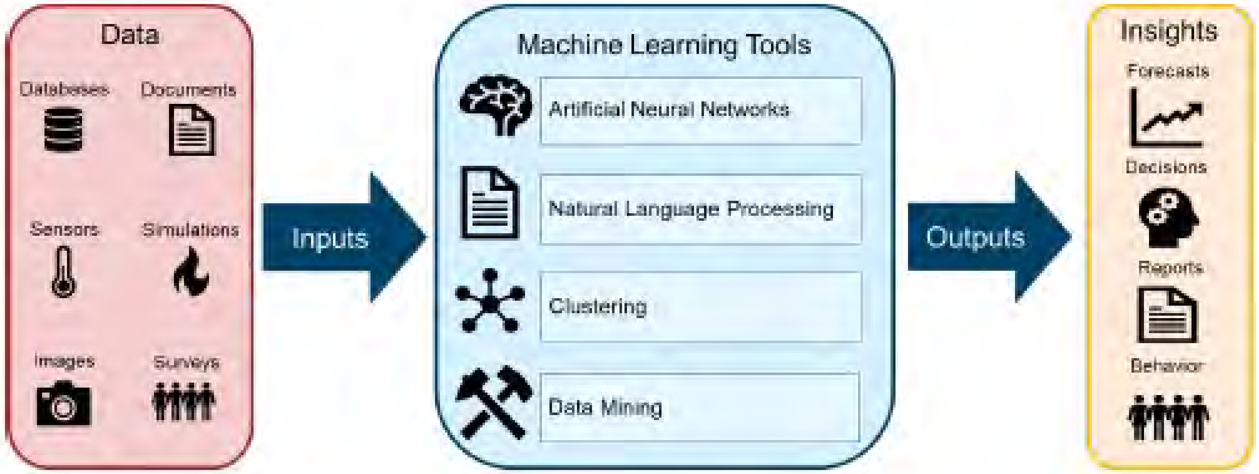
U.S. NRC AI Workshop
August 2021

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

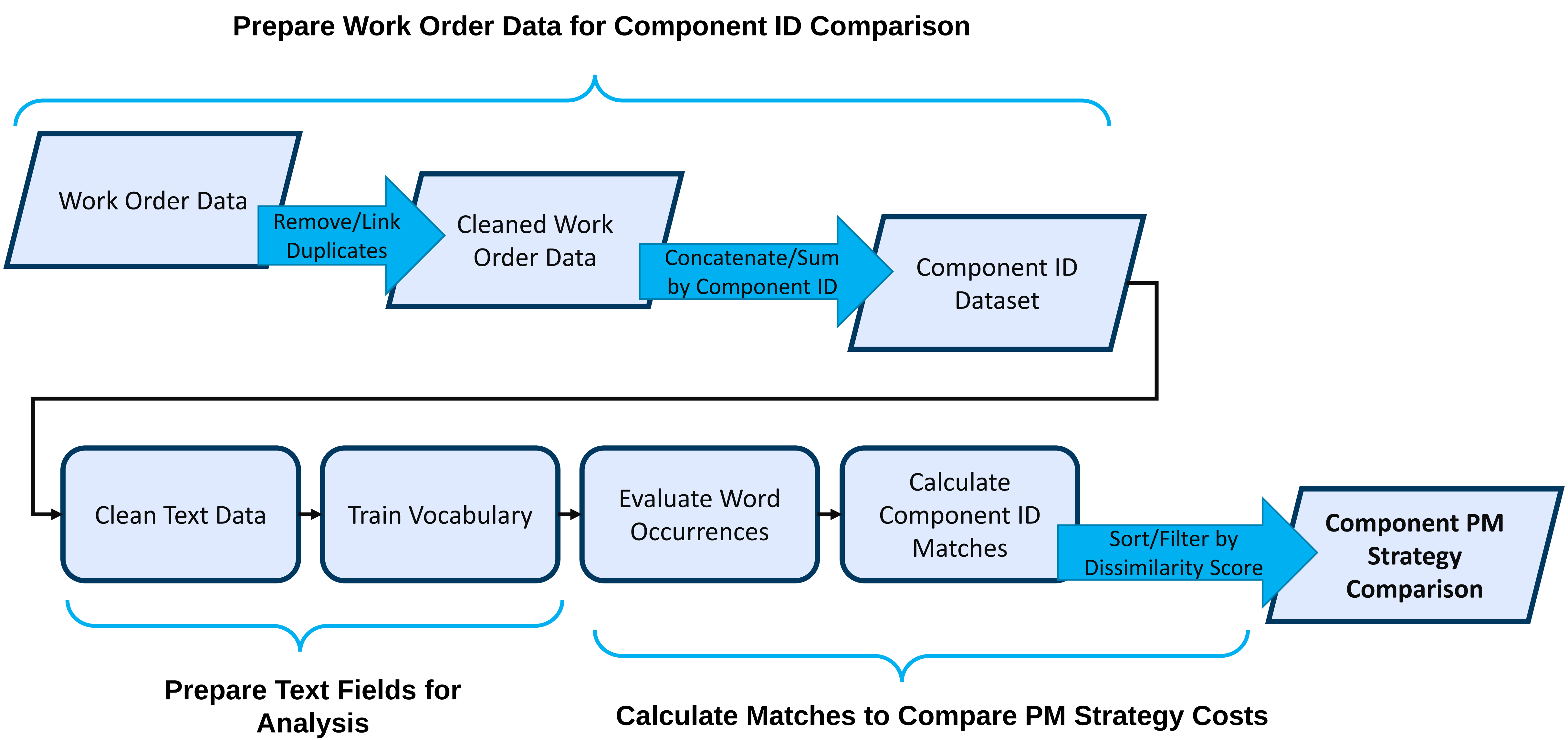
1. Collect work orders for each Component ID and combine into the **Component ID Dataset**
2. Train word models to obtain vocabulary in the dataset for both text fields
3. Evaluate word and phrase occurrences for each Component ID in the dataset
4. Compute the pairwise cosine distance for each Component 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

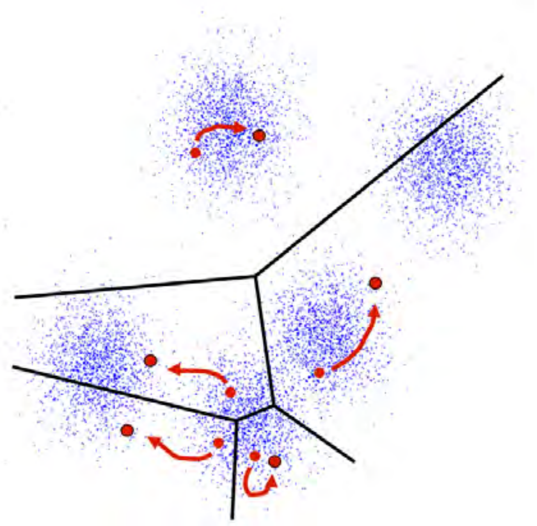
ACRONYM TRANSLATION MATRICES

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

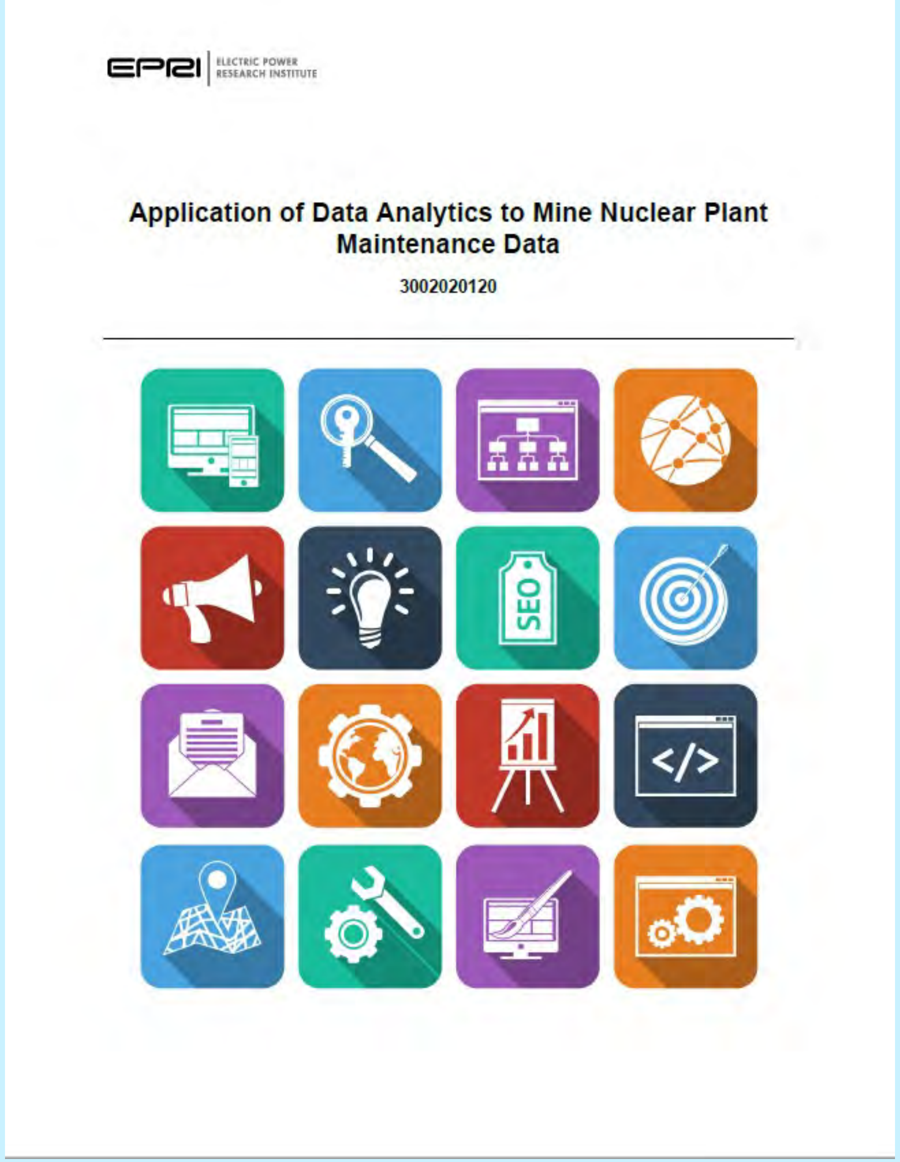
PMID	PMFREQ	PMDESC	PMHRSAVG	PMCSTAVG
RE500178	182	LUBRICATE BEARINGS eWP	3.09	0
RE500070	364	Perform Annual Pump Maintenance - Packing Inspection and Adjustment	8.22	0

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

Together…Shaping the Future of Electricity