ADVANCING NRC EVENT ANALYSIS

Machine Learning Approaches for Classification of Human Performance-Related Events in Licensee Event Reports

Matthew Homiack

Reactor Systems Engineer (Data Scientist)
U.S. Nuclear Regulatory Commission
Office of Nuclear Regulatory Research
Division of Systems Analysis
Accident Analysis Branch

Matthew.Homiack@nrc.gov

Explored many key technical considerations for use of AI/ML in regulatory decision-making through development of an LER classifier.

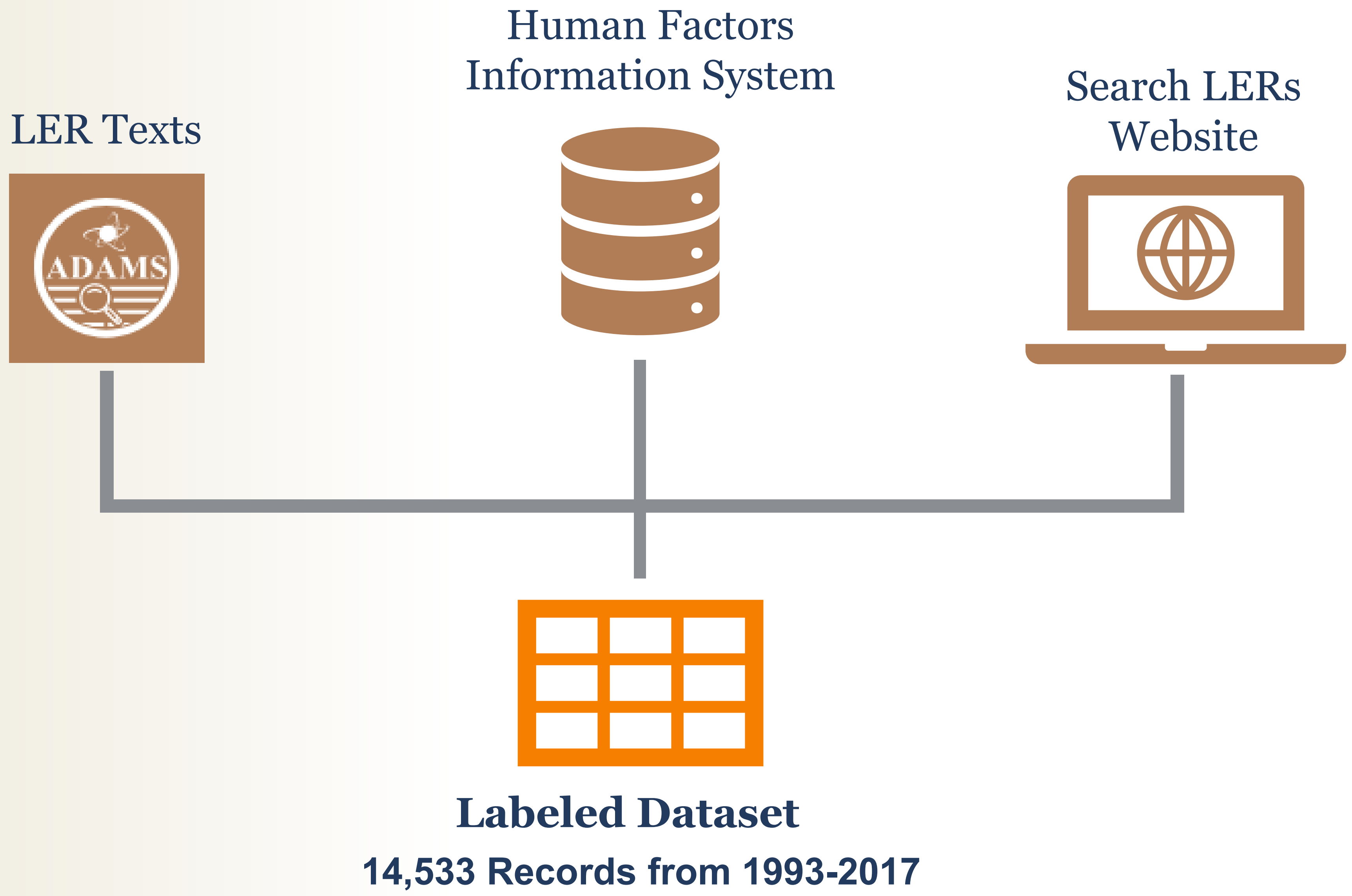
Machine Learning Pipeline

Data Collection and Curation → Preprocessing → Feature Extraction → Model Development → Model Evaluation

Technical Considerations

Explainability	Trustworthiness	Bias
Robustness	Ethics	Security
Risk Analysis	Test, Evaluation, Verification, and Validation	Assurance Processes
Model Maintenance	Domain Adaptation	Data Drift
Field Performance Degradation	Life Cycle Management	Data Quality, Quantity, Applicability, and Uncertainty

How data is collected and curated shapes its quality, quantity, and applicability, and carries ethical implications.



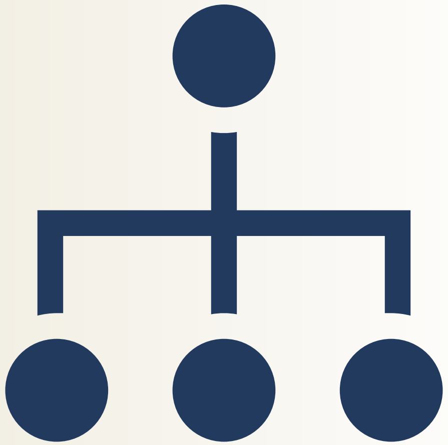
Preprocessing can improve data quality and applicability, but may introduce bias.

,\nOn the evening of December 2,\n1994, an operator was in the
process of \nhanging a Protective Tagging Record (PTR) for work on
Emergency Service\ni\nWater System B pump in support of
scheduled refuel outage activities.\nThis\ntag required the operator
to open 600V electrical breaker 12610 at load\n’\ncanter 71L-26
located in the East Electrical Bay.\nThe operator located the\ncorrect
breaker, but dropped the PTR tag.\nAfter picking up the tag and
\nupon returning his attention to the breaker, but without
reverifying\ncorrect breaker 12610, the operator inadvertently
pushed the \”TRIP\” button\nfor breaker 12606.\

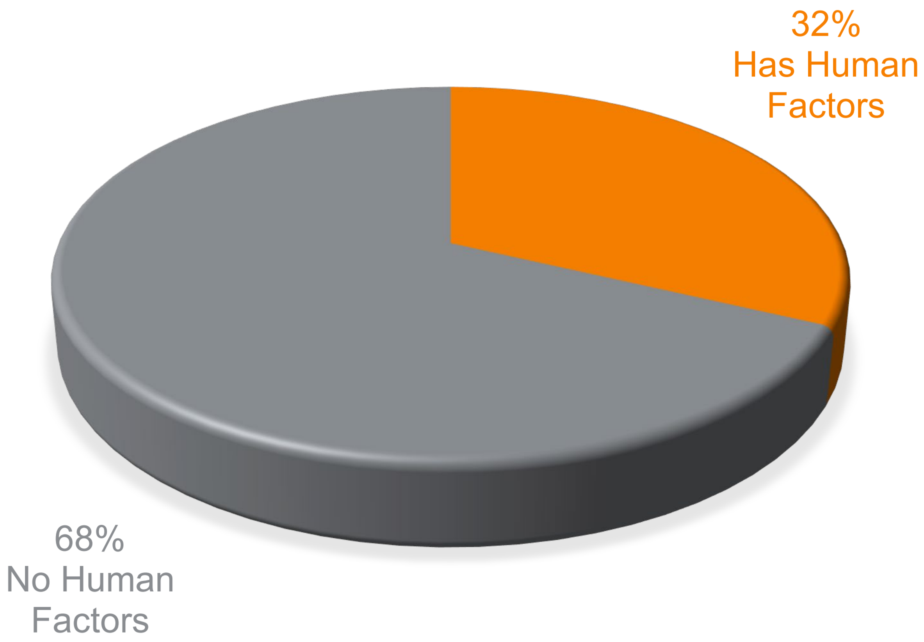
Textual information removed during preprocessing shown in orange.

Feature extraction and model development influence the explainability of AI decisions and address bias.

Naïve Bayes Model



Class Imbalance



Token-level contributions per class illustrate the internal logic of the model and enhance explainability.

LERs with Human Factors	LERs with no Human Factors
Briefing	Lightning
Crew	Weld
Workers	Corrosion
Knowledge	Capacitor
Human	Manufacturing

Trustworthiness is supported by test and evaluation strategies that align system behavior with intended use.

Performance metrics for high-recall, human-in-the-loop system design

$$\text{False Negative Rate} = \frac{\text{False Negatives}}{\text{True Positives} + \text{False Negatives}}$$

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

$$\text{Precision} = \frac{\text{False Negatives}}{\text{True Positives} + \text{False Negatives}}$$

		Predicted	
		Negative	Positive
Actual	Negative	True Negative	False Positive
	Positive	False Negative	True Positive

Putting the ML model into use would save 17% in classification effort per year.

Illustration based on average workload of 137 LER reviews/year

21 True Negatives
(properly classified LERs that are not of interest)

2 False Negatives
(acceptably low proportion of mis-classified LERs that are of interest)

43 True Positives
(properly classified LERs that are of interest)

71 False Positives
(mis-classified LERs that are not of interest)

Savings → **23** LER reviews/year

Remaining Workload → **114** LER reviews/year

Lessons and insights from this project will be used to shape future NRC research efforts.

AI/ML success for regulatory decision-making depends on carefully considering all applicable technical considerations in each stage of the development pipeline.

Naïve Bayes demonstrates strong performance for human factors classification in LER texts.

Collaboration between data scientists and domain experts is essential for ensuring alignment with real-world needs.