

Nuclear Safety Analysis using NLP

Advisors: Dr. Najmedin Meshkati and Dr. Yolanda Gil (Viterbi School of Engineering)

Students: Anood (ADS), Shruthakeerthy (CS), Samarth (ADS), Chenfeng (ADS), Rushit (CS), Vaidehi (CS), Abhishek (CS), Adarsh (CS), Siddarth (CS), Yuwen (ADS)

Problem Statement



Diablo Canyon Power Plant, California

- Most of Nuclear and Aviation incidents are caused by safety culture problems.
- Safety reports are too lengthy (900+ pages).
- NLP techniques are needed to extract valuable information.

Data and Resources Available

**Diablo Canyon Independent Safety Committee
Thirty-Second Annual Report on the Safety of Diablo Canyon Nuclear
Power Plant Operations
July 1, 2021—June 30, 2022**



Robert J. Budnitz, Chair
Per F. Peterson, Vice-Chair
Peter Lam, Member

Approved: September 28, 2022

932 pages!

*Pocket Guide
to INPO 12-012:*
**Traits of a
Healthy Nuclear
Safety Culture**

INPO®

Procedures contained inconsistent guidance for conservative decision-making. (Pg. 489)

LA 1 / WP 1

Evaluation of an update of local seismology was limited to its impact on the Long Term Seismic Program and did not include its impact on the plant's design basis. (Pg. 614)

PI 1 / CO 1 / WP 1, 3

Leaders are not consistently setting, modeling, and reinforcing clear standards and expectations for conservative decision-making, resulting in a station culture that favors production-oriented interpretation of the license basis. (Pg. 489)

LA 1, 2, 4 / WP 1 / CL 2

Inadequate staffing of Emergency Response Organization (ERO) personnel on shift occurred due to the fact that an operator had filled an on-shift ERO position for three consecutive shifts while the operator's formal qualification for being able to use a self-contained oxygen breathing apparatus for respiratory protection had expired. (Pg. 488)

PA 1, 2 / QA 4 / LA 1, 6

A licensing basis change on the 230kV offsite power system was not accompanied by a review for whether the change might create an unreviewed safety question, which would have required prior approval by the NRC. (Pg. 613)

LA 4, 5 / CO 3

Station leadership had not effectively used performance metrics for self-identifying deficiencies for early indication, tracking, and resolution of safety culture performance deficiencies. (Pg. 615)

PA 3 / LA 7, 8

Leadership Safety Values and Actions

Problem Identification and Resolution

Personal Accountability

Work Processes

Continuous Learning

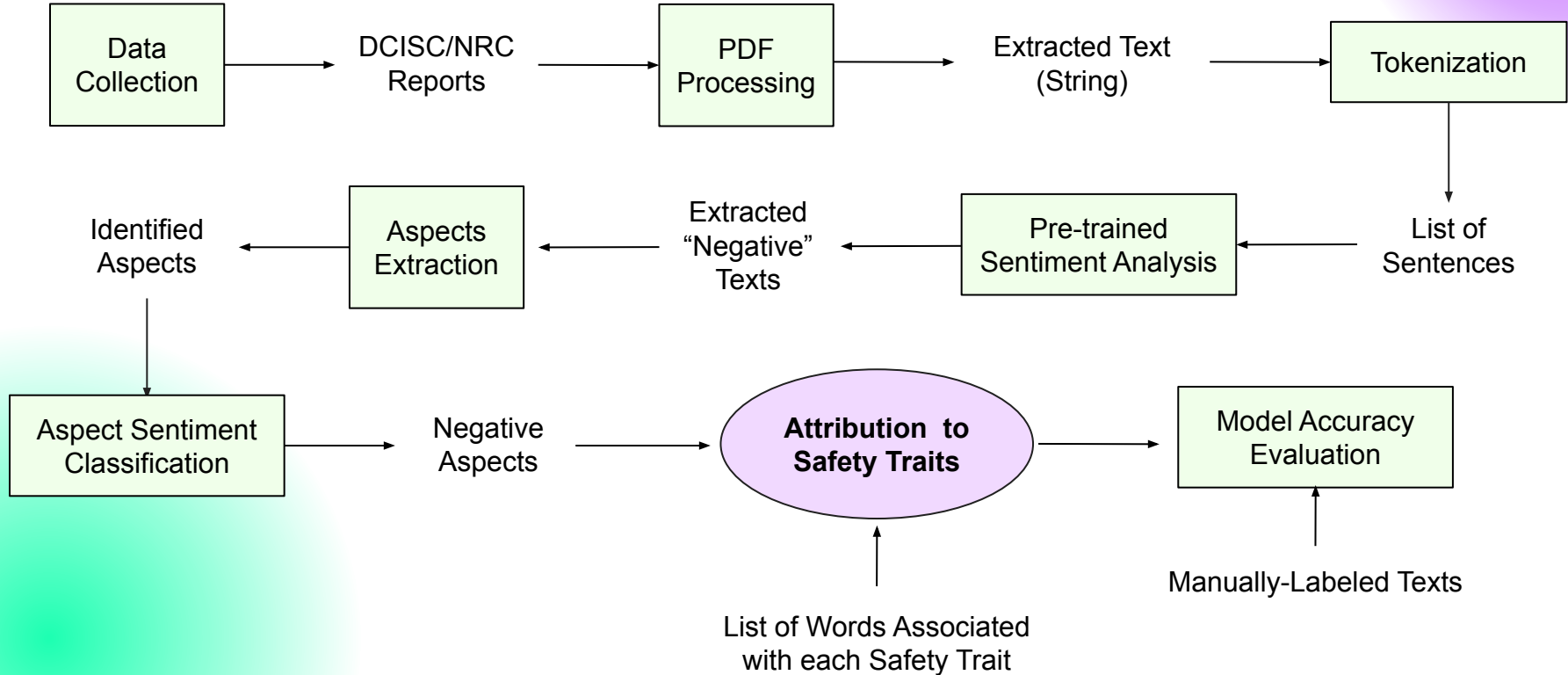
Environment for Raising Concerns

Effective Safety Communications

Respectful Work Environment

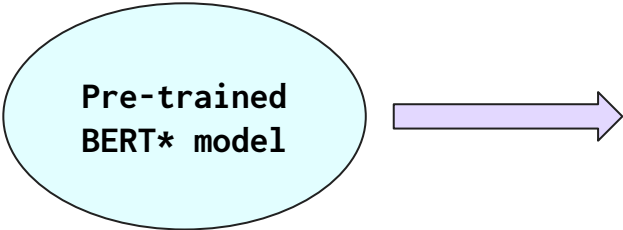
Questioning Attitude

General Workflow



Initial Sentiment Analysis

**Pre-trained
BERT* model**

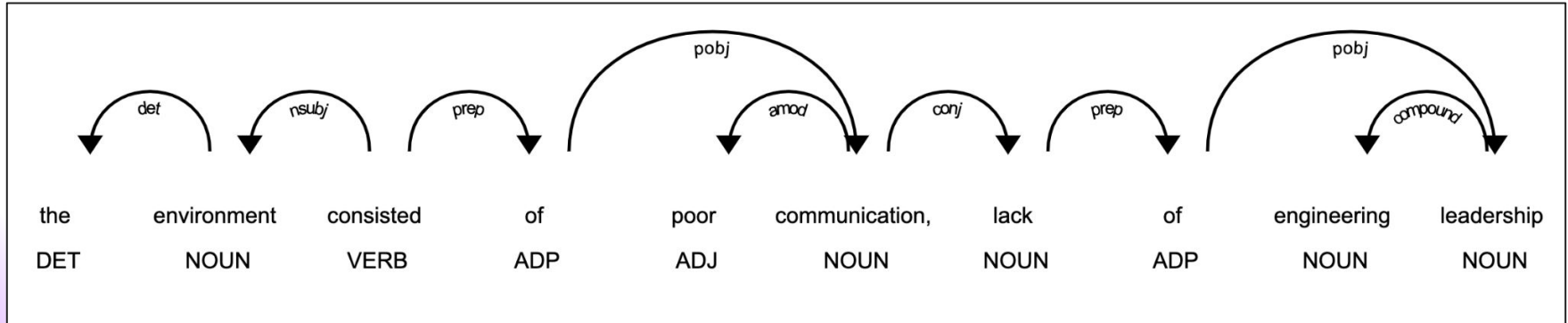


```
{'sentence': 'Operations performance with respect to human error rate has  
been Red (Unsatisfactory) since July 2013.',  
'page': 12,  
'label': 'negative',  
'confidence': 0.9728230237960815},  
  
{'sentence': 'Also, equipment problems due to aging have led to an  
increasingly negative trend in the station's Deficient Critical Component  
Backlog Orders.',  
'page': 196,  
'label': 'negative',  
'confidence': 0.9721506237983704},  
  
{'sentence': 'Mr. Wardell replied that the root cause of the safety  
system functional failure was due to lack of clear standards for risk',  
'page': 421,  
'label': 'negative',  
'confidence': 0.9389176964759827},
```

1) Rule-Based Approach

“The RCE team determined that the environment consisted of poor communication, lack of engineering leadership, too much reliance on vendor designs, time pressure, and distractions.”

-DCISC 24th Annual Report (2014), Page 139



```
{'communication': [poor], 'leadership': ['lack']}
```


2) Machine Learning Approach

- Supervised approach is challenging due to the need for large labelled data.
- We are currently exploring several unsupervised learning approaches.

“Text Categorization using SVD and LSA algorithms” or “LSA Topic Modeling with Cosine Similarity Mapping”:

1. Preprocess input texts by removing stop words and lemmatizing.
2. Extract features using TF-IDF vectorization and create a matrix.
3. Decompose matrix using SVD and LSA to obtain seedwords.
4. Map input texts to topics based on cosine similarity with topic vectors.

Thanks!

Do you have any suggestions?

