Digitalization of Risk Assessment and Evaluation of Incident Reports for Companies Operating in Alberta

Presented By:

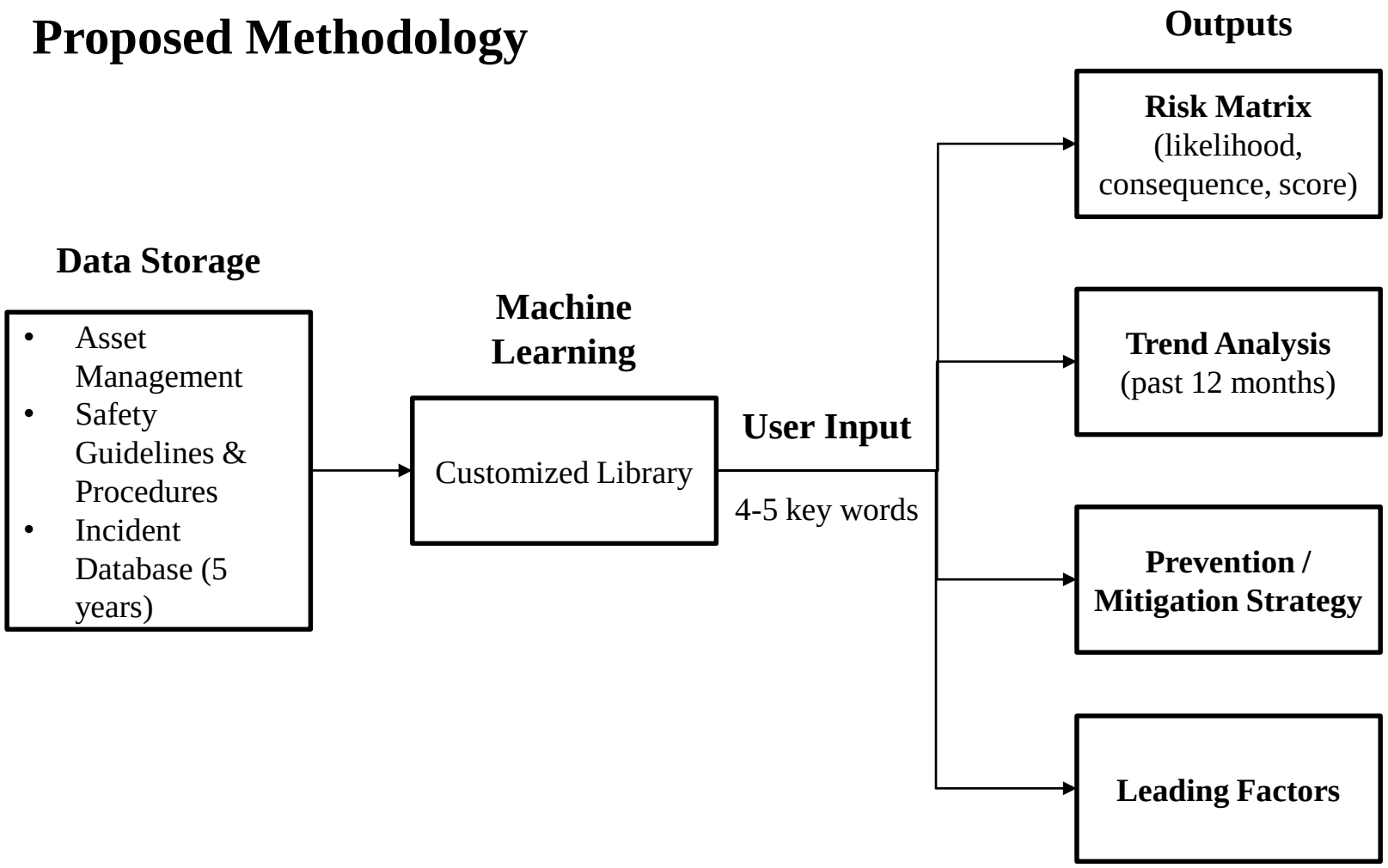
Daniel Kurian, Dr. Fereshteh Sattari, Dr. Lianne Lefsrud

Why Digitalization of Incident Reports is Important?

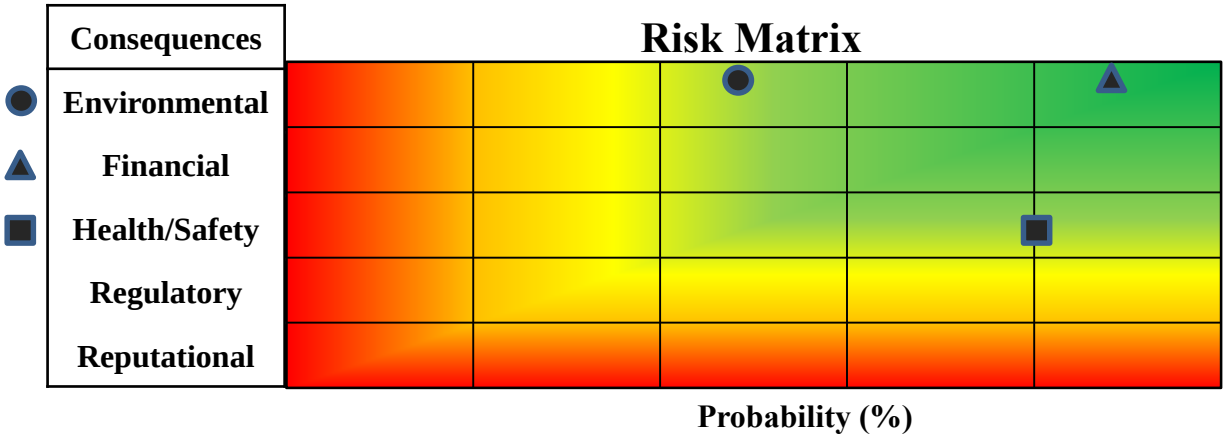
- **Incidents still occur and are costly:**
  - In 2016, costs & claim payments cost Canadians ~$8.7 billion[1]
  - Production loss, absenteeism, medical costs, and compensation equate to 4% of the annual global gross domestic products[2]
- **Gaps in current methods of incident reporting:**
  - Typos and misinformation
  - Inadequate follow-ups for incidents
  - Time consuming[3]
  - Bias (different interpretations)[4]

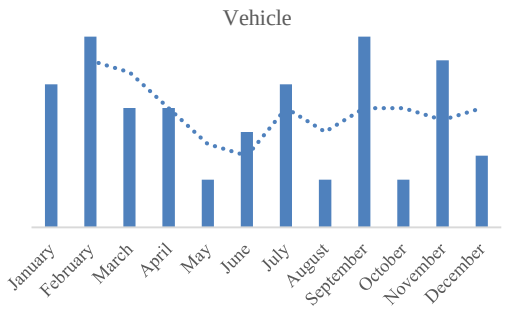
1. 2016 Statistics from Canadian Centre for Occupational Health and Safety, https://www.ccohs.ca/events/mourning/.
2. Takala, J., Hamalainen, P., Saarela, K.L., Yun, L.Y., Manickam, K., Jin, T.W., Heng, P., Tjong, C., Kheng, L.G., Lim, S., and Lin, G.S. (2014). Global estimates of the burden of injury and illness at work in 2012. *Journal of Occupational and Environmental Hygiene*, 11: 326e337.
3. Nijs Jan Duijm, "Recommendations on the use and design of risk matrices," Safety Science, 76, pp. 21–31, 2015.
4. Philip Thomas, J. Eric Bickel, and Reidar B. Bratvold, "The Risk of Using Risk Matrices," SPE Annual Technical Conference and Exhibition, 2013.

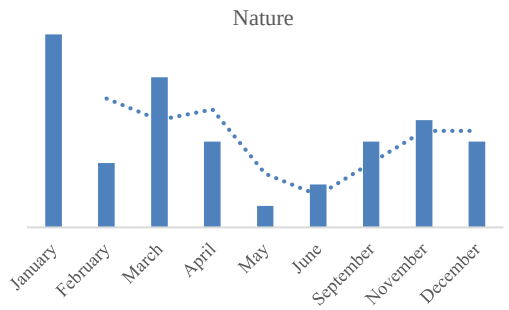
Proposed Methodology

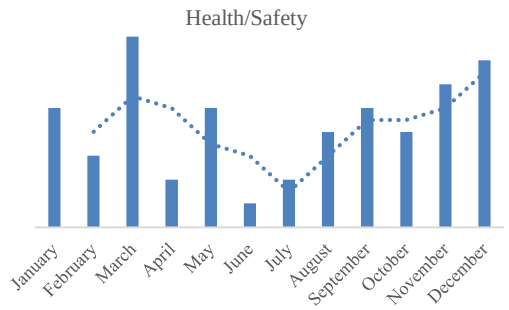


Results: Risk Matrix and Trend Analysis









Results: Prevention / Mitigation Strategies & Leading Factors

- **Prevention / Mitigation Strategies**
  - Was the road salted / graveled?
  - Drive at a speed appropriate to road and weather conditions.
  - Provide employees with an option to take defensive driving (or training to drive in winter).

- **Leading Factors**
  - Icy road conditions
  - Inadequate risk recognition
  - Factors to affect situational awareness

UNIVERSITY OF ALBERTA

Current Research: Machine Learning Algorithm

```
###PRIMARY CLASSIFICATION
X1_train, X1_test, y1_train, y1_test = train_test_split(X, y1, test_size=0.33, random_state=7)

from sklearn.pipeline import Pipeline
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.svm import LinearSVC

text_clf1 = Pipeline([('tfidf', TfidfVectorizer()),('clf', LinearSVC()),])

text_clf1.fit(X1_train, y1_train)
predictions1 = text_clf1.predict(X1_test)


from sklearn import metrics

print(metrics.confusion_matrix(y1_test,predictions1))
print(metrics.classification_report(y1_test,predictions1))
print("LinearSVC Accuracy: ", metrics.accuracy_score(y1_test,predictions1))
```

- **TfidfVectorizer**
  - Tokenizing, text pre-processing, removing stop words
- **Train a classifier**
  - Naïve Bayes, Linear SVC (SVM), Neural Networks

UNIVERSITY OF ALBERTA

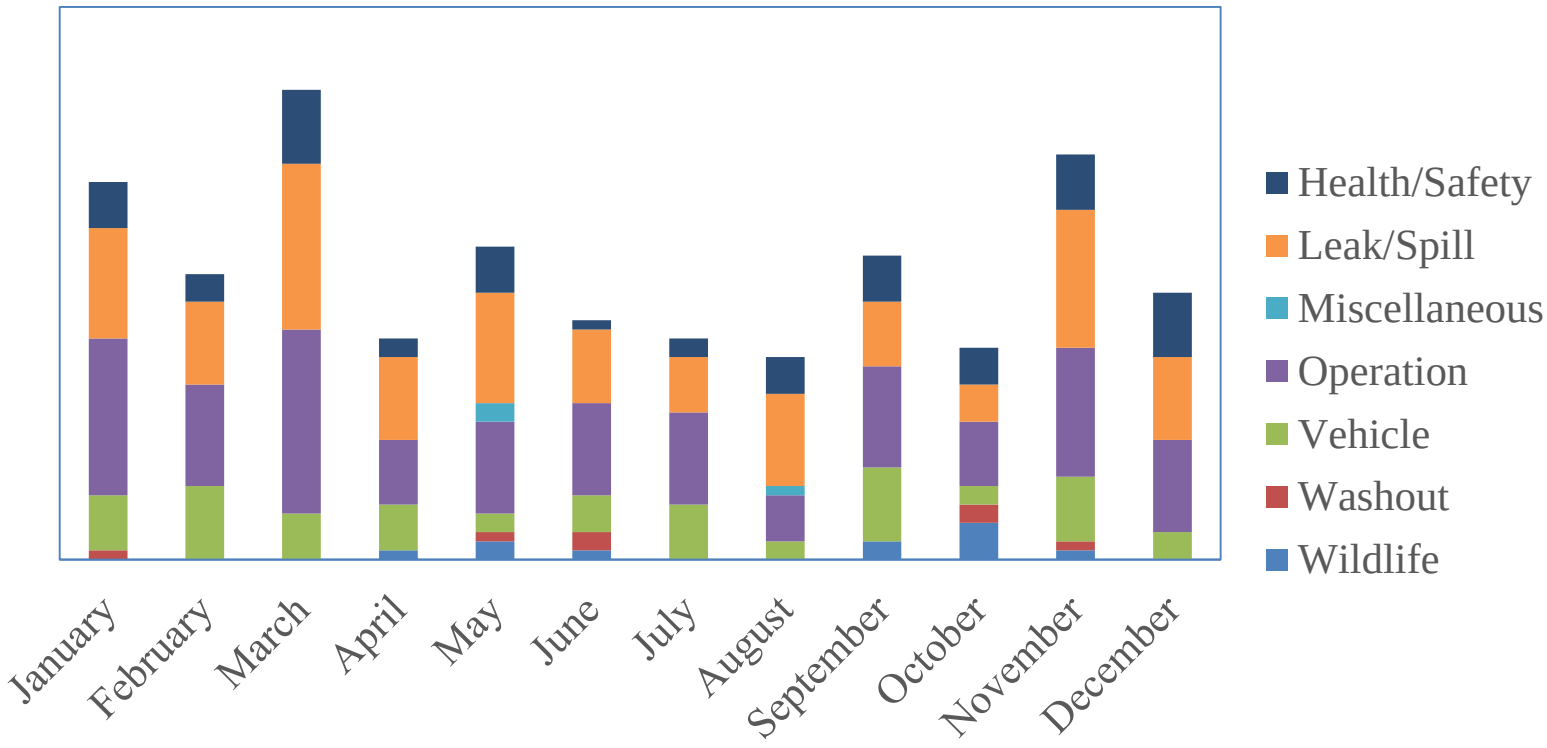
Current Results (Accuracy for different classifiers)

Classification Method	Accuracy
Support Vector Classification (SVC)	43.41%
Linear SVC	70.34%
MLP Classifier (Neural Network)	67.03%
Nearest Neighbors	60.44%
Decision Tree	59.07%
Random Forest	59.34%
Adaboost	54.12%
Multinomial Naïve Bayes	53.85%
Logistic Regression	60.16%

Current Results (Linear SVC Classification)

		Predicted						
		Comm.	**Health/S**	**Leak/Spill**	**Misc.**	**Operation**	**Uncat.**	**Vehicle**
Actual	**Comm.**	5	1	0	0	3	0	0
	Health/S	0	46	1	0	11	0	7
	Leak/Spill	0	3	19	0	8	0	0
	Misc.	1	1	0	6	3	0	3
	Oper.	2	6	3	1	132	2	12
	Uncat.	0	3	0	0	17	1	2
	Vehicle	0	6	2	0	10	0	47

Current Results (Frequency of Incidents)



Conclusion

- **Use incident database to create a customized keyword library**
- **Classify incidents using key words**
- **Assign risk score, evaluate trends, suggest tailored improvements**

Next Steps:

- **What are the biggest issues with your current systems of incident reporting?**
- **What would you need/want in an automated incident classification and reporting system?**

UNIVERSITY OF ALBERTA

Thank you