Preventing Falls: A Multimodal Causal Inference and NLP Approach

Faculty Advisor : James Butler
Liberty Mutual Team : Jesse Jacobs and Jared Haite

Mohammed Ali Benchekroun

Julien Pinede

Business Problem

6.8 million people in emergency departments annually

20% Of falls results in a **severe injury** like head trauma

Costs the US economy **144B$** each year

Liberty Mutual → Insured company

Liberty Mutual provides risk control consultation services to employers to improve safety and reduce injuries
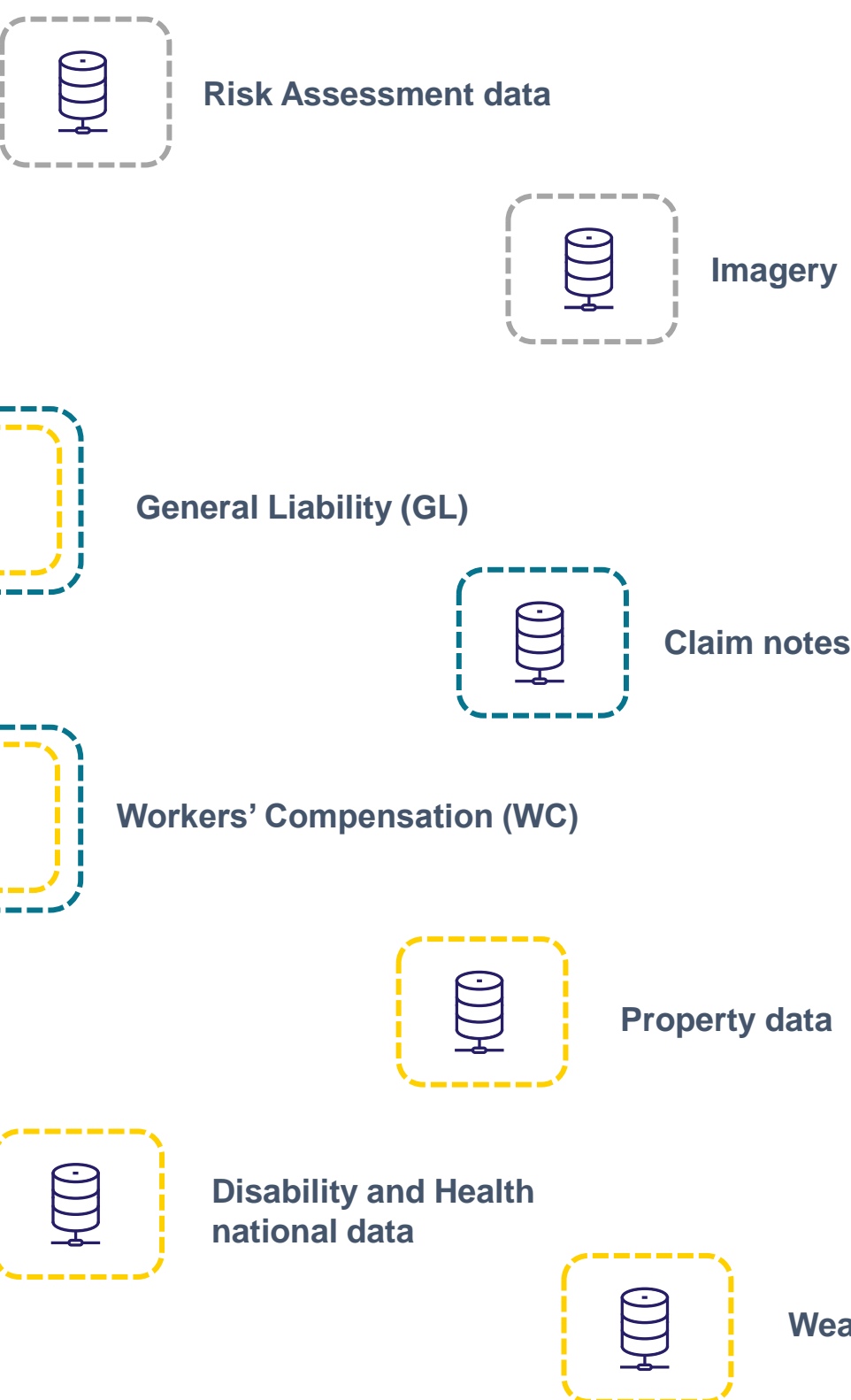
- Identify insights into the drivers and interactions that result in fall risk
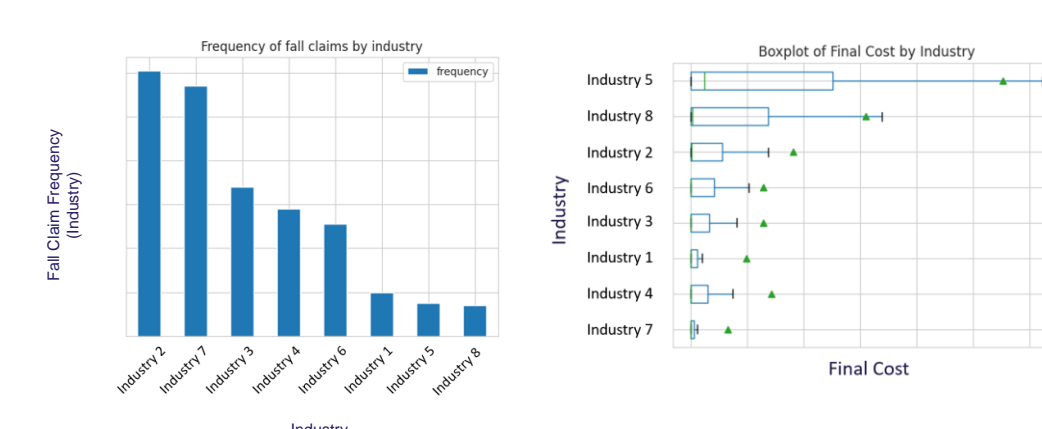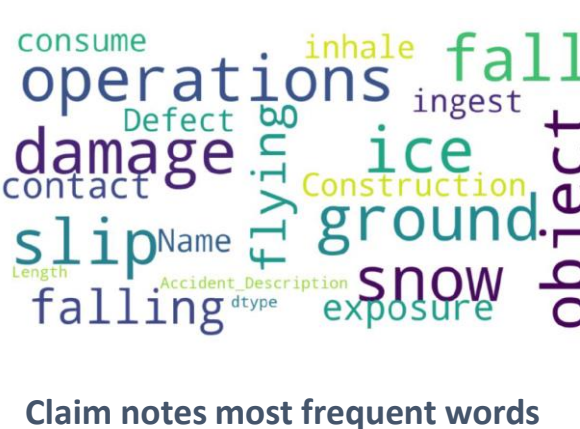- Lower loss ratio through reduced frequency and severity of falls claims
- Allow for better underwriting, pricing, and capital reserves for these claims
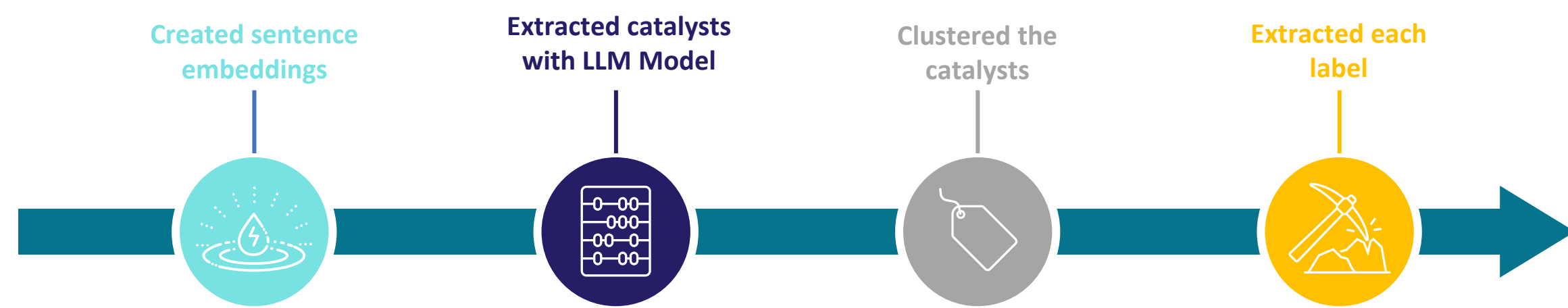
Dataset and Methodology

Datasets

- Risk Assessment data
- Imagery
- General Liability (GL)
- Claim notes
- Workers' Compensation (WC)
- Property data
- Disability and Health national data
- Weather data

Univariate Data Analysis

Claim notes most frequent words

Outcome
- Identified data distribution for processing (outliers, imbalance,…)
- Performed T test and Kruskal-Wallis test to assess feature distribution for Falls vs non Falls claims

Fall Claims Catalyst – NLP Pipeline

Created sentence embeddings → Extracted catalysts with LLM Model → Clustered the catalysts → Extracted each label

- Identifed the catalysts of fall accidents from unstructured text data
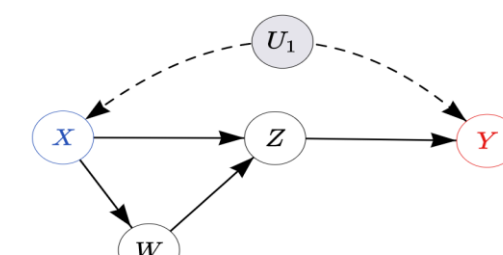- Performed feature creation for multivariate analysis

Falls Risk Modelling

1. Random Forest
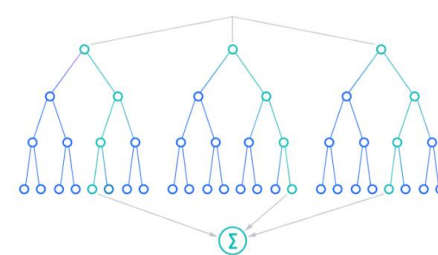
Finding **correlations** in large dataset and make prediction.

Outperformed Logistic regression for each model

2. Causal Inference

Infering **causality** rather than just correlation.

We tried Propensity Score Matching and Double Machine Learning

- Developed an efficient way to infer risk scores
- Identified features with high predictive power
- Assessed added value of joining dataset from a modelling point of view
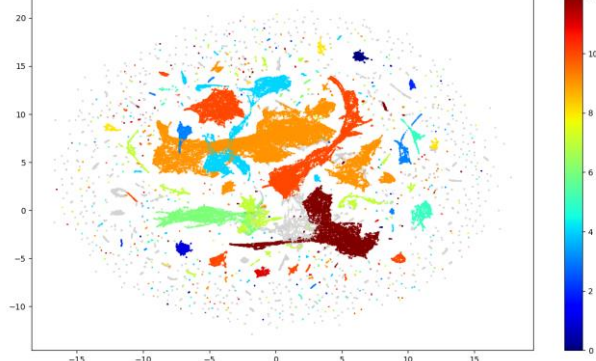
Results

Natural Language Processing (NLP)

Clusters for WC Accident Description

| Cluster	Label*
1	Choked_langoustine
2	Walking_lobby
3	Sleepwalker_slipped_pool
4	Drink_bar

*Examples and figures are for illustrative purposes to preserve confidentiality

Clusters for GL Accident Description (All Claims)

For more specifics insights, we clustered the catalysts by industry. We obtained **interpretable labels** that provide context on the Falls.

Random Forest
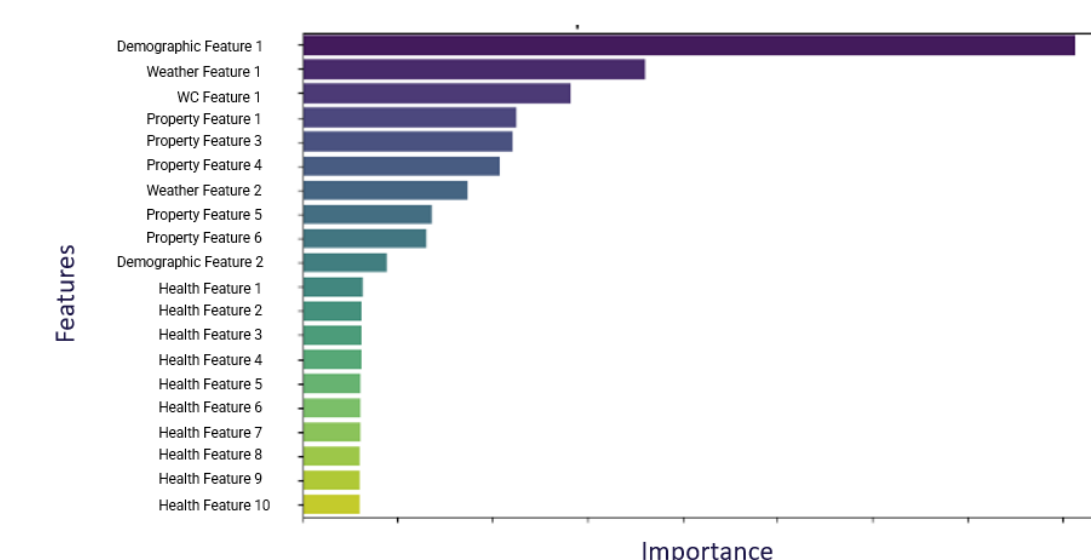
Metric	Baseline Model	August Model
Sensitivity	0.14	**0.85**
Specificity	0.92	**0.97**
F-measure	0.18	**0.81**

Improved the Baseline model (Random forest trained on GL data) by **450%**

Causal Inference

We computed the **Average Treatment Effect** with and without confounders to compute the bias induced by confounders. We also computed a treatment effect with a **random treatment** to assess the causal relationship significance. We obtain the **magnitude and direction** of the causal effect of each treatment on Falls.

Causality	Treatment Effect (No confounders)	Treatment Effect (Confounders)	Interpretation
Age on Fall Claim	0.014	0.01	If you are 1 year older, it increases (on average) your risk to fall by 1%
Loss amount on Fall Claim	937	526	If you have a fall claim, the loss amount (on average) is $526 more compared to a non-fall claim

*Figures have been modified for illustrative purposes to preserve confidentiality

Business Impact

Deploy automated fall risk scores at scale based on multimodal data

Target at-risk clients with interventions that are more accurately tied to their falls catalysts

More successfully underwrite customer's policies based on their risk for falls claims

$20M In estimated annual business value when deployed

- Improve the effectiveness of Liberty's risk consultations for falls risk
- Improve customer retention based on our successful interactions and interventions
- Reduce the amount of fall claims through a more efficient prevention process

Significant parts of the results (e.g. features) were removed for confidentiality purposes