Actuaries Institute.

AI powered claims models

Using Large Language Models to predict injury compensation claim outcomes

Data Science Masterclass

Presented by Michael McLean and Nikolay Nikolaev

Important notice for all participants

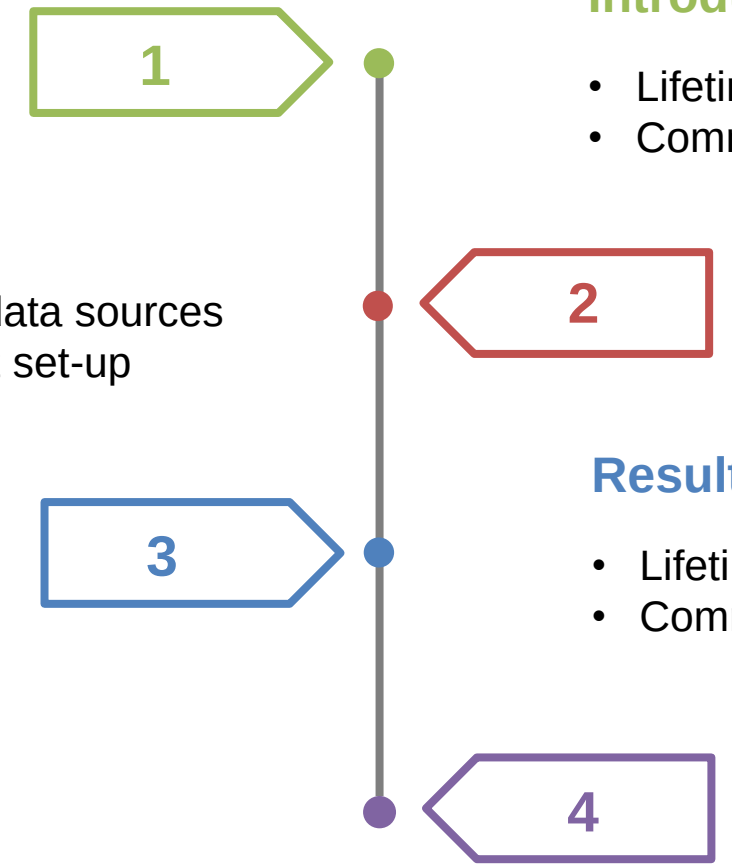
This meeting is being conducted in accordance with Institute's Code of Conduct and attended by members in their professional capacity.

It is acknowledged that professional members in their employed capacity, may be active market participants in their respective industries who may compete with each other as defined by competition law.

Participants are, therefore, reminded that in accordance with their competition law compliance obligations they should not:

- discuss any matter that may be perceived as being cooperation by competitors in a market to influence that market;
- discuss any matters that could be regarded as fixing, maintaining or controlling prices, allocation of customers or territories, coordinating bids and/or restricting output or acquisitions in any circumstances;
- share commercially sensitive information relating to their employer; or
- share information for an anti-competitive purpose.

Table of Contents



Introduction

- Lifetime Cost of Claim (LCC) modelling
- Common Law Lodgement modelling

Methodology

- Claim information & data sources
- Modelling experiment set-up

Results

- Lifetime Cost of Claim (LCC) modelling
- Common Law Lodgement modelling

Outlook

- Future work
- Conclusion

Actuaries Institute.

Introduction

Lifetime Cost of Claim (LCC) model for accident compensation claims

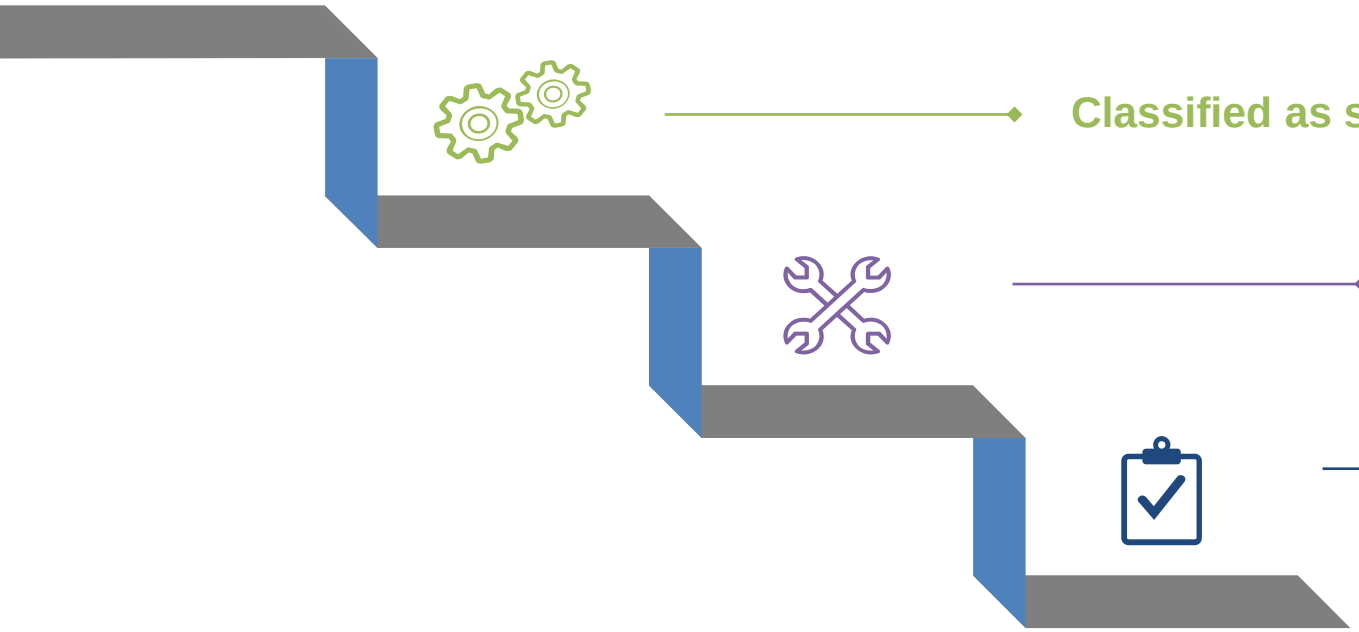
- Case reserves
- Claim management
- Strategic intervention

Our research

- Can unstructured text add value when modelling LCC?
- Test on one component of the LCC model we built for a large Scheme

Actuaries Institute.

Lifetime Cost of Claim modelling



All claims start as No-Fault claims

Classified as severe or non-severe

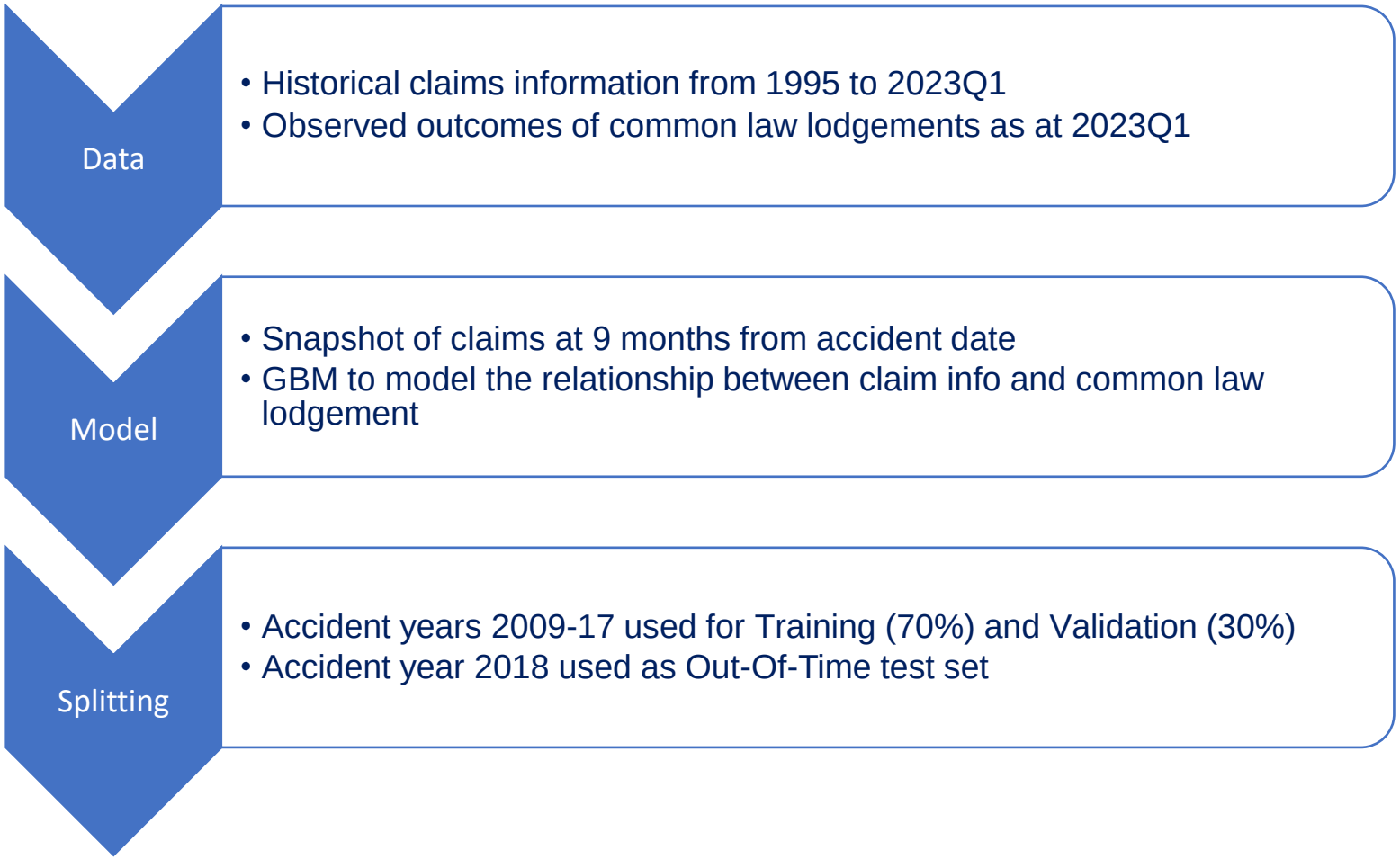
No-Fault claims can become a Common Law claim

Common Law claims

- Model probability of a Common Law lodgement
- Size given lodgement

Actuaries Institute.

Predicting the probability of CL lodgement



Actuaries Institute.

Data

Claims header file
All info relating to claimant, injury

Payments
Transactions payment data

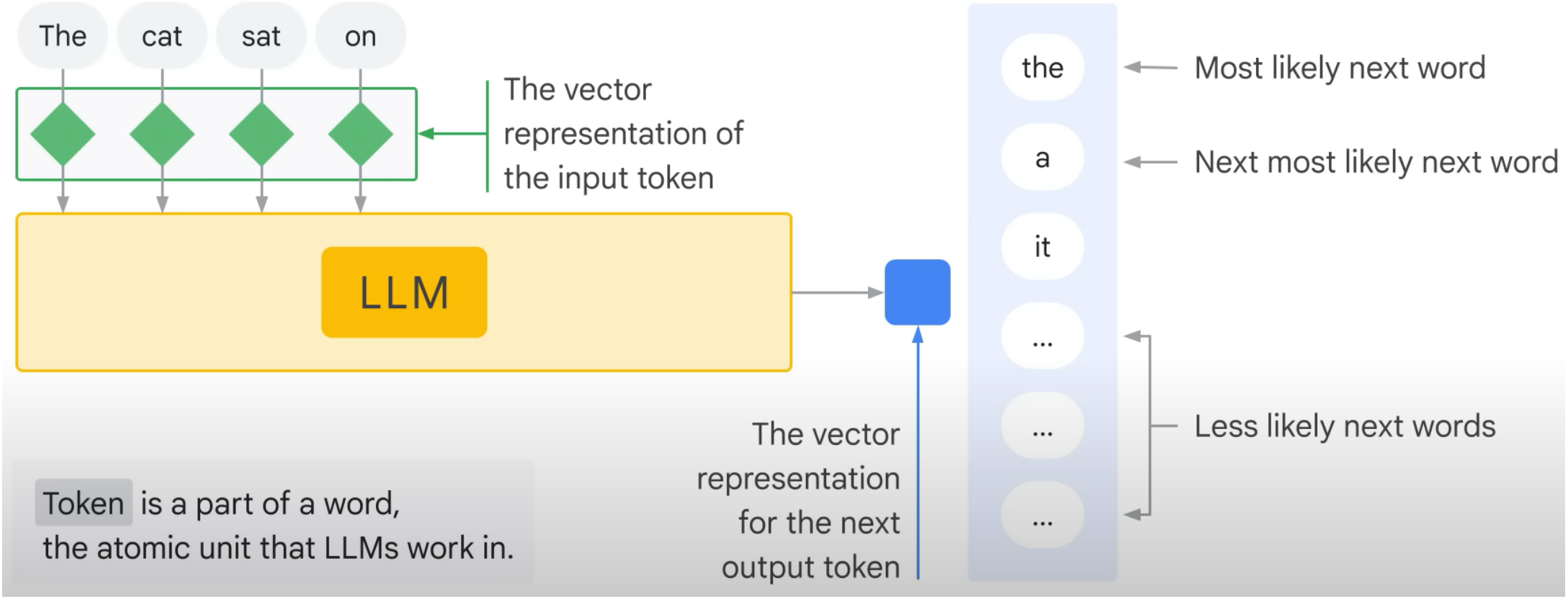
Road safety data
Info on crash, vehicles, drivers involved, drugs, alcohol

Free form text data
Five distinct types (e.g. case notes, external documents, phone calls)

Census data
SEIFA index, vehicle density, remoteness

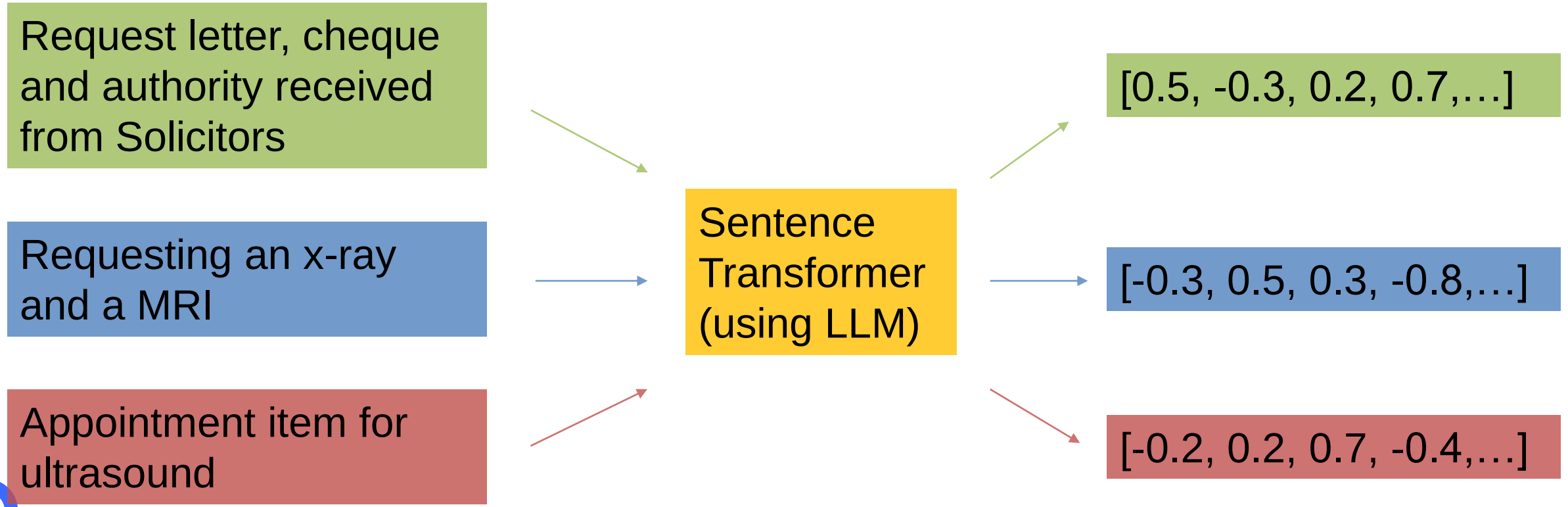
Actuaries Institute.

Large Language Models



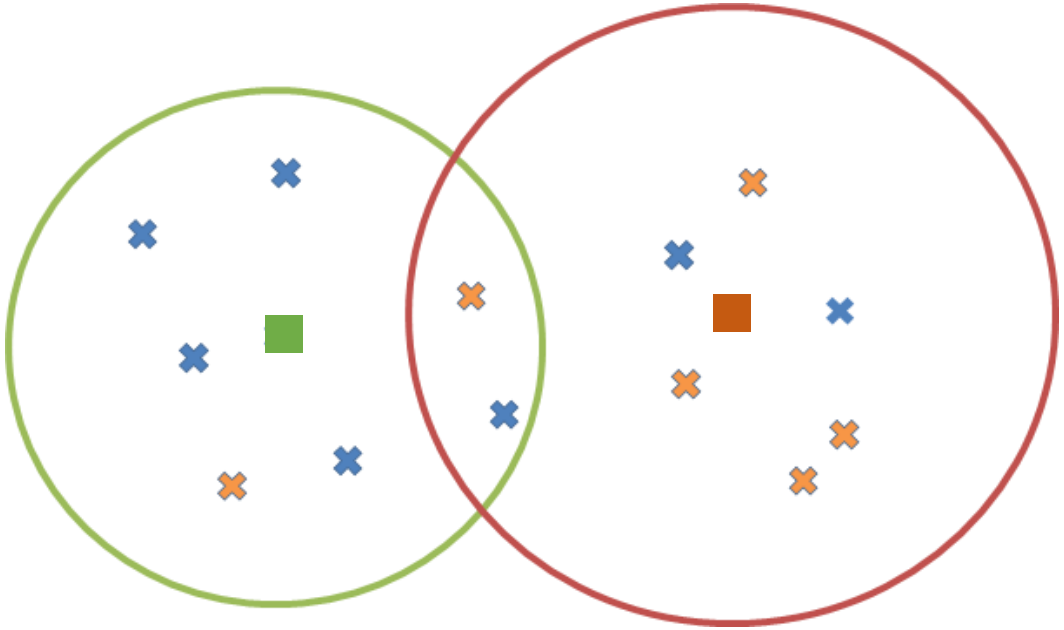
Actuaries Institute.

Sentence embeddings



Actuaries Institute.

Use LLM for finding similar claims

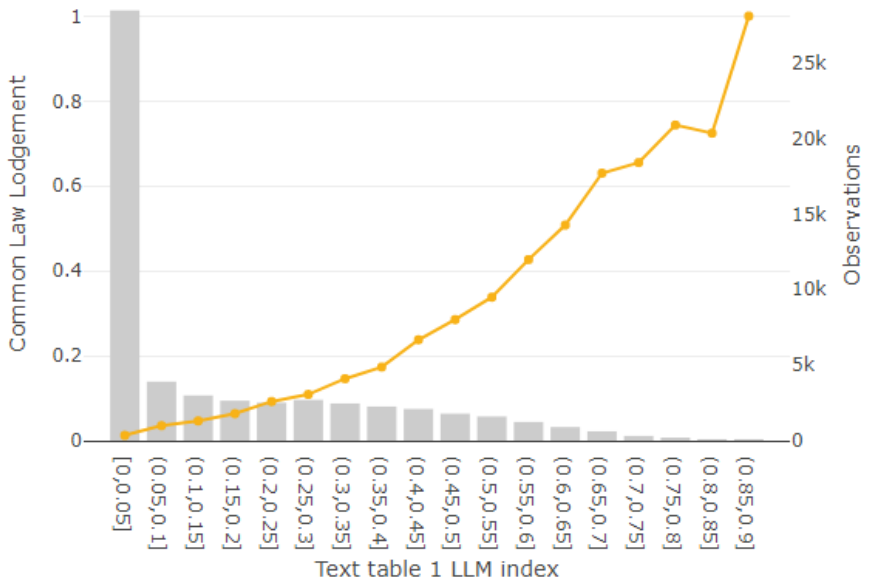


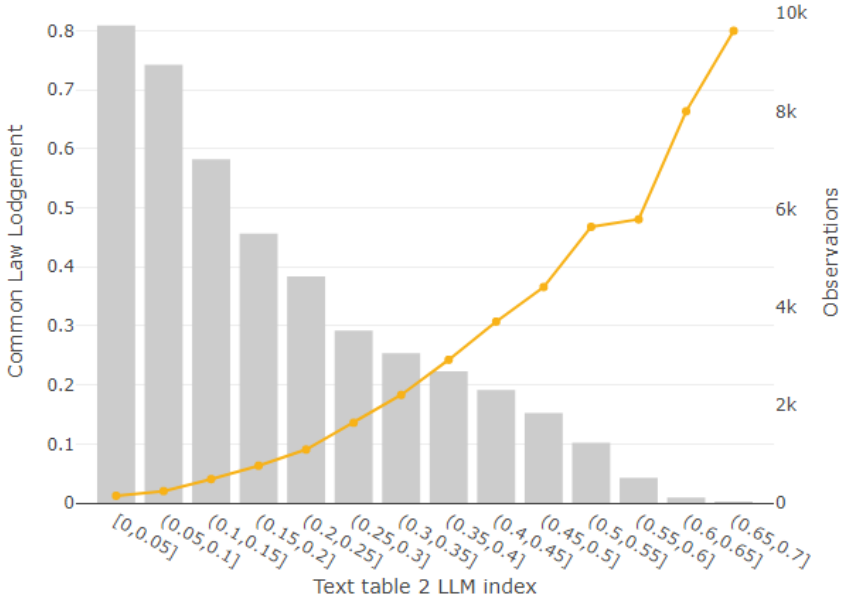
Intuition: Can historical claims with similar claim descriptions help predict the probability of a common law claim?

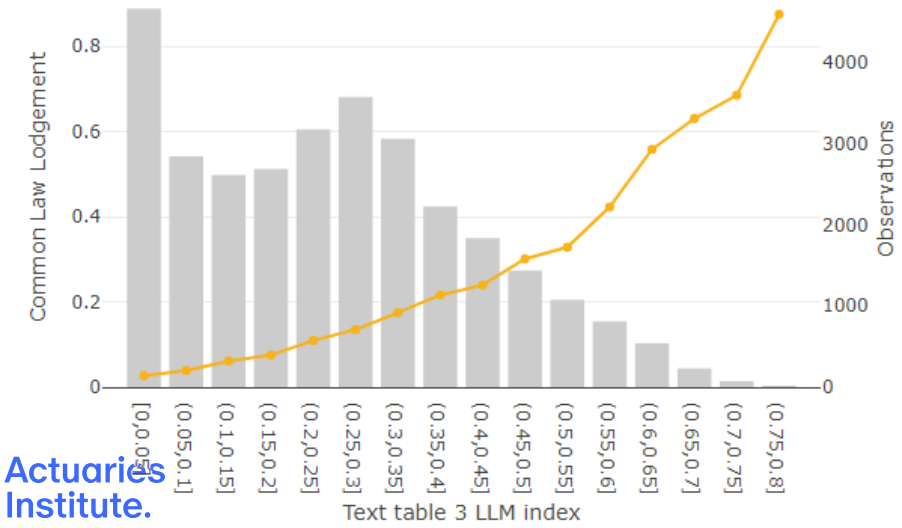
- Embed claim texts into numerical vectors that capture the meaning using open source LLMs (mpnet, gte)
- Average embeddings per claim and text type to represent claim as a whole
- For each claim
  - Find other claims that have similar text descriptions
  - Derive a score based on how many neighbours lodged a CL
- Use the derived score as a predictor in the modelling

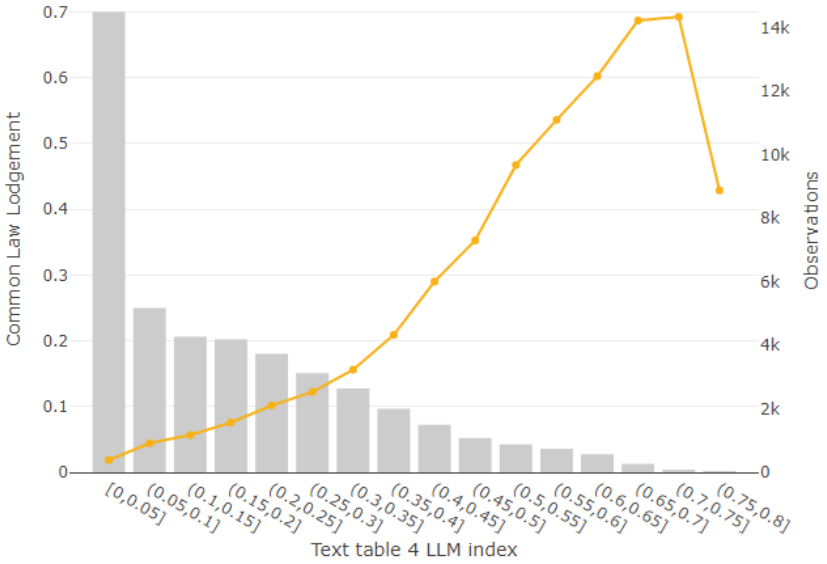
Actuaries Institute.

LLM based text score









Text-based neighbours strongly predict CL probability

Actuaries Institute.

Experimental setup

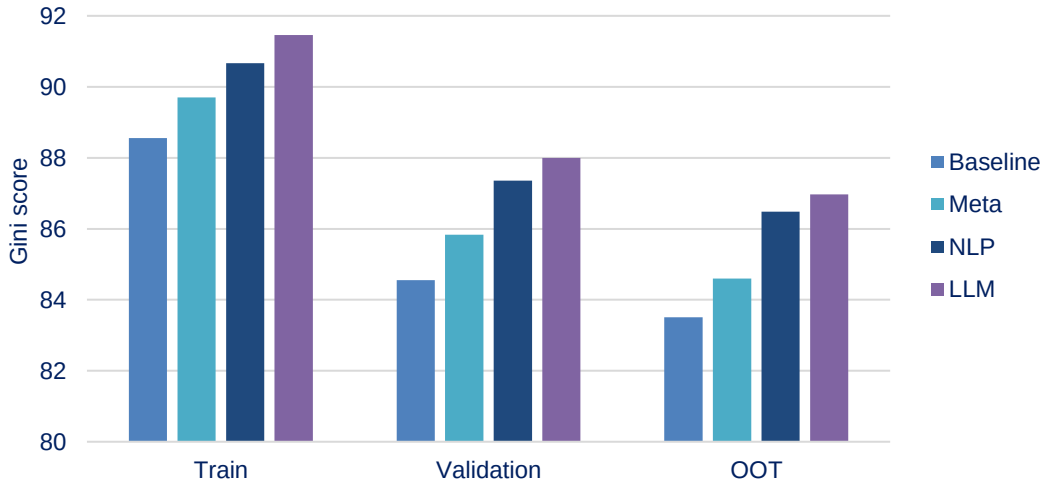
	Features			
Models	*Structured data*	*Text meta-data*	*NLP*	*Text data embeddings*
#1: Baseline	■			
#2: Meta	■	■		
#3: NLP	■	■	■	
#4: LLM	■	■	■	■

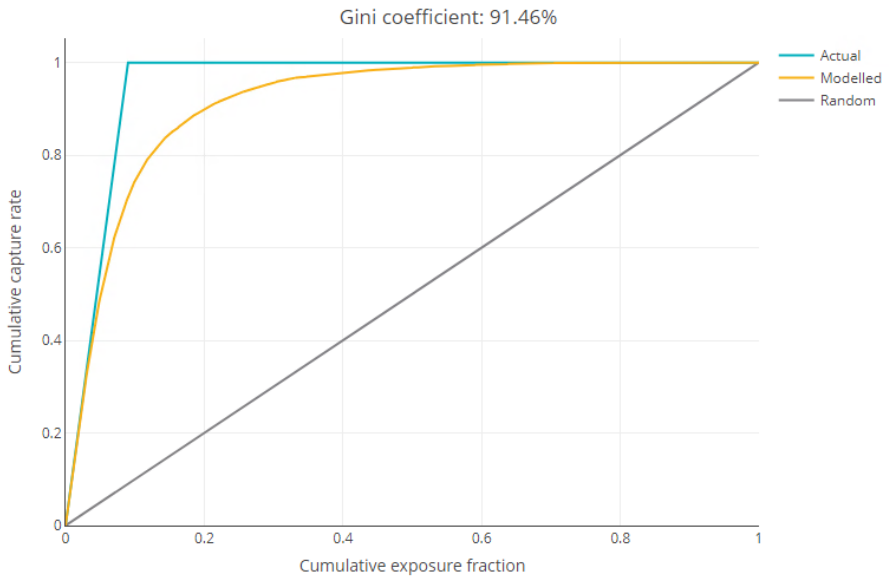
Evaluation: Gini coefficient

Actuaries Institute.

Results

- Model performance improves with each new set of features
- Results demonstrate the strong predictive power of claim text information
- Using LLM features results in the strongest model



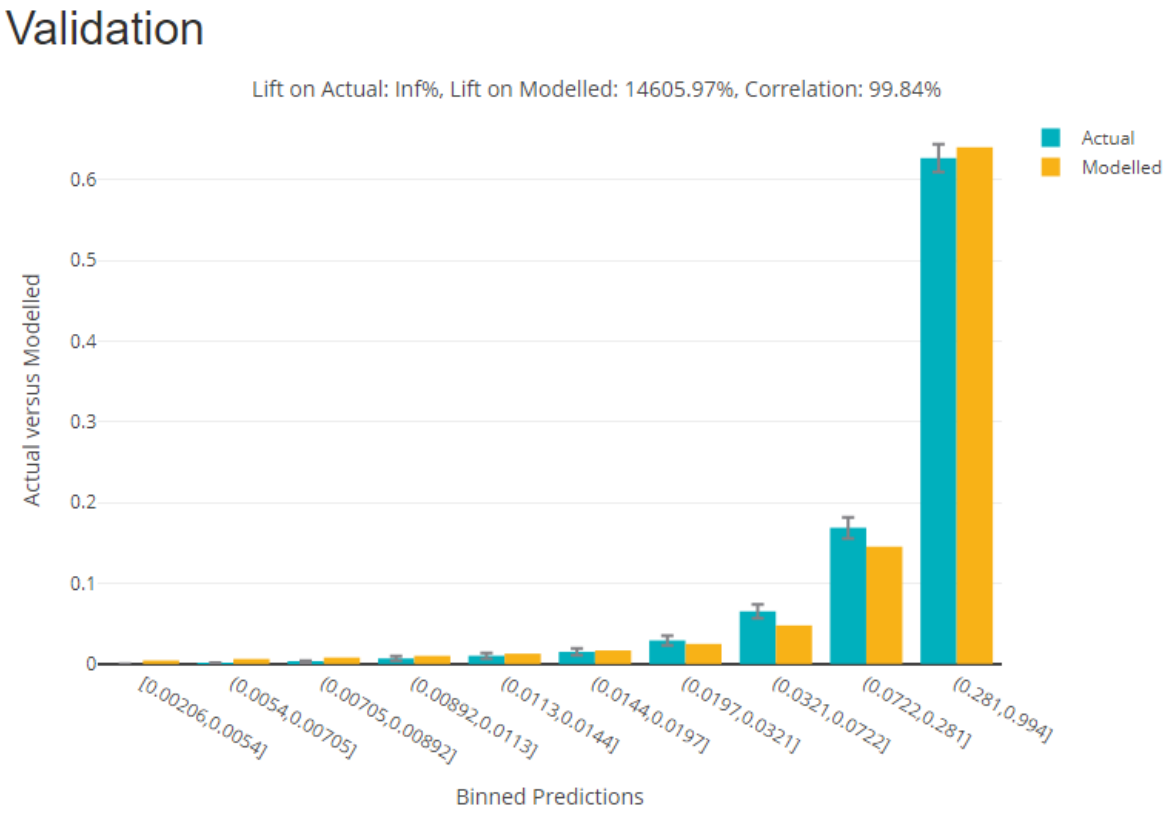


	% Common Law Lodgements Identified				
Validation	% of claims	Baseline	Meta	NLP	LLM
	2%	20%	20%	20%	20%
	5%	42%	43%	45%	45%
	10%	64%	64%	68%	68%
	25%	87%	89%	89%	91%

Actuari
Institut

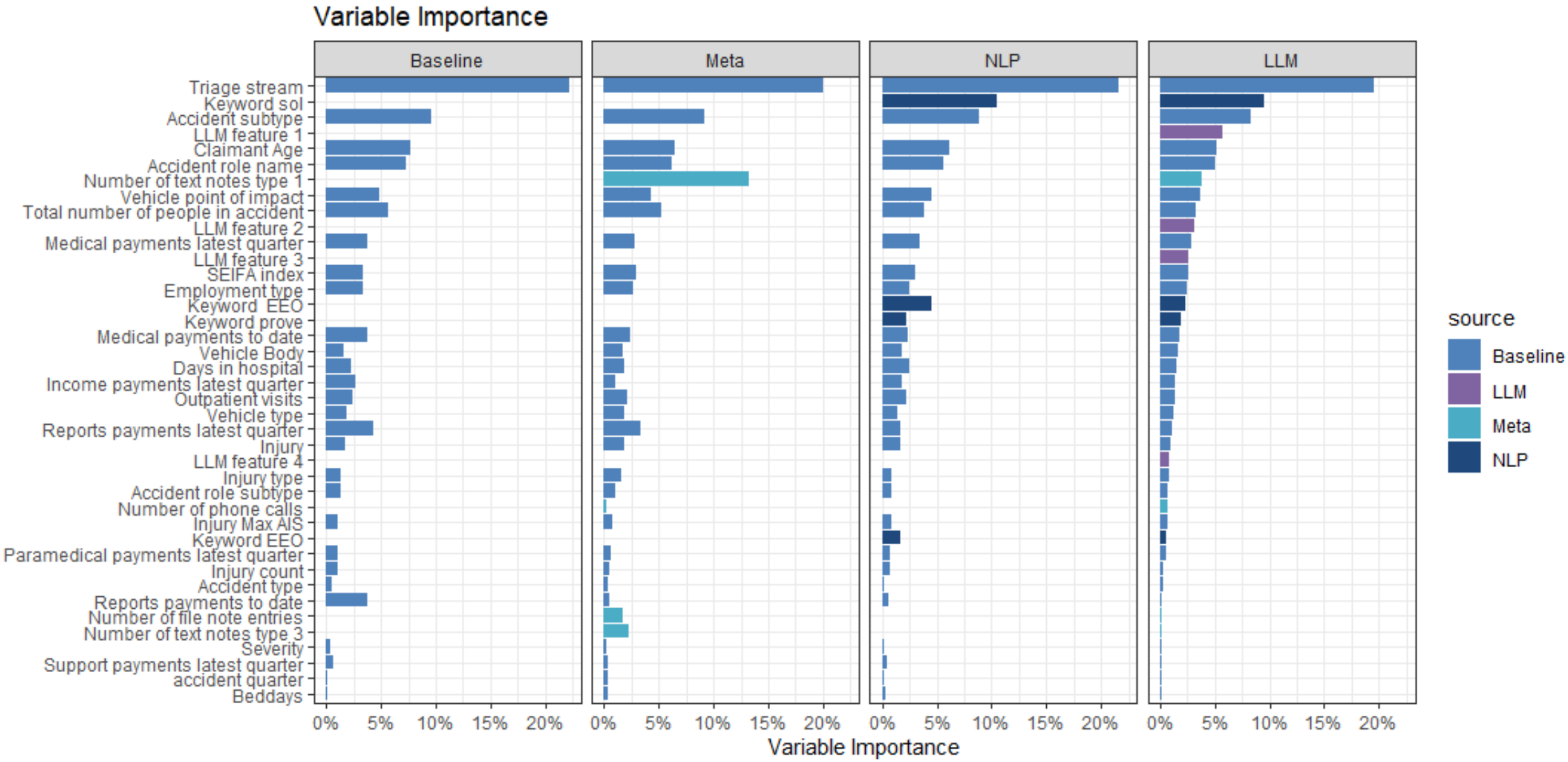
Risk differentiation

- LLM based model achieves the best results and validates well on unseen data
- Well calibrated predictions for the probability of a claim to become a common law claim
- Strong risk differentiation achieving a high model lift



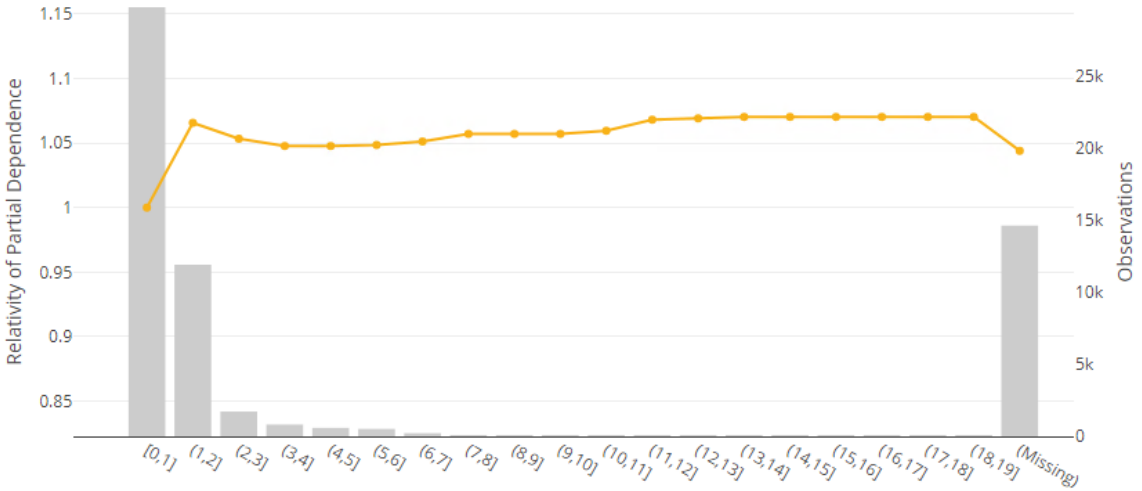
Actuaries Institute.

Feature Importance

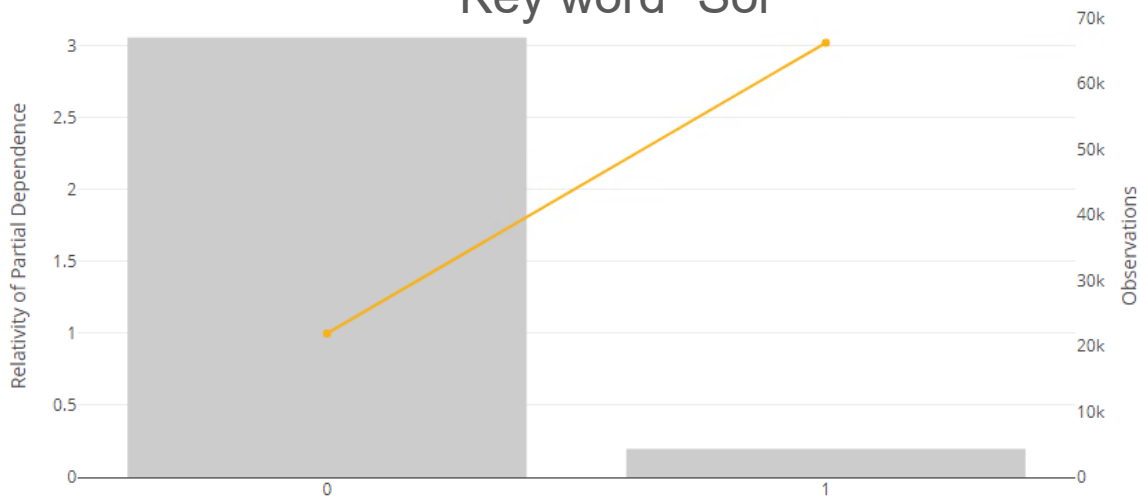


Partial Dependence

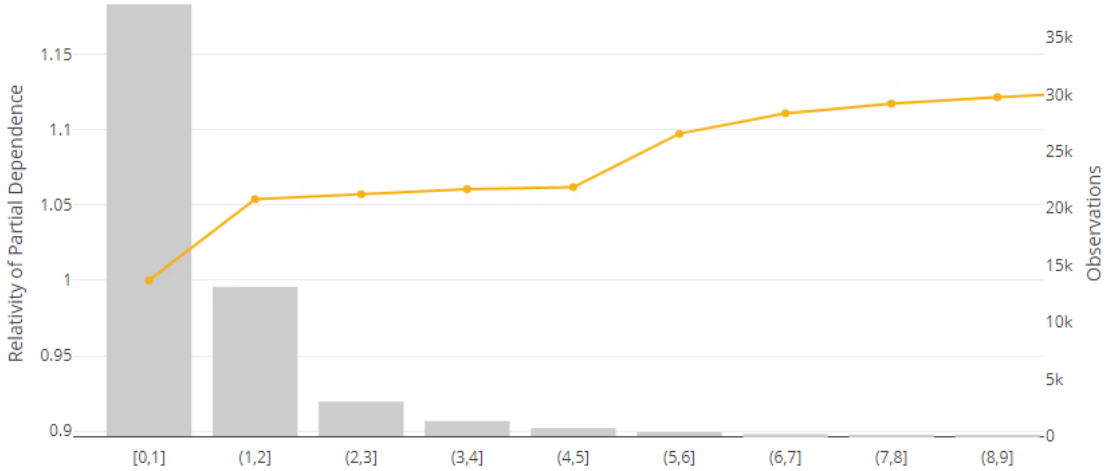
LLM Feature 1



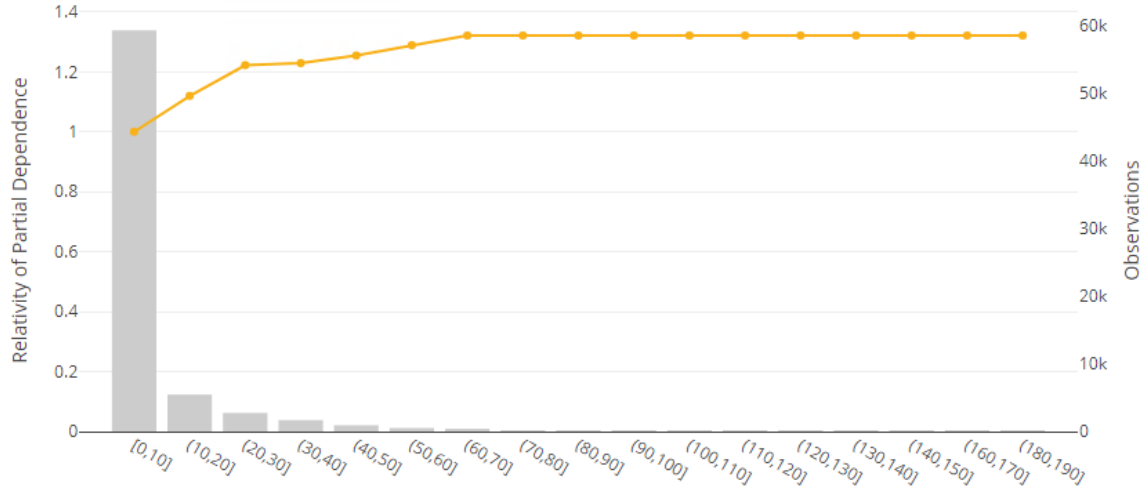
Key word “Sol”



LLM Feature 2

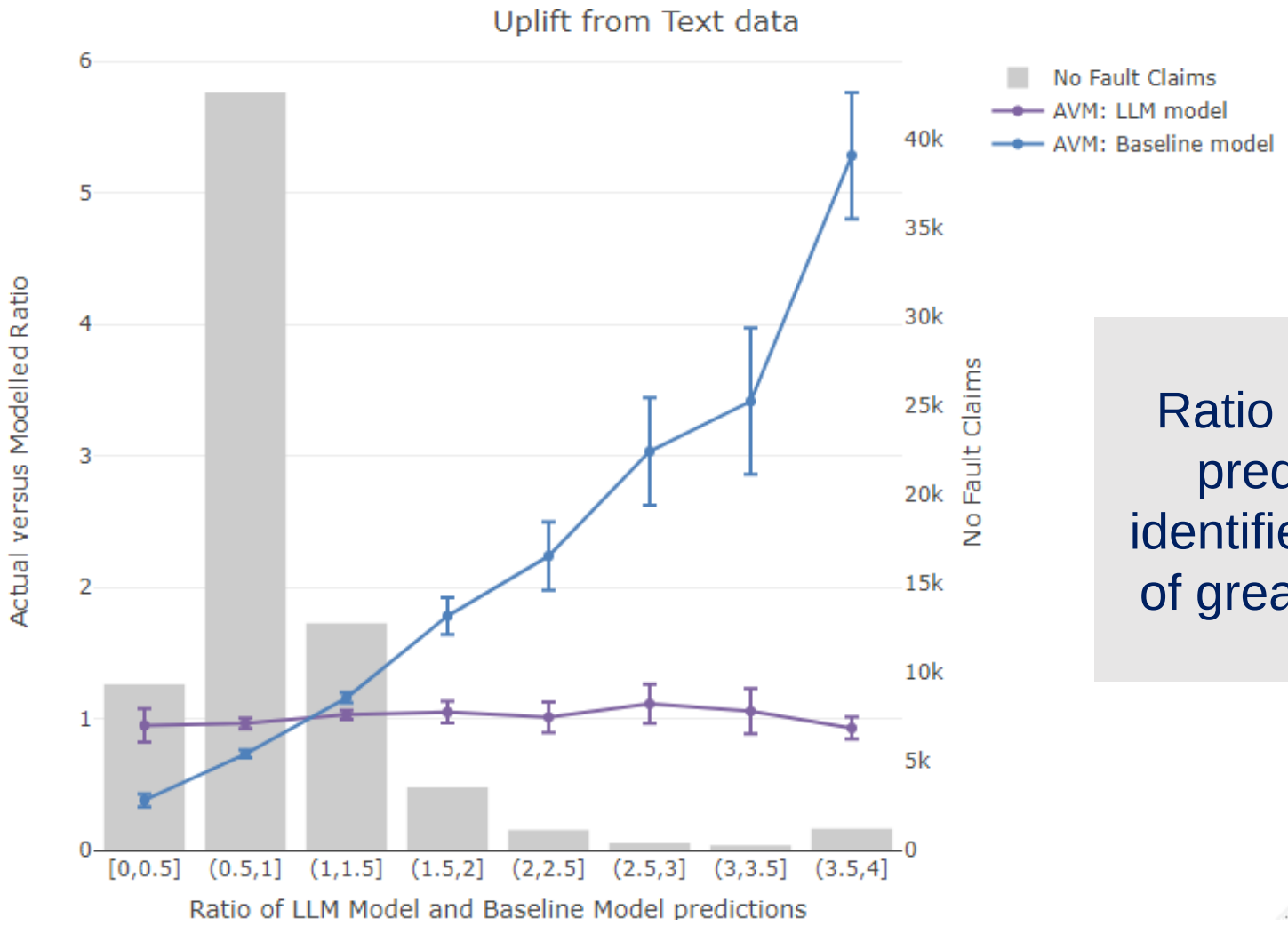


Count of external document notes



institute.

Model Comparison



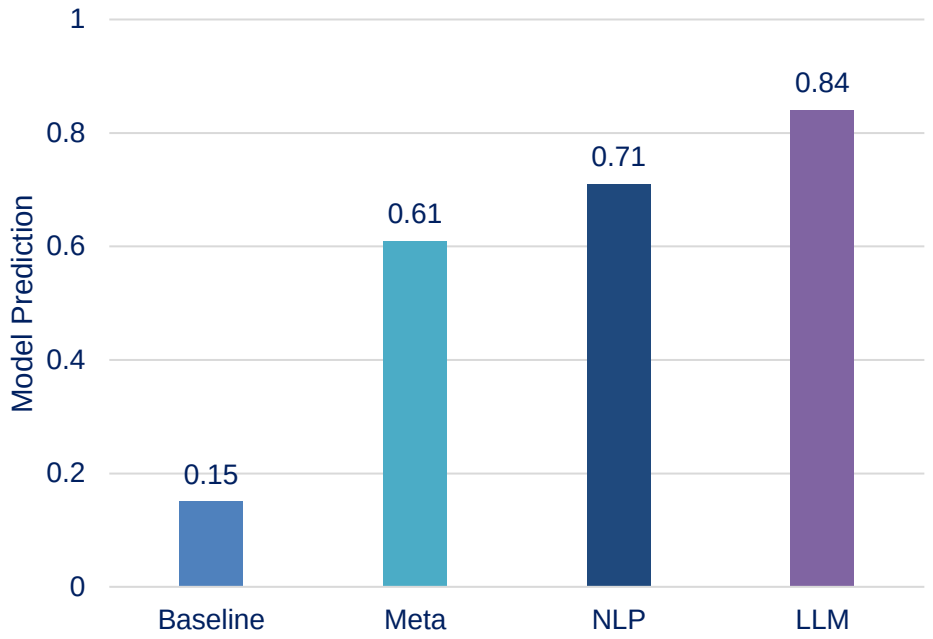
Ratio of model predictions identifies regions of greatest uplift

Actuaries Institute.

Example – Claim 1

Outcome: Common Law Lodgement

Injury detail	Fractures – Limb
Role	Passenger/Pillion
Total sum to date	$24k
Days in hospital	9
Age	19
Text records counts	0-20
LLM 10-NN scores	0.7
Common keywords	TAXI, Support, Form, General, approval

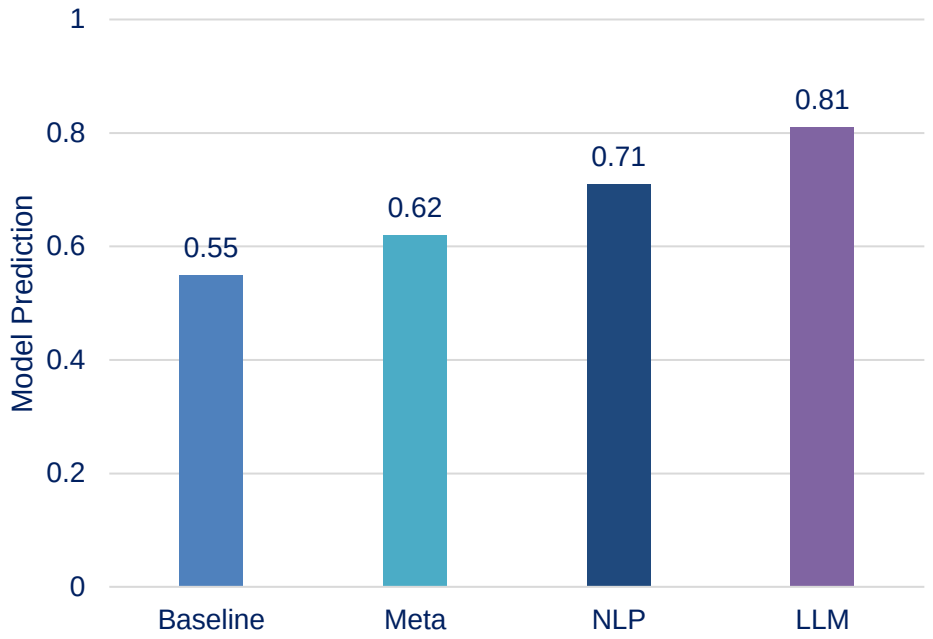


Actuaries Institute.

Example – Claim 2

Outcome: Common Law Lodgement

Injury detail	Brain Injury (Mild) / Head Injury (Ill defined)
Role	Bicyclist
Total sum to date	$109k
Days in hospital	23
Age	49
Text records counts	70+
LLM 10-NN scores	0.9
Common keywords	Prove, Uploaded, Received, Correspondence, Benefits, Care, Dr, Report, LOE, Form, Rehabilitation, Support, Treating, Certificate, Services, RTW, letter, Practitioner, employer, Income

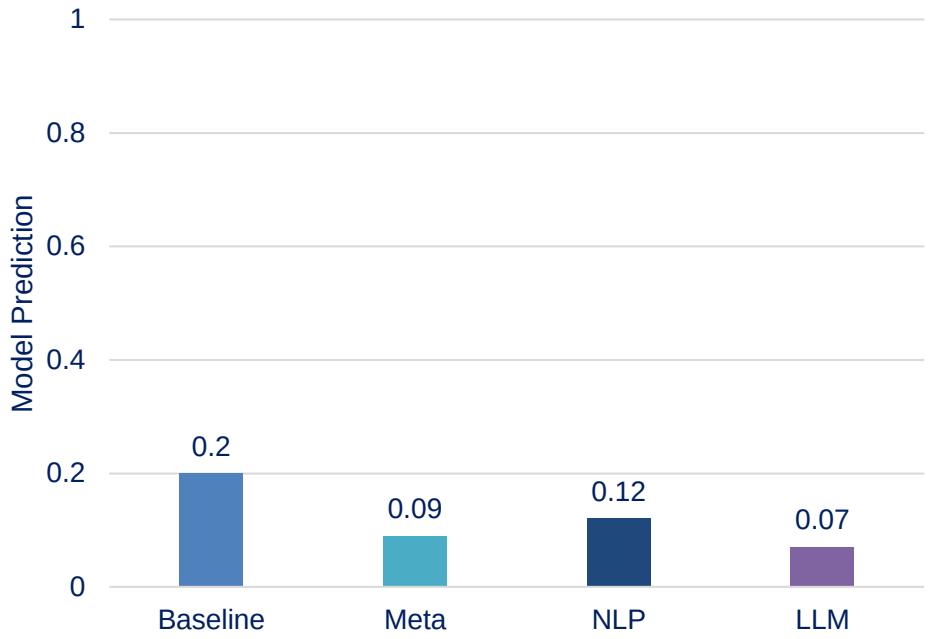


Actuaries Institute.

Example – Claim 3

Outcome: No Common Law Lodgement

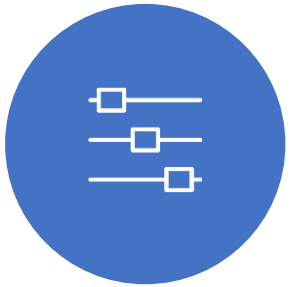
Injury detail	Fractures – Limb
Role	Bicyclist
Total sum to date	$83k
Days in hospital	11
Age	51
Text records counts	0-7
LLM 10-NN scores	0.1
NLP	Prove
Common keywords	Police, Report, Incident, Confidential



Actuaries Institute.

Future work

Utilize more unstructured text data – documents, e-forms, medical reports etc.



Fine-tuning the large language model or method of aggregation of the embeddings

Use commercial private versions of more powerful models (e.g. chat GPT) instead of the smaller open source LLMs

Apply to other components of lifetime cost of claims model (e.g. cost of No-Fault claims)

Actuaries Institute.

Conclusions

- Unstructured text data significantly improves compensation claims predictive model performance
- Schemes, insurers and claims service providers have a valuable asset which can be utilized at scale with potential significant improvements in claim management and reserving
- Large Language Models are a powerful tool for extracting signal out of unstructured data
- LLM field is emerging and improving rapidly - expect better results in the future with advances in technology

Q&A

Join at

slido.com

#2554 956

Feedback

Join at

slido.com

#2554 956